Electronic Acknowledgement Receipt

EFS ID:	26656814
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Customer Number:	91337
Filer:	Thomas Stephen Chlebeck/Darlene Morin
Filer Authorized By:	Thomas Stephen Chlebeck
Attorney Docket Number:	N0484.71090US04
Receipt Date:	16-AUG-2016
Filing Date:	
Time Stamp:	15:58:13
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$1600
RAM confirmation Number	081716INTEFSW15590000
Deposit Account	8543
Authorized User	Wolf Greenfield

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal of New Application	N048471090US04-TRN-TSC.pdf	31876 05b89e5cd8914913b3bbc7d64eb28dbb84bb6f22	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-TSC.pdf	30112 d4eb56f1ac992215efbd90195490e6c565a57f16	no	1
Warnings:					
Information:					
3	Oath or Declaration filed	N048471090US04-INVDEC-TSC.pdf	2481504 1364be4c8e8a3556124c67d7edcbfc0e2190aa90	no	2
Warnings:					
Information:					
4	Application Data Sheet	N048471090US04-ADS-TSC.pdf	29969 260319af839cf422c1c29d8939394aa5750692f0	no	7
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
5		N048471090US04-CONAPP-TSC.pdf	208047 aa632b5b0985713b837daa8b95a28a25da07d138	yes	67

Multipart Description/PDF files in .zip description

Document Description	Start	End
Specification	1	62
Claims	63	66

	Abstract		67	67	
Warnings:					
Information:					
6	Drawings-other than black and white line drawings	N048471090US04-NDRW-TSC.pdf	2095041 e2b39339ff768cac3af27c7e45141c7e6d132fba	no	10
Warnings:					
Information:					
7	Fee Worksheet (SB06)	fee-info.pdf	35306 97b953f57fea795c8aacaa685d3605b2e55caef4	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	4911855		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	William
Middle Name::	F.
Family Name::	Ganong
Name Suffix::	III
City of Residence::	Brookline
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	118 Sumner Road
City of mailing address::	Brookline
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02146

Inventor Number::	2
Name Prefix::	Mr.
Given Name::	Paul
Middle Name::	Adrian
Family Name::	Van Mulbregt
City of Residence::	Wayland
State or Province of Residence::	MA
Country of Residence::	US

Street of mailing address::	30 Bow Road
City of mailing address::	Wayland
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	01778

Inventor Number::	3
Name Prefix::	Mr.
Given Name::	Vladimir
Family Name::	Sejnoha
City of Residence::	Lexington
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	10 Flintlock Road
City of mailing address::	Lexington
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02420

Inventor Number::	4
Given Name::	Glen
Middle Name::	Edward
Family Name::	Wilson
City of Residence::	Maynard
State or Province of Residence::	MA

Country of Residence::	US
Street of mailing address::	13 Sherman Street
City of mailing address::	Maynard
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	01754

Correspondence Information

Correspondence Customer Number::	91337

Application Information

Application Type::	Regular
Subject Matter::	Utility
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Attorney Docket Number::	N0484.71090US04
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets::	10
Small Entity?::	No
Petition included?::	No
Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy	No

Order pursuant to 37 CFR 5.2::

This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013::	No

Representative Information

Representative Customer Number::	91337

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::
This Application	Continuation of	13/795933	03/12/13

Foreign Priority Information

Applicant Information

Applicant Number::	1
Applicant Type::	Assignee
Organization Name::	Nuance Communications, Inc.
Street of mailing address::	One Wayside Road
City of mailing address::	Burlington
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	01803

Assignee Information Including Non-Applicant Assignee Information

Assignee Number::	1
Organization Name::	Nuance Communications, Inc.
Street of mailing address::	One Wayside Road
City of mailing address::	Burlington
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	01803

Authorization or Opt-Out of Authorization to Permit Access

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

Signature:

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Thomas S. Chlebeck/	Date (YYYY-MM-DD)	2016-08-16
Name	Thomas S. Chlebeck	Registration Number	70,616

CLAIMS

1. A mobile device comprising:

at least one input configured to receive acoustic input from an environment of the mobile device while the mobile device is operating in a low power mode; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command, the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command,

wherein the mobile device is configured to initiate responding to the voice command when the at least one processing stage determines that the acoustic input includes a voice command.

2. The mobile device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage to evaluate whether the acoustic input includes a voice command and a second processor configured to perform at least one second processing stage to evaluate whether the acoustic input includes a voice command.

3. The mobile device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

4. The mobile device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

5. The mobile device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

Petitioner Exhibit 1004, Page 11 of 1052

6. The mobile device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

7. The mobile device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

8. The mobile device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

9. The mobile device of claim 7, wherein the mobile device is further configured to provide, prior to activating the second processor or turning on a display of the mobile device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. The mobile device of claim 3, wherein the mobile device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

11. The mobile device of claim 1, wherein at least one task specified by the voice command is performed with the mobile device remaining in the idle mode.

12. The mobile device of claim 1, wherein the mobile device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. The mobile device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device.

Petitioner Exhibit 1004, Page 12 of 1052

14. The mobile device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device.

15. At least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a mobile device to receive acoustic input from the environment of the mobile device while the mobile device is operating in a low power mode, the method comprising:

performing, while in the low power mode, at least one processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command, the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command; and

initiating responding to the voice command when the at least one processing stage determines that the acoustic input includes a voice command.

16. The at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

17. The at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

18. The at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. The at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the mobile device remaining in the idle mode.

Petitioner Exhibit 1004, Page 13 of 1052

20. The at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device.

Petitioner Exhibit 1004, Page 14 of 1052

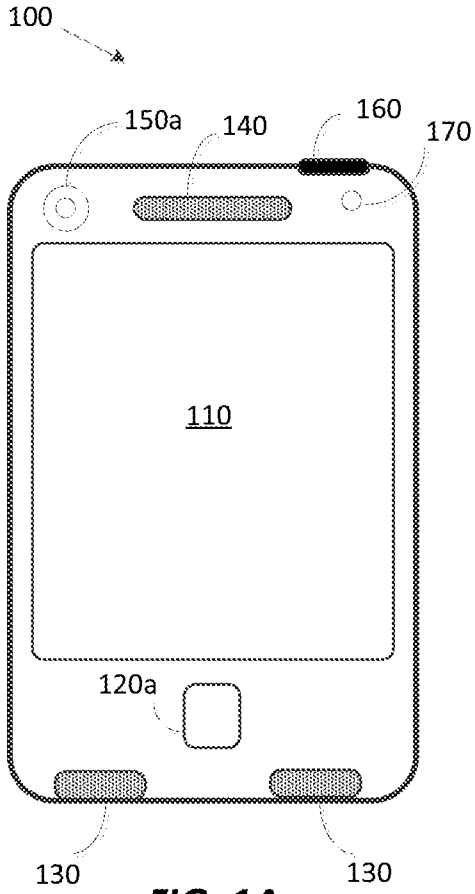


FIG. 1A

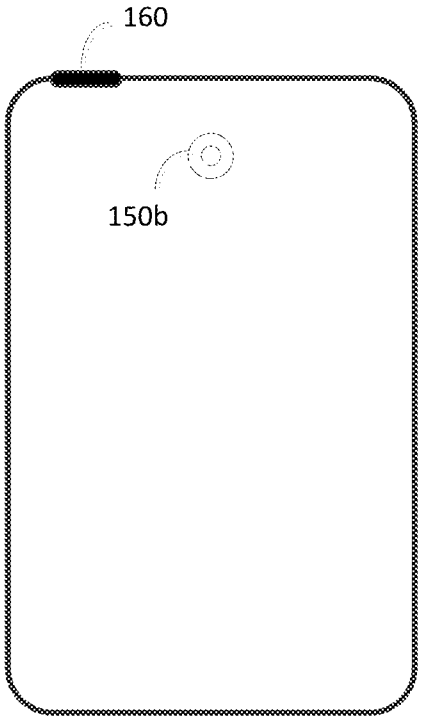


FIG. 1B

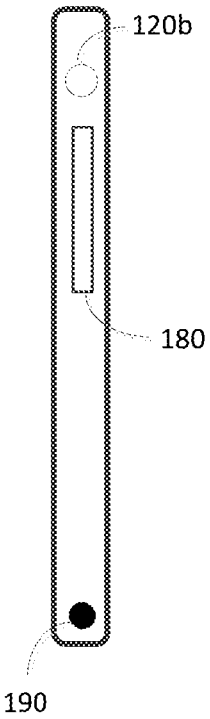


FIG. 1C

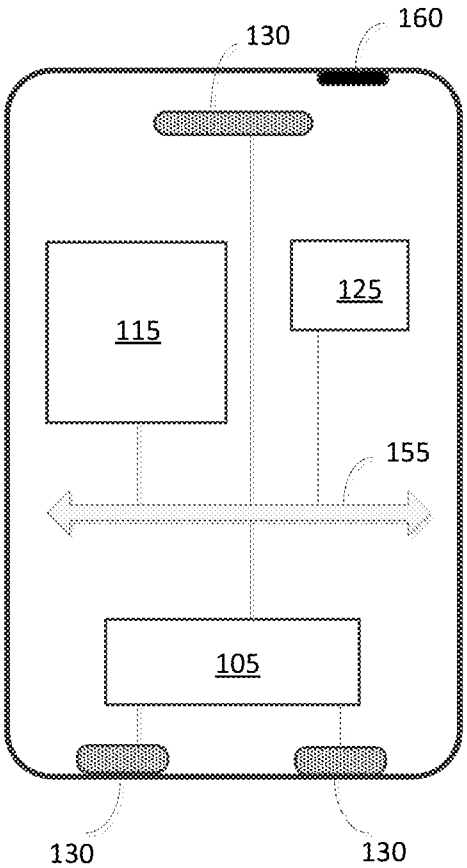


FIG. 1D

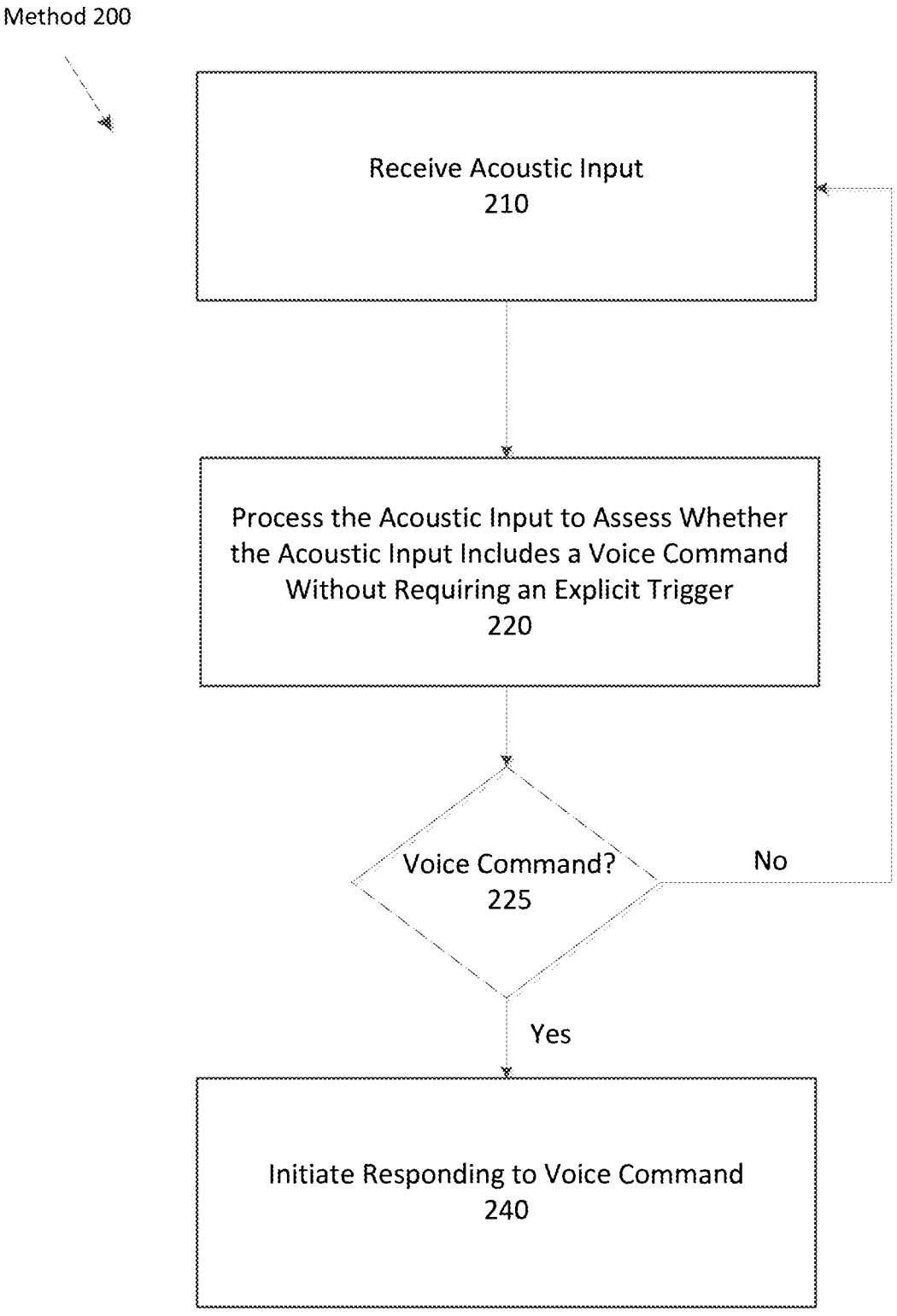

Method 200
Receive Acoustic Input
210
Process the Acoustic Input to Assess Whether the Acoustic Input Includes a Voice Command Without Requiring an Explicit Trigger
220
Voice Command?
225
No
Yes
Initiate Responding to Voice Command
240

FIG. 2

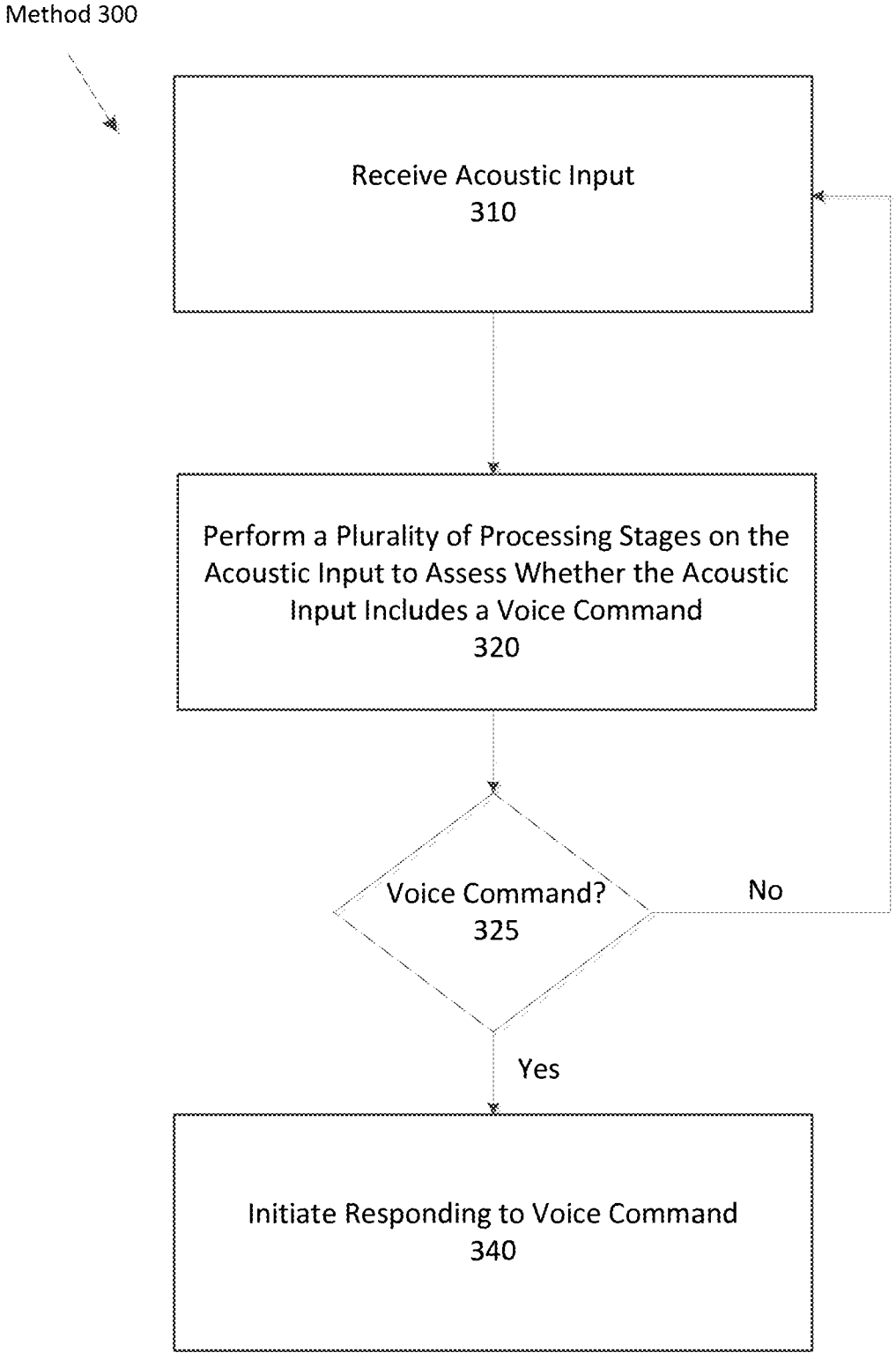

Method 300
Receive Acoustic Input
310
Perform a Plurality of Processing Stages on the Acoustic Input to Assess Whether the Acoustic Input Includes a Voice Command
320
Voice Command?
325
No
Yes
Initiate Responding to Voice Command
340

FIG. 3

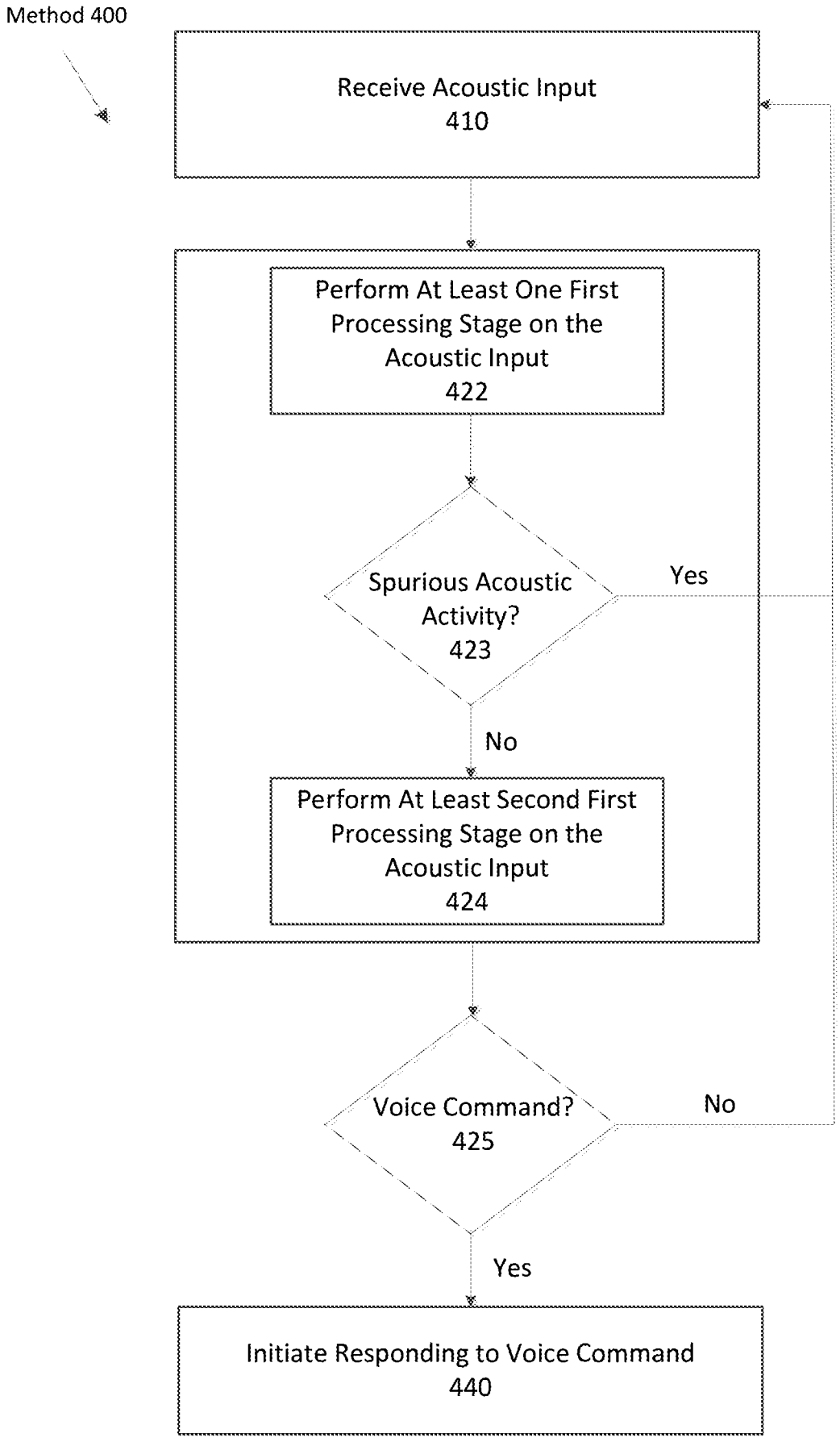

Method 400
Receive Acoustic Input
410
Perform At Least One First Processing Stage on the Acoustic Input
422
Spurious Acoustic Activity?
423
Yes
No
Perform At Least Second First Processing Stage on the Acoustic Input
424
Voice Command?
425
No
Yes
Initiate Responding to Voice Command
440

FIG. 4

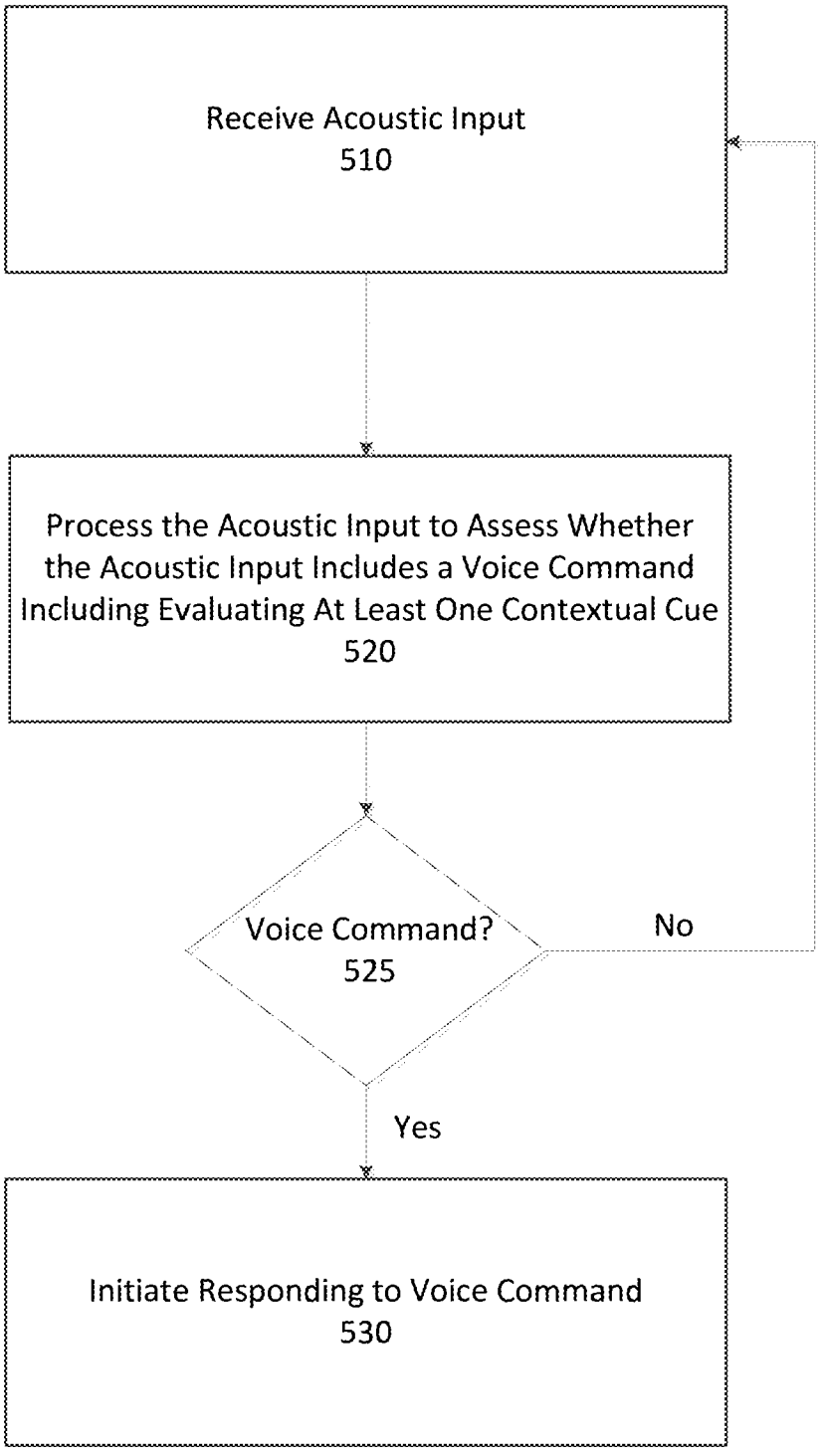

Receive Acoustic Input
510
Process the Acoustic Input to Assess Whether the Acoustic Input Includes a Voice Command Including Evaluating At Least One Contextual Cue
520
Voice Command?
525
No
Yes
Initiate Responding to Voice Command
530

FIG. 5

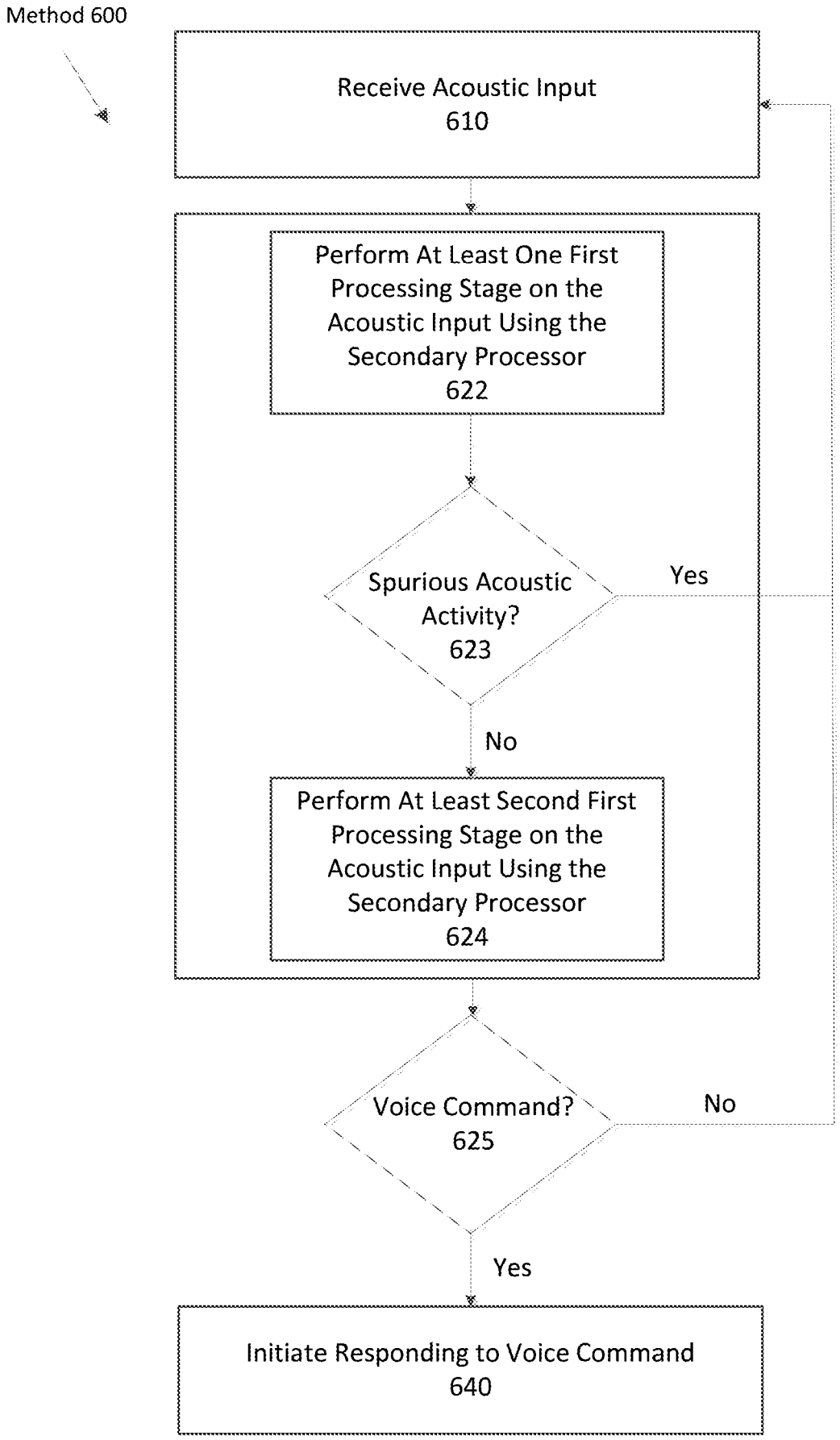

Method 600
Receive Acoustic Input
610
Perform At Least One First Processing Stage on the Acoustic Input Using the Secondary Processor
622
Spurious Acoustic Activity?
623
Yes
No
Perform At Least Second First Processing Stage on the Acoustic Input Using the Secondary Processor
624
Voice Command?
625
No
Yes
Initiate Responding to Voice Command
640

FIG. 6

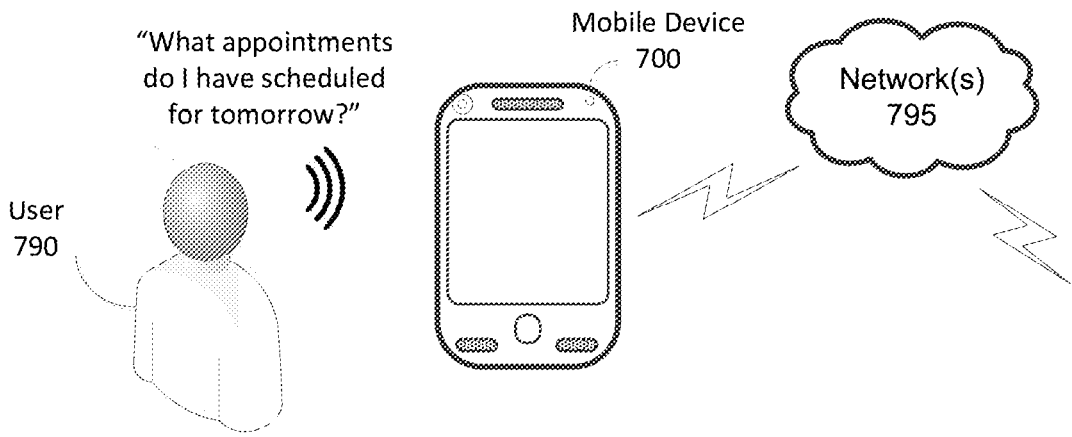


FIG. 7A

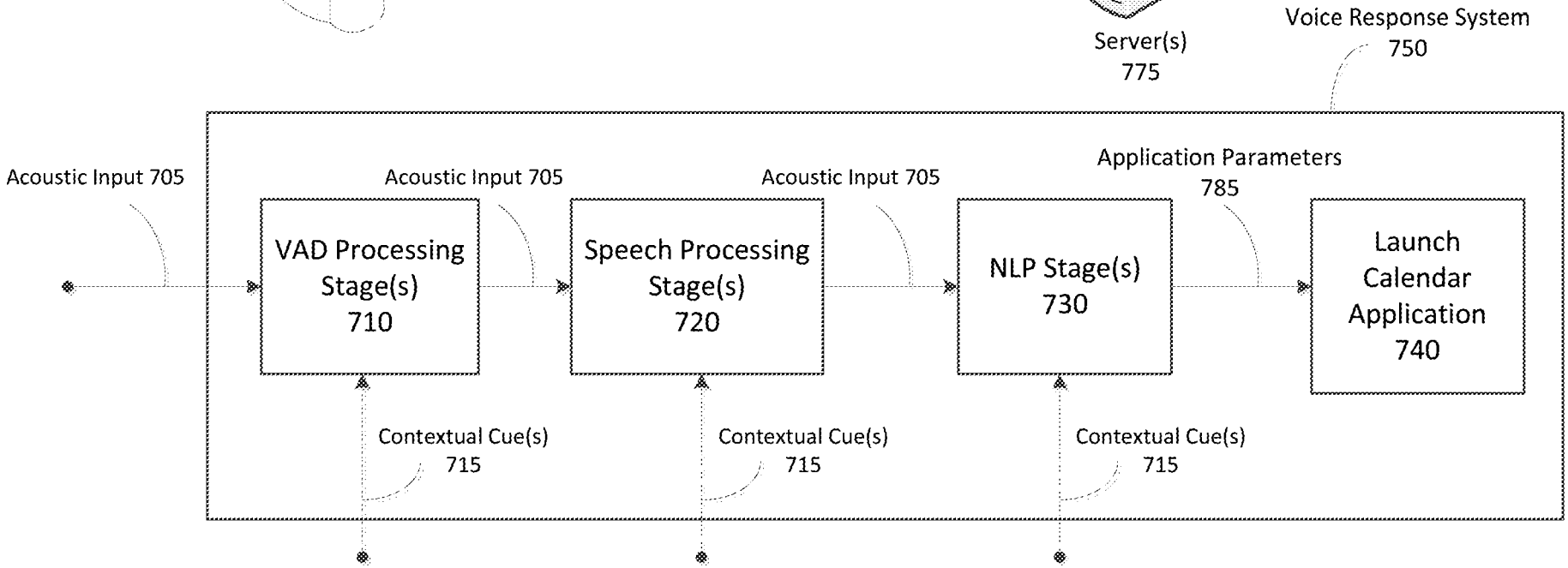


FIG. 7B

Petitioner Exhibit 1004, Page 21 of 1052

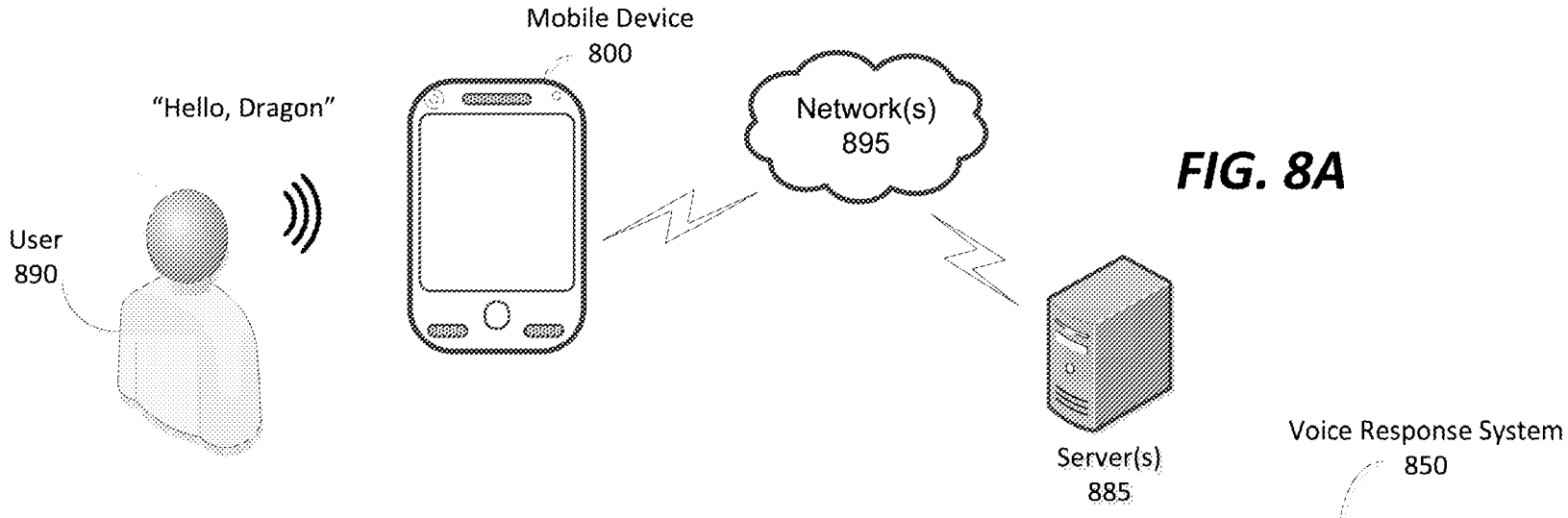


FIG. 8A

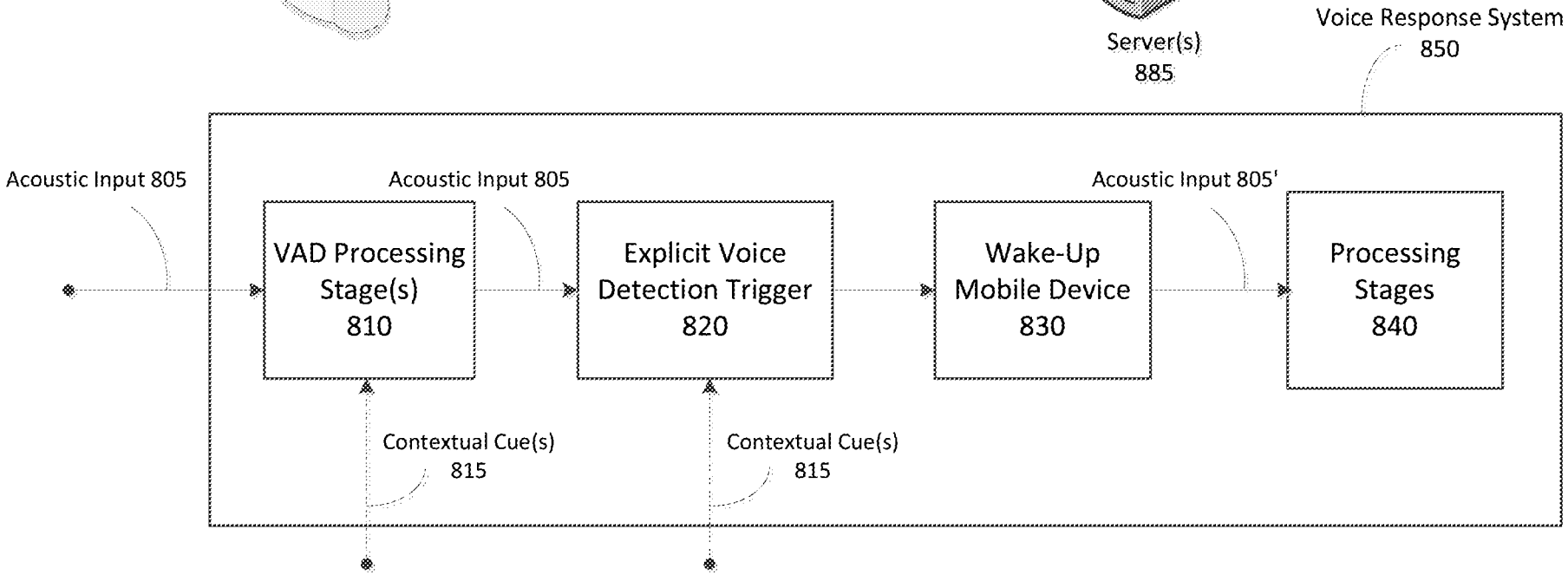


FIG. 8B

Petitioner Exhibit 1004, Page 22 of 1052

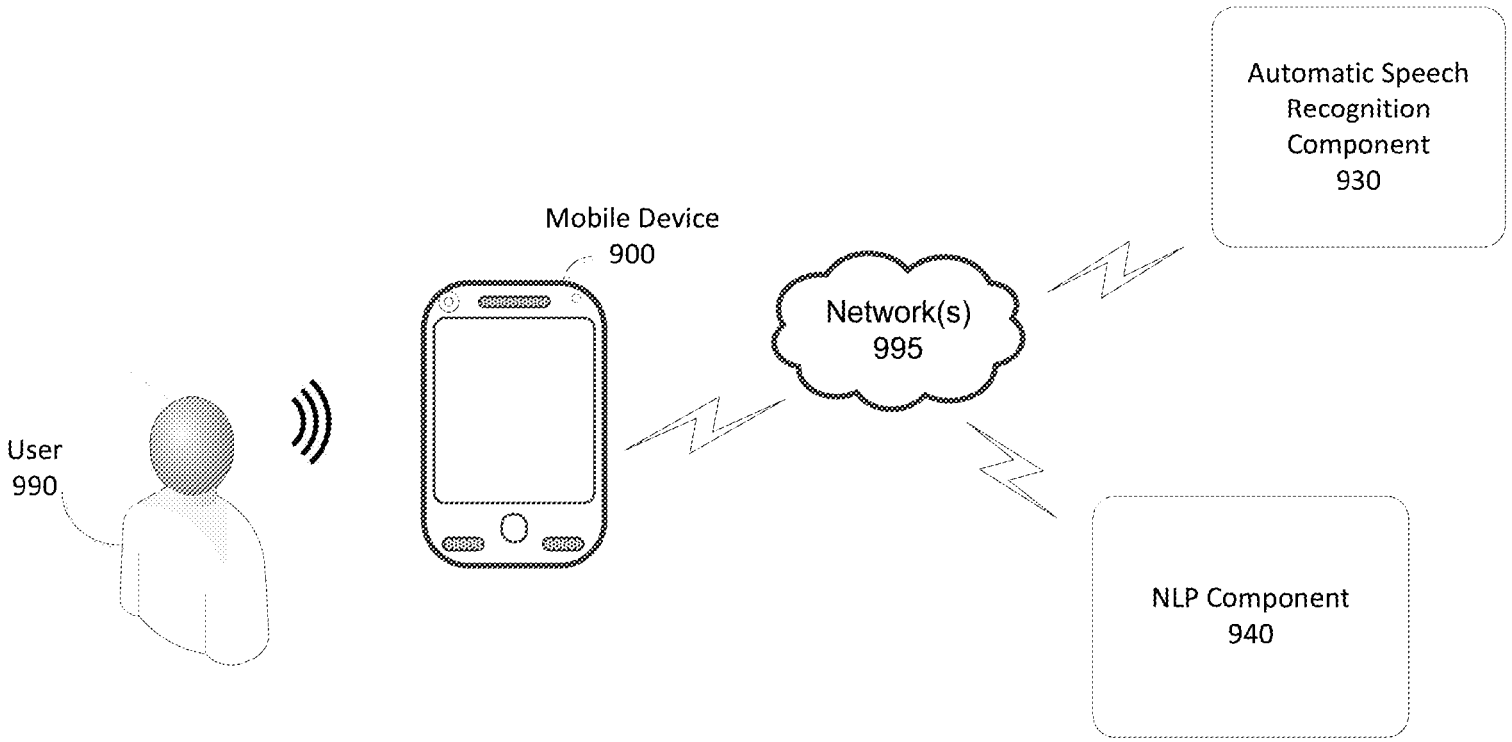


FIG. 9

Petitioner Exhibit 1004, Page 23 of 1052

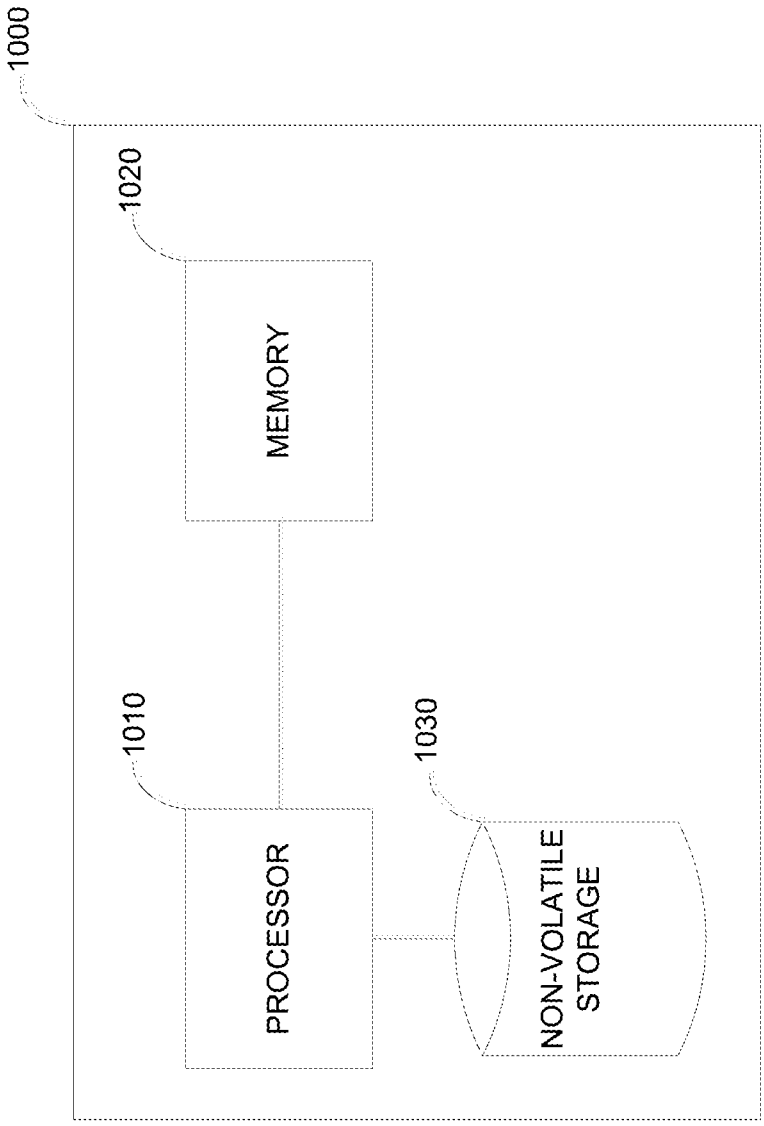


FIG. 10

PTO/SB/17 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

FEE TRANSMITTAL

Complete if known	
Application Number	Not Yet Assigned
Filing Date	Herewith
First Named Inventor	William F. Ganong, III
Examiner Name	Not Yet Assigned
Art Unit	N/A
Practitioner Docket No.	N0484.71090US04

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($) 1,600.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☐ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☐ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17 ☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	280	140*	70	600	300	150	720	360	180	1,600.00
Design	180	90	45	120	60	30	460	230	115	
Plant	180	90	45	380	190	95	580	290	145	
Reissue	280	140	70	600	300	150	2,160	1,080	540	
Provisional	260	130	65	0	0	0	0	0	0	

* The $140 small entity status filing fee for a utility application is further reduced to $70 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	80	40	20
Each independent claim over 3 (including Reissues)	420	210	105
Multiple dependent claims	780	390	195

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**	**Multiple Dependent Claims**	
				Fee ($)	**Fee Paid ($)**
20 - 20 or HP =	x	=			

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
2 - 3 or HP =	x	=	

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
58 - 100 =	/50 =	(round **up** to a whole number) x	=	

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge):	

SUBMITTED BY

Signature	/Thomas S. Chlebeck/	Registration No. (Attorney/Agent)	70,616	Telephone	617.646.8000
Name (Print/Type)	Thomas S. Chlebeck			Date	2016-08-16

4922877.1

PTO/AIA/15 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	N0484.71090US04
First Named Inventor	William F. Ganong, III
Title	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Express Mail Label No.	

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

1. [X] **Fee Transmittal Form** (PTO/SB/17 or equivalent)
2. [ ] **Applicant asserts small entity status.** See 37 CFR 1.27
3. [ ] **Applicant certifies micro entity status.** See 37 CFR 1.29. Applicant must attach form PTO/SB/15A or B or equivalent.
4. [X] **Specification** [*Total Pages* 67] Both the claims and abstract must start on a new page. *(See MPEP § 608.01(a) for information on the preferred arrangement)*
5. [X] **Drawing(s)** (35 U.S.C. 113) [*Total Sheets* 10]
6. **Inventor's Oath or Declaration** [*Total Pages* 2] *(including substitute statements under 37 CFR 1.64 and assignments serving as an oath or declaration under 37 CFR 1.63(e))*
   - a. [ ] Newly executed (original or copy)
   - b. [X] A copy from a prior application (37 CFR 1.63(d))
7. [X] **Application Data Sheet** ** See note below.* See 37 CFR 1.76 (PTO/AIA/14 or equivalent)
8. **CD-ROM or CD-R** in duplicate, large table, or Computer Program (*Appendix*)
   - [ ] Landscape Table on CD
9. **Nucleotide and/or Amino Acid Sequence Submission** (*if applicable, items a. – c. are required*)
   - a. [ ] Computer Readable Form (CRF)
   - b. [ ] Specification Sequence Listing on:
     - i. [ ] CD-ROM or CD-R (2 copies); or
     - ii. [ ] Paper
   - c. [ ] Statements verifying identity of above copies

ADDRESS TO: **Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450**

ACCOMPANYING APPLICATION PAPERS

10. [ ] **Assignment Papers** (cover sheet & document(s))
    Name of Assignee
11. [ ] **37 CFR 3.73(c) Statement** (*when there is an assignee*) [ ] **Power of Attorney**
12. [ ] **English Translation Document** (*if applicable*)
13. [ ] **Information Disclosure Statement** (PTO/SB/08 or PTO-1449)
    - [ ] Copies of citations attached
14. [ ] **Preliminary Amendment**
15. [ ] **Return Receipt Postcard** (*MPEP § 503*) (*Should be specifically itemized*)
16. [ ] **Certified Copy of Priority Document(s)** (*if foreign priority is claimed*)
17. [ ] **Nonpublication Request** Under 35 U.S.C. 122(b)(2)(B)(i). Applicant must attach form PTO/SB/35 or equivalent.
18. [ ] **Other:**

***Note:** (1) Benefit claims under 37 CFR 1.78 and foreign priority claims under 1.55 **must** be included in an Application Data Sheet (ADS).
(2) For applications filed under 35 U.S.C. 111, the application must contain an ADS specifying the applicant if the applicant is an assignee, person to whom the inventor is under an obligation to assign, or person who otherwise shows sufficient proprietary interest in the matter. See 37 CFR 1.46(b).

19. CORRESPONDENCE ADDRESS

[X] The address associated with Customer Number: 91337 **OR** [ ] Correspondence address below

Name					
Address					
City		State		Zip Code	
Country		Telephone		Email	

Signature	/Thomas S. Chlebeck/	Date	2016-08-16
Name (Print/Type)	Thomas S. Chlebeck	Registration No. (Attorney/Agent)	70,616

DECLARATION (37 C.F.R. § 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 C.F.R. § 1.76)

Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

As the below named inventor, I declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 13/795960 filed on March 12, 2013

☐ As amended on _____ (if applicable):

The above-identified application was made or authorized to be made by me.

I believe the inventor(s) named below to be the original and first inventor(s) of the subject matter which is claimed and for which a patent is sought.

I believe I am the original inventor or an original joint inventor of a claimed invention in the application.

I have reviewed and understand the contents of the above-identified application, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in 37 C.F.R. § 1.56.

All statements made herein of my own knowledge are true, and all statements made herein on information and belief are believed to be true.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. § 1001 by fine or imprisonment of not more than five (5) years, or both.

LEGAL NAME OF SOLE OR FIRST INVENTOR

Inventor one: William F. Ganong III Date: 3/14/2013

Signature: _____ Citizen of: USA

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor two: Paul A. Van Mulbregt Date: 14 March 2013

Signature: _____ Citizen of: New Zealand

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor three: Vladimir Sejnoha Date: 3/14/2013

Signature: _____ Citizen of: Canada

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor four: Glenn Wilson Date: _____

Signature: _____ Citizen of: USA

Petitioner Exhibit 1004, Page 27 of 1052

DECLARATION (37 C.F.R. § 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 C.F.R. § 1.76)

Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

As the below named inventor, I declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 13/795960 filed on March 12, 2013

☐ As amended on _____ (if applicable);

The above-identified application was made or authorized to be made by me.

I believe the inventor(s) named below to be the original and first inventor(s) of the subject matter which is claimed and for which a patent is sought.

I believe I am the original inventor or an original joint inventor of a claimed invention in the application.

I have reviewed and understand the contents of the above-identified application, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in 37 C.F.R. § 1.56.

All statements made herein of my own knowledge are true, and all statements made herein on information and belief are believed to be true.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. § 1001 by fine or imprisonment of not more than five (5) years, or both.

LEGAL NAME OF SOLE OR FIRST INVENTOR

Inventor one: William F. Ganong III Date: _____

Signature: _____ Citizen of: USA

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor two: Paul A Van Mulbregt Date: _____

Signature: _____ Citizen of: New Zealand

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor three: Vladimir Sejnoha Date: _____

Signature: _____ Citizen of: Canada

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor four: Glen Wilson Date: 3/22/2013

Signature: _____ Citizen of: USA

Petitioner Exhibit 1004, Page 28 of 1052

ABSTRACT

According to some aspects, a method of monitoring an acoustic environment of a mobile device, at least one computer readable medium encoded with instructions that, when executed, perform such a method and/or a mobile device configured to perform such a method is provided. The method comprises receiving acoustic input from the environment of the mobile device while the mobile device is operating in the low power mode, detecting whether the acoustic input includes a voice command based on performing a plurality of processing stages on the acoustic input, wherein at least one of the plurality of processing stages is performed while the mobile device is operating in the low power mode, and using at least one contextual cue to assist in detecting whether the acoustic input includes a voice command.

Petitioner Exhibit 1004, Page 29 of 1052

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Filer:	Thomas Stephen Chlebeck/Darlene Morin
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Utility application filing	1011	1	280	280
Utility Search Fee	1111	1	600	600
Utility Examination Fee	1311	1	720	720
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**1600**

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Related Applications

[0001] This Application claims the benefit under 35 U.S.C. §120 and is a continuation of U.S. Application Serial No. 13/795,933, entitled "METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND," filed on March 12, 2013, which is herein incorporated by reference in its entirety.

BACKGROUND

[0002] Many mobile communications devices, such as smart phones, are equipped with a voice response system (e.g., a virtual assistant or agent) that can recognize speech and respond to voice commands to perform desired tasks (perform an Internet search, make a phone call, provide directions, answer questions, make recommendations, schedule appointments, etc.). However, engaging the voice response system conventionally requires one or more manual actions by the user before the system is engaged and ready to respond to speech input from the user. For example, the user may have to activate an icon (e.g., via touch) to launch a virtual assistant application, or manipulate a software or hardware interface control on the mobile device to engage the voice response system (e.g., activate a microphone display icon, press a button, activate a switch, etc.).

[0003] Such manual actions requiring the user's hands, referred to herein as "manual triggers," complicate the interaction with a mobile device and, in some instances, may be prohibitive (e.g., when a user's hands are otherwise occupied). Voice triggers have been implemented to eliminate at least some of the manual actions required to activate a voice response system in an attempt to provide generally hands-free access to the voice response system. However, conventional voice response systems are responsive to voice triggers in limited contexts such as when the mobile device is active (i.e., awake), and require an explicit trigger word or phrase to engage the mobile device's voice response capabilities. As such, a user must speak a specific and predetermined word or phrase, referred to as an explicit voice trigger, to engage the voice response system and conventionally can only do so when the mobile device is active. That is, conventional voice response systems are unresponsive when the mobile device is asleep.

[0004] When a mobile device is operating in a low power mode (e.g., in a sleep, hibernate or idle mode), the actions required to engage the voice response system typically become even more extensive. In particular, the user need first wake-up the mobile device itself before the voice response system can be engaged using manual action or an explicit voice trigger. For example, a user may have to press a button to turn-on the display and/or enable one or more processors, may have to manipulate one or more controls to ready the mobile device for use, and/or may have to input a passcode if the mobile device has been inactive for a certain period of time.

[0005] Thus, wake-up actions may further hamper the use of a voice response system in ways that may be inconvenient or annoying under normal circumstances and prohibitive in others (e.g., while operating a vehicle, or engaging in other tasks that occupy the user's hands). Conventionally, these wake-up actions are unavoidable. Moreover, to engage a voice response system from a low power mode, one or more wake-up actions must then be followed by one or more manual and/or explicit voice triggers to engage the voice response system to be ready to respond to a user's speech.

SUMMARY

[0006] Some embodiments include a method of monitoring an acoustic environment of a mobile device, the method comprising receiving, by the mobile device, acoustic input from the environment of the mobile device, detecting whether the acoustic input includes a voice command from a user without requiring receipt of an explicit trigger from the user, and initiating responding to the detected voice command.

[0007] Some embodiments include at least one computer-readable medium encoded with instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a mobile device to receive acoustic input from the environment of the mobile device, the method comprising detecting whether the acoustic input includes a voice command from a user without requiring receipt of an explicit trigger from the user, and initiating responding to the detected voice command.

[0008] Some embodiments include a mobile device comprising at least one input for receiving acoustic input from the environment of the mobile device, and at least one processor configured to detect whether the acoustic input includes a voice command from a user without requiring receipt of an explicit trigger from the user, and initiate

responding to the detected voice command.

[0009] Some embodiments include a method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is operating in a low power mode, the method comprising receiving acoustic input from the environment of the mobile device while the mobile device is operating in the low power mode, detecting whether the acoustic input includes a voice command based on performing a plurality of processing stages on the acoustic input, wherein at least one of the plurality of processing stages is performed while the mobile device is operating in the low power mode, and using at least one contextual cue to assist in detecting whether the acoustic input includes a voice command.

[0010] Some embodiments include at least one computer readable medium encoded with instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a mobile device to receive acoustic input from the environment of the mobile device while the mobile device is operating in a low power mode, the method comprising detecting whether the acoustic input includes a voice command based on performing a plurality of processing stages on the acoustic input, wherein at least one of the plurality of processing stages is performed while the mobile device is operating in the low power mode, and using at least one contextual cue to assist in detecting whether the acoustic input includes a voice command.

[0011] Some embodiments include a mobile device comprising at least one input to receive acoustic input from an environment of the mobile device while the mobile device is operating in a low power mode, at least one processor configured to detect whether the acoustic input includes a voice command based on performing a plurality of processing stages on the acoustic input, wherein at least one of the plurality of processing stages is performed while the mobile device is operating in the low power mode, and use at least one contextual cue to assist in detecting whether the acoustic input includes a voice command

[0012] Some embodiments include a method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is operating in a low power mode, the mobile device having a first processor and second processor, the method comprising receiving acoustic input while the mobile device is operating in the low power mode, performing at least one first processing stage on the acoustic input using the first processor, prior to engaging the second processor to process

the acoustic input, to evaluate whether the acoustic input includes a voice command, performing at least one second processing stage on the acoustic input using the second processor to evaluate whether the acoustic input includes a voice command if further processing is needed to determine whether the acoustic input includes a voice command, and initiating responding to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

[0013] Some embodiments include at least one computer readable medium storing with instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a mobile device to receive acoustic input from the environment of the mobile device while the mobile device is operating in a low power mode, the mobile device having a first processor and second processor, the method comprising performing at least one first processing stage on the acoustic input using the first processor, prior to engaging the second processor to process the acoustic input, to evaluate whether the acoustic input includes a voice command, performing at least one second processing stage on the acoustic input using the second processor to evaluate whether the acoustic input includes a voice command if further processing is needed to determine whether the acoustic input includes a voice command, and initiating responding to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

[0014] Some embodiments include a mobile device comprising at least one input to receive acoustic input from an environment of the mobile device while the mobile device is operating in the low power mode, a first processor to perform at least one first processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command, a second processor to perform at least one second processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command, wherein the at least one first processing stage is performed prior to engaging the second processor, wherein the at least one second processing stage is performed if further processing is needed to determine whether the acoustic input includes a voice command, and wherein response to the voice command is initiated when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

BRIEF DESCRIPTION OF DRAWINGS

[0015] Various aspects and embodiments of the application will be described with reference to the following figures.

[0016] FIGS. 1A, 1B and 1C illustrate an exemplary mobile device on which voice response techniques may be implemented;

[0017] FIG. 1D illustrates exemplary system components of an illustrative mobile device on which voice response techniques may be implemented;

[0018] FIG. 2 is a flow chart illustrating a method of determining whether acoustic input received from the environment of a mobile device includes a voice command, without using an explicit trigger, in accordance with some embodiments;

[0019] FIG. 3 is a flow chart illustrating a multi-stage processing method for determining whether acoustic input received from the environment of a mobile device includes a voice command, in accordance with some embodiments;

[0020] FIG. 4 is a flow chart illustrating a multi-stage processing method for determining whether acoustic input received from the environment of a mobile device includes a voice command, using at least one first processing stage followed by at least one second processing stage, in accordance with some embodiments;

[0021] FIG. 5 is a flow chart illustrating a method for determining whether acoustic input received from the environment of a mobile device includes a voice command using at least one contextual cue, in accordance with some embodiments;

[0022] FIG. 6 is a flow chart illustrating a method for determining whether acoustic input received from the environment of a mobile device includes a voice command using multiple processors, in accordance with some embodiments;

[0023] FIGS. 7A and 7B illustrate an example of a voice response system for determining whether acoustic input received from the environment of a mobile device includes a voice command, without requiring an explicit trigger, in accordance with some embodiments;

[0024] FIGS. 8A and 8B illustrate an example of a voice response system for determining whether acoustic input received from the environment of a mobile device includes a voice command when the mobile device is operating in a low power mode, in accordance with some embodiments;

Petitioner Exhibit 1004, Page 36 of 1052

[0025] FIG. 9 illustrates a system including a mobile device capable of accessing one or more network resources to assist in assessing whether acoustic input includes a voice command, in accordance with some embodiments; and

[0026] FIG. 10 illustrates an exemplary computer system on which techniques described herein may be implemented.

DETAILED DESCRIPTION

[0027] As discussed above, conventional voice response systems require one or more explicit triggers to engage the voice response system. An "explicit trigger" refers herein to one or more specific, designated and predetermined actions required to engage a voice response system, and includes manual triggers (i.e., actions performed on the mobile device via a user's hands) and explicit voice triggers (i.e., speaking a specific, designated word or phrase to engage the voice response system).

[0028] The inventors have appreciated that user interaction with a mobile device may be improved if the user could interact with a voice response system of the mobile device simply by speaking one or more desired voice commands, without having to issue one or more explicit triggers. The term "voice command" refers herein to any type of actionable voice input from a user to a voice response system, including, but not limited to, voice requests (e.g., "Schedule my doctor's appointment for tomorrow at 9am," "Call Tom," "Remind me to pick-up kids tonight," "Please recommend a nearby Chinese restaurant," "Search for attractions in Boston," "Take dictation," etc.), voice queries (e.g., "Where is the nearest gas station?" "What is the temperature in Boston?" "How do I get to the Mass Pike from here?" "What appointments do I have scheduled today?" "How many tablespoons are in a cup?" etc.), and explicit voice triggers (e.g., "Hello, Dragon") to engage a voice response system.

[0029] Mobile devices are generally power-limited in that they are frequently expected to operate for relatively extended periods of time off of battery power. To conserve power, mobile devices frequently enter a low power mode (e.g., a sleep or hibernate mode) after some duration of inactivity. A low power mode refers generally to any mode or state that is entered to conserve power that conventionally requires one or more manual wake-up actions to exit (e.g., to activate the mobile device). Mobile devices may have more than one low power mode and different mobile devices may power down different components when entering a low power mode. However, mobile

Petitioner Exhibit 1004, Page 37 of 1052

devices will typically have some variation of a low power mode that conventionally requires manual action to exit the low power mode and transition to an active mode.

[0030] Entering a low power mode may include turning off the display, shutting down one or more processors (e.g., disabling the clock tree associated with one or more processors) and/or deactivating components that would otherwise draw power if left operational. Automatically entering a low power mode after a designated interval of time may significantly extend the period that a mobile device can operate on battery power before requiring a recharge. In view of the relative importance of conserving power in a mobile device, many mobile devices default to a relatively short duration of inactivity before powering down and going to sleep, such that a mobile device may frequently be in a low power mode when a user wants to access functionality on the mobile device.

[0031] As discussed above, one or more manual actions must be performed to wake-up the mobile device before the device can be utilized. Conventional wake-up triggers are manual in that they require the user's hands to activate the mobile device. Though many mobile devices are designed to at least partially wake-up when certain events occur (e.g., when an incoming telephone call is received, when an alert or alarm is set and executes), the user generally must perform the appropriate manual triggers to achieve a user initiated wake-up of the mobile device if the user would like to engage with a voice response system. Such manual triggers may be inconvenient, annoying and/or, in some circumstances, prohibitive (e.g., when the user is driving or the user's hands are otherwise occupied).

[0032] The inventors have recognized that user experience improvements may result if a user could engage with a mobile device via voice activation even when the mobile device is operating in a low power mode. Moreover, the inventors have appreciated that voice activation functionality may facilitate hands-free interaction with a mobile device in virtually any situation, independent of whether the mobile device is operating in an active or in a low power mode (e.g., a sleep mode). That is, from the perspective of the responsiveness of the mobile device, the inventors have appreciated the utility of providing voice activation functionality that makes it appear that the mobile device is "always-on" by monitoring the acoustic environment (e.g., the voice response system of the mobile device can be voice initiated or activated whether the device is in an active mode or in a low power mode).

Petitioner Exhibit 1004, Page 38 of 1052

[0033] The term "always-on" is used herein to describe the ability of a voice response system of a mobile device to be generally responsive to voice (e.g., capable of detecting when a voice command has been spoken) both when a mobile device is active and when the device has entered a low power mode. To achieve such "always-on" functionality, the acoustic environment of the mobile device may need to be monitored both in active and low power modes to obtain acoustic input and determine whether the acoustic input includes a voice command or whether the acoustic input resulted from spurious acoustic activity. The term "spurious acoustic activity" refers generally to any acoustic activity (including speech) detected in the acoustic environment of the mobile device that does not correspond to a voice command upon which the voice response system can or should act.

[0034] According to some embodiments, the acoustic environment of a mobile device is monitored to receive acoustic input when the mobile device is in a low power mode and to detect when the acoustic input includes a voice command. Rendering a mobile device responsive to voice when the mobile device is powered down in a low power mode (e.g., when the mobile device is asleep or otherwise in an inactive state that conventionally requires one or more wake-up actions) presents challenges, particularly in view of the relatively strict power consumption constraints of mobile devices. Mobile devices typically inhabit the same noisy environments of their users. Recruiting the mobile device's relatively power intensive resources each time acoustic input is received by the mobile device (e.g., when a microphone detects acoustic activity) to assess whether a voice command has been spoken could utilize significant enough processing power to frustrate the efforts of powering the mobile device down into a low power mode in the first place.

[0035] The inventors have recognized the importance of minimizing power consumption to the extent reasonable when performing processing on received acoustic input to determine whether the acoustic input includes a voice command. To this end, some embodiments utilize a hierarchical approach to monitoring and processing acoustic input that applies relatively low power processing to received acoustic information first and generally higher power processing as needed to evaluate the likelihood that the acoustic information includes a voice command. According to some embodiments, a plurality of processing stages are performed that evaluate whether acoustic input includes a voice command, wherein at least one subsequent stage of the plurality of processing

stages is performed only if one or more previous processing stages is unable conclude that the acoustic input corresponds to spurious acoustic activity.

[0036] Achieving viable "always-on" operation may involve attempting to generally optimize false positive and false negative rates while keeping power consumption to a minimum to a reasonable extent. According to some embodiments, the one or more processing stages used to determine if acoustic input includes a voice command are selected in an attempt to avoid over inclusion that can result in noticeable reduction in battery life and/or the nuisance of a mobile device that too frequently wakes-up in response to spurious acoustic activity, either result of which may be unsatisfactory to the user. On the other hand, over exclusion may result in a voice response system that too frequently misses voice commands such that the mobile device is perceived by the user as non-responsive. Some embodiments include techniques designed to strike an acceptable balance in this regard.

[0037] The inventors have recognized that considering one or more contextual cues may facilitate improved assessment of whether acoustic information includes a voice command. The term "contextual cue" refers generally to information other than the properties or characteristics of the specific acoustic input being evaluated as containing a voice command. For example, a user may pick-up their mobile device or bring it closer to their mouth when preparing to speak a voice command. Thus, motion may be used as a contextual cue to influence the evaluation of whether acoustic input likely includes a voice command. Other contextual cues that may be considered, alone or in any combination, include (but are not limited to) location of the mobile device, time of day, a power state of the mobile device, occurrence of an event on the mobile device, recent activity on the mobile device, acoustic environment, etc. Contextual cues may be used to bias one or more processing stages, aid in the selection and/or omission of one or more processing stages, operate as a processing stage itself, or otherwise assist in determining whether acoustic input includes a voice command, as discussed in further detail below.

[0038] As discussed above, conserving power may be an important component of implementing "always-on" functionality. The inventors have appreciated that relatively low power processors on a mobile device, such as a secondary, specialized and/or low power processor may be utilized to perform one or more initial processing stages to determine whether received acoustic input corresponds to spurious acoustic activity, or

Petitioner Exhibit 1004, Page 40 of 1052

whether higher level (and typically higher power) processing is needed to conclude that the acoustic information includes a voice command. If the one or more initial processing stages cannot dispense with the acoustic input as spurious acoustic activity, the mobile device's primary processor (e.g., main central processing unit (CPU)) may be engaged to perform one or more subsequent processing stages. Thus, a relatively substantial amount of the acoustic input received by the mobile device may be evaluated and identified as spurious acoustic activity (and thereafter ignored) without needing to activate the primary processor, as discussed in further detail below.

[0039] Numerous exemplary techniques for processing "acoustic input" obtained from monitoring the acoustic environment of a mobile phone are described herein. Different techniques may process acoustic input at different stages before and after the acoustic input has been transformed, reduced, supplemented or otherwise modified. For example, some processing stages may apply techniques to acoustic input in analog form, while other stages may process acoustic input in digital form. Some processing stages may operate on acoustic input as an audio signal, while other stages may operate on acoustic input that has been recognized, at least partially, as its constituent words.

[0040] Thus, "acoustic input" refers generally to information received from the acoustic environment or information derived therefrom. When multiple processing stages are described as processing a given acoustic input, it should be understood that the acoustic input operated on by each may be in different form and contain different information (or may be in the same form and contain essentially the same information) as the given acoustic input may have undergone different amounts and types of modifications and processing when a respective processing stage operates on it. Thus, the acoustic input refers generically to any of the forms in which it may take, unless expressly specified otherwise.

[0041] Following below are more detailed descriptions of various concepts related to, and embodiments of, methods and apparatus for implementing techniques for engaging a voice response system. It should be appreciated that various aspects described herein may be implemented in any of numerous ways. Examples of specific implementations are provided herein for illustrative purposes only. In addition, the various aspects described in the embodiments below may be used alone or in any combination, and are not limited to the combinations explicitly described herein.

Petitioner Exhibit 1004, Page 41 of 1052

[0042] FIGS. 1A-1C illustrate an exemplary mobile device on which one or more voice activation techniques may be implemented. Mobile device 100 is illustrated as having a number of components characteristic of a "smart phone," but it should be appreciated that mobile device 100 may be any generally portable device capable of wireless communications and which includes an input capable of receiving acoustic input (e.g., one or more microphones). Mobile device 100 includes one or more transducers 130 for converting acoustic energy to electrical energy and vice versa. For example, transducers 130 may include one or more speakers and/or one or more microphones arranged on the mobile device to allow input/output (I/O) of acoustic information. Different mobile devices may include differing numbers and arrangements of transducers, any of which may be suitable to facilitate acoustic I/O.

[0043] Mobile device 100 may also include one or more wireless communication components 160. For example, mobile communication devices such as cellular telephones will typically have a wireless transceiver capable of communicating with one or more cellular networks. Alternatively, or in addition to, wireless communication component 160 may include a wireless transceiver capable of communicating with one or more other networks or external devices. For example, wireless communication component 160 may include a component configured to communication via the IEEE 802.11 standard (Wi-Fi) to connect to a local area network (LAN), wide area network (WAN) such as the Internet, and/or may include a Bluetooth® transceiver to connect to a Bluetooth® compatible device. Wireless communication component 160 may also include a Global Positioning System (GPS) to communicate with a satellite network to determine the location of the mobile device, or may include one or more other wireless communication transceivers.

[0044] To simplify the illustration of exemplary mobile device 100, wireless communication component 160 is used to represent any wireless communication component 160 implemented on mobile device 100 and may include one or multiple components that may be either separate or integrated components, may be independent system components, or may be different system components in communication with one or more other wireless components. That is, wireless communication component 160 denotes one or any combination of components that may be implemented on a mobile device that allows the mobile device 100 to communicate wirelessly with one or more networks, systems and/or other devices.

Petitioner Exhibit 1004, Page 42 of 1052

[0045] Mobile device 100 may also include a display 110 for visually presenting information to the user. Display 110 may also be configured as an input device, for example, by responding to touch, pen input, etc. Mobile device 100 may include one or more other input components 120 that allow a user to interface with the mobile device. For example, button 120a may be used to initiate wake-up of the mobile device when the mobile device has entered a low power mode such as a sleep mode and/or may be used as a selection button or may include any other functionality when the mobile device is active. Button 120b may be a power button, volume button, or any other input component having any desired functionality. Different mobile devices may have different numbers and arrangements of input components 120 depending on the design of the mobile device.

[0046] Mobile device 100 may also include other components such as one or more cameras 150, one or more light emitting diodes (LEDs) 170, one or more motion detectors (discussed in further detail below), and/or any other components. Mobile device 100 may also include a power port 190 for connecting the mobile device to a power source (e.g., wall power) and/or a data port for providing/receiving data to/from another device connected via a data cable. According to some mobile device designs, the data and power port are the same port and can be used both for power and for data.

[0047] It should be appreciated that mobile device 100 is merely an example of a mobile device suitable for use with any one or combination of voice activation techniques described herein. The mobile device components are described to illustrate common features of smart phones, but no particular one or combination of these components are required. Techniques described herein are not limited for use with any particular mobile device having any particular number or arrangement of components and any mobile device having the ability to receive acoustic input may be suitable, as the aspects described herein are not limited in this respect.

[0048] FIG. 2 is a flow chart illustrating a method of monitoring an acoustic environment of a mobile device to detect voice commands, in accordance with some embodiments. As discussed above, the inventors have appreciated the benefits of a mobile device having a voice response system that evaluates received acoustic input to ascertain whether a user has spoken a voice command, without requiring an explicit trigger. In act 210, acoustic input from the environment of the mobile device is received by the mobile device. For example, one or more microphones may sense acoustic

Petitioner Exhibit 1004, Page 43 of 1052

activity in the environment and obtain the resulting acoustic input for further processing to assess whether the acoustic input includes a voice command.

[0049] When the voice response system is enabled to monitor the acoustic environment, act 210 may be performed continuously or periodically at a frequency that provides an appearance of continuous monitoring, even though it may not be strictly continuous. The acoustic environment may be monitored to obtain an acoustic input over a desired interval of time before processing the acoustic input, the acoustic environment may be sampled periodically to obtain the acoustic input, an event may trigger collecting and processing of acoustic input (e.g., the microphone may indicate that sound of a given threshold has been detected), or any other method of monitoring the acoustic environment may be used such that acoustic input is received for further processing.

[0050] The acoustic environment may continue to be monitored while received acoustic input is being evaluated (e.g., act 210 may be essentially performed on a continuous, near continuous or ongoing basis while other processing is performed). New acoustic input received while prior acoustic input is being evaluated may be buffered or processed in parallel with the prior acoustic input. According to some embodiments, the new acoustic input is processed to assist in the evaluation of whether the prior acoustic input includes a voice command. In this manner, new acoustic may be used to inform the evaluation of prior received acoustic input. Also, information ascertained about the prior received acoustic input may be used to inform the evaluation of the new acoustic input received subsequently. That is, received acoustic input may be used in a feed-forward or feed-back configuration to inform the processing of acoustic input received before or after, as discussed in further detail below. Some embodiments include both feeding forward and feeding back information about acoustic input that has been received by the mobile device to better evaluate whether a voice command has been received from the user. As such, act 210 may, in some embodiments, reflect a generally ongoing process when the voice response system is enabled to monitor the acoustic environment.

[0051] According to some embodiments, a voice response system may be enabled to monitor the acoustic environment both in active and low power modes, while in some embodiments, monitoring the acoustic environment is limited to either active or lower power modes. According to some embodiments, whether the voice response

Petitioner Exhibit 1004, Page 44 of 1052

system monitors the acoustic environment in active modes, low power modes or both is user selectable. In this manner, a user can choose when the mobile device should operate in an "always-on" mode and can select what operating modes the mobile device should perform "always-on" functionality.

[0052] In act 220, the acoustic input is processed to determine whether the acoustic input includes a voice command, without requiring an explicit trigger to do so. That is, the user is not required to manually engage the voice response system (e.g., by performing one or more manual triggers such as manipulating one or more interface controls by hand), nor is the user required to speak an explicit voice trigger to notify the voice response system that the user is uttering or will immanently utter an actionable voice command (though in some embodiments, the user can optionally use an explicit voice trigger if the user so desires, while not requiring the user to do so). Thus, according to some embodiments, the user is free to simply begin speaking to the mobile device to engage the voice response system.

[0053] Implementing "always-on" functionality presents challenges with respect to balancing false positive and false negative rates. If the former is too high (e.g., the voice response system is too inclusive), the mobile device will too frequently respond to spurious acoustic activity, and if the latter is too high (e.g., the voice response system is too exclusive), the mobile device will be perceived as unresponsive. According to some embodiments, a multistage processing approach may be used to facilitate balancing the false positive and false negative rates while minimizing (to the extent reasonable) the amount of power consumed in determining whether the acoustic input corresponds to spurious acoustic activity or includes a voice command.

[0054] According to some embodiments, act 220 may include performing one or more voice activity detection (VAD) processing stages that evaluate whether the acoustic input has the characteristics of voice/speech or whether the acoustic input is more likely the result of non-voice acoustic activity in the environment. VAD techniques refer generally to those that analyze one or more properties or characteristics of acoustic input (e.g., signal characteristics of acoustic input) to evaluate whether the one or more properties/characteristics are suggestive of speech, some techniques of which are described in further detail below. The result of performing one or more VAD processing stages may include assessing a likelihood that the acoustic input includes voice content, which assessment may be used to determine whether the acoustic input can be ignored as

Petitioner Exhibit 1004, Page 45 of 1052

spurious acoustic activity, or whether the acoustic input should be further processed to determine the content of the acoustic input (e.g., determine the content and/or understand the content of speech).

[0055] With respect to evaluating the content of acoustic input containing speech, act 220 may include one or more speech processing stages configured to learn something about the content of the acoustic input. For example, in embodiments that employ one or more VAD processing stages, a speech processing stage (or multiple speech processing stages) may be performed on acoustic input for which voice content has been detected via the one or more VAD processing stages. The speech processing stage(s) may evaluate the acoustic input to determine the content of speech contained therein to assess whether the user has spoken a voice command.

[0056] Speech processing stages may include, but are not limited to, any one or combination of automatic speech recognition (ASR) processes such as limited vocabulary ASR (e.g., evaluating the acoustic input to determine whether a user spoke an explicit voice trigger, evaluating the acoustic input to determine whether the acoustic input includes certain words characteristic of a voice command, etc.), unrestricted and/or large vocabulary ASR, etc. Speech processing stages may include one or more language processes that seek to understand whether words in the acoustic input are characteristic of a voice command and/or to understand the meaning of what the user spoke. For example, speech processing stages may include one or more statistical classifiers, natural language processing (NLP), or any suitable language techniques designed to categorize or understand the content to facilitate determining whether a voice command has been spoken. Some non-limiting examples of speech processing stages that may be utilized are discussed in further detail below.

[0057] According to some embodiments that implement a multi-stage approach, when performing act 220, generally low power processing stages may be performed before engaging one or more higher power processing stages. By doing so, a significant amount of acoustic input may be dispensed with as corresponding to spurious acoustic activity while consuming relatively little power to generally minimize the impact on the battery. Subsequent processing stage(s) may be reserved for assessing acoustic input for which a conclusion could not be reached by earlier (and generally lower power) processing stages regarding whether the acoustic input includes a voice command or corresponds to spurious acoustic activity. In this manner, the sequence of stages may be

Petitioner Exhibit 1004, Page 46 of 1052

selected to generally optimize false positive and false negative rates while attempting to keep power consumption to a minimum.

[0058] Other techniques may be used to assist in minimizing false positive and false negative rates while keeping power consumption relatively low, when performing act 220. For example, act 220 may utilize one or more contextual cues to influence the determination of whether the acoustic input includes a voice command, some exemplary contextual cues of which are described in further detail below. Alternatively or in addition to, one or more processing stages may be performed on a secondary processor (e.g., a lower power processor) and a primary processor (e.g., the main CPU) engaged only if needed to conclude whether the acoustic input includes a voice command, as discussed in further detail below. It should be appreciated that any one or combination of techniques described herein may be used to determine whether the acoustic input includes a voice command, as the aspects are limited to using any particular technique or combination of techniques.

[0059] If it is determined that the acoustic input includes a voice command, the voice response system may initiate one or more processes to respond to the voice command (act 230). For example, the voice response system may perform further language processing to understand what the voice command means and engage the necessary procedures/components required to undertake carrying out the directives of the voice command. Otherwise, the mobile device may discontinue further processing of the acoustic input and ignore it as spurious acoustic activity (e.g., non-speech sounds, background noise, speech not corresponding to a voice command or, according to some embodiments, speech from one or more people that are not the user of the mobile device, as discussed in further detail below). The voice response system may then continue to monitor the acoustic environment to obtain further acoustic input (e.g., the voice response system may return to or continue to perform act 210).

[0060] Initiating one or more further processes when acoustic input is determined to include a voice command may include, but is not limited to, engaging one or more language processing stages to understand the meaning of the voice command, initiating one or more tasks needed to carry out the voice command such as initiating a search, launching one or more applications or processes to, for example, initiate a search, schedule an appointment, update a calendar, create an alert, alarm or other electronic reminder, generate a text or email, make a telephone call, access a website, etc.,

responding to the user with a request for more information regarding the voice command or to confirm an understanding of the voice command, and/or initiating or performing any other task that the voice response system is capable of initiating, engaging and/or performing, either locally on the mobile device and/or remotely via one or more networks that the mobile device is capable of connecting to and interacting with. Initiating further processing may include evaluating or modifying the evaluation of subsequently received acoustic input, for example, when the detected voice command includes an explicit voice trigger.

[0061] As discussed above, the voice activation method described in FIG. 2 may be performed to monitor the acoustic environment to detect voice commands when a mobile device is operating in an active mode, when the mobile device is operating in a low power mode, or both. According to some embodiments, the manner in which acoustic input is processed may depend on whether the mobile device is operating in an active or low power mode. For example, if the mobile device is in an active mode, the display may be on, the primary processor enabled and/or other components active such that these resources may be more readily utilized without first having to activate them (e.g., if the primary processor and/or other components are already engaged, certain processing stages may be performed with less overhead and/or a number of contextual cues may be more readily accessible without incurring relatively large amounts of overhead to engage the components that provide the contextual cues and process this information).

[0062] As such, in an active mode, use of certain processing stages early on may be justified, wherein this might not be the case when operating in a low power mode (e.g., because of overhead, user expectations regarding a mobile device that is asleep, power consumption concerns, perception of the intrusiveness of the voice response system to the user, etc.). When the mobile device is operating in a low power mode, the processing strategy for determining whether acoustic input includes a voice command may proceed more conservatively with regard to utilizing resources and/or performing higher level processing stages, reserving further processing for when it is necessary. However, it should be appreciated that, while the operating mode may be advantageously used to tailor the processing strategy in some circumstances, the choice and/or order of processing stages need not depend on the operating mode of the mobile device and evaluating acoustic input may proceed in a similar or same manner independent of the

Petitioner Exhibit 1004, Page 48 of 1052

operating mode of the mobile device, as the aspects are not limited in this respect. Various techniques for selecting which processing stages to perform and in what order to determine whether the acoustic input includes a voice command are discussed in further detail below.

[0063] As discussed above, mobile devices are generally power limited and therefore conserving power may be a priority on such devices. FIG. 3 illustrates a method of using a multi-stage approach to relatively low power monitoring of the acoustic environment of a mobile device to detect uttered voice commands, in accordance with some embodiments. Method 300 may be utilized both when a mobile device is in an active mode, when the mobile device is in a low power mode (e.g., asleep, hibernating, idle, etc.), or both. Initially, acoustic input to be evaluated is received in act 310 (e.g., acoustic input is received in any of the ways described above in connection with act 210, or any other suitable manner). In act 320, a plurality of processing stages are performed on the acoustic input to evaluate whether the acoustic input likely includes a voice command, or whether the acoustic input is likely the result of spurious acoustic activity. The plurality of stages may be performed generally in a sequence (e.g., serially), may be performed in parallel, or a combination of serial and parallel processing.

[0064] A processing stage that determines that the acoustic input likely corresponds to spurious acoustic activity may terminate further processing of the acoustic input to avoid consuming additional power. A processing stage that concludes that the acoustic input likely corresponds to a voice command (e.g., to a desired level of confidence) may initiate further processing to respond to the voice command (act 330). When processing one or more stages sequentially, a processing stage that neither concludes that the acoustic input includes a voice command, nor that the acoustic input corresponds to spurious acoustic activity, may engage a subsequent processing stage to proceed with further evaluation of the acoustic input. Thus, additional processing may be reserved for circumstances in which it is needed to arrive at a conclusion about the nature of the acoustic input to conserve power.

[0065] Selecting a subsequent stage may depend on the stage itself, a confidence that a processing stage (or one or more previous stages) has in its assessment, one or more contextual cues and/or any other suitable factor. According to some embodiments, the sequence of processing stages utilized are dynamically selected with an intent to

Petitioner Exhibit 1004, Page 49 of 1052

optimize false positive and false negative rates while consuming as little power as reasonable, some techniques of which are discussed in further detail below. However, in some embodiments, the sequence of stages that are utilized is essentially static and processing of the acoustic input generally proceeds in a prescribed manner until the acoustic input has been either determined to include a voice command or to correspond to spurious acoustic activity.

[0066] When performing processing stages in parallel, the results from the parallel processing stages may be used together to assess the likelihood that an acoustic input includes a voice command. For example, the results from parallel processing stages may be utilized in a voting scheme, weighted decision or other suitable manner that evaluates the information from a plurality of processing stages to assist in determining whether acoustic input includes a voice command.

[0067] According to some embodiments, both sequential/serial processing and parallel processing are used with respect to the plurality of processing stages performed to assess whether acoustic input includes a voice command. For example, a plurality of processing stages may be performed in parallel and the results may be used to determine if further processing is needed and/or which subsequent one or more processing stages should be selected for performance. Sequential and parallel processing can be used together in other ways, as the aspects are not limited in this respect.

[0068] FIG. 4 illustrates a multi-stage method for determining whether acoustic input includes a voice command wherein at least some of the processing stages are performed in a sequence. Method 400 may be similar in some respects to the method illustrated in FIG. 3 in that acoustic input is received and processed using a plurality of processing stage to determine whether it includes a voice command, or whether the acoustic input likely resulted from spurious acoustic activity. Method 400 provides a hierarchical approach to multi-stage processing to facilitate conserving the power resources of the mobile device.

[0069] In particular, performing a plurality of processing stages to process the acoustic input comprises performing at least one first processing stage (act 422) prior to performing at least one second processing stage (act 424). The at least one first processing stage may include one or more stages designed to determine whether the acoustic input may potentially include a voice command, or conclude that the acoustic input corresponds to spurious acoustic activity. If the at least one first processing stage

Petitioner Exhibit 1004, Page 50 of 1052

determines that the acoustic input corresponds to spurious acoustic activity, further processing of the acoustic input may be terminated and further acoustic input received from monitoring the acoustic environment evaluated (e.g., a voice response system may return to or continue to perform act 410).

[0070] Otherwise, if the at least one first processing stage is not able to conclude that the acoustic input corresponds to spurious acoustic activity, at least one second processing stage may be performed to assess whether the acoustic input includes a voice command, or corresponds to spurious acoustic activity. The at least one second processing stage may include one or more processing stages configured to conclude whether the acoustic input includes a voice command, or instead corresponds to spurious acoustic activity. If the at least one second processing stage determines that acoustic input includes a voice command, processes to respond to the voice command may be initiated (act 430). For example, if the acoustic input is determined to include an explicit voice trigger, the voice response system may be readied to expect one or more voice commands to act upon. If the acoustic input includes an actionable voice command, initiation of the processes to perform the actions needed to respond to the voice command may be invoked.

[0071] According to some embodiments, the at least one first processing stage comprises one or more relatively low power processing stages, and the at least one second processing stage comprises one or more generally higher power processing stages. In this manner, the processing stages that require more processing power may be reserved for circumstances where the at least one relatively low power processing stage is not able to conclude that the acoustic input corresponds to spurious acoustic activity so that further evaluation may be necessary.

[0072] According to some embodiments, the at least one first processing stage comprises one or more VAD processing stages configured to determine whether the acoustic input likely contains voice content or whether the acoustic input more likely contains non-speech acoustic activity. The at least one second processing stage may include one or more speech processing stages configured to determine the content of the speech and/or the meaning of the speech assessed as being present in the acoustic input by the one or more VAD processing stages. In this manner, speech processing stages may be reserved for acoustic input assessed as including voice content, thus conserving power to an extent reasonable by first detecting the presence of speech before

Petitioner Exhibit 1004, Page 51 of 1052

determining its content. Some exemplary VAD and speech processing stages are discussed in further detail below.

[0073] According to some embodiments, the at least one first processing stage is performed on a secondary processor (e.g., a relatively low power digital signal processor (DSP) on the mobile device) and the at least one second processing stage is performed on a primary processor (e.g., a main central processing unit (CPU) of the mobile device). In this manner, the secondary processor may be able to dispense with some appreciable percentage of the received acoustic input without having to engage the primary processor, thus conserving power resources and allocating them on an as-needed basis. This may be particularly useful when the mobile device is operating in a low power mode wherein the primary processor is powered down or at least partially disabled or deactivated. Thus, monitoring the acoustic environment may be achieved without having to frequently wake-up relatively power intensive components of the mobile device to evaluate whether a voice command has been received.

[0074] The above described examples of multi-stage processing may facilitate the implementation of an "always-on" voice response system that can monitor and respond to voice commands while generally minimizing power consumption, both when the mobile device is active and/or when the mobile device is in a low power mode (where reducing power consumption may be or perceived to be more critical or important by a user). Some exemplary processing stages that may be used to implement multi-stage voice response/activation are described below, though aspects related to monitoring acoustic input and evaluating whether the acoustic input includes a voice command are not limited to the examples provided herein, nor are any particular processing stages required as any one or combination of processing stages may be suitable for implementing an "always-on" voice response system.

[0075] When a mobile device is operating in a low power mode, the voice response system may transition the mobile device from the low power mode to an active mode in a number of ways. In general, the voice response system may be configured to keep as many components in a low power mode as possible to reduce the impact on the battery of the mobile device. As such, the voice response system may incrementally or progressively transition the mobile device from a low power mode to an active mode by activating resources only as needed. For example, the voice response system may activate a secondary processor to perform one or more processing stages while leaving

Petitioner Exhibit 1004, Page 52 of 1052

the primary processor in a low power mode until and/or unless the primary processor is needed.

[0076] One or more processors may include a low power state whereby certain processes may be executed without fully activating the processor (e.g., one or more low power clocks may allow for low power operation), and the voice response system may utilize such low power operating states to the extent possible before bringing the processor to a full power state or otherwise rendering the processor fully operational. Some components, such as the display, may remain in low power mode until the voice response system is confident acoustic input includes a voice command. It should be appreciated that there are many suitable approaches for transitioning a mobile device from a low power mode to an active mode, either progressively or more instantaneously, and the manner in which wake-up of a mobile device is performed may rely on the processing strategy used to determine whether acoustic input includes a voice command.

[0077] It should be appreciated that incremental or staged wake-up may also include progressive wake-up of other components on the mobile device (e.g., components other than the one or more processors) as appropriate. For example, a mobile device may include multiple microphones to monitor the acoustic environment of the mobile device. When operating in a low power mode, one or more of the microphones may be powered down or turned off to conserver power, while at least one microphone remains at least partially on so as to continue to monitor the acoustic environment to receive acoustic input when the mobile device is operating in a low power mode. When the voice response system determines that the acoustic input likely includes speech and/or a voice command, the voice response system may turn on one or more additional microphones.

[0078] It should be appreciated that the processing stage at which additional microphone(s) are activated is not limited to any particular stage in the processing. For example, the voice response system may activate one or more additional microphones when one or more VAD processing stages has determined that the acoustic input likely contains speech. The voice response system may activate one or more additional microphones when an explicit voice trigger is detected, or when one or more keywords has been recognized. Additionally, one or more additional microphones may be activated after relatively early processing determines that the acoustic input exceeds certain low level threshold inquiries such as amplitude, duration, energy, etc.

Petitioner Exhibit 1004, Page 53 of 1052

Accordingly, any of the processing stages described herein may trigger activating one or more additional microphones, as the aspects are not limited in this respect.

[0079] Any of a variety of mobile device components capable of providing one or more contextual cues may also be activated as part of a staged or incremental wake-up when the mobile device is operating in a low power mode including, but not limited to, a GPS system, an accelerometer, or a clock to provide location information, motion information and time of day, respectively. Any of various system resources may be activated to, for example, check for recent activity, check for an occurrence of an event, check the battery power, etc. to incrementally wake-up portions of the mobile device on an as-needed basis. Network components may also be activated in a staged wake-up when the mobile device is operating in a low power mode. As discussed above, one or more processing stages may involve transmitting at least a portion of the acoustic input over the network to be processed by one or more network servers, and the network transceiver and/or other components needed to communicate with an appropriate network may be activated while the mobile device is operating in a low power mode. Thus, network resources such as ASR and/or NLP resources may be utilized while the mobile device otherwise remains in a low power mode.

[0080] There are a number and variety of processing stages that may be used to assist in deciding whether acoustic input received by a mobile device includes a voice command or corresponds to spurious acoustic activity. Following below are descriptions of some non-limiting examples of processing stages that may be used alone or in any combination in a processing strategy for detecting when a voice command has been spoken. Other processing stages beside the illustrative processing stages described herein may be used, as the aspects are not limited in this respect.

[0081] A first class of processing stages that may be utilized include VAD processing stages configured to detect the presence of voice content in acoustic input. For example, audio signals receiving from one or more microphones in response to acoustic activity in the environment may be processed to determine if the audio signals contain voice content. Performing one or more VAD processing stages may include evaluating the amplitude of the audio signal generated by the microphone. It is often the case that a user uttering a voice command will speak at least at some minimal volume in the vicinity of the mobile device such that low amplitude audio signals below a given threshold may be regarded as acoustic input arising from background or ambient acoustic

Petitioner Exhibit 1004, Page 54 of 1052

activity. Alternatively or in addition to, the duration of acoustic activity may be evaluated to rule out spikes of acoustic activity or other acoustic activity of a duration too brief to likely correspond to speech.

[0082] VAD processing stage(s) may also include evaluation of the acoustic input to assess whether the acoustic input is characteristic of speech. For example, the energy of acoustic input received by a microphone on the mobile device may be evaluated to assess whether the energy content is suggestive of speech. Spectral slope analysis, evaluation of signal periodicity, cepstral analysis, or any other technique that evaluates one or more properties of the acoustic input to determine whether the acoustic input is characteristic of speech. VAD processing stages may include analyzing the acoustic input for phoneme content to determine the likelihood that the acoustic input includes speech. For example, one or more VAD processing stages may utilize a limited number of stored phoneme models (or a complete set) and compare the acoustic input to the phoneme models to evaluate whether the acoustic input has phonemic content. For example, if portions of the acoustic input match one or more of the phoneme models utilized, it may be ascertained that the acoustic input likely includes phonemic content characteristic of speech.

[0083] The confidence such a processing stage has that acoustic input includes phonemic content, and thus that the acoustic input include voice content, may relate to how well portions of the acoustic input match the set of phoneme models utilized and/or how many matches can be found. To reduce the amount of processing required, one or more such processing stages may be configured to conclude that voice content is present when a certain number of matches are found and/or the quality of one or more matches are sufficient to strongly suggest speech, discontinuing further processing in this respect. One or more phone loop techniques may be used to ascertain the phonemic content of the acoustic input. Such phonemic processing may be performed relatively inexpensively as a precursor to ASR and using less computing power than required to recognize words. According to some embodiments, results of phoneme analysis may be propagated to subsequent ASR processing stages.

[0084] In embodiments that include one or more VAD processing stages, the processing stages may be performed in sequence (e.g., a low level amplitude check may be performed as a threshold inquiry, followed by one or more evaluations of the audio signal characteristics if the amplitude is sufficient to suggest it may be associated with

Petitioner Exhibit 1004, Page 55 of 1052

voice), in parallel (e.g., one or more signal characteristic processing stages may be performed relatively simultaneously or as part of the same processing stage, and the results may together be evaluated to assess whether the acoustic input is suggestive of speech), or a combination of both. According to some embodiments, one or more VAD processing stages may be utilized to determine whether it is believed that acoustic input contains speech, and if not, further processing of the acoustic input may be terminated.

[0085] Another class of processing stages that may be performed to determine whether acoustic input includes a voice command are speech processing stages, which herein refer to processing stages that evaluate or assess the content of speech contained in an acoustic input (e.g., as opposed to determining whether it is believed voice content is present, as performed by typical VAD processing stages). Evaluating the content of acoustic input may include, but is not limited to, automatic speech recognition (ASR) (e.g., determining what words are present in an acoustic input containing speech), classification (e.g., categorizing speech), language understanding (e.g., ascertaining the semantic content of speech) and/or speaker verification (e.g., determining whether speech was uttered by a particular speaker), further details of which are discussed below. Speech processing stages may be used in any number and in any combination, as the aspects are not limited in this respect.

[0086] Limited vocabulary ASR is one example of speech processing that may be suitable in determining whether a voice command has been spoken. As a general principle, the processing required to recognize speech increases as the size of the vocabulary of words that can be recognized increases. Accordingly, using a restricted vocabulary may allow for ASR to be performed quickly and efficiently while consuming relatively little power.

[0087] Limited vocabulary ASR may be used to perform explicit voice trigger detection. For example, an exemplary speech processing stage may include performing ASR using a vocabulary restricted to the words in the explicit voice trigger phrase (which may include as few as a single word.). For example, for the explicit voice trigger "Hello, Dragon," the vocabulary may be restricted to the two words "Hello" and "Dragon." By limiting the vocabulary to the words permitted in an explicit voice trigger, ASR may be performed using little processing to assess whether the acoustic input includes a voice command (e.g., whether the acoustic input includes the explicit voice trigger).

[0088] Limited vocabulary ASR may also be used in contexts other than detecting an explicit voice trigger, alternatively or in addition to explicit voice trigger detection. For example, limited vocabulary ASR may be performed using a restricted vocabulary having a desired number of key words that are frequently uttered people when speaking a voice command. For example, terms such as "what," "where," "how," etc., may be frequently used when speaking a voice query. Action words such as "search," "schedule," "locate," "call," "contact," "remind," etc., may also be common words uttered when speaking a voice command. It should be appreciated that any word deemed suggestive of a voice command may be included the limited vocabulary to facilitate relatively fast, relatively low power ASR to obtain information about whether acoustic input includes a voice command.

[0089] A limited vocabulary in this respect may also be built, learned or modified based on a user's behavior. For example, words spoken by a user in past voice commands may be added to a limited vocabulary. According to some embodiments, the words spoken by a user in voice commands recognized as such may be stored and each time a given word is again spoken by the user, the use count for that word may be incremented. The time at which a given word is spoken may also be recorded so that frequency of use may be determined. In this manner, a record of how many times and/or how frequently a user speaks particular words when uttering voice commands may be obtained. When a word that does not appear in the limited vocabulary reaches some threshold count and/or is spoken frequently enough (e.g., spoken enough times within a given interval of time), the word may be included in the limited vocabulary. One or more words included in the limited vocabulary may be removed if it is determined that a user does not use them regularly or at all when speaking voice commands. Other suitable techniques may be used to modify a limited vocabulary based on a user's behavior to generally tailor the limited vocabulary to the user, as this concept is not limited to any particular technique or set of techniques.

[0090] According to some embodiments, limited vocabulary ASR may be followed by more comprehensive ASR (e.g., large vocabulary ASR, generally unrestricted ASR, etc.). For example, limited vocabulary ASR may detect that an explicit voice trigger was uttered and invoke more comprehensive ASR to recognize other words present in the acoustic input, or to recognize words present in acoustic input received subsequent to detecting the explicit voice trigger. According to some

Petitioner Exhibit 1004, Page 57 of 1052

embodiments, limited ASR may be used to recognize one or more key words suggestive of a voice command and, if it is determined that the recognized word(s) provide sufficient evidence of a voice command, more comprehensive ASR may be performed to recognize the remainder of the speech contained in the acoustic input, either to confirm that a voice command is present or to prepare for responding to a voice command, or both.

[0091] According to some embodiments, generally unrestricted ASR (e.g., ASR performed without restricting or limiting the available vocabulary of the ASR component located on the mobile device) may be performed without first performing limited vocabulary ASR, as the aspects are not limited in this respect. Speech processing stages may also include utilizing ASR resources not available on the mobile device. For example, some embodiments include a speech processing stage that transmits the acoustic input wirelessly to a network where it can be recognized using a desired ASR component, for example, located on one or more servers connected to the network. This may allow for relatively computationally expensive ASR to be performed without incurring these processing costs on the mobile device itself. Limited ASR may also be performed on a network server instead of on the mobile device, as the use of local or network resources is not limited to any particular processing stage. In general, any one or combination of ASR processing stages may be utilized to assist in determining whether acoustic input includes a voice command.

[0092] Speech processing stages that may be utilized to evaluate whether a voice command has been spoken include one or more statistical classification stages to categorize the acoustic input as either corresponding or not corresponding to a voice command. Any type of statistical classifier may utilized and in any combination, including but not limited to binary or multiclass classification, linear classifiers, hidden Markov models (HMMs), Bayesian networks, support vector machines, neural networks, decision trees, etc. According to some embodiments, training data is utilized to cluster utterances into classes (e.g., voice commands and speech that does not correspond to a voice commands). Words recognized in acoustic input may be compared to the trained clusters to classify the utterance as either likely being associated with a voice command or likely not.

[0093] Such classification may be a binary decision, or the classifier (e.g., clusters) may include more numerous classifications. According to some embodiments,

Petitioner Exhibit 1004, Page 58 of 1052

a binary classification may be used to assess whether speech corresponds to a voice command or not and, when it is determined that speech corresponds to a voice command, further analysis may be performed to understand the content of the voice command (e.g., one or more natural language processing stages may be performed). Alternately, a multiclass classification may be performed to better understand how the voice command should be responded to. Any of numerous suitable statistical classifiers may be used to determine whether speech corresponds to a voice command or not, and statistical classifier processing stages are not limited to any one, particular or combination of techniques.

[0094] Speech processing stages that may be utilized to evaluate the content of input also include one or more natural language processing stages to ascertain the semantic meaning of speech recognized using one or more ASR processing stages. NLP stages may be used either to evaluate whether speech contained in acoustic input corresponds to a voice command, or to determine the meaning of the voice command and/or intent of the user so that the voice command can be carried out.

[0095] It should be appreciated that any processing stage configured to evaluate whether acoustic input includes a voice command (e.g., any of the exemplary processing stages described herein) may be performed on the mobile device or remotely via a network connection to one or more network servers). In some circumstances, delegating one or more processing stages to network resources may further generally optimizing false positive and false negative rates while generally minimizing power consumption and/or may bring to bear resources not available on the mobile device.

[0096] As discussed above, generally optimizing false positive and false negative rates to ensure that a mobile device is perceived as responsive to voice commands, while avoiding too frequently attempting to respond to spurious acoustic activity may be a consideration when implementing "always-on" functionality. This may have increased importance when the mobile device is operating in a low power mode where unnecessarily waking-up the mobile device may not only be particularly frustrating for the user from the perspective of loss of battery power, but may become annoying to the user who may then prefer to disable this functionality.

[0097] The inventors have appreciated that considering one or more contextual cues may improve false positive and/or false negative rates. In particular, the inventors have appreciated that considering other factors other than the content of received

Petitioner Exhibit 1004, Page 59 of 1052

acoustic input may improve accurately detecting whether the acoustic input includes a voice command when used in conjunction with techniques for evaluating the acoustic input itself (e.g., in conjunction with any of the above described processing stages and techniques).

[0098] FIG. 5 illustrates a method of using at least one contextual cue to assist in determining whether acoustic input includes a voice command. Method 500 may be similar in some respects to methods 200, 300 or 400 in that acoustic information is obtained (510) and then processed in some manner to evaluate whether the acoustic input includes a voice command (e.g., using any one or combination of techniques described herein). In act 520, at least one contextual cue is used to assist in determining whether the acoustic input includes a voice command. Example contextual cues include, but are not limited to, the time of day, the location of the mobile device, motion of the mobile device, recent activity on the mobile device, biometric information, status of the power source, information about the acoustic environment, etc., further details of which are provided below.

[0099] Contextual cues may be used in numerous ways to influence the determination of whether acoustic input includes a voice command. For example, one or more contextual cues may increase or decrease a confidence value associated with one or more processing stages. As illustration, movement of the mobile device may influence the decisions of one or more VAD processing stages when it is suggestive of a user having picked up the mobile device, as such action may indicate that proximately received acoustic input includes voice content.

[00100] Contextual cues may be used to determine which processing stages are performed and in what order (e.g., the selection and/or sequence of processing stages may be based, at least in part, on evaluating one or more contextual cues). For example, a location cue indicating that the mobile device is located in a vehicle may be used to initiate a particular noise cancellation processing stage that would not otherwise be performed if the mobile device is determined to be located in the office or home or other location having a different acoustic profile. A low battery condition may be used to favor selection of specific (or additional) low power processing stages and/or discourage selection of higher power processing stages. In some instances, one or more contextual cues may be used to terminate further processing altogether (e.g., a determination that received speech was not uttered by the user of the mobile device may trigger ceasing

Petitioner Exhibit 1004, Page 60 of 1052

further processing). Additionally, a contextual cue may be used to initiate obtaining one or more other contextual cues. Contextual cues may be used in any suitable way to assist in determining and/or influencing the evaluation of whether an acoustic input includes a voice command, as the aspects are not limited in this respect.

[00101] Following below are more detailed descriptions of various exemplary contextual cues. Examples of specific contextual cues are provided herein for illustrative purposes as any contextual cue may be utilized in any suitable combination and are not limited to the type or combinations explicitly described herein.

[00102] The inventors have appreciated that the time of day may be an indicator of whether acoustic input includes a voice command, and that the time in which acoustic input was received may be used to influence the evaluation of whether the user has uttered a voice command. For example, acoustic input received when people normally are sleeping may suggest that the acoustic input does not include a voice command. This information may be used to discourage activating the voice response system unless other indicators suggest more strongly that the acoustic input in fact does include a voice command. Acoustic input received at other times of day may suggest that it is more likely that the acoustic input includes a voice command, and this information may be used to influence the evaluation (e.g., to bias the decision as to whether a voice command has been uttered to some desired extent).

[00103] According to some embodiments, the system may record the time of day that a given user has uttered voice commands in the past. By doing so, a distribution of when a particular user has issued voice commands can be compiled. Such a distribution indicates when a user is more and less likely to issue voice commands. As such, when acoustic input is received, the time may be obtained and compared to the distribution to see whether receipt of the acoustic input corresponds to a time when the user frequently or infrequently utters voice commands. This comparison may be used to influence the determination of whether the acoustic input likely includes a voice command.

[00104] As an example, if the time of receipt of the acoustic input corresponds generally to a time of day when the user frequently utter voice commands, the determination may be biased towards finding a voice command to a desired extent (e.g., in proportion to how frequently the user utters voice command at the current time of day). On the other hand, if the current time corresponds to a time of day when the user infrequently utters voice commands, the comparison may be used to influence the

Petitioner Exhibit 1004, Page 61 of 1052

evaluation to discourage, to an extent desired, the conclusion that the acoustic input includes a voice command. It should be appreciated that a history of the times of past voice commands may be collected and utilized in other ways to influence the determination of whether acoustic input includes a voice command, or historical data may not be utilized, as such techniques are not limited in this respect.

[00105] According to some embodiments, the voice response system may obtain the current time when acoustic input is received and access a user's calendar to check whether the user may have an appointment, meeting or event scheduled that would make it more or less likely that the user would be uttering a voice command. The voice response system may be capable of accessing a calendar or schedule that allows the user to specify certain times when the user does not want "always-on" functionality to be operational, but would like the mobile device to enable and disable this functionality automatically. Thus, when acoustic input is received, the system may initially check to see if the current time is within an interval of time that the user has indicated that the user would like "always-on" functionality disabled before proceeding with further processing. The time at which acoustic input is received may be utilized in other ways to influence whether a user has uttered a voice command, as the aspects are not limited to the examples described herein.

[00106] Many mobile devices are equipped with one or more components that can detect motion of the mobile device, typically by sensing acceleration (e.g., using a gyroscope or other component that responds to acceleration forces). The inventors have appreciated that such motion detectors may be used as a contextual cue to facilitate improved performance with respect to detecting a voice command. For example, users often lift a mobile device or otherwise bring the mobile device closer to their mouth just prior to or concurrent with uttering a voice command, and this action may be used to facilitate improved evaluation of acoustic input. In particular, response from a motion detector can be used as a cue that acoustic input is likely to include a voice command (e.g., when acoustic input is received in close temporal proximity to a response from a motion detector, it may be assumed more likely that the user spoke a voice command).

[00107] Motion detector (e.g., accelerometer) response may be incorporated into the evaluation by using motion detector response in conjunction with, to supplement, augment or otherwise influence one or more processing stages (or the results of one or more processing stages) to bias the evaluation in favor of the acoustic input including a

Petitioner Exhibit 1004, Page 62 of 1052

voice command. For example, motion response may be used to reduce the threshold of one or more VAD processing stages, relax the requirements of one or more ASR and/or natural language processing stages, provide a vote in favor of the acoustic input including a voice command in a voting scheme, etc. Response from one or more motion detectors may be used in any suitable way to influence the determination of whether acoustic input includes a voice command, as the aspects are not limited in this respect.

[00108] Many mobile devices are equipped with location determination capabilities such as GPS. The inventors have appreciated that the location of the mobile device may be used as a cue to influence the determination of whether acoustic input includes a voice command. For example, a user may more frequently issue voice commands while driving then when at the office or home, or vice versa. The voice response system may incorporate this information into evaluating of acoustic input to influence the decision regarding whether a voice command has been spoken. The system may also be configured to be more inclined or more reluctant to conclude that acoustic input includes a voice command in certain locations, independent of how frequently a user utters voice commands in a specific location.

[00109] A voice response system may have stored information regarding general areas associated with home, office or other locations such that when the location determination component, such as a GPS, generates coordinates located within these areas, the voice response system can ascertain that the mobile device is located in a known location. With respect to a vehicle, the voice response system may be configured to recognize when the location of the mobile device is changing rapidly and use this information to infer that the mobile device is located in a moving vehicle.

[00110] Determining location is not limited to using an electronic component that determines the coordinates of the mobile device, and location of the mobile device may be ascertained through inference using other cues. For example, the voice response system may conclude that the mobile device is presently located in the user's vehicle by detecting connection to a hands free docking station, detecting the presence of a wireless device (e.g., a Bluetooth device) located in the vehicle, or determining that the mobile device is capable of communicating with one or more of the vehicle's on-board systems. The voice response system may also ascertain that the mobile device is located in a particular location by detecting the presence of a wireless access point (e.g., the user's home network, office LAN, etc.).

Petitioner Exhibit 1004, Page 63 of 1052

[00111] Techniques that generally infer the location of the mobile device due to proximity, connection and/or ability to communicate with other devices, components and/or networks may be used to ascertain the mobile device's location wherever suitable and wherever such inferences can be made, as the aspects are not limited in this respect. In general, location may be used in any of a variety of ways to assist in determining whether acoustic input includes a voice command, as the aspects are not limited in this respect.

[00112] According to some embodiments, the voice response system may learn a user's behavior regarding where the user is most likely to speak a voice command. For example, when it has been determined that the user has spoken a voice command, the location may be recorded and a distribution obtained that captures where the user frequently utters voice commands. When acoustic input is received, a present location of the mobile device may be compared to this distribution to either bias in favor of or discourage the determination that the acoustic input includes a voice command based on whether the mobile device is located where a user frequently or infrequently speaks voice commands. Other techniques for learning a user's habits with respect to uttering voice commands may be used, as the aspects are not limited in this respect.

[00113] The occurrence of an event and/or recent activity on the mobile device may be used as a contextual cue to influence the decision regarding whether acoustic input includes a voice command. For example, certain events on a mobile device may be likely to trigger a voice response from the user. For example, when an alarm on the mobile device goes off, a user may be interested in addressing the alarm via voice commands such as "Snooze alarm," "Turn alarm off," or a command to set the alarm for a later time (e.g., "Set alarm for 8 o'clock," "Set alarm for an hour from now," etc.). After a calendar alert has activated, the user may be inclined to take some sort of action such as "Show me my Calendar," or in response to a reminder to call John, the user may instruct the mobile device to "Call John Doe," to initiate a telephone call. In response to an incoming telephone call, the user may want to handle the interaction via voice with instructions such as "Answer call," "Send to voicemail," "Mute phone," etc. A user may want to respond via voice when a text is received by commanding the mobile device to "Respond to latest text," or may want to respond to an alert that a voicemail was just received by speaking the command "Listen to voicemail."

Petitioner Exhibit 1004, Page 64 of 1052

[00114] A user's own activity on a mobile device may also be used to influence a decision about whether acoustic input includes a voice command. For example, if a user opens a text or email, the user may indicate the desire to speak the information by commanding the mobile device to "Take dictation." Opening a calendar or scheduling application make be cue that the user may want to schedule something using voice (e.g., "Schedule lunch with Jane at noon tomorrow"). In addition, how recently a user interacted with a mobile device may also be considered as a contextual cue. For example, the length of time since the user last activated the mobile device, how long ago the user last uttered a voice command, how long ago the user performed one or more tasks on the mobile device, etc., may also be considered as a contextual cue to influence the evaluation of one or more processing stages.

[00115] The above described events and activities are merely exemplary and any event or combination of events or activity on the mobile device may be used as a contextual cue. The term event information is used herein to describe information obtained about an event occurring on the mobile device and recent activity enacted on the mobile device. Thus, according to some embodiments, a voice response system monitors one or more events and/or recent activity on the mobile device and uses event information obtained as a result to influence the evaluation of acoustic input received proximate in time to the event information. Event information may be utilized to either bias in favor or discourage a determination that acoustic input includes a voice command.

[00116] As discussed above, conserving power may be an important consideration when implementing "always-on" functionality. Accordingly, in some embodiments, remaining battery power may be used to assist in the optimal selection of processing stages and/or in choosing what order the processing stages should be performed. For example, if the mobile device has sufficient battery power (e.g., the mobile device has been recently charged), the voice response system may be willing to consume more power to achieve lower false positive and/or false negative rates. As such, the choice of processing stages may be selected to achieve this result. However, when battery power is low, power conservation may become increasingly important. In such circumstances, an indication of a low battery state may be used to discourage passing acoustic information on for further processing such that additional power consumption is incurred only in situations where the confidence is very high that the acoustic input includes a voice

Petitioner Exhibit 1004, Page 65 of 1052

command (e.g., this may be achieved by modifying thresholds in one or more processing stages when a low battery condition is detected).

[00117] Furthermore, if the system detects that the mobile device is plugged into a power source such that the choice of processing stages is generally not or significantly less constrained with respect to power consumption, the voice response system may select processing stages that generally will determine whether there is a voice command with the best accuracy, while relaxing constraints on the amount of processing power used to make the determination.

[00118] The acoustic environment of the mobile device may also be used as a contextual cue. For example, when acoustic input has been received and assessed as resulting from spurious acoustic activity, the acoustic input may be stored as acoustic information reflecting the acoustic environment and later referenced to assist in evaluating subsequently obtained acoustic input. According to some embodiments, the stored acoustic information, or information derived therefrom (e.g., spectral information, salient components believed to correspond to noise, etc.) may be compared to or subtracted from received acoustic input to eliminate at least portions of the background acoustic environment. The acoustic information may also be obtained by periodically sampling the acoustic environment to obtain information that can be processed and stored as a reference with respect to the acoustic environment of the mobile device. Acoustic information regarding the environment may also be used to determine a location of the mobile device by comparing acoustic information to stored templates of environments in which the mobile device is frequently located (e.g., a car).

[00119] Voice has been used as a biometric signature to facilitate verifying or authenticating the identity of a speaker electronically. Techniques for performing such speaker recognition often utilize a stored "voice print" of the speaker which can be compared to a received audio signal to assess whether the characteristics of the audio signal match those captured by the voice print. A voice print is typically comprised of one or more characteristics that have a facility for distinguishing one speaker from another. When acoustic input is received, one or more characteristics may be extracted and compared to the voice print to assess whether it is believed the acoustic input came from the same speaker from which the voice print was obtained.

[00120] Speaker recognition techniques may be used as part of the process of determining whether acoustic input includes an actionable voice command. According

Petitioner Exhibit 1004, Page 66 of 1052

to some embodiments, the voice response system may be configured to respond only to voice commands spoken by a particular user of the mobile device (e.g., the owner). As such, when acoustic input has been determined to likely contain speech (e.g., using one or more VAD techniques), the acoustic input may undergo speaker recognition to determination whether the speech came from the user or from one or more other speakers. The system may be configured to disregard the acoustic input if it is determined that it did not originate from the specific user, whether it includes a voice command or not. In some embodiments in which the system is capable of performing speaker recognition, this capability may be enabled and disabled by the user to allow for other speakers to speak voice commands that can be acted upon.

[00121] Mobile devices may include multiple microphones. Such configurations may be utilized to compute the direction from which a given user is speaking. For example, time delay of arrival (TDOA) techniques may be used to compute the direction from which a given user is speaking based, at least in part, on the delay between the times a user's speech is received at two different microphones on the mobile device, provided there is some sufficient distance between the microphones on the mobile device. For example, transducers 130 in FIG. 1A may include at least one microphone located on the upper portion and on the lower portion to provide a spread sufficient to perform TDOA computations.

[00122] When a given user speaks, the direction from which the speech is incoming may be computed. It should be appreciated that any technique may be used to compute a direction of the user, as there are no limitations in this respect. When acoustic input is received by the mobile device, the direction of the acoustic information may be computed and compared to the stored direction. If the acoustic input is arriving from a different direction, the voice response system may disregard the acoustic input as not including a voice command from the given user. Alternatively or in addition to, the stored direction may be used to filter out acoustic information received from other directions such that acoustic input includes more signal from the general direction of the user to improve subsequent processing of the acoustic input.

[00123] It should be appreciated that the above described contextual and other cues are merely exemplary and other contextual cues or cues of a different nature may be used to assist in detecting a voice command by influencing the evaluation of acoustic input received from the acoustic environment of the mobile device, as the aspects are not

Petitioner Exhibit 1004, Page 67 of 1052

limited for use with any particular contextual cue or combination of contextual cues. Contextual cues may be used at any stage of processing where information that suggests an acoustic input is more or less likely to include a voice command may be useful.

[00124] FIG. 1D illustrates an example configuration of system components internal to a mobile device, such as mobile device 100 illustrated in FIG. 1A. Exemplary system components of a mobile device may include a primary processor 115, a secondary processor 125 and an audio codec 105, all illustrated for convenience and clarity of illustration as being interconnected via a common bus 155. It should be appreciated that a given mobile device may include more or fewer processing components interconnected in any of numerous ways using any type of interconnection, and the particular configuration illustrated in FIG. 1D is provided to demonstrate some concepts related to generally low power "always-on" functionality, in accordance with some embodiments.

[00125] It should be appreciated that any of the techniques described herein may be implemented on a mobile device having a single processor (or performed on a single processor even when multiple processors are available), and the following discussion with respect to FIG. 1D is provided to describe exemplary techniques for exploiting the processing capabilities of mobile devices that have multiple processors. As such, the configuration in FIG. 1D is used merely to illustrate a mobile device having a primary and secondary processor that are capable of processing information and of exchanging information via some type of interconnection or series of interconnections or couplings and not to limit the type of mobile device on which techniques described herein may be implemented. Techniques described herein with respect to monitoring the acoustic environment of a mobile device to detect voice commands may be used with any suitable system configuration having any number of processing components interconnected in any suitable manner.

[00126] Processor 115 may be the system's primary or main processor and generally operates as the central processing unit for the mobile device and, therefore may include a high speed processor that may consume relatively large amounts of power when operating (e.g., when the clock tree is enabled and the processor is executing instructions). Accordingly, when a mobile device enters a low power mode (e.g., powers down into a sleep or idle mode), the primary processor may be powered down to a desired extent (e.g., one or more clocks in the clock tree may be disabled to conserve

Petitioner Exhibit 1004, Page 68 of 1052

power). When the mobile device wakes-up (e.g., enters an active mode), the primary processor may be activated or powered up to perform processing tasks that are needed.

[00127] Processor 125 may be a secondary processor such as a digital signal processor (DSP), co-processor and/or specialized processor to handle some portion or aspect of the processing tasks performed by the mobile device, to assist the primary processor in performing certain functionality, to handle certain specialized processing tasks, etc. For example, the secondary processor may include a digital signal processor configured to process signals from one or more wireless communications components 160, for example, to handle at least some processing involved in a cellular communications. The secondary processor may include a coprocessor of the primary processor or a specialized processor such as a graphics processing unit (GPU) to handle display capabilities.

[00128] The secondary processor may be a lower power processor than the primary processor in that less power may consumed by the secondary processor when enabled and executing instructions. In some embodiments, the secondary processor can be configured to operate at a lower power and/or may be simpler and require less power to power-up and/or keep enabled/powered-up, or otherwise be generally attractive from a power standpoint. In some embodiments, it may be simpler, more efficient and/or less power intensive to implement certain processing stages on the secondary processor, or there may be other reasons besides the operating characteristics of the secondary processor that render the secondary processor advantageous from either a power standpoint, an efficiency standpoint, or both.

[00129] As discussed above, the inventors have recognized that utilizing a secondary processor to perform one or more processing stages when determining whether acoustic input includes a voice command may afford lower power "always on" functionality. FIG. 6 illustrates a method of utilizing a system configuration of a mobile device that includes a primary and secondary processor to monitor the acoustic environment to detect voice commands, in accordance with some embodiments. In act 610, the acoustic environment of the mobile device is monitored and acoustic input received. The acoustic input may be processed by one or more components of the mobile device, for example, one or more transducers 130 (e.g., a microphone) may detect acoustic activity and process the acoustic input, audio codec 105 may convert the

Petitioner Exhibit 1004, Page 69 of 1052

acoustic input from analog to digital, perform one or more operations on the acoustic input, etc.

[00130] In act 622, a secondary processor performs at least one first processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command or likely corresponds to spurious acoustic activity. If after performing the at least one first processing stage, it is determined that the acoustic input corresponds to spurious acoustic activity, the acoustic input may be ignored and further acoustic input may be received for evaluation (e.g., act 610 may be repeated or continue to be performed).

[00131] The at least one first processing stage may include one or more VAD processing stages and/or process one or more contextual cues to assess whether the acoustic input is likely to include voice and, if the acoustic input is determined to be the result of spurious acoustic activity, the voice response system may discontinue further processing of the acoustic input. The at least one first processing stage may also include one or more speech processing stages, provided the secondary processor has the processing power and/or functionality implemented to do so. For example, the secondary processor may be configured to perform limited vocabulary ASR on the acoustic input such as detecting an explicit voice trigger or keyword spotting.

[00132] By first performing the at least one first processing stage on the secondary processor, it may be possible to rule out substantial amounts of acoustic input without needing to utilize the primary processor. When the mobile device is operating in a low power mode with the primary processor generally inactive or disabled, it may be possible to dispense with appreciable amounts of acoustic input without needing to "wake-up" the primary processor, thus achieving reduction in the power requirements of implementing "always-on" functionality. As such, the primary processor may be reserved for circumstances where additional processing is required to evaluate the acoustic input (e.g., when the secondary processor cannot itself reach a conclusion with enough certainty).

[00133] In embodiments that check for explicit voice triggers, and wherein the secondary processor is configured to perform such explicit voice trigger detection, it may be determined that acoustic input includes a voice command without engaging the primary processor or, if in low power mode, without enabling, activating and/or waking up the primary processor at all. The voice response system may then be readied to

Petitioner Exhibit 1004, Page 70 of 1052

process subsequent acoustic input expected to follow the explicit voice trigger, or to further process the acoustic input if it includes an actionable voice command in addition to the explicit voice trigger. The further processing may engage the primary processor to assist in understanding the voice command and/or to carry out the directives of the voice command.

[00134] In act 624, if after performing the one or more first processing stages, it cannot be concluded that the acoustic input corresponds to spurious acoustic activity (act 623), at least one second processing stage may be performed on the acoustic input by the primary processor to further evaluate whether the acoustic input includes a voice command. The at least one second processing stage may include one or more processing stages that require more processing power, more complex processing, and/or capabilities or functionality that the secondary processor is not configured to perform. The at least one second processing stage may include one or more speech processing stages that recognize the content of speech in the acoustic input and/or ascertain the semantic content of speech in the acoustic input.

[00135] While it may be advantageous to reserve the primary processor for circumstances where the secondary processor cannot reach a conclusion and/or where more computing power, resources or functionality are needed, it should be appreciated that any of the exemplary speech processing stages described above, either alone or in any combination, may be performed by the primary processor to ascertain whether the acoustic input includes a voice command. When it is determined that the acoustic input includes a voice command, further processing needed to respond to the voice command may be initiated (act 640).

[00136] According to some embodiments, the at least one first processing stage and/or the at least one second processing stage includes connecting to a network and transmitting at least some of the acoustic input over the network to be processed by one or more servers connected to the network. For example, the acoustic input may be sent to an ASR server to recognize the speech and/or to a natural language understanding server to ascertain the semantic meaning of the acoustic input. While connecting to the network and transmitting the acoustic input over the network consumes power (e.g., when in a low power mode, the appropriate wireless transceiver may need to be activated and the necessary processing components (e.g., a wireless codec) enabled to provide the acoustic input to one or more network servers), once received by the one or more servers,

Petitioner Exhibit 1004, Page 71 of 1052

the acoustic input can be processed using network resources without further exhausting the battery power of the mobile device.

[00137] Thus, utilizing network resources to perform one or more processing stages may result in a net savings in battery power of the mobile device. That is, in some circumstances, utilizing the computing resources of the network may justify incurring the power consumption needed to transmit the acoustic input over the network and receive relevant results back (e.g., the power savings resulting from utilizing network resources may justify the processing power needed to connect and maintain a network communication). Network resources may be utilized to perform one or more processing stages on the acoustic input when the mobile device is operating in an active mode, when the mobile device is operating in a low power mode, or both. In an active mode, the components needed to perform the network communication may already be active and the overhead in this respect may be relatively minimal. When in a low power mode, the voice response system may activate only those components needed to perform the network communication and allow the remainder of the components that have been powered down to remain in a low power mode to limit the power consumption to the extent possible.

[00138] While it may be advantageous in many circumstances to perform certain processing stages on the mobile device (e.g., one or more VAD processing stages and/or limited vocabulary ASR) and utilize network resources for more computationally intensive processing stages such as more complex ASR and/or NLP, there are no limitations in this respect, and acoustic input may be transmitted over the network at any stage of processing and network resources may be utilized to perform any one or combination of processing stages suitable for a given processing strategy.

[00139] According to some embodiments, one or more VAD processing stages are performed on the secondary processor and, if voice is detected, the primary processor is engaged to perform one or more speech processing stages. It should be appreciated that while, from a power standpoint, it may be advantageous to perform VAD processing stages on the secondary processor, one or more VAD processing stages may be performed on the primary processor, as the aspects of utilizing two or more processors is not limited to performing any particular processing stage or combination of processing stages on the secondary processor or the primary processor.

Petitioner Exhibit 1004, Page 72 of 1052

[00140] The secondary processor may be utilized in any suitable manner to implement a processing strategy that facilitates achieving relatively low power "always-on" functionality while having satisfactory false positive and false negative rates. Some processing strategies may be guided by the principal that further processing of acoustic input should only be performed and/or additional system components engaged or activated only when necessary to reach a conclusion regarding whether the acoustic input includes a voice command or is the result of spurious acoustic activity. However, a secondary and primary processor resources may be utilized in any manner, as the aspects are not limited in this respect.

[00141] FIG. 7A and 7B illustrate an example of a voice response system 750 configured to monitor the acoustic environment of a mobile device 700 to receive an acoustic input and determine whether the acoustic input includes a voice command, without requiring the user to speak an explicit voice trigger. In particular, a user 790 may speak the command "What appointments do I have scheduled for tomorrow?" to a mobile device 700 without first needing to alert the voice response system using an explicit voice trigger. According to some embodiments, the voice response system may be responsive to such voice commands (i.e., voice commands spoken without the user uttering an explicit voice trigger) when the mobile device is in an active mode, a low power mode, or both.

[00142] Mobile device 700 may be the user's smartphone or other mobile device configured to receive voice input from the user. Mobile device 700 may be capable of connecting to one or more networks 795, including, but not limited to a cellular network, one or more LANs, one or more WANs, the Internet, etc. Voice response system 750 may include hardware, software or a combination of both and is at least partially implemented on the mobile device 700 to monitor the acoustic environment of mobile device 700 to receive acoustic input 705, and determine whether acoustic input 705 includes a voice command, examples of which are shown in FIG. 7B for the exemplary voice command shown in FIG. 7A.

[00143] Voice response system 750 may process acoustic input 705 using one or more VAD processing stages 710 to determine whether the acoustic input includes speech content or corresponds to non-speech acoustic activity. As discussed above, VAD processing stages may include any number of processing stages that evaluate properties of the acoustic input to assess whether the properties are characteristic of

Petitioner Exhibit 1004, Page 73 of 1052

speech, including but not limited to initial threshold inquiries regarding amplitude, energy and/or duration of the acoustic input, analysis of signal characteristics such as spectrum analysis, periodicity, spectral slope, cepstral analysis, etc.

[00144] According to some embodiments, the number and type of VAD processing stages are selected to minimize the number of processing stages needed to either conclude that the acoustic input includes voice content, or conclude that the acoustic input is the result of non-speech acoustic activity. For example, if a given VAD processing stage concludes with sufficiently high confidence that the acoustic input includes voice content, further VAD processing stages may be omitted in favor proceeding directly to one or more speech processing stages to evaluate the content of the speech. If a given VAD processing stage evaluates the likelihood of the acoustic input as being sufficiently low, further processing may be discontinued and the acoustic input categorized as non-speech input.

[00145] However, if a given VAD processing stage cannot conclude one way or another, further VAD processing stages may be performed to reach a conclusion as to whether the acoustic input includes voice content or not. The result of a given VAD processing stage may determine which subsequent VAD processing stage is selected. For example, results from certain types of VAD processing stages may suggest which subsequent VAD processing stage should be performed. In this manner, the number, type and order of processing stages may be dynamically determined to facilitate relatively low power VAD processing of the acoustic input. According to some embodiments, multiple VAD processing stages may be performed in parallel and the results from the parallel stages compared to evaluate whether the acoustic input includes voice content. For example, a voting scheme, which may be weighted or not, may be used based on results from multiple VAD processing stages.

[00146] In the example illustrated in FIG. 7A and 7B, VAD processing stage(s) 710 determine that acoustic input 705 includes voice content and the voice response system further evaluates acoustic input 705 using one or more speech processing stages 720 to determine whether the acoustic input includes a voice command. As discussed above, speech processing stages may include ASR, classification (e.g., using one or more statistical classifiers), NLP, etc. For example, according to some embodiments, acoustic input 705 may undergo limited vocabulary ASR to perform keyword spotting, any technique for which may be used to identify whether acoustic input 705 contains any

Petitioner Exhibit 1004, Page 74 of 1052

words deemed suggestive of a voice command and/or to identify words needed to perform classification. Other ASR techniques may be utilized depending on the processing strategy being used to recognize one or more words in speech contained in the acoustic input. If no words can be recognized with sufficient confidence, it may be the case that VAD processing stage(s) 710 incorrectly concluded that the acoustic input includes speech. In such an event, the voice response system may discontinue further processing to avoid consuming further power.

[00147] When one or more words have been recognized using ASR, the recognized words may undergo classification. At this stage in the processing, acoustic input 705 may include one or more recognized words as a result of performing one or more ASR processing stages, and one or more statistical classifiers may be performed on acoustic input 705 to categorize the acoustic input as corresponding or not corresponding to a voice command based, at least in part, on the one or more recognized words. Other speech processing stages may be utilized, either alternatively or in addition to the exemplary speech processing stages described above, as the aspects are not limited in this respect.

[00148] In the exemplary processing sequence in FIG. 7B operating on acoustic input 705, the one or more statistical classifiers may have concluded that the speech in acoustic input 705 corresponds to a voice command. Having so concluded, certain portions of the mobile device may be activated in situations in which the mobile device 700 is operating in low power mode, some exemplary wake-up techniques of which are discussed in connection with FIG. 8A and 8B below. Generally speaking, essential components of the mobile device needed to efficiently determine the meaning of the voice command and to respond to the directives of the voice command may be engaged or activated pursuant to the determination that acoustic input 705 includes a voice command. However, voice response system wait to activate one or more components of the mobile device 700 until the meaning of the voice command is evaluated to better understand what components may be required to carry out the directives of the voice command.

[00149] After concluding that a voice command is present, voice response system 750 may perform NLP stages 730 to evaluate the semantic content of the acoustic input to understand what the user intended the voice response system to do when speaking the voice command. In this respect, the acoustic input may be fully recognized to the extent

Petitioner Exhibit 1004, Page 75 of 1052

that previous ASR stages were unable to (or not configured to) fully recognize the speech contained in the acoustic input prior to being processed by NLP stage(s) 730 (or large vocabulary and/or unrestricted ASR may be performed as part of the NLP processing). In the example shown in FIG. 7B, NLP stage(s) 730 may ascertain that the user would like to view his/her calendar to check what appointments are scheduled for tomorrow. As a result, voice response system 750 may check to see what tomorrow's date is and launch a calendar application (see process 740) and pass to the calendar application any appropriate parameters 785, such as tomorrow's date so that the calendar can display the day that the user is interested in and/or list appointments on the calendar scheduled on the indicated date.

[00150] In some circumstances, it may be the case that voice response 750 may need further information from the user and may ask the user for further information, or may confirm that the voice command was understood correctly by formulating confirmation speech via speech synthesis and providing the confirmation to the user via one or more speakers on mobile device 700. Voice response system may then proceed upon confirmation from the user. As denoted by the dotted arrows in FIG. 7B, one or more contextual cues 715 may be evaluated at any of the stages of processing to assist in detecting a voice command by influencing the evaluation of one or more processing stages being performed on the acoustic input. As shown in FIG. 7A, mobile device 700 is capable of communicating over network(s) 795 and voice response system 750 may elect to have network resources perform one or more processing stages on the acoustic input. For example, one or more ASR stages and/or one or more NLP stages may be performed by one or more servers 775 connected to network(s) 795.

[00151] It should be appreciated that when a voice command is detected and the intent of the command ascertained, the further processing that is initiated will depend on the content of the command and/or the operating mode of the mobile device. When the mobile device is operating in a low power mode, some commands may provoke a transition of the mobile device into an active mode, while other commands may be carried out without further waking up the mobile device. For example, with respect to the latter, some mobile devices are capable of rendering music while in a low power mode. Voice commands such as "next track," "previous track," "repeat track," "pause music," "decrease volume," "increase volume," etc. may be performed without having to exit a low power mode. Thus, the acoustic input may be processed in a low power mode

Petitioner Exhibit 1004, Page 76 of 1052

(e.g., where certain components are activated on an "as-needed" basis) to detect the voice command, and the voice command may be carried out without needing to further transition the mobile device into an active mode. In this example, the fact that a music application or player is executing on the mobile device in a low power mode may also operate as a contextual cue to bias the evaluation of the acoustic input to assist in detecting voice commands related to the music player (e.g., the music player being operational may be used to select a processing stage that includes limited vocabulary ASR, wherein the limited vocabulary is selected to include terms frequently associated with controlling a music player such as one or any combination of "track," "volume," "resume," "pause," "repeat," "skip," "shuffle," etc., or any other word or term deemed suggestive of a voice command to control the music player).

[00152] According to some embodiments, tasks that the voice response system is configured to respond to may have an associated indication as to whether the mobile device should be transitioned to an active mode when the corresponding task is identified and/or may include information that specifies which components should be activated to perform the corresponding tasks. In this way, the manner in which the voice response system responds to voice commands may be tailored to consume less power by activating the mobile device or some subset of components of the mobile device as needed to perform the one or more tasks or actions that the user intended to be performed.

[00153] It should be appreciated that any number of other processing stages not specifically mentioned may be performed alternatively or in addition to, and the type and sequence of stages described in connection with FIG. 7B are merely exemplary. For example, in embodiments that allow explicit voice triggers, explicit voice trigger detection may be performed. In the context of the exemplary voice command illustrated in FIGS. 7A and 7B, explicit voice trigger detection would have failed because the user elected to simply begin speaking a voice command in this particular instance. As such, electing to perform explicit voice trigger detection would have consumed power unnecessarily. Allowing for different modes of operation may help to optimize performance in this respect.

[00154] As discussed above, some embodiments determine whether acoustic input includes a voice command without requiring an explicit voice trigger to do so, while some embodiments may require an explicit voice trigger. Embodiments in which an

Petitioner Exhibit 1004, Page 77 of 1052

explicit trigger phrase is not required may operate either by evaluating, but no relying on, whether the acoustic input includes an explicit trigger phrase, or by simply foregoing any detection of an explicit voice trigger. For example, embodiments that do not require an explicit voice trigger may still perform one or more processing stages that detect whether an explicit voice trigger is present to more quickly come to a conclusion in the event the user elects to speak an explicit trigger phrase. However, the absence of an explicit trigger phrase would not end the evaluation due to the voice response system's ability to also detect and respond to voice commands even in the absence of an explicit voice trigger. In other embodiments where explicit voice triggers are not required, no processing to detect an explicit voice trigger is performed (e.g., in voice response implementations that do not have or do not consider explicit voice triggers.

[00155] There may be advantages and disadvantages to each of the above described capabilities with respect to flexibility and processing strategy. For example, not requiring an explicit voice trigger but allowing for them may provide the user with the greatest flexibility. In particular, the user is free to either engage the voice response system either by uttering an explicit voice trigger or by simply speaking the voice command that the user would like the voice response system to respond to (allowing for explicit voice triggers also provides a user with a more failsafe mechanism to engage the voice response system in the event that implicit wake-up is not responding in a given situation). However, this flexibility may come at the cost of increased processing needed to determine whether acoustic input includes a voice command as the system will likely need to both determine whether an explicit voice trigger has been spoken and perform one or more other processing stages to determine whether acoustic input includes a voice command in the absence of an explicit voice trigger, as discussed in further detail below.

[00156] Embodiments that do not require an explicit voice trigger, but also do not allow for explicit voice triggers may perform implicit wake-up more efficiently in this respect because processing stages that evaluate whether an explicit voice trigger has been spoken may be eliminated in favor of processing stages that analyze the acoustic input to detect a voice command by evaluating other characteristics, instead of having to do both. Requiring an explicit voice trigger on the other hand, may provide the most efficient detection of a voice command with generally optimal false positive and false negative rates, but at the cost of potentially impacting the user experience in a negative way by

Petitioner Exhibit 1004, Page 78 of 1052

requiring the user to first speak an explicit voice trigger because speaking the command that the user would like carried out.

[00157] According to some embodiments, a voice response system may be configured to operate in more than one of the above described ways. By having multiple modes of operation in this respect, the mode that a voice response system operates in may be selected by the user, or selected by the mobile device itself. With respect to the former, in some environments, the user may prefer to operate in one mode or the other. For example, in some circumstances reducing false positives may be more important than reducing false negatives, or vice versa. Also, in some circumstances, battery conservation may be of heightened importance such that the user may prefer the mode that requires the least power consumption. Other factors may also contribute to a user wanting to choose a particular mode, either temporarily, or as a more permanent preference. With respect to the latter, the voice response system may switch between modes automatically according to a schedule (which may be user defined), based on the context or environment of the mobile device and/or may learn the user's habits or behaviors and switch between modes accordingly.

[00158] FIGS. 8A and 8B illustrate an example of a voice response system 850 configured to monitor the acoustic environment of a mobile device 800 to receive an acoustic input 805 and determine whether the acoustic input includes a voice command. In this example, mobile device 800 is operating in a low power mode. For example, mobile device 800 may have been inactive for some period of time such that, to conserve battery power, the mobile device has turned off the display, powered down one or more processors and/or disabled certain system components. Conventionally, waking-up the mobile device so that its functionality is available requires one or more manual actions to re-engage the mobile device. Absent these manual actions, conventional mobile devices are rendered unresponsive to voice when operating in low power mode.

[00159] Voice response system 850 is configured to be responsive to voice even when the mobile device 800 is operating in a low power mode. In the example shown in FIGS. 8A and 8B, user 890 has spoken the words "Hello, Dragon" to wake-up the mobile device and engage the voice response system, or otherwise utilize functionality of the mobile device. "Hello, Dragon," in this example, represents an explicit voice trigger understood by voice response system 850. The user's speech may be detected by one or more microphones, located on mobile device, that has been kept at least partially on and

Petitioner Exhibit 1004, Page 79 of 1052

enabled in order to monitor the acoustic environment of the mobile device. In this manner, acoustic input 805 may be received for processing even when the mobile device is operating in a low power mode.

[00160] As shown in FIG. 8B, one or more VAD processing stages 810 may be performed on acoustic input 805. As discussed above, any one or combination of VAD techniques may be applied and any desired order. In this respect, VAD processing stages 810 may be similar in some respects to VAD processing stages 710 in that both may attempt to generally minimize the power required to reach a conclusion as to whether the acoustic input includes voice content or not, and proceed accordingly based upon that evaluation.

[00161] In some embodiments, one or more generally low level acoustic input processing stages may be performed by the microphone (e.g., assessing amplitude levels, determining signal energy characteristics, etc.). In particular, some microphones (e.g., MEMS microphones) include an integrated circuit coupled to the transducer that performs some minimal level of processing of the electrical signals from the transducer. For example, the microphone IC may apply a gain or otherwise amplify the electrical signals as required and/or the microphone IC may perform some level of noise removal and/or filtering to enhance/suppress characteristics of the transducer signal.

[00162] The microphone IC may be configured to perform one or more VAD techniques that facilitate determining whether acoustic input sensed by the microphone includes a voice command or is merely unrelated acoustic activity. For example, the microphone IC may evaluate one or any combination of signal amplitude, signal energy, signal duration to assess whether the received signal merits passing on to further processing stages. In this manner, at least some acoustic activity may be disregarded by the microphone itself without consuming further processing power from the mobile device. Some mobile devices implement an audio codec to convert between analog and digital signals. One or more VAD processing stages may also be performed by the audio codec as well to provide some measure of VAD up front to potentially avoid further processing.

[00163] In the example in FIGS. 8A and 8B, VAD processing stage(s) 810 conclude that acoustic input 805 includes voice content, and the voice response system, based on the assessment that voice is present, decides to invoke explicit voice trigger detection 820 to determine whether the detected speech in acoustic input 805

Petitioner Exhibit 1004, Page 80 of 1052

corresponds to an explicit voice trigger. Generally speaking, detection of an explicit voice trigger involves determining whether the user spoke specific words in a specific order. An explicit voice trigger is often selected to consist of one or more words that are not expected to be used frequently during normal conversation and/or a word ordering having a relatively low expectancy of occurrence. For example, the explicit voice trigger "Hello Dragon" or "Wake-up Dragon" includes the word "Dragon," which alone may have a relatively low occurrence rate in normal speech and thus useable as a trigger phrase, preceded by the word "Hello" or "Wake-up" to provide an explicit voice trigger that is relatively unlikely to be spoken except by a user intending to engage the voice response system.

[00164] Explicit voice trigger detection may be performed using any number of techniques such as a rule-based approach, using a grammar, etc. In some embodiments, word and/or word order may be strictly enforced. That is, a user may be constrained to speak the correct words in the correct order for the system to determine that the user has spoken the explicit trigger phrase. According to some embodiments, some flexibility may be built into explicit voice trigger detection with respect to word choice and/or word order to provide the user with some leeway in this respect. For example, both "Hello Dragon" and "Wake-up Dragon" may be recognized as acceptable explicit voice triggers, providing some flexibility with respect to word choice. As another example, "Dragon Hello" and/or "Dragon Wake-up" may also be recognized as acceptable utterances, providing some flexibility with respect to word order. As a general matter, the more flexibility a user is given with respect to uttering an acceptable explicit voice trigger, the more processing will be required to detect an explicit voice trigger, resulting in increased power consumption. Thus, there may be a trade-of between providing flexibility and implementing the most efficient process in this respect.

[00165] Mobile device 800 may include a single processor, or may include multiple processors. When mobile device 800 includes a single processor (or includes multiple processors but only a single processor capable of performing one or more processing stages), voice response system 800 may activate the processor to the extent needed to perform the various processing stages in a given processing strategy. In a low power mode when multiple processors are available, a primary processor may be powered down into an idle or sleep state, and a secondary processor may be utilized to perform at least a portion of VAD processing stage(s) 810 and/or at least a portion of

Petitioner Exhibit 1004, Page 81 of 1052

explicit voice detection 820. In this manner, at least some significant processing may be performed without having to wake-up or otherwise activate the primary processor (e.g., the clock system of the primary processor may remain substantially at the powered-down state of the low power mode the mobile device is operating in). According to some embodiments, VAD processing and explicit voice detection may be performed by the secondary processor without having to wake-up the primary processor to do so.

[00166] In the example in FIGS. 8A and 8B, explicit voice detection 820 determines that the user spoke a permissible explicit voice trigger. In response, voice response system 850 may transition the mobile device from the low power mode to an active mode (e.g.. the voice response system may incrementally wake-up the mobile device as appropriate, or fully activate the mobile device). The extent to which the voice response system wakes-up the mobile device may depend upon the design of the voice response system. According to some embodiments, the voice response system may wake-up the mobile device to the same extent the mobile device is activated when a user performs the manual actions required to wake-up the device (e.g., the voice response system may turn on the display, activate the primary processor and/or otherwise ready the mobile device to perform any of its available functions). According to some embodiments, the voice response system maintains certain components in low power mode. For example, the voice response system may keep the display off but activate the primary processor to the extent that it wasn't already engaged after detecting the explicit voice trigger, or the voice response system may choose to keep the primary processor deactivated to the same extent it was after detecting the explicit voice trigger until it is needed to process further acoustic information (e.g., to process one or more actionable voice commands following the explicit voice trigger). This applies equally to circumstances in which only a single processor is utilized and when multiple processors are utilized.

[00167] According to some embodiments, voice response system 850 may provide an indication to the user that the explicit voice trigger was detected. For example, the voice response system may synthesize a voice acknowledgement to be rendered to the user via one or more speakers on the mobile device, and/or may turn-on, flash or otherwise utilize an LED on the mobile device to indicate that the explicit voice trigger was received and/or understood. Voice response system may indicate in other ways that the explicit voice trigger was detected and that the system is presently waiting for

Petitioner Exhibit 1004, Page 82 of 1052

subsequent speech input to act upon or is presently processing any remainder of speech contained in acoustic input 805.

[00168] The voice response system may also generally provide feedback to the user that acoustic input has been received. For example, the mobile device may beep or provide some other audible indication that acoustic input has been received. The voice response system may wait to provide such an indication until some level of confidence has been reached that the acoustic input includes the user's speech. Receipt of acoustic input may also be performed using visual indicators such as using one or more LEDs, flashing the display, or via non-visual indicators such as vibration to let the user know that acoustic input was received. In some embodiments, one or more indicators may immediately provide feedback to the user based on any acoustic activity sensed by the mobile device. For example, one or more LEDs on the mobile device may be powered in correspondence to the amplitude of incoming acoustic information, thereby providing a meter of sorts to show the user the intensity of acoustic information being sensed by the mobile device.

[00169] Pursuant to the explicit voice trigger being detected, voice response system 850 may perform one or more processing stages 840 on subsequently received acoustic input 805' or any remainder of acoustic input 805 that may include actionable voice commands that the user expects the voice response system to respond to. Processing stages 840 may include any one or combination of processing stages described herein that are configured to understand the content of a user's voice command. For example, processing stages 840 may include one or more speech processing stages such as ASR or NLP to recognize and understand the voice command so as to take any actions needed to carry out the directives of the user's command.

[00170] It should be appreciated that any one or combination of processing stages 840 may be performed, and such processes may be performed locally or using network resources, as the aspect are not limited in this respect. As denoted by the dotted arrows in FIG. 8B, one or more contextual cues 815 may be evaluated at any of the stages of processing to assist in detecting a voice command by influencing the evaluation of one or more processing stages being performed on the acoustic input. However, the voice response system need not utilized contextual cues, as the aspects are not limited in this respect.

[00171] In some circumstances, a voice response system may detect that acoustic input includes a voice command, but may be unable to ascertain the intent of the user and/or the action that the user would like performed. For example, one or more ASR and/or one or more NLP stages may detect with a high degree of likelihood that acoustic input includes a voice command, but may be unable to map the voice command to one or more specific tasks or actions. According to some embodiments, a default action may be performed when a voice command is detected, but where there is sufficient uncertainty with respect to the intent of the voice command and/or the voice command cannot be suitably mapped to a specific task or set of tasks. The default action may be, for example, an internet search using one or more words recognized in the acoustic input. Alternatively, the default action may include formulating a question to be rendered to the user to obtain further information and/or clarification regarding the voice command. It should be appreciated that any default action may be assigned to voice commands that have been detected, but for which a specific task or set of tasks are not identified (e.g., with a sufficient level of certainty), as the aspects are not limited in this respect.

[00172] Such "catch-all" functionality allows a voice response system to be more responsive to a user even in the presence of uncertainty with respect to the user's intent, either by performing a generic action or by engaging in a dialog with the user to clarify the user's intent with respect to the voice command. According to some embodiments, such a "catch-all" functionality may be disabled when the mobile device is operating in a low power mode to avoid the mobile device taking action whenever a sequence of words is identified. In other embodiments, "catch-all" functionality may be enabled in low power mode, active mode or both, as the aspects are not limited in this respect.

[00173] As discussed above, according to some embodiments, one or more processing stages may be performed by resource(s) available over a network. FIG. 9 illustrates exemplary network resources that may be utilized by a voice response system in assessing whether acoustic input includes a voice command. For example, mobile device 900 may communicate with automatic speech recognition (ASR) component 930 and/or natural language processing (NLP) component 940 via network(s) 995 to perform one or more processing stages. Each of these network resources may be implemented in software, hardware, or a combination of software and hardware. Components implemented in software may comprise sets of processor-executable instructions that may be executed by one or more processors of one or more network computers, such as a

Petitioner Exhibit 1004, Page 84 of 1052

network server or multiple network servers. Each of ASR component 930 and NLP component 940 may be implemented as a separate component, or any combination of these components may be integrated into a single component or a set of distributed components implemented on one or multiple network computers (e.g., network servers). In addition, any one of ASR component 930 and NLP component 940 may be implemented as a set of multiple software and/or hardware components.

[00174] According to some embodiments, acoustic input may be transmitted to ASR component 930 to be recognized. The acoustic input may be processed in any suitable manner prior to providing the acoustic input to ASR component 930. For example, the acoustic input may be pre-processed to remove information, format the acoustic input or modify the acoustic input in preparation for ASR (e.g., the acoustic input may be formatted to conform with a desired audio format and/or prepared for streaming as an audio stream or prepared as an appropriate audio file) so that the acoustic input can be provided as an audio input to ASR component 930 (e.g., transmitted over a network).

[00175] ASR component 930 may be configured to process received audio input (e.g., audio input representing the acoustic input) to form a textual representation of the audio input (e.g., a textual representation of the constituent words in the acoustic input that can be further processed to understand the meaning of the constituent words). Such processing to produce a textual representation may be performed in any suitable way. In some embodiments, ASR component 930 may convert speech to a representation other than a textual representation, or the speech may not be recognized as words, but instead a sequence or collection of abstract concepts. It should be appreciated that the words and/or concepts represented in the acoustic input may be represented and/or stored as data in any suitable form, as aspects are not limited in this respect. ASR component 930 may transmit the recognized acoustic input over the network to be received by the mobile device and/or ASR component 930 may provide the recognized acoustic input to NLP component 940 to assist in understanding the semantic content of the acoustic input to evaluate whether the acoustic input includes a voice command. According to some embodiments, ASR component 930 may be configured to perform limited vocabulary recognition and/or generally unrestricted recognition using one or any combination of suitable acoustic models, vocabularies/lexicons, grammars, language models, etc. to perform recognition, as aspects are not limited for use with any particular ASR

Petitioner Exhibit 1004, Page 85 of 1052

implementation. Any of the ASR processing stages described herein may be performed by ASR component 930.

[00176] As discussed above, one or more processing stages may be performed by NLP component 940 configured to process acoustic input (e.g., a textual representation of the acoustic input or a portion of the acoustic input) to ascertain the semantic content of the acoustic input. NLP component 940 may use any suitable language understanding techniques to ascertain the meaning of the acoustic input so as to assist in assessing whether the acoustic input includes a voice command and/or what tasks or actions are specified therein. For example, NLP component may be configured to identify and extract grammatical and/or syntactical components of the acoustic input, such as parts of speech, or words or phrases belonging to known semantic categories, to facilitate an understanding of the acoustic input. For example, NLP component 940 may identify action words (e.g., action words suggestive of specific tasks or actions), subject words, topic words, and/or any other type or category of words the NLP component 940 may deem relevant to ascertaining the semantic form or content of the acoustic input to facilitate determining whether the acoustic input includes a voice command and/or what tasks or actions are specified therein (e.g., understand the intent of the user).

[00177] NLP component 140 may use, employ or implement one or more statistical models to extract semantic entities from natural language input. In general, some statistical models can be described as a functional component designed and/or trained to analyze new inputs based on probabilistic patterns observed in prior training inputs. In this sense, statistical models differ from "rule-based" models, which typically apply hard-coded deterministic rules to map from inputs having particular characteristics to particular outputs. By contrast, a statistical model may operate to determine a particular output for an input with particular characteristics by considering how often (e.g., with what probability) training inputs with those same characteristics (or similar characteristics) were associated with that particular output in the statistical model's training data. For example, in the context of a voice response system, the input on which the model operates may correspond to acoustic input (or a representation derived therefrom) and the output may include an identification that a voice command is present and/or identification of one or more tasks or actions specified by the voice command and/or one or more associated parameters.

Petitioner Exhibit 1004, Page 86 of 1052

[00178] To supply the probabilistic data that allows a statistical model to extrapolate from the tendency of particular input characteristics to be associated with particular outputs in past examples, statistical models are typically trained (or "built") on training corpuses with a sufficiently large numbers of example inputs. Typically, the example inputs are labeled with the known outputs with which they should be associated (e.g., a corpus of acoustic input may be labeled with the correct task the instruction specifies), often by a human labeler (e.g., with expert knowledge of the domain), or sometimes automatically. Characteristics of interest (known as "features") are identified ("extracted") from the inputs, and the statistical model learns the probabilities with which different features are associated with different outputs, based on how often training inputs with those features are associated with those outputs. When the same features are extracted from a new input (e.g., an input that has not been labeled with a known output, for example, by a human), the statistical model can then use the learned probabilities for the extracted features (as learned from the training data) to determine which output is most likely correct for the new input. In the context of evaluating recognized acoustic input, the features may correspond to the words, concepts, associations, relationships, etc. identified in the acoustic input, and the statistical model may determine whether such features are suggestive of a voice command and/or which one or more tasks or actions are most likely specified based on these features and their relationships.

[00179] NLP component 940 may also include (or alternatively include) one or more rule-based models that provide a set of rules as to how to map words or phrases in acoustic input to voice commands and/or corresponding task(s) and/or that map words or phrases in acoustic input to parameters of an identified task. For example, NLP component 940 may include a rule-based natural language processing component to extract relevant facts about the constituent words recognized in the acoustic input, link facts to concepts or otherwise assist in assessing whether a voice command is present and/or identifying one or more tasks specified in acoustic input. In a rule-based system, a linguist and/or other individual may create a plurality of rules that can specify what words or combinations or words evidence that recognized acoustic input includes a voice command and/or specifies a particular task. Rule-based techniques can be employed by NLP component 940 to determine whether acoustic input includes a voice command

Petitioner Exhibit 1004, Page 87 of 1052

and/or identify one or more tasks or actions specified in the acoustic input and/or any parameters associated with identified tasks.

[00180] It should be appreciated that an NLP component may utilize one or any combination of techniques described above to ascertain whether a voice command is present and/or identify task(s) in acoustic input and/or identify any parameters associated with identified task(s), as aspects are not limited for use with any particular technique or combination of techniques. In accordance with one embodiment described herein, a hybrid approach is used in which NLP component 940 employs both a statistical model and a rules-based component to evaluate whether acoustic input includes a voice command and/or to identify one or more tasks specified in the acoustic input. Such a hybrid approach can be implemented in any of numerous ways, as aspects that employ a hybrid approach are not limited to any particular implementation technique. In accordance with some embodiments, the statistic model and rules-based component of the NLP component 940 may operate in parallel on the same acoustic input, and in some embodiments they may operate serially. In either parallel or serial operation, the statistic model and the rules-based component may operate independently or may exchange information to facilitate evaluating the semantic content of the acoustic input (e.g., to assess whether a voice command is present and/or to identify task(s) specified in the acoustic input, and/or parameters associated with an identified task).

[00181] According to some embodiments, a rules-based component may perform a first pass on acoustic input to identify semantic entities by, for example, extracting relevant facts regarding the language used, linking words or phrases into concepts, and/or identifying other characteristics of known tasks. Thereafter, the statistical model may be used to process the semantic entities to determine what one or more tasks the acoustic input most likely relates to. That is, the outputs of the rules-based component can be used as features for the statistical model. For example, when a rule fires or otherwise indicates that the rule has been met, this may be a feature evaluated by the statistical model. Some rules may be quite specific, so that a firing of the rule indicates with high probability that the determination expressed thereby is accurate (e.g., detection of the word "call" in combination with identifying a proper name may indicate with high probability that the acoustic input includes a voice command to make a telephone call).

[00182] According to some embodiments, a statistical model may operate to identify the one or more tasks most likely specified by the acoustic input and the rule-

Petitioner Exhibit 1004, Page 88 of 1052

based component may be used to identify one or more parameters from the acoustic input associated with the identified task. A statistical model and a rule-based component may be used in other suitable ways, as the aspects are not limited in this respect. In accordance with some embodiments, a determination can be made about whether to use a statistical model or a rule-based component depending upon the content of the acoustic input. For example, some voice commands that are very structured may lend themselves to rules, whereas types of voice commands with significant variability in how they are expressed in natural language may be better suited to the use of a statistical model.

[00183] Again, the descriptions provided above for combining the use of a rule-based component with a statistical model are merely illustrative, as the techniques described herein are not limited to any particular implementation or choice of techniques. The techniques described herein can be implemented using solely a statistical model, solely a rule-based component, or any combination, as they are not limited to any particular implementation technique. That is, processing by NLP component 940 may be as simplistic or as sophisticated as desired to suitably evaluate whether acoustic input includes a voice command and/or to identify one or more tasks when they are present in acoustic input.

[00184] As discussed above, when a mobile device is operating in a low power mode, in some circumstances it may make sense from a power consumption perspective to transmit the acoustic input over the network and allow available network resources (e.g., ASR component 930 and/or NLP component 940) to evaluate whether the acoustic input includes a voice command. In some embodiments, only the components needed to conduct such a network transaction need be activated to access and utilize the network resources, while other components may remain in a low power state. It should be appreciated that any of the above techniques described in connection with ASR component 930 and/or NLP component 940 may alternatively be (or in addition to) implemented on the mobile device.

[00185] An illustrative implementation of a computer system 1000 that may be used to implement one or more of the voice response techniques described herein is shown in FIG. 10. Computer system 1000 may include one or more processors 1010 and one or more non-transitory computer-readable storage media (e.g., memory 1020 and one or more non-volatile storage media 1030). The processor 1010 may control writing data to and reading data from the memory 1020 and the non-volatile storage device 1030

Petitioner Exhibit 1004, Page 89 of 1052

in any suitable manner, as the aspects of the invention described herein are not limited in this respect. Processor 1010, for example, may be a processor on a mobile device.

[00186] To perform functionality and/or techniques described herein, the processor 1010 may execute one or more instructions stored in one or more computer-readable storage media (e.g., the memory 1020, storage media, etc.), which may serve as non-transitory computer-readable storage media storing instructions for execution by processor 1010. Computer system 1000 may also include any other processor, controller or control unit needed to route data, perform computations, perform I/O functionality, etc. For example, computer system 1000 may include any number and type of input functionality to receive data and/or may include any number and type of output functionality to provide data, and may include control apparatus to perform I/O functionality.

[00187] In connection with determining whether acoustic input includes a voice command, one or more programs configured to receive acoustic input, perform one or more processing stages on acoustic input or otherwise execute functionality described herein may be stored on one or more computer-readable storage media of computer system 1000. In particular, some portions or all of a voice response system may be implemented as instructions stored on one or more computer-readable storage media. Processor 1010 may execute any one or combination of such programs that are available to the processor by being stored locally on computer system 1000 or accessible over a network. Any other software, programs or instructions described herein may also be stored and executed by computer system 1000. Computer system 1000 may be the computer system on a mobile device. Computer system 1000 may also be implemented as a standalone computer, server, part of a distributed computing system, and may be connected to a network and capable of accessing resources over the network and/or communicate with one or more other computers connected to the network (e.g., computer system 1000 may be part of the network resources available to a voice response system implemented on a mobile device).

[00188] The terms "program" or "software" are used herein in a generic sense to refer to any type of computer code or set of processor-executable instructions that can be employed to program a computer or other processor to implement various aspects of embodiments as discussed above. Additionally, it should be appreciated that according to one aspect, one or more computer programs that when executed perform methods of the

Petitioner Exhibit 1004, Page 90 of 1052

disclosure provided herein need not reside on a single computer or processor, but may be distributed in a modular fashion among different computers or processors to implement various aspects of the disclosure provided herein.

[00189] Processor-executable instructions may be in many forms, such as program modules, executed by one or more computers or other devices. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Typically, the functionality of the program modules may be combined or distributed as desired in various embodiments.

[00190] Also, data structures may be stored in one or more non-transitory computer-readable storage media in any suitable form. For simplicity of illustration, data structures may be shown to have fields that are related through location in the data structure. Such relationships may likewise be achieved by assigning storage for the fields with locations in a non-transitory computer-readable medium that convey relationship between the fields. However, any suitable mechanism may be used to establish relationships among information in fields of a data structure, including through the use of pointers, tags or other mechanisms that establish relationships among data elements.

[00191] Also, various inventive concepts may be embodied as one or more processes, of which multiple examples have been provided. The acts performed as part of each process may be ordered in any suitable way. Accordingly, embodiments may be constructed in which acts are performed in an order different than illustrated, which may include performing some acts concurrently, even though shown as sequential acts in illustrative embodiments.

[00192] All definitions, as defined and used herein, should be understood to control over dictionary definitions, and/or ordinary meanings of the defined terms.

[00193] As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified.

Petitioner Exhibit 1004, Page 91 of 1052

Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

[00194] The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

[00195] Use of ordinal terms such as "first," "second," "third," etc., in the claims to modify a claim element does not by itself connote any priority, precedence, or order of one claim element over another or the temporal order in which acts of a method are performed. Such terms are used merely as labels to distinguish one claim element having a certain name from another element having a same name (but for use of the ordinal term).

[00196] The phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of "including," "comprising," "having," "containing", "involving", and variations thereof, is meant to encompass the items listed thereafter and additional items.

[00197] Having described several embodiments of the techniques described herein in detail, various modifications, and improvements will readily occur to those skilled in the art. Such modifications and improvements are intended to be within the spirit and

Petitioner Exhibit 1004, Page 92 of 1052

scope of the disclosure. Accordingly, the foregoing description is by way of example only, and is not intended as limiting. The techniques are limited only as defined by the following claims and the equivalents thereto.

What is claimed is:

Petitioner Exhibit 1004, Page 93 of 1052

Electronic Acknowledgement Receipt

EFS ID:	26656814
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Customer Number:	91337
Filer:	Thomas Stephen Chlebeck/Darlene Morin
Filer Authorized By:	Thomas Stephen Chlebeck
Attorney Docket Number:	N0484.71090US04
Receipt Date:	16-AUG-2016
Filing Date:	
Time Stamp:	15:58:13
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$1600
RAM confirmation Number	081716INTEFSW15590000
Deposit Account	232825
Authorized User	Wolf Greenfield

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal of New Application	N048471090US04-TRN-TSC.pdf	31876 05b89e5cd8914913b3bbc7d64eb28dbb84bb6f22	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-TSC.pdf	30112 d4eb56f1ac992215efbd90195490e6c565a57f16	no	1
Warnings:					
Information:					
3	Oath or Declaration filed	N048471090US04-INVDEC-TSC.pdf	2481504 1364be4c8e8a3556124c67d7edcbfc0e2190aa90	no	2
Warnings:					
Information:					
4	Application Data Sheet	N048471090US04-ADS-TSC.pdf	29969 260319af839cf422c1c29d8939394aa5750692f0	no	7
Warnings:					
This is not an USPTO supplied ADS fillable form					
Information:					
5		N048471090US04-CONAPP-TSC.pdf	208047 aa632b5b0985713b837daa8b95a28a25da07d138	yes	67

Multipart Description/PDF files in .zip description

Document Description	Start	End
Specification	1	62
Claims	63	66

	Abstract		67	67	
Warnings:					
Information:					
6	Drawings-other than black and white line drawings	N048471090US04-NDRW-TSC.pdf	2095041 e2b39339ff768cac3af27c7e45141c7e6d132fba	no	10
Warnings:					
Information:					
7	Fee Worksheet (SB06)	fee-info.pdf	35306 97b953f57fea795c8aacaa685d3605b2e55caef4	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	4911855		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 15238238 Document Date: 08/16/2016

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

Electronic Acknowledgement Receipt

EFS ID:	26717745
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Customer Number:	91337
Filer:	Thomas Stephen Chlebeck/Kyle Spencer
Filer Authorized By:	Thomas Stephen Chlebeck
Attorney Docket Number:	N0484.71090US04
Receipt Date:	23-AUG-2016
Filing Date:	
Time Stamp:	10:48:27
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-TSC.pdf	24538 7e87b379d5fe5b974bdcdf36fe6f5dc9010fe1f5	no	1

Warnings:

Information:					
2		N048471090US04-IDS-TSC.pdf	42979 33e0b36b3c771c2fe36f81e7e6465dee5314d65f	yes	5

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Transmittal Letter	1	3
Information Disclosure Statement (IDS) Form (SB08)	4	5

Warnings:	
Information:	
Total Files Size (in bytes):	67517

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2653
Examiner Name	Not Yet Assigned
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Form PTO – 1449

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Thomas S. Chlebeck/		
Printed name	Thomas S. Chlebeck		
Date	2016-08-22	Reg. No.	70,616

Certificate of Electronic Filing Under 37 CFR 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: August 23, 2016 Signature: ___/Kyle D. Spencer/___ (Kyle D. Spencer)

4936779.1

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong, III	
Sheet 1 of 2	GROUP ART UNIT: 2653	EXAMINER: Not Yet Assigned

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		5,428,707		Gould et al.	06-27-1995
		5,774,859		Houser et al.	06-30-1998
		5,802,305		McKaughan et al.	09-01-1998
		5,983,186		Miyazawa et al.	11-09-1999
		6,070,140		Tran	05-30-2000
		6,092,043		Squires et al.	07-18-2000
		6,397,186		Bush et al.	05-28-2002
		6,408,396		Forbes	06-18-2002
		6,411,926		Chang	06-25-2002
		6,449,496		Beith et al.	09-10-2002
		6,594,630		Zlokarnik et al.	07-15-2003
		6,756,700		Zeng	06-29-2004
		6,941,265		Bi et al.	09-06-2005
		6,965,786		Qu et al.	11-15-2005
		7,114,090		Kardach et al.	09-26-2006
		7,567,827		Kim	07-28-2009
		7,574,361		Yeager et al.	04-05-2007
		7,720,683		Vermeulen et al.	05-18-2010
		7,774,204		Mozer et al.	08-10-2010
		8,056,070		Goller et al.	11-08-2011
		8,181,046		Marcu et al.	05-15-2012
		8,190,420		Kadirkamanathan et al.	05-29-2012
		8,285,545		Lee et al.	10-09-2012
		8,548,176		Bright	10-01-2013
		8,620,389		Schrager	12-31-2013
		8,666,751		Murthi et al.	03-04-2014
		8,977,255		Freeman et al.	03-10-2015
		8,977,255		Freeman et al.	03-10-2015
		9,112,984		Sejnoha et al.	08-18-2015
		9,361,885		Ganong, III et al.	06-07-2016
		2003-0040339	A1	Chang	02-27-2003
		2003-0120486	A1	Brittan et al.	06-26-2003

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

**a copy of this reference is not provided as it was previously cited by or submitted to the office in a prior application, Serial No. 13/795,933, filed March 12, 2013, and relied upon for an earlier filing date under 35 U.S.C. 120 (continuation, continuation-in-part, and divisional applications).

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
				FIRST NAMED INVENTOR: William F. Ganong, III	
Sheet	2	of	2	GROUP ART UNIT: 2653	EXAMINER: Not Yet Assigned

		2007-0129949	A1	Alberth, JR. et al.	06-07-2007
		2010-0009719	A1	Oh et al.	01-14-2010
		2010-0124896	A1	Kumar	05-20-2010
		2010-0185448	A1	Meisel	07-22-2010
		2010-0289994	A1	Nonaka	11-18-2010
		2011-0054899	A1	Phillips et al.	03-03-2011
		2012-0329389	A1	Royston et al.	12-27-2012
		2013-0080167	A1	Mozer	03-28-2013
		2013-0080171	A1	Mozer et al.	03-28-2013
		2013-0289994	A1	Newman et al.	10-31-2013
		2013-0339028	A1	Rosner et al.	12-19-2013
		2014-0012573	A1	Hung et al.	01-09-2014
		2014-0012586	A1	Rubin et al.	01-09-2014
		2014-0274203	A1	Ganong, III et al.	09-18-2014
		2014-0274211	A1	Sejnoha et al.	09-18-2014
		2014-0278435	A1	Ganong, III et al.	09-18-2014
		2015-0340042	A1	Sejnoha et al.	11-26-2015

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
	**	International Search Report and Written Opinion for PCT/US2014/024270 mailed June 16, 2014 (N0484.71090WO00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

**a copy of this reference is not provided as it was previously cited by or submitted to the office in a prior application, Serial No. 13/795,933, filed March 12, 2013, and relied upon for an earlier filing date under 35 U.S.C. 120 (continuation, continuation-in-part, and divisional applications).

4936766.1

DOCKET NO.: N0484.71090US04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Not Yet Assigned
Art Unit:	2653

CERTIFICATE OF ELECTRONIC FILING UNDER 37 C.F.R. § 1.8

The undersigned hereby certifies that this paper, along with any paper referred to as being attached or enclosed, is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4), on the 23rd day of August, 2016.

___/Kyle D. Spencer/___
Kyle D. Spencer

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed before the mailing of a first Office Action on the merits in the above-identified case.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of

presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*13/795,933	3/12/2013	Ganong et al.	N0484.71090US01
*14/814,192	7/30/2015	Sejnoha et al.	N0484.71090US03

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
PCT/US2014/024270	**June 16, 2014**	**International Search Report and Written Opinion**	**N0484.71090WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

Petitioner Exhibit 1004, Page 104 of 1052

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Thomas S. Chlebeck/
Thomas S. Chlebeck, Reg. No. 70,616
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: N0484.71090US04
Date: 2016-08-22
xNDDx

Petitioner Exhibit 1004, Page 105 of 1052

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 15/238,238

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A			N/A	280
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A			N/A	600
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	720
TOTAL CLAIMS (37 CFR 1.16(i))	20 minus 20 =	*			OR	x 80 =	0.00
INDEPENDENT CLAIMS (37 CFR 1.16(h))	2 minus 3 =	*				x 420 =	0.00
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).						0.00
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))							0.00
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL			TOTAL	1600

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/238,238	08/16/2016	2653	1600	N0484.71090US04	20	2

CONFIRMATION NO. 5864

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

FILING RECEIPT



Date Mailed: 08/30/2016

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

William F. Ganong III, Brookline, MA;
Paul Adrian Van Mulbregt, Wayland, MA;
Vladimir Sejnoha, Lexington, MA;
Glen Edward Wilson, Maynard, MA;

Applicant(s)

Nuance Communications, Inc., Burlington, MA;

Assignment For Published Patent Application

Nuance Communications, Inc., Burlington, MA

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 13/795,933 03/12/2013

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 08/25/2016

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 15/238,238**

Projected Publication Date: 12/08/2016

Non-Publication Request: No

Early Publication Request: No

Title

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Preliminary Class

379

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04

CONFIRMATION NO. 5864

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

PUBLICATION NOTICE

OC000000087752101

Title:METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Publication No.US-2016-0358605-A1
Publication Date:12/08/2016

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Office of Public Records. The Office of Public Records can be reached by telephone at (703) 308-9726 or (800) 972-6382, by facsimile at (703) 305-8759, by mail addressed to the United States Patent and Trademark Office, Office of Public Records, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently http://pair.uspto.gov/. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Managment, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

DOCKET NO.: N0484.71090US04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Huy C. Ho
Art Unit:	2644

CERTIFICATE OF ELECTRONIC FILING UNDER 37 C.F.R. § 1.8

The undersigned hereby certifies that this paper, along with any paper referred to as being attached or enclosed, is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4), on the 30th day of May, 2017.

___/Kyle D. Spencer/___
Kyle D. Spencer

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed before the mailing of a first Office Action on the merits in the above-identified case.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of

presentation of the references should not be construed as an indication of the importance of the references.

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

Petitioner Exhibit 1004, Page 112 of 1052

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Thomas S. Chlebeck/
Thomas S. Chlebeck, Reg. No. 70,616
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: N0484.71090US04
Date: 2017-05-30
xNDDx

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong
Art Unit	2644
Examiner Name	Huy C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

Form PTO – 1449

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Thomas S. Chlebeck/		
Printed name	Thomas S. Chlebeck		
Date	2017-05-30	Reg. No.	70,616

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: May 30, 2017 Signature: ___/Kyle D. Spencer/___ (Kyle D. Spencer)

5463613.1

Electronic Acknowledgement Receipt

EFS ID:	29339584
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Thomas Stephen Chlebeck/Kyle Spencer
Filer Authorized By:	Thomas Stephen Chlebeck
Attorney Docket Number:	N0484.71090US04
Receipt Date:	30-MAY-2017
Filing Date:	16-AUG-2016
Time Stamp:	14:14:00
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-TSC.pdf	24914 7040fa2aa2a4b367f648b36a3bda5e24293f8df2	no	1

Warnings:

Information:					
2		N048471090US04-IDS-TSC.pdf	28761 b05689f05f859b4be23c960f27cfb00f8ba71c5e	yes	4

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Transmittal Letter	1	3
	Information Disclosure Statement (IDS) Form (SB08)	4	4

Warnings:	
Information:	
Total Files Size (in bytes):	53675

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: Huy C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		
		9,646,610	B2	Macho	05-09-2017
		2010-0121636	A1	Burke et al.	05-13-2010
		2012-0034904	A1	LeBeau et al.	02-09-2012
		2012-0035924	A1	Jitkoff et al.	02-09-2012
		2012-0127072	A1	Kim	05-24-2012

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

5463597.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 06/01/2017

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
HO, HUY C

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
06/01/2017	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@wolfgreenfield.com
N0484_eOfficeAction@WolfGreenfield.com
IP.Inbox@nuance.com

Office Action Summary	Application No. 15/238,238	Applicant(s) GANONG ET AL.	
	Examiner HUY C. HO	Art Unit 2644	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *08/16/2016*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-20* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-20* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *08/16/2016* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Double Patenting

2. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438,

164 USPQ 619 (CCPA 1970); *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on nonstatutory double patenting provided the reference application or patent either is shown to be commonly owned with the examined application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. See MPEP § 717.02 for applications subject to examination under the first inventor to file provisions of the AIA as explained in MPEP § 2159. See MPEP §§ 706.02(l) (1) - 706.02(l)(3) for applications not subject to examination under the first inventor to file provisions of the AIA. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

The USPTO Internet website contains terminal disclaimer forms which may be used. Please visit www.uspto.gov/patent/patents-forms. The filing date of the application in which the form is filed determines what form (e.g., PTO/SB/25, PTO/SB/26, PTO/AIA/25, or PTO/AIA/26) should be used. A web-based eTerminal Disclaimer may be filled out completely online using web-screens. An eTerminal Disclaimer that meets all requirements is auto-processed and approved

immediately upon submission. For more information about eTerminal Disclaimers, refer to www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

3. Claim 1 is rejected on the ground of nonstatutory double patenting as being unpatentable over claim 1 of U.S. Patent No. 9,361,885. Although the claims at issue are not identical, they are not patentably distinct from each other because claim 1 of the U.S. Patent No. 9,361,885 is narrower in scope to claim 1 of the current application.

The following is the comparison between the claims to underline the similar limitations.

Claim 1 (The U.S. Patent No. 9,361,885) A method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is operating in an idle mode, the mobile device having a first processor and a second processor, the method comprising:

receiving acoustic input while the mobile device is operating in the idle mode;

performing at least one first processing stage on the acoustic input using the first processor, prior to engaging the second processor to process the acoustic input, to evaluate whether the acoustic input includes a voice command;

performing at least one second processing stage on the acoustic input using the second processor to evaluate whether the acoustic input includes a voice command if further processing is needed to determine whether the acoustic input includes a voice command;

wherein performing the at least one first processing stage or the at least one second processing stage includes, while the mobile device is operating in the idle mode, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command; and

initiating responding to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

Claim 1 (Current Application), A mobile device comprising:

at least one input configured to receive acoustic input from an environment of the mobile device while the mobile device is operating in a low power mode; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command,

the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command,

wherein the mobile device is configured to initiate responding to the voice command when the at least one processing stage determines that the acoustic input includes a voice command.

Correction is required.

Claim Rejections - 35 USC § 103

4. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

5. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

7. Claims 1-12 and 15-19 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in view of Mozer (U.S. Patent Publication No. 20103/0080167).

Regarding claim 1, Freeman teaches a mobile device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the mobile device while the mobile device is operating in a low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*, wherein the mobile device is configured to initiate responding to the voice command when the at least one processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520)*.

Freeman does not teach "the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command".

Mozer teaches "the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command" (*Mozer, Figure 1B, Paragraph(s) [31]*).

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 15, Freeman teaches at least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a mobile device to receive acoustic input from the environment of the mobile device while the mobile device is operating in a low power mode (*Freeman, the Abstract, Figure 1, Col. 1 Lines 60-65*), the method comprising:

performing, while in the low power mode, at least one processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*; and

initiating responding to the voice command when the at least one processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520)*.

Freeman does not teach "the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command".

Mozer teaches "the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command" (*Mozer, Figure 1B, Paragraph(s) [31])*.

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and

method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 2, Freeman, as modified by Mozer, teaches the mobile device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage to evaluate whether the acoustic input includes a voice command and a second processor configured to perform at least one second processing stage to evaluate whether the acoustic input includes a voice command (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 3, Freeman, as modified by Mozer, teaches the mobile device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 4, Freeman, as modified by Mozer, teaches the mobile device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor (*Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 5, Freeman, as modified by Mozer, teaches the mobile device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech (*Freeman, Col. 4 Lines 45-62, Col. 5 Lines 13-33*).

Regarding claim 6, Freeman, as modified by Mozer, teaches the mobile device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger (*Mozer, Paragraph(s) [33]*).

Regarding claim 7, Freeman, as modified by Mozer, teaches the mobile device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command (*Mozer, Paragraph(s) [36]*).

Regarding claim 8, Freeman, as modified by Mozer, teaches the mobile device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger (*Mozer, Paragraph(s) [33], [36]*).

Regarding claim 9, Freeman, as modified by Mozer, teaches the mobile device of claim 7, wherein the mobile device is further configured to provide, prior to activating the second processor or turning on a display of the mobile device, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Mozer, Paragraph(s) [26]-[28], [46]-[47]*).

Regarding claim 10, Freeman, as modified by Mozer, teaches the mobile device of claim 3, wherein the mobile device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 5, boxes 512)*.

Regarding claim 11, Freeman, as modified by Mozer, teaches the mobile device of claim 1, wherein at least one task specified by the voice command is performed with the mobile device remaining in the idle mode (*Freeman, Figure 5, boxes 506, 508, 510)*.

Regarding claim 12, Freeman, as modified by Mozer, teaches the mobile device of claim 1, wherein the mobile device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input (*Freeman, Figure 5, box 506)*.

Regarding claim 16, Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second

processor being engaged to perform at least one second processing stage (*Mozer, Figure 1B, Paragraph(s) [31]-[33], [36]*).

Regarding claim 17, Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger (*Mozer, Paragraph(s) [33]*).

Regarding claim 18, Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Freeman, Figure 5, boxes 504, 506, 508*).

Regarding claim 19, Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the mobile device remaining in the idle mode (*Freeman, Figure 5, boxes 506, 508, 510*).

8. Claims 13, 14 and 20 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in views of Mozer (U.S. Patent Publication No. 20103/0080167) and LeBeau et al. (U.S. Patent Publication No. 2012/0034904).

Regarding claim 13, Freeman, as modified by Mozer, does not teach the mobile device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 14, Freeman, as modified by Mozer, does not teach the mobile device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]*).

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 20, Freeman, as modified by Mozer, does not teach the at least one non-transitory computer readable medium

of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the mobile device, an occurrence of an event on the mobile device, recent activity on the mobile device, and state of a battery powering the mobile device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]*).

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and

intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUY C. HO whose telephone number is (571)270-1108. The examiner can normally be reached on Monday - Friday, 8:00 a.m. - 5:00 p.m., EST.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Kathy Wang-Hurst can be reached on 571-270-5371. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HUY C HO/
Primary Examiner, Art Unit 2644

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15238238	GANONG ET AL.
	Examiner	Art Unit
	HUY C HO	2644

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	04/20/2017								
	1	✓								
	2	✓								
	3	✓								
	4	✓								
	5	✓								
	6	✓								
	7	✓								
	8	✓								
	9	✓								
	10	✓								
	11	✓								
	12	✓								
	13	✓								
	14	✓								
	15	✓								
	16	✓								
	17	✓								
	18	✓								
	19	✓								
	20	✓								

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong, III	
Sheet 1 of 2	GROUP ART UNIT: 2653	EXAMINER: Not Yet Assigned

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		5,428,707		Gould et al.	06-27-1995
		5,774,859		Houser et al.	06-30-1998
		5,802,305		McKaughan et al.	09-01-1998
		5,983,186		Miyazawa et al.	11-09-1999
		6,070,140		Tran	05-30-2000
		6,092,043		Squires et al.	07-18-2000
		6,397,186		Bush et al.	05-28-2002
		6,408,396		Forbes	06-18-2002
		6,411,926		Chang	06-25-2002
		6,449,496		Beith et al.	09-10-2002
		6,594,630		Zlokarnik et al.	07-15-2003
		6,756,700		Zeng	06-29-2004
		6,941,265		Bi et al.	09-06-2005
		6,965,786		Qu et al.	11-15-2005
		7,114,090		Kardach et al.	09-26-2006
		7,567,827		Kim	07-28-2009
		7,574,361		Yeager et al.	04-05-2007
		7,720,683		Vermeulen et al.	05-18-2010
		7,774,204		Mozer et al.	08-10-2010
		8,056,070		Goller et al.	11-08-2011
		8,181,046		Marcu et al.	05-15-2012
		8,190,420		Kadirkamanathan et al.	05-29-2012
		8,285,545		Lee et al.	10-09-2012
		8,548,176		Bright	10-01-2013
		8,620,389		Schrager	12-31-2013
		8,666,751		Murthi et al.	03-04-2014
		8,977,255		Freeman et al.	03-10-2015
		8,977,255		Freeman et al.	03-10-2015
		9,112,984		Sejnoha et al.	08-18-2015
		9,361,885		Ganong, III et al.	06-07-2016
		2003-0040339	A1	Chang	02-27-2003
		2003-0120486	A1	Brittan et al.	06-26-2003

EXAMINER:	DATE CONSIDERED:
/HUY C HO/	04/20/2017

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

**a copy of this reference is not provided as it was previously cited by or submitted to the office in a prior application, Serial No. 13/795,933, filed March 12, 2013, and relied upon for an earlier filing date under 35 U.S.C. 120 (continuation, continuation-in-part, and divisional applications).

4936766.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong, III	
Sheet 2 of 2	GROUP ART UNIT: 2653	EXAMINER: Not Yet Assigned

		2007-0129949	A1	Alberth, JR. et al.	06-07-2007
		2010-0009719	A1	Oh et al.	01-14-2010
		2010-0124896	A1	Kumar	05-20-2010
		2010-0185448	A1	Meisel	07-22-2010
		2010-0289994	A1	Nonaka	11-18-2010
		2011-0054899	A1	Phillips et al.	03-03-2011
		2012-0329389	A1	Royston et al.	12-27-2012
		2013-0080167	A1	Mozer	03-28-2013
		2013-0080171	A1	Mozer et al.	03-28-2013
		2013-0289994	A1	Newman et al.	10-31-2013
		2013-0339028	A1	Rosner et al.	12-19-2013
		2014-0012573	A1	Hung et al.	01-09-2014
		2014-0012586	A1	Rubin et al.	01-09-2014
		2014-0274203	A1	Ganong, III et al.	09-18-2014
		2014-0274211	A1	Sejnoha et al.	09-18-2014
		2014-0278435	A1	Ganong, III et al.	09-18-2014
		2015-0340042	A1	Sejnoha et al.	11-26-2015

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
	**	International Search Report and Written Opinion for PCT/US2014/024270 mailed June 16, 2014 (N0484.71090WO00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:
/HUY C HO/	04/20/2017

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

**a copy of this reference is not provided as it was previously cited by or submitted to the office in a prior application, Serial No. 13/795,933, filed March 12, 2013, and relied upon for an earlier filing date under 35 U.S.C. 120 (continuation, continuation-in-part, and divisional applications).

4936766.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

Notice of References Cited	Application/Control No. 15/238,238	Applicant(s)/Patent Under Reexamination GANONG ET AL.	
	Examiner HUY C. HO	Art Unit 2644	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-8,977,255 B2	03-2015	Freeman; Daniel	H04M1/72522	455/425
*	B	US-2013/0080167 A1	03-2013	Mozer; Todd F.	G10L17/22	704/246
*	C	US-2012/0034904 A1	02-2012	LeBeau; Michael J.	G10L15/265	455/414.1
	D	US-				
	E	US-				
	F	US-				
	G	US-				
	H	US-				
	I	US-				
	J	US-				
	K	US-				
	L	US-				
	M	US-				

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15238238	GANONG ET AL.
	Examiner	Art Unit
	HUY C HO	2644

CPC- SEARCHED		
Symbol	Date	Examiner
G10L15/20; G10L15/22; G10L15/24; G10L15/25; G10L2015/223; G10L2015/226; H04M2250/74; H04W52/0251; H04W52/0254.	4/20/2017	HH

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner
379	88.01	4/20/2017	HH

SEARCH NOTES		
Search Notes	Date	Examiner
EAST Database Search	4/20/2017	HH
Inventor Search	4/20/2017	HH

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner
379	88.02	4/20/2017	HH
704	246-255	4/20/2017	HH

	/HUY C HO/ Primary Examiner.Art Unit 2644

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 5864

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
15/238,238	08/16/2016	455	2644	N0484.71090US04

APPLICANTS

Nuance Communications, Inc., Burlington, MA;

INVENTORS

William F. Ganong III, Brookline, MA;
Paul Adrian Van Mulbregt, Wayland, MA;
Vladimir Sejnoha, Lexington, MA;
Glen Edward Wilson, Maynard, MA;

**** CONTINUING DATA ***************************

This application is a CON of 13/795,933 03/12/2013

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED ****

08/25/2016

Foreign Priority claimed ☐ Yes ☑ No 35 USC 119(a-d) conditions met ☐ Yes ☑ No Verified and Acknowledged /HUY C HO/ Examiner's Signature	☐ Met after Allowance HH Initials	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
		MA	10	20	2

ADDRESS

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

TITLE

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

FILING FEE RECEIVED		
1600	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

BIB (Rev. 05/07).

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	33784	(379/88.01-88.02|G10L15/20|G10L15/22|G10L15/24|G10L15/25|G10L2015/223|G10L2015/226|H04M2250/74|H04W52/0251|H04W52/0254|).ccls,cpc.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:11
L2	1	("15".src. and "238238".ap.) or "15238238".RLAN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/04/20 16:12
L3	25130	1 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:13
L4	4617	3 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
L5	2692	4 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
L6	563	5 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
L7	362	6 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
L8	17	7 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT;	OR	ON	2017/04/20 16:33

			IBM_TDB			
L9	9564	(704/246-255).ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
L10	8758	9 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
L11	1733	10 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
L12	1338	11 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
L13	212	12 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
L14	104	13 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
L15	3	14 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/03/18 21:04
S2	104	(("GANONG") near3 ("William")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S3	16	(("VAN") near3 ("MULBREGT") near3 ("Paul")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S4	38	(("SEJNOHA") near3 ("Vladimir")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S5	33	(("WILSON") near3 ("Glen")).INV.	US-PGPUB;	OR	OFF	2017/04/20

			USPAT; USOCR			10:20
S6	220	(("NUANCE") near3 ("COMMUNICATIONS") near3 ("INC")).AANM.	USPAT	OR	OFF	2017/04/20 10:20
S7	43	("8977255" | "9112984" | "9361885" | "20070129949" | "20120329389" | "7574361" | "20100124896" | "20030040339" | "20100289994" | "20140274211" | "20140278435" | "8548176" | "20100185448" | "20110054899" | "20130080167" | "20130080171" | "20130289994" | "20130339028" | "20140012573" | "20140012586" | "5428707" | "5774859" | "5802305" | "5983186" | "6070140" | "6092043" | "6397186" | "6408396" | "6411926" | "6449496" | "6594630" | "6756700" | "6941265" | "6965786" | "7114090" | "7567827" | "7720683" | "7774204" | "8056070" | "8181046" | "8190420" | "8285545" | "8666751").PN.	US-PGPUB; USPAT	OR	OFF	2017/04/20 11:21
S8	5	"20120034904"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:10
S9	9	"9361885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:32

4/20/2017 4:37:30 PM
C:\Users\hho3\Documents\EAST\Workspaces\15238238.wsp

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 09/27/2017

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
HO, HUY C

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
09/27/2017	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@wolfgreenfield.com
N0484_eOfficeAction@WolfGreenfield.com
IP.Inbox@nuance.com

Applicant-Initiated Interview Summary	Application No.	Applicant(s)	
	15/238,238	GANONG ET AL.	
	Examiner	Art Unit	
	HUY C. HO	2644	

All participants (applicant, applicant's representative, PTO personnel):

(1) *Huy Ho*.

(2) *Thomas S. Chlebeck (Reg. No. 70,616)*.

(3) *Marc Johannes (Reg. No. 64,978)*.

(4) _____.

Date of Interview: *09/20/2017*.

Type: ☒ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☒ No.
If Yes, brief description: *N/A*.

Issues Discussed ☐101 ☐112 ☐102 ☒103 ☒Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: *1*.

Identification of prior art discussed: *Freeman and Mozer*.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Some amendments to claim 1 were proposed and discussed over the prior art references of record; no agreement was reached at this time.

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HUY C HO/ Primary Examiner, Art Unit 2644	

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record
A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews
Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.
All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiner's responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicant's correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:
- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:

1) A brief description of the nature of any exhibit shown or any demonstration conducted,
2) an identification of the claims discussed,
3) an identification of the specific prior art discussed,
4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
5) a brief identification of the general thrust of the principal arguments presented to the examiner,
   (The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
6) a general indication of any other pertinent matters discussed, and
7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicant's record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiner's version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, "Interview Record OK" on the paper recording the substance of the interview along with the date and the examiner's initials.

Electronic Acknowledgement Receipt

EFS ID:	30544702
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Marc S. Johannes/Eileen MacKenzie
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US04
Receipt Date:	02-OCT-2017
Filing Date:	16-AUG-2016
Time Stamp:	19:43:21
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$200
RAM confirmation Number	100317INTEFSW19434900
Deposit Account	232825
Authorized User	Wolf Greenfield

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-MSJ.pdf	25148 58e144c1d45f8ebaa273edde43db505b9dbd8621	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-MSJ.pdf	31316 a24d545b636ac478e9ea43ada7f33ea8158f47c6	no	1
Warnings:					
Information:					
3	Extension of Time	N048471090US04-EXT-MSJ.pdf	23516 ad89b873778839d8e026bba4be61b60e418d439e	no	1
Warnings:					
Information:					
4		N048471090US04-AMN-MSJ.pdf	139860 4debfeac18e4cfc01d0261455c7b09e45e92fe39	yes	12

Multipart Description/PDF files in .zip description

Document Description	Start	End
Amendment/Req. Reconsideration-After Non-Final Reject	1	1
Claims	2	5
Applicant Arguments/Remarks Made in an Amendment	6	12

Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	31079 911159f949e5dade0dbe178f2b89231946c48514	no	2

Warnings:	
Information:	
Total Files Size (in bytes):	250919

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/AIA/22 (03-13)
Approved for use through 7/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) N0484.71090US04
Application Number 15/238,238-Conf. #5864	Filed August 16, 2016
For METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	
Art Unit 2644	Examiner H. C. Ho

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	Micro Entity Fee	
[x] One month (37 CFR 1.17(a)(1))	$200	$100	$50	$ 200.00
[ ] Two months (37 CFR 1.17(a)(2))	$600	$300	$150	$
[ ] Three months (37 CFR 1.17(a)(3))	$1,400	$700	$350	$
[ ] Four months (37 CFR 1.17(a)(4))	$2,200	$1,100	$550	$
[ ] Five months (37 CFR 1.17(a)(5))	$3,000	$1,500	$750	$

[ ] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

[ ] A check in the amount of the fee is enclosed.

[x] Payment by credit card. ~~Form PTO-2038 is attached.~~

[ ] The Director has already been authorized to charge fees in this application to a Deposit Account.

[x] The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 23/2825 .

[x] Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

[ ] applicant.

[ ] attorney or agent of record. Registration number _____.

[x] attorney or agent acting under 37 CFR 1.34. Registration number 64,978 .

/Marc S. Johannes/ — Signature
October 2, 2017 — Date

Marc S. Johannes — Typed or printed name
617.646.8000 — Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

[ ] * Total of 1 forms are submitted.

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: October 2, 2017 Electronic Signature for: /Eileen M. MacKenzie/

5778930.1

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 CFR §1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: October 2, 2017 Electronic Signature for: ____/Eileen M. MacKenzie/ ___

AMENDMENT IN RESPONSE TO NON-FINAL OFFICE ACTION UNDER 37 C.F.R. § 1.111

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Office Action dated June 1, 2017, please amend the above-identified U.S. patent application as follows:

Amendments to the Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 6 of this paper.

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [X] Fee Transmittal Form
  - [ ] Fee Attached
- [X] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [X] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [X] Other Enclosure(s) (please Identify below):

  Copy of Interview Agenda

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Marc S. Johannes/		
Printed name	Marc S. Johannes		
Date	October 2, 2017	Reg. No.	64,978

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: October 2, 2017 Electronic Signature for: /Eileen M. MacKenzie/

5778929.1

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Marc S. Johannes/Eileen MacKenzie
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 1 month with $0 paid	1251	1	200	200
Miscellaneous:				
			Total in USD ($)	**200**

AMENDMENTS TO THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims:

1. (Currently amended) A ~~mobile~~ device comprising:

at least one input configured to receive acoustic input from an environment of the ~~mobile~~ device while the ~~mobile~~ device is operating in a low power mode; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger~~ evaluate whether the acoustic input includes a voice command~~, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, ~~including~~ transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server ~~at least~~ to recognize a voice command included in ~~evaluate whether~~ the acoustic input ~~includes a voice command~~,

wherein the ~~mobile~~ device is configured to respond ~~initiate responding~~ to ~~the~~ voice commands recognized by the at least one server ~~command when the at least one processing stage determines that the acoustic input includes a voice command~~.

2. (Currently amended) The ~~mobile~~ device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage ~~to evaluate whether the acoustic input includes a voice command~~ and a second processor configured to perform at least one second processing stage ~~to evaluate whether the acoustic input includes a voice command~~.

3. (Currently amended) The ~~mobile~~ device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

Petitioner Exhibit 1004, Page 160 of 1052

4. (Currently amended) The ~~mobile~~ device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

5. (Currently amended) The ~~mobile~~ device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

6. (Currently amended) The ~~mobile~~ device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes [[an]] the explicit voice trigger.

7. (Currently amended) The ~~mobile~~ device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

8. (Currently amended) The ~~mobile~~ device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

9. (Currently amended) The ~~mobile~~ device of claim 7, wherein the ~~mobile~~ device is further configured to provide, prior to activating the second processor or turning on a display of the ~~mobile~~ device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. (Currently amended) The ~~mobile~~ device of claim 3, wherein the ~~mobile~~ device is further configured to transition to an active mode to respond to the voice command when either the

Petitioner Exhibit 1004, Page 161 of 1052

at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

11. (Currently amended) The ~~mobile~~ device of claim 1, wherein at least one task specified by the voice command is performed with the ~~mobile~~ device remaining in the idle mode.

12. (Currently amended) The ~~mobile~~ device of claim 1, wherein the ~~mobile~~ device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. (Currently amended) The ~~mobile~~ device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the ~~mobile~~ device, location of the ~~mobile~~ device, an occurrence of an event on the ~~mobile~~ device, recent activity on the ~~mobile~~ device, and state of a battery powering the ~~mobile~~ device.

14. (Currently amended) The ~~mobile~~ device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the ~~mobile~~ device, location of the ~~mobile~~ device, an occurrence of an event on the ~~mobile~~ device, recent activity on the ~~mobile~~ device, and state of a battery powering the ~~mobile~~ device.

15. (Currently amended) At least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a ~~mobile~~ device to receive acoustic input from the environment of the ~~mobile~~ device while the ~~mobile~~ device is operating in a low power mode, the method comprising:

performing, while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger ~~evaluate whether the acoustic input includes a voice command~~, the at least one processing stage, upon detecting that the acoustic

Petitioner Exhibit 1004, Page 162 of 1052

input includes the explicit voice trigger, ~~including~~ transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server ~~at least~~ to recognize a voice command included in ~~evaluate whether~~ the acoustic input ~~includes a voice command~~; and

~~initiating~~ responding to ~~the~~ voice commands recognized by the at least one server ~~command when the at least one processing stage determines that the acoustic input includes a voice command~~.

16. (Original) The at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

17. (Currently amended) The at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes [[an]] the explicit voice trigger.

18. (Original) The at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. (Currently amended) The at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the ~~mobile~~ device remaining in the idle mode.

20. (Currently amended) The at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the ~~mobile~~ device, location of the ~~mobile~~ device, an occurrence of an event on the ~~mobile~~ device, recent activity on the ~~mobile~~ device, and state of a battery powering the ~~mobile~~ device.

Petitioner Exhibit 1004, Page 163 of 1052

REMARKS

In response to the Office Action dated June 1, 2017, the Assignee of the present application (Nuance Communications, Inc.) respectfully requests reconsideration. Claims 1-20 were previously pending in this application. By this amendment, claims 1-15, 17, and 19-20 are amended. No claims are added or canceled. As a result, claims 1-20 are pending for examination with claims 1 and 15 being independent. No new matter has been added.

Telephone Interview Summary

The Assignee's representatives, Thomas S. Chlebeck and Marc S. Johannes, thank Examiner Ho for his courtesy extended in granting and conducting a telephone interview on September 20, 2017. The substance of that interview is summarized herein.

Assignee's representatives discussed embodiments of the specification and proposed amendments included in the enclosed agenda. Assignee's representatives further discussed how the proposed claims distinguish over the cited references. In particular, Assignee's representatives discussed techniques described in the specification to provide examples of how a device in low power mode detects whether acoustic input includes an explicit voice trigger (such as "Hello, Dragon") and then transmits at least a portion of the acoustic input to a server to determine if a voice command is included (or to recognize the command) (as recited in the amended claims).

The Examiner questioned how the proposed amendments distinguishes the keywords described in Mozer. Assignee's representative explained that the keywords in Mozer are merely words identifiable by the first stage recognizer from user conversation that are categorized and counted to determine if the conversation involves a topic that the system should contribute to. The keywords are not, however, a designated phrase spoken by the user to indicate that the speech is being directed to the device to engage the voice response system. As described in the specification, an "explicit voice trigger" means a "specific and predetermined word or phrase" that is spoken "to engage the voice response system." The Examiner appreciated the distinction, but indicated that a further review of the Freeman and Mozer references would be required before taking a definitive position. The Examiner suggested the Assignee's representatives submit corresponding amendments, which are submitted herein.

Petitioner Exhibit 1004, Page 164 of 1052

Double Patenting Rejection

The Office Action rejected claim 1 for purportedly claiming the same invention as that of claim 1 of prior U.S. Patent No. 9,361,885. Claim 1 is amended herein. As a result, it is believed that the basis of this double patenting rejection no longer exists. Accordingly, withdrawal of this rejection is respectfully requested.

Rejections Under 35 U.S.C. §103

The Office Action rejects independent claims 1 and 15 under 35 U.S.C. §103(a) as purportedly being unpatentable over Freeman (U.S. Patent No. 8,977,255) in view of Mozer (U.S. Patent Publication No. 2013/0080167). The Assignee respectfully requests reconsideration.

As discussed during the telephone interview, claims 1 and 15 are amended herein to advance prosecution. In particular, claim 1 is amended to recite that the at least one processor is configured to "perform, while in the low power mode, at least one processing stage on the acoustic input to ***detect whether the acoustic input includes an explicit voice trigger***, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input." Claim 15 is amended to recite "performing, while in the low power mode, at least one processing stage on the acoustic input to ***detect whether the acoustic input includes an explicit voice trigger***, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input."

As discussed during the telephone interview, neither Freeman nor Mozer disclose or suggest the above quoted limitations in claims 1 and 15, and the above amendments distinguish over the alleged combination of Freeman and Mozer. Specifically, Mozer describes a stage 1 recognizer that seeks to identify keywords in a conversation and to classify those keywords. Mozer at ¶ 33, 42-44. When enough keywords of the same category are identified, the stage 1 recognizer may signal to the stage 2 recognizer to begin performing more sophisticated recognition techniques to assess whether

Petitioner Exhibit 1004, Page 165 of 1052

the user is discussing a topic to which the system may contribute. *Id.* These keywords are not "explicit voice triggers," as defined in the specification. In particular, the specification states that an explicit voice trigger is a "predetermined word or phrase" that is spoken "to engage the voice response system." Specification at [0003]. Furthermore, the specification defines explicit voice triggers as a spoken "designated word or phrase to engage the voice response system." Specification at [0027]. The keywords in Mozer do not meet these requirements. Specifically, the keywords in Mozer are merely words that a conversant uses in a regular conversation, ***not*** a predetermined word or phrase. Moreover, Mozer's keywords are not designated words or phrases used to engage the voice response system. Indeed, the users in Mozer are not even endeavoring to engage the voice response system.

During the telephone conversation, Assignee's representative discussed the examples of "Hello, Dragon" and "Alexa," explaining that explicit voice triggers are designated words or phrases known by the user and a voice response system that allows the user to engage the voice response system. As further discussed, and as described in the specification, an "explicit voice trigger is often selected to consist of one or more words that are not expected to be used frequently during normal conversation and/or a word ordering having a relatively low expectancy of occurrence." Specification at [0164]. By making the explicit voice trigger "unlikely to be spoken except by a user intending to engage the voice response system," false positives can be reduced. Accordingly, the keywords of Mozer not only fail to meet the requirements of an explicit voice trigger, they also fail to have the hallmark characteristics of an explicit voice trigger. Thus, one of ordinary skill in the art would understand that the keywords of Mozer are not explicit voice triggers. Interpreting the keywords as explicit voice triggers is not only inconsistent with the specification, doing so is contrary to it.

Accordingly, the alleged combination of Freeman and Mozer fails to disclose or suggest performing "at least one processing stage on the acoustic input to ***detect whether the acoustic input includes an explicit voice trigger***," as recited in claims 1 and 15 (emphasis added). Therefore, claims 1 and 15, and each claim that depends respectively therefrom, patentably distinguish over the alleged combination and are in allowable condition.

Petitioner Exhibit 1004, Page 166 of 1052

In view of the foregoing, the Assignee respectfully requests withdrawal of the rejections under 35 U.S.C. §103.

General Comments on Dependent Claims

Because each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Assignee believes that it is unnecessary at this time to argue the further distinguishing features of all of the dependent claims. However, the Assignee does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does the Assignee concur that the basis for the rejection of any of the dependent claims is proper. Therefore, the Assignee reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

Petitioner Exhibit 1004, Page 167 of 1052

CONCLUSION

In view of the foregoing remarks, the application is believed to be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes for any reason that the application is not in condition for allowance, he is requested to contact the undersigned to discuss any outstanding issues relating to the allowability of the application.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, the Assignee hereby requests any necessary extension of time. If there is a fee occasioned by this response, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Dated: October 2, 2017

Respectfully submitted,

By /Marc S. Johannes/
Marc S. Johannes
Registration No. 64,978
Thomas S. Chlebeck
Registration No.: 70,616
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

COPY

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Huy C. Ho
Art Unit:	2644

CERTIFICATE OF FACSIMILE TRANSMISSION 37 C.F.R. §1.8(a)

The undersigned hereby certifies that this document is being transmitted via facsimile to the attention of Examiner Huy C. Ho, FAX number (571) 273-8300, at the United States Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450, in accordance with 37 C.F.R. §1.6(d), on the _____ day of September, 2017.

Signature: _____

INTERVIEW AGENDA

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

The Assignee's representatives thank Examiner Ho for agreeing to a telephone interview scheduled for Wednesday, September 20, 2017 at 3:00 PM ET. The tentative interview participants include Examiner Ho and the Assignee's representatives, Marc Johannes (Reg. No. 64,978) and Thomas S. Chlebeck (Reg. No. 70,616). The undersigned will call Examiner Ho at (571) 270-1108 at the proposed time.

During the telephone interview, the Assignee's representatives would like to discuss the proposed amendment below to advance prosecution of this application.

1. (Currently amended) A mobile device comprising:

at least one input configured to receive acoustic input from an environment of the mobile device while the mobile device is operating in a low power mode; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger ~~evaluate whether the acoustic input includes a voice command~~, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, ~~including~~ transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command,

wherein the mobile device is configured to initiate responding to the voice command when the at least one processing stage determines that the acoustic input includes a voice command.

If the Examiner has any comments or questions on this agenda, the Examiner is requested to call the undersigned at the telephone number listed below.

Dated: 2017-09-19

Respectfully submitted,

By /Thomas S. Chlebeck/
Thomas S. Chlebeck
Registration No.: 70,616
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Petitioner Exhibit 1004, Page 170 of 1052

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Examiner Name	H. C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

[ ] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)200.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[ ] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[ ] Charge fee(s) indicated below [ ] Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17 [X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	280	140*	70	600	300	150	720	360	180	
Design	180	90	45	120	60	30	460	230	115	
Plant	180	90	45	380	190	95	580	290	145	
Reissue	280	140	70	600	300	150	2,160	1,080	540	
Provisional	260	130	65	0	0	0	0	0	0	

* The $140 small entity status filing fee for a utility application is further reduced to $70 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	80	40	20
Each independent claim over 3 (including Reissues)	420	210	105
Multiple dependent claims	780	390	195

Total Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
20	- or HP =			x =	

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**

Indep. Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
	- or HP =			x =	

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		**Extra Sheets**		**Number of each additional 50 or fraction thereof**		**Fee ($)**		**Fee Paid ($)**
	- 100 =		/50 =		(round **up** to a whole number) x		=	

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1251 Extension for response within first month	200.00

SUBMITTED BY

Signature	/Marc S. Johannes/	Registration No. (Attorney/Agent)	64,978	Telephone	617.646.8000
Name (Print/Type)	Marc S. Johannes			Date	October 2, 2017

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: October 2, 2017 Electronic Signature for: /Eileen M. MacKenzie/

5778928.1

Electronic Acknowledgement Receipt

EFS ID:	30582291
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Thomas Stephen Chlebeck/Trish McDonald
Filer Authorized By:	Thomas Stephen Chlebeck
Attorney Docket Number:	N0484.71090US04
Receipt Date:	05-OCT-2017
Filing Date:	16-AUG-2016
Time Stamp:	21:29:18
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$180
RAM confirmation Number	100617INTEFSW21294300
Deposit Account	232825
Authorized User	Wolf Greenfield

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-TSC.pdf	24972 8b430f6952f0c32fe9692832b5086903a9d11f18	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-TSC.pdf	31058 0e21f81968a0de4aadeef36bcc75d3fa84c8474d	no	1
Warnings:					
Information:					
3		N048471090US04-IDS-TSC.pdf	31081 9ef4e48799e0501da2c050a32551a88674c079bb	yes	4

Multipart Description/PDF files in .zip description

Document Description	Start	End
Transmittal Letter	1	3
Information Disclosure Statement (IDS) Form (SB08)	4	4

Warnings:					
Information:					
4	Other Reference-Patent/App/Search documents	N048471090CN00_OA_Dated_07-28-2017.pdf	545151 c2521d594df48fcc6aa01ce7056e317ab6484868	no	15
Warnings:					
Information:					
5	Foreign Reference	CN105009204A.pdf	2177666 a96d10c13ceb22f2f3de266f9b559fc80d787524	no	29
Warnings:					
Information:					

6	Foreign Reference	EP2932500B1.pdf	1732727 f162ed64eff763feedf956bca415dc860c4e5d4f	no	26
Warnings:					
Information:					
7	Fee Worksheet (SB06)	fee-info.pdf	30844 a71a19abcc9b88a972f8e9db4b04b6d7210c0273	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	4573499	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: Huy C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		2003-0216909	A1	Davis et al.	11-20-2003

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		CN	105009204	A	Amazon Tech Inc.	10-28-2015	Y-abstract only
		EP	2932500	B1	Amazon Tech Inc.	02-15-2017	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action and Translation thereof for Chinese Application No. 201480013903.1 dated July 28, 2017 (N0484.71090CN00).	Y

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

5675651.1

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	Huy C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Form PTO – 1449

  Copies of Cited References

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Thomas S. Chlebeck/		
Printed name	Thomas S. Chlebeck		
Date	2017-10-05	Reg. No.	70,616

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: October 5, 2017 Electronic Signature: /Trish McDonald/

5675653.1

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Examiner Name	Huy C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)180.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☐ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☐ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17 ☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	280	140*	70	600	300	150	720	360	180	
Design	180	90	45	120	60	30	460	230	115	
Plant	180	90	45	380	190	95	580	290	145	
Reissue	280	140	70	600	300	150	2,160	1,080	540	
Provisional	260	130	65	0	0	0	0	0	0	

* The $140 small entity status filing fee for a utility application is further reduced to $70 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	80	40	20
Each independent claim over 3 (including Reissues)	420	210	105
Multiple dependent claims	780	390	195

Total Claims		**Extra Claims**		**Fee ($)**		**Fee Paid ($)**
_____	- 20 or HP =	_____	x	_____	=	_____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims

Fee ($)	**Fee Paid ($)**
_____	_____

Indep. Claims		**Extra Claims**		**Fee ($)**		**Fee Paid ($)**
_____	- 3 or HP =	_____	x	_____	=	_____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		**Extra Sheets**		**Number of each additional 50 or fraction thereof**		**Fee ($)**		**Fee Paid ($)**
_____	- 100 =	_____	/50 =	_____	(round **up** to a whole number) x	_____	=	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1806 Submission of an Information Disclosure Statement	180.00

SUBMITTED BY

Signature	/Thomas S. Chlebeck/	Registration No. (Attorney/Agent)	70,616	Telephone	617.646.8000
Name (Print/Type)	Thomas S. Chlebeck			Date	2017-10-05

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: October 5, 2017 Electronic Signature: /Trish McDonald/

5675652.1

Espacenet

Bibliographic data: CN105009204 (A) — 2015-10-28

SPEECH RECOGNITION POWER MANAGEMENT

Inventor(s): BASYE KENNETH JOHN; SECKER-WALKER HUGH EVAN; DAVID TONY; KNESER REINHARD; ADAMS JEFFREY PENROD; SALVADOR STAN WEIDNER; KRISHNAMOORTHY MAHESH ± (BASYE KENNETH JOHN, ; SECKER-WALKER HUGH EVAN, ; DAVID TONY, ; KNESER REINHARD, ; ADAMS JEFFREY PENROD, ; SALVADOR STAN WEIDNER, ; KRISHNAMOORTHY MAHESH)

Applicant(s): AMAZON TECH INC ± (AMAZON TECHNOLOGIES, INC)

Classification: **- international:** ***G10L15/26; G10L25/78***
- cooperative: **G10L15/28**; G10L15/30; G10L2015/088; G10L25/78

Application number: CN2013864898 20131209 Global Dossier

Priority number(s): WO2013US73913 20131209 ; US201213711510 20121211

Also published as: EP2932500 (A1) EP2932500 (B1) JP2016505888 (A) US2014163978 (A1) US9704486 (B2) more

Abstract of CN105009204 (A)

Power consumption for a computing device may be managed by one or more keywords. For example, if an audio input obtained by the computing device includes a keyword, a network interface module and/or an application processing module of the computing device may be activated. The audio input may then be transmitted via the network interface module to a remote computing device, such as a speech recognition server. Alternately, the computing device may be provided with a speech recognition engine configured to process the audio input for on-device speech recognition.

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105009204 A
(43)申请公布日 2015.10.28

(21)申请号 201380064898.2
(22)申请日 2013.12.09
(30)优先权数据
13/711,510 2012.12.11 US
(85)PCT国际申请进入国家阶段日
2015.06.11
(86)PCT国际申请的申请数据
PCT/US2013/073913 2013.12.09
(87)PCT国际申请的公布数据
WO2014/093238 EN 2014.06.19
(71)申请人 亚马逊技术有限公司
地址 美国内华达州
(72)发明人 肯尼思·约翰·巴斯耶
休·埃文·塞克-瓦尔克
托尼·大卫 赖因哈德·克内泽尔
杰弗瑞·彭罗德·亚当斯
斯坦·魏德纳·萨尔瓦多
马赫什·克里希纳穆尔蒂
(74)专利代理机构 北京天昊联合知识产权代理有限公司 11112
代理人 顾丽波 李荣胜
(51)Int.Cl.
G10L 15/26(2006.01)
G10L 25/78(2006.01)

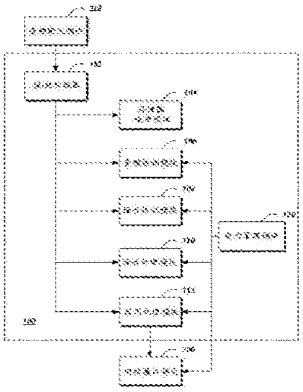
权利要求书2页 说明书18页 附图7页

(54)发明名称

语音识别功率管理

(57)摘要

计算装置的功耗可通过一个或多个关键字进行管理。例如，如果所述计算装置获得的音频输入包含关键字，那么所述计算装置的网络接口模块和/或应用处理模块可被启用。所述音频输入然后可经由所述网络接口模块传输到远程计算装置，诸如，语音识别服务器。或者，所述计算装置可配有语音识别引擎，所述语音识别引擎配置来处理所述音频输入以便进行装置上的语音识别。



CN 105009204 A

1. 一种系统，其包括：

音频输入模块，其配置来接收音频输入；

与所述音频输入模块通信的音频检测模块，所述音频检测模块配置来确定所述音频输入的音量，并且至少部分基于所确定的音量，使语音检测模块启用；

所述语音检测模块配置来确定指示所述音频输入包括语音的可能性的第一分数，并且至少部分基于所述分数，使唤醒字识别模块启用；

所述唤醒字识别模块配置来确定指示所述音频输入包括唤醒字的可能性的第二分数；以及

网络接口模块，其配置来将所获得的音频输入的至少一部分传输到远程计算装置。

2. 根据权利要求 1 所述的系统，其中所述音频输入装置包括麦克风，所述音频检测模块包括第一数字信号处理器，所述语音检测模块包括第二数字信号处理器，并且所述唤醒字识别模块包括微处理器。

3. 根据权利要求 1 所述的系统，其中：

所述语音检测模块还配置来使用以下各者中的至少一个来确定所述第一分数：隐马尔可夫模型、高斯混合模型、多个频谱带中的能量或者多个频谱带中的信噪比；并且

所述唤醒字识别模块还配置来使用以下各者中的至少一个来确定所述第二分数：应用处理模块、隐马尔可夫模型以及高斯混合模型。

4. 根据权利要求 1 所述的系统，其中：

所述唤醒字识别模块还配置来至少部分基于所述第一分数来使所述音频检测模块停用；并且

所述唤醒字识别模块还配置来至少部分基于所述第二分数来使所述语音检测模块停用。

5. 一种操作第一计算装置的计算机实施的方法，所述方法包括：

接收音频输入；

从所述音频输入确定一个或多个值，其中所述一个或多个值包括以下各者中的至少一个：

指示所述音频输入的能量水平的第一值；或

指示所述音频输入包括语音的可能性的第二值；

至少部分基于所述一个或多个值，启用所述第一计算装置的第一模块；

通过所述第一模块来执行操作，其中所述操作包括以下各者中的至少一个：

确定所述音频输入包括唤醒字；

对所述音频输入的至少一部分执行语音识别，以获得语音识别结果；或

使所述音频输入的至少一部分传输到第二计算装置。

6. 根据权利要求 5 所述的计算机实施的方法，其中：

所述第一模块包括处理器，所述处理器可在低功率状态与高功率状态之间切换；并且

所述处理器只有在处于所述高功率状态时才执行所述操作。

7. 根据权利要求 6 所述的计算机实施的方法，其中启用所述第一模块包括将所述处理器从所述低功率状态切换到所述高功率状态。

8. 根据权利要求 6 所述的计算机实施的方法，其还包括停用所述第一模块，其中停用

Petitioner Exhibit 1004, Page 180 of 1052

所述第一模块包括将所述处理器从所述高功率状态切换到所述低功率状态。

9. 根据权利要求 6 所述的计算机实施的方法，其中所述处理器包括数字信号处理器或微处理器中的至少一个。

10. 根据权利要求 5 所述的计算机实施的方法，其中所述操作包括使所述音频输入的至少一部分传输到所述第二计算装置，并且还包括从所述第二计算装置接收语音识别结果。

11. 根据权利要求 10 所述的计算机实施的方法，其还包括 ：

至少部分基于所述一个或多个值，启用所述第一计算装置的第二模块，其中所述第二模块配置来实施语音识别应用 ；以及

通过所述语音识别应用来处理所述语音识别结果。

12. 一种装置，其包括 ：

第一处理器，其配置来 ：

确定一个或多个值，其中所述一个或多个值包括以下各者中的至少一个 ：指示音频输入的能量水平的第一值，或指示所述音频输入包括语音的可能性的第二值 ；以及

至少部分基于所述一个或多个值，使第二处理器启用 ；

所述第二处理器，其配置来执行操作，其中所述操作包括以下各者中的至少一个 ：

确定所述音频输入包括唤醒字 ；

对所述音频输入的至少一部分执行语音识别，以获得语音识别结果 ；或

使所述音频输入的至少一部分传输到第二装置。

13. 根据权利要求 12 所述的装置，其中所述第一处理器包括数字信号处理器或微处理器中的至少一个。

14. 根据权利要求 12 所述的装置，其还包括配置来存储所述音频输入的存储器缓冲模块。

15. 根据权利要求 12 所述的装置，其还包括与所述第一处理器通信的音频输入模块，其中所述音频输入模块配置来获得所述音频输入。

Petitioner Exhibit 1004, Page 181 of 1052

语音识别功率管理

背景技术

[0001] 计算装置可包含语音识别能力。例如，计算装置可捕获音频输入并使用声学模型和语言模型来识别语音。声学模型用来生成假设，其中基于语音的声学特征，声音子字单元（例如，音素等）对应于语音。语言模型基于所说语音所用语言的词汇特征来确定使用声学模型生成的哪个假设最有可能是语音的抄本。计算装置也能够针对具体的语音识别应用来处理识别的语音。例如，有限的语法或自然语言处理技术可用来处理语音。

[0002] 附图简述

[0003] 参考以下详细说明，结合附图，将更易于了解、同时更好地理解本公开的各方面以及随之而来的很多优点，在附图中：

[0004] 图 1 为描绘说明性功率管理子系统的示意图。

[0005] 图 2 为描绘包括功率管理子系统的说明性用户计算装置的示意图。

[0006] 图 3 为描绘用于语音识别功率管理的说明性例行程序的流程图，所述语音识别功率管理可由图 1 的功率管理子系统实施。

[0007] 图 4A、图 4B 和图 4C 为描绘分布式语音识别系统的说明性操作的状态图。

[0008] 图 5 为描绘说明性用户接口的直观图，所述用户接口可由包括功率管理子系统的用户计算装置提供。

[0009] 详细说明

[0010] 在语音识别的一些当前方法中，语音识别能力被分配到分布式计算环境中的一个或多个计算装置上。在这些方法的一个具体实例中，第一计算装置可配置来捕获音频输入，并且可通过网络将音频输入传输到第二计算装置。第二计算装置可对音频输入执行语音识别，并且生成音频输入中包含的语音的抄本。语音的抄本随后可通过网络从第二计算装置传输回到第一计算装置。在其他当前方法中，第一计算装置可配置来捕获音频输入并独立地转录音频输入。

[0011] 在这些和其他当前方法中，第一计算装置可配置来保持在持续活动状态。在这种持续活动状态下，第一计算装置可连续不断地维持到第二计算装置的网络连接。第一计算装置也可继续为用来实施自身的语音识别能力的任何硬件提供电力。这些方法的一个缺点尤其在于，第一计算装置用来维持持续活动状态而消耗的能量可能是无法接受的。对于依赖电池电力的移动计算装置而言，这种能量需求特别成问题。当前方法中还有其他问题。

[0012] 因此，本公开的方面涉及语音识别的功率管理。计算装置可配有功率管理子系统，所述功率管理子系统选择性地启用或停用计算装置的一个或多个模块。此启用可响应于音频输入，所述音频输入包括一个或多个预先指定的所说字词，本文中有时称为“关键字”。促使启用一个或多个部件的关键字在本文中有时称为“唤醒字”，而促使停用一个或多个部件的关键字在本文中有时称为“休眠字”。在一个特定实例中，计算装置可包括选择性地启用的网络接口模块，所述网络接口模块在启用时会消耗能量以使得计算装置连接到第二计算装置，诸如，语音识别服务器或其他计算装置。功率管理子系统可处理音频输入以确定音频输入是否包含唤醒字，并且响应于确定音频输入包括唤醒字而启用网络接口模块。启用网

Petitioner Exhibit 1004, Page 182 of 1052

络接口模块之后，功率管理子系统可使音频输入传输到语音识别服务器以进行处理。

[0013] 功率管理子系统自身可包括一个或多个选择性地启用的模块。在一些实施例中，一个或多个选择性地启用的模块被实施为专用硬件（诸如集成电路、数字信号处理器或其他类型的处理器），其可从功能相对较少的低功率停用状态切换到功能相对较多的高功率启用状态，并且反之亦然。在其他实施例中，一个或多个模块被实施为软件，其包含由一个或多个通用处理器执行的计算机可执行代码。通过启用（或停用）通用处理器可启用（或停用）软件模块，所述通用处理器配置来或能够执行软件中包含的计算机可执行代码。在另外的实施例中，功率管理系统包括一个或多个硬件模块以及一个或多个软件模块。

[0014] 功率管理子系统可还包括与一个或多个选择性地启用的模块通信的控制模块。此控制模块在本文中有时称为“功率管理模块”，并且可包括上述硬件或软件中的任一个。功率管理模块可使启用或停用功率管理子系统的模块。在一些实施例中，至少部分基于计算装置中包括的音频输入模块获得的音频输入的特征，功率管理模块启用或停用一个或多个模块。例如，功率管理子系统的模块可确定一个或多个值，例如，所述值可包括音频输入的能量水平或音量；对应于语音出现在音频输入中的可能性的分数；对应于关键字出现在语音中的可能性的分数；以及其他值。模块可将一个或多个值传送到功率管理模块，所述功率管理模块可与另一模块通信以使该模块启用，或者与从中接收一个或多个值的模块通信以使所述模块和/或其他模块停用。然而，在其他实施例中，第一选择性地启用的模块可直接与第二选择性地启用的模块通信，以使该模块启用。在此类实施例中，无需功率管理模块。在另外的实施例中，功率管理子系统可配有一个或多个模块，其中一个或多个模块中的至少一些彼此通信，但不与功率管理模块通信。

[0015] 在一个示例性实施方案中，功率管理子系统可包括音频检测模块，所述音频检测模块可配置来确定计算装置获得的音频输入的能量水平或音量。尽管音频检测模块可持续监测音频输入，但功率管理子系统的其余部件可保持在低功率非活动状态，直到被启用为止（被功率管理模块或另一不同模块启用）。如果音频检测模块确定音频输入满足阈值能量水平或音量，则语音检测模块可被启用以确定音频输入是否包含语音。如果语音检测模块确定音频输入包含语音，则功率管理子系统中包括的语音处理模块可被启用。语音处理模块可确定语音是否包含唤醒字，并且可任选地对语音分类，以确定特定的用户是否说出所述唤醒字。如果语音处理模块确定语音包含唤醒字，那么应用处理模块可被启用，所述应用处理模块可实施存储在计算装置的存储器中的语音识别应用模块。例如，语音识别应用可包括智能代理前端，诸如，如2011年1月10日申请并于2012年1月19日公开为美国公布号2012/0016678的“智能自动助理（Intelligent Automated Assistant）”中所描述的智能代理前端。该专利申请的公开内容以全文引用的方式并入本文中。选择性地启用的网络接口模块也可被启用，如上文所述，并且音频输入可被传输到远程计算装置，以便进行处理。下文相对于图3更详细地论述了此示例性实施方案。或者，功率管理子系统可响应于检测到唤醒字而启用处理单元，所述处理单元实施计算装置的任何装置上的语音识别能力。

[0016] 通过选择性地启用计算装置的模块，功率管理子系统可有利地提高计算装置的能量效率。通过选择性地启用一个或多个自身的模块，功率管理子系统还可提高计算装置的能量效率。尽管对于依赖电池电力的计算装置而言，此类实施方案尤其有利，但可能需要功率管理的所有计算装置均可受益于本公开的原理。

Petitioner Exhibit 1004, Page 183 of 1052

[0017] 现在转到图1，示出可包括在计算装置中的说明性功率管理子系统100。功率管理子系统100可包括模/数转换器102；存储器缓冲模块104；音频检测模块106；语音检测模块108；语音处理模块110；应用处理模块112；以及功率管理模块120。存储器缓冲模块104可与音频检测模块106、语音检测模块108、语音处理模块110、应用处理模块112以及网络接口模块206通信。功率管理模块120同样可与音频检测模块106、语音检测模块108、语音处理模块110、应用处理模块112以及网络接口模块206通信。

[0018] 模/数转换器102可接收来自音频输入模块208的音频输入。下文相对于图2还详细地论述音频输入模块208。模/数转换器102可配置来将模拟音频输入转换成数字音频输入，以便由功率管理子系统100的其他部件处理。在音频输入模块208获得数字音频输入（例如，音频输入模块208包括数字麦克风或其他数字音频输入装置）的实施例中，模/数转换器102可任选地从功率管理子系统100中省去。因此，音频输入模块208可将音频输入直接提供到功率管理子系统100的其他模块。

[0019] 存储器缓冲模块104可包含配置来存储数字音频输入的一个或多个存储器缓冲区。音频输入模块208获得的音频输入（并且如果是模拟音频输入，则由模/数转换器102转换成数字形式）可被记录到存储器缓冲模块104。记录到存储器缓冲模块104的音频输入可由功率管理子系统100的其他模块存取，以便由这些模块处理，如本文中进一步论述。

[0020] 存储器缓冲模块104的一个或多个存储器缓冲区可包括硬件存储器缓冲区、软件存储器缓冲区或者两者。一个或多个存储器缓冲区可具有相同的容量或不同的容量。存储器缓冲模块104的存储器缓冲区可经选择以存储，其他模块可取决于所述音频输入而启用。例如，如果只有音频检测模块106活动，那么音频输入可被存储到容量相对较小的硬件存储器缓冲区。然而，如果其他模块被启用，诸如，语音检测模块108、语音处理模块110、应用处理模块112和/或网络接口模块206，那么音频输入可存储到容量相对较大的软件存储器缓冲区。在一些实施例中，存储器缓冲模块104包含环形缓冲区，其中音频输入可按照音频输入模块208获得其的顺序进行记录和重写。

[0021] 音频检测模块106可处理音频输入，以确定音频输入的能量水平。在一些实施例中，音频检测模块106包括低功率数字信号处理器（或其他类型的处理器），所述处理器配置来确定所获得的音频输入的能量水平（诸如，音量、强度、振幅等）并且将音频输入的能量水平与能量水平阈值相比较。能量水平阈值可根据用户输入来设置，或者可由功率管理子系统100自动设置，如下文相对于图3进一步论述。在一些实施例中，音频检测模块106还配置来确定音频输入的能量水平至少在阈值持续时间内满足阈值。在此类实施例中，持续时间相对较短的高能量音频输入可对应于相对不可能包含语音的突发噪音，它们可被忽略并且不被功率管理子系统100的其他部件处理。

[0022] 如果音频检测模块106确定所获得的音频输入的能量水平满足能量水平阈值，那么它可与功率管理模块120通信，以指导功率管理模块120启用语音检测模块108。或者，音频检测模块106可将能量水平传送到功率管理模块120，并且功率管理模块120可将所述能量水平与能量水平阈值相比较（且任选地与阈值持续时间相比较），以确定是否启用语音检测模块108。在另一替代方案中，音频检测模块106可直接与语音检测模块108通信以将其启用。任选地，响应于音频检测模块106确定音频输入的能量水平满足阈值，功率管理模块120（或音频检测模块106）可指导音频输入模块208提高其采样率（无论是以帧速率

Petitioner Exhibit 1004, Page 184 of 1052

还是位速率测量）。

[0023] 语音检测模块 108 可处理音频输入，以确定音频输入是否包含语音。在一些实施例中，语音检测模块 108 包括低功率数字信号处理器（或其他类型的处理器），所述处理器配置来实施一个或多个技术，以便确定音频输入是否包含语音。在一些实施例中，语音检测模块 108 应用话音活动检测 (VAD) 技术。此类技术可基于音频输入的各个量化方面确定语音是否存在于音频输入中，所述量化方诸如，音频输入的一个或多个帧之间的频谱斜率；一个或多个频谱带中的音频输入的能量水平；一个或多个频谱带中的音频输入的信噪比；或其他量化方面。在其他实施例中，语音检测模块 108 实施受限分类器，所述分类器配置来将语音与背景噪音区分开来。所述分类器可由诸如线性分类器、支持向量机以及决策树的技术实施。在另外的实施例中，语音检测模块 108 应用隐马尔可夫模型 (Hidden Markov Model，HMM) 或高斯混合模型 (Gaussian Mixture Model，GMM) 技术，以将音频输入与一个或多个声学模型相比较，所述声学模型可包括对应于语音、噪音（诸如环境噪音或背景噪音）或静默的模型。其他技术可用来确定语音是否存在于音频输入中。

[0024] 使用上述任一技术，语音检测模块 108 均可确定一个分数或置信度水平，其值对应于语音实际存在于音频输入中的可能性（本文中所用的“可能性”可指常见用法，不管某件事是可能的还是统计上的用法）。如果分数满足阈值，那么语音检测模块 108 可确定语音存在于音频输入中。然而，如果分数不满足阈值，那么语音检测模块 108 可确定音频输入中没有语音。

[0025] 语音检测模块 108 可将其关于语音是否存在于音频输入中的确定传送到功率管理模块 120。如果语音存在于音频输入中，那么功率管理模块 120 可启用语音处理模块 110（或者，语音检测模块 108 可直接与语音处理模块 110 通信）。如果语音不存在于音频输入中，那么功率管理模块 120 可停用语音检测模块 108。或者，语音检测模块 108 可将分数传送到功率管理模块 120，于是功率管理模块 120 可确定是启用语音处理模块 110 还是停用语音检测模块 108。

[0026] 语音处理模块 110 可处理音频输入，以确定语音中是否包含关键字。在一些实施例中，语音处理模块 110 包括微处理器，所述微处理器配置来检测语音中的关键字，诸如，唤醒字或休眠字。语音处理模块 110 可配置来使用 HMM 技术、GMM 技术或其他语音识别技术来检测关键字。

[0027] 通过确定关键字是否是在一个或多个音素或词语之前或之后立即说出来的，语音处理模块 110 能够将意外包含关键字的语音与有意说出的关键字分开。例如，如果关键字是“ten”，那么语音处理模块 110 能够将独立地说出“ten”的用户与意外说出“ten”（作为词语“Tennessee”、词语“forgotten”、词语“stent”或短语“ten bucks”的一部分）的用户区分开。

[0028] 语音处理模块 110 还可配置来确定语音是否与包括功率管理子系统 100 的计算装置的特定用户相关联，或者语音是否对应于背景噪音；电视音频；音乐；或者用户以外的人的语音，以及其他分类。这一功能可用以下技术实施：诸如线性分类器、支持向量机和决策树，以及用于对音频输入进行分类的其他技术。

[0029] 使用上述任一技术，语音处理模块 110 均可确定一个分数或置信度水平，其值对应于关键字实际存在于语音中的可能性。如果分数满足阈值，那么语音处理模块 110 可确

Petitioner Exhibit 1004, Page 185 of 1052

定关键字存在于语音中。然而，如果分数不满足阈值，那么语音处理模块 110 可确定语音中没有关键字。

[0030] 语音处理模块 110 可将其关于关键字是否存在于语音中的确定传送到功率管理模块 120。如果关键字存在于语音中并且该关键字是唤醒字，那么功率管理模块 120 可启用应用处理模块 112 和网络接口模块 206（或者，语音处理模块 110 可直接与这些其他模块通信）。如果音频输入中不存在关键字（或者关键字是休眠字），那么功率管理模块 120 可停用语音处理模块 110 和语音检测模块 108。或者，语音处理模块 110 可将分数传送到功率管理模块 120，于是功率管理模块 120 可确定是启用应用处理模块 112 和网络接口模块 206，还是停用语音处理模块 110 和语音检测模块 108。在一些实施例中，只有在语音处理模块 110 确定特定用户说出了包含关键字的语音时，才会发生这些启用和 / 或停用。

[0031] 应用处理模块 112 可包括配置来实施语音识别应用的微处理器，所述语音识别应用是由包含功率管理子系统的计算装置提供的。语音识别应用可以包括可能需要语音识别的任何应用，诸如听写应用、消息应用、智能代理前端应用或任何其他应用。语音识别应用也可配置来将语音格式化（例如，将语音压缩），以便通过网络传输到远程计算装置，诸如语音识别服务器。

[0032] 在一些实施例中，应用处理模块 112 包括专用微处理器，用于实施语音识别应用。在其他实施例中，应用处理模块 112 包括通用微处理器，所述微处理器也可实施由包括功率管理子系统 100 的计算装置提供的其他软件，诸如图 2 所示的处理单元 202，将在下文进一步论述。

[0033] 网络接口模块 206 在被启用时可通过一个或多个有线或无线网络提供连接。启用之后，网络接口模块 206 可通过网络将记录到存储器缓冲模块 104 的所接收音频输入传输到远程计算装置，诸如语音识别服务器。远程计算装置可将识别结果（例如，智能代理查询的抄本或响应）返回到包括网络接口模块 206 的计算装置，于是网络接口模块 206 可将接收的识别结果提供给应用处理模块 112 以便处理。下文相对于图 2 进一步论述网络接口模块 206。

[0034] 在不脱离本公开范围的情况下，功率管理子系统 100 的模块可进行组合或重新布置。上述任何模块的功能均可分配到多个模块，或与不同的模块相组合。如上所述，任何或所有模块均可实现在一个或多个集成电路中、一个或多个通用微处理器中，或者实现在一个或多个专用数字信号处理器或其他专用微处理硬件中。一个或多个模块也可实现在由计算装置中包括的处理单元 202 实施的软件中，如下文相对于图 2 进一步论述。此外，一个或多个模块可从功率管理子系统 100 中整个省去。

[0035] 现在转到图 2，示出了其中可包括功率管理子系统 100 的用户计算装置 200。用户计算装置 200 包括处理单元 202；非瞬时计算机可读媒体驱动器 204；网络接口模块 206；如图 1 所示的功率管理子系统 100；以及音频输入模块 208，它们均可借助通信总线而互相通信。用户计算装置 200 也可包括电源 218，所述电源可为用户计算装置 200 的各种部件提供电力，诸如处理单元 202；非瞬时计算机可读媒体驱动器 204；网络接口模块 206；如图 1 所示的功率管理子系统 100；以及音频输入模块 208。

[0036] 处理单元 202 可包括一个或多个通用微处理器，所述微处理器配置来与存储器 210 来回通信，以实施存储于其中的各种软件模块，诸如用户接口模块 212，操作系统 214 以

Petitioner Exhibit 1004, Page 186 of 1052

及语音识别应用模块216。处理单元202还可与功率管理子系统100通信，并且可还实施在软件中实现的功率管理子系统100的任何模块。因此，处理单元202可配置来实施任何或所有以下模块：音频检测模块106；语音检测模块108；语音处理模块110；应用处理模块112以及功率管理模块120。此外，处理单元202可配置来实施装置上的自动语音识别能力，用户计算装置200可具有所述能力。

[0037] 存储器210通常包括RAM、ROM和/或其他持续或非瞬时计算机可读存储媒体。用户接口模块212可配置来借助用户计算装置200的显示器（未示出）来呈现用户接口。用户接口模块212还可配置来处理经由用户输入装置（未示出）接收的用户输入，诸如鼠标、键盘、触摸屏、小键盘等。由用户接口模块212呈现的用户接口可为用户提供机会来定制功率管理子系统100的操作和/或用户计算装置200实施的其他操作。下文将相对于图5进一步论述用户接口的实例。存储器210可另外存储操作系统214，所述操作系统提供计算机程序指令，以便由处理单元202用在用户计算装置200的一般管理和操作中。存储器210还可包含计算机程序指令，应用处理模块112和/或处理单元202执行所述计算机程序指令，以便实施语音识别应用模块216的一个或多个实施例。如上所述，语音识别应用模块216可以是可使用语音识别结果的任何应用，诸如听写应用、消息应用、智能代理应用前端；或可有利地使用语音识别结果的任何其他应用。在一些实施例中，存储器210还可包含可由处理单元202实施的自动语音识别引擎（未示出）。

[0038] 非瞬时计算机可读媒体驱动器204可包括本领域已知的任何电子数据存储器。在一些实施例中，非瞬时计算机可读媒体驱动器204存储一个或多个关键字模型（例如，唤醒字模型或休眠字模型），其中功率管理子系统100可将音频输入与所述关键字模型相比较。非瞬时计算机可读媒体驱动器204还可存储一个或多个声学模型和/或语言模型，用于实施用户计算装置200的任何装置上的语音识别能力。关于语言模型和声学模型的其他信息可见2012年8月16日提交的标题为“判别式语言模型修剪(DISCRIMINATIVE LANGUAGE MODEL PRUNING)”美国专利申请号13/587,799，以及2012年8月22日提交的标题为“无监督声学模型训练(UNSUPERVISED ACOUSTIC MODEL TRAINING)”的美国专利申请号13/592,157。这两个申请的公开内容均以全文引用的方式并入本文中。

[0039] 网络接口模块206可将用户计算装置200连接到一个或多个网络，诸如网络410，将在下文相对于图4A、图4B和图4C进一步论述。处理单元202和功率管理子系统100因而可接收来自远程计算装置的指令和信息，所述远程计算装置也可通过网络410进行通信，诸如语音识别服务器420，如下文也进一步论述。在一些实施例中，网络接口模块206包括无线网络接口，所述无线网络接口通过一个或多个无线网络为用户计算装置200提供连接。

[0040] 在一些实施例中，网络接口模块206被选择性地启用。当网络接口模块206处于停用或“休眠”状态时，它可受限制地连接到或不连接到网络或计算系统，以便省电。在一些实施例中，网络接口模块206默认处于停用状态，并且响应于来自功率管理子系统100的信号而变得启用。当网络接口模块206处于启用状态时，它可相对大量地连接到网络或计算系统，因此网络接口模块206使得用户计算装置200能够将音频输入发送到远程计算装置和/或从远程计算装置（例如，语音识别服务器420）接收关键字确认、语音识别结果或者停用指令。

Petitioner Exhibit 1004, Page 187 of 1052

[0041] 在一个特定的非限制性实例中，响应于功率管理子系统 100 确定音频输入包含唤醒字，网络接口模块 206 可被启用。功率管理子系统 100 可使音频输入经由启用的网络接口模块 206 而传输到远程计算装置（诸如语音识别服务器 420）。任选地，在使随后接收的音频输入传输到远程计算装置之前，功率管理子系统 100 可获得来自远程计算装置的唤醒字的确认。响应于接收到来自远程计算装置的停用指令，响应于确定自从获得满足能量水平阈值的音频输入之后已过了至少预定量的时间，或者响应于接收到包含休眠字的音频输入，功率管理子系统 100 之后可将启用的网络接口模块 206 停用。

[0042] 音频输入模块 208 可包括音频输入装置，诸如麦克风或麦克风阵列，不管是模拟的还是数字的。麦克风或麦克风阵列可实施为定向麦克风或麦克风定向阵列。在一些实施例中，音频输入模块 208 接收音频并且将音频提供到功率管理子系统 100 进行处理，大体上如上文相对于图 1 所论述。音频输入模块 208 还可接收来自功率管理子系统 100 的指令，以设置用于获得音频的采样率（无论是帧速率还是位速率）。音频输入模块 208 也可（或改为）包括一个或多个压电元件和 / 或微机电系统 (MEMS)，其可将声能转换成电信号，以便由功率管理子系统 100 处理。音频输入模块 208 还可配有放大器、整流器以及所需的其他音频处理部件。

[0043] 用户计算装置 200 可配有一个或多个另外的输入装置，诸如光传感器、位置传感器、图像捕获装置等。图 2 中并未示出此类另外的输入装置，以免模糊本公开的原理。在一些实施例中，另外的输入装置可检测条件的发生或不发生。关于此类条件的信息可被提供到功率管理子系统 100，以确定用户计算装置 200 或功率管理子系统 100 的一个或多个部件应被启用还是停用。在一个实施例中，另外的输入装置包括配置来检测光水平的光传感器。在光传感器检测到的光水平不满足阈值时，功率管理模块 120 才可充当网络接口模块 206、才可被启用。在另一实施例中，另外的输入装置包括配置有面部识别能力的图像捕获装置。在此实施例中，只有在图像捕获装置识别与用户计算装置 200 相关联的用户的面部时，网络接口模块 206 才可被启用。关于用输入装置控制语音识别能力的更多信息可见 2002 年 1 月 30 日提交的标题为“自动语音识别系统和方法 (AUTOMATIC SPEECH RECOGNITION SYSTEM AND METHOD)”的美国专利申请号 10/058,730，该申请在 2003 年 7 月 31 日被公开为美国专利公开号 2003/0144844，其公开内容以全文引用的方式并入本文中。关于控制语音识别能力的进一步信息可见 2012 年 12 月 4 日颁布的标题为“使用物理现象检测器来控制语音识别引擎的操作 (USING A PHYSICAL PHENOMENON DETECTOR TO CONTROL OPERATION OF A SPEECH RECOGNITION ENGINE)”的美国专利号 8,326,636。该专利的公开内容以全文引用的方式并入本文中。

[0044] 可提供另外的输入装置，其可包含用户输入装置，诸如，鼠标、键盘、触摸屏、小键盘等。同样，可提供输出装置，诸如，显示器、扬声器、耳机等。在一个特定实例中，用户计算装置 200 可包括一个或多个输出装置，其配置来采用音频形式（例如，通过文本到语音）或视觉形式（例如，通过显示器）呈现语音识别结果。此类输入和输出装置已为本领域所熟知，并且无需在本文中进一步详细论述，且未在图 2 中示出，以避免模糊本发明的原理。

[0045] 电源 218 可向用户计算装置 200 的各个部件提供电力。电源 218 可包括无线或便携式电源，诸如，一次性或可充电电池或电池组；或者可包括有线电源，诸如，配置来插入插座中的交流 (AC) 电源。在一些实施例中，电源 218 将其可供应的功率水平传送到功率管理

Petitioner Exhibit 1004, Page 188 of 1052

子系统 100（例如，剩余电池寿命的百分数、电源 218 是否插入插座等）。在一些实施例中，至少部分基于电源所指示的功率水平，功率管理子系统 100 选择性地启用或停用一个或多个模块。例如，如果用户计算装置 200 插入插座中，那么功率管理子系统 100 可启用网络接口模块 206 并让其处于启用状态。如果用户计算装置 200 使用电池电力，那么功率管理子系统 100 可选择性地启用和停用网络接口模块 206，如上所述。

[0046] 现在转到图 3，示出了说明性例行程序 300，其中功率管理子系统 100 的模块可被选择性地启用以处理音频输入。当说明性例行程序 300 中随后启用的模块可相对大量地处理需求和 / 或功率消耗时，说明性例行程序 300 代表处理和 / 或功率消耗的升级。

[0047] 说明性例行程序 300 可在框 302 处开始，此时，音频输入模块 208 监测音频输入。在框 304 处，音频输入模块 208 可接收音频输入。在框 306 处，接收的音频输入可被记录到存储器缓冲模块 104。在框 308 处，音频检测模块 106 可确定音频输入的能量水平是否满足能量水平阈值（以及任选地，音频输入的能量水平至少是否在阈值持续时间内满足能量水平阈值）。如果音频输入的能量水平并不满足能量水平阈值，那么在框 310 中，音频输入模块 208 可继续监测音频输入，直至接收到另一音频输入为止。

[0048] 返回到框 308，如果音频检测模块 106 确定音频输入的能量水平满足阈值，那么在框 312 处，功率管理模块 120 可启用语音检测模块 108（或者，音频检测模块 106 可直接启用语音检测模块 108，并且在接下来的框中，功率管理模块 120 也可省去）。在框 314 处，语音检测模块 108 可确定语音是否存在于所获得的音频输入中，大体如上文相对于图 1 所论述。如果语音检测模块 108 确定语音不存在于（或不可能存在于）音频输入中，那么在框 316 处，功率管理模块 120 可停用语音检测模块 108。在框 310 中，音频输入模块 208 随后可继续监测音频输入，直至接收到另一音频输入为止。

[0049] 返回到框 314，如果语音检测模块 108 确定音频输入包含语音，那么在框 318 处，功率管理模块 120 可启用语音处理模块 110。如上所述，在框 320 处，语音处理模块 110 可确定唤醒字是否存在于语音中。如果语音处理模块 110 确定唤醒字不存在于语音中（或不可能存在于语音中），那么在框 322 处，语音处理模块 110 可被停用。语音检测模块 108 也可在框 316 处被停用。在框 310 中，音频输入装置 208 随后可继续监测音频输入，直至接收到另一音频输入为止。

[0050] 返回到框 320，如果在一些实施例中，语音处理模块 110 确定唤醒字存在于语音中，那么用户 401，在框 324 中，语音处理模块 110 任选地确定语音是否与特定用户相关联（例如，用户是否说出唤醒字），大体如上文相对于图 1 所论述。如果语音并不与特定用户相关联，那么在框 322 处，语音处理模块 110 可被停用。语音检测模块 108 也可在框 316 处被停用。在框 310 中，音频输入装置 208 随后可继续监测音频输入，直至接收到另一音频输入为止。如果语音与特定用户相关联，那么说明性例行程序 300 可进行到框 326。在其他实施例中，框 324 可被省去，并且响应于语音处理模块 110 确定唤醒字存在于语音中，说明性例行程序 300 可直接从框 320 进行到框 326。

[0051] 在框 326 处，功率管理模块 120 可启用应用处理模块 112，所述应用处理模块可实施图 2 所示的语音识别应用模块 216。在框 328 处，功率管理模块 120 也可启用网络接口模块 206。网络接口模块 206 被启用之后，记录到存储器缓冲模块 104 的音频输入可经由网络接口模块 206 通过网络进行传输。在一些实施例中，当网络接口模块 206 活动时，随后获得

Petitioner Exhibit 1004, Page 189 of 1052

的音频输入从音频输入模块 208 直接提供到应用处理模块 112 和 / 或网络接口模块 206，以用于传输到远程计算装置。然而，在其他实施例中，在将音频输入提供到网络接口模块 206 以便通过网络 410 传输到远程计算装置之前，语音检测模块 108、语音处理模块 110 以及应用处理模块 112 中的任一个或全部均对所述音频输入进行处理。

[0052] 在未示出的一些实施例中，功率管理子系统 100 等待，直到远程计算装置返回唤醒字存在于第一音频输入中的确认，以便传输随后的音频输入用于识别。如果远程计算装置没有提供唤醒字的确认，或者如果经由网络接口模块 206 接收到停用指令，那么网络接口模块 206 以及功率管理子系统 100 的一个或多个模块可被停用。

[0053] 由于功率管理子系统 100 的许多操作可生成概率确定而非准确确定，因此，在说明性例行程序 300 期间可能出现误差。在一些情况下，功率管理子系统 100 的特定模块可提供“误报”，从而导致一个或多个模块被错误地启用。例如，在框 314 处，语音检测模块 108 可错误地确定存在语音，或者在框 320 处，语音处理模块 110 可错误地确定语音包含唤醒字，或在框 324 处，错误地确定语音属于用户。功率管理子系统 100 的模块之间的自适应阈值和交叉验证可有利地用来减少误报。本文中论述了自适应阈值的两个实例，但其他类型的自适应阈值是可能的。如上所述，在框 314 处，语音检测模块可确定语音存在于音频输入中。然而，语音处理模块 110 由于其较好的处理能力而可比语音检测模块 108 更准确地识别语音，它可确定实际上音频输入中没有语音。因此，语音处理模块 110 可指导语音检测模块 108 提高用于确定语音存在于音频输入中的分数阈值，以便减少将来的误报。同样地，如果远程计算装置（诸如语音识别服务器 420）包含语音识别能力，那么远程计算装置可将语音中不存在唤醒字的指示传输到用户计算装置 200，即使语音处理模块 110 可能已经指示存在唤醒字。因此，用于确定唤醒字存在于语音中的语音处理模块 110 的分数阈值可以提高，以便减少将来的误报。此外，可提供用户接口，从而使得用户可提高一个或多个分数阈值来减少误报，如下文相对于图 5 进一步描述。

[0054] 在其他情况下，特定部件可提供“漏报”，从而功率管理子系统 100 的部件不被启用和 / 或网络接口模块 206 不被启用，即使用户已经说出唤醒字也是如此。例如，在框 314 处，语音检测模块 108 可错误地确定不存在语音，或者在框 320 处，语音处理模块 110 可错误地确定语音不包含唤醒字，或在框 324 处，错误地确定语音不属于用户。为了降低漏报的可能性，功率管理子系统 100 可定期降低阈值分数，例如，在框 314、320 和 / 或 324 中，降低满足阈值所需的分数。阈值可继续降低，直至获得一个或多个误报为止，如上所述。一旦获得一个或多个误报，阈值可不再进一步降低，或者可稍微提高。此外，可相应地提供用户接口，从而使得用户可降低一个或多个分数阈值来减少漏报，如下文相对于图 5 进一步描述。

[0055] 在一些实施例中，如果在框 314、320 或 324 中的任一框处获得否定结果，那么并非所有启用的部件均被停用。例如，如果唤醒字在框 320 处未被识别出，那么语音处理模块 110 可在框 322 处被停用，但语音检测模块 108 可保持启用。此外，在一些实施方案中，可跳过某些框。在一些实施例中，在框 314 或 320 处满足阈值的分数会促使一个或多个随后的框被跳过。例如，如果在框 320 处，语音处理模块 110 非常肯定地确定唤醒字存在于语音中，那么说明性例行程序 300 可直接跳到框 326

[0056] 此外，在一些实施例中，用户计算装置 200 可包括被配置成由处理单元 202 执行的自动语音识别引擎。由于此类装置上的语音识别的功率消耗可能尤其高，因此，响应于语音

Petitioner Exhibit 1004, Page 190 of 1052

处理模块 110 确定用户已经说出唤醒字，处理单元 202 可只实施自动语音识别引擎来识别语音。

[0057] 现在参考图 4A、图 4B 和图 4C，说明性环境 400 中示出了分布式语音识别服务的示例性操作。环境 400 可包括用户 401 ；如上所述的用户计算装置 200 ；网络 410 ；语音识别服务器 420 ；以及数据暂存器 430。

[0058] 网络 410 可以是任何有线网络、无线网络或其组合。此外，网络 410 可以是个人区域网、局域网、广域网、电缆网络、卫星网络、蜂窝电话网或其组合。计算机通信领域的技术人员已熟知借由因特网或上述任何其他类型的通信网络进行通信的协议和装置，并且因此，无需在本文中更详细地描述。

[0059] 语音识别服务器 420 通常可以是能够通过网络 410 进行通信的任何计算装置。在一些实施例中，语音识别服务器 420 被实施为一个或多个服务器计算装置，但其他实施方式是可能的。语音识别服务器 420 可能能够通过网络 410 从用户计算装置 200 接收音频输入。此音频输入可用多种方式进行处理，取决于语音识别服务器 420 的实施方式。在一些实施例中，语音识别服务器 420 处理从用户计算装置 200 接收的音频输入，以确认唤醒字存在（例如，通过将音频输入与唤醒字的已知模型相比较），并且将确认传输到用户计算装置 200。语音识别服务器 420 还可配置来使用已知的说话人鉴别技术来鉴别说出唤醒字的用户 401。

[0060] 语音识别服务器 420 可处理从用户计算装置 200 接收的音频输入，以根据音频输入来确定语音识别结果。例如，音频输入可包含用于智能代理来处理的口头查询 ；将被转录成文本的语音 ；或者适用于语音识别应用的其他语音。语音识别服务器 420 可通过网络 410 将语音识别结果传输到用户计算装置 200。关于分布式语音识别应用的进一步信息可见于 2012 年 2 月 14 日颁布的标题为“用于无线装置的托管话音识别系统 (Hosted voice recognition system for wireless devices)”的美国专利号 8,117,268，该专利的公开内容以全文引用的方式并入本文中。

[0061] 语音识别服务器 420 可与数据暂存器 430 本地或远程通信。数据暂存器 430 可实现为以下形式 ：硬盘驱动器、固态存储器，和 / 或语音识别服务器 420 可存取的任何其他类型的非瞬时计算机可读存储媒体。在不脱离本公开的精神和范围的情况下，数据存储器 430 也可以是本领域已知的分布式或跨分区的多个存储装置。此外，在一些实施例中，数据暂存器 430 被实施为基于网络的电子存储服务。

[0062] 数据暂存器 430 可包括一个或多个唤醒字模型。在一些实施例中，唤醒字模型专用于用户 401，而在其他实施例中，接收到用户计算装置 200 确定包含唤醒字的音频输入之后，语音识别服务器可将音频输入与数据暂存器 430 中存储的唤醒字的已知模型相比较。如果音频输入极其类似于已知模型，那么语音识别服务器 420 可将唤醒字的确认传输到用户计算装置 200，于是用户计算装置 200 可获得待由语音识别服务器 420 处理的其他音频输入。

[0063] 数据暂存器 430 还可包含用于语音识别的一个或多个声学和 / 或语言模型。这些模型可包括通用模型，以及专用模型。模型可专用于用户 401 ；专用于用户计算装置 200 和 / 或语音识别服务器 420 实施的语音识别应用 ；或者可具有其他特定用途。关于语言模型和声学模型的其他信息可见 2012 年 8 月 16 日提交的标题为“判别式语言模型

Petitioner Exhibit 1004, Page 191 of 1052

修剪(DISCRIMINATIVE LANGUAGE MODEL PRUNING)”的美国专利申请号13/587,799,以及2012年8月22日提交的标题为“无监督声学模型训练(UNSUPERVISED ACOUSTIC MODEL TRAINING)”的美国专利申请号13/592,157。这两个申请案的公开内容先前以引用的方式并入。

[0064] 数据暂存器430还可包含数据,该数据响应于语音识别服务器420接收的音频输入中含有的查询。语音识别服务器420可识别音频输入中包含的语音,鉴别语音中包含的查询,并且处理查询以鉴别数据暂存器430中的响应数据。语音识别服务器420随后可通过网络410将包含响应数据的智能代理响应提供到用户计算装置200。数据暂存器430中还可包含其他数据。

[0065] 将认识到,很多上述装置都是任选的,而且环境400的实施例可以或可不结合装置。此外,装置无需不同或离散。装置也可以在环境400中重新组织。例如,语音识别服务器420可表示为单个物理服务器计算装置,或者可以分成实现本文所述功能的多个物理服务器。此外,用户计算装置200可具有语音识别服务器420的一些或所有语音识别功能。

[0066] 另外,应注意,在一些实施例中,用户计算装置200和/或语音识别服务器420可由实施在托管计算环境中的一个或多个虚拟机执行。托管计算环境可包含一个或多个快速配置和释放的计算资源,所述计算资源可包括计算装置、网络装置和/或存储装置。托管计算环境也可称为云计算环境。托管计算环境的一个或多个计算装置可包括如上所述的功率管理子系统100。

[0067] 具体参考图4A,示出了可确认唤醒字的说明性操作。用户401可能说出唤醒字502。用户计算装置200可获得可能包含用户语音的音频输入(1),并且确定唤醒字402存在于语音中(2),大体如上文相对于图3所论述。音频输入还可包含话音命令或查询。响应于确定语音包含唤醒字,应用处理模块112和用户计算装置200的网络接口模块206可被启用(3),并且音频输入通过网络410传输到语音识别服务器420(4)。语音识别服务器420可确认唤醒字存在于音频输入中(5),并且可通过网络410将确认传输到用户计算装置200(6)。

[0068] 现转到图4B,响应于接收来自语音识别服务器420的唤醒字的确认,用户计算装置200可继续获得音频输入(7),所述音频输入将被提供到语音识别服务器420以用于处理。例如,获得的音频输入可包含智能代理查询404,用于由语音识别服务器420处理。或者,获得的音频输入可包含将被语音识别服务器420转录的语音(例如,用于由应用处理模块112执行的听写、文字处理或消息应用)。用户计算装置200可通过网络410将音频输入传输到语音识别服务器420(8)。任选地,将生成语音识别结果的语音识别应用的识别符可被提供给语音识别服务器420,因此,语音识别服务器420可生成结果,所述结果具体用于应用处理模块112实施的语音识别应用。语音识别服务器420可识别音频输入中包含的语音(9),并且由此生成语音识别结果(10)。例如,语音识别结果可包含语音的抄本、语音中包含的查询的智能代理响应,或任何其他类型的结果。这些语音识别结果可通过网络410从语音识别服务器420传输到用户计算装置200(11)。响应于接收到结果,应用处理模块112可使以声音形式(例如,通过文本到语音)或以视觉形式(例如,通过用户计算装置200的显示器)来呈现结果(12)。

[0069] 现在参考图4C,用户计算装置200可继续获得音频输入(13),所述音频输入将被

Petitioner Exhibit 1004, Page 192 of 1052

提供到语音识别服务器 420 以用于处理。用户计算装置 200 可通过网络 410 将音频输入传输到语音识别服务器 420(14)。语音识别服务器可识别音频输入中包含的任何语音 (15)。响应于识别出语音,语音识别服务器 420 可确定用户 401 不再对着用户计算装置 200 说话,并且停止任何随后的语音识别 (16)。例如,用户 401 可说出并不对应于结构化命令或查询的字词,诸如无向的自然语言语音 406。语音识别服务器 420 也可分析语音的速度、仔细程度、音调变化或清晰度,以确定该语音并不指向用户计算装置 200,并且不应被处理成语音识别结果。

[0070] 其他类型的音频输入也可促使语音识别服务器 420 停止随后的语音识别。或者,语音识别服务器 420 可确定接收的音频输入并不包含语音。响应于接收到并不包含指向用户计算装置 200 的语音的一个或多个音频输入,语音识别服务器 420 可确定语音识别结果不应被生成,并且语音识别应停止。此外,音频输入可包含预先确定的休眠字,所述休眠字可以是用户 401 选择的。如果语音识别服务器 420 检测到休眠字,那么语音识别服务器 420 可停止对音频输入执行语音识别。此外,语音识别服务器 420 可确定多个用户 401 存在于用户计算装置 200 的附近 (例如,通过对用户计算装置 200 获得的多个音频输入执行说话人鉴别)。如果鉴别的用户 401 的数量满足阈值 (所述阈值可以是大于一的任何数量的用户 401),那么语音识别服务器 420 可确定用户计算装置 200 获得的任何音频输入不太可能被处理成语音识别结果。

[0071] 响应于确定用户 401 的语音并不指向用户计算装置 200(或由于上述任何其他原因,确定不应执行随后的语音识别),语音识别服务器 420 可通过网络 410 将停用指令传输到用户计算装置 200(17)。响应于接收到停用指令,用户计算装置 200 可停用网络接口模块 206 以及功率管理子系统 100 的一个或多个部件 (18),诸如应用处理模块 112、语音处理模块 110 和 / 或语音检测模块 108。其他条件也可促使语音识别服务器 420 将停用指令传输到用户计算装置 200。例如,返回到图 4A,如果语音识别服务器 420 确定唤醒字并未存在于在状态 (1) 处接收到的音频输入中,那么语音识别服务器 420 可将停用指令传输到用户计算装置 200。或者,语音识别服务器 420 可确定自从上次从用户计算装置 200 接收到包含语音的音频输入,已经过去了阈值时间量,并且可因此将停用指令传输到用户计算装置 200。可确定其他标准,用于将停用指令传输到用户计算装置 200。

[0072] 再次返回到图 4A,在接收到确定包含唤醒字的随后音频输入之后,用户计算装置 200 可启用功率管理子系统 100 的部件和网络接口模块 206,并且将音频输入传输到语音识别服务器 420。本文中示出的示例性操作因此可自己重复。

[0073] 图 4A、图 4B 和图 4C 中示出的示例性操作用于说明性目的。本文中示出的示例性操作中可省去一个或多个状态,或者可添加额外的状态。在一个特定的实例中,在传输语音识别服务器 420 将为其生成语音识别结果的音频输入之前,用户计算装置 200 无需从语音识别服务器 420 获得唤醒字的确认。此外,在停用网络接口模块 206 和 / 或功率管理子系统 100 的一个或多个部件诸如应用处理模块 112、语音处理模块 110 或语音检测模块 108 之前,用户计算装置 200 无需获得停用指令。事实上,功率管理子系统 100 可 (通过音频检测模块 106) 确定,自从用户计算装置 200 已获得能量水平满足能量水平阈值的音频输入之后,已经过去了至少阈值时间量。或者,用户计算装置 200 可 (通过语音检测模块 108) 确定,自从获得包含语音的音频输入之后,已经过去了至少阈值时间量。响应于确定已经过去

Petitioner Exhibit 1004, Page 193 of 1052

了阈值时间量，功率管理子系统 100 可使网络接口模块 206 停用，并且可停用其本身的一个或多个部件，如上文相对于图 3 所述。

[0074] 此外，功率管理子系统 100 可配置来识别用户 401 选择并说出的休眠字，采用大体类似于图 3 所示的识别唤醒字的方式。如果休眠字被功率管理子系统 100 检测到（例如，被语音处理模块 110 检测到），那么网络接口模块 206 和 / 或功率管理子系统 100 的一个或多个部件可被停用。同样地，如果用户计算装置 200 包含其自身装置上的语音识别能力，那么响应于检测到休眠字，它们可被停用。

[0075] 图 5 描绘出说明性用户接口 500，其可由用户计算装置 200 提供用于定制功率管理子系统 100 和用户计算装置 200 的操作。在一个实施例中，用户接口模块 212 通过用户接口 500 处理用户输入，并将其提供到功率管理子系统 100。

[0076] 能量水平阈值元件 502 可使得用户能够指定阈值能量水平，在该阈值能量水平，语音检测模块 108 应被启用，如图 3 中的框 308 所示。例如，如果用户计算装置 200 处于相对嘈杂的环境，或者如果用户计算装置 200 正在经历音频检测模块 106 确定的相当大量的“误报”，那么用户 401 可能会希望提高语音处理模块 108 被启用时所处的能量水平阈值。如果用户 401 处于相对安静的环境，或者如果用户计算装置 200 正在经历相当大量的漏报，那么用户 401 可能会希望降低语音检测模块 108 被启用时所处的能量水平阈值。如上所述，能量水平阈值可对应于音量阈值、强度阈值、振幅阈值或与音频输入相关的其他阈值。

[0077] 关键字置信度阈值元件 504 可使得用户能够指定阈值分数，在该阈值分数，语音处理模块 110 确定关键字存在。同样地，鉴别置信度阈值元件可使得用户能够指定阈值分数，在该阈值分数，语音处理模块 110 确定用户说出关键字。在一个实施例中，响应于语音处理模块 110 识别出唤醒字（例如，语音处理模块 110 确定满足阈值的分数，所述分数对应于语音中包含唤醒字的可能性），应用处理模块 112 和网络接口模块 206 被启用。在另一实施例中，响应于至少通过对应于唤醒字与用户相关联的可能性的阈值分数，语音处理模块 110 确定唤醒字与用户 401 相关联，应用处理模块 112 和网络接口模块 206 被启用。在另外的实施例中，响应于语音处理模块 110 至少通过阈值分数识别唤醒字并且至少通过阈值分数确定唤醒字与用户 401 相关联，应用处理模块 112 和网络接口模块 206 被启用。其他阈值元件可被提供来使得用户 401 能够设置个别阈值，用于启用功率管理子系统 100 的任何或所有个别部件。另外的阈值元件可被提供来使得用户能够指定说明性例行程序 300 的一个或多个框可被跳过的分数，大体如上文相对于图 3 所论述。

[0078] 用户接口 500 还可包含一个或多个计时元件 508A 和 508B。每个计时元件均可用来设置阈值时间间隔，在该阈值时间间隔，网络接口模块 206 和 / 或功率管理子系统 100 的一个或多个部件自动停用。参考计时元件 508A，如果功率管理子系统 100 确定自从用户计算装置 200 获得能量水平满足能量水平阈值的音频输入之后，已经过去了至少阈值时间间隔，那么网络接口模块 206 可自动停用，此外，功率管理子系统 100 的应用处理模块 112、语音处理模块 110、语音检测模块 108 也是如此。另外的计时元件也可用来设置阈值时间间隔，在该阈值时间间隔之后，语音识别服务器 420 将停用指令自动发送到网络接口模块 206 和功率管理子系统 100，大体如上文相对于图 4C 所论述。也可提供用于功率管理子系统 100 的其他模块的计时元件。

[0079] 继续参考图 5，用户 401 可选择唤醒字是否应由语音识别服务器 420 通过服务器确

Petitioner Exhibit 1004, Page 194 of 1052

认元件 510 来进行确认。在一些实施例中，在语音处理模块 110 检测到唤醒字之后，如果唤醒字的确认从语音识别服务器 420 接收到，应用处理模块 112 和网络接口模块 206 才保持启用。如果用户 401 要求唤醒字的服务器确认，那么随后获得的音频输入可以不传输到语音识别服务器 420，除非唤醒字被确认。然而，如上所述，未必要求确认。如果用户 401 并不要求唤醒字的服务器确认，那么用户计算装置 200 可将随后获得的一个或多个音频输入传输到语音处理模块 110，所述语音处理模块检测语音中的唤醒字和 / 或确定语音与用户 401 相关联。

[0080] 用户 401 也可选择是否要求用说话人鉴别元件 512 进行说话人鉴别。如果用户 401 要求说话人鉴别，那么语音处理模块 110 和 / 或语音识别服务器 420 可用来确定包含对应于唤醒字的语音的音频输入是否与用户 401 相关联。响应于语音处理模块 110 确定用户 401 就是语音的说话人，应用处理模块 112 和网络接口模块 206 可被启用。同样地，响应于从语音识别服务器 420 接收到用户 401 就是唤醒字的说话人的确认，网络接口模块 206 可保持启用状态。然而，如果用户 401 并不要求说话人鉴别，那么语音处理模块 110 和语音识别服务器 420 均无需鉴别说话人。

[0081] 用户接口 500 还可包含装置上的识别选择元件 514，其中用户 401 可选择用户计算装置 200 是否自己生成语音识别结果，或者音频输入是否发送到语音识别服务器 420 以用于处理成语音识别结果。如果用户计算装置 200 并不包含装置上的语音识别能力，那么装置上的识别选择元件 514 可任选地被禁用或停止。此外，如果电源 218 下降至低于阈值供电水平（例如，电池电量百分数），那么装置上的识别选择元件 514 可自动取消选择（并且装置上的语音识别能力自动禁用），因为处理单元 202 和 / 或应用处理模块 112 实施的装置上的语音识别能力可能会要求相对较大的电耗。

[0082] 唤醒字方框 516 和休眠字方框 518 可包括用户接口元件，借此，用户 401 可记录用户 401 说出的唤醒字或休眠字，并将其回放。当用户 401 记录唤醒字或休眠字时，网络接口模块 206 可自动启用，从而使得包含用户语音的音频输入可被提供到语音识别服务器 420。语音识别服务器 420 可返回记录的唤醒字或休眠字的抄本，从而用户可确定语音识别服务器 420 是否正确理解了记录的唤醒字或休眠字。或者，当用户 401 记录唤醒字或休眠字时，用户计算装置 200 的任何装置上的语音识别能力均可被启用，以对记录的用户 401 的语音进行转录。说出的唤醒字或休眠字的频谱指示也可由用户接口 500 提供。任选地，唤醒字方框 516 和休眠字方框 518 可包含唤醒字或休眠字的暗示，并且可指示用户 401 提供的唤醒字或休眠字的质量，所述质量可反映出唤醒字或休眠字会产生误报或漏报的可能性。关于提示关键字的进一步信息可见于 2012 年 11 月 6 日提交的标题为“唤醒字评估 (WAKE WORD EVALUATION)”的美国专利申请号 13/670,316。该申请案的公开内容以全文引用的方式并入本文中。

[0083] 出于说明的目的，已将本公开的各方面论述为硬件实施方式。然而，如上所述，功率管理子系统 100 可部分或完全由处理单元 202 实施。例如，功率管理子系统 100 的一些或全部功能可实施为处理单元 202 执行的软件指令。在一个特定的非限制性实例中，语音处理模块 110、应用处理模块 112 以及功率管理模块 120 的功能可实施为处理单元 202 执行的软件。因此，处理单元 202 可配置来响应于检测到唤醒字而选择性地启用和 / 或停用网络接口模块 206。可能还有其他的实施方式。

Petitioner Exhibit 1004, Page 195 of 1052

[0084] 根据实施例，本文所述的例行程序或算法中的任一个的某些动作、事件或功能可用不同的顺序来执行，并可被添加、合并或完全忽略（例如，并非所有描述的操作或事件都是实践算法所必需的）。此外，在某些实施例中，操作或事件可同时执行（例如，通过多线程处理、中断处理，或者多个处理器或处理器芯，或者在其他并行架构上），而非连续执行。

[0085] 结合本文公开的实施例描述的各种说明性逻辑块、模块、例行程序和算法步骤可实施为电子硬件、计算机软件，或是这两个的组合。为了清楚地说明硬件和软件的这种互换性，上文总体上就其功能描述了各种说明性部件、块、模块和步骤。此类功能是实施为硬件还是软件，取决于整个系统上施加的特定应用和设计限制。所述功能可用不同的方式实施，以便用于每个特定的应用，但此类实施决策不应被解释为脱离本公开的范围。

[0086] 除非另行说明，或者根据上下文可用其他方式理解，否则本文中所用的条件性语言，诸如“可”、“可以”、“可能”、“也许”、“例如”等，总体上意图传达以下内容：某些实施例包含（但其他实施例不包含）某些特征、元件和/或步骤。因此，这些条件性语言总体上并不意图暗示：特征、元件和/或步骤是一项或多项实施例无论如何都需要的，或者在有或没有作者输入或提示的情况下，一项或多项实施例必然包含决策逻辑，不论这些特征、元件和/或步骤均包含其中还是将在任何特定的实施例中执行。术语“包括”、“包含”、“具有”等是同义词，以开放的形式使用，并且并不排除额外的元件、特征、动作、操作等等。此外，术语“或”具有包含意义（而并非排除意义），因此（例如）在用于连接一系列元件时，术语“或”意味着这一系列元件中的一个、一些或所有元件。

[0087] 除非另行说明，否则结合性语言（诸如短语“X、Y和Z中的至少一个”）将在上下文中被理解为大体传达一个项目、术语等可以是X、Y或Z或者其组合。因此，这些结合性语言通常并不意图暗示某些实施例要求分别存在X中的至少一个、Y中的至少一个以及Z中的一个。

[0088] 虽然上文的具体实施方式已展示、描述并指出应用于各种实施例的新颖特征，但是应理解，在不脱离本发明的精神的情况下，可对所示装置或算法的形式和细节进行各种省略、取代和改变。应认识到，本文所述的本发明的某些实施例可采用不提供本文所述的所有特征和优点的方式实现，因为一些特征可独立于其他特征单独使用或实践。本文公开的本发明的范围由所附权利要求书、而非上述说明指示。权利要求书的范围涵盖其等效意义和范围内的所有变化。

[0089] 条款：

[0090] 1. 一种系统，其包括：

[0091] 音频输入模块，其配置来接收音频输入；

[0092] 与所述音频输入模块通信的音频检测模块，所述音频检测模块配置来确定所述音频输入的音量，并且至少部分基于所确定的音量，使语音检测模块启用；

[0093] 所述语音检测模块配置来确定指示所述音频输入包括语音的可能性的第一分数，并且至少部分基于所述分数，使唤醒字识别模块启用；

[0094] 所述唤醒字识别模块配置来确定指示所述音频输入包括唤醒字的可能性的第二分数；以及

[0095] 网络接口模块，其配置来将所获得的音频输入的至少一部分传输到远程计算装置。

Petitioner Exhibit 1004, Page 196 of 1052

[0096] 2. 根据条款 1 所述的系统，其中所述音频输入装置包括麦克风，所述音频检测模块包括第一数字信号处理器，所述语音检测模块包括第二数字信号处理器，并且所述唤醒字识别模块包括微处理器。

[0097] 3. 根据条款 1 所述的系统，其中：

[0098] 所述语音检测模块还配置来使用以下各者中的至少一个来确定所述第一分数：隐马尔可夫模型、高斯混合模型、多个频谱带中的能量或者多个频谱带中的信噪比；并且

[0099] 所述唤醒字识别模块还配置来使用以下各者中的至少一个来确定所述第二分数：应用处理模块、隐马尔可夫模型以及高斯混合模型。

[0100] 4. 根据条款 1 所述的系统，其中：

[0101] 所述唤醒字识别模块还配置来至少部分基于所述第一分数来使所述音频检测模块停用；并且

[0102] 所述唤醒字识别模块还配置来至少部分基于所述第二分数来使所述语音检测模块停用。

[0103] 5. 一种操作第一计算装置的计算机实施的方法，所述方法包括：

[0104] 接收音频输入；

[0105] 从所述音频输入确定一个或多个值，其中所述一个或多个值包括以下各者中的至少一个：

[0106] 指示所述音频输入的能量水平的第一值；或

[0107] 指示所述音频输入包括语音的可能性的第二值；

[0108] 至少部分基于所述一个或多个值，启用所述第一计算装置的第一模块；

[0109] 通过所述第一模块来执行操作，其中所述操作包括以下各者中的至少一个：

[0110] 确定所述音频输入包括唤醒字；

[0111] 对所述音频输入的至少一部分执行语音识别，以获得语音识别结果；或

[0112] 使所述音频输入的至少一部分传输到第二计算装置。

[0113] 6. 根据条款 5 所述的计算机实施的方法，其中：

[0114] 所述第一模块包括处理器，所述处理器可在低功率状态与高功率状态之间切换；并且

[0115] 所述处理器只有在处于所述高功率状态时才执行所述操作。

[0116] 7. 根据条款 6 所述的计算机实施的方法，其中启用所述第一模块包括将所述处理器从所述低功率状态切换到所述高功率状态。

[0117] 8. 根据条款 6 所述的计算机实施的方法，其还包括停用所述第一模块，其中停用所述第一模块包括将所述处理器从所述高功率状态切换到所述低功率状态。

[0118] 9. 根据条款 6 所述的计算机实施的方法，其中所述处理器包括数字信号处理器或微处理器中的至少一个。

[0119] 10. 根据条款 5 所述的计算机实施的方法，其中所述第一模块包括配置来由微处理器执行的软件模块。

[0120] 11. 根据条款 10 所述的计算机实施的方法，其中启用所述第一模块包括使所述微处理器执行所述软件模块。

[0121] 12. 根据条款 5 所述的计算机实施的方法，其中所述操作包括使所述音频输入的

Petitioner Exhibit 1004, Page 197 of 1052

至少一部分传输到所述第二计算装置，并且还包括从所述第二计算装置接收语音识别结果。

[0122] 13. 根据条款 12 所述的计算机实施的方法，其中所述语音识别结果包括以下各者中的至少一个：所述音频输入的至少一部分的抄本以及对包括在所述音频输入的至少一部分中的智能代理查询的响应。

[0123] 14. 根据条款 12 所述的计算机实施的方法，其还包括：

[0124] 至少部分基于所述一个或多个值，启用所述第一计算装置的第二模块，其中所述第二模块配置来实施语音识别应用；以及

[0125] 通过所述语音识别应用来处理所述语音识别结果。

[0126] 15. 一种装置，其包括：

[0127] 第一处理器，其配置来：

[0128] 确定一个或多个值，其中所述一个或多个值包括以下各者中的至少一个：指示音频输入的能量水平的第一值，或指示所述音频输入包括语音的可能性的第二值；以及

[0129] 至少部分基于所述一个或多个值，使第二处理器启用；

[0130] 所述第二处理器配置来执行操作，其中所述操作包括以下各者中的至少一个：

[0131] 确定所述音频输入包括唤醒字；

[0132] 对所述音频输入的至少一部分执行语音识别，以获得语音识别结果；或

[0133] 使所述音频输入的至少一部分传输到第二装置。

[0134] 16. 根据条款 15 所述的装置，其中所述第一处理器包括数字信号处理器或微处理器中的至少一个。

[0135] 17. 根据条款 15 所述的装置，其中所述第二处理器包括数字信号处理器或微处理器中的至少一个。

[0136] 18. 根据条款 15 所述的装置，其还包括配置来存储所述音频输入的存储器缓冲模块。

[0137] 19. 根据条款 18 所述的装置，其中配置来存储所述音频输入的所述存储器缓冲模块包括环形缓冲区。

[0138] 20. 根据条款 15 所述的装置，其还包括与所述第一处理器通信的音频输入模块，其中所述音频输入模块配置来获得所述音频输入。

[0139] 21. 一种系统，其包括：

[0140] 音频输入模块，其配置来获得音频输入；

[0141] 与所述音频输入模块通信的第一模块；以及

[0142] 与所述第一模块通信的第二模块；

[0143] 其中所述第一模块配置来：

[0144] 至少部分基于所述音频输入来确定一个或多个值，以及

[0145] 至少部分基于所述一个或多个值，使所述第二模块启用；并且

[0146] 其中所述第二模块配置来对所述音频输入的至少一部分执行语音识别。

[0147] 22. 根据条款 21 所述的系统，其中所述一个或多个值包括所述音频输入的音量。

[0148] 23. 根据条款 22 所述的系统，其中只有在所述音频输入的所述音量在至少阈值持续时间内满足音量阈值时，所述第二模块才被启用。

Petitioner Exhibit 1004, Page 198 of 1052

[0149] 24. 根据条款 21 所述的系统，其中所述一个或多个值包括所述音频输入包括语音的可能性。

[0150] 25. 根据条款 21 所述的系统，其中所述一个或多个值包括指示以下可能性的分数：所述音频输入包括唤醒字。

[0151] 26. 根据条款 25 所述的系统，其中所述一个或多个值还包括指示以下可能性的分数：所述唤醒字由与所述唤醒字相关联的用户说出。

[0152] 27. 根据条款 21 所述的系统，其中所述第二模块配置来通过为所述音频输入的至少一部分生成语音识别结果对所述音频输入的至少一部分执行语音识别。

[0153] 28. 根据条款 21 所述的系统，其中所述第二模块配置来通过以下操作对所述音频输入的至少一部分执行语音识别：

[0154] 将所述音频输入传输到远程计算装置；以及

[0155] 从所述远程计算装置接收所述音频输入的至少一部分的语音识别结果。

Petitioner Exhibit 1004, Page 199 of 1052

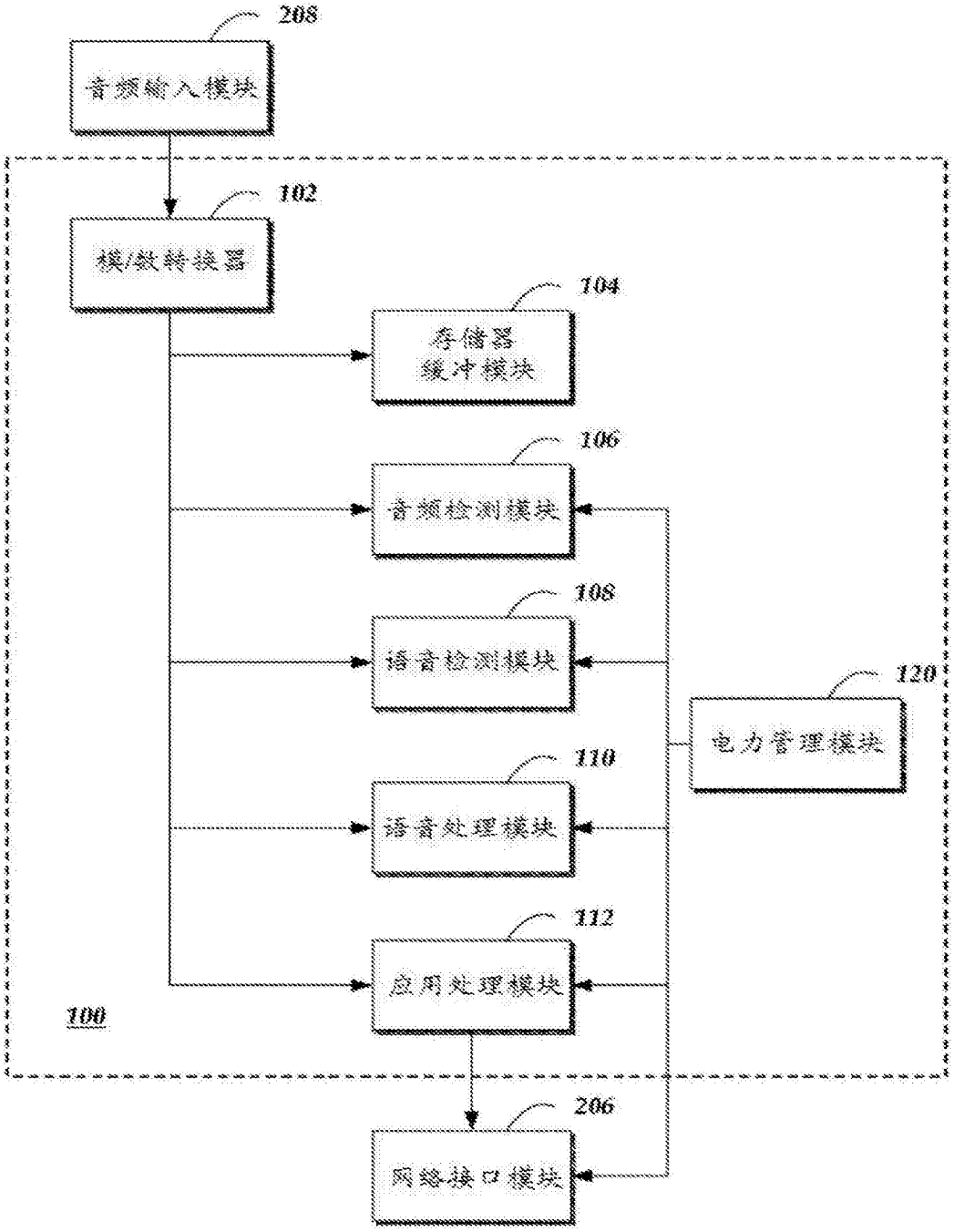


图 1

Petitioner Exhibit 1004, Page 200 of 1052

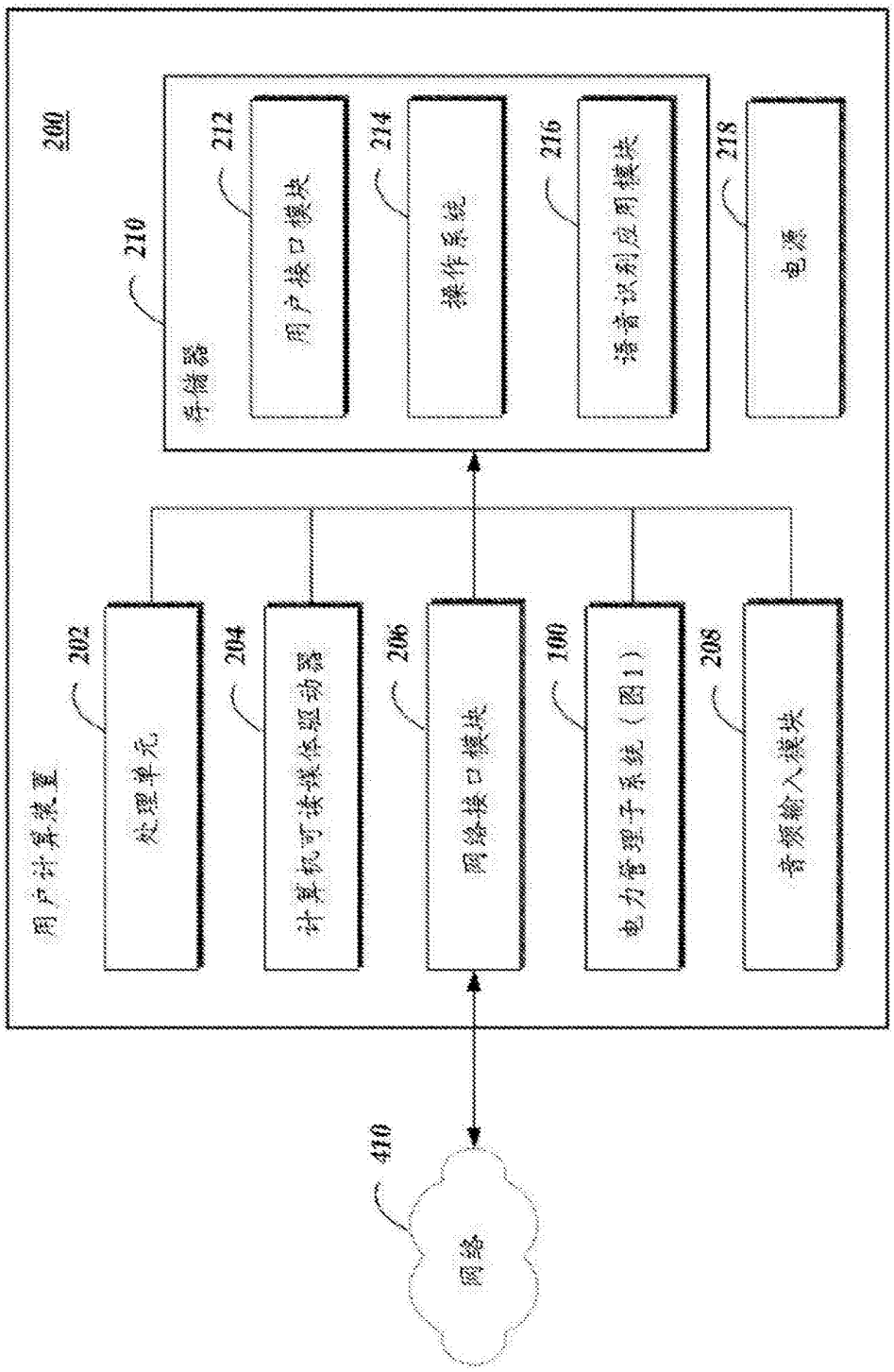


图 2

Petitioner Exhibit 1004, Page 201 of 1052

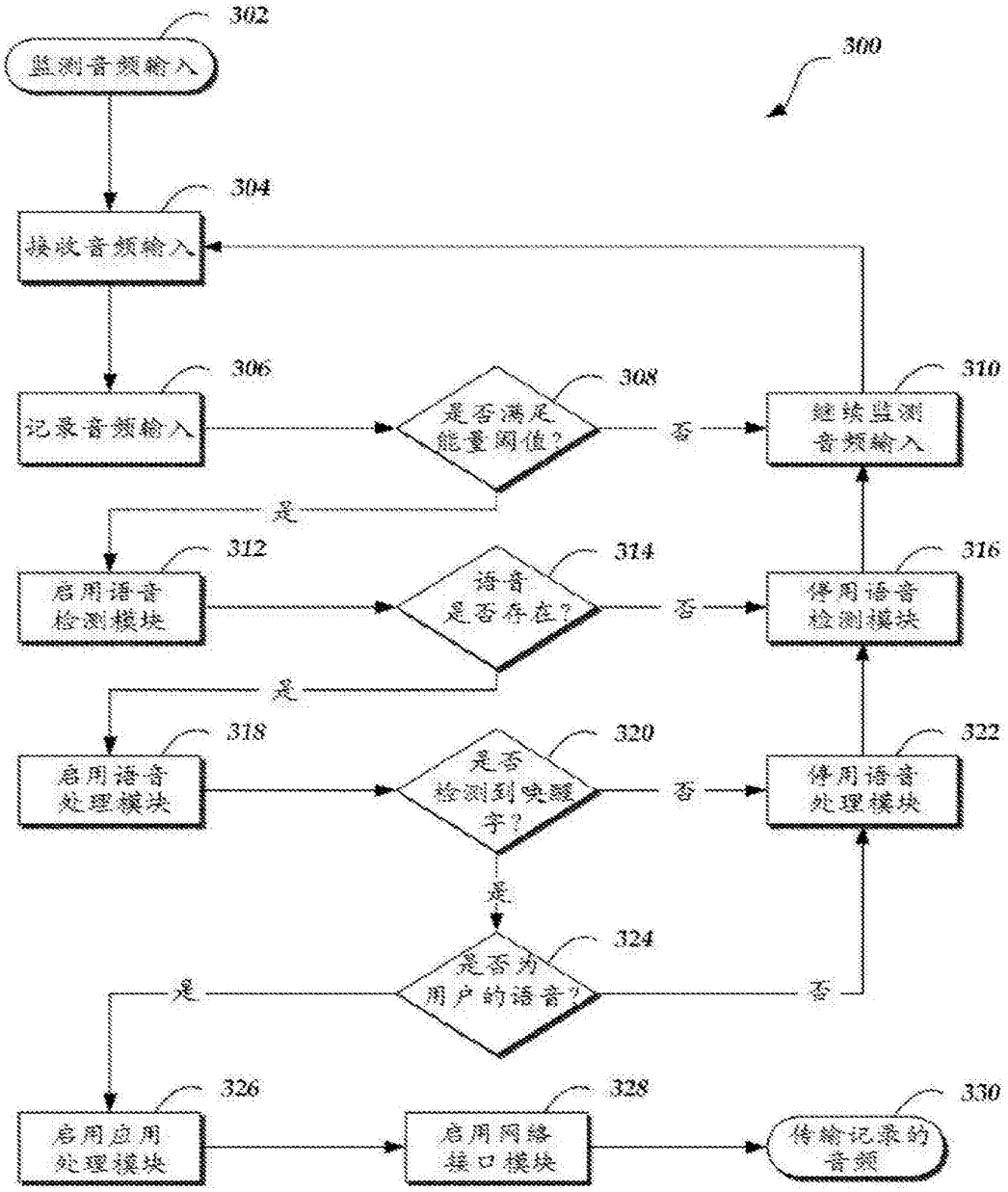


图 3

Petitioner Exhibit 1004, Page 202 of 1052

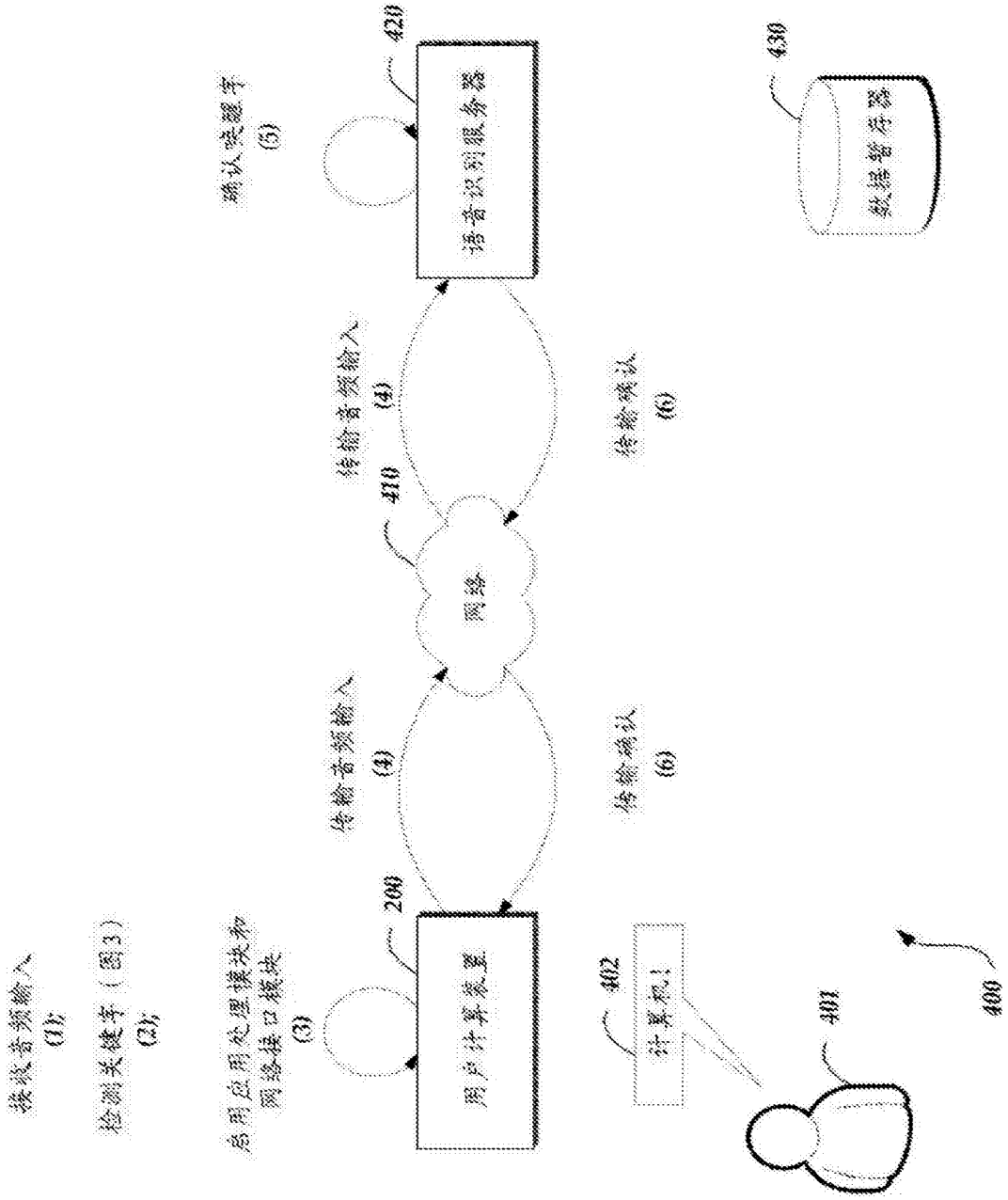


图 4A

Petitioner Exhibit 1004, Page 203 of 1052

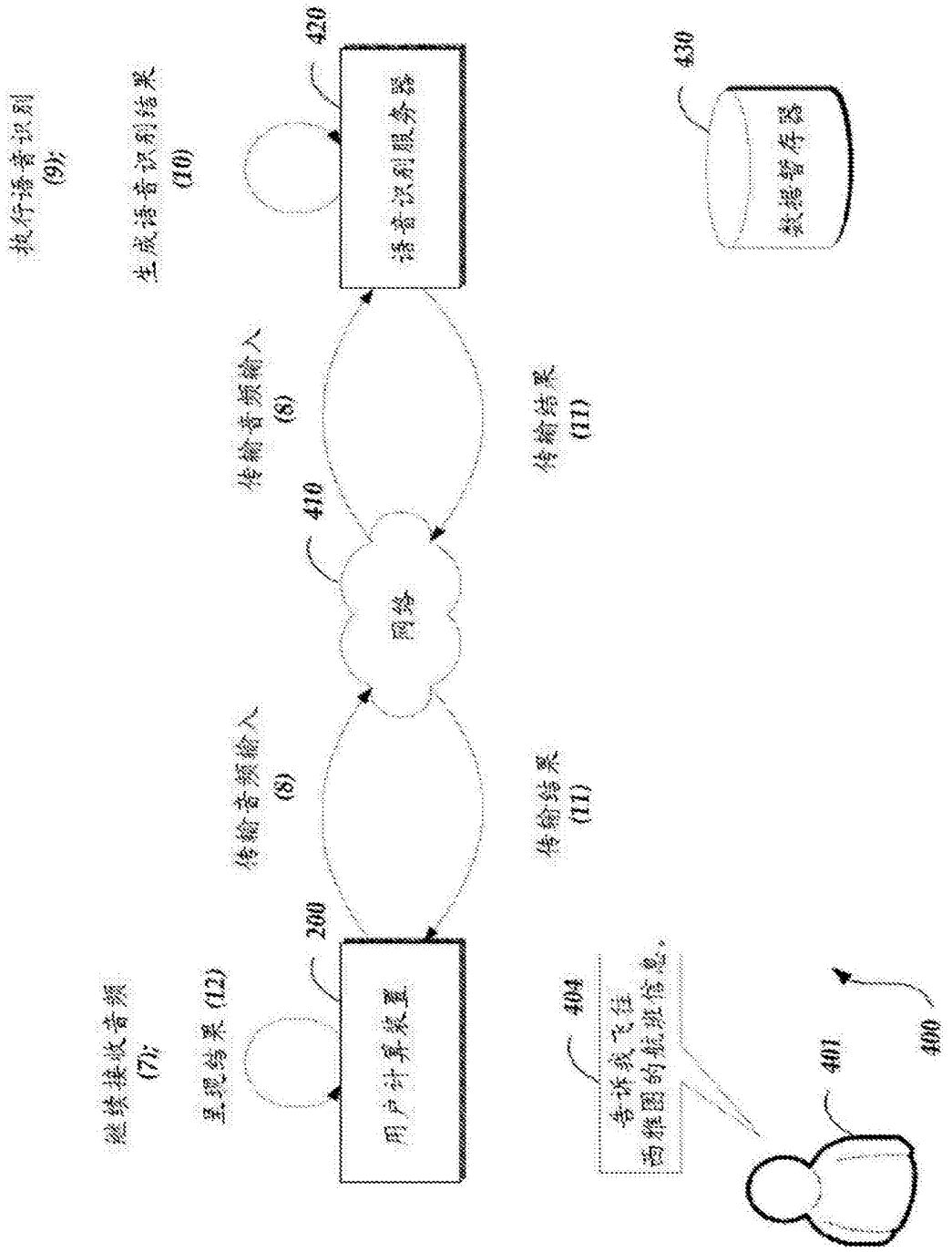


图 4B

Petitioner Exhibit 1004, Page 204 of 1052

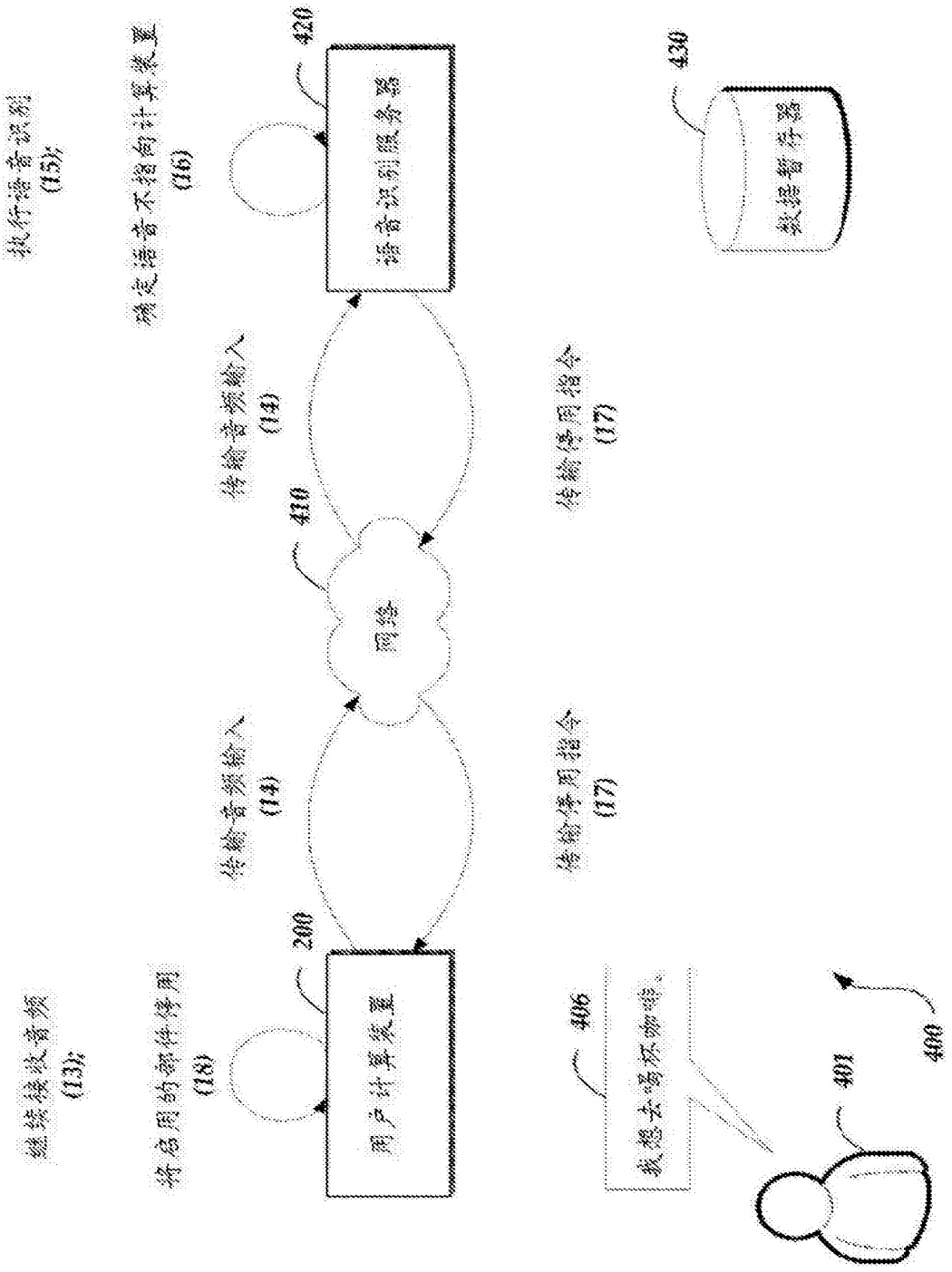


图 4C

Petitioner Exhibit 1004, Page 205 of 1052

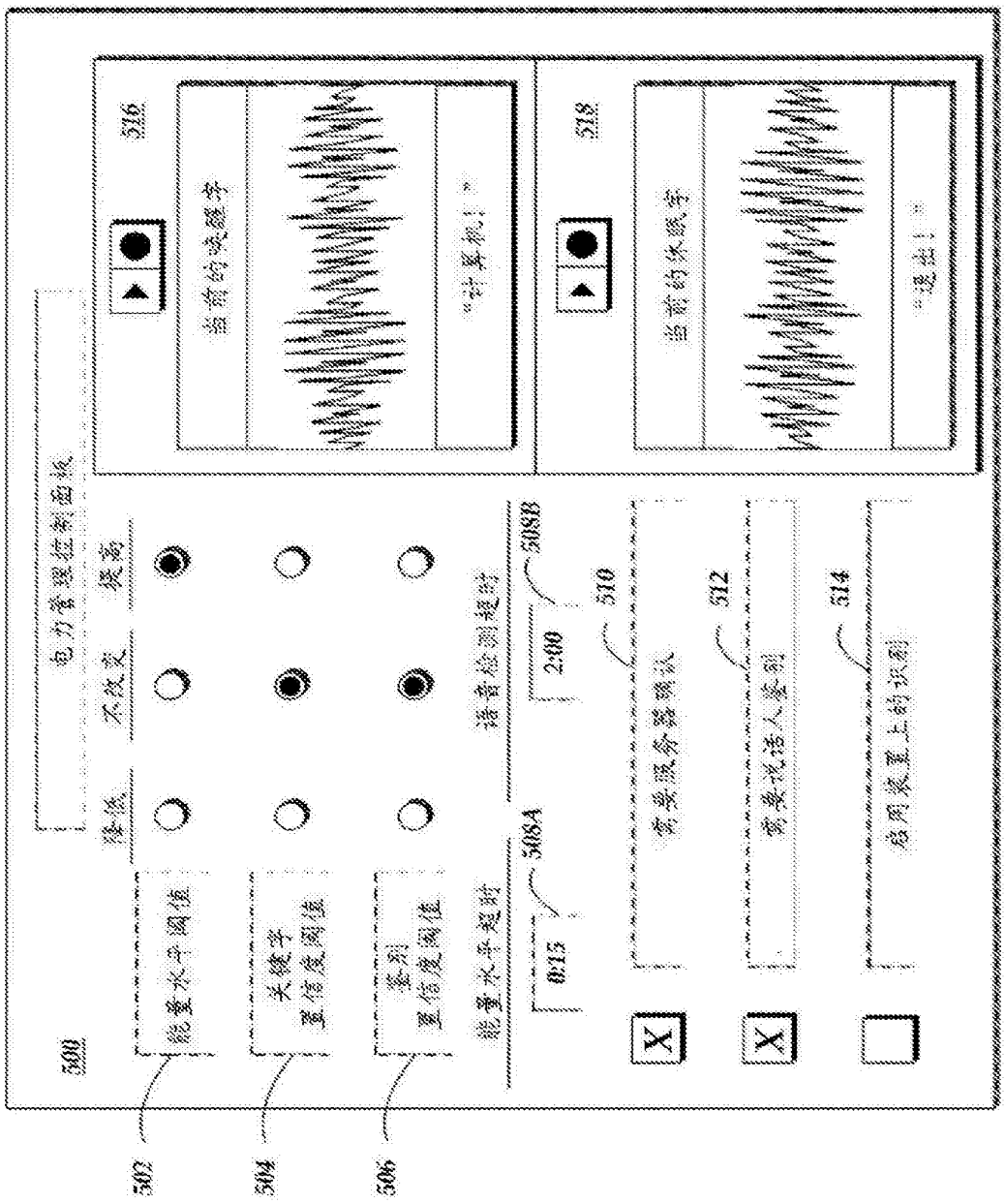


图 5

Petitioner Exhibit 1004, Page 206 of 1052

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Thomas Stephen Chlebeck/Trish McDonald
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
Submission- Information Disclosure Stmt	1806	1	180	180
		Total in USD ($)		**180**

DOCKET NO.: N0484.71090US04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Huy C. Ho
Art Unit:	2644

CERTIFICATE OF ELECTRONIC FILING UNDER 37 C.F.R. § 1.8

The undersigned hereby certifies that this paper, along with any paper referred to as being attached or enclosed, is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4), on the 5th day of October, 2017.

Electronic Signature: /Trish McDonald/

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed more than three months after the filing date of this application and after the mailing date of a first Office Action, but before the mailing date of any of a Final Action under 37 C.F.R. § 1.113, a Notice of Allowance under 37 C.F.R. § 1.311, or an action that otherwise closes prosecution in this application.

Please charge our Credit Card in the amount of $180.00 covering the fee set forth in 37 C.F.R. § 1.17(p).

Petitioner Exhibit 1004, Page 209 of 1052

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
CN 201480013903.1	**07-28-2017**	**Office Action**	**N0484.71090CN00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

Petitioner Exhibit 1004, Page 210 of 1052

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Thomas S. Chlebeck/
Thomas S. Chlebeck, Reg. No. 70,616
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: N0484.71090US04
Date: 2017-10-05
xNDDx

Petitioner Exhibit 1004, Page 211 of 1052

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 932 500 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
15.02.2017 Bulletin 2017/07

(21) Application number: **13811335.2**

(22) Date of filing: **09.12.2013**

(51) Int Cl.:
G10L 15/26 *(2006.01)* ***G10L 15/28*** *(2013.01)*
G10L 25/78 *(2013.01)*

(86) International application number:
PCT/US2013/073913

(87) International publication number:
WO 2014/093238 (19.06.2014 Gazette 2014/25)

(54) **SPEECH RECOGNITION POWER MANAGEMENT**

SPRACHERKENNUNGSLEISTUNGSVERWALTUNG

GESTION D'ALIMENTATION PAR RECONNAISSANCE DE LA PAROLE

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **11.12.2012 US 201213711510**

(43) Date of publication of application:
21.10.2015 Bulletin 2015/43

(73) Proprietor: **Amazon Technologies, Inc.**
Reno, NV 89507 (US)

(72) Inventors:
- **BASYE, Kenneth, John**
  Sutton, MA 01590 (US)
- **SECKER-WALKER, Hugh, Evan**
  Seattle, WA 98109-5210 (US)
- **DAVID, Tony**
  Seattle, WA 98109-5210 (US)
- **KNESER, Reinhard**
  Seattle, WA 98109-5210 (US)
- **ADAMS, Jeffrey, Penrod**
  Seattle, WA 98109-5210 (US)
- **SALVADOR, Stan, Weidner**
  Seattle, WA 98109-5210 (US)
- **KRISHNAMOORTHY, Mahesh**
  Seattle, WA 98109-5210 (US)

(74) Representative: **D Young & Co LLP**
120 Holborn
London EC1N 2DY (GB)

(56) References cited:
US-A- 5 983 186
US-A1- 2003 216 909
US-A1- 2010 277 579
US-A1- 2010 292 987

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Printed by Jouve, 75001 PARIS (FR)

EP 2 932 500 B1

Description

BACKGROUND

[0001] Computing devices may include speech recognition capabilities. For example, a computing device can capture audio input and recognize speech using an acoustic model and a language model. The acoustic model is used to generate hypotheses regarding which sound subword units (e.g., phonemes, etc.) correspond to speech based on the acoustic features of the speech. The language model is used to determine which of the hypotheses generated using the acoustic model is the most likely transcription of the speech based on lexical features of the language in which the speech is spoken. A computing device may also be capable of processing the recognized speech for specific speech recognition applications. For example, finite grammars or natural language processing techniques may be used to process speech. US 2003/216909 discloses (see the Abstract) that a subset of values is used to discriminate voice activity in a signal. The subset of values belongs to a larger set of values representing a segment of a signal, the larger set of values being suitable for speech recognition.

SUMMARY

[0002] The subject matter claimed herein is defined in the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] The various aspects and many of the attendant advantages of the present disclosure will become more readily appreciated as the same become better understood by reference to the following detailed description when taken in conjunction with the accompanying drawings, wherein:

FIG. 1 is a schematic diagram depicting an illustrative power management subsystem.

FIG. 2 is a schematic diagram depicting an illustrative user computing device including a power management subsystem.

FIG. 3 is flow diagram depicting an illustrative routine for speech recognition power management which may be implemented by the power management subsystem of FIG. 1.

FIG. 4A, FIG. 4B, and FIG. 4C are state diagrams depicting an illustrative operation of a distributed speech recognition system.

FIG. 5 is a pictorial diagram depicting an illustrative user interface that may be provided by a user computing device that includes a power management subsystem.

DETAILED DESCRIPTION

[0004] In some current approaches to speech recognition, speech recognition capabilities are allocated among one or more computing devices in a distributed computing environment. In a particular example of these approaches, a first computing device may be configured to capture audio input, and may transmit the audio input over a network to a second computing device. The second computing device may perform speech recognition on the audio input and generate a transcription of speech included in the audio input. The transcription of the speech then may be transmitted over the network from the second computing device back to the first computing device. In other current approaches, the first computing device may be configured to capture audio input and transcribe the audio input on its own.

[0005] In these and other current approaches, the first computing device may be configured to remain in a persistently active state. In such a persistently active state, the first computing device may continuously maintain a network connection to the second computing device. The first computing device may also continue to power any hardware used to implement its own speech recognition capabilities. One drawback of these approaches, among others, is that the first computing device may consume unacceptable amounts of energy to maintain the persistently active state. Such energy demands may prove especially problematic for mobile computing devices that rely on battery power. Still other problems are present in current approaches.

[0006] Accordingly, aspects of the present disclosure are directed to power management for speech recognition. A computing device may be provided with a power management subsystem that selectively activates or deactivates one or more modules of the computing device. This activation may be responsive to an audio input that includes one or more pre-designated spoken words, sometimes referred to herein as "keywords." A keyword that prompts the activation of one or more components may be activated is sometimes referred to herein as a "wakeword," while a keyword that prompts the deactivation of one or more components is sometimes referred to herein as a "sleepword." In a particular example, the computing device may include a selectively activated network interface module that, when activated, consumes energy to provide the computing device with connectivity to a second computing device, such as a speech recognition server or other computing device. The power management subsystem may process an audio input to determine that the audio input includes a wakeword, and activate the network interface module in response to determining that the audio input comprises the wakeword. With the network interface module activated, the power management subsystem may cause transmission of the audio input to a speech recognition server for processing.

[0007] The power management subsystem may itself include one or more selectively activated modules. In

Petitioner Exhibit 1004, Page 213 of 1052

some embodiments, one or more of the selectively activated modules are implemented as dedicated hardware (such as an integrated circuit, a digital signal processor or other type of processor) that may be switched from a low-power, deactivated state with relatively lesser functionality, to a high-power, activated state with relatively greater functionality, and vice versa. In other embodiments, one or more modules are implemented as software that includes computer-executable code carried out by one or more general-purpose processors. A software module may be activated (or deactivated) by activating (or deactivating) a general-purpose processor configured to or capable of carrying out the computer-executable code included in the software. In still further embodiments, the power management system includes both one or more hardware modules and one or more software modules.

[0008] The power management subsystem may further include a control module in communication with the one or more selectively activated modules. Such a control module is sometimes referred to herein as a "power management module," and may include any of the hardware or software described above. The power management module may cause the activation or deactivation of a module of the power management subsystem. In some embodiments, the power management module activates or deactivates one or more modules based at least in part on a characteristic of audio input obtained by an audio input module included in the computing device. For example, a module of the power management subsystem may determine one or more values, which values may include, for example, an energy level or volume of the audio input; a score corresponding to a likelihood that speech is present in the audio input; a score corresponding to a likelihood that a keyword is present in the speech; and other values. The module may communicate the one or more values to the power management module, which may either communicate with another module to cause activation thereof or communicate with the module from which the one or more values were received to cause deactivation of that module and/or other modules. In other embodiments, however, a first selectively activated module may communicate directly with a second selectively activated module to cause activation thereof. In such embodiments, no power management module need be present. In still further embodiments, a power management subsystem may be provided with one or more modules, wherein at least some of the one or more modules are in communication with each other but not with the power management module.

[0009] In an example implementation, the power management subsystem may include an audio detection module, which may be configured to determine an energy level or volume of an audio input obtained by the computing device. While the audio detection module may persistently monitor for audio input, the remaining components of the power management subsystem may remain in a low-power, inactive state until activated (either by the power management module or by a different module). If the audio detection module determines that an audio input meets a threshold energy level or volume, a speech detection module may be activated to determine whether the audio input includes speech. If the speech detection module determines that the audio input includes speech, a speech processing module included in the power management subsystem may be activated. The speech processing module may determine whether the speech includes a wakeword, and may optionally classify the speech to determine if a particular user spoke the wakeword. If the speech processing module determines that the speech includes the wakeword, an application processing module may be activated, which application processing module may implement a speech recognition application module stored in memory of the computing device. The speech recognition application may include, for example, an intelligent agent frontend, such as that described in "Intelligent Automated Assistant," which was filed on January 10, 2011 and published as U.S. Publication No. 2012/0016678 on January 19, 2012. The selectively activated network interface module may also be activated as discussed above, and the audio input may be transmitted to a remote computing device for processing. This example implementation is discussed in greater detail below with respect to FIG. 3. Alternately, the power management subsystem may, responsive to detecting the wakeword, activate a processing unit that implements any on-device speech recognition capabilities of the computing device.

[0010] By selectively activating modules of the computing device, the power management subsystem may advantageously improve the energy efficiency of the computing device. The power management subsystem may further improve the energy efficiency of the computing device by selectively activating one or more of its own modules. While such implementations are particularly advantageous for computing devices that rely on battery power, all computing devices for which power management may be desirable can benefit from the principles of the present disclosure.

[0011] Turning now to FIG. 1, an illustrative power management subsystem 100 that may be included in a computing device is shown. The power management subsystem 100 may include an analog/digital converter 102; a memory buffer module 104; an audio detection module 106; a speech detection module 108; a speech processing module 110; an application processing module 112; and a power management module 120. The memory buffer module 104 may be in communication with the audio detection module 106; the speech detection module 108; the speech processing module 110; the application processing module 112; and a network interface module 206. The power management module 120 may likewise be in communication with audio detection module 106; the speech detection module 108; the speech processing module 110; the application processing module 112; and the network interface module 206.

Petitioner Exhibit 1004, Page 214 of 1052

[0012] The analog/digital converter 102 may receive an audio input from an audio input module 208. The audio input module 208 is discussed in further detail below with respect to FIG. 2. The analog/digital converter 102 may be configured to convert analog audio input to digital audio input for processing by the other components of the power management subsystem 100. In embodiments in which the audio input module 208 obtains digital audio input (e.g., the audio input module 208 includes a digital microphone or other digital audio input device), the analog/digital converter 102 may optionally be omitted from the power management subsystem 100. Thus, the audio input module 208 may provide audio input directly to the other modules of the power management subsystem 100.

[0013] The memory buffer module 104 may include one or more memory buffers configured to store digital audio input. The audio input obtained by the audio input module 208 (and, if analog, converted to digital form by the analog/digital converter 102) may be recorded to the memory buffer module 104. The audio input recorded to the memory buffer module 104 may be accessed by other modules of the power management subsystem 100 for processing by those modules, as discussed further herein.

[0014] The one or more memory buffers of the memory buffer module 104 may include hardware memory buffers, software memory buffers, or both. The one or more memory buffers may have the same capacity, or different capacities. A memory buffer of the memory buffer module 104 may be selected to store an audio input depending on which other modules are activated. For example, if only the audio detection module 106 is active, an audio input may be stored to a hardware memory buffer with relatively small capacity. However, if other modules are activated, such as the speech detection module 108; the speech processing module 110; the application processing module 112; and/or the network interface module 206, the audio input may be stored to a software memory buffer of relatively larger capacity. In some embodiments, the memory buffer module 104 includes a ring buffer, in which audio input may be recorded and overwritten in the order that it is obtained by the audio input module 208.

[0015] The audio detection module 106 may process audio input to determine an energy level of the audio input. In some embodiments, the audio detection module 106 includes a low-power digital signal processor (or other type of processor) configured to determine an energy level (such as a volume, intensity, amplitude, etc.) of an obtained audio input and for comparing the energy level of the audio input to an energy level threshold. The energy level threshold may be set according to user input, or may be set automatically by the power management subsystem 100 as further discussed below with respect to FIG. 3. In some embodiments, the audio detection module 106 is further configured to determine that the audio input has an energy level satisfying a threshold for at least a threshold duration of time. In such embodiments, high-energy audio inputs of relatively short duration, which may correspond to sudden noises that are relatively unlikely to include speech, may be ignored and not processed by other components of the power management subsystem 100.

[0016] If the audio detection module 106 determines that the obtained audio input has an energy level satisfying an energy level threshold, it may communicate with the power management module 120 to direct the power management module 120 to activate the speech detection module 108. Alternately, the audio detection module 106 may communicate the energy level to the power management module 120, and the power management module 120 may compare the energy level to the energy level threshold (and optionally to the threshold duration) to determine whether to activate the speech detection module 108. In another alternative, the audio detection module 106 may communicate directly with the speech detection module 108 to activate it. Optionally, the power management module 120 (or audio detection module 106) may direct the audio input module 208 to increase its sampling rate (whether measured in frame rate or bit rate) responsive to the audio detection module 106 determining that the audio input has an energy level satisfying a threshold.

[0017] The speech detection module 108 may process audio input to determine whether the audio input includes speech. In some embodiments, the speech detection module 108 includes a low-power digital signal processor (or other type of processor) configured to implement one or more techniques to determine whether the audio input includes speech. In some embodiments, the speech detection module 108 applies voice activity detection (VAD) techniques. Such techniques may determine whether speech is present in an audio input based on various quantitative aspects of the audio input, such as the spectral slope between one or more frames of the audio input; the energy levels of the audio input in one or more spectral bands; the signal-to-noise ratios of the audio input in one or more spectral bands; or other quantitative aspects. In other embodiments, the speech detection module 108 implements a limited classifier configured to distinguish speech from background noise. The classifier may be implemented by techniques such as linear classifiers, support vector machines, and decision trees. In still other embodiments, the speech detection module 108 applies Hidden Markov Model (HMM) or Gaussian Mixture Model (GMM) techniques to compare the audio input to one or more acoustic models, which acoustic models may include models corresponding to speech, noise (such as environmental noise or background noise), or silence. Still other techniques may be used to determine whether speech is present in the audio input.

[0018] Using any of the techniques described above, the speech detection module 108 may determine a score or a confidence level whose value corresponds to a likelihood that speech is actually present in the audio input (as used herein, "likelihood" may refer to common usage,

Petitioner Exhibit 1004, Page 215 of 1052

whether something is likely, or the usage in statistics). If the score satisfies a threshold, the speech detection module 108 may determine that speech is present in the audio input. However, if the score does not satisfy the threshold, the speech detection module 108 may determine that there is no speech in the audio input.

[0019] The speech detection module 108 may communicate its determination as to whether speech is present in the audio input to the power management module 120. If speech is present in the audio input, the power management module 120 may activate the speech processing module 110 (alternately, the speech detection module 108 may communicate directly with the speech processing module 110). If speech is not present in the audio input, the power management module 120 may deactivate the speech detection module 108. Alternately, the speech detection module 108 may communicate the score to the power management module 120, whereupon the power management module 120 may determine whether to activate the speech processing module 110 or deactivate the speech detection module 108.

[0020] The speech processing module 110 may process the audio input to determine whether a keyword is included in the speech. In some embodiments, the speech processing module 110 includes a microprocessor configured to detect a keyword in the speech, such as a wakeword or sleepword. The speech processing module 110 may be configured to detect the keyword using HMM techniques, GMM techniques, or other speech recognition techniques.

[0021] The speech processing module 110 may be able to separate speech that incidentally includes a keyword from a deliberate utterance of the keyword by determining whether the keyword was spoken immediately before or after one or more other phonemes or words. For example, if the keyword is "ten," the speech processing module 110 may be able to distinguish the user saying "ten" by itself from the user saying "ten" incidentally as part of the word "Tennessee," the word "forgotten," the word "stent," or the phrase "ten bucks."

[0022] The speech processing module 110 may further be configured to determine whether the speech is associated with a particular user of a computing device in which the power management subsystem 100 is included, or whether the speech corresponds to background noise; audio from a television; music; or the speech of a person other than the user, among other classifications. This functionality may be implemented by techniques such as linear classifiers, support vector machines, and decision trees, among other techniques for classifying audio input.

[0023] Using any of the techniques described above, the speech processing module 110 may determine a score or confidence level whose value corresponds to a likelihood that a keyword is actually present in the speech. If the score satisfies a threshold, the speech processing module 110 may determine that the keyword is present in the speech. However, if the score does not satisfy the threshold, the speech processing module 110 may determine that there is no keyword in the speech.

[0024] The speech processing module 110 may communicate its determination as to whether a keyword is present in the speech to the power management module 120. If the keyword is present in the speech and the keyword is a wakeword, the power management module 120 may activate application processing module 112 and the network interface module 206 (alternately, the speech processing module 110 may communicate directly with these other modules). If the keyword is not present in the audio input (or the keyword is a sleepword), the power management module 120 may deactivate the speech processing module 110 and the speech detection module 108. Alternately, the speech processing module 110 may communicate the score to the power management module 120, whereupon the power management module 120 may determine whether to activate the application processing module 112 and the network interface module 206 or deactivate the speech processing module 110 and the speech detection module 108. In some embodiments, these activations and/or deactivations only occur if the speech processing module 110 determines that a particular user spoke the speech that includes the keyword.

[0025] The application processing module 112 may include a microprocessor configured to implement a speech recognition application provided with a computing device in which the power management subsystem is included. The speech recognition application may include any application for which speech recognition may be desirable, such as a dictation application, a messaging application, an intelligent agent frontend application, or any other application. The speech recognition application may also be configured to format the speech (e.g., by compressing the speech) for transmission over a network to a remote computing device, such as a speech recognition server.

[0026] In some embodiments, the application processing module 112 includes a dedicated microprocessor for implementing the speech recognition application. In other embodiments, the application processing module 112 includes a general-purpose microprocessor that may also implement other software provided with a computing device in which the power management subsystem 100 is included, such as the processing unit 202 shown in FIG. 2, which is discussed further below.

[0027] The network interface module 206, when activated, may provide connectivity over one or more wired or wireless networks. Upon its activation, the network interface module 206 may transmit the received audio input recorded to the memory buffer module 104 over a network to a remote computing device, such as a speech recognition server. The remote computing device may return recognition results (e.g., a transcription or response to an intelligent agent query) to the computing device in which the network interface module 206 is included, whereupon the network interface module 206 may provide the received recognition results to the ap-

Petitioner Exhibit 1004, Page 216 of 1052

plication processing module 112 for processing. The network interface module 206 is discussed further below with respect to FIG. 2.

[0028] The modules of the power management subsystem 100 may be combined or rearranged without departing from the scope of the present disclosure. The functionality of any module described above may be allocated among multiple modules, or combined with a different module. As discussed above, any or all of the modules may be embodied in one or more integrated circuits, one or more general-purpose microprocessors, or in one or more special-purpose digital signal processors or other dedicated microprocessing hardware. One or more modules may also be embodied in software implemented by a processing unit 202 included in a computing device, as discussed further below with respect to FIG. 2. Further, one or more of the modules may be omitted from the power management subsystem 100 entirely.

[0029] Turning now to FIG. 2, a user computing device 200 in which a power management subsystem 100 may be included is illustrated. The user computing device 200 includes a processing unit 202; a non-transitory computer-readable medium drive 204; a network interface module 206; the power management subsystem 100 as shown in FIG. 1; and an audio input module 208, all of which may communicate with one another by way of a communication bus. The user computing device 200 may also include a power supply 218, which may provide power to the various components of the user computing device 200, such as the processing unit 202; the non-transitory computer-readable medium drive 204; the network interface module 206; the power management subsystem 100 as shown in FIG. 1; and the audio input module 208.

[0030] The processing unit 202 may include one or more general-purpose microprocessors configured to communicate to and from the memory 210 to implement various software modules stored therein, such as a user interface module 212, operating system 214, and speech recognition application module 216. The processing unit 202 may also communicate with the power management subsystem 100 and may further implement any modules of the power management subsystem 100 embodied in software. Accordingly, the processing unit 202 may be configured to implement any or all of the audio detection module 106; the speech detection module 108; the speech processing module 110; the application processing module 112; and the power management module 120. Further, the processing unit 202 may be configured to implement on-device automatic speech recognition capabilities that may be provided with the user computing device 200.

[0031] The memory 210 generally includes RAM, ROM, and/or other persistent or non-transitory computer-readable storage media. The user interface module 212 may be configured to present a user interface via a display of the user computing device 200 (not shown). The user interface module 212 may be further configured to process user input received via a user input device (not shown), such as a mouse, keyboard, touchscreen, keypad, etc. The user interface presented by the user interface module 212 may provide a user with the opportunity to customize the operation of the power management subsystem 100 and/or other operations implemented by the user computing device 200. An example of a user interface is discussed further below with respect to FIG. 5. The memory 210 may additionally store an operating system 214 that provides computer program instructions for use by the processing unit 202 in the general administration and operation of the user computing device 200. The memory 210 can further include computer program instructions that the application processing module 112 and/or processing unit 202 executes in order to implement one or more embodiments of a speech recognition application module 216. As discussed above, the speech recognition application module 216 may be any application that may use speech recognition results, such as a dictation application; a messaging application; an intelligent agent application frontend; or any other application that may advantageously use speech recognition results. In some embodiments, the memory 210 may further include an automatic speech recognition engine (not shown) that may be implemented by the processing unit 202.

[0032] The non-transitory computer-readable medium drive 204 may include any electronic data storage known in the art. In some embodiments, the non-transitory computer-readable medium drive 204 stores one or more keyword models (e.g., wakeword models or sleepword models) to which an audio input may be compared by the power management subsystem 100. The non-transitory computer-readable medium drive 204 may also store one or more acoustic models and/or language models for implementing any on-device speech recognition capabilities of the user computing device 200. Further information regarding language models and acoustic models may be found in U.S. Patent App. No. 13/587,799, entitled "DISCRIMINATIVE LANGUAGE MODEL PRUNING," filed on August 16, 2012; and in U.S. Patent App. No. 13/592,157, entitled "UNSUPERVISED ACOUSTIC MODEL TRAINING," filed on August 22, 2012.

[0033] The network interface module 206 may provide the user computing device 200 with connectivity to one or more networks, such as a network 410, discussed further below with respect to FIG: 4A, FIG. 4B, and FIG. 4C. The processing unit 202 and the power management subsystem 100 may thus receive instructions and information from remote computing devices that may also communicate via the network 410, such as a speech recognition server 420, as also discussed further below. In some embodiments, the network interface module 206 comprises a wireless network interface that provides the user computing device 200 with connectivity over one or more wireless networks.

[0034] In some embodiments, the network interface

Petitioner Exhibit 1004, Page 217 of 1052

module 206 is selectively activated. While the network interface module 206 is in a deactivated or "sleeping" state, it may provide limited or no connectivity to networks or computing systems so as to conserve power. In some embodiments, the network interface module 206 is in a deactivated state by default, and becomes activated responsive to a signal from the power management subsystem 100. While the network interface module 206 in an activated state, it may provide a relatively greater amount of connectivity to networks or computing systems, such that the network interface module 206 enables the user computing device 200 to send audio input to a remote computing device and/or receive a keyword confirmation, speech recognition result, or deactivation instruction from the remote computing device, such as a speech recognition server 420.

[0035] In a particular, non-limiting example, the network interface module 206 may be activated responsive to the power management subsystem 100 determining that an audio input includes a wakeword. The power management subsystem 100 may cause transmission of the audio input to a remote computing device (such as a speech recognition server 420) via the activated network interface module 206. Optionally, the power management subsystem 100 may obtain a confirmation of a wakeword from a remote computing device before causing the transmission of subsequently received audio inputs to the remote computing device. The power management subsystem 100 may later deactivate the activated network interface module 206 in response to receiving a deactivation instruction from the remote computing device, in response to determining that at least a predetermined amount of time has passed since an audio input satisfying an energy level threshold has been obtained, or in response to receiving an audio input that includes a sleepword.

[0036] The audio input module 208 may include an audio input device, such as a microphone or array of microphones, whether analog or digital. The microphone or array of microphones may be implemented as a directional microphone or directional array of microphones. In some embodiments, the audio input module 208 receives audio and provides the audio to the power management subsystem 100 for processing, substantially as discussed above with respect to FIG. 1. The audio input module 208 may also receive instructions from the power management subsystem 100 to set a sampling rate (whether in frame rate or bitrate) for obtaining audio. The audio input module 208 may also (or instead) include one or more piezoelectric elements and/or micro-electrical-mechanical systems (MEMS) that can convert acoustic energy to an electrical signal for processing by the power management subsystem 100. The audio input module 208 may further be provided with amplifiers, rectifiers, and other audio processing components as desired.

[0037] One or more additional input devices such as light sensors, position sensors, image capture devices, or the like may be provided with the user computing device 200. Such additional input devices are not shown in FIG. 2 so as not to obscure the principles of the present disclosure. In some embodiments, an additional input device may detect the occurrence or non-occurrence of a condition. Information pertaining to such conditions may be provided to the power management subsystem 100 to determine whether one or more components of the user computing device 200 or the power management subsystem 100 should be activated or deactivated. In one embodiment, the additional input device includes a light sensor configured to detect a light level. The power management module 120 may only act network interface module 206 may only be activated if the light level detected by the light sensor does not satisfy a threshold. In another embodiment, the additional input device includes an image capture device configured with facial recognition capabilities. In this embodiment, the network interface module 206 may only be activated if the image capture device recognizes the face of a user associated with the user computing device 200. More information on controlling speech recognition capabilities with input devices may be found in U.S. Patent App. No. 10/058,730, entitled "AUTOMATIC SPEECH RECOGNITION SYSTEM AND METHOD," filed on January 30, 2002, which published as U.S. Patent Pub. No. 2003/0144844 on July 31, 2003. Further information on controlling speech recognition capabilities may be found in U.S. Patent No. 8,326,636, entitled "USING A PHYSICAL PHENOMENON DETECTOR TO CONTROL OPERATION OF A SPEECH RECOGNITION ENGINE," which issued on December 4, 2012.

[0038] Still further input devices may be provided, which may include user input devices such as mice, keyboards, touchscreens, keypads, etc. Likewise, output devices such as displays, speakers, headphones, etc. may be provided. In a particular example, one or more output devices configured to present speech recognition results in audio format (e.g., via text-to-speech) or in visual format (e.g., via a display) may be included with the user computing device 200. Such input and output devices are well known in the art and need not be discussed in further detail herein, and are not shown in FIG. 2 so as to avoid obscuring the principles of the present disclosure.

[0039] The power supply 218 may provide power to the various components of the user computing device 200. The power supply 218 may include a wireless or portable power supply, such as a disposable or rechargeable battery or battery pack; or may include a wired power supply, such as an alternating current (AC) power supply configured to be plugged into an electrical outlet. In some embodiments, the power supply 218 communicates the level of power that it can supply to the power management subsystem 100 (e.g., a percentage of battery life remaining, whether the power supply 218 is plugged into an electrical outlet, etc.). In some embodiments, the power management subsystem 100 selectively activates or deactivates one or more modules based at least in part

Petitioner Exhibit 1004, Page 218 of 1052

on the power level indicated by the power supply. For example, if the user computing device 200 is plugged in to an electrical outlet, the power management subsystem 100 may activate the network interface module 206 and leave it in an activated state. If the user computing device 200 is running on battery power, the power management subsystem 100 may selectively activate and deactivate the network interface module 206 as discussed above.

[0040] Turning now to FIG. 3, an illustrative routine 300 is shown in which modules of the power management subsystem 100 may be selectively activated for processing an audio input. The illustrative routine 300 represents an escalation of processing and/or power consumption, as modules that are activated later in the illustrative routine 300 may have relatively greater processing requirements and/or power consumption.

[0041] The illustrative routine 300 may begin at block 302 as the audio input module 208 monitors for audio input. The audio input module 208 may receive an audio input at block 304. At block 306, the received audio input may be recorded to the memory buffer module 104. At block 308, the audio detection module 106 may determine whether the audio input has an energy level that satisfies an energy level threshold (and, optionally, whether the audio input has an energy level that satisfies an energy level threshold for at least a threshold duration). If the audio input's energy level does not satisfy the energy level threshold, the audio input module 208 may continue to monitor for audio input in block 310 until another audio input is received.

[0042] Returning to block 308, if the audio detection module 106 determines that the audio input has an energy level satisfying a threshold, the power management module 120 may activate the speech detection module 108 at block 312 (alternately, the audio detection module 106 may directly activate the speech detection module 108, and the power management module 120 may be omitted as well in the following blocks). At block 314, the speech detection module 108 may determine whether speech is present in the obtained audio input, substantially as discussed above with respect to FIG. 1. If the speech detection module 108 determines that speech is not present (or not likely to be present) in the audio input, the power management module 120 may deactivate the speech detection module 108 at block 316. The audio input module 208 may then continue to monitor for audio input in block 310 until another audio input is received.

[0043] Returning to block 314, if the speech detection module 108 determines that the audio input includes speech, the power management module 120 may activate the speech processing module 110 at block 318. As discussed above, the speech processing module 110 may determine whether a wakeword is present in the speech at block 320. If the speech processing module 110 determines that the wakeword is not present in the speech (or not likely to be present in the speech), the speech processing module 110 may be deactivated at block 322. The speech detection module 108 may also be deactivated at block 316. The audio input device 208 may then continue to monitor for audio input in block 310 until another audio input is received.

[0044] Returning to block 320, if in some embodiments, the speech processing module 110 determines that the wakeword is present in the speech, user 401 the speech processing module 110 optionally determines in block 324 whether the speech is associated with a particular user (e.g., whether the wakeword was spoken by the user), substantially as discussed above with respect to FIG. 1. If the speech is not associated with the particular user, the speech processing module 110 may be deactivated at block 322. The speech detection module 108 may also be deactivated at block 316. The audio input device 208 may then continue to monitor for audio input in block 310 until another audio input is received. If the speech is associated with the particular user, the illustrative routine 300 may proceed to block 326. In other embodiments, block 324 may be omitted, and the illustrative routine 300 may proceed directly from block 320 to block 326 responsive to the speech processing module 110 determining that a wakeword is present in the speech.

[0045] At block 326, the power management module 120 may activate the application processing module 112, which may implement the speech recognition application module 216 shown in FIG. 2. The power management module 120 may also activate the network interface module 206 at block 328. With the network interface module 206 activated, the audio input recorded to the memory buffer module 104 may be transmitted over a network via the network interface module 206. In some embodiments, while the network interface module 206 is active, subsequently obtained audio inputs are provided from the audio input module 208 directly to the application processing module 112 and/or the network interface module 206 for transmission to the remote computing device. However, in other embodiments, any or all of the speech detection module 108, speech processing module 110, and application processing module 112 process the audio input before providing it to the network interface module 206 to be transmitted over the network 410 to a remote computing device.

[0046] In some embodiments, not shown, the power management subsystem 100 waits until the remote computing device returns a confirmation that the wakeword is present in the first audio input to transmit subsequent audio inputs for recognition. If no confirmation of the wakeword is provided by the remote computing device, or if a deactivation instruction is received via the network interface module 206, the network interface module 206 and one or more modules of the power management subsystem 100 may be deactivated.

[0047] As many of the operations of the power management subsystem 100 may generate probabilistic rather than exact determinations, errors may occur during the illustrative routine 300. In some instances, a particular module of the power management subsystem 100 may

Petitioner Exhibit 1004, Page 219 of 1052

provide a "false positive," causing one or more modules to be incorrectly activated. For example, the speech detection module 108 may incorrectly determine that speech is present at block 314, or the speech processing module 110 may incorrectly determine that the speech includes the wakeword at block 320or that the speech belongs to the user at block 324. Adaptive thresholding and cross-validation among the modules of the power management subsystem 100 may be advantageously used to reduce false positives. Two examples of adaptive thresholding are discussed herein, but other types of adaptive thresholding are possible. As discussed above, the speech detection module may determine that speech is present in an audio input at block 314. However, the speech processing module 110, which may recognize speech more precisely than the speech detection module 108 owing to its superior processing power, may determine that in fact no speech is present in the audio input. Accordingly, the speech processing module 110 may direct the speech detection module 108 to increase its score threshold for determining that speech is present in the audio input, so as to reduce future false positives. Likewise, if the remote computing device (such as a speech recognition server 420) includes speech recognition capabilities, the remote computing device may transmit to the user computing device 200 an indication that no wakeword was present in the speech, even though the speech processing module 110 may have indicated that the wakeword was present. Accordingly, the score threshold of the speech processing module 110 for determining that the wakeword is present in the speech may be increase, so as to reduce future false positives. Further, a user interface may be provided so that a user may increase one or more score thresholds to reduce false positives, as further described below with respect to FIG. 5.

[0048] In other instances, a particular component may provide a "false negative," such that components of the power management subsystem 100 are not activated and/or the network interface module 206 is not activated, even though the user has spoken the wakeword. For example, the speech detection module 108 may incorrectly determine that no speech is present at block 314, or the speech processing module 110 may incorrectly determine that the speech does not include the wakeword at block 320 or that the speech does not belong to the user at block 324. To reduce the likelihood of false negatives, the power management subsystem 100 may periodically lower the threshold scores, e.g., lower the score required to satisfy the thresholds in blocks 314, 320, and/or 324. The threshold may continue to be lowered until one or more false positives are obtained, as described above. Once one or more false positives are obtained, the threshold may not be lowered further, or may be slightly increased. Further, a user interface may accordingly be provided so that a user may decrease one or more score thresholds to reduce false negatives, as further described below with respect to FIG. 5.

[0049] In some embodiments, not all activated components are deactivated if a negative result is obtained at any of blocks 314, 320, or 324. For example, if a wakeword is not recognized at block 320, the speech processing module 110 may be deactivated at block 322, but the speech detection module 108 may remain activated. Additionally, blocks may be skipped in some implementations. In some embodiments, a score satisfying a threshold at either of blocks 314 or 320 prompts one or more subsequent blocks to be skipped. For example, if the speech processing module 110 determines with very high confidence that the wakeword is present in the speech at block 320, the illustrative routine 300 may skip directly to block 326

[0050] Further, in some embodiments, the user computing device 200 may include an automatic speech recognition engine configured to be executed by the processing unit 202. As such on-device speech recognition may have especially high power consumption, the processing unit 202 may only implement the automatic speech recognition engine to recognize speech responsive to the speech processing module 110 determining that the wakeword has been spoken by a user.

[0051] With reference now to FIG. 4A, FIG. 4B, and FIG. 4C, example operations of a distributed speech recognition service are shown in the illustrative environment 400. The environment 400 may include a user 401; a user computing device 200 as described above; a network 410; a speech recognition server 420; and a data store 430.

[0052] The network 410 may be any wired network, wireless network or combination thereof. In addition, the network 410 may be a personal area network, local area network, wide area network, cable network, satellite network, cellular telephone network, or combination thereof. Protocols and devices for communicating via the Internet or any of the other aforementioned types of communication networks are well known to those skilled in the art of computer communications, and thus need not be described in more detail herein.

[0053] The speech recognition server 420 may generally be any computing device capable of communicating over the network 410. In some embodiments, the speech recognition server 420 is implemented as one or more server computing devices, though other implementations are possible. The speech recognition server 420 may be capable of receiving audio input over the network 410 from the user computing device 200. This audio input may be processed in a number of ways, depending on the implementation of the speech recognition server 420. In some embodiments, the speech recognition server 420 processes the audio input received from the user computing device 200 to confirm that a wakeword is present (e.g., by comparing the audio input to a known model of the wakeword), and transmits the confirmation to the user computing device 200. The speech recognition server 420 may further be configured to identify a user 401 that spoke the wakeword using known speaker

Petitioner Exhibit 1004, Page 220 of 1052

identification techniques.

[0054] The speech recognition server 420 may process the audio input received from the user computing device 200 to determine speech recognition results from the audio input. For example, the audio input may include a spoken query for an intelligent agent to process; speech to be transcribed to text; or other speech suitable for a speech recognition application. The speech recognition server 420 may transmit the speech recognition results over the network 410 to the user computing device 200. Further information pertaining to distributed speech recognition applications may be found in U.S. Patent No. 8,117,268, entitled "Hosted voice recognition system for wireless devices" and issued on February 14, 2012, the disclosure of which is hereby incorporated by reference in its entirety.

[0055] The speech recognition server 420 may be in communication either locally or remotely with a data store 430. The data store 430 may be embodied in hard disk drives, solid state memories, and/or any other type of non-transitory, computer-readable storage medium accessible to the speech recognition server 420. The data store 430 may also be distributed or partitioned across multiple storage devices as is known in the art without departing from the spirit and scope of the present disclosure. Further, in some embodiments, the data store 430 is implemented as a network-based electronic storage service.

[0056] The data store 430 may include one or more models of wakewords. In some embodiments, a wakeword model is specific to a user 401, while in other embodiments, the Upon receiving an audio input determined by the user computing device 200 to include a wakeword, the speech recognition server may compare the audio input to a known model of the wakeword stored in the data store 430. If the audio input is sufficiently similar to the known model, the speech recognition server 420 may transmit a confirmation of the wakeword to the user computing device 200, whereupon the user computing device 200 may obtain further audio input to be processed by the speech recognition server 420.

[0057] The data store 430 may also include one or more acoustic and/or language models for use in speech recognition. These models may include general-purpose models as well as specific models. Models may be specific to a user 401; to a speech recognition application implemented by the user computing device 200 and/or the speech recognition server 420; or may have other specific purposes. Further information regarding language models and acoustic models may be found in U.S. Patent App. No. 13/587,799, entitled "DISCRIMINATIVE LANGUAGE MODEL PRUNING," filed on August 16, 2012; and in U.S. Patent App. No. 13/592,157, entitled "UNSUPERVISED ACOUSTIC MODEL TRAINING," filed on August 22, 2012.

[0058] The data store 430 may further include data that is responsive to a query contained in audio input received by the speech recognition server 420. The speech recognition server 420 may recognize speech included in the audio input, identify a query included in the speech, and process the query to identify responsive data in the data store 430. The speech recognition server 420 may then provide an intelligent agent response including the responsive data to the user computing device 200 via the network 410. Still further data may be included in the data store 430.

[0059] It will be recognized that many of the devices described above are optional and that embodiments of the environment 400 may or may not combine devices. Furthermore, devices need not be distinct or discrete. Devices may also be reorganized in the environment 400. For example, the speech recognition server 420 may be represented as a single physical server computing device, or, alternatively, may be split into multiple physical servers that achieve the functionality described herein. Further, the user computing device 200 may have some or all of the speech recognition functionality of the speech recognition server 420.

[0060] Additionally, it should be noted that in some embodiments, the user computing device 200 and/or speech recognition server 420 may be executed by one more virtual machines implemented in a hosted computing environment. The hosted computing environment may include one or more rapidly provisioned and released computing resources, which computing resources may include computing, networking and/or storage devices. A hosted computing environment may also be referred to as a cloud computing environment. One or more of the computing devices of the hosted computing environment may include a power management subsystem 100 as discussed above.

[0061] With specific reference to FIG. 4A, an illustrative operation by which a wakeword may be confirmed is shown. The user 401 may speak the wakeword 502. The user computing device 200 may obtain the audio input that may include the user's speech (1) and determine that the wakeword 402 is present in the speech (2), substantially as discussed above with respect to FIG. 3. The audio input may also include a voice command or query. Responsive to determining that the speech includes the wakeword, the application processing module 112 and the network interface module 206 of the user computing device 200 may be activated (3) and the audio input transmitted (4) over the network 410 to the speech recognition server 420. The speech recognition server 420 may confirm (5) that the wakeword is present in the audio input, and may transmit (6) a confirmation to the user computing device 200 over the network 410.

[0062] Turning now to FIG. 4B, responsive to receiving the confirmation of the wakeword from the speech recognition server 420, the user computing device 200 may continue to obtain audio input (7) to be provided to the speech recognition server 420 for processing. For example, the obtained audio input may include an intelligent agent query 404 for processing by the speech recognition server 420. Alternately, the obtained audio input may in-

Petitioner Exhibit 1004, Page 221 of 1052

clude speech to be transcribed by the speech recognition server 420 (e.g., for use with a dictation, word processing, or messaging application executed by the application processing module 112). The user computing device 200 may transmit the audio input (8) over the network 410 to the speech recognition server 420. Optionally, an identifier of the speech recognition application for which speech recognition results are to be generated may be provided to the speech recognition server 420, so that the speech recognition server 420 may generate results specifically for use with the speech recognition application implemented by the application processing module 112. The speech recognition server 420 may recognize speech (9) included in the audio input and generate speech recognition results (10) therefrom. The speech recognition results may include, for example, a transcription of the speech, an intelligent agent response to a query included in the speech, or any other type of result. These speech recognition results may be transmitted (11) from the speech recognition server 420 to the user computing device 200 over the network 410. In response to receiving the results, the application processing module 112 may cause presentation of the results (12) in audible format (e.g., via text-to-speech) or in visual format (e.g., via a display of the user computing device 200).

[0063] With reference now to FIG. 4C, the user computing device 200 may continue to obtain audio input (13) to be provided to the speech recognition server 420 for processing. The user computing device 200 may transmit the audio input (14) over the network 410 to the speech recognition server 420. The speech recognition server may recognize any speech (15) included in the audio input. Responsive to recognizing the speech, the speech recognition server 420 may determine that the user 401 is no longer speaking to the user computing device 200 and stop (16) any subsequent speech recognition. For example, the user 401 may speak words that do not correspond to a structured command or query, such as undirected natural language speech 406. The speech recognition server 420 may also analyze the speech's speed, carefulness, inflection, or clarity to determine that the speech is not directed to the user computing device 200 and should not be processed into speech recognition results.

[0064] Other types of audio inputs may also prompt the speech recognition server 420 to stop subsequent speech recognition. Alternately, the speech recognition server 420 may determine that the received audio input does not include speech. Responsive to receiving one or more audio inputs that do not include speech directed to the user computing device 200, the speech recognition server 420 may determine that speech recognition results should not be generated and that the speech recognition should stop. Further, the audio input may include a predetermined sleepword, which may be selected by the user 401. If the speech recognition server 420 detects the sleepword, the speech recognition server 420 may stop performing speech recognition on the audio input. Further, the speech recognition server 420 may determine that multiple users 401 are present in the vicinity of the user computing device 200 (e.g., by performing speaker identification on multiple audio inputs obtained by the user computing device 200). If the number of identified users 401 satisfies a threshold (which may be any number of users 401 greater than one), the speech recognition server 420 may determine that any audio inputs obtained by the user computing device 200 are not likely intended to be processed into speech recognition results.

[0065] Responsive to determining that the speech of the user 401 is not directed to the user computing device 200 (or determining that subsequent speech recognition should not be performed for any of the other reasons discussed above), the speech recognition server 420 may transmit a deactivation instruction (17) over the network 410 to the user computing device 200. In response to receiving the deactivation instruction, the user computing device 200 may deactivate (18) its network interface module 206 and one or more components of the power management subsystem 100, such as the application processing module 112, the speech processing module 110, and/or the speech detection module 108. Other conditions may also prompt the speech recognition server 420 to transmit the deactivation instruction to the user computing device 200. For example, returning to FIG. 4A, if the speech recognition server 420 determines that a wakeword is not present in the audio input received at state (1), the speech recognition server 420 may transmit a deactivation instruction to the user computing device 200. Alternately, the speech recognition server 420 may determine that a threshold amount of time has passed since it last received an audio input including speech from the user computing device 200, and may accordingly transmit a deactivation instruction to the user computing device 200. Still other criteria may be determined for transmitting a deactivation instruction to the user computing device 200.

[0066] Returning again to FIG. 4A, upon receiving a subsequent audio input determined to include a wakeword, the user computing device 200 may activate the components of the power management subsystem 100 and the network interface module 206 and transmit the audio input to the speech recognition server 420. The example operations shown herein may thus repeat themselves.

[0067] The example operations depicted in FIG. 4A, FIG. 4B, and FIG. 4C are provided for illustrative purposes. One or more states may be omitted from the example operations shown herein, or additional states may be added. In a particular example, the user computing device 200 need not obtain a confirmation of the wakeword from the speech recognition server 420 before transmitting an audio input for which speech recognition results are to be generated by the speech recognition server 420. Additionally, the user computing device 200 need not obtain a deactivation instruction before deactivating its network interface module 206 and/or one or more of

Petitioner Exhibit 1004, Page 222 of 1052

the components of its power management subsystem 100, such as the application processing module 112, speech processing module 110, or speech detection module 108. Rather, power management subsystem 100 may determine (via the audio detection module 106) that at least a threshold amount of time has passed since an audio input having an energy level satisfying an energy level threshold has been obtained by the user computing device 200. Alternately, the user computing device 200 may determine (via the speech detection module 108) that at least a threshold amount of time has passed since an audio input that includes speech has been obtained. Responsive to determining that the threshold amount of time has passed, the power management subsystem 100 may cause deactivation of the network interface module 206, and may deactivate one or more of its own components as described above with respect to FIG. 3.

[0068] Further, the power management subsystem 100 may be configured to recognize a sleepword selected and spoken by the user 401, in a manner substantially similar to how the wakeword is identified in FIG. 3. If the sleepword is detected by the power management subsystem 100 (e.g., by the speech processing module 110), the network interface module 206 and/or one or more of the components of the power management subsystem 100 may be deactivated. Likewise, if the user computing device 200 includes its own on-device speech recognition capabilities, they may be deactivated responsive to the sleepword being detected.

[0069] FIG. 5 depicts an illustrative user interface 500 that may be provided by a user computing device 200 for customizing operations of the power management subsystem 100 and of the user computing device 200. In one embodiment, the user interface module 212 processes user input made via the user interface 500 and provides it to the power management subsystem 100.

[0070] The energy level threshold element 502 may enable a user to specify a threshold energy level at which the speech detection module 108 should be activated, as shown in block 308 of FIG. 3. For example, if the user computing device 200 is in a relatively noisy environment or if the user computing device 200 is experiencing a significant number of "false positives" determined by the audio detection module 106, the user 401 may wish to increase the energy level threshold at which the speech processing module 108 is activated. If the user 401 is in a relatively quiet environment or if the user computing device 200 is experiencing a significant number of false negatives, the user 401 may wish to decrease the energy level threshold at which the speech detection module 108 is activated. As discussed above, the energy level threshold may correspond to a volume threshold, intensity threshold, amplitude threshold, or other threshold related to the audio input.

[0071] The keyword confidence threshold element 504 may enable a user to specify a threshold score at which the speech processing module 110 determines that a keyword is present. Likewise, the identification confidence threshold element may enable a user to specify a threshold score at which the speech processing module 110 determines that the user spoke the keyword. In one embodiment, the application processing module 112 and the network interface module 206 are activated responsive to the speech processing module 110 recognizing a wakeword (e.g., the speech processing module 110 determining a score that satisfies a threshold, which score corresponds to a likelihood that the wakeword is included in the speech).. In another embodiment, the application processing module 112 and the network interface module 206 are activated responsive to the speech processing module 110 determining that the wakeword is associated with the user 401 with at least the threshold score corresponding to a likelihood that the wakeword is associated with the user. In a further embodiment, the application processing module 112 and the network interface module 206 are activated responsive to the speech processing module 110 both recognizing the wakeword with at least the threshold score and determining that the wakeword is associated with the user 401 with at least the threshold score. Other threshold elements may be provided to enable the user 401 to set individual thresholds for activating any or all of the individual components of the power management subsystem 100. Further threshold elements may be provided to enable the user to specify scores at which one or more blocks of the illustrative routine 300 may be skipped, substantially as discussed above with respect to FIG. 3.

[0072] The user interface 500 may further include one or more timer elements 508A and 508B. Each timer element may be used to set a threshold time interval at which the network interface module 206 and/or one or more components of the power management subsystem 100 are automatically deactivated. With reference to timer element 508A, if the power management subsystem 100 determines that at least a threshold interval of time has passed since an audio input having an energy level satisfying an energy level threshold has been obtained by the user computing device 200, the network interface module 206 may be automatically deactivated, in addition to the application processing module 112, the speech processing module 110, and the speech detection module 108 of the power management subsystem 100. Further timer elements may also be used to set a threshold time interval after which the speech recognition server 420 automatically sends a deactivation instruction to the network interface module 206 and the power management subsystem 100, substantially as discussed above with respect to FIG. 4C. Timer elements for other modules of the power management subsystem 100 may also be provided.

[0073] With continued reference to FIG. 5, the user 401 can select whether the wakeword should be confirmed by the speech recognition server 420 with server confirmation element 510. In some embodiments, the application processing module 112 and network interface module 206 only remains activated after the speech process-

Petitioner Exhibit 1004, Page 223 of 1052

ing module 110 detects the wakeword if a confirmation of the wakeword is received from the speech recognition server 420. If the user 401 requires server confirmation of the wakeword, subsequently obtained audio inputs may not be transmitted to the speech recognition server 420 unless the wakeword is confirmed. However, as discussed above, confirmation is not necessarily required. If the user 401 does not require server confirmation of the wakeword, the user computing device 200 may transmit one or more audio inputs obtained subsequent to the speech processing module 110 detecting the wakeword in the speech and/or determining that the speech is associated with the user 401.

[0074] The user 401 may also select whether speaker identification is required with speaker identification element 512. If the user 401 requires speaker identification, the speech processing module 110 and/or the speech recognition server 420 may be used to determine whether an audio input including speech corresponding to a wakeword is associated with the user 401. The application processing module 112 and network interface module 206 may be activated responsive to the speech processing module 110 determining that the user 401 is the speaker of the speech. Likewise, the network interface module 206 may remain in an activated state responsive to receiving a confirmation from the speech recognition server 420 that the user 401 is indeed the speaker of the wakeword. If the user 401 does not require speaker identification, however, neither the speech processing module 110 nor the speech recognition server 420 need identify the speaker.

[0075] The user interface 500 may also include an on-device recognition selection element 514, wherein the user 401 may select whether the user computing device 200 generates speech recognition results by itself, or whether audio inputs are routed to the speech recognition server 420 for processing into speech recognition results. The on-device recognition selection element 514 may be optionally disabled or grayed out if the user computing device 200 does not include on-device speech recognition capabilities. Further, the on-device recognition selection element 514 may be automatically deselected (and on-device speech recognition capabilities automatically disabled) if the power supply 218 drops below a threshold power supply level (e.g., a battery charge percentage), as on-device speech recognition capabilities as implemented by the processing unit 202 and/or the application processing module 112 may require a relatively large power draw.

[0076] The wakeword pane 516 and sleepword pane 518 may include user interface elements whereby the user 401 may record and cause playback of a wakeword or sleepword spoken by the user 401. When the user 401 records a wakeword or sleepword, the network interface module 206 may be automatically activated so that the audio input including the user's speech may be provided to the speech recognition server 420. The speech recognition server 420 may return a transcription of the recorded wakeword or sleepword so that the user may determine whether the recorded wakeword or sleepword was understood correctly by the speech recognition server 420. Alternately, when the user 401 records a wakeword or sleepword, any on-device speech recognition capabilities of the user computing device 200 may be activated to transcribe the recorded speech of the user 401. A spectral representation of the spoken wakeword or sleepword may also be provided by the user interface 500. Optionally, the wakeword pane 516 and sleepword pane 518 may include suggestions for wakewords or sleepwords, and may also indicate a quality of a wakeword or sleepword provided by the user 401, which quality may reflect a likelihood that the wakeword or sleepword is to produce false positive or false negative. Further information regarding suggesting keywords may be found in U.S. Patent App. No. 13/670,316, entitled "WAKE WORD EVALUATION," which was filed on November 6, 2012.

[0077] Various aspects of the present disclosure have been discussed as hardware implementations for illustrative purposes. However, as discussed above, the power management subsystem 100 may be partially or wholly implemented by the processing unit 202. For example, some or all of the functionality of the power management subsystem 100 may be implemented as software instructions executed by the processing unit 202. In a particular, non-limiting example, the functionality of the speech processing module 110, the application processing module 112, and the power management module 120 may be implemented as software executed by the processing unit 202. The processing unit 202 may accordingly be configured to selectively activate and/or deactivate the network interface module 206 responsive to detecting a wakeword. Still further implementations are possible.

[0078] Depending on the embodiment, certain acts, events, or functions of any of the routines or algorithms described herein can be performed in a different sequence, can be added, merged, or left out altogether (e.g., not all described operations or events are necessary for the practice of the algorithm). Moreover, in certain embodiments, operations or events can be performed concurrently, e.g., through multi-threaded processing, interrupt processing, or multiple processors or processor cores or on other parallel architectures, rather than sequentially.

[0079] The various illustrative logical blocks, modules, routines, and algorithm steps described in connection with the embodiments disclosed herein can be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. The described functionality can be implemented in varying

Petitioner Exhibit 1004, Page 224 of 1052

ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the disclosure.

[0080] Conditional language used herein, such as, among others, "can," "could," "might," "may," "e.g.," and the like, unless specifically stated otherwise, or otherwise understood within the context as used, is generally intended to convey that certain embodiments include, while other embodiments do not include, certain features, elements and/or steps. Thus, such conditional language is not generally intended to imply that features, elements and/or steps are in any way required for one or more embodiments or that one or more embodiments necessarily include logic for deciding, with or without author input or prompting, whether these features, elements and/or steps are included or are to be performed in any particular embodiment. The terms "comprising," "including," "having," and the like are synonymous and are used inclusively, in an open-ended fashion, and do not exclude additional elements, features, acts, operations, and so forth. Also, the term "or" is used in its inclusive sense (and not in its exclusive sense) so that when used, for example, to connect a list of elements, the term "or" means one, some, or all of the elements in the list.

[0081] Conjunctive language such as the phrase "at least one of X, Y and Z," unless specifically stated otherwise, is to be understood with the context as used in general to convey that an item, term, etc. may be either X, Y, or Z, or a combination thereof. Thus, such conjunctive language is not generally intended to imply that certain embodiments require at least one of X, at least one of Y and at least one of Z to each is present.

[0082] While the above detailed description has shown, described, and pointed out novel features as applied to various embodiments, it can be understood that various omissions, substitutions, and changes in the form and details of the devices or algorithms illustrated can be made without departing from the disclosure. The claimed subject matter is defined by the appended claims.

Claims

1. A computer-implemented method of operating a first computing device (202), the method comprising:

   receiving an audio input;
   determining one or more values from the audio input, wherein the one or more values comprise at least one of:

      a first value indicating an energy level of the audio input; or
      a second value indicating a likelihood that the audio input comprises speech;

   activating a first module (110, 112; 202) of the first computing device based at least in part on the one or more values;
   performing an operation, by the first module, wherein the operation comprises:

      determining that the audio input comprises a wakeword and causing activation of a network interface module (206) in response to determining that the audio input comprises a wakeword; and
      causing transmission of at least a portion of the audio input to a second computing device via the activated network interface module.

2. The computer-implemented method of Claim 1, wherein:

   the first module comprises a processor that is switchable between a low-power state and a high-power state; and
   the processor only performs the operation when it is in the high-power state.

3. The computer-implemented method of Claim 2, wherein activating the first module comprises switching the processor from the low-power state to the high-power state.

4. The computer-implemented method of Claim 2 or Claim 3, further comprising deactivating the first module, wherein deactivating the first module comprises switching the processor from the high-power state to the low-power state.

5. The computer-implemented method of any one of Claims 2 to 4, wherein the processor comprises at least one of a digital signal processor or a microprocessor.

6. The computer-implemented method of any one of Claims 1 to 5, wherein the operation further comprises receiving speech recognition results from a second computing device.

7. The computer-implemented method of Claim 6 further comprising:

   activating a second module of the first computing device based at least in part on the one or more values, wherein the second module is configured to implement a speech recognition application; and
   processing the speech recognition results with the speech recognition application.

8. A device (200) comprising:

Petitioner Exhibit 1004, Page 225 of 1052

a first processor (106, 108) configured to:

determine one or more values, wherein the one or more values comprise at least one of a first value indicating an energy level of an audio input or a second value indicating a likelihood that the audio input comprises speech; and
cause activation of a second processor based at least in part on the one or more values;

the second processor (110,112; 202) configured to perform an operation, wherein the operation comprises:

determining that the audio input comprises a wakeword and causing activation of a network interface module (206) in response to determining that the audio input comprises a wakeword; and
causing transmission of at least a portion of the audio input to a second device via the activated network interface module.

9. The device of Claim 8, wherein the first processor comprises at least one of a digital signal processor or a microprocessor.

10. The device of Claim 8 or Claim 9, further comprising a memory buffer module configured to store the audio input.

11. The device of any one of Claims 8 to 10, further comprising an audio input module in communication with the first processor, wherein the audio input module is configured to obtain the audio input.

12. The device of any one of Claims 8 to 11, wherein:

the first value comprises a volume of the audio input;
the second value comprises a first score; and

determining that the audio input comprises a wakeword comprises determining that the audio input comprises a wakeword based at least in part on determining a second score indicating a likelihood that the audio input comprises a wakeword.

13. The device of Claim 12, wherein the first processor comprises an audio detection module and a speech detection module, and wherein the second processor comprises a wakeword recognition module.

14. The device of Claim 12 or Claim 13, wherein:

the speech detection module is configured to determine the first score using at least one of a hidden Markov model, a Gaussian mixture model, energies in a plurality of spectral bands, or signal to noise ratios in a plurality of spectral bands; and
the wakeword recognition module is further configured to determine the second score using at least one of an application processing module, a hidden Markov model, and a Gaussian mixture model.

15. The device of any one of Claims 12 to 14, wherein the wakeword recognition module is configured:

to cause deactivation of the audio detection module based at least in part on the first score; and
to cause deactivation of the speech detection module based at least in part on the second score.

Patentansprüche

1. Computerimplementiertes Verfahren zum Betreiben eines ersten Datenverarbeitungsgerätes (202), wobei das Verfahren Folgendes aufweist:

Empfangen einer Audioeingabe;
Bestimmen eines oder mehrerer Werte aus der Audioeingabe, wobei der eine oder die mehreren Werte mindestens eines der Folgenden aufweisen:

einen ersten Wert, der ein Energieniveau der Audioeingabe angibt; oder
einen zweiten Wert, der eine Wahrscheinlichkeit, dass die Audioeingabe Sprache aufweist, angibt;

Aktivieren eines ersten Moduls (110, 112; 202) des ersten Datenverarbeitungsgerätes mindestens teilweise basierend auf dem einen oder den mehreren Werten;
Durchführen einer Operation, durch das erste Modul, wobei die Operation Folgendes aufweist:

Bestimmen, dass die Audioeingabe ein Weckwort aufweist, und Veranlassen der Aktivierung eines Netzwerkschnittstellenmoduls (206) als Reaktion auf das Bestimmen, dass die Audioeingabe ein Weckwort aufweist; und
Veranlassen der Übertragung von mindestens einem Abschnitt der Audioeingabe an ein zweites Datenverarbeitungsgerät über das aktivierte Netzwerkschnittstellenmodul.

Petitioner Exhibit 1004, Page 226 of 1052

2. Computerimplementiertes Verfahren nach Anspruch 1, wobei:

das erste Modul einen Prozessor aufweist, der zwischen einem Zustand geringer Leistung und einem Zustand hoher Leistung schaltbar ist; und der Prozessor die Operation nur durchführt, wenn er sich in dem Zustand hoher Leistung befindet.

3. Computerimplementiertes Verfahren nach Anspruch 2, wobei das Aktivieren des ersten Moduls das Schalten des Prozessors aus dem Zustand geringer Leistung in den Zustand hoher Leistung aufweist.

4. Computerimplementiertes Verfahren nach Anspruch 2 oder 3, welches ferner das Deaktivieren des ersten Moduls aufweist, wobei das Deaktivieren des ersten Moduls das Schalten des Prozessors aus dem Zustand hoher Leistung in den Zustand geringer Leistung aufweist.

5. Computerimplementiertes Verfahren nach einem der Ansprüche 2 bis 4, wobei der Prozessor mindestens entweder einen digitalen Signalprozessor oder einen Mikroprozessor aufweist.

6. Computerimplementiertes Verfahren nach einem der Ansprüche 1 bis 5, wobei die Operation ferner das Empfangen von Spracherkennungsergebnissen von einem zweiten Datenverarbeitungsgerät aufweist.

7. Computerimplementiertes Verfahren nach Anspruch 6, welches ferner Folgendes aufweist:

Aktivieren eines zweiten Moduls des ersten Datenverarbeitungsgerätes mindestens teilweise basierend auf dem einen oder den mehreren Werten, wobei das zweite Modul zum Implementieren einer Spracherkennungsanwendung konfiguriert ist; und Verarbeiten der Spracherkennungsergebnisse mit der Spracherkennungsanwendung.

8. Gerät (200), welches Folgendes aufweist:

einen ersten Prozessor (106, 108), der zu Folgendem konfiguriert ist:

Bestimmen eines oder mehrerer Werte, wobei der eine oder die mehreren Werte mindestens entweder einen ersten Wert, der ein Energieniveau einer Audioeingabe angibt, oder einen zweiten Wert, der eine Wahrscheinlichkeit, dass die Audioeingabe Sprache aufweist, angibt, aufweisen; und

Veranlassen der Aktivierung eines zweiten Prozessors mindestens teilweise basierend auf dem einen oder den mehreren Werten;

den zweiten Prozessor (110, 112; 202), der zum Durchführen einer Operation konfiguriert ist, wobei die Operation Folgendes aufweist:

Bestimmen, dass die Audioeingabe ein Weckwort aufweist, und Veranlassen der Aktivierung eines Netzwerkschnittstellenmoduls (206) als Reaktion auf das Bestimmen, dass die Audioeingabe ein Weckwort aufweist; und
Veranlassen der Übertragung von mindestens einem Abschnitt der Audioeingabe an ein zweites Gerät über das aktivierte Netzwerkschnittstellenmodul.

9. Gerät nach Anspruch 8, wobei der erste Prozessor mindestens entweder einen digitalen Signalprozessor oder einen Mikroprozessor aufweist.

10. Gerät nach Anspruch 8 oder 9, welches ferner ein Speicherpuffermodul aufweist, das zum Speichern der Audioeingabe konfiguriert ist.

11. Gerät nach einem der Ansprüche 8 bis 10, welches ferner ein Audioeingabemodul in Kommunikation mit dem ersten Prozessor aufweist, wobei das Audioeingabemodul zum Erhalten der Audioeingabe konfiguriert ist.

12. Gerät nach einem der Ansprüche 8 bis 11, wobei:

der erste Wert eine Lautstärke der Audioeingabe aufweist;
der zweite Wert eine erste Punktzahl aufweist; und
das Bestimmen, dass die Audioeingabe ein Weckwort aufweist, das Bestimmen, dass die Audioeingabe ein Weckwort aufweist, mindestens teilweise basierend auf dem Bestimmen einer zweiten Punktzahl, die eine Wahrscheinlichkeit, dass die Audioeingabe ein Weckwort aufweist, angibt, aufweist.

13. Gerät nach Anspruch 12, wobei der erste Prozessor ein Audioerkennungsmodul und ein Spracherkennungsmodul aufweist, und wobei der zweite Prozessor ein Weckworterkennungsmodul aufweist.

14. Gerät nach Anspruch 12 oder 13, wobei:

das Spracherkennungsmodul zum Bestimmen der ersten Punktzahl unter Verwendung von mindestens entweder einem Hidden-Markov-Modell, einem Gaußschen Mischverteilungs-

Petitioner Exhibit 1004, Page 227 of 1052

modell, Energien in mehreren Spektralbändern oder Signal-Rausch-Verhältnissen in mehreren Spektralbändern konfiguriert ist; und
das Weckworterkennungsmodul ferner zum Bestimmen der zweiten Punktzahl unter Verwendung von mindestens entweder einem Anwendungsverarbeitungsmodul, einem Hidden-Markov-Modell oder einem Gaußschen Mischverteilungsmodell konfiguriert ist.

15. Gerät nach einem der Ansprüche 12 bis 14, wobei das Weckworterkennungsmodul zu Folgendem konfiguriert ist:

Veranlassen der Deaktivierung des Audioerkennungsmoduls mindestens teilweise basierend auf der ersten Punktzahl; und
Veranlassen der Deaktivierung des Spracherkennungsmoduls mindestens teilweise basierend auf der zweiten Punktzahl.

Revendications

1. Procédé mis en oeuvre par ordinateur pour le fonctionnement d'un premier dispositif informatique (202), le procédé comprenant les étapes suivantes :

recevoir une entrée audio ;
déterminer une ou plusieurs valeurs provenant de l'entrée audio, la ou les valeurs comprenant au moins un élément parmi :

une première valeur indiquant un niveau d'énergie de l'entrée audio ; ou
une deuxième valeur indiquant une probabilité que l'entrée audio comprenne de la parole ;

activer un premier module (110, 112 ; 202) du premier dispositif informatique au moins en partie en fonction de la ou des valeurs ;
effectuer une opération, par le premier module, l'opération comprenant les étapes suivantes :

déterminer que l'entrée audio comprend un mot d'activation et provoquer l'activation d'un module d'interface réseau (206) en réponse à la détermination que l'entrée audio comprend un mot d'activation ; et
provoquer la transmission d'au moins une partie de l'entrée audio à un deuxième dispositif informatique par le biais du module d'interface réseau activé.

2. Procédé mis en oeuvre par ordinateur de la revendication 1, dans lequel :

le premier module comprend un processeur qui est commutable entre un état à basse puissance et un état à haute puissance ; et
le processeur effectue l'opération uniquement quand il se trouve dans l'état à haute puissance.

3. Procédé mis en oeuvre par ordinateur de la revendication 2, dans lequel l'activation du premier module comprend la commutation du processeur de l'état à basse puissance à l'état à haute puissance.

4. Procédé mis en oeuvre par ordinateur de la revendication 2 ou la revendication 3, comprenant en outre la désactivation du premier module, la désactivation du premier module comprenant la commutation du processeur de l'étape à haute puissance à l'état à basse puissance.

5. Procédé mis en oeuvre par ordinateur de l'une quelconque des revendications 2 à 4, dans lequel le processeur comprend un processeur de signal numérique et/ou un microprocesseur.

6. Procédé mis en oeuvre par ordinateur de l'une quelconque des revendications 1 à 5, dans lequel l'opération comprend en outre la réception de résultats de reconnaissance de la parole provenant d'un deuxième dispositif informatique.

7. Procédé mis en oeuvre par ordinateur de la revendication 6 comprenant en outre :

l'activation d'un deuxième module du premier dispositif informatique au moins en partie en fonction de la ou des valeurs, le deuxième module étant configuré pour mettre en oeuvre une application de reconnaissance de la parole ; et
le traitement des résultats de reconnaissance de la parole avec l'application de reconnaissance de la

8. Dispositif (200) comprenant :

un premier processeur (106, 108) configuré pour :

déterminer une ou plusieurs valeurs, la ou les valeurs comprenant une première valeur indiquant un niveau d'énergie d'une entrée audio et/ou une deuxième valeur indiquant une probabilité que l'entrée audio comprenne de la parole ; et
provoquer l'activation d'un deuxième processeur au moins en partie en fonction de la ou des valeurs ;

le deuxième processeur (110, 112 ; 202) configuré pour effectuer une opération, l'opération

Petitioner Exhibit 1004, Page 228 of 1052

comprenant les étapes suivantes :

déterminer que l'entrée audio comprend un mot d'activation et provoquer l'activation d'un module d'interface réseau (206) en réponse à la détermination que l'entrée audio comprend un mot d'activation ; et
provoquer la transmission d'au moins une partie de l'entrée audio à un deuxième dispositif par le biais du module d'interface réseau activé.

9. Dispositif de la revendication 8, dans lequel le premier processeur comprend un processeur de signal numérique et/ou un microprocesseur.

10. Dispositif de la revendication 8 ou la revendication 9, comprenant en outre un module de mémoire tampon configuré pour stocker l'entrée audio.

11. Dispositif de l'une quelconque des revendications 8 à 10, comprenant en outre un module d'entrée audio en communication avec le premier processeur, le module d'entrée audio étant configuré pour obtenir l'entrée audio.

12. Dispositif de l'une quelconque des revendications 8 à 11, dans lequel :

la première valeur comprend un volume de l'entrée audio ;
la deuxième valeur comprend un premier score ; et

la détermination que l'entrée audio comprend un mot d'activation comprend la détermination que l'entrée audio comprend un mot d'activation au moins en partie en fonction de la détermination d'un deuxième score indiquant une probabilité que l'entrée audio comprenne un mot d'activation.

13. Dispositif de la revendication 12, dans lequel le premier processeur comprend un module de détection audio et un module de détection de la parole, et dans lequel le deuxième processeur comprend un module de reconnaissance de mot d'activation.

14. Dispositif de la revendication 12 ou la revendication 13, dans lequel :

le module de détection de la parole est configuré pour déterminer le premier score en utilisant au moins un élément parmi un modèle de Markov caché, un modèle de mélange gaussien, des énergies dans une pluralité de bandes spectrales, ou des rapports signal/bruit dans une pluralité de bandes spectrales ; et
le module de reconnaissance de mot d'activation est en outre configuré pour déterminer le deuxième score en utilisant au moins un élément parmi un module de traitement d'application, un modèle de Markov caché, et un modèle de mélange gaussien.

15. Dispositif de l'une quelconque des revendications 12 à 14, dans lequel le module de reconnaissance de mot d'activation est configuré :

pour provoquer la désactivation du module de détection audio au moins en partie en fonction du premier score ; et
pour provoquer la désactivation du module de détection de la parole au moins en partie en fonction du deuxième score.

Petitioner Exhibit 1004, Page 229 of 1052

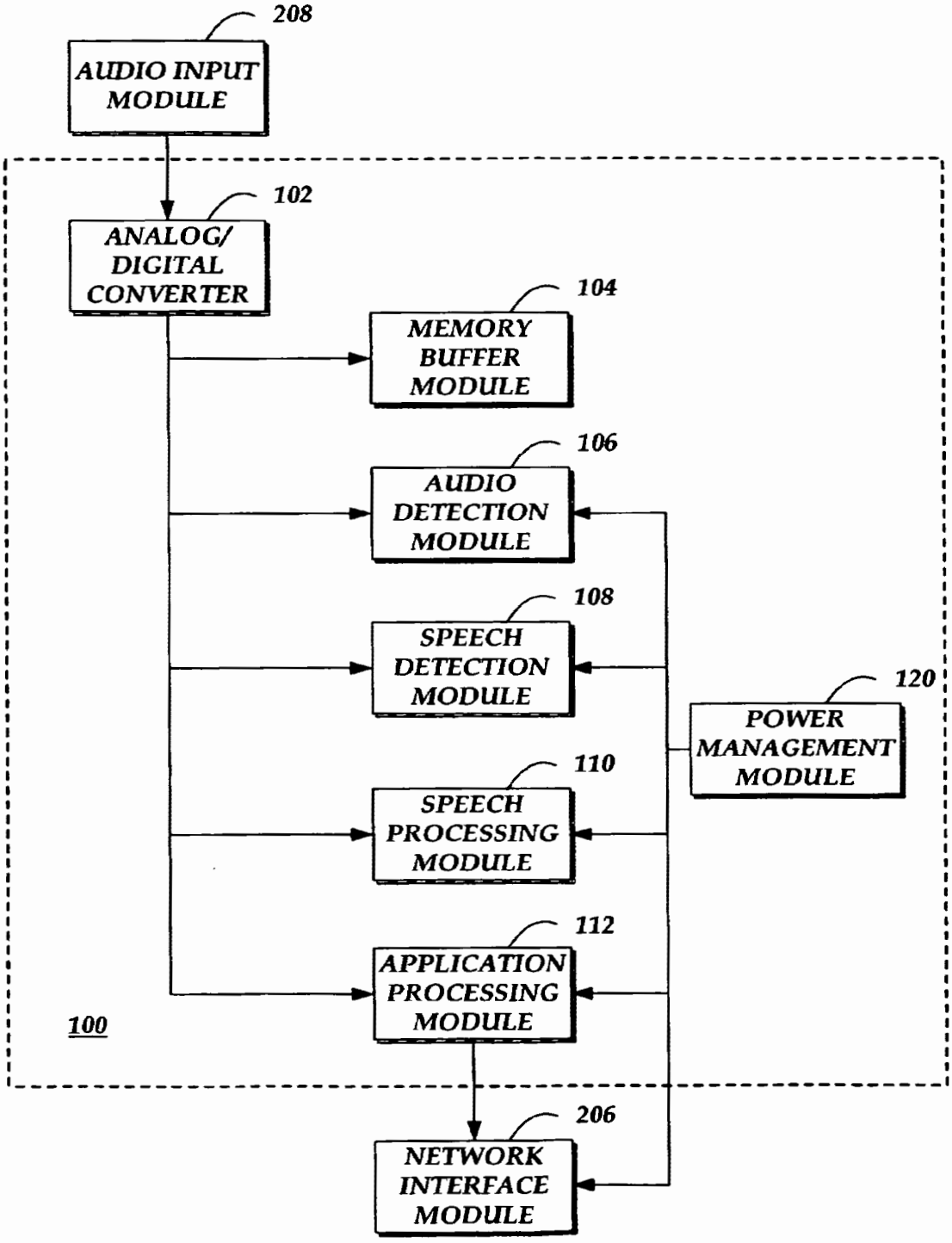


Fig. 1.

Petitioner Exhibit 1004, Page 230 of 1052

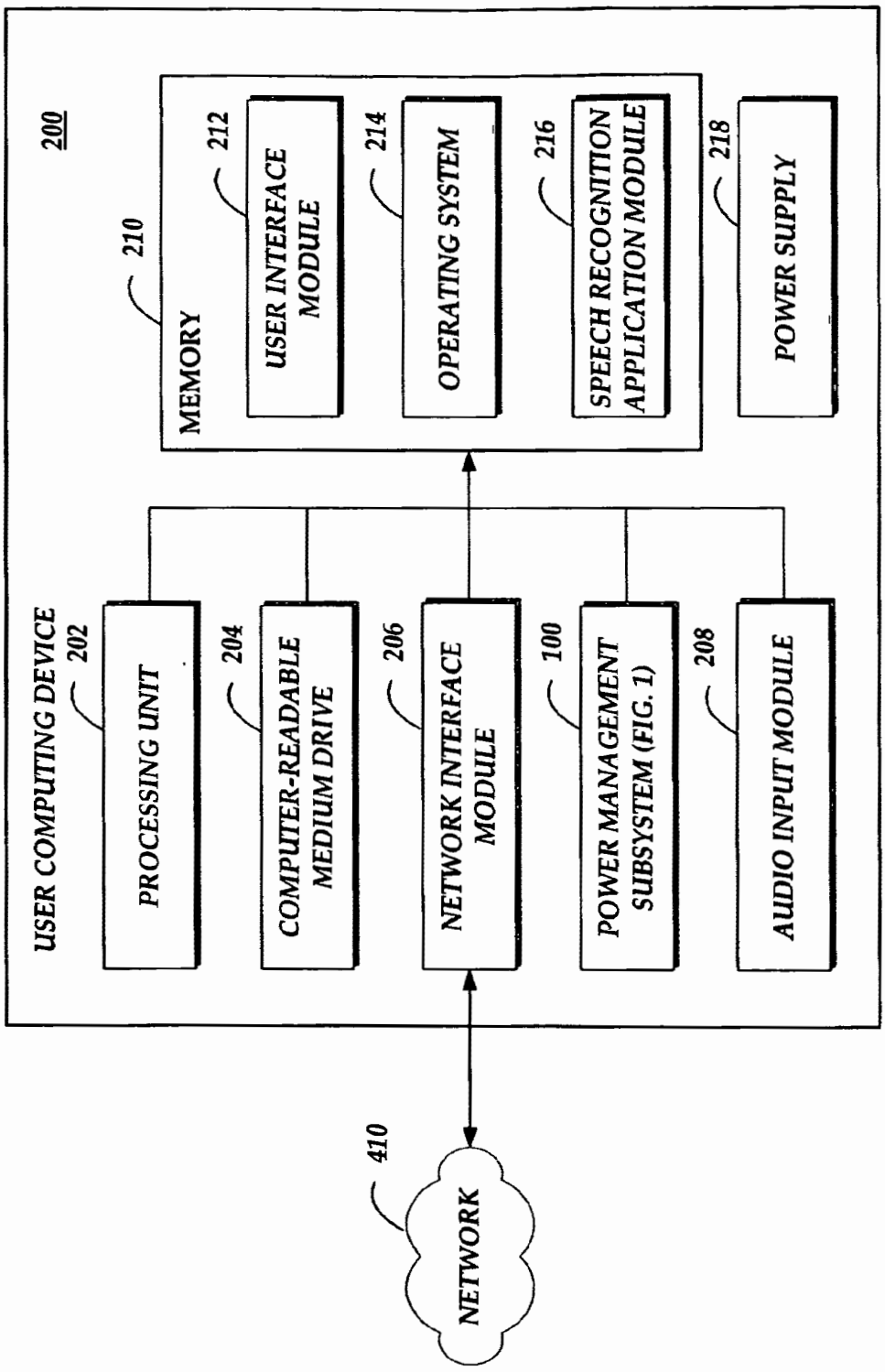


Fig. 2.

Petitioner Exhibit 1004, Page 231 of 1052

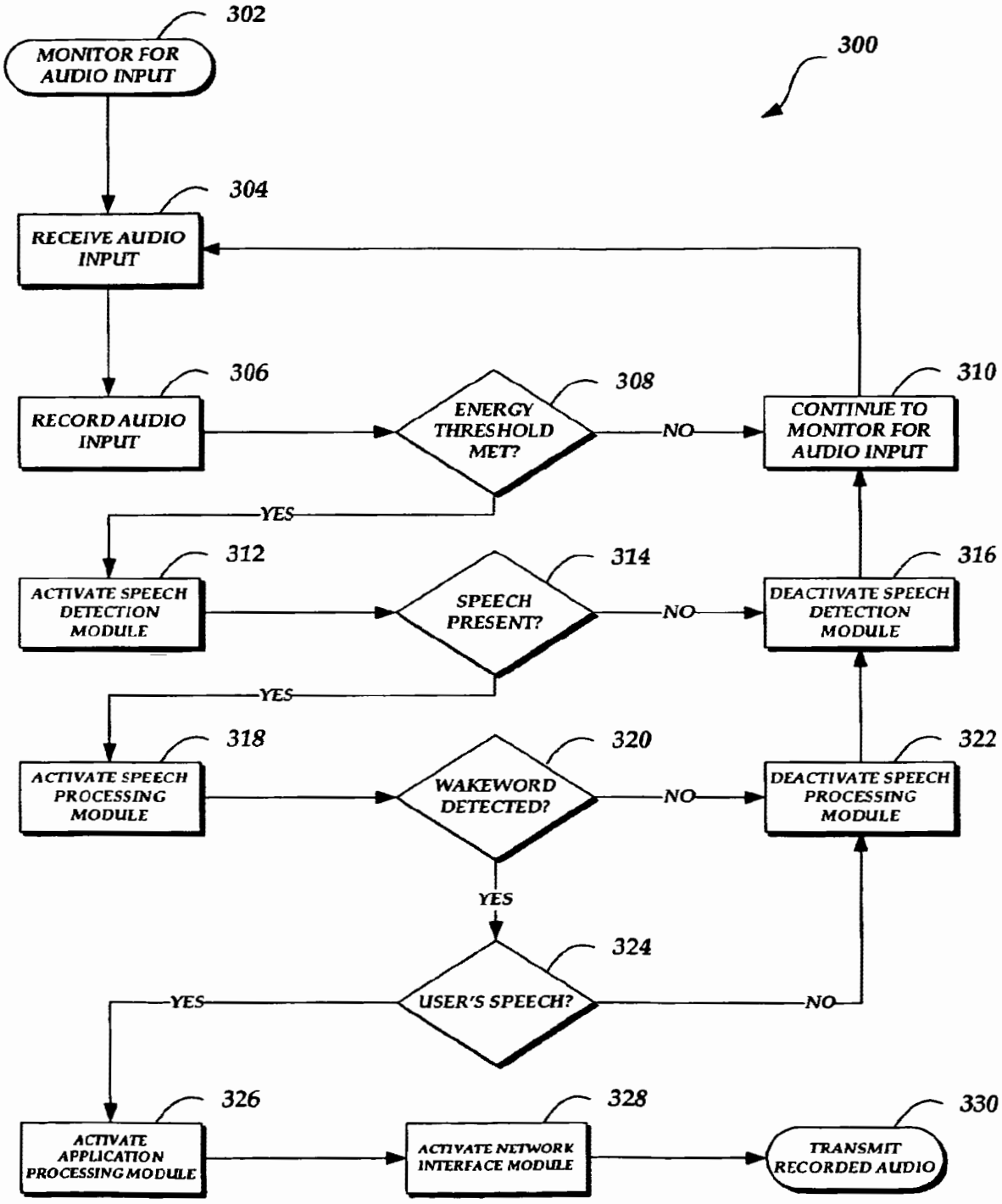


Fig.3.

Petitioner Exhibit 1004, Page 232 of 1052

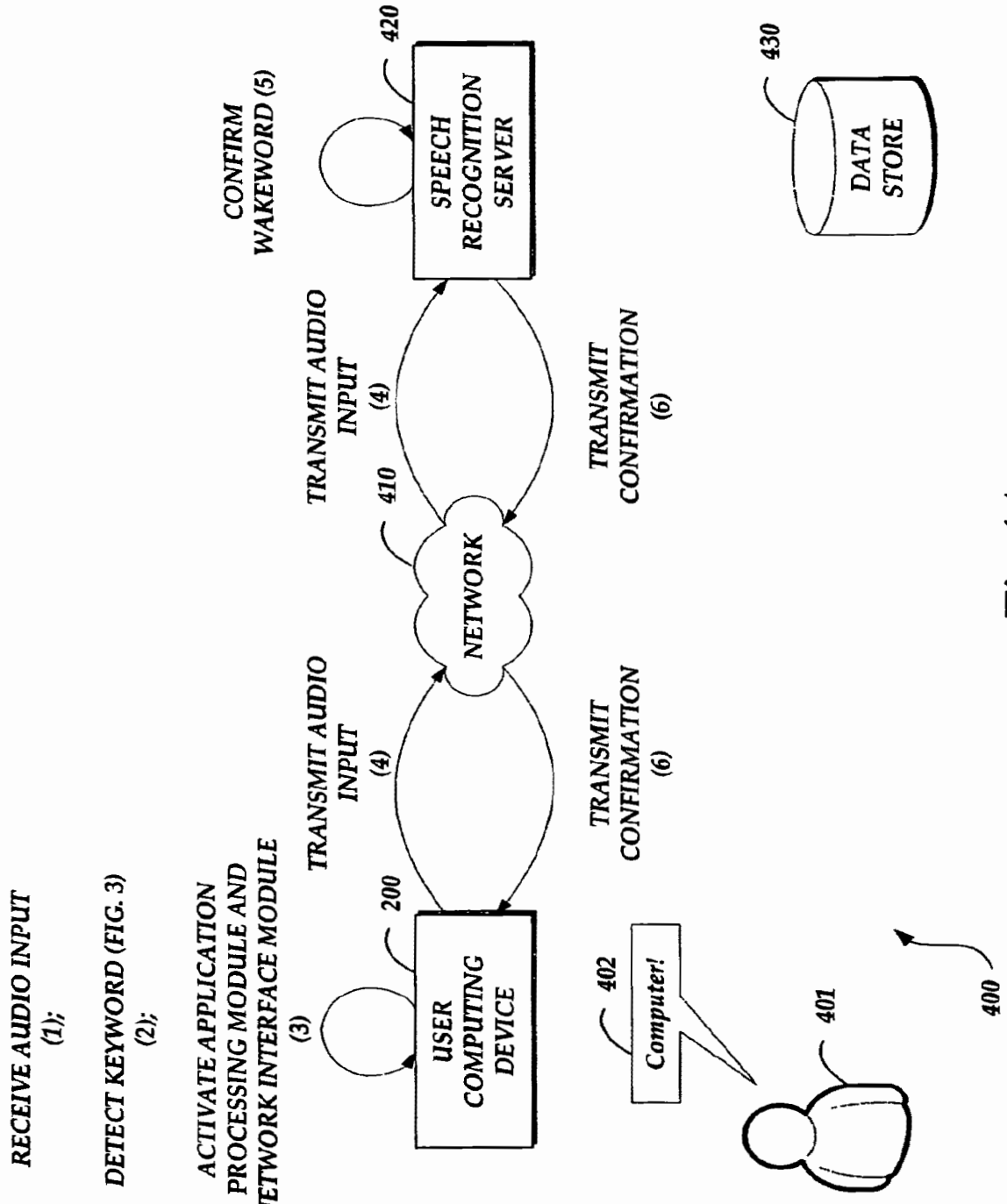


Fig.4A.

Petitioner Exhibit 1004, Page 233 of 1052

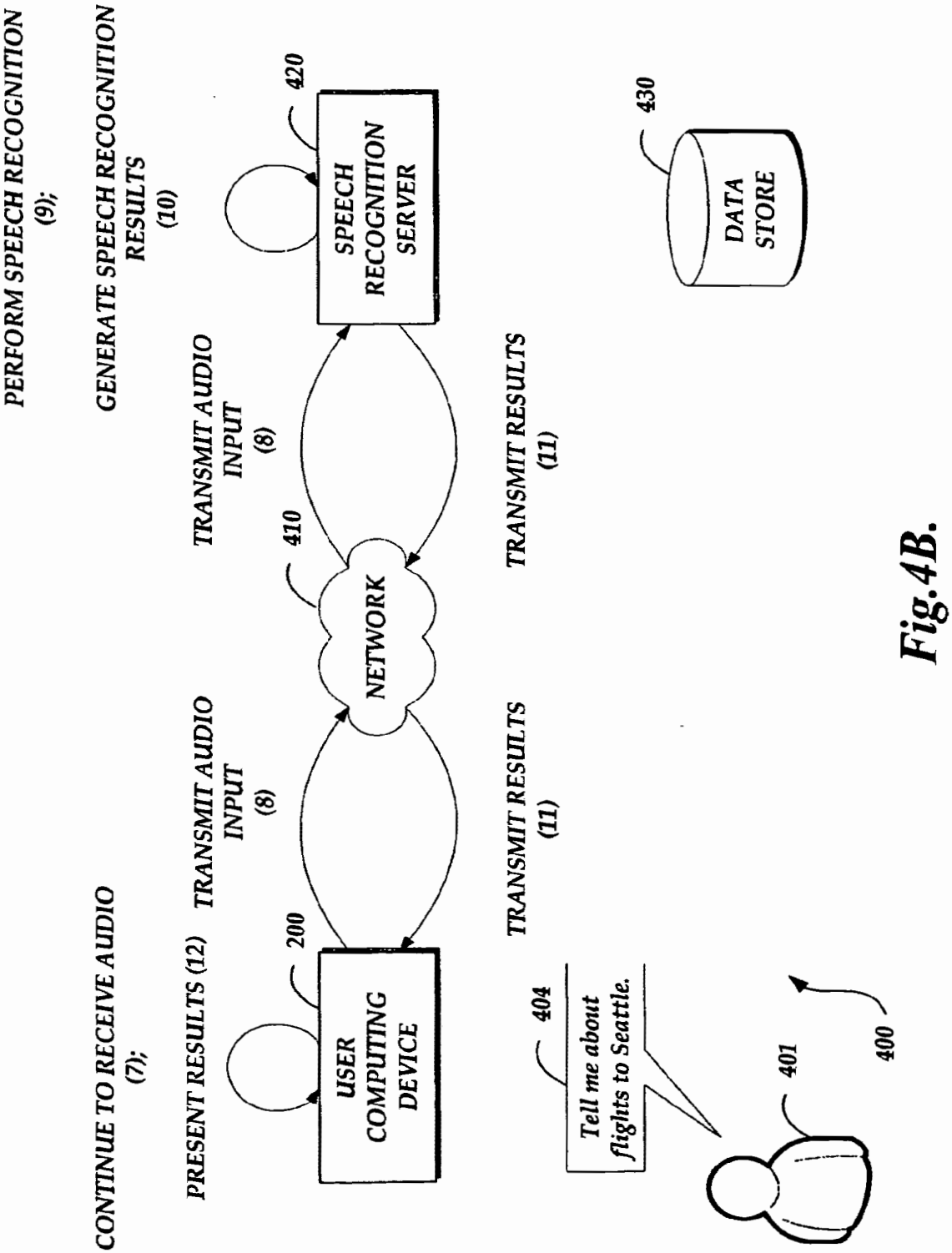


Fig.4B.

Petitioner Exhibit 1004, Page 234 of 1052

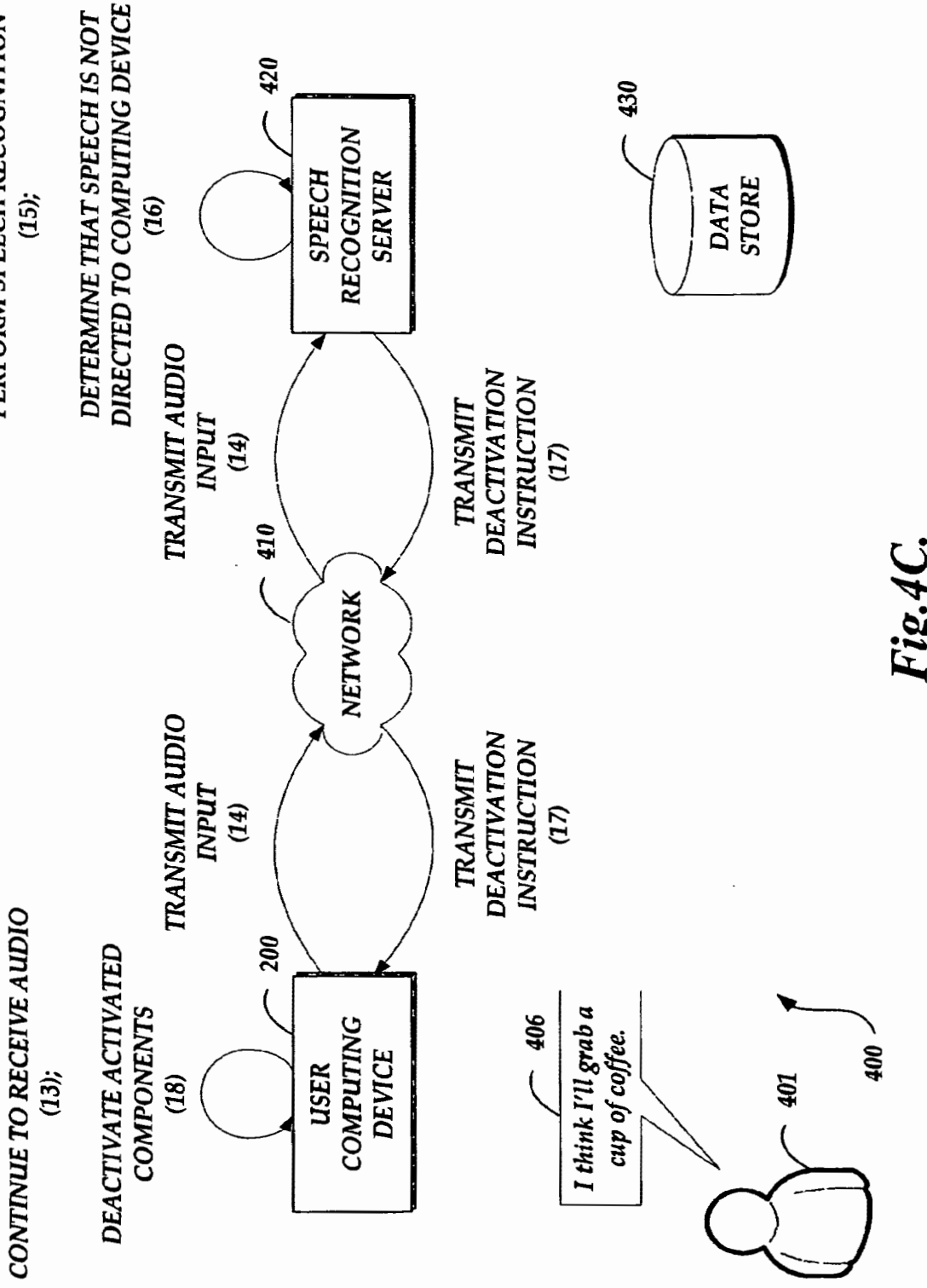


Fig.4C.

Petitioner Exhibit 1004, Page 235 of 1052

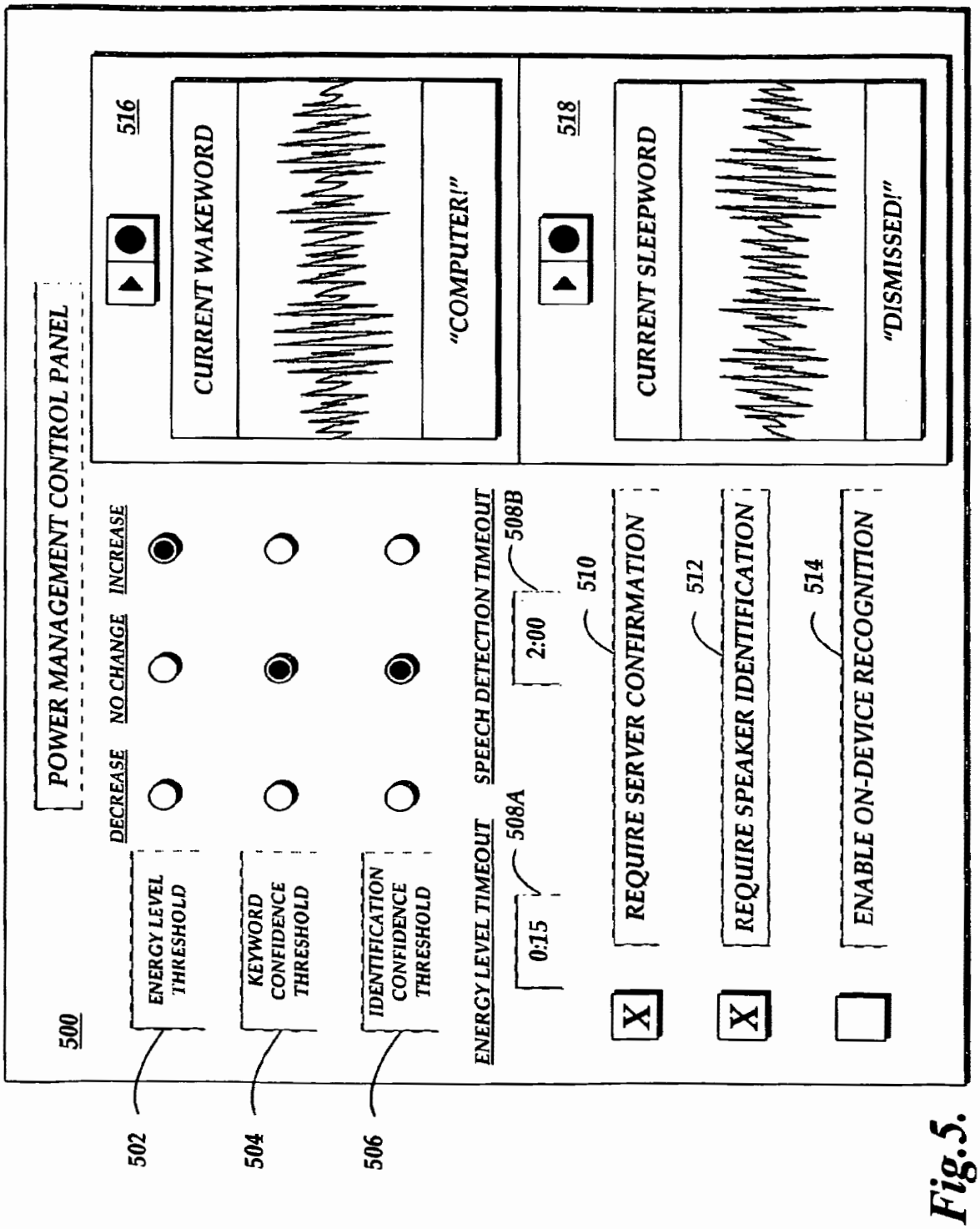


Fig.5.

Petitioner Exhibit 1004, Page 236 of 1052

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 2003216909 A **[0001]**
- US 20120016678 A **[0009]**
- US 58779912 A **[0032] [0057]**
- US 59215712 A **[0032] [0057]**
- US 05873002 A **[0037]**
- US 20030144844 A **[0037]**
- US 8326636 B **[0037]**
- US 8117268 B **[0054]**
- US 67031612 A **[0076]**

Petitioner Exhibit 1004, Page 237 of 1052

Bibliographic Data

Application No: 15238238

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Huy C. Ho/ (Examiner's Signature) | HH (Initials)

Title: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
08/16/2016	455	2644	N0484.71090US04
RULE			

APPLICANTS

Nuance Communications, Inc., Burlington, MA,

INVENTORS

William F. Ganong, Brookline, MA, UNITED STATES

Paul Adrian Van Mulbregt, Wayland, MA, UNITED STATES

Vladimir Sejnoha, Lexington, MA, UNITED STATES

Glen Edward Wilson, Maynard, MA, UNITED STATES

CONTINUING DATA

15238238 is a CON of 13795933 03/12/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

08/25/2016

STATE OR COUNTRY

UNITED STATES

ADDRESS

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

FILING FEE RECEIVED

$1,600

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 01/18/2018

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
01/18/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@wolfgreenfield.com
N0484_eOfficeAction@WolfGreenfield.com
IP.Inbox@nuance.com

Office Action Summary	Application No. 15/238,238	Applicant(s) GANONG ET AL.	
	Examiner HUY C. HO	Art Unit 2644	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *10/02/2017*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-20* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-20* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *08/16/2016* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Response to Arguments

2. Applicant's arguments filed 10/02/2017 have been fully considered but they are moot because the claims have been amended with new limitations, and the argued features are not persuasive for the following reasons.

With respect to the arguments regarding to the new amended claim features for the independent claims 1 and 15 that neither Freeman nor Mozer disclose or suggest the recited limitations, i.e., "perform, while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input."

The examiner respectfully disagrees because Mozer teaches "perform, while in the low power mode, at least one processing

stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input" (*Mozer, Paragraph(s) [33], [35]-[36]: stage 1 recognizer at low power operation mode detects acoustic input signals using triggered keywords; Figures 1B, 9, Paragraph(s) [31], [38], [56]: stage 1 recognizer activate or transmit or communicate with stage 2 recognizer as a server to further process the triggered keywords in the acoustic input signals).*

Therefore, Mozer teaches and discloses the argued features, i.e., "perform, while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input".

Double Patenting

3. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on nonstatutory double patenting provided the reference application or patent either is shown to be commonly owned with the examined application, or claims an

invention made as a result of activities undertaken within the scope of a joint research agreement. See MPEP § 717.02 for applications subject to examination under the first inventor to file provisions of the AIA as explained in MPEP § 2159. See MPEP §§ 706.02(l) (1) - 706.02(l)(3) for applications not subject to examination under the first inventor to file provisions of the AIA. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

The USPTO Internet website contains terminal disclaimer forms which may be used. Please visit www.uspto.gov/patent/patents-forms. The filing date of the application in which the form is filed determines what form (e.g., PTO/SB/25, PTO/SB/26, PTO/AIA/25, or PTO/AIA/26) should be used. A web-based eTerminal Disclaimer may be filled out completely online using web-screens. An eTerminal Disclaimer that meets all requirements is auto-processed and approved immediately upon submission. For more information about eTerminal Disclaimers, refer to www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

4. Claim 1 is rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1 and 5 of U.S. Patent No. 9,361,885. Although the claims at issue are not identical, they are not patentably distinct from each other

because claims 1 and 5 of the U.S. Patent No. 9,361,885 reflects the recited limitations of claim 1 of the current application as in the following comparison.

Claim 1 (The Patent No. 9,361,885), A method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is operating in an idle mode, the mobile device having a first processor and a second processor, the method comprising:

receiving acoustic input while the mobile device is operating in the idle mode;

performing at least one first processing stage on the acoustic input using the first processor, prior to engaging the second processor to process the acoustic input, to evaluate whether the acoustic input includes a voice command;

performing at least one second processing stage on the acoustic input using the second processor to evaluate whether the acoustic input includes a voice command if further processing is needed to determine whether the acoustic input includes a voice command;

wherein performing the at least one first processing stage or the at least one second processing stage includes, while the mobile device is operating in the idle mode, transmitting at least a portion of the acoustic input to at least one server via

a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command; and

initiating responding to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

Claim 5 (The Patent No. 9,361,885), The method of claim 1, wherein performing the at least one first processing stage comprises performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

Claim 1 (The Current Application), A mobile device comprising:

at least one input configured to receive acoustic input from an environment of the mobile device while the mobile device is operating in a low power mode; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input to evaluate whether the acoustic input includes a voice command,

the at least one processing stage including transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server at least to evaluate whether the acoustic input includes a voice command,

wherein the mobile device is configured to initiate responding to the voice command when the at least one processing stage determines that the acoustic input includes a voice command.

Correction is required.

Claim Rejections - 35 USC § 103

5. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

6. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

7. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

8. Claims 1-12 and 15-19 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in view of Mozer (U.S. Patent Publication No. 20103/0080167).

Regarding claim 1, (Currently Amended) Freeman teaches a device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512),*

wherein the device is configured to respond to the voice command (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520).*

Freeman does not teach "detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input" and "voice commands recognized by the at least one server".

Mozer teaches detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the

acoustic input" and "voice commands recognized by the at least one server" (*Mozer, Paragraph(s) [33], [35]-[36]: stage 1 recognizer at low power operation mode detects acoustic input signals using triggered keywords; Figures 1B, 9, Paragraph(s) [31], [38], [56]: stage 1 recognizer activate or transmit or communicate with stage 2 recognizer as a server to further process the triggered keywords in the acoustic input signals).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 15, (Currently Amended) Freeman teaches at least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode

(*Freeman, the Abstract, Figure 1, Col. 1 Lines 60-65*), the method comprising:

performing, while in the low power mode, at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*; and

responding to voice commands (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520)*.

Freeman does not teach "detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input" and "voice commands recognized by the at least one server".

Mozer teaches detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the

acoustic input" and "voice commands recognized by the at least one server" (*Mozer, Paragraph(s) [33], [35]-[36]: stage 1 recognizer at low power operation mode detects acoustic input signals using triggered keywords; Figures 1B, 9, Paragraph(s) [31], [38], [56]: stage 1 recognizer activate or transmit or communicate with stage 2 recognizer as a server to further process the triggered keywords in the acoustic input signals).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 2, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage

(*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 3, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 4, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor (*Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 5, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice

activity and/or phone loop analysis to evaluate whether the acoustic input includes speech (*Freeman, Col. 4 Lines 45-62, Col. 5 Lines 13-33).*

Regarding claim 6, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger (*Mozer, Paragraph(s) [33]).*

Regarding claim 7, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command (*Mozer, Paragraph(s) [36]).*

Regarding claim 8, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the

acoustic input includes an explicit voice trigger (*Mozer, Paragraph(s) [33], [36]*).

Regarding claim 9, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Mozer, Paragraph(s) [26]-[28], [46]-[47]*).

Regarding claim 10, (Currently Amended) Freeman, as modified by Mozer, teaches the mobile device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 5, boxes 512*).

Regarding claim 11, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the idle mode (*Freeman, Figure 5, boxes 506, 508, 510*).

Regarding claim 12, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input (*Freeman, Figure 5, box 506)*.

Regarding claim 16, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Mozer, Figure 1B, Paragraph(s) [31]-[33], [36])*.

Regarding claim 17, (Currently Amended) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger (*Mozer, Paragraph(s) [33])*.

Regarding claim 18, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the

low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Freeman, Figure 5, boxes 504, 506, 508)*.

Regarding claim 19, (Currently Amended) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the idle mode (*Freeman, Figure 5, boxes 506, 508, 510)*.

9. Claims 13, 14 and 20 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in views of Mozer (U.S. Patent Publication No. 20103/0080167) and LeBeau et al. (U.S. Patent Publication No. 2012/0034904).

Regarding claim 13, (Currently Amended) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the mobile device, an occurrence of an event on the device,

recent activity on the mobile device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 14, (Currently Amended) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43])*.

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source,

etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 20, (Currently Amended) Freeman, as modified by Mozer, does not teach the at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43])*.

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUY C. HO whose telephone number is (571)270-1108. The examiner can normally be reached on Monday - Friday, 8:00 a.m. - 5:00 p.m., EST.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based

collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Kathy Wang-Hurst can be reached on 571-270-5371. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HUY C HO/
Primary Examiner, Art Unit 2644

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15238238	GANONG ET AL.
	Examiner	Art Unit
	HUY C HO	2644

CPC- SEARCHED		
Symbol	Date	Examiner
G10L15/20; G10L15/22; G10L15/24; G10L15/25; G10L2015/223; G10L2015/226; H04M2250/74; H04W52/0251; H04W52/0254.	4/20/2017	HH

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner
379	88.01	4/20/2017	HH

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

SEARCH NOTES		
Search Notes	Date	Examiner
EAST Database Search	4/20/2017	HH
Inventor Search	4/20/2017	HH
Search update	1/10/2018	HH

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner
379	88.02	4/20/2017	HH
704	246-255	4/20/2017	HH

	/HUY C HO/ Primary Examiner.Art Unit 2644

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15238238	GANONG ET AL.
	Examiner	Art Unit
	HUY C HO	2644

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	04/20/2017	01/10/2018							
	1	✓	✓							
	2	✓	✓							
	3	✓	✓							
	4	✓	✓							
	5	✓	✓							
	6	✓	✓							
	7	✓	✓							
	8	✓	✓							
	9	✓	✓							
	10	✓	✓							
	11	✓	✓							
	12	✓	✓							
	13	✓	✓							
	14	✓	✓							
	15	✓	✓							
	16	✓	✓							
	17	✓	✓							
	18	✓	✓							
	19	✓	✓							
	20	✓	✓							

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: Huy C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		9,646,610	B2	Macho	05-09-2017
		2010-0121636	A1	Burke et al.	05-13-2010
		2012-0034904	A1	LeBeau et al.	02-09-2012
		2012-0035924	A1	Jitkoff et al.	02-09-2012
		2012-0127072	A1	Kim	05-24-2012

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER: /HUY C HO/	DATE CONSIDERED: 01/10/2018

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

5463597.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	0	EP2932500	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/01/10 15:59
L2	31	"2932500"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/01/10 15:59
L3	3	"20140163978"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/01/10 16:03
L4	4	"9646610"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/01/10 16:04
L16	968	("20010047264" | "20020002461" | "20020010584" | "20020031262" | "20020032564" | "20020035474" | "20020042707" | "20020046025" | "20020059066" | "20020059068" | "20020067308" | "20020069063" | "20020077817" | "20020091511" | "20020099552" | "20020103641" | "20020116171" | "20020116185" | "20020116189" | "20020128827" | "20020133347" | "20020135565" | "20020138265" | "20020143551" | "20020154160" | "20020164000" | "20020184189" | "20020198714" | "20030078766" | "20030088414" | "20030097210" | "20030099335" | "20030120494" | "20030135740" | "20030167335" | "20030190074" | "20030212961" | "20030233230" | "20030234824" | "20040054535" | "20040054690" | "20040085162" | "20040106432" | "20040135701" | "20040145607" | "20040162741" | "20040176958" | "20040186714" | "20040193420" | "20040199387" | "20040220798" | "20040236778" | "20050002507" | "20050015772" | "20050027385" | "20050033582" | "20050045373" | "20050049880" | "20050055403" | "20050058438" | "20050071332" | "20050080625" | "20050086059" | "20050091118" | "20050100214" | "20050102614" | "20050108001" | "20050108074" | "20050114124" | "20050114140" | "20050119897" | "20050125235" | "20050132301" | "20050143972" | "20050152602" | "20050165607" | "20050182616" | "20050182628" | "20050182629" | "20050192801" | "20050196733" | "20050203747" | "20050203991" | "20050228665" | "20050273626" | "20050288936" | "20050289463" | "20060009973" | "20060018492" | "20060061488" | "20060095848" | "20060106592" | "20060106594").PN. OR ("20060106595" | "20060111906" | "20060117002" | "20060122834" | "20060143007" | "20060156252" | "20060193518" | "20060217967" | "20060235700" | "20060252457" | "20060293886" | "20070006098" | "20070021956" | "20070027732" | "20070033003" | "20070038436" | "20070041361" | "20070050191" | "20070050712" | "20070055514" | "20070055525" | "20070055529" | "20070058832" | "20070073540" | "20070088556" | "20070094026" | "20070100790" | "20070106674" | "20070118377" | "20070118378" | "20070124149" | "20070135949" | "20070174188" | "20070185754" | "20070185917" | "20070198269" | "20070208569" | "20070208579" | "20070208726" | "20070211071" | "20070225980" | "20070276651" | "20070276714" | "20070276810" | "20070282595" | "20080012950" | "20080015864" | "20080021708" | "20080034032" | "20080052063" | "20080052073" | "20080056579" | "20080071544" | "20080077384" | "20080079566" | "20080082332" | "20080082338" | "20080082390" | "20080082651" | "20080091406" | "20080118143" | "20080120112" | "20080120342" | "20080126100" | "20080129520" | "20080131006" | "20080140413" | "20080140416" | "20080140652" | "20080140657" | "20080189114" | "20080221880" | "20080221889" | "20080221903" | "20080228463" | "20080228490" | "20080228496" | "20080240569" | "20080247519" | "20080249770" | "20080256613" | "20080270118" | "20080281510" | "20080300878" | "20080313335" | "20080319763" | "20090006100" | "20090006343" | "20090011709" | "20090018835" | "20090028435" | "20090030800" | "20090055179").PN. OR ("20090058823" |	US-PGPUB; USPAT; USOCR	OR	OFF	2018/01/10 16:43

"20090070097" | "20090076792" | "20090076796" | "20090077165" | "20090094033" | "20090100049" | "20090112572" | "20090112677" | "20090123071" | "20090125477" | "20090144049" | "20090150156" | "20090157401" | "20090164441" | "20090171664" | "20090187577" | "20090191895" | "20090204409" | "20090222488" | "20090239552" | "20090253463" | "20090271109" | "20090271175" | "20090271178" | "20090287583" | "20090290718" | "20090299745" | "20090299849" | "20090300391" | "20090306980" | "20090307162" | "20100005081" | "20100019834" | "20100023320" | "20100036660" | "20100042400" | "20100049514" | "20100057457" | "20100070899" | "20100081456" | "20100088020" | "20100088100" | "20100131273" | "20100138215" | "20100138416" | "20100145700" | "20100146442" | "20100161554" | "20100204986" | "20100217604" | "20100228540" | "20100231474" | "20100235341" | "20100257160" | "20100262599" | "20100277579" | "20100278320" | "20100278453" | "20100280983" | "20100286985" | "20100299133" | "20100299142" | "20100312547" | "20100318576" | "20100324905" | "20100332235" | "20100332280" | "20100332348" | "20110047072" | "20110054901" | "20110060807" | "20110076994" | "20110082688" | "20110112827" | "20110112921" | "20110119049" | "20110125540" | "20110130958" | "20110131036" | "20110131045" | "20110143811" | "20110144999" | "20110161076" | "20110161309" | "20110175810" | "20110184730" | "20110195758" | "20110218855" | "20110224972" | "20110231182" | "20110231188" | "20110231474").PN. OR ("20110260861" | "20110264643" | "20110279368" | "20110288861" | "20110298585" | "20110306426" | "20110314404" | "20120002820" | "20120016678" | "20120020490" | "20120022787" | "20120022857" | "20120022860" | "20120022868" | "20120022869" | "20120022870" | "20120022872" | "20120022874" | "20120022876" | "20120023088" | "20120034904" | "20120035908" | "20120035924" | "20120035931" | "20120035932" | "20120042343" | "20120084086" | "20120137367" | "20120149394" | "20120173464" | "20120197998" | "20120214517" | "20120245719" | "20120245944" | "20120265528" | "20120271635" | "20120271676" | "20120309363" | "20120310642" | "20120310649" | "20120311583" | "20120311584" | "20120311585" | "20120330661" | "20130006638" | "20130110515" | "20130110518" | "20130110519" | "20130110520" | "20130111348" | "20130111487" | "20130115927" | "20130117022" | "20130185074" | "20130185081" | "20130325443" | "3704345" | "3828132" | "3979557" | "4278838" | "4282405" | "4310721" | "4348553" | "4653021" | "4680805" | "4688195" | "4692941" | "4718094" | "4724542" | "4726065" | "4727354" | "4776016" | "4783807" | "4811243" | "4819271" | "4827518" | "4827520" | "4829576" | "4829583" | "4833712" | "4839853" | "4852168" | "4862504" | "4878230" | "4903305" | "4905163" | "4914586" | "4914590" | "4944013" | "4955047" | "4965763" | "4974191" | "4975975" | "4977598" | "4992972" | "5007098" | "5010574" | "5020112" | "5021971" | "5022081" | "5027406" | "5031217" | "5032989" | "5040218" | "5047614" | "5057915").PN. OR ("5072452" | "5091945" | "5127053" | "5127055" | "5128672" | "5133011" | "5142584" | "5164900" | "5165007" | "5179627" | "5179652" | "5194950" | "5197005" | "5199077" | "5202952" | "5208862" | "5216747" | "5220639" | "5220657" | "5222146" | "5230036" | "5235680" | "5267345" | "5268990" | "5282265" | "5289562" | "5291286" | "5293448" | "5293452" | "5297170" | "5301109" | "5303406" | "5309359" | "5317507" | "5317647" | "5325297" | "5325298" | "5327498" | "5333236" | "5333275" | "5345536" | "5349645" | "5353377" | "5377301" | "5377303" | "5384892" | "5384893" | "5386494" | "5386556" | "5390279" | "5396625" | "5400434" | "5404295" | "5412756" | "5412804" | "5412806" | "5418951" | "5424947" | "5434777" | "5444823" | "5455888" | "5469529" | "5471611" | "5475587" | "5479488" | "5491758" | "5491772" | "5493677" | "5495604" | "5500905" | "5502790" | "5502791" | "5515475" | "5536902" | "5537618" | "5544264" | "5555343" | "5574823" | "5577164" | "5577241" | "5578808" | "5579436" | "5581655" | "5584024" | "5596676" | "5596994" | "5608624" | "5613036" | "5617507" | "5619694" | "5621859" | "5621903" | "5642464" | "5642519" | "5644727" | "5649060" | "5664055" | "5675819" | "5682539" | "5687077" | "5696962" | "5701400" | "5706442" | "5710886" | "5712957" | "5715468" | "5721827" | "5727950" | "5729694" | "5732216" | "5732390" | "5734750" | "5734791" | "5737734" | "5742705" | "5748974" | "5749081" | "5757979" | "5759101" | "5777614" | "5790978" | "5794050" | "5794182" | "5794207" | "5794237").PN. OR ("5799276" | "5812697" | "5812698" | "5822743" | "5825881" | "5826261" | "5828999" | "5835893" | "5839106" | "5841902" | "5845255" | "5857184" | "5860063" | "5862223" | "5864806" | "5864844" | "5867799" | "5873056" | "5875437" | "5878394" | "5884323" | "5895464" | "5895466" | "5899972" | "5909666" | "5913193" | "5915236" | "5915249" | "5920836" | "5920837" | "5930408" | "5930769" | "5933822" | "5936926" | "5940811" | "5941944" | "5943670" | "5948040" | "5956699" | "5960394" | "5960422" | "5963924" | "5966126" | "5970474" | "5974146" | "5982891" | "5983179" | "5987132" | "5987140" | "5987404" | "5987440" | "5991441" | "5999908" | "6016471" | "6023684" | "6024288" | "6026345" | "6026375" | "6026388" | "6026393" | "6029132" | "6038533" | "6052656" | "6055514" | "6055531" | "6064767" | "6064959" | "6064960" | "6069648" | "6070139" | "6070147" | "6073097" | "6076051" | "6076088" | "6078914" | "6081750" | "6081774" | "6088671" | "6088731" | "6092043" | "6094649" | "6101468" | "6105865" | "6108627" | "6119101" | "6122616" | "6125356" | "6144938" | "6161084" | "6173261" | "6173279" | "6177905" | "6188999" | "6195641" | "6205456" | "6208971" | "6226403" | "6233559" | "6233578" | "6240303" | "6246981" | "6249606" | "6259826" | "6260011" | "6260013" | "6260024" | "6266637" | "6275824" | "6282507" | "6285786" | "6289124" | "6308149" | "6311189" | "6317594" | "6317707" | "6317831" | "6321092" | "6324512" | "6334103" | "6353794" | "6356854" | "6356905" | "6366883" | "6366884" | "6397186").PN. OR ("6421672" | "6434522" | "6434524" | "6442518" | "6446076" | "6449620" | "6453281" | "6453292" | "6460029" | "6466654" | "6477488" | "6487534" | "6489951" | "6499013" | "6501937" | "6505158" | "6505175" | "6505183" | "6510417" | "6513063" | "6519565" | "6523061" | "6523172" | "6526382" | "6526395" | "6532444" | "6532446" | "6546388" | "6553344" | "6556983" | "6584464" | "6598022" | "6598039" | "6601026" | "6601234" | "6604059" | "6615172" | "6615175" | "6615220" | "6622121" | "6622136" | "6625583" | "6628808" | "6631346" | "6633846" | "6643401" | "6647260" | "6650735" | "6654740" | "6665639" | "6665640" | "6665641" | "6680675" | "6684187" | "6691064" | "6691090" | "6691111" | "6691151" | "6697780" | "6697824" | "6701294" | "6711585" | "6718324" | "6721728" | "6735632" | "6742021" | "6757362" | "6757718" | "6766320" | "6772123" | "6778951" | "6778952" | "6778962" | "6778970" | "6792082" | "6807574" | "6810379" | "6813491" | "6829603" | "6832194" | "6839464" | "6839669" | "6842767" | "6847966" | "6847979" | "6851115" | "6859931" | "6865533" | "6895257" | "6895380" | "6895558" | "6901399" | "6912498" | "6912499" | "6924828" | "6928614" | "6931384" | "6934684" | "6937975" | "6937986" | "6957076" | "6960734" | "6964023" | "6980949" | "6980955" | "6985865" | "6988071" | "6996531" | "6999927" | "7020685" | "7024363" | "7027974" | "7031530" | "7036128" | "7050977" | "7058569" | "7062428" | "7069560" | "7072686" | "7084758" | "7085723" | "7092887" | "7092928" | "7093693" | "7107204").PN. OR ("7127046" | "7127403" | "7136710" | "7137126" | "7139714" | "7139722" | "7152070" | "7177798" | "7197460" | "7200559" | "7203646" | "7216073" | "7216080" | "7225125" | "7228278" | "7233790" | "7233904" | "7266496" | "7269556" | "7277854" | "7290039" | "7292579" | "7299033" | "7302392" | "7302686" | "7310600" | "7315818" | "7324947" | "7349953" | "7362738" | "7376556" | "7376645" | "7379874" | "7386449" | "7389224" | "7392185" | "7398209" | "7403938" | "7409337" | "7415100" | "7418392" | "7426467" | "7427024" | "7447635" | "7454351" | "7460652" | "7467087" | "7475010" | "7483832" | "7483894" | "7487089" | "7496498" | "7496512" | "7502738" | "7508373" | "7522927" | "7523108" | "7526466" | "7528713" | "7529671" | "7529676" | "7539656" | "7543232" | "7546382" | "7548895" | "7552055" | "7555431" | "7558730" | "7571106" | "7577522" | "7580551" | "7580576" | "7599918" | "7603381" | "7620549" | "7624007" | "7634409" | "7636657" | "7640160" | "7647225" | "7649454" | "7657424" | "7664638" | "7672841" | "7676026" | "7684985" | "7684990" | "7693715" | "7693719" | "7693720" | "7698131" | "7702500" | "7702508" | "7707027" | "7707032" | "7707267" | "7711565" | "7711672" | "7716056" | "7720674" | "7720683" | "7721301" | "7725307" | "7725318" | "7725320" | "7725321" | "7729904" | "7729916" | "7734461" | "7747616" | "7752152" | "7756868" |

		"7774204" | "7783486" | "7801729" | "7809570" | "7809610" | "7818176" | "7818291" | "7822608" | "7823123" | "7826945" | "7831426" | "7840400" | "7840447").PN. OR ("7853574" | "7853664" | "7873519" | "7873654" | "7881936" | "7885844" | "7890652" | "7912702" | "7917367" | "7917497" | "7920678" | "7925525" | "7930168" | "7930197" | "7949529" | "7949534" | "7974844" | "7974972" | "7983915" | "7983917" | "7983997" | "7986431" | "7987151" | "7996228" | "7999669" | "8000453" | "8005664" | "8005679" | "8015006" | "8024195" | "8032383" | "8036901" | "8041570" | "8041611" | "8050500" | "8055708" | "8065155" | "8065156" | "8069046" | "8069422" | "8073681" | "8078473" | "8082153" | "8095364" | "8099289" | "8107401" | "8112275" | "8112280" | "8117037" | "8122353" | "8131557" | "8138912" | "8140335" | "8165886" | "8166019" | "8188856" | "8190359" | "8195467" | "8204238" | "8205788" | "8219407" | "8285551" | "8285553" | "8290777" | "8290778" | "8290781" | "8296146" | "8296153" | "8296383" | "8301456" | "8311834" | "8370158" | "8371503" | "8374871" | "8447612" | "8498857" | "RE34562").PN. OR ("8977255").URPN.				
S1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/03/18 21:04
S2	104	(("GANONG") near3 ("William")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S3	16	(("VAN") near3 ("MULBREGT") near3 ("Paul")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S4	38	(("SEJNOHA") near3 ("Vladimir")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S5	33	(("WILSON") near3 ("Glen")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:20
S6	220	(("NUANCE") near3 ("COMMUNICATIONS") near3 ("INC")).AANM.	USPAT	OR	OFF	2017/04/20 10:20
S7	43	("8977255" | "9112984" | "9361885" | "20070129949" | "20120329389" | "7574361" | "20100124896" | "20030040339" | "20100289994" | "20140274211" | "20140278435" | "8548176" | "20100185448" | "20110054899" | "20130080167" | "20130080171" | "20130289994" | "20130339028" | "20140012573" | "20140012586" | "5428707" | "5774859" | "5802305" | "5983186" | "6070140" | "6092043" | "6397186" | "6408396" | "6411926" | "6449496" | "6594630" | "6756700" | "6941265" | "6965786" | "7114090" | "7567827" | "7720683" | "7774204" | "8056070" | "8181046" | "8190420" | "8285545" | "8666751").PN.	US-PGPUB; USPAT	OR	OFF	2017/04/20 11:21
S8	5	"20120034904"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:10
S9	9	"9361885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:32
S10	33784	(379/88.01-88.02|G10L15/20|G10L15/22|G10L15/24|G10L15/25|G10L2015/223|G10L2015/226|H04M2250/74|H04W52/0251|H04W52/0254|).ccls.cpc.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:11
S11	1	("15".src. and "238238".ap.) or "15238238".RLAN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/04/20 16:12
S12	25130	S10 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:13

S13	4617	S12 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S14	2692	S13 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S15	563	S14 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S16	362	S15 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S17	17	S16 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:33
S18	9564	(704/246-255).ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S19	8758	S18 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S20	1733	S19 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S21	1338	S20 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S22	212	S21 and (trigger)	US-PGPUB; USPAT;	OR	ON	2017/04/20 16:36

			USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB			
S23	104	S22 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S24	3	S23 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S25	3	"9646610"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:41
S26	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/09/20 13:42
S27	8	"20100121636"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:42
S28	5	"20120035924"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:43
S29	6	"20120127072"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:45
S30	2	"20030216909"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2018/01/10 15:20
S31	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2018/01/10 15:20

1/10/2018 5:05:13 PM
C:\Users\hho3\Documents\EAST\Workspaces\15238238.wsp

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: Huy C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		2003-0216909	A1	Davis et al.	11-20-2003

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		CN	105009204	A	Amazon Tech Inc.	10-28-2015	Y-abstract only
		EP	2932500	B1	Amazon Tech Inc.	02-15-2017	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Chinese Office Action and Translation thereof for Chinese Application No. 201480013903.1 dated July 28, 2017 (N0484.71090CN00).	Y

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:
/HUY C HO/	01/10/2018

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

5675651.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 03/14/2018

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
03/14/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@wolfgreenfield.com
N0484_eOfficeAction@WolfGreenfield.com
IP.Inbox@nuance.com

Applicant-Initiated Interview Summary	Application No.	Applicant(s)	
	15/238,238	GANONG ET AL.	
	Examiner	Art Unit	
	HUY C. HO	2644	

All participants (applicant, applicant's representative, PTO personnel):

(1) *HUY C. HO*.

(2) *THOMAS S. CHLEBECK (Reg. No. 70,616)*.

(3) *MARC JOHANNES (Reg. No. 64,978)*.

(4) _____.

Date of Interview: *03/09/2018*.

Type: ☒ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☒ No.
If Yes, brief description: *N/A*.

Issues Discussed ☐101 ☐112 ☐102 ☒103 ☒Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: *1*.

Identification of prior art discussed: *Mozer*.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Discussions for claim 1 with respect to a proposed amended features, i.e., wherein the explicit voice trigger is a specific and predetermined word or phrase that, when detected, is sufficient to engage the device, over the prior art reference Mozer. However, no agreement was reached at this time.

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HUY C HO/ Primary Examiner, Art Unit 2644	

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record
A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews
Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.
All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiner's responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicant's correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:
- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:

1) A brief description of the nature of any exhibit shown or any demonstration conducted,
2) an identification of the claims discussed,
3) an identification of the specific prior art discussed,
4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
5) a brief identification of the general thrust of the principal arguments presented to the examiner,
   (The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
6) a general indication of any other pertinent matters discussed, and
7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicant's record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiner's version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, "Interview Record OK" on the paper recording the substance of the interview along with the date and the examiner's initials.

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 CFR §1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 18, 2018 Electronic Signature for: /Trish McDonald/

RESPONSE AFTER FINAL ACTION UNDER 37 C.F.R. § 1.116

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Final Office Action dated January 18, 2018, please enter the following in the above-identified U.S. patent application:

Amendments to the Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 6 of this paper.

Doc Code: A.NE.AFCP
Document Description: After Final Consideration Pilot Program Request

PTO/SB/434 (05-13)

CERTIFICATION AND REQUEST FOR CONSIDERATION UNDER THE AFTER FINAL CONSIDERATION PILOT PROGRAM 2.0

Practitioner Docket No.:	Application No.:	Filing Date:
N0484.71090US04	15/238,238-Conf. #5864	August 16, 2016

First Named Inventor:	Title:
William F. Ganong, III	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

APPLICANT HEREBY CERTIFIES THE FOLLOWING AND REQUESTS CONSIDERATION UNDER THE AFTER FINAL CONSIDERATION PILOT PROGRAM 2.0 (AFCP 2.0) OF THE ACCOMPANYING RESPONSE UNDER 37 CFR 1.116.

1. The above-identified application is (i) an original utility, plant, or design nonprovisional application filed under 35 U.S.C. 111(a) [a continuing application (*e.g.*, a continuation or divisional application) is filed under 35 U.S.C. 111(a) and is eligible under (i)], or (ii) an international application that has entered the national stage in compliance with 35 U.S.C. 371(c).

2. The above-identified application contains an outstanding final rejection.

3. Submitted herewith is a response under 37 CFR 1.116 to the outstanding final rejection. The response includes an amendment to at least one independent claim, and the amendment does not broaden the scope of the independent claim in any aspect.

4. This certification and request for consideration under AFCP 2.0 is the only AFCP 2.0 certification and request filed in response to the outstanding final rejection.

5. Applicant is willing and available to participate in any interview requested by the examiner concerning the present response.

6. This certification and request is being filed electronically using the Office's electronic filing system (EFS-Web).

7. Any fees that would be necessary consistent with current practice concerning responses after final rejection under 37 CFR 1.116, *e.g.*, extension of time fees, are being concurrently filed herewith. [There is no additional fee required to request consideration under AFCP 2.0.]

8. By filing this certification and request, applicant acknowledges the following:

- Reissue applications and reexamination proceedings are not eligible to participate in AFCP 2.0.
- The examiner will verify that the AFCP 2.0 submission is compliant, *i.e.*, that the requirements of the program have been met (see items 1 to 7 above). For compliant submissions:
  - The examiner will review the response under 37 CFR 1.116 to determine if additional search and/or consideration (i) is necessitated by the amendment and (ii) could be completed within the time allotted under AFCP 2.0. If additional search and/or consideration is required but cannot be completed within the allotted time, the examiner will process the submission consistent with current practice concerning responses after final rejection under 37 CFR 1.116, *e.g.*, by mailing an advisory action.
  - If the examiner determines that the amendment does not necessitate additional search and/or consideration, or if the examiner determines that additional search and/or consideration is required and could be completed within the allotted time, then the examiner will consider whether the amendment places the application in condition for allowance (after completing the additional search and/or consideration, if required). If the examiner determines that the amendment does not place the application in condition for allowance, then the examiner will contact the applicant and request an interview.
    - The interview will be conducted by the examiner, and if the examiner does not have negotiation authority, a primary examiner and/or supervisory patent examiner will also participate.
    - If the applicant declines the interview, or if the interview cannot be scheduled within ten (10) calendar days from the date that the examiner first contacts the applicant, then the examiner will proceed consistent with current practice concerning responses after final rejection under 37 CFR 1.116.

Signature	/Thomas S. Chlebeck/	Date	2018-04-18
Name (Print/Typed)	Thomas S. Chlebeck	Practitioner Registration No.	70,616

***Note**: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.*

☐ * Total of 1 forms are submitted.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 18, 2018 Electronic Signature for: /Trish McDonald/

REMARKS

In response to the Office Action mailed January 18, 2018, Applicant respectfully requests reconsideration in view of the amendments and the following remarks. Claims 1-20 were previously pending in this application. By this amendment, claims 1 and 15 are amended. No claims are added or canceled. As a result, claims 1-20 are pending for examination with claims 1 and 15 being independent claims. No new matter has been added.

Telephone Interview Summary

The Assignee's representatives, Thomas S. Chlebeck and Marc S. Johannes, thank Examiner Ho for his courtesy extended in granting and conducting a telephone interview on March 9, 2018. The substance of that interview is summarized herein.

Assignee's representatives discussed embodiments of the specification and proposed amendments included in the enclosed agenda. Assignee's representatives further discussed how the proposed claims distinguish over the cited references. In particular, Assignee's representatives discussed techniques described in the specification to provide examples of how a device in low power mode detects whether acoustic input includes an explicit voice trigger (which is a specific word or phrase that, when detected, is sufficient to engage the device (such as "Hello, Dragon")) and then transmits at least a portion of the acoustic input to a server to determine if a voice command is included (or to recognize the command).

Assignee's representatives explained that the proposed amendments explicitly recite the meaning given to the term "explicit voice trigger" in the specification, which distinguishes the keywords described in Mozer because the keywords in Mozer are merely a collection of words that the first stage recognizer listens for in a conversation between users that, if a certain number of words from a given category are heard, the first stage recognizer takes further action. None of the keywords, alone or taken collectively, are a specific word or phrase that, when detected, is sufficient to engage the device. During the telephone interview, the Examiner questioned if the "trigger word" in ¶[23 of Mozer is the same as the claimed "explicit voice trigger." Assignee's representatives explained the "trigger word" of Mozer appears to be an explicit voice trigger, but pointed out that the "trigger word" in Mozer fails to meet the requirement of independent claims 1

Petitioner Exhibit 1004, Page 279 of 1052

and 15 that "the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in-the acoustic input."

The Examiner appeared to appreciate the distinctions explained by Assignee's representative, but he indicated that he would need to verify that it would overcome the rejection based on the Freeman and Mozer references. The Examiner suggested the Assignee's representatives submit corresponding amendments under the After Final Consideration Pilot 2.0 program, which is done herein, with additional amendments.

Double Patenting Rejection

The Office Action rejected claim 1 for purportedly claiming the same invention as that of claims 1 and 5 of prior U.S. Patent No. 9,361,885. While the Assignee does not agree, a Terminal Disclaimer is filed herewith to address the Examiner's concern. Accordingly, withdrawal of this rejection is respectfully requested.

Rejections Under 35 U.S.C. §103

The Office Action rejects independent claims 1 and 15 under 35 U.S.C. §103(a) as purportedly being unpatentable over Freeman (U.S. Patent No. 8,977,255) in view of Mozer (U.S. Patent Publication No. 2013/0080167). The Assignee respectfully requests reconsideration. While the Assignee believes that the claims, as previously presented, patentably distinguish over the alleged combination of Freeman and Mozer, the independent claims are amended herein to recite the meaning of the term explicit voice trigger given to this term by the specification.

As discussed during the telephone interview, claims 1 and 15 are amended herein to address the Examiner's concern. In particular, claim 1 is amended to recite that the at least one processor is configured to "perform, while in the low power mode, at least one processing stage on the acoustic input to ***detect whether the acoustic input includes an explicit voice trigger***, wherein the explicit voice trigger is ***a specific word or phrase that, when detected, is sufficient to engage the device***, the at least one processing stage, ***upon detecting that the acoustic input includes the explicit voice***

Petitioner Exhibit 1004, Page 280 of 1052

trigger, requesting, via a network, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command." Claim 15 is amended to recite "performing, while in the low power mode, at least one processing stage on the acoustic input to ***detect whether the acoustic input includes an explicit voice trigger***, wherein the explicit voice trigger is ***a specific word or phrase that, when detected, is sufficient to engage the device***, the at least one processing stage, ***upon detecting that the acoustic input includes the explicit voice trigger, requesting, via a network, that at least one server perform recognition on at least a portion of the acoustic input*** to determine whether the acoustic input includes at least one voice command."

As discussed during the telephone interview, neither Freeman nor Mozer disclose or suggest the above quoted limitations in claims 1 and 15, and the above amendments distinguish over the alleged combination of Freeman and Mozer. Specifically, Mozer describes a stage 1 recognizer that seeks to identify keywords in a conversation and to classify those keywords. Mozer at ¶ 33, 42-44. When enough keywords of the same category are identified, the stage 1 recognizer may signal to the stage 2 recognizer to begin performing more sophisticated recognition techniques to assess whether the user is discussing a topic to which the system may contribute. *Id.* These keywords are therefore not the same as an "explicit voice trigger," which is recited as being "a specific word or phrase that, when detected, is sufficient to engage the device." Specifically, the keywords in Mozer are merely words that a conversant uses in a regular conversation, ***not*** a specific word or phrase used to engage the voice response system. Indeed, the users in Mozer are merely conversing and not trying to engage the voice response system. Moreover, none of the keywords are sufficient to engage the device.

During the telephone interview, the Examiner questioned if the "trigger word" in ¶ 23 of Mozer is the same as the claimed "explicit voice trigger … a specific word or phrase that, when detected, is sufficient to engage the device." Assignee's representatives explained the "trigger word" of Mozer is arguably an explicit voice trigger, but the "trigger word" in Mozer fails to meet the requirement in previously pending independent claims 1 and 15 that "the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one

Petitioner Exhibit 1004, Page 281 of 1052

server to recognize a voice command included in-the acoustic input." Therefore, that Mozer discloses the "trigger word" in ¶ 23 has no bearing on whether the combination of Freeman and Mozer meets each limitation of claims 1 and 15 because Mozer still fails to disclose ***requesting, via a network, that at least one server perform recognition on at least a portion of the acoustic input*** to determine whether the acoustic input includes at least one voice command ***upon detecting that the acoustic input includes the explicit voice trigger***, as required by amended claims 1 and 15.

Accordingly, the keywords of Mozer are not an "explicit voice trigger" and the system of Mozer does not request that at least one server perform recognition on at least a portion of the acoustic input upon detecting a trigger word. Thus, the alleged combination of Freeman and Mozer fails to disclose or suggest performing "at least one processing stage on the acoustic input to ***detect whether the acoustic input includes an explicit voice trigger***, wherein the explicit voice trigger is ***a specific word or phrase that, when detected, is sufficient to engage the device***, the at least one processing stage, ***upon detecting that the acoustic input includes the explicit voice trigger, requesting, via a network, that at least one server perform recognition on at least a portion of the acoustic input*** to determine whether the acoustic input includes at least one voice command," as recited in claims 1 and 15 (emphasis added). Therefore, claims 1 and 15, and each claim that depends respectively therefrom, patentably distinguish over the alleged combination and are in allowable condition.

In view of the foregoing, the Assignee respectfully requests withdrawal of the rejections under 35 U.S.C. §103.

General Comments on Dependent Claims

Because each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Assignee believes that it is unnecessary at this time to argue the further distinguishing features of all of the dependent claims. However, the Assignee does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does the Assignee concur that the basis for the rejection of any of the dependent claims is proper. Therefore, the Assignee reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

Petitioner Exhibit 1004, Page 282 of 1052

CONCLUSION

In view of the foregoing amendments and remarks, reconsideration is respectfully requested. This application should now be in condition for allowance; a notice to this effect is respectfully requested. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the Applicant's attorney at the telephone number listed below.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, Applicant hereby requests any necessary extension of time. Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Dated: 2018-04-18

Respectfully submitted,

By /Thomas S. Chlebeck/
Thomas S. Chlebeck
Registration No.: 70,616
Marc S. Johannes
Registration No.: 64,978
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Petitioner Exhibit 1004, Page 283 of 1052

AMENDMENTS TO THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims:

1. (Currently amended) A device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and

at least one processor configured to perform, while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger, wherein the explicit voice trigger is a specific word or phrase that, when detected, is sufficient to engage the device, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, ~~transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a~~ requesting, via a network, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command ~~included in the acoustic input~~,

wherein the device is configured to respond to voice commands recognized by the at least one server.

2. (Previously presented) The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.

3. (Previously presented) The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

4. (Previously presented) The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

Petitioner Exhibit 1004, Page 284 of 1052

5. (Previously presented) The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

6. (Previously presented) The device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger.

7. (Previously presented) The device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

8. (Previously presented) The device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

9. (Previously presented) The device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. (Previously presented) The device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

Petitioner Exhibit 1004, Page 285 of 1052

11. (Previously presented) The device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the idle mode.

12. (Previously presented) The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. (Previously presented) The device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

14. (Previously presented) The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

15. (Currently amended) At least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:

performing, while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger, wherein the explicit voice trigger is a specific word or phrase that, when detected, is sufficient to engage the device, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, ~~transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a~~ requesting, via a network, that at least one server perform

Petitioner Exhibit 1004, Page 286 of 1052

recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command ~~included in the acoustic input~~; and

responding to voice commands recognized by the at least one server.

16. (Original) The at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

17. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger.

18. (Original) The at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the idle mode.

20. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

Petitioner Exhibit 1004, Page 287 of 1052

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [X] Fee Transmittal Form
  - [ ] Fee Attached
- [X] Amendment/Reply
  - [X] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [X] Power of Attorney, Revocation Change of Correspondence Address
- [X] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [X] Other Enclosure(s) (please Identify below):
  - After Final Consideration Pilot Program
  - Statement under 37 CFR 3.73(c)

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Thomas S. Chlebeck/		
Printed name	Thomas S. Chlebeck		
Date	2018-04-18	Reg. No.	70,616

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 18, 2018 Electronic Signature for: /Trish McDonald/

6036262.1

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Examiner Name	H. C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)160.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☒ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☐ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17 ☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
20	- or HP =		x	=	

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**

Indep. Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
	- or HP =		x	=	

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		**Extra Sheets**		**Number of each additional 50 or fraction thereof**		**Fee ($)**		**Fee Paid ($)**
	- 100 =		/50 =		(round **up** to a whole number) x		=	

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1814 Statutory disclaimer, including terminal ...	160.00

SUBMITTED BY

Signature	/Thomas S. Chlebeck/	Registration No. (Attorney/Agent)	70,616	Telephone	617.646.8000
Name (Print/Type)	Thomas S. Chlebeck			Date	2018-04-18

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 18, 2018 Electronic Signature for: /Trish McDonald/

6230110.1

Electronic Acknowledgement Receipt

EFS ID:	32378544
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Thomas Stephen Chlebeck/Trish McDonald
Filer Authorized By:	Thomas Stephen Chlebeck
Attorney Docket Number:	N0484.71090US04
Receipt Date:	18-APR-2018
Filing Date:	16-AUG-2016
Time Stamp:	18:00:22
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$160
RAM confirmation Number	041918INTEFSW18011100
Deposit Account	232825
Authorized User	Wolf Greenfield

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.20 (Post Issuance fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-TSC.pdf	25293 fb2e1eb053385a4107abf8e6c2f5fb6d273ca2af	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-TSC.pdf	31023 dc10df2d91d36602e7636b4d642541e395416563	no	1
Warnings:					
Information:					
3	Assignee showing of ownership per 37 CFR 3.73	N048471090US04-373-TSC.pdf	31028 a67893fbf4fbed3126165d0e9b6f0214ff26bdd2	no	2
Warnings:					
Information:					
4	Power of Attorney	N048471090US04-POA-TSC.PDF	47000 a24d238133af45b0863c4f33196d5ed16b2e2f92	no	1
Warnings:					
Information:					
5	After Final Consideration Program Request	N048471090US04-RAFPP-TSC.pdf	37816 c8fe7b2bf99f3d0461b773814ff83b7c4816e13d	no	1
Warnings:					
Information:					
6		N048471090US04-116AMN-TSC.pdf	57846 5cc9be332365b4f3a2b0d6aa73e21e1c534ae7a5	yes	10
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	

	Response After Final Action		1		1
	Claims		2		5
	Applicant Arguments/Remarks Made in an Amendment		6		10
Warnings:					
Information:					
7	Terminal Disclaimer Filed	N048471090US04-TD-TSC.pdf	20368 8919253912f11af6581400ca1f3aae8c4ec2b7ac	no	1
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	30817 98e117c067a7a9b11fcf76cc50472db5a3c9b00e	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	281191		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Thomas Stephen Chlebeck/Trish McDonald
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
STATUTORY OR TERMINAL DISCLAIMER	1814	1	160	160
		Total in USD ($)		**160**

PTO/AIA/96 (08-12)
Approved for use through 01/31/2013. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Nuance Communications, Inc.

Application No./Patent No.: 15/238,238 Filed/Issue Date: August 16, 2016

Titled: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Nuance Communications, Inc. (Name of Assignee) , a Corporation (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. [X] The assignee of the entire right, title, and interest.

2. [ ] An assignee of less than the entire right, title, and interest (check applicable box):

   [ ] (The extent (by percentage) of its ownership interest is _____ %). Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   [ ] There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. [ ] The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. [ ] The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2, or 3 above (not option 4) is evidence by either (choose **one** of options A or B below):

A. [ ] An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____ , Frame _____ , or for which a copy thereof is attached.

B. [X] A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

   1. From: William F. Ganong To: Nuance Communications, Inc.
      The document was recorded in the United States Patent and Trademark Office at
      Reel 040342 , Frame 0893 , or for which a copy thereof is attached.

   2. From: Paul Adrian Van Mulbregt To: Nuance Communications, Inc.
      The document was recorded in the United States Patent and Trademark Office at
      Reel 040342 , Frame 0893 , or for which a copy thereof is attached.

PTO/AIA/96 (08-12)
Approved for use through 01/31/2013. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: Vladimir Sejnoha To: Nuance Communications, Inc.
The document was recorded in the United States Patent and Trademark Office at
Reel 040342 , Frame 0893 , or for which a copy thereof is attached.

4. From: Glen Edward Wilson To: Nuance Communications, Inc.
The document was recorded in the United States Patent and Trademark Office at
Reel 040342 , Frame 0893 , or for which a copy thereof is attached.

5. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

6. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☐ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Thomas S. Chlebeck/	2018-04-18
Signature	Date
Thomas S. Chlebeck	70,616
Printed or Typed Name	Title or Registration Number

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 18, 2018 Electronic Signature for: /Trish McDonald/

PTO/AIA/26 (04-14)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TERMINAL DISCLAIMER TO OBVIATE A DOUBLE PATENTING REJECTION OVER A "PRIOR" PATENT	Docket Number (Optional) N0484.71090US04

In re Application of: William F. Ganong, III et al.

Application No.: 15/238,238-Conf. #5864

Filed: August 16, 2016

For: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

The applicant, Nuance Communications, Inc., owner of 100 percent interest in the instant application hereby disclaims, except as provided below, the terminal part of the statutory term of any patent granted on the instant application which would extend beyond the expiration date of the full statutory term of **prior patent** No. 9,361,885 as the term of said **prior patent** is presently shortened by any terminal disclaimer. The applicant hereby agrees that any patent so granted on the instant application shall be enforceable only for and during such period that it and the **prior patent** are commonly owned. This agreement runs with any patent granted on the instant application and is binding upon the grantee, its successors or assigns.

In making the above disclaimer, the applicant does not disclaim the terminal part of the term of any patent granted on the instant application that would extend to the expiration date of the full statutory term of the **prior patent**, "as the term of said **prior patent** is presently shortened by any terminal disclaimer," in the event that said **prior patent** later:

expires for failure to pay a maintenance fee;
is held unenforceable;
is found invalid by a court of competent jurisdiction;
is statutorily disclaimed in whole or terminally disclaimed under 37 CFR 1.321;
has all claims canceled by a reexamination certificate;
is reissued; or
is in any manner terminated prior to the expiration of its full statutory term as presently shortened by any terminal disclaimer.

Check either box 1 or 2 below, if appropriate.

1. ☐ The undersigned is the applicant. If the applicant is an assignee, the undersigned is authorized to act on behalf of the assignee.

I hereby acknowledge that any willful false statements made are punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

2. ☒ The undersigned is an attorney or agent of record. Reg. No. 70,616

/Thomas S. Chlebeck/ — Signature
2018-04-18 — Date

Thomas S. Chlebeck — Typed or printed name

Attorney for Applicant(s) — Title
617.646.8000 — Telephone Number

☒ Terminal disclaimer fee under 37 CFR 1.20(d) included.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 18, 2018 Electronic Signature for: /Trish McDonald/

PTO/AIA/80 (07-12)
Approved for use through 11/30/2014. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

[x] Practitioners associated with the Customer Number: 91337

OR

[ ] Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number	Name	Registration Number

as attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned **only** to the undersigned according to the USPTO assignment records or assignment documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

[x] The address associated with Customer Number: 91337

OR

[ ] Firm or Individual Name			
Address			
City		State	Zip
Country			
Telephone		Email	

Assignee Name and Address:

Nuance Communications, Inc.
One Wayside Road
Burlington, Massachusetts 01803

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/SB/96 or equivalent) is required to be filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of the practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee

Signature		Date	11/8/2012
Name	Leanne J. Fitzgerald, Esq.	Telephone	781 565-4792
Title	Associate General Counsel		

2887608.1

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 15/238,238	Filing Date 08/16/2016	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	04/18/2018	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 20	Minus	** 20	= 0		x $ =	0
	Independent (37 CFR 1.16(h))	* 2	Minus	*** 3	= 0		x $ =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $ =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $ =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	SLIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	shanda E ross
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Application Number	Application/Control No.	Applicant(s)/Patent under Reexamination	
	15/238,238	GANONG ET AL.	

Document Code - DISQ	Internal Document – DO NOT MAIL

TERMINAL DISCLAIMER	☒ APPROVED	☐ DISAPPROVED
Date Filed : 4/18/18	This patent is subject to a Terminal Disclaimer	

Approved/Disapproved by:
Lawana Hixon

U.S. Patent and Trademark Office

Application Number	Application/Control No.	Applicant(s)/Patent under Reexamination	
	15/238,238	GANONG ET AL.	

Document Code - DISQ	Internal Document – DO NOT MAIL

TERMINAL DISCLAIMER	☒ APPROVED	☐ DISAPPROVED
Date Filed : 4/18/18	This patent is subject to a Terminal Disclaimer	

Approved/Disapproved by:
Lawana Hixon

U.S. Patent and Trademark Office

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04

CONFIRMATION NO. 5864

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

POA ACCEPTANCE LETTER

OC000000098950624

Date Mailed: 04/24/2018

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 04/18/2018.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/nbekele/

AFCP 2.0 Decision	Application No.	Applicant(s)
	15/238,238	GANONG ET AL.
	Examiner	Art Unit
	HUY C. HO	2644

This is in response to the After Final Consideration Pilot request filed 18 April 2018.

1. **Improper Request** – The AFCP 2.0 request is improper for the following reason(s) and the after final amendment submitted with the request will be treated under pre-pilot procedure.

   ☐ An AFCP 2.0 request form PTO/SB/434 (or equivalent document) was not submitted.

   ☐ A non-broadening amendment to at least one independent claim was not submitted.

   ☐ A proper AFCP 2.0 request was submitted in response to the most recent final rejection.

   ☐ Other:

2. **Proper Request**

   A. After final amendment submitted with the request will not be treated under AFCP 2.0.
   The after final amendment cannot be reviewed and a search conducted within the guidelines of the pilot program.

   ☐ The after final amendment will be treated under pre-pilot procedure.

   B. Updated search and/or completed additional consideration.
   The examiner performed an updated search and/or completed additional consideration of the after final amendment within the time authorized for the pilot program. The result(s) of the updated search and/or completed additional consideration are:

   ☐ 1. All of the rejections in the most recent final Office action are overcome and a Notice of Allowance is issued herewith.

   ☒ 2. The after final amendment would not overcome all of the rejections in the most recent final Office action. See attached interview summary for further details.

   ☐ 3. The after final amendment was reviewed, and it raises a new issue(s). See attached interview summary for further details.

   ☐ 4. The after final amendment raises new issues, but would overcome all of the rejections in the most recent final Office action. A decision on determining allowability could not be made within the guidelines of the pilot. See attached interview summary for further details, including any newly discovered prior art.

   ☒ 5. Other: See the Interview Summary and the Advisory Action

   Examiner Note: Please attach an interview summary when necessary as described above.

Petitioner Exhibit 1004, Page 303 of 1052

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 06/05/2018

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
HO, HUY C

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
06/05/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@wolfgreenfield.com
N0484_eOfficeAction@WolfGreenfield.com
IP.Inbox@nuance.com

Examiner-Initiated Interview Summary	Application No.	Applicant(s)	
	15/238,238	GANONG ET AL.	
	Examiner	Art Unit	
	HUY C. HO	2644	

All participants (applicant, applicant's representative, PTO personnel):

(1) *HUY C. HO*.

(2) *THOMAS S. CHLEBECK*.

(3) *MARC S. JOHANNES*.

(4) _____.

Date of Interview: *29 May 2018*.

Type: ☒ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☒ No.
If Yes, brief description: *N/A*.

Issues Discussed ☐101 ☐112 ☐102 ☒103 ☒Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: *1*.

Identification of prior art discussed: *Mozer*.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Discussed the argued features,i.e.,"explicit voice trigger is a specific word or phrase" and "requesting via a network that at least one server perform recognition on at least a portion of the acoustic input to determinine weather the acoustic input includes at least a voice command", over the prior art reference Mozer. The argued features are written such that they still read upon the cited prior art reference as taught and described from the paragraphs [32], [36]-[38], Figure 5, [55] and Figure 8 paragraphs [66]-[67].

Applicant recordation instructions: It is not necessary for applicant to provide a separate record of the substance of interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HUY C HO/ Primary Examiner, Art Unit 2644	

Advisory Action Before the Filing of an Appeal Brief	**Application No.** 15/238,238	**Applicant(s)** GANONG ET AL.	
	Examiner HUY C. HO	**Art Unit** 2644	**AIA (First Inventor to File) Status** No

--The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

THE REPLY FILED 18 April 2018 FAILS TO PLACE THIS APPLICATION IN CONDITION FOR ALLOWANCE.

NO NOTICE OF APPEAL FILED

1. ☒ The reply was filed after a final rejection. No Notice of Appeal has been filed. To avoid abandonment of this application, applicant must timely file one of the following replies: (1) an amendment, affidavit, or other evidence, which places the application in condition for allowance;

(2) a Notice of Appeal (with appeal fee) in compliance with 37 CFR 41.31; or (3) a Request for Continued Examination (RCE) in compliance with 37 CFR 1.114 if this is a utility or plant application. Note that RCEs are not permitted in design applications. The reply must be filed within one of the following time periods:

a) ☒ The period for reply expires *3* months from the mailing date of the final rejection.

b) ☐ The period for reply expires on: (1) the mailing date of this Advisory Action; or (2) the date set forth in the final rejection, whichever is later. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of the final rejection.

c) ☐ A prior Advisory Action was mailed more than 3 months after the mailing date of the final rejection in response to a first after-final reply filed within 2 months of the mailing date of the final rejection. The current period for reply expires months from the mailing date of *the prior Advisory Action* or SIX MONTHS from the mailing date of the final rejection, whichever is earlier.

Examiner Note: If box 1 is checked, check either box (a), (b) or (c). ONLY CHECK BOX (b) WHEN THIS ADVISORY ACTION IS THE FIRST RESPONSE TO APPLICANT'S FIRST AFTER-FINAL REPLY WHICH WAS FILED WITHIN TWO MONTHS OF THE FINAL REJECTION. ONLY CHECK BOX (c) IN THE LIMITED SITUATION SET FORTH UNDER BOX (c). See MPEP 706.07(f).

Extensions of time may be obtained under 37 CFR 1.136(a). The date on which the petition under 37 CFR 1.136(a) and the appropriate extension fee have been filed is the date for purposes of determining the period of extension and the corresponding amount of the fee. The appropriate extension fee under 37 CFR 1.17(a) is calculated from: (1) the expiration date of the shortened statutory period for reply originally set in the final Office action; or (2) as set forth in (b) or (c) above, if checked. Any reply received by the Office later than three months after the mailing date of the final rejection, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

NOTICE OF APPEAL

2. ☐ The Notice of Appeal was filed on _____. A brief in compliance with 37 CFR 41.37 must be filed within two months of the date of filing the Notice of Appeal (37 CFR 41.37(a)), or any extension thereof (37 CFR 41.37(e)), to avoid dismissal of the appeal. Since a Notice of Appeal has been filed, any reply must be filed within the time period set forth in 37 CFR 41.37(a).

AMENDMENTS

3. ☒ The proposed amendments filed after a final rejection, but prior to the date of filing a brief, will not be entered because
   a) ☒ They raise new issues that would require further consideration and/or search (see NOTE below);
   b) ☐ They raise the issue of new matter (see NOTE below);
   c) ☐ They are not deemed to place the application in better form for appeal by materially reducing or simplifying the issues for appeal; and/or
   d) ☐ They present additional claims without canceling a corresponding number of finally rejected claims.

   NOTE: _____. (See 37 CFR 1.116 and 41.33(a)).

4. ☐ The amendments are not in compliance with 37 CFR 1.121. See attached Notice of Non-Compliant Amendment (PTOL-324).

5. ☐ Applicant's reply has overcome the following rejection(s): _____.

6. ☐ Newly proposed or amended claim(s) _____ would be allowable if submitted in a separate, timely filed amendment canceling the non-allowable claim(s).

7. ☒ For purposes of appeal, the proposed amendment(s): (a) ☒ will not be entered, or (b) ☐ will be entered, and an explanation of how the new or amended claims would be rejected is provided below or appended.

AFFIDAVIT OR OTHER EVIDENCE

8. ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

9. ☐ The affidavit or other evidence filed after final action, but before or on the date of filing a Notice of Appeal will not be entered because applicant failed to provide a showing of good and sufficient reasons why the affidavit or other evidence is necessary and was not earlier presented. See 37 CFR 1.116(e).

10. ☐ The affidavit or other evidence filed after the date of filing the Notice of Appeal, but prior to the date of filing a brief, will not be entered because the affidavit or other evidence failed to overcome all rejections under appeal and/or appellant fails to provide a showing of good and sufficient reasons why it is necessary and was not earlier presented. See 37 CFR 41.33(d)(1).

11. ☐ The affidavit or other evidence is entered. An explanation of the status of the claims after entry is below or attached.

REQUEST FOR RECONSIDERATION/OTHER

12. ☒ The request for reconsideration has been considered but does NOT place the application in condition for allowance because: See Continuation Sheet.

13. ☐ Note the attached Information *Disclosure Statement*(s). (PTO/SB/08) Paper No(s). _____

14. ☒ Other: Please see the AFCP 2.0 Decision.

STATUS OF CLAIMS

15. The status of the claim(s) is (or will be) as follows:

Claim(s) allowed: .
Claim(s) objected to: .
Claim(s) rejected: 1-20.
Claim(s) withdrawn from consideration: .

/HUY C HO/
Primary Examiner, Art Unit 2644

Continuation of 12. does NOT place the application in condition for allowance because: The newly amended features have been added into the independent claims 1 and 15, however the new amended feautures have been discussed during the interview dated 05/29/2018 (see more details from the Interview Summary attached herein) over the prior art reference of record, i.e., Mozer, and they do not overcome the prior art rejection since Mozer still teaches and discloses the argued features above as from paragraphs [32], [36]-[38], [55] and [66]-[67], Figures 5 and 8.

Examiner-Initiated Interview Summary	Application No.	Applicant(s)	
	15/238,238	GANONG ET AL.	
	Examiner	Art Unit	
	HUY C. HO	2644	

All participants (applicant, applicant's representative, PTO personnel):

(1) *HUY C. HO*.

(2) *THOMAS S. CHLEBECK*.

(3) *MARC S. JOHANNES*.

(4) _____.

Date of Interview: *29 May 2018*.

Type: ☒ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☒ No.
If Yes, brief description: *N/A*.

Issues Discussed ☐101 ☐112 ☐102 ☒103 ☒Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: *1*.

Identification of prior art discussed: *Mozer*.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Discussed the argued features,i.e.,"explicit voice trigger is a specific word or phrase" and "requesting via a network that at least one server perform recognition on at least a portion of the acoustic input to determinine weather the acoustic input includes at least a voice command", over the prior art reference Mozer. The argued features are written such that they still read upon the cited prior art reference as taught and described from the paragraphs [32], [36]-[38], Figure 5, [55] and Figure 8 paragraphs [66]-[67].

Applicant recordation instructions: It is not necessary for applicant to provide a separate record of the substance of interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HUY C HO/ Primary Examiner, Art Unit 2644	

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number	Filing Date	☐ To be Mailed
15/238,238	08/16/2016	

ENTITY: ☒ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*	X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*	X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT **06/18/2018**

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	* 20	Minus	** 20	= 0	X $100 =	0
Independent (37 CFR 1.16(h))	* 3	Minus	***3	= 0	X $460 =	0
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	0

AMENDMENT

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

LIE
MINNIE JACKSON

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

REMARKS

In response to the Office Action dated January 18, 2018 and the Advisory Action dated June 5, 2018, Assignee respectfully requests reconsideration in view of the amendments and the following remarks. Claims 1-20 were previously pending in this application. By this amendment, claims 1, 11, 15, and 19 are amended. No claims are added or canceled. As a result, claims 1-20 are pending for examination with claims 1 and 15 being independent claims. No new matter has been added.

Telephone Interview Summary

The Assignee's representatives, Thomas S. Chlebeck and Andrew J. Tibbetts, thank Examiner Ho for his courtesy extended in granting and conducting a telephone interview that took place in two telephone calls on May 29 and May 30, 2018. The substance of the interview is summarized in the discussion below.

In brief, during the interview, Assignee's representatives discussed with the Examiner the amendments and arguments set out in the Amendment After Final filed on April 18, 2018. The Examiner appeared to agree that the amendments suggested a distinction between the claim and the references, but believed that the precise wording of the amendments could be changed to better capture that distinction. The Assignee's representatives discussed changes to the wording of the claims to streamline and simplify the language as well as more clearly indicate a relationship between elements of the claims, which the Examiner indicated may help advance prosecution. Amendments of the type discussed with the Examiner are made to the claims herein.

Double Patenting Rejection

The Office Action rejects claim 1 for purportedly claiming the same invention as that of claims 1 and 5 of prior U.S. Patent No. 9,361,885. While the Assignee does not agree, a Terminal Disclaimer was filed April 18, 2018 to address the Examiner's concern. Accordingly, withdrawal of this rejection is respectfully requested.

Rejections Under 35 U.S.C. §103

The Office Action rejects independent claims 1 and 15 under 35 U.S.C. §103(a) as purportedly being unpatentable over Freeman (U.S. Patent No. 8,977,255) in view of Mozer (U.S. Patent Publication No. 2013/0080167). The Assignee respectfully requests reconsideration. While the Assignee believes that the claims, as previously presented, patentably distinguish over the alleged combination of Freeman and Mozer, the independent claims are amended herein to streamline and simplify claim language and highlight distinctions from the proposed combination.

A. Independent Claim 1 Patentably Distinguishes Over Any Combination of the Cited References.

Claim 1 is amended herein to recite that the at least one processor is configured to "perform at least one processing stage" that includes "in response to detecting ***while in the low power mode*** that the acoustic input includes an explicit voice trigger, ***requesting, via a network while in the low power mode, that at least one server perform recognition*** on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (emphasis added).

Neither Freeman nor Mozer would have taught or suggested the above-quoted limitations if claim 1.

For example, neither Freeman nor Mozer, nor any combination of the references, would have taught or suggested performing both "detecting … that the acoustic input includes an explicit voice trigger" ***and*** "requesting, via a network … that at least one server perform recognition" while "in [a] low power mode," as recited by amended claim 1. At best, Mozer describes that its "Stage 2" analysis may include interacting with a server, but Mozer would not have taught or suggested that this interaction with the server occurs in a low power mode (Mozer, ¶ 31). Accordingly, no combination of the cited references would have taught or suggested such an interaction with the server in a low power mode, and thus the combination would not have taught or suggested the quoted limitations of claim 1.

As another reason that the amended claim distinguishes over any combination of the cited references, no such combination would have taught or suggested "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command." The Office

Action cites to Mozer in alleging that the combination would have taught or suggested limitations of previously-pending claim 1 regarding such a "voice command" (Office Action, pp. 10-11). However, the system of Mozer does not request anything (including either its "stage 1" or its "stage 2" recognizers) to determine whether acoustic input include a voice command. Mozer describes a "stage 1" recognizer that seeks to count keywords in a conversation between users and to classify those keywords (Mozer, ¶ 33, 42-44). When enough keywords of the same category are counted, the stage 1 recognizer may signal to the stage 2 recognizer to begin performing more sophisticated recognition techniques to assess whether the users are discussing a topic to which the system may contribute (Mozer, ¶ 33, 42-44). Mozer's discussion of identifying a topic of received speech would not have taught or suggested requesting that a server determine whether acoustic input includes at least one voice command. The Office Action does not allege that Freeman includes any teaching or suggestion regarding such recognition of a voice command. Accordingly, even if the references were combined in the manner proposed by the Office Action, no such combination would have taught or suggested the quoted limitations of claim 1.

For at least these reasons, claim 1, and each claim that depends therefrom, patentably distinguishes over the proposed combination of the cited references. Withdrawal of the rejections of claim 1 and its dependents under 35 U.S.C. §103 is respectfully requested.

B. Independent Claim 1 Patentably Distinguishes Over Any Combination of the Cited References.

Claim 15 is amended to recite "performing at least one processing stage on the acoustic input" that includes "in response to detecting ***while in the low power mode*** that the acoustic input includes an explicit voice trigger, ***requesting, via a network while in the low power mode, that at least one server perform recognition*** on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (emphasis added).

For reasons that should be appreciated from the foregoing, no combination of the cited references would have taught or suggested the quoted language of claim 15. For at least these reasons, claim 15, and each claim that depends therefrom, patentably distinguishes over the proposed combination of the cited references. Withdrawal of the rejections of claim 15 and its dependents under 35 U.S.C. §103 is respectfully requested.

General Comments on Dependent Claims

Because each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Assignee believes that it is unnecessary at this time to argue the further distinguishing features of all of the dependent claims. However, the Assignee does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does the Assignee concur that the basis for the rejection of any of the dependent claims is proper. Therefore, the Assignee reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

CONCLUSION

In view of the foregoing amendments and remarks, reconsideration is respectfully requested. This application should now be in condition for allowance; a notice to this effect is respectfully requested. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the Assignee's attorney at the telephone number listed below.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, Assignee hereby requests any necessary extension of time. Assignee believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Dated: 2018-06-18

Respectfully submitted,

By /Thomas S. Chlebeck/
Thomas S. Chlebeck
Registration No.: 70,616
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [X] Fee Transmittal Form
  - [ ] Fee Attached
- [X] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [X] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [X] Other Enclosure(s) (please Identify below):

  Request for Continued Examination

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Thomas S. Chlebeck/		
Printed name	Thomas S. Chlebeck		
Date	2018-06-18	Reg. No.	70,616

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: June 18, 2018 Electronic Signature for: /Eileen M. MacKenzie/

6391451.1

Electronic Acknowledgement Receipt

EFS ID:	32930715
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Thomas Stephen Chlebeck/Eileen MacKenzie
Filer Authorized By:	Thomas Stephen Chlebeck
Attorney Docket Number:	N0484.71090US04
Receipt Date:	18-JUN-2018
Filing Date:	16-AUG-2016
Time Stamp:	20:44:25
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$1900
RAM confirmation Number	061918INTEFSW20454300
Deposit Account	232825
Authorized User	Wolf Greenfield

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-TSC.pdf	25318 f2d06c7ceba75daed44dfd7b7271bb95d801a161	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-TSC.pdf	32194 79c132276e1b98631cae07347ca533d4ac0c9c1b	no	1
Warnings:					
Information:					
3	Extension of Time	N048471090US04-EXT-TSC.pdf	23584 0ff440276de222be90d4551a6e256acb38341fa1	no	1
Warnings:					
Information:					
4	Request for Continued Examination (RCE)	N048471090US04-RCE-TSC.pdf	27368 1e00d343732dba7c8eb7c856906a56a24a51e6f1	no	1
Warnings:					
This is not a USPTO supplied RCE SB30 form.					
Information:					
5		N048471090US04-AMN-TSC.pdf	52991 a49f0a11f00d27888fa7d863438aafcb1bd7302a	yes	11

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Amendment Submitted/Entered with Filing of CPA/RCE	1	1
	Claims	2	6

	Applicant Arguments/Remarks Made in an Amendment		7	11	
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	32905 fa5249bce754c9d014bf82a1469c7d03441ab8b0	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	194360		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Examiner Name	H. C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

[ ] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)1,900.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[X] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[ ] Charge fee(s) indicated below

[ ] Charge fee(s) indicated below, **except for the filing fee**

[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

[ ] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

[X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
20 - 20 or HP =	x	=	

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
2 - 3 or HP =	x	=	

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
- 100 =	/50 =	(round **up** to a whole number) x		=

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1252 Extension for response within second month	600.00
1801 Request for continued examination (RCE) - 1st ...	1,300.00

SUBMITTED BY

Signature	/Thomas S. Chlebeck/	Registration No. (Attorney/Agent)	70,616	Telephone	617.646.8000
Name (Print/Type)	Thomas S. Chlebeck			Date	2018-06-18

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: June 18, 2018 Electronic Signature for: /Eileen M. MacKenzie/

6391450.1

PTO/AIA/22 (03-13)
Approved for use through 7/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) N0484.71090US04
Application Number 15/238,238-Conf. #5864	Filed August 16, 2016
For METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	
Art Unit 2644	Examiner H. C. Ho

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	Micro Entity Fee	
☐ One month (37 CFR 1.17(a)(1))	$200	$100	$50	$
☒ Two months (37 CFR 1.17(a)(2))	$600	$300	$150	$ 600.00
☐ Three months (37 CFR 1.17(a)(3))	$1,400	$700	$350	$
☐ Four months (37 CFR 1.17(a)(4))	$2,200	$1,100	$550	$
☐ Five months (37 CFR 1.17(a)(5))	$3,000	$1,500	$750	$

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

☐ A check in the amount of the fee is enclosed.

☒ Payment by credit card. ~~Form PTO-2038 is attached~~.

☐ The Director has already been authorized to charge fees in this application to a Deposit Account.

☒ The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 23/2825 .

☒ Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

☐ applicant.

☐ attorney or agent of record. Registration number _____.

☒ attorney or agent acting under 37 CFR 1.34. Registration number 70,616 .

/Thomas S. Chlebeck/ — Signature
2018-06-18 — Date

Thomas S. Chlebeck — Typed or printed name
617.646.8000 — Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

☐ * Total of 1 forms are submitted.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: June 18, 2018 Electronic Signature for: /Eileen M. MacKenzie/

6391453.1

AMENDMENTS TO THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims:

1. (Currently amended) A device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and

at least one processor configured to perform at least one processing stage on the acoustic input, the at least one processing stage including:

in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command, ~~while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger, and the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input,~~

wherein the device is configured to respond to voice commands recognized by the at least one server.

2. (Previously presented) The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.

3. (Previously presented) The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

4. (Previously presented) The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

5. (Previously presented) The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

6. (Previously presented) The device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger.

7. (Previously presented) The device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

8. (Previously presented) The device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

9. (Previously presented) The device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. (Previously presented) The device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least

one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

11. (Currently amended) The device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the low power ~~idle~~ mode, and wherein the explicit voice trigger is a specific word or phrase that, when detected, is sufficient to engage the device.

12. (Previously presented) The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. (Previously presented) The device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

14. (Previously presented) The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

15. (Currently amended) At least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:

performing[[,]] at least one processing stage on the acoustic input, the at least one processing stage including:

in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command, ~~while in the low power mode, at least one processing stage on the acoustic input to detect whether the acoustic input includes an explicit voice trigger, the at least one processing stage, upon detecting that the acoustic input includes the explicit voice trigger, transmitting at least a portion of the acoustic input to at least one server via a network for processing by the at least one server to recognize a voice command included in the acoustic input~~; and

responding to voice commands recognized by the at least one server.

16. (Original) The at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

17. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger.

18. (Original) The at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. (Currently amended) The at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the low power ~~idle~~ mode.

20. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue

selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Thomas Stephen Chlebeck/Eileen MacKenzie
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 2 months with $0 paid	1252	1	600	600
Miscellaneous:				
RCE- 1ST REQUEST	1801	1	1300	1300
		Total in USD ($)		**1900**

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 CFR §1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: June 18, 2018 Electronic Signature for: /Eileen M. MacKenzie/

AMENDMENT FILED WITH REQUEST FOR CONTINUED EXAMINATION

Mail Stop RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Final Office Action dated January 18, 2018 and the Advisory Action dated June 5, 2018, please enter the following in the above-identified U.S. patent application:

Amendments to the Claims begin on page 2 of this paper.

Remarks/Arguments begin on page 7 of this paper.

Request for Continued Examination (RCE) Transmittal		
Address to: Mail Stop RCE Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450	Application Number	15/238,238-Conf. #5864
	Filing Date	August 16, 2016
	First Named Inventor	William F. Ganong, III
	Art Unit	2644
	Examiner Name	H. C. Ho
	Attorney Docket Number	N0484.71090US04

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.

Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, to any international application that does not comply with the requirements of 35 U.S.C 371, or to any design application. See Instruction Sheet for RCEs (not to be submitted to the USPTO on page 2.)

1. Submission required under 37 CFR 1.114 Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).
   - a. [ ] Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.
     - i. [ ] Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
     - ii. [ ] Other _____
   - b. [x] Enclosed
     - i. [x] Amendment/Reply
     - ii. [ ] Affidavit(s)/ Declaration(s)
     - iii. [ ] Information Disclosure Statement (IDS)
     - iv. [x] Other Extension of Time; Fee Transmittal
2. Miscellaneous
   - a. [ ] Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)
   - b. [ ] Other _____
3. Fees The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.
   - a. [ ] The Director is hereby authorized to charge the following fees, any underpayment of fees, or credit any overpayments, to Deposit Account No. 23/2825 .
     - i. [x] RCE fee required under 37 CFR 1.17(e)
     - ii. [x] Extension of time fee (37 CFR 1.136 and 1.17)
     - iii. [ ] Other _____
   - b. [ ] Check in the amount of $ _____ enclosed
   - c. [x] Payment by credit card (Form PTO-2038 enclosed)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED			
Signature	/Thomas S. Chlebeck/	Date	2018-06-18
Name (Print/Type)	Thomas S. Chlebeck	Registration No.	70,616

~~***CERTIFICATE OF MAILING OR TRANSMISSION***~~			
~~I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date shown below.~~			
~~Signature~~			
~~Name (Print/Type)~~		~~Date~~	

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: June 18, 2018 Electronic Signature for: /Eileen M. MacKenzie/

6391452.1

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/03/18 21:04
S2	104	(("GANONG") near3 ("William")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S3	16	(("VAN") near3 ("MULBREGT") near3 ("Paul")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S4	38	(("SEJNOHA") near3 ("Vladimir")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S5	33	(("WILSON") near3 ("Glen")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:20
S6	220	(("NUANCE") near3 ("COMMUNICATIONS") near3 ("INC")).AANM.	USPAT	OR	OFF	2017/04/20 10:20
S7	43	("8977255" | "9112984" | "9361885" | "20070129949" | "20120329389" | "7574361" | "20100124896" | "20030040339" | "20100289994" | "20140274211" | "20140278435" | "8548176" | "20100185448" | "20110054899" | "20130080167" | "20130080171" | "20130289994" | "20130339028" | "20140012573" | "20140012586" | "5428707" | "5774859" | "5802305" | "5983186" | "6070140" | "6092043" | "6397186" | "6408396" | "6411926" | "6449496" | "6594630" | "6756700" | "6941265" | "6965786" | "7114090" | "7567827" | "7720683" | "7774204" | "8056070" | "8181046" | "8190420" | "8285545" | "8666751").PN.	US-PGPUB; USPAT	OR	OFF	2017/04/20 11:21
S8	5	"20120034904"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:10
S9	9	"9361885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:32
S10	33784	(379/88.01-88.02|G10L15/20|G10L15/22|G10L15/24|G10L15/25|G10L2015/223|G10L2015/226|H04M2250/74|H04W52/0251|H04W52/0254|).ccls.cpc.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:11
S11	1	("15".src. and "238238".ap.) or "15238238".RLAN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/04/20 16:12
S12	25130	S10 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT;	OR	ON	2017/04/20 16:13

			USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB			
S13	4617	S12 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S14	2692	S13 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S15	563	S14 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S16	362	S15 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S17	17	S16 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:33
S18	9564	(704/246-255).ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S19	8758	S18 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S20	1733	S19 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S21	1338	S20 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS;	OR	ON	2017/04/20 16:36

			EPO; JPO; DERWENT; IBM_TDB			
S22	212	S21 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S23	104	S22 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S24	3	S23 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S25	3	"9646610"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:41
S26	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/09/20 13:42
S27	8	"20100121636"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:42
S28	5	"20120035924"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:43
S29	6	"20120127072"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:45
S30	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/03/09 10:35
S31	12	"20140012573"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	ON	2018/03/09 10:53

			DERWENT; IBM_TDB			
S32	10	"20100009719"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2018/03/09 10:55
S33	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/05/30 09:37

8/23/2018 11:42:32 AM
C:\Users\hho3\Documents\EAST\Workspaces\15238238.wsp

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15238238	GANONG ET AL.
	Examiner	**Art Unit**
	HUY C HO	2644

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	04/20/2017	01/10/2018	08/20/2018						
	1	✓	✓	✓						
	2	✓	✓	✓						
	3	✓	✓	✓						
	4	✓	✓	✓						
	5	✓	✓	✓						
	6	✓	✓	✓						
	7	✓	✓	✓						
	8	✓	✓	✓						
	9	✓	✓	✓						
	10	✓	✓	✓						
	11	✓	✓	✓						
	12	✓	✓	✓						
	13	✓	✓	✓						
	14	✓	✓	✓						
	15	✓	✓	✓						
	16	✓	✓	✓						
	17	✓	✓	✓						
	18	✓	✓	✓						
	19	✓	✓	✓						
	20	✓	✓	✓						

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15238238	GANONG ET AL.
	Examiner	Art Unit
	HUY C HO	2644

CPC- SEARCHED		
Symbol	Date	Examiner
G10L15/20; G10L15/22; G10L15/24; G10L15/25; G10L2015/223; G10L2015/226; H04M2250/74; H04W52/0251; H04W52/0254.	4/20/2017	HH

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner
379	88.01	4/20/2017	HH

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

SEARCH NOTES		
Search Notes	Date	Examiner
EAST Database Search	4/20/2017	HH
Inventor Search	4/20/2017	HH
Search update	1/10/2018	HH
Search update	8/20/2018	HH

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner
379	88.02	4/20/2017	HH
704	246-255	4/20/2017	HH

Bibliographic Data

Application No: 15/238,238

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Huy C. Ho/ (Examiner's Signature) | HH (Initials)

Title: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
08/16/2016	455	2644	N0484.71090US04
RULE			

APPLICANTS

Nuance Communications, Inc., Burlington, MA,

INVENTORS

William F. Ganong III, Brookline, MA, UNITED STATES

Paul Adrian Van Mulbregt Wayland, MA, UNITED STATES

Vladimir Sejnoha Lexington, MA, UNITED STATES

Glen Edward Wilson Maynard, MA, UNITED STATES

CONTINUING DATA

This application is a CON of 13795933 03/12/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

08/25/2016

STATE OR COUNTRY

UNITED STATES

ADDRESS

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

FILING FEE RECEIVED

$1,600

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 08/29/2018

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
HO, HUY C

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
08/29/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Patents_eOfficeAction@wolfgreenfield.com
N0484_eOfficeAction@WolfGreenfield.com
IP.Inbox@nuance.com

Office Action Summary	Application No. 15/238,238	Applicant(s) GANONG ET AL.	
	Examiner HUY C. HO	Art Unit 2644	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *06/18/2018*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-20* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-20* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *08/16/2016* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Continued Examination Under 37 CFR 1.114

2. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 06/18/2018 has been entered.

Response to Arguments

3. Applicant's arguments filed 06/18/2018 have been fully considered but they are moot because the claims have been amended, and the arguments are not persuasive for the following reasons.

With respect to the arguments that neither Freeman nor Mozer teach or suggest the argued features, i.e., "perform at least one processing stage" that includes "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command", the examiner respectfully disagrees because:

Mozer teaches at stage 1 recognizer recognizes keywords at low power mode (*Mozer, Figure 6, Box 602, Paragraph(s) [33], [66]*). Once the keywords are recognized, if further details are needed, then stage 1 necessitates to stage 2 to recognize more detailed information in the acoustic input that has the keywords. Stage 1 requests further recognition in stage 2 while the device is still in the low power mode (*Mozer, Figure 6, Box 602, Paragraph(s) [33], [66]; Figure 6, Box 604, Para(s) [67]-[68])*. Stage 2 may be located in a server *(Mozer, Figure 8, device 802-2, stage 2 recognizer 104, Para(s) [64])*.

As such, the argued features/limitations were written such that they read upon the cited references as explained above.

Claim Rejections - 35 USC § 103

4. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

5. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

7. Claims 1-12 and 15-19 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in view of Mozer (U.S. Patent Publication No. 20103/0080167).

Regarding claim 1, (Currently Amended) Freeman teaches a device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to perform at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512),*

wherein the device is configured to respond to the voice command (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520).*

Freeman does not teach "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a

portion of the acoustic input to determine whether the acoustic input includes at least one voice command".

Mozer teaches "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger (*Mozer, Figure 6, Paragraph(s) [33], [66])*, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 6, Paragraph(s) [33], [67]-[68]: the device is still in low power mode of stage 1 transition to stage 2 to request for further recognition of detailed information from the acoustic input; Figure 8, Paragraph(s) [64]: the stage 2 located in a server).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 15, (Currently Amended) Freeman teaches at least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode (*Freeman, the Abstract, Figure 1, Col. 1 Lines 60-65*), the method comprising:

performing, at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*; and

responding to voice commands (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520)*;

responding to the voice command (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520)*.

Freeman does not teach "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command".

Mozer teaches "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger (*Mozer, Figure 6, Paragraph(s) [33], [66]*), requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 6, Paragraph(s) [33], [67]-[68]: the device is still in low power mode of stage 1 transition to stage 2 to request for further recognition of detailed information from the acoustic input; Figure 8, Paragraph(s) [64]: the stage 2 located in a server).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 2, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the at

least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 3, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 4, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor (*Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 5, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis

on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech (*Freeman, Col. 4 Lines 45-62, Col. 5 Lines 13-33)*.

Regarding claim 6, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger (*Mozer, Paragraph(s) [33]*).

Regarding claim 7, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command (*Mozer, Paragraph(s) [36]*).

Regarding claim 8, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the

second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger (*Mozer, Paragraph(s) [33], [36]*).

Regarding claim 9, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Mozer, Paragraph(s) [26]-[28], [46]-[47]*).

Regarding claim 10, (Previously Presented) Freeman, as modified by Mozer, teaches the mobile device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 5, boxes 512*).

Regarding claim 11, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 1, wherein at least one task specified by the voice command is performed with

the device remaining in the low power mode, and wherein the explicit voice trigger is a specific word or phrase that, when detected, is sufficient to engage the device (*Mozer, Paragraph(s) [65]-[66]: specific words such as "vacation", "travel", etc.).*

Regarding claim 12, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input (*Freeman, Figure 5, box 506).*

Regarding claim 16, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Mozer, Figure 1B, Paragraph(s) [31]-[33], [36]).*

Regarding claim 17, (Previously Presented) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on

the acoustic input to detect whether the acoustic input includes the explicit voice trigger (*Mozer, Paragraph(s) [33]*).

Regarding claim 18, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Freeman, Figure 5, boxes 504, 506, 508*).

Regarding claim 19, (Currently Amended) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the low power mode (*Freeman, Figure 5, boxes 506, 508, 510*).

8. Claims 13, 14 and 20 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in views of Mozer (U.S. Patent Publication No. 20103/0080167) and LeBeau et al. (U.S. Patent Publication No. 2012/0034904).

Regarding claim 13, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the mobile device, an occurrence of an event on the device, recent activity on the mobile device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43])*.

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source,

etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 14, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43])*.

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 20, (Previously Presented) Freeman, as modified by Mozer, does not teach the at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43])*.

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUY C. HO whose telephone number is (571)270-1108. The examiner can normally be reached on Monday - Friday, 8:00 a.m. - 5:00 p.m., EST.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Kathy Wang-Hurst can be reached on 571-270-5371. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on

access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HUY C HO/
Primary Examiner, Art Unit 2644

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: November 29, 2018 Signature: /Anna K. Oshiro/ (Anna K. Oshiro)

AMENDMENT IN RESPONSE TO NON-FINAL OFFICE ACTION

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Office Action dated August 29, 2018, please amend the above-identified U.S. patent application as follows:

Amendments to the Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 7 of this paper.

Electronic Acknowledgement Receipt

EFS ID:	34438562
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Andrew J. Tibbetts/Anna K. Oshiro
Filer Authorized By:	Andrew J. Tibbetts
Attorney Docket Number:	N0484.71090US04
Receipt Date:	29-NOV-2018
Filing Date:	16-AUG-2016
Time Stamp:	16:08:19
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-AJT.pdf	25956 1ad7e227b98caebb324e5f8169b701ec1d96da12	no	1

Warnings:

Information:					
2		N048471090US04-AMN-AJT.pdf	72958 eb6c4c42b07d72942758cf2aa4f3db415344687e	yes	13

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Amendment/Req. Reconsideration-After Non-Final Reject	1	1
Claims	2	6
Applicant Arguments/Remarks Made in an Amendment	7	13

Warnings:	
Information:	
Total Files Size (in bytes):	98914

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [x] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [ ] Other Enclosure(s) (please Identify below):

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Andrew J. Tibbetts/		
Printed name	Andrew J. Tibbetts		
Date	November 29, 2018	Reg. No.	65,139

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: November 29, 2018 Electronic Signature for Anna K. Oshiro: /Anna K. Oshiro/

REMARKS

In response to the Office Action dated August 29, 2018, Applicant respectfully requests reconsideration. Each of the rejections set forth in the Office Action has been carefully considered and is addressed below.

I. Claim Rejections under 35 U.S.C. §103

Independent claims 1 and 15 were rejected under 35 U.S.C. §103 as purportedly being obvious over U.S. Patent No. 8,977,255 ("Freeman") in view of U.S. Patent Application Publication No. 2013/0080167 ("Mozer"). Applicant respectfully requests reconsideration.

A. Discussion of Mozer

Mozer describes a two-stage speech recognition system in which a stage 1 recognizer classifies keywords recognized in an acoustic input signal into pre-defined classifications (¶0033). "Pre-defined classifications may be topics that describe different areas of interest, such as travel, purchases, entertainment, research, food, or electronics" (¶0033). The stage 1 recognizer "classifies recognized keywords into classifications, and when one classification has enough keywords classified with it, the stage 1 recognizer 102 may trigger [a] stage 2 recognizer" (¶0033). For example, the stage 1 recognizer may classify the words "vacation" and "flight" under a classification of "travel" (¶0066). The stage 1 recognizer then selects the "travel" classification and activates the stage 2 recognizer (¶0066).

After being activated, the stage 2 recognizer "may output a response to" a detected meaning in an acoustic signal (¶0038). The stage 2 recognizer may decipher sentences in a conversation between two people, and output information to the device based on the deciphered sentences (¶0067). For example, after the stage 1 recognizer determines a classification of "travel," the stage 2 recognizer may decipher a sentence "maybe we should do an activity in Ireland," and output pictures of Ireland to the device (¶0067).

B. Claim 1 Patentably Distinguishes Over Freeman and Mozer

No combination of Freeman and Mozer would have taught or suggested "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting … that at least one server perform recognition on at least a portion of the acoustic input," as recited in claim 1.

The Office Action asserts that Figure 6, ¶0033, and ¶¶0066-68 of Mozer describe the above-quoted language of claim 1 (see Office Action p. 6). The cited passage of Mozer describes that Mozer's "stage 1" recognizer triggers Mozer's "stage 2" recognizer in response to the stage 1 recognizer classifying enough words into a particular classification and thereby determining that the input speech corresponds to a particular topic (¶0033). By citing to Mozer's functionality for initiating recognition by the "stage 2" recognizer in response to this topic identification, the Office Action suggests that determining a topic of input speech by determining whether a certain number of words fall within a classification can be properly interpreted as "detecting that the acoustic input includes an explicit voice trigger," as recited in claim 1. This is not correct.

One of ordinary skill would understand the plain meaning of "explicit voice trigger" included in an "acoustic input" to be "voice" content that "triggers" an action when that voice content is "explicit[ly]" included in the acoustic input. This plain meaning is consistent with the Specification as filed.

Mozer does not describe such an "explicit voice trigger" because Mozer does not describe that its topic identification through word classification includes any detection of such a "voice" content that "triggers" an action when that voice content is "explicit[ly]" included in acoustic input. Mozer's process included actions taken in response to a post-processing of speech content, by analyzing words and phrases in combination with one another to determine whether more than a threshold number of words relate to a particular topic. No one word or phrase in Mozer, or any other explicit speech content, triggers any action. Determining a topic of speech input based on whether a certain number of words fall within a classification does not teach or suggest detecting whether the speech input includes an "explicit voice trigger."

Thus, Mozer does not teach or suggest "in response to ***detecting … that the acoustic input includes an explicit voice trigger***, requesting … that at least one server perform recognition on at least a portion of the acoustic input" as recited in claim 1.

The Office Action does not allege that Freeman teaches or suggests the above-quoted language of claim 1. Accordingly, even if the references would have been combined in the manner alleged by the Office Action (which the Assignee does not concede), no combination of the cited references would have taught or suggested at least the quoted language of claim 1.

For at least this reason, no combination of Mozer and Freeman would have taught or suggested all limitations of independent claim 1. Accordingly, withdrawal of the rejection under § 103 of claim 1 and the claims that depend therefrom is respectfully requested.

C. Claim 15 Patentably Distinguishes Over Mozer and Freeman

No combination of Mozer and Freeman would have taught or suggested "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, the explicit voice trigger comprising a predetermined phrase or word that triggers voice response processing: requesting…that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" as recited in claim 15.

i. The combination would not have taught or suggested "in response to detecting…that the acoustic input includes an explicit voice trigger comprising a predetermined phrase or word that triggers voice response processing: requesting…that at least one server perform recognition."

As should be appreciated from the foregoing discussion, the cited passages of Mozer do not describe detecting whether an acoustic input includes an "explicit voice trigger" that comprises "a predetermined phrase or word that triggers voice response processing," as recited by amended claim 15.

The Office Action does not allege that Freeman teaches or suggests the above-quoted language of claim 15. Accordingly, even if the references would have been combined in the manner

alleged by the Office Action (which the Assignee does not concede), no combination of the cited references would have taught or suggested at least the quoted language of claim 15.

ii. The combination would not have taught or suggested "requesting … that at least one server perform speech recognition…to determine whether the acoustic input includes at least one voice command."

The Office Action asserts that Figure 6, ¶0033, ¶0064, and ¶¶0067-68 of Mozer describe the above-quoted language of claim 15. The cited passage describes that a stage 2 recognizer is triggered by a stage 1 recognizer to interpret sentences in audio of a conversation, and to obtain and provide to the speakers information that might be relevant to the conversation (¶0067). In Mozer, the information is not obtained in response to any user request, but rather is obtained when the system determines, from the speech, that the speakers may be discussing a topic and the system could helpfully contribute by providing information about that topic (¶0067). Mozer gives the example of recognizing a sentence "maybe we should do an activity in Ireland" and, in response, outputting pictures of Ireland or coupons for a soccer game in Ireland to a device (¶0067).

By citing to these passages, the Office Action asserts that interpreting a conversation, and making a determination of whether helpful information could be provided, can be properly interpreted as determining "whether the acoustic input includes at least one voice command," as recited in claim 15. This is not correct. As should be appreciated from the foregoing, in Mozer's system there is no "command" included in the voice content. The speech is simply audio of a conversation, and the conversation is not directed at the system and does not include a request that the system do anything. The system is thus passively monitoring, and makes its own determination of whether an action should be taken. The system is not looking for any "command" included in the voice content.

Thus, the cited passage of Mozer does not teach or suggest "requesting…that at least one server perform speech recognition … ***to determine whether the acoustic input includes at least one voice command***" as recited in claim 15 (emphasis supplied).

The Office Action does not allege that Freeman teaches or suggests the above-quoted language of claim 15. Accordingly, even if the references would have been combined in the manner

alleged by the Office Action (which the Assignee does not concede), no combination of the cited references would have taught or suggested at least the quoted language of claim 15.

iii. Conclusion

For at least these reasons, claim 15 patentably distinguishes over any combination of Mozer and Freeman. Accordingly, withdrawal of the rejection under § 103 of claim 15 and the claims that depend therefrom is respectfully requested.

D. Newly Added Claim 21 Patentably Distinguishes Over Mozer and Freeman

Newly added claim 21 recites, in part, "in response to detecting…that the acoustic input includes an explicit voice trigger, requesting that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command." For reasons that should be appreciated from the foregoing, independent claim 21 patentably distinguishes over Mozer and Freeman.

Additionally, claim 21 recites, in part, "***in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command***, respond to at least one voice command in the acoustic input recognized by the at least one server" (emphasis supplied). Recognizer 2 of Mozer does not determine whether an acoustic input includes a voice command. It follows that recognizer 1 would not receive an indication from recognizer 2 that an acoustic input includes a voice command.

Freeman describes voice activation of a single device without interaction with a separate server. Thus, Freeman does not respond to a voice command "in response to receiving an indication from the at least one server" as recited in claim 21.

For at least these reasons, independent claim 21 patentably distinguishes over Mozer and Freeman.

II. General Comments on Dependent Claims

As each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Applicant believes that arguing the further distinguishing features of all

of the dependent claims is unnecessary at this time. However, Applicant does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does Applicant concur that the basis for the rejection of any of the dependent claims is proper. Therefore, Applicant reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

CONCLUSION

In view of the foregoing amendments and remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call Applicant's representative at the telephone number indicated below to discuss any outstanding issues relating to the allowability of the application.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, Applicant hereby requests any necessary extension of time. If there is a fee occasioned by this response, including an extension fee, please charge any deficiency to Deposit Account No. 23/2825 under Docket No. N0484 .71090US04, from which the undersigned is authorized to draw.

Dated: November 29, 2018

Respectfully submitted,

By /Andrew J. Tibbetts/

Andrew J. Tibbetts, Reg. No. 65,139
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

AMENDMENTS TO THE CLAIMS

Applicant submits below a complete listing of the current claims, including marked-up claims with insertions indicated by underlining and deletions indicated by strikeouts and/or double bracketing. This listing of claims replaces all prior versions, and listings, of claims in the application:

Listing of the Claims

1. (Previously presented) A device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and

at least one processor configured to perform at least one processing stage on the acoustic input, the at least one processing stage including:

in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command,

wherein the device is configured to respond to voice commands recognized by the at least one server.

2. (Previously presented) The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.

3. (Previously presented) The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

4. (Previously presented) The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

5. (Previously presented) The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

6. (Previously presented) The device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger.

7. (Previously presented) The device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

8. (Previously presented) The device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.

9. (Previously presented) The device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. (Previously presented) The device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

11. (Previously presented) The device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the low power mode, and wherein the explicit voice trigger is a specific word or phrase that, when detected, is sufficient to engage the device.

12. (Previously presented) The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. (Previously presented) The device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

14. (Previously presented) The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

15. (Currently Amended) At least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:

performing at least one processing stage on the acoustic input, the at least one processing stage including:

in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger comprising a predetermined word or phrase that triggers voice response processing: [[,]]

requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command; and

responding to voice commands recognized by the at least one server.

16. (Original) The at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

17. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger.

18. (Original) The at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the low power mode.

20. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

21. (New) A device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and

at least one processor configured to:

in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command; and

in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server.

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number	Filing Date	☐ To be Mailed
15/238,238	08/16/2016	

ENTITY: ☒ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*	X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*	X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT **11/29/2018**

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	* 21	Minus	** 20	= 1	X $100 =	100
Independent (37 CFR 1.16(h))	* 3	Minus	***3	= 0	X $460 =	0
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	**100**

AMENDMENT

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

LIE
ROSA WEST

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	04/20/2017	01/10/2018	08/20/2018	02/24/2019					
	1	✓	✓	✓	✓					
	2	✓	✓	✓	✓					
	3	✓	✓	✓	✓					
	4	✓	✓	✓	✓					
	5	✓	✓	✓	✓					
	6	✓	✓	✓	✓					
	7	✓	✓	✓	✓					
	8	✓	✓	✓	✓					
	9	✓	✓	✓	✓					
	10	✓	✓	✓	✓					
	11	✓	✓	✓	✓					
	12	✓	✓	✓	✓					
	13	✓	✓	✓	✓					
	14	✓	✓	✓	✓					
	15	✓	✓	✓	✓					
	16	✓	✓	✓	✓					
	17	✓	✓	✓	✓					
	18	✓	✓	✓	✓					
	19	✓	✓	✓	✓					
	20	✓	✓	✓	✓					
	21				✓					

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

CPC - Searched*		
Symbol	Date	Examiner
G10L15/20; G10L15/22; G10L15/24; G10L15/25; G10L2015/223; G10L2015/226; H04M2250/74; H04W52/0251; H04W52/0254.	4/20/2017	HH

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner
379	88.01	4/20/2017	HH

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
EAST Database Search	4/20/2017	HH
Inventor Search	4/20/2017	HH
Search update	1/10/2018	HH
Search update	8/20/2018	HH
Search update	02/24/2019	HH

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner
379	88.02	4/20/2017	HH
704	246-255	4/20/2017	HH

/HUY C HO/ Primary Examiner.Art Unit 2644	

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 02/28/2019

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
02/28/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

IP.Inbox@nuance.com
N0484_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@wolfgreenfield.com

Office Action Summary	Application No. 15/238,238	Applicant(s) Ganong et al.	
	Examiner HUY C HO	Art Unit 2644	AIA Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 29 November 2018.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☑ This action is **FINAL.** 2b) ☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-21 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1-21 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 16 August 2016 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Response to Arguments

2. Applicant's arguments filed 11/29/2018 have been fully considered but they are not persuasive for the following reasons.

With respect to the arguments that no combination of Freeman and Mozer would have taught or suggested "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting ... that at least one server perform recognition on at least a portion of the acoustic input," as recited in claims 1, 15 and new claim 21.

The examiner respectfully disagrees because:

First of all, the main reference Freeman teaches a portable electronic device operates in a low power mode for detecting an audio input and checking for any possible voice command in the audio input (Freeman, Figure 5, boxes 504, 506, 508 and 510). Freeman's voice command could read upon the claimed "explicit voice trigger" because in a broadest reasonable interpretation for "explicit voice trigger", the voice trigger can be meant as a word or a phrase that provides as a trigger or an initiator,

or an ignitor for a next action. Meanwhile, the voice command in the audio input taught by Freeman could be a word, a phrase that provides an initiating word/phrase or a triggering word/phrase to ignite or trigger for next actions. Thus, Freeman teaches and suggests the argued features, i.e., “explicit voice trigger” in claims 1, 15 and new claim 21.

However, for the sake of the explicit and clear rejection in the office action, reference Mozer was used for this purpose. In fact, Mozer teaches a device comprises speech recognizers as stage 1 and stage 2 recognizers (Mozer, Figure 1B). The stage 1 recognizer is designed to recognize a limited set of keywords that may trigger for a next action or a next stage. The device uses stage 1 recognizer to conserve batter energy at low power mode (Mozer, pp [32]). The keywords read on the “explicit voice trigger” because the keywords are recognized by a stage 1 speech recognizer (Mozer, Figure 1B). The keywords are words/phrases that trigger for the next action as shown in stage 2 recognizer (Mozer, Figure 4, pp [33]). Stage 2 recognizer is triggered by keywords from stage 1 recognizer to receive an acoustic input signal and to determine further information from the acoustic input signal and to output responses per the determination (Mozer, Figure 5, pp [36]-[40]). Besides, the stage 2 recognizer can be located on a server (Mozer, pp [31]). Therefore, Mozer discloses the argued features/limitations, i.e., “in response to

detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting ... that at least one server perform recognition on at least a portion of the acoustic input," as recited in claims 1, 15 and new claim 21.

In conclusion, the argued features/limitations were written such that they read upon the cited references as explained above.

Claim Rejections - 35 USC § 103

3. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

4. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the

contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

6. Claims 1-12, 15-19 and 21 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in view of Mozer (U.S. Patent Publication No. 20103/0080167).

Regarding claim 1, (Previously Presented) Freeman teaches a device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to perform at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512),*

wherein the device is configured to respond to the voice command (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520).*

Freeman does not teach "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command".

Mozer teaches "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger (*Mozer, Figures 4 and 6, Paragraph(s) [32]-[33], [66]*), requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 6, Paragraph(s) [33], [67]-[68]: the device is still in low power mode of stage 1 transition to stage 2 to request for further recognition of detailed information from the acoustic input; Figure 8, Paragraph(s) [64]: the stage 2 located in a server).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could

be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 15, (Currently Amended) Freeman teaches at least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode (*Freeman, the Abstract, Figure 1, Col. 1 Lines 60-65*), the method comprising:

performing, at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*; and

responding to voice commands (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520)*;

responding to the voice command (*Freeman, Figure 4, boxes 408, 410; Figure 5, boxes 516, 520)*.

Freeman does not teach "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger comprising a predetermined word or phrase that

triggers voice response processing, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command".

Mozer teaches "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger comprising a predetermined word or phrase that triggers voice response processing (*Mozer, Figures 4 and 6, Paragraph(s) [32]-[33], [66])*, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 6, Paragraph(s) [33], [67]-[68]: the device is still in low power mode of stage 1 transition to stage 2 to request for further recognition of detailed information from the acoustic input; Figure 8, Paragraph(s) [64]: the stage 2 located in a server).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by

multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 21, (New) Freeman teaches a device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*.

Freeman does not teach "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command; and in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server".

Mozer teaches "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger" (*Mozer, Figures 4 and 6, Paragraph(s) [32]-[33], [66]*), requesting that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command; and in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server" (*Mozer, Figures 4 and 6, Paragraph(s)[32]- [33], [67]-[68]: the device is still in low power mode of stage 1 transition to stage 2 to request for further recognition of detailed information from the acoustic input; Figure 8, Paragraph(s) [64]: the stage 2 located in a server).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus

providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 2, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 3, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 4, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor (*Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 5, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech (*Freeman, Col. 4 Lines 45-62, Col. 5 Lines 13-33).*

Regarding claim 6, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes the explicit voice trigger (*Mozer, Paragraph(s) [33]).*

Regarding claim 7, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command (*Mozer, Paragraph(s) [36]).*

Regarding claim 8, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger (*Mozer, Paragraph(s) [33], [36]*).

Regarding claim 9, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Mozer, Paragraph(s) [26]-[28], [46]-[47]*).

Regarding claim 10, (Previously Presented) Freeman, as modified by Mozer, teaches the mobile device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 5, boxes 512*).

Regarding claim 11, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein at

least one task specified by the voice command is performed with the device remaining in the low power mode, and wherein the explicit voice trigger is a specific word or phrase that, when detected, is sufficient to engage the device (*Mozer, Paragraph(s) [65]-[66]: specific words such as "vacation", "travel", etc.).*

Regarding claim 12, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input (*Freeman, Figure 5, box 506).*

Regarding claim 16, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Mozer, Figure 1B, Paragraph(s) [31]-[33], [36]).*

Regarding claim 17, (Previously Presented) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on

the acoustic input to detect whether the acoustic input includes the explicit voice trigger (*Mozer, Paragraph(s) [33]*).

Regarding claim 18, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Freeman, Figure 5, boxes 504, 506, 508*).

Regarding claim 19, (Previously Presented) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the low power mode (*Freeman, Figure 5, boxes 506, 508, 510*).

8. Claims 13, 14 and 20 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in views of Mozer (U.S. Patent Publication No. 20103/0080167) and LeBeau et al. (U.S. Patent Publication No. 2012/0034904).

Regarding claim 13, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the mobile device, an occurrence of an event on the device, recent activity on the mobile device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient

sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 14, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to

modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 20, (Previously Presented) Freeman, as modified by Mozer, does not teach the at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches “performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on

the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Conclusion

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS

of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUY C HO whose telephone number is (571)270-1108. The examiner can normally be reached on M-F 8AM-5PM.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner’s supervisor, KATHY WANG-HURST can be reached on (571)270-5371. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval

(PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HUY C HO/
Primary Examiner, Art Unit 2644

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/03/18 21:04
S2	104	(("GANONG") near3 ("William")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S3	16	(("VAN") near3 ("MULBREGT") near3 ("Paul")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S4	38	(("SEJNOHA") near3 ("Vladimir")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S5	33	(("WILSON") near3 ("Glen")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:20
S6	220	(("NUANCE") near3 ("COMMUNICATIONS") near3 ("INC")).AANM.	USPAT	OR	OFF	2017/04/20 10:20
S7	43	("8977255" | "9112984" | "9361885" | "20070129949" | "20120329389" | "7574361" | "20100124896" | "20030040339" | "20100289994" | "20140274211" | "20140278435" | "8548176" | "20100185448" | "20110054899" | "20130080167" | "20130080171" | "20130289994" | "20130339028" | "20140012573" | "20140012586" | "5428707" | "5774859" | "5802305" | "5983186" | "6070140" | "6092043" | "6397186" | "6408396" | "6411926" | "6449496" | "6594630" | "6756700" | "6941265" | "6965786" | "7114090" | "7567827" | "7720683" | "7774204" | "8056070" | "8181046" | "8190420" | "8285545" | "8666751").PN.	US-PGPUB; USPAT	OR	OFF	2017/04/20 11:21
S8	5	"20120034904"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:10
S9	9	"9361885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:32
S10	33784	(379/88.01-88.02|G10L15/20|G10L15/22|G10L15/24|G10L15/25|G10L2015/223|G10L2015/226|H04M2250/74|H04W52/0251|H04W52/0254|).ccls,cpc.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:11
S11	1	("15".src. and "238238".ap.) or "15238238".RLAN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/04/20 16:12
S12	25130	S10 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT;	OR	ON	2017/04/20 16:13

			USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB			
S13	4617	S12 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S14	2692	S13 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S15	563	S14 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S16	362	S15 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S17	17	S16 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:33
S18	9564	(704/246-255).ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S19	8758	S18 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S20	1733	S19 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S21	1338	S20 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS;	OR	ON	2017/04/20 16:36

			EPO; JPO; DERWENT; IBM_TDB			
S22	212	S21 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S23	104	S22 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S24	3	S23 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S25	3	"9646610"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:41
S26	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/09/20 13:42
S27	8	"20100121636"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:42
S28	5	"20120035924"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:43
S29	6	"20120127072"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:45
S30	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/03/09 10:35
S31	12	"20140012573"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	ON	2018/03/09 10:53

			DERWENT; IBM_TDB			
S32	10	"20100009719"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2018/03/09 10:55
S33	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/05/30 09:37

2/ 24/ 2019 5:54:51 PM
C:\ Users\ hho3\ Documents\ EAST\ Workspaces\ 15238238.wsp

Bibliographic Data

Application No: 15/238,238

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Huy C. Ho/ HH

Examiner's Signature Initials

Title: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
08/16/2016	455	2644	N0484.71090US04
RULE			

APPLICANTS

Nuance Communications, Inc., Burlington, MA,

INVENTORS

William F. Ganong III, Brookline, MA, UNITED STATES

Paul Adrian Van Mulbregt Wayland, MA, UNITED STATES

Vladimir Sejnoha Lexington, MA, UNITED STATES

Glen Edward Wilson Maynard, MA, UNITED STATES

CONTINUING DATA

This application is a CON of 13795933 03/12/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

08/25/2016

STATE OR COUNTRY

UNITED STATES

ADDRESS

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

FILING FEE RECEIVED

$1,600

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 04/16/2019

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
04/16/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

IP.Inbox@nuance.com
N0484_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@wolfgreenfield.com

Applicant-Initiated Interview Summary	**Application No.** 15/238,238	**Applicant(s)** Ganong et al.	
	Examiner HUY C HO	**Art Unit** 2644	**AIA (FITF) Status** No

All participants (applicant, applicants representative, PTO personnel):

(1) Huy C. Ho.

(2) Saad Alam (Reg. No. 77,456).

(3) Andrew Tibbetts (Reg. No. 65,139).

(4) _____.

Date of Interview: 11 April 2019.

Type: ☑ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☑ No.
If Yes, brief description: _____.

Issues Discussed ☐101 ☐112 ☐102 ☑103 ☑Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: 1,15 and 21.

Identification of prior art discussed: Mozer.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Discussed limitations/features, i.e., detecting while in a low power mode that the acoustic input includes an explicit voice trigger. Some amendments were discussed and proposed to clarify the languages. However, no agreement is reached at this time.

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HUY C HO/ Primary Examiner, Art Unit 2644	

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record
A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) 1.133 Interviews
Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.
All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiners responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicants correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:
- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:
1) A brief description of the nature of any exhibit shown or any demonstration conducted,-
2) an identification of the claims discussed,
3) an identification of the specific prior art discussed,
4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
5) a brief identification of the general thrust of the principal arguments presented to the examiner,
(The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
6) a general indication of any other pertinent matters discussed, and
7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicants record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiners version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, Interview Record OK on the paper recording the substance of the interview along with the date and the examiners initials.

Electronic Acknowledgement Receipt

EFS ID:	35882970
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Saad Alam/Eileen MacKenzie
Filer Authorized By:	Saad Alam
Attorney Docket Number:	N0484.71090US04
Receipt Date:	30-APR-2019
Filing Date:	16-AUG-2016
Time Stamp:	21:00:57
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$1900
RAM confirmation Number	050119INTEFSW21014500
Deposit Account	232825
Authorized User	Eileen MacKenzie

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-SA.pdf	25932 f74e5b57009a1fb0445a1183696f9b96150e4f24	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-SA.pdf	32476 0f420841f1d2c282204aeaeba753a440aea6fbf5	no	1
Warnings:					
Information:					
3	Request for Continued Examination (RCE)	N048471090US04-RCE-SA.pdf	27929 92a6e28d8b399ff62da31e609d0c98a70c4245e0	no	1
Warnings:					
This is not a USPTO supplied RCE SB30 form.					
Information:					
4		N048471090US04-AMN-SA.pdf	75768 c43ad05ea961c2c2ec5b9afd2446aa1e86085779	yes	15

Multipart Description/PDF files in .zip description

Document Description	Start	End
Applicant Arguments/Remarks Made in an Amendment	7	15
Claims	2	6
Amendment Submitted/Entered with Filing of CPA/RCE	1	1

Warnings:

Information:

5	Fee Worksheet (SB06)	fee-info.pdf	30754 3eb146458a06ad433be42f4879f7dbe3d1c1f9ec	no	2

Warnings:	
Information:	
Total Files Size (in bytes):	192859

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [x] Amendment/Reply
  - [x] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Request for Continued Examination (RCE)

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Saad Alam/		
Printed name	Saad Alam		
Date	April 30, 2019	Reg. No.	77,456

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 30, 2019 Electronic Signature for: /Eileen M. MacKenzie/

Request for Continued Examination (RCE) Transmittal		
Address to: Mail Stop RCE Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450	Application Number	15/238,238-Conf. #5864
	Filing Date	August 16, 2016
	First Named Inventor	William F. Ganong, III
	Art Unit	2644
	Examiner Name	H. C. Ho
	Attorney Docket Number	N0484.71090US04

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.

Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, to any international application that does not comply with the requirements of 35 U.S.C 371, or to any design application. See Instruction Sheet for RCEs (not to be submitted to the USPTO on page 2.)

1. Submission required under 37 CFR 1.114 Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).
   - a. [ ] Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.
     - i. [ ] Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
     - ii. [ ] Other _____
   - b. [X] Enclosed
     - i. [X] Amendment/Reply
     - ii. [ ] Affidavit(s)/ Declaration(s)
     - iii. [ ] Information Disclosure Statement (IDS)
     - iv. [X] Other Fee Transmittal
2. Miscellaneous
   - a. [ ] Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)
   - b. [ ] Other _____
3. Fees The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.
   - a. [ ] The Director is hereby authorized to charge the following fees, any underpayment of fees, or credit any overpayments, to Deposit Account No. 23/2825 .
     - i. [x] RCE fee required under 37 CFR 1.17(e)
     - ii. [ ] Extension of time fee (37 CFR 1.136 and 1.17)
     - iii. [ ] Other _____
   - b. [ ] Check in the amount of $ _____ enclosed
   - c. [x] Payment by credit card ~~(Form PTO-2038 enclosed)~~

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED			
Signature	/Saad Alam/	Date	April 30, 2019
Name (Print/Type)	Saad Alam	Registration No.	77,456

~~***CERTIFICATE OF MAILING OR TRANSMISSION***~~			
~~I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date shown below.~~			
~~Signature~~			
~~Name (Print/Type)~~		~~Date~~	

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 30, 2019 Electronic Signature for: /Eileen M. MacKenzie/

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Examiner Name	H. C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)1,900.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☒ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☐ Charge fee(s) indicated below

☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

☐ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
20 - or HP =	_____	x _____	= _____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - or HP =	_____	x _____	= _____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____	= _____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1820 Request for continued examination (RCE) - 2nd and ...	1,900.00

SUBMITTED BY

Signature	/Saad Alam/	Registration No. (Attorney/Agent)	77,456	Telephone	617.646.8000
Name (Print/Type)	Saad Alam			Date	April 30, 2019

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 30, 2019 Electronic Signature for: /Eileen M. MacKenzie/

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Saad Alam/Eileen MacKenzie
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
RCE- 2ND AND SUBSEQUENT REQUEST	1820	1	1900	1900
		Total in USD ($)		**1900**

REMARKS

In response to the Office Action dated February 28, 2019, Applicant respectfully requests reconsideration. Each of the rejections set forth in the Office Action has been carefully considered and is addressed below. To further prosecution of the application, amendments have been made to the claims herein. The claims as presented are believed to be in condition for allowance.

I. Interview Summary

Applicant's representatives, Saad Alam and Andrew J. Tibbetts, thank Examiner Ho for conducting the interview on April 11, 2019. The substance of that interview is summarized in the discussion below.

During the interview, Applicant's representatives discussed some embodiments of the technology described in connection with FIG. 4 of the application. Next, the discussion turned to the rejection of the claims over the Freeman and Mozer references. In particular, Applicant's representatives explained distinctions of techniques described in the application from those of Mozer, as detailed below. The Examiner agreed that there was a distinction between the technique and the references, but suggested amending the claims to better capture the distinction. Applicant's representatives discussed changes to the claims to better capture the distinction. The Examiner agreed to reconsider the rejection in view of amendments discussed during the telephone interview. Accordingly, the claims are amended herein along the lines discussed during the telephone interview.

II. Claim Rejections under 35 U.S.C. §103

Independent claims 1, 15 and 21 were rejected under 35 U.S.C. §103 as allegedly being obvious over U.S. Patent No. 8,977,255 ("Freeman") in view of U.S. Patent Application Publication No. 2013/0080167 ("Mozer"). Applicant respectfully requests reconsideration in view of the telephone interview, the foregoing amendments, and the following remarks.

A. Discussion of Some Embodiments From The Specification

Some embodiments described in the present application relate to a technique for a device to determine whether an acoustic input includes a voice command, while the device remains in a low power mode (Published Specification, ¶0035). The device determines whether the acoustic input includes a trigger word or phrase (e.g., "Hello Dragon") by (1) performing automatic speech recognition (ASR) on the acoustic input to recognize spoken words or phrases in the acoustic input; and (2) determining whether any of the recognized words or phrases match the trigger word or phrase (¶0087). For example, the device may determine whether any of the phrases recognized in the acoustic input match the trigger phrase "Hello Dragon" (¶0087).

In response to determining that a trigger word phrase was spoken in the acoustic input, the device may (while remaining in the low power mode) invoke more comprehensive ASR to determine whether a voice command is present in the acoustic input (¶0090). To invoke more comprehensive ASR, the device may transmit (while in the low power mode) the acoustic input to a server may perform more computationally expensive ASR to determine whether the acoustic input includes a voice command (¶0091). When the server determines that the acoustic input includes a voice command, the server may notify the device, such that the device responds to the voice command recognized by the server (¶0070).

The foregoing discussion is provided to assist the Examiner in appreciating some applications for various inventive aspects of the present disclosure. However, this discussion may not apply to each of the independent claims, and the language of the independent claims may differ in material respects from the discussion provided above. Thus, Applicant respectfully requests that careful consideration be given to the language of each of the independent claims and that each be addressed on its own merits, without relying on the discussion provided above. In this respect, Applicant does not rely on the discussion provided above to distinguish any of the claims over the cited art. Rather, Applicant relies only upon the language of the claims themselves and the arguments related specifically thereto.

B. Discussion of Mozer

Mozer describes a "background speech recognizer" (Mozer, ¶0020) that passively listens to a conversation between two or more speakers and, after inferring a topic of the conversation, displays information that might be relevant to the speakers (¶0066). Mozer thus describes that a device "may classify keywords that are recognized into pre-defined classifications" that are "topics that describe different areas of interest, such as travel, purchases, entertainment, research, food, or electronics" (Mozer, ¶0033). In Mozer's system, once "a certain number of keywords for a specific classification are detected, then it may be determined that a topic associated with the classification is being discussed" and "***when one classification has enough key words classified with it, then stage 1 recognizer 102 may trigger stage 2 recognizer 104***" (¶0033, emphasis added). As an example, the device selects a classification of "travel" based on the words "vacation" and "flight" in an acoustic input and determines "***that the travel classification should be selected and the stage 2 recognizer 104 should be activated***" (¶0066, emphasis supplied). This "stage 2 recognizer" is a server-based recognizer, which analyzes the conversation based on the identified topic and determines relevant information to display (¶0067).

In particular, Mozer describes that the server "receives the trigger to activate and also may receive information that a conversation is occurring about the classification of 'travel'" (¶0067). "At this point, stage 2 recognizer 104 may take over listening to the conversation" (¶0067). The "stage 2 recognizer 104 may be able to decipher whole sentences and may hear a sentence 'Maybe we should do an activity in Ireland'" and "the classification of 'travel' may be used to determine content for the response" (¶0067). "At this point, a response may be determined that pictures of Ireland should be output with a coupon for a soccer game in Ireland" and "pictures of Ireland may be output to an interface, such as the tablet computer screen" (¶0067).

C. Claim 21 Patentably Distinguishes Over Freeman and Mozer.

Independent claim 21 is amended herein, based on the discussion with the Examiner in the telephone interview, to more clearly differentiate the cited references. Even if one of ordinary skill would have been led to combine the cited references in *some* manner (which Applicant does not concede), no such combination would have taught or suggested "perform[ing] automatic speech

recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input" and "determin[ing], while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase," and "***in response to determining that the at least one first word or phrase matches the trigger word or phrase***, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (emphasis added). As an additional reason that claim 21 patentably distinguishes over the cited references, no combination would have taught or suggested "***in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server***" (emphasis added).

i. Freeman and Mozer fail to describe "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition."

The Office Action concedes that Freeman would not teach or suggest "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting … that at least one server perform recognition" as recited in previously presented claim 21 (see Office Action p. 6). It follows that Freeman would not have taught or suggested the above-quoted language of claim 21.

The Office Action alleges that figures 4 and 6, ¶¶0032-33, and ¶0066 of Mozer would have disclosed the "detecting while in the low power mode that the acoustic input includes an explicit voice trigger," as recited by previously-pending claim 21 (see Office Action p. 6). As discussed during the interview and highlighted in the discussion of the reference above, the cited passage explicitly describes that the device ("stage 1 recognizer") activates the server ("stage 2 recognizer") when the device determines that a conversation relates to a particular topic/classification (e.g., travel, food, electronics) (¶0033, ¶0066). In terms of activating the server, the device does not take any action when any particular word or phrase is recognized. For example, Mozer describes that the device may recognize the words "flight" and "vacation," but Mozer's device does not take any action based only on those words. Mozer's device does not (1) match those words to a trigger word or phrase, or (2) activate the server in response to detecting that the acoustic input includes those

words (¶0066). In fact, Mozer's device does not take ***any*** particular action in response to recognition of any particular word/phrase, or in response to determining that any particular word/phrase "matches" a "trigger word or phrase." Rather, Mozer's device activates the server only when a classification of a conversation is determined (e.g., "travel") (¶0066). Mozer's server uses the classification identified by the device for determining content to provide to the device, and thus relies on the classification to begin processing. In Mozer, the device does not activate the server "***in response to determining that [] at least one word or phrase matches [a] trigger word or phrase***" (emphasis supplied), as recited in claim 21.

Accordingly, even if the references would have been combined in the manner alleged by the Office Action (which Applicant does not concede), the references would not have taught or suggested the above-quoted language of amended claim 21.

ii. Freeman and Mozer fail to describe "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server."

The server of Mozer does not determine whether an acoustic input includes any voice command. As can be appreciated from the discussion of the reference above, Mozer's server analyzes sentences of a conversation and outputs content (e.g., pictures and coupons) based on the sentences and a classification received from the device (¶0067). The server does not determine whether an acoustic input includes a voice command. As the server does not determine whether acoustic input includes a voice command, the device of Mozer does not receive any indication from the server that the acoustic input includes a voice command. It follows that Mozer does not teach or suggest "***in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command***, respond to the at least one voice command" (emphasis added) as recited in claim 21.

Freeman describes voice activation of a single device without interaction with a separate server. Thus Freeman does not describe that a device responds to a voice command "in response to receiving an indication from the at least one server," as recited in claim 21.

Accordingly, even if the references would have been combined in the manner alleged by the Office Action (which Applicant does not concede), the references would not have taught or suggested the above-quoted language of claim 21.

iii. Conclusion

For at least these reasons, claim 21 patentably distinguishes over any combination of Mozer and Freeman. Accordingly, withdrawal of the rejection under §103 of claim 21 is respectfully requested.

D. Each of Independent Claims 1 and 15 Patentably Distinguishes Over Freeman and Mozer.

Each of claims 1 and 15 recites, in part, "performing automatic speech recognition (ASR), while in the lower power mode, to recognize at least one word or phrase in the acoustic input; determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase; in response to determining that the at least one word or phrase matches the trigger word or phrase, requesting … that at least one server perform recognition … respond to voice commands recognized by the at least one server." For reasons that should be appreciated from the foregoing, each of claims 1 and 15 patentably distinguishes over Freeman and Mozer. Accordingly, withdrawal of the rejections under §103 of claims 1 and 15, and the claims that depend therefrom, is respectfully requested.

III. General Comments on Dependent Claims

As each of the dependent claims depends from a base claim that is believed to be in condition for allowance, Applicant believes that arguing the further distinguishing features of all of the dependent claims is unnecessary at this time. However, Applicant does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does Applicant concur that the basis for the rejection of any of the dependent claims is proper. Therefore, Applicant reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

CONCLUSION

In view of the foregoing amendments and remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the Applicant's representative at the telephone number indicated below to discuss any outstanding issues relating to the allowability of the application.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, Applicant hereby requests any necessary extension of time. If there is a fee occasioned by this response, including an extension fee, please charge any deficiency to Deposit Account No. 23/2825 under Docket No. N0484 .71090US04, from which the undersigned is authorized to draw.

Dated: April 30, 2019

Respectfully submitted,

By /Saad Alam/
Andrew J. Tibbetts, Reg. No. 65,139
Saad Alam, Reg. No. 77,456
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

INTERVIEW AGENDA

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Applicant's representatives, Saad Alam (Reg. No. 77,456) and Andrew Tibbetts (Reg. No. 65,139), thank Examiner Ho for his courtesy in agreeing to conduct a telephone interview scheduled for Thursday, April 11th at 11:00 AM (EST) in connection with the above-identified matter. Applicant's representatives will call the Examiner at 571-270-1108 at the scheduled time.

Applicant's representatives would like to discuss at least the §103 rejection of claim 15 and the manner in which Applicant's representatives believe claims 15 and 21 patentably distinguish over the references. Applicant's representatives would like to discuss at least the limitations:

- "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger comprising a predetermined word or phrase that triggers voice response processing, requesting … that at least one server perform recognition on at least a portion of the acoustic input," as recited by claim 15, and
- "in response to receiving an indication from the at least on server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server," as recited by claim 21.

As part of that discussion, Applicant's representatives would like to discuss the remarks made in the Response dated Nov. 29, 2018 regarding these limitations, the Examiner's interpretation of this claim language, and whether the Examiner has any suggestions on how the claims might be amended to more clearly distinguish the references.

If the Examiner has any questions concerning the Interview Agenda, he is invited to contact the undersigned at the telephone number listed below.

Dated: April 9, 2019

Respectfully submitted,

Electronic signature: /Saad Alam/
Saad Alam
Registration No.: 77,456
Andrew J. Tibbetts
Registration No.: 65,139
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: April 30, 2019 Electronic Signature for: /Eileen M. MacKenzie/

AMENDMENT FILED WITH REQUEST FOR CONTINUED EXAMINATION

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Final Office Action dated February 28, 2019, please enter the following in the above-identified U.S. patent application:

Amendments to the Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 7 of this paper.

AMENDMENTS TO THE CLAIMS

Applicant submits below a complete listing of the current claims, including marked-up claims with insertions indicated by underlining and deletions indicated by strikeouts and/or double bracketing. This listing of claims replaces all prior versions, and listings, of claims in the application:

Listing of the Claims

1. (Currently Amended) A device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and

at least one processor configured to perform at least one processing stage on the acoustic input, the at least one processing stage including:

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input;

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase;

in response to determining that the at least one first word or phrase matches the trigger word or phrase, ~~detecting while in the low power mode~~ that ~~the acoustic input includes an explicit voice trigger,~~ requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command,

wherein the device is configured to respond to voice commands recognized by the at least one server.

2. (Previously presented) The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.

3. (Previously presented) The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

4. (Previously presented) The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

5. (Previously presented) The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

6. (Currently Amended) The device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to determine whether the at least one first word or phrase matches the trigger word or phrase ~~detect whether the acoustic input includes the explicit voice trigger~~.

7. (Previously presented) The device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

8. (Currently Amended) The device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to determine whether the at least one first word or phrase matches the trigger word or phrase ~~detect whether the acoustic input includes an explicit voice trigger~~.

9. (Previously presented) The device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. (Previously presented) The device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

11. (Currently Amended) The device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the low power mode, and wherein the trigger word or phrase ~~explicit voice trigger is a specific word or phrase that~~, when detected, is sufficient to engage the device.

12. (Previously presented) The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. (Previously presented) The device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

14. (Previously presented) The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

15. (Currently Amended) At least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:

performing at least one processing stage on the acoustic input, the at least one processing stage including:

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input;

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase;

in response to determining that the at least one first word or phrase matches the trigger word or phrase ~~detecting while in the low power mode that the acoustic input includes an explicit voice trigger comprising a predetermined word or phrase that triggers voice response processing~~:

requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command; and

responding to voice commands recognized by the at least one server.

16. (Original) The at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

17. (Currently Amended) The at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the at least one first word or phrase matches the trigger word or phrase ~~detect whether the acoustic input includes the explicit voice trigger~~.

18. (Original) The at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the low power mode.

20. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

21. (Currently Amended) A device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and

at least one processor configured to:

perform automatic speech recognition (ASR), while in a low power mode, to recognize at least one word or phrase in the acoustic input;

determine, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase;

in response to determining that the at least one first word or phrase matches the trigger word or phrase ~~detecting while in the low power mode that the acoustic input includes an explicit voice trigger~~, ~~requesting~~ request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command; and

in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server.

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 15/238,238	Filing Date 08/16/2016	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $80 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $420 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	04/30/2019	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 21	Minus	** 21	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 3	Minus	*** 3	= 0		x $460 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	SLIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/JULIET MCMILLAN/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

REMARKS

In response to the Office Action dated February 28, 2019, Applicant respectfully requests reconsideration. Each of the rejections set forth in the Office Action has been carefully considered and is addressed below. To further prosecution of the application, amendments have been made to the claims herein. The claims as presented are believed to be in condition for allowance.

I. Interview Summary

Applicant's representatives, Saad Alam and Andrew J. Tibbetts, thank Examiner Ho for conducting the interview on April 11, 2019. The substance of that interview is summarized in the discussion below.

During the interview, Applicant's representatives discussed some embodiments of the technology described in connection with FIG. 4 of the application. Next, the discussion turned to the rejection of the claims over the Freeman and Mozer references. In particular, Applicant's representatives explained distinctions of techniques described in the application from those of Mozer, as detailed below. The Examiner agreed that there was a distinction between the technique and the references, but suggested amending the claims to better capture the distinction. Applicant's representatives discussed changes to the claims to better capture the distinction. The Examiner agreed to reconsider the rejection in view of amendments discussed during the telephone interview. Accordingly, the claims are amended herein along the lines discussed during the telephone interview.

II. Claim Rejections under 35 U.S.C. §103

Independent claims 1, 15 and 21 were rejected under 35 U.S.C. §103 as allegedly being obvious over U.S. Patent No. 8,977,255 ("Freeman") in view of U.S. Patent Application Publication No. 2013/0080167 ("Mozer"). Applicant respectfully requests reconsideration in view of the telephone interview, the foregoing amendments, and the following remarks.

A. Discussion of Some Embodiments From The Specification

Some embodiments described in the present application relate to a technique for a device to determine whether an acoustic input includes a voice command, while the device remains in a low power mode (Published Specification, ¶0035). The device determines whether the acoustic input includes a trigger word or phrase (e.g., "Hello Dragon") by (1) performing automatic speech recognition (ASR) on the acoustic input to recognize spoken words or phrases in the acoustic input; and (2) determining whether any of the recognized words or phrases match the trigger word or phrase (¶0087). For example, the device may determine whether any of the phrases recognized in the acoustic input match the trigger phrase "Hello Dragon" (¶0087).

In response to determining that a trigger word phrase was spoken in the acoustic input, the device may (while remaining in the low power mode) invoke more comprehensive ASR to determine whether a voice command is present in the acoustic input (¶0090). To invoke more comprehensive ASR, the device may transmit (while in the low power mode) the acoustic input to a server may perform more computationally expensive ASR to determine whether the acoustic input includes a voice command (¶0091). When the server determines that the acoustic input includes a voice command, the server may notify the device, such that the device responds to the voice command recognized by the server (¶0070).

The foregoing discussion is provided to assist the Examiner in appreciating some applications for various inventive aspects of the present disclosure. However, this discussion may not apply to each of the independent claims, and the language of the independent claims may differ in material respects from the discussion provided above. Thus, Applicant respectfully requests that careful consideration be given to the language of each of the independent claims and that each be addressed on its own merits, without relying on the discussion provided above. In this respect, Applicant does not rely on the discussion provided above to distinguish any of the claims over the cited art. Rather, Applicant relies only upon the language of the claims themselves and the arguments related specifically thereto.

B. Discussion of Mozer

Mozer describes a "background speech recognizer" (Mozer, ¶0020) that passively listens to a conversation between two or more speakers and, after inferring a topic of the conversation, displays information that might be relevant to the speakers (¶0066). Mozer thus describes that a device "may classify keywords that are recognized into pre-defined classifications" that are "topics that describe different areas of interest, such as travel, purchases, entertainment, research, food, or electronics" (Mozer, ¶0033). In Mozer's system, once "a certain number of keywords for a specific classification are detected, then it may be determined that a topic associated with the classification is being discussed" and "***when one classification has enough key words classified with it, then stage 1 recognizer 102 may trigger stage 2 recognizer 104***" (¶0033, emphasis added). As an example, the device selects a classification of "travel" based on the words "vacation" and "flight" in an acoustic input and determines "***that the travel classification should be selected and the stage 2 recognizer 104 should be activated***" (¶0066, emphasis supplied). This "stage 2 recognizer" is a server-based recognizer, which analyzes the conversation based on the identified topic and determines relevant information to display (¶0067).

In particular, Mozer describes that the server "receives the trigger to activate and also may receive information that a conversation is occurring about the classification of 'travel'" (¶0067). "At this point, stage 2 recognizer 104 may take over listening to the conversation" (¶0067). The "stage 2 recognizer 104 may be able to decipher whole sentences and may hear a sentence 'Maybe we should do an activity in Ireland'" and "the classification of 'travel' may be used to determine content for the response" (¶0067). "At this point, a response may be determined that pictures of Ireland should be output with a coupon for a soccer game in Ireland" and "pictures of Ireland may be output to an interface, such as the tablet computer screen" (¶0067).

C. Claim 21 Patentably Distinguishes Over Freeman and Mozer.

Independent claim 21 is amended herein, based on the discussion with the Examiner in the telephone interview, to more clearly differentiate the cited references. Even if one of ordinary skill would have been led to combine the cited references in *some* manner (which Applicant does not concede), no such combination would have taught or suggested "perform[ing] automatic speech

recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input" and "determin[ing], while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase," and "***in response to determining that the at least one first word or phrase matches the trigger word or phrase***, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (emphasis added). As an additional reason that claim 21 patentably distinguishes over the cited references, no combination would have taught or suggested "***in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server***" (emphasis added).

i. Freeman and Mozer fail to describe "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition."

The Office Action concedes that Freeman would not teach or suggest "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger, requesting … that at least one server perform recognition" as recited in previously presented claim 21 (see Office Action p. 6). It follows that Freeman would not have taught or suggested the above-quoted language of claim 21.

The Office Action alleges that figures 4 and 6, ¶¶0032-33, and ¶0066 of Mozer would have disclosed the "detecting while in the low power mode that the acoustic input includes an explicit voice trigger," as recited by previously-pending claim 21 (see Office Action p. 6). As discussed during the interview and highlighted in the discussion of the reference above, the cited passage explicitly describes that the device ("stage 1 recognizer") activates the server ("stage 2 recognizer") when the device determines that a conversation relates to a particular topic/classification (e.g., travel, food, electronics) (¶0033, ¶0066). In terms of activating the server, the device does not take any action when any particular word or phrase is recognized. For example, Mozer describes that the device may recognize the words "flight" and "vacation," but Mozer's device does not take any action based only on those words. Mozer's device does not (1) match those words to a trigger word or phrase, or (2) activate the server in response to detecting that the acoustic input includes those

words (¶0066). In fact, Mozer's device does not take ***any*** particular action in response to recognition of any particular word/phrase, or in response to determining that any particular word/phrase "matches" a "trigger word or phrase." Rather, Mozer's device activates the server only when a classification of a conversation is determined (e.g., "travel") (¶0066). Mozer's server uses the classification identified by the device for determining content to provide to the device, and thus relies on the classification to begin processing. In Mozer, the device does not activate the server "***in response to determining that [] at least one word or phrase matches [a] trigger word or phrase***" (emphasis supplied), as recited in claim 21.

Accordingly, even if the references would have been combined in the manner alleged by the Office Action (which Applicant does not concede), the references would not have taught or suggested the above-quoted language of amended claim 21.

ii. Freeman and Mozer fail to describe "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server."

The server of Mozer does not determine whether an acoustic input includes any voice command. As can be appreciated from the discussion of the reference above, Mozer's server analyzes sentences of a conversation and outputs content (e.g., pictures and coupons) based on the sentences and a classification received from the device (¶0067). The server does not determine whether an acoustic input includes a voice command. As the server does not determine whether acoustic input includes a voice command, the device of Mozer does not receive any indication from the server that the acoustic input includes a voice command. It follows that Mozer does not teach or suggest "***in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command***, respond to the at least one voice command" (emphasis added) as recited in claim 21.

Freeman describes voice activation of a single device without interaction with a separate server. Thus Freeman does not describe that a device responds to a voice command "in response to receiving an indication from the at least one server," as recited in claim 21.

Accordingly, even if the references would have been combined in the manner alleged by the Office Action (which Applicant does not concede), the references would not have taught or suggested the above-quoted language of claim 21.

iii. Conclusion

For at least these reasons, claim 21 patentably distinguishes over any combination of Mozer and Freeman. Accordingly, withdrawal of the rejection under §103 of claim 21 is respectfully requested.

D. Each of Independent Claims 1 and 15 Patentably Distinguishes Over Freeman and Mozer.

Each of claims 1 and 15 recites, in part, "performing automatic speech recognition (ASR), while in the lower power mode, to recognize at least one word or phrase in the acoustic input; determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase; in response to determining that the at least one word or phrase matches the trigger word or phrase, requesting … that at least one server perform recognition … respond to voice commands recognized by the at least one server." For reasons that should be appreciated from the foregoing, each of claims 1 and 15 patentably distinguishes over Freeman and Mozer. Accordingly, withdrawal of the rejections under §103 of claims 1 and 15, and the claims that depend therefrom, is respectfully requested.

III. General Comments on Dependent Claims

As each of the dependent claims depends from a base claim that is believed to be in condition for allowance, Applicant believes that arguing the further distinguishing features of all of the dependent claims is unnecessary at this time. However, Applicant does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does Applicant concur that the basis for the rejection of any of the dependent claims is proper. Therefore, Applicant reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

CONCLUSION

In view of the foregoing amendments and remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the Applicant's representative at the telephone number indicated below to discuss any outstanding issues relating to the allowability of the application.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, Applicant hereby requests any necessary extension of time. If there is a fee occasioned by this response, including an extension fee, please charge any deficiency to Deposit Account No. 23/2825 under Docket No. N0484 .71090US04, from which the undersigned is authorized to draw.

Dated: April 30, 2019

Respectfully submitted,

By /Saad Alam/
Andrew J. Tibbetts, Reg. No. 65,139
Saad Alam, Reg. No. 77,456
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

INTERVIEW AGENDA

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Applicant's representatives, Saad Alam (Reg. No. 77,456) and Andrew Tibbetts (Reg. No. 65,139), thank Examiner Ho for his courtesy in agreeing to conduct a telephone interview scheduled for Thursday, April 11th at 11:00 AM (EST) in connection with the above-identified matter. Applicant's representatives will call the Examiner at 571-270-1108 at the scheduled time.

Applicant's representatives would like to discuss at least the §103 rejection of claim 15 and the manner in which Applicant's representatives believe claims 15 and 21 patentably distinguish over the references. Applicant's representatives would like to discuss at least the limitations:

- "in response to detecting while in the low power mode that the acoustic input includes an explicit voice trigger comprising a predetermined word or phrase that triggers voice response processing, requesting … that at least one server perform recognition on at least a portion of the acoustic input," as recited by claim 15, and
- "in response to receiving an indication from the at least on server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server," as recited by claim 21.

As part of that discussion, Applicant's representatives would like to discuss the remarks made in the Response dated Nov. 29, 2018 regarding these limitations, the Examiner's interpretation of this claim language, and whether the Examiner has any suggestions on how the claims might be amended to more clearly distinguish the references.

If the Examiner has any questions concerning the Interview Agenda, he is invited to contact the undersigned at the telephone number listed below.

Dated: April 9, 2019

Respectfully submitted,

Electronic signature: /Saad Alam/
Saad Alam
Registration No.: 77,456
Andrew J. Tibbetts
Registration No.: 65,139
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Publication number: CN103021409 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20130403
Application number: CN201210455215
Application date: 20121113
Priority: CN201210455215201211 3CN20121455215;
Assignee: ANHUI USTC IFLYTEK CO ;
Assignee[std]: ANHUI USTC IFLYTEK CO LTD ;
Inventor[std]: ZHOU JIAN ; PAN ZHIBING ; MA BING ; ZHANG KAI ; ZHU GUOZHENG ; SHU YINDONG ; WANG HAIKUN ;
International class[8]: G10L15/22; G10L15/06; H04N5/232;
Family members: CN103021409 A CN103021409 B

Title: VOICE ACTIVATING PHOTOGRAPHING SYSTEM

Title: 一种语音启动拍照系统

Abstract:

The invention provides a voice activating photographing system. The voice activating photographing system comprises an operating system module (10), an activating word customizing module (11), a photographing module (12), an audio module (13) and a storage module (14), wherein the voice activating module runs in the operating system module (10) and includes voice input (11), voice activating algorithm (2) and activating execution (3); the voice activating algorithm (2) is used for acquiring voice signals from the voice input (1), carrying out voice activating, and outputting the result to the activating execution (3), thus finishing the activating operation; and the voice activating algorithm (2) is executed by acoustics characteristic extracting (4), activating word detecting (5), activating word confirming (6), construction of an activating word detecting network (7), a training acoustics module (8) and construction of an activating word confirming network (9). According to the voice activating photographing system, command words can be customized according to the user requirement, human-computer exchange is more humanized, the accuracy in activating is higher, and the interference resistance stronger.

Abstract:

一种语音启动拍照系统，包括：操作系统模块（10）、唤醒词定制模块（11）、拍照模块（12）、音频模块（13）和存储模块（14）；其中：操作系统模块（10）中运行语音唤醒模块；所述语音唤醒模块包括语音输入（1）、语音唤醒算法（2）和唤醒执行（3）；语音唤醒算法（2）获取语音输入（1）的语音信号，进行语音唤醒处理

A speech start camera system, comprising: an operating system module (10), Wake word custom module (11), camera module (12), an audio module (13) and memory module (14); wherein: the operating system module (10) Voice wake run module; the voice wake module includes voice input (1), voice wake algorithm (2) and wake Execution (3); Voice wake algorithm (2) Get Voice

后，将结果输出给唤醒执行（3），从而完成唤醒操作；所述语音唤醒算法（2）通过声学特征提取（4）、唤醒词检测（5）、唤醒词确认（6）、构建唤醒词检测网络（7）、训练声学模型（8）和构建唤醒词确认网络（9）来实现。本发明可以根据用户需求进命令词定制，人机交换更人性化启动的准确度更高，抗干扰性更强。

Input (1) of the voice signal, voice wake After treatment, the resulting output to wake execution (3), thereby completing the wakeup operations; the voice wake-up algorithm (2) by the acoustic feature extraction (4), Wake term test (5), Wake word confirmation (6), Construction wake Words detect network (7), trained acoustic model (8), and build a network wake-word confirmation (9) to achieve. The present invention can enter the command word customized according to user needs, more accurate and more user-friendly human-computer exchange started, stronger anti-interference.

Claims:

CN 1. 一种语音启动拍照系统，其特征在于包括：操作系统模块（10)、唤醒词定制模块 (11)、拍照模块（12)、音频模块（13)和存储模块（14);其中：操作系统模块（10)中运行语音唤醒模块；所述语音唤醒模块包括语音输入（1)、语音 唤醒算法（2)和唤醒执行（3)；语音唤醒算法（2)获取语音输入（1)的语音信号，进行语音 唤醒处理后，将结果输出给唤醒执行（3)，从而完成唤醒操作；所述语音唤醒算法（2)通过 声学特征提取（4)、唤醒词检测（5)、唤醒词确认（6)、构建唤醒词检测网络（7)、训练声学模 型（8)和构建唤醒词确认网络（9)来实现，具体实现过程如下： 第一步，声学特征提取（4)：通过语音输入（1)获取语音信号输入，提取具有区分性的、 并且是基于人耳听觉特性提取的特征，通常选取语音识别中用到的MFCC (Mel-Frequency Cepstrum Coefficient，美尔频率倒谱系数）特征作为声学特征； 第二步，唤醒词检测（5):将提取得到的声学特征，采用训练的声学模型（8)在唤醒词检 测网络（7)上计算声学得分，如果声学得分最优的路径中包含要检测的唤醒词，则确定已检 出唤醒词，进入第三步操作，否则回到第一步重新进行提取声学特征（4)； 第三步，唤醒词确认（6):将提取得到的声学特征，采用训练的声学模型（8)在唤醒词确 认网络（9)上进行唤醒词确认，得到最终确认得分；判断该检出的唤醒词是否为真实的唤 醒词，即将该唤醒词的最终确认得分和预先设定的门限，如果最终确认得分大于等于门限， 则认为该唤醒词是真实的唤醒词，语音唤醒成功，将结果输出给唤醒执行（3)，从而完成语 音唤醒操作；如果最终确认得分小于门限，则认为该唤醒词为虚假的唤醒词，重新回到第一 步重新进行声学特征提取（4); 唤醒词定制模块（11)，接收用户输入的唤醒词，输出操作系统模块（10)中语音唤醒模 块所需的声学模型（8)和唤醒词检测网络（7)； 拍照模块（12)，具有拍照功能的系统模块，接收操作系统模块（10)的拍照指令，并在 拍

CN 1. A speech start camera system, comprising: an operating system module (10), Wake word custom module (11), camera module (12), an audio module (13) and storage module (14); wherein: the operating system module (10) running voice wake-up module; the voice wake-up module includes voice input (1), voice wake algorithm (2) and wake Execution (3); Voice wake algorithm (2) Get Voice Input (1) of the voice signal, voice wake-up processing, and outputs the result to wake execution (3), thereby completing the wake-up; the voice wake-up algorithm (2) by the acoustic feature extraction (4), Wake-term detection (5), Wake word confirmation (6), to build a network wake-word detection (7), trained acoustic model (8), and build a network wake-word confirmation (9) to achieve the specific process is as follows: first, the acoustic feature extraction (4): (1) obtain a voice signal input through the voice input, extract distinguishing nature, and is based on the human auditory characteristics of the extracted features, usually used in selecting the speech recognition MFCC (Mel-Frequency Cepstrum Coefficient, Mayer Frequency Cepstral Coefficients) feature as the acoustic characteristics; the second step, wake-word test (5): the extracted acoustic features obtained, preclude the use of trained acoustic model (8) in the wake of the word detecting network (7) to calculate the acoustic score, If the acoustic score the best path contains a wake-up word to be detected, it is determined that the word has been detected wake, the third step, otherwise return to the first step in re-extracted acoustic features (4); the third step, wake word Confirm (6): the extracted acoustic features obtained using the acoustic model training (8) in the wake of the word confirm confirmation wake word on the network

照后将图像的数字信号发送到存储模块（14）； 音频模块（13），具有录音采集和音频播报功能，录音功能主要是采集环境中音频数据， 并传送到操作系统模块（10），播音功能主要是接收系统模块（10)发出提示反馈后，播报提 示音功能； 存储模块（14），具有永久存储能力的存储介质，对拍照模块生成的数字图像信息进行 存储。

2.根据权利要求1所述的语音启动拍照系统，其特征在于：所述声学模型（8)的训 练分为两部分，分别为音素声学模型和废料模型（即Garbage模型）；音素声学模型采用传 统的语音识别中的声学模型训练方法，选取数据库，利用基于MLE (Maximum Likelihood Estimation，最大似然估计）和MPE (Minimum Phone Error，最小音素错误）区分性训练准 则下得到；Garbage模型用于吸收除唤醒词之外的无关语音，使用和训练音素模型同样的 数据库，通过计算各音素模型之间的相似性，将各音素分为20类，使用每类音素对应的所 有训练数据融合起来，采用MLE准则训练对应的Garbage模型，便得到20类Garbage模型。

3.根据权利要求1所述的语音启动拍照系统，其特征在于：所述唤醒词检测网络（7)的 实现采用最优得分路径计算得出，所述最优得分路径的计算公式是： F = Srgmaxr(W)P(^IF) 2 CN 其中X代表从输入语音中提取的声学特征向量，W代表得分最大的最优词序列；条件概 率P (XI ff)为声学模型得分，通过训练好的声学模型（8)计算得到；先验概率P (W)为语言模 型得分，即为对不同的声学模型所加的PenaltyP(X)为全概率，当声学模型和唤醒词检测网 络确定下来后即是定值。

4.根据权利要求1所述的语音启动拍照系统，其特征在于：所述唤醒词确认网络（9)实 现是： a.将检出的唤醒词解码到音素一级，并记录所有的得分（Sc0rephmel, SCOreph_2^-，Scoreph。neN)，其中 N 为唤醒词中总的音素个数，Scorephonel, Scorephone2,…，Scoreph。neN 分别表 示该唤醒词中所有音素的是解码得分，其中下标表示音素的N个音素的标识； b.使用和唤醒词检测同样的特征，得到相应的声学得分，并精确到帧一级(Scoreframel, Scoreframe2,…，ScoreframeM)，其中M为该句特征总的时长，以帧为单位； c.计算得到唤醒词每个音素的确认得分，计算方式如下：^imd CMphomi = [Scorephonei — [Scoreframek)/(Kiend — Ktstorr) k=Kistan 其中Kistart和Kimd分别为第i个音素的起始时间和结束时间； CMphonei表示第i个音素的确认得分，下标phonei表示第i个音素，Scorephonei如上面 所示第i个Phone的解码得分，Scoreframek表示使用唤醒词确认网络解码得到的第k帧的得 分； d.计算得到该唤醒词的最终确

(9), to give final confirmation score; determining whether the detected wake-up word for real wake-up word, will the wake-up word final score and a threshold set in advance, if the final score is greater than equal to the threshold, it is considered that the wake-up word is a real wake-up word, voice wake-up success, and outputs the result to wake Execution (3), thereby completing the voice wake-up; if a final confirmation score less than the threshold, it is considered that the wake-up word for word a false wake, the first step back to re-acoustic feature extraction (4); wakeup word custom module (11), receiving user Wake-term input and output operating system module (10) in the required modules wake voice acoustic model (8) and wake-up word detection network (7); camera module (12), having a camera system module, the receiver operating system modules (10) of camera commands and sends the digital signal after the camera image to the memory module (14); audio module (13), having a record collection and audio broadcast function, voice recording mainly audio data collection environment, and transfer to the operating system module (10), the main function after receiving broadcast system module (10) to alert feedback, broadcast beep function; a storage module (14), a permanent storage capacity of the storage medium, the digital image information generated by the camera module for storage.

.

2 according to claim 1, wherein the speech start camera system, characterized in that: said acoustic model (8) of the training is divided into two portions, respectively, acoustic models and phoneme model waste (ie Garbage model); phoneme acoustic models using the traditional speech recognition acoustic model training method, select a database, use-based MLE (Maximum Likelihood Estimation, maximum likelihood estimation) and MPE (Minimum Phone Error, the smallest phoneme error) Division of By training criteria obtained; Garbage model except for absorbing the same wake-up word database-independent voice, using and training the phoneme models, by calculating each of the similarity between the phoneme models, each phoneme is divided into 20 classes, each class using All the training data corresponding phonemes integrate using MLE training guidelines Garbage corresponding model, it will have 20 categories Garbage models.

认得分，计算方式如下所示： I N C^w,/=T7 Z 5.根据权利要求1所述的语音启动拍照系统，其特征在于：所述操作系统模块（10)中 的语音唤醒模块能够被移植到被移植到Windows、Linux、Android、IOS系统平台。

.

3 speech according to claim 1, wherein the camera system is started, characterized in that: said wakeup word detection network (7) is realized using the optimal route calculated score, the score is calculated optimal path is: F = Srgmaxr (W) P (IF) 2 CN wherein X represents acoustic features extracted from the input speech vector, W maximum score representative of the optimal word sequence; conditional probability P (XI ff) for the acoustic model score, by trained acoustic models (8) is calculated; the prior probability P (W) is the language model score, that is, different acoustic models applied PenaltyP (X) for the whole probability, when Wake-word acoustic model and finalized after detecting network is constant.

.

4 according to claim 1, wherein the voice start camera system, wherein: the wake-up word Verify that the network (9) implementation is:. A wake will be checked to decode words phoneme-level, and record all scores (Sc0rephmel, SCOreph_2 -, Scoreph.neN), where N is the total number of wake-up word phoneme, Scorephonel, Scorephone2,..., Scoreph. neN words denote the wake of all the phonemes are decoding score, where the subscript represents N phonemes phonemes identification;. b detected using the same words and wake-up feature to give the corresponding acoustic score, and frame-accurate level (Scoreframel, Scoreframe2,..., ScoreframeM), where M is the total time for the sentence length feature, in frame units; c calculated wake-word recognition score for each phoneme, calculated as follows: imd CMphomi = [Scorephonei - [Scoreframek.) / (Kiend - Ktstorr) k = Kistan where Kistart and Kimd are the i-th phoneme start time and end time; CMphonei represents the i-th phoneme recognition score, the subscript phonei represents the i-th phoneme, Scorephonei as above Phone shown decoding the i-th score, Scoreframek confirm that the use of the word wake scoring the first k network decoded frame; d calculated to obtain the final confirmation of the wake-up word score, calculated as follows:. INC w, / = T7 Z 5. Voice of claim 1, wherein the camera system to start, characterized in that: the operating system module (10) in the wake of the speech module can be ported to be ported to Windows, Linux, Android, IOS platform.

Petitioner Exhibit 1004, Page 442 of 1052

Description:

CN 一种语音启动拍照系统 技术领域 [0001] 本发明涉及一种语音启动拍照系统，可以用于含有拍照功能的消费类电子产品领 域。

背景技术 [0002] 随着消费类电子产品快速的发展，各种包含拍照功能的设备使用越来越频繁，人 们对拍照的人性化要求越越来越高，语音作为人类最基本的方式，将语音识别技术应用到 消费类电子产品中，实现通过自然语音来控制此类产品功能是未来发展的趋势。

发明内容 [0003] 本发明技术解决问题：提供一种语音启动拍照系统，克服了现有按钮式、触控式启 动方式带来的不便利，操作方式受限和拍照抖动等不足，具有操作简单更人性化，启动拍照 不受限，防抖动等优点。现有的一些语音启动技术，是利用特点语音如口哨，误启动率搞，输 入语音命令单一等缺陷，本发明可以根据用户需求进命令词定制，人机交换更人性化启动 的准确度更高，抗干扰性更强。

[0004] 本发明是通过以下技术方案实现的：一种语音启动拍照系统，包括：操作系统模 块10、唤醒词定制模块11、拍照模块12、音频模块13和存储模块14 ；其中： [0005] 操作系统模块10中运行语音唤醒模块；所述语音唤醒模块包括语音输入1、语音唤醒算法2和唤醒执行3 ；语音唤醒算法2获取语音输入1的语音信号，进行语音唤醒处理 后，将结果输出给唤醒执行3，从而完成唤醒操作；所述语音唤醒算法2通过声学特征提取 4、唤醒词检测5、唤醒词确认6、构建唤醒词检测网络7、训练声学模型8和构建唤醒词确认 网络9来实现，具体实现过程如下： [0006] 第一步，声学特征提取4 ：通过语音输入1获取语音信号输入，提取具有区分性的、 并且是基于人耳听觉特性提取的特征，通常选取语音识别中用到的MFCC (Mel-Frequency Cepstrum Coefficient，美尔频率倒谱系数）特征作为声学特征； [0007] 第二步，唤醒词检测5 ：将提取得到的声学特征，采用训练的声学模型8在唤醒词 检测网络7上计算声学得分，如果声学得分最优的路径中包含要检测的唤醒词，则确定已 检出唤醒词，进入第三步操作，否则回到第一步重新进行提取声学特征4 ； [0008] 第三步，唤醒词确认6 ：将提取得到的声学特征，采用训练的声学模型8在唤醒词 确认网络9上进行唤醒词确认，得到最终确认得分；判断该检出的唤醒词是否为真实的唤 醒词，即将该唤醒词的最终确认得分和预先设定的门限，如果最终确认得分大于等于门限， 则认为该唤醒词是真实的唤醒词，语音唤醒成功，将结果输出给唤醒执行

CN a voice-activated camera system TECHNICAL FIELD

[0001] The present invention relates to a voice-activated camera system that can be used in consumer electronics products containing the camera function.

[0002] With the rapid development of consumer electronics products, a variety of devices containing camera more frequently, people taking pictures of the increasingly high requirements of humanity, as the most basic human voice way, the application of speech recognition technology to consumer electronics products, through natural speech to control functions such products is the future trend of development.

Disclosure

[0003] The present technology to solve the problem: to provide a voice activated camera system, to overcome the existing push-button, bringing a touch start-up mode is not convenient, the operation mode is limited and insufficient camera jitter, Simple operation more humane, start taking pictures is not limited, Stabilize advantages. Some of the existing voice-activated technology is the use of voice characteristics such as whistles, the rate of false start out, enter a single voice command and other defects, the present invention can enter the command word customized according to user needs, more user-friendly human-computer exchange of higher accuracy of the start of stronger anti-interference.

[0004] The present invention is achieved by the following technical solutions: one voice activated camera system, comprising: an operating system module 10, a wake-up word customization module 11, a camera module 12, an audio module 13 and the memory module 14; wherein:

[0005] The operating system module 10 runs wake voice module; the voice wake-up module includes voice input 1, 2 and wake voice wake algorithm execution 3; voice wake Algorithm 2 Get 1 voice input voice signals for voice wake-up post-treatment, The results are output to the wake-up implementation of three, thus completing the wakeup operations; the voice wake-up acoustic feature extraction algorithm 2 by 4, 5 wake-word

3，从而完成语音 唤醒操作；如果最终确认得分小于门限，则认为该唤醒词为虚假的唤醒词，重新回到第一步 重新进行声学特征提取4； [0009] 唤醒词定制模块11，语音唤醒模块的资源生成模块接收用户输入的唤醒词，输出 操作系统模块10中语音唤醒模块所需的声学模型8和唤醒词检测网络7； 4 CN [0010] 拍照模块12，具有拍照功能的系统模块，接收操作系统模块10的拍照指令，并在 拍照后将图像的数字信号发送到存储模块14； [0011] 音频模块13，具有录音采集和音频播报功能，录音功能主要是采集环境中音频数 据，并传送到操作系统模块10，播音功能主要是接收系统模块10发出提示反馈后，播报提 示音功能； [0012] 存储模块14，具有永久存储能力的存储介质，对拍照模块生成的数字图像信息进 行存储。

[0013] 所述声学模型8的训练分为两部分，分别为音素声学模型和废料模型（即Garbage 模型）；音素声学模型采用传统的语音识别中的声学模型训练方法，选取数据库，利用基于 MLE (Maximum Likelihood Estimation，最大似然估计）和 MPE (Minimum Phone Error, 最小音素错误）区分性训练准则下得到；Garbage模型用于吸收除唤醒词之外的无关语音， 使用和训练音素模型同样的数据库，通过计算各音素模型之间的相似性，将各音素分为20 类，使用每类音素对应的所有训练数据融合起来，采用MLE准则训练对应的Garbage模型， 便得到20类Garbage模型。

[0014] 所述唤醒词检测网络7的实现采用最优得分路径计算得出，所述最优得分路径的 计算公式是：[00151 W = argi職I W) W [0016] 其中X代表从输入语音中提取的声学特征向量，W代表得分最大的最优词序列；条 件概率P(XlW)为声学模型得分，通过训练好的声学模型8计算得到；先验概率P(W)为语言 模型得分，即为对不同的声学模型所加的PenaltyP(X)为全概率，当声学模型和唤醒词检测网络确定下来后即是定值。

[0017] 所述唤醒词确认网络9实现是： [0018] a.将检出的唤醒词解码到音素一级，并记录所有的得分（Sc0rephmel, Scoreph0ne2,... ，Scoreph。neN)，其中 N 为唤醒词中总的音素个数，Scorephonel, Scorephone2,...，Scoreph。neN 分别表 示该唤醒词中所有音素的是解码得分，其中下标表示音素的N个音素的标识； [0019] b.使用和唤醒词检测同样的特征，得到相应的声学得分，并精确到帧一级 (Scoreframel, Scoreframe2,...，Sc0reframeM)，其中M为该句特征总的时长，以帧为单位； [0020] c.计算得到唤醒词每个音素的确认得分，计算方式如下： ^iend [0021] CMphonei = (Scoreplwnej - Scorefmmek) / (Kiend [0022] 其中

detection, Wake words confirm 6, construct wake word detection network 7, 8, and build the acoustic model training wake word Verify that the network 9 to achieve the specific process is as follows:

[0006] a first step, the acoustic feature extraction 4: 1 to obtain a voice through the voice input signal is input, extracts distinguishing nature, and is based on the human auditory characteristics extracted features, usually Select speech recognition used in MFCC (Mel-Frequency Cepstrum Coefficient, Mel Frequency Cepstral) feature as the acoustic characteristics;

[0007] The second step, wake word detection 5: The extracted acoustic features obtained using acoustic training Model 8 in the wake of the word detecting network on 7 acoustic score is calculated, if the acoustic score optimal path contains a wake-up word to be detected, it is determined that the word has been detected wake, the third step, otherwise return to the first step in re extracting acoustic features 4;

[0008] The third step, wake word confirm 6: extract acoustic features obtained using the acoustic model training of 8 on the network in the wake of the word confirm confirm 9. Wake word to give final confirmation score; determining the subject the wake-up word whether real wake-up word, will the wake-up word final score with a preset threshold, if the final confirmation of the score than or equal to the threshold, it is considered that the wake-up word is a real wake-up word, voice wake-up is successful, Results output to wake execution 3, thus completing the voice wake-up operation; if final score is less than the threshold, that the wake-up word for word a false wake, the first step back to re-acoustic feature extraction 4;

[0009] wake customized word module 11, resource generation module voice module receives wake wake term user input, output operating system module 10 Wake voice acoustic model required modules 8 and wake-word detecting network 7; 4 CN

[0010] camera module 12, with pictures system module function, the receiver operating system modules photographing instruction 10, and transmits the image photographed in the digital signal after the storage module 14;

[0011] audio module 13, with the acquisition and recording audio broadcast function, the main

Kistot和Kimd分别为第i个音素的起始时间和结束时间； [0023] CMphonei表示第i个音素的确认得分，下标phonei表示第i个音素，Scorephonei如上 面所示第i个Phone的解码得分，Scoreframek表示使用唤醒词确认网络解码得到的第k帧的 得分； [0024] d.计算得到该唤醒词的最终确认得分，计算方式如下所示： 1 N . [0025] CMword=-YjCMphonm iv j=i [0026] 所述操作系统模块中的语音唤醒模块能够被移植到被移植到Windows、Linux、 2 CN Android、IOS系统平台。

[0027] 本发明与现有技术相比的优点在于： [0028] (1)本发明通过用户特定的语音唤醒词作为触发源，即使是在嘈杂的环境下，也无 需用户利用双手操作，仅通过语音命令快速实现唤醒功能，进行下一步交互操作。

[0029] (2)本发明带来的价值是，使用本系统后可通过说约定的语音唤醒词开启语音唤 醒功能，无需事先暂停音频播放，同时通过实际测试验证，正确识别唤醒率可以达到90%以 上。

[0030] (3)本发明实现，成本低，代码移植方便，具有很好的推广应用价值。

[0031] (4)本发明中的语音唤醒功能均通过软件算法实现，并且算法可以很方便的被移 植到Windows、Linux、Android、IOS 等系统平台。

附图说明 [0032] 图1为本发明的结构示意图； [0033] 图2为本发明中的语音唤醒模块实现框图； [0034] 图3为本发明的构建唤醒词检测网络示意图； [0035] 图4为本发明的构建唤醒词确认网络示意图。

具体实施方式 [0036] 如图1所示，本发明的语音启动拍照系统包括：操作系统模块10、唤醒词定制模块 11、拍照模块12、音频模块13和存储模块14、操作系统模块10中运行语音唤醒模块。

[0037] 如图2所示，语音唤醒模块包括语音输入1、语音唤醒算法2和唤醒执行3。语音 唤醒算法2实现主要由声学特征提取4、唤醒词检测5、唤醒词确认6、构建唤醒词检测网络 7、训练声学模型8和构建唤醒词确认网络9完成，具体实现过程是： [0038] (1)训练声学模型8 ：声学模型的训练分为两部分，分别为音素声学模型和废料模 型（即Garbage模型)。音素声学模型采用传统的语音识别中的声学模型训练方法，选取合适 的数据库，利用基于MLE(Maximum Likelihood Estimation,最大似然估计)和MPE(Minimum Phone Error，最小音素错误）区分性训练准则下的得到。Garbage模型用于

function is to collect environmental record audio data, and transferred to the operating system module 10, the main function is to receive broadcast system to alert the feedback module 10, the broadcast beep function;

[0012] The storage module 14, with a permanent storage capacity of the storage medium, for photographing module generated storing the digital image information.

[0013] The acoustic model training 8 is divided into two parts, namely phoneme acoustic model and scrap models (ie Garbage model); phoneme acoustic models using the traditional speech recognition acoustic model training method, select the database, use based on the following MLE (Maximum Likelihood Estimation, maximum likelihood estimation) and MPE (Minimum Phone Error, the smallest phoneme error) discriminative training guidelines are; Garbage model-independent voice except for absorbing wake of the word, the use and training of phoneme models The same database, each phoneme model by calculating the similarity between, each phoneme is divided into 20 categories, each category using the corresponding phoneme integrate all training data, using the MLE criteria Garbage corresponding training model, it will have 20 categories Garbage Model.

[0014] wake up the word detection networks 7 uses the best path score is calculated, the formula for calculating the optimal score path is: [00151 W = argi grade IW) W [0016] wherein X represents extracted from the input speech acoustic feature vector, W on behalf of the score's largest optimal word sequence; conditional probability P (XlW) for the acoustic model score 8 by trained acoustic model is calculated; the prior probability P (W) is the language score model, is different acoustic models applied PenaltyP (X) for the whole probability, when the word acoustic model and wake finalized after detecting network is constant.

[0017] 9 Wake words the acknowledgment network implementations are:. [0018] a detection of the wake-up word decoding to phoneme level, and record all scores (Sc0rephmel, Scoreph0ne2,..., Scoreph.neN), where N is the total number of wake-up word phoneme, Scorephonel, Scorephone2,..., Scoreph. neN words denote the wake of all the phonemes are decoding score, where the subscript represents N

吸收除唤醒词 之外的无关语音，使用和训练音素模型同样的数据库，通过计算各音素模型之间的相似性， 将各音素分为20类，使用每类音素对应的所有训练数据融合起来，采用MLE准则训练对应 的Garbage模型，这样即得到20类Garbage模型。Garbage模型采用了聚类的音素训练数 据混合训练，有两种用途，在唤醒词检测网络中用来吸收除唤醒词之外的其他语音，在唤醒 词确认网络中用来计算确认网络的得分。

[0039] (2)声学特征提取4:通过语音输入1获取语音信号输入，提取可以具有一 定区分性的，并且是基于人耳听觉特性提取的特征，一般选取语音识别中用到的MFCC (Mel-Frequency Cepstrum Coeff icient，美尔频率倒谱系数）特征。

[0040] (3)唤醒词检测5 ：将提取得到的声学特征，使用声学模型8在唤醒词检测网络7 上计算声学得分，如果得分最优的路径中包含要检测的唤醒词，则检出唤醒词，进入下一步 操作；否则重新提取声学特征操作。为了保证唤醒词能够被正常的检出，同时无效的语音又 能被有效的吸收。唤醒检测网络的构建主要由用户选定的唤醒词和Garbage模型组成，如 2 CN 图3所示，这种网络在语音识别中也称为识别网络，由于唤醒检测网路结构非常简单，可以 通过简单的程序或则手工构建都可以。由于实际使用环境的复杂性，在很多情况下，接收到 的唤醒语音被噪声污染了，此时唤醒语音对应的声学的特征在音素声学模型上的得分就会 降低的很多，而由于Garbage模型是使用较多音素混合训练得到，其本身不是很精确，声学 特征在Garbage模型上的得分降低的幅度有限，此时唤醒语音就被Garbage模型误吸收，系 统唤醒率就会降低。

[0041] 为了防止上述情况的发生，在唤醒词检测网络上解码时，对Garbage所在的弧的 解码得分做一定的惩罚，即Penalty，使其不能和音素声学模型公平竞争，以保证有被噪声 污染的唤醒语音也能被正常检出。具体的惩罚幅度需要针对不同的唤醒词作经验性的调 m iF. o [0042] 唤醒词检测网络7的实现方法是采用最优得分路径计算得出。

[0043] 最优得分路径的获取采用经典的贝叶斯公式，如下所示： [0044] [0045] 上式中X代表从输入语音中提取的声学特征向量，W代表得分最大的最优词序列。

条件概率P (XIW)为声学模型得分，可以通过训练好的音素声学模型和废料模型计算得到， 先验概率P(W)为语言模型得分，这里可以理解为对不同的声学模型所加的Penalty。P(X) 为全概率，当声学模型和唤醒词检测网络确定下来后就是定值，因此公

phonemes phonemes identification;. [0019] b use to detect and wake-word the same features, to give the corresponding acoustic score, and accurate to the frame a (Scoreframel, Scoreframe2,..., Sc0reframeM), which features M for the total sentence length of time, in units of frames; [0020] c is calculated for each phoneme confirmation wake word score, calculated as follows: iend [. 0021] CMphonei = (Scoreplwnej - Scorefmmek) / (Kiend [0022] wherein Kistot and Kimd are the i-th phoneme start time and end time; [0023] CMphonei represents confirmation of the i-th phoneme score represents the subscript phonei the i-th phoneme, decoding Scorephonei as shown above the i-th Phone scoring, Scoreframek Claim word score represents the wake-up network of the k-th decoded frame;. [0024] d calculated final confirmation of the wake-up word score calculation Information is as follows:. 1 N [0025] CMword = -YjCMphonm iv j = i [0026] The operating system modules voice wake modules can be ported to be ported to Windows, Linux, 2 CN Android, IOS platform.

[0027] advantages of the present invention over the prior art in that: [0028] (1) The term of the present invention as a wakeup trigger sources via user-specific voice, even in a noisy environment, without requiring the user with both hands operation, only to quickly realize the wake-up function using a voice command, the next step interaction.

[0029] (2) is to bring value to the present invention, the use of the system can be agreed upon by saying the word turn voice voice wake wake-up function, without prior pause audio playback, verified by actual test, the rate of correct identification wake It can reach 90 percent.

[0030] (3) of the present invention is to realize, low cost, easy code migration, and has good application value.

[0031] (4) of the present invention, the voice wake-up function through the software algorithm, and the algorithm can easily be ported to Windows, Linux, Android, IOS and other platforms.

BRIEF DESCRIPTION [0032] FIG. 1 a schematic diagram of the structure of the present invention; [0033] FIG. 2 is a schematic of the voice wake module block diagram; 3 is to build word detection network diagram wake invention [0034] FIG;

式（1)可写为： [0046] W = aí.g max P{W)P(X \ W) C2) [0047] (4)唤醒词确认6 ：由于声学模型的本身存在不精确性以及实际使用环境的复杂 性,通过唤醒词检测环节得到的唤醒词不一定是真实的唤醒词。为了能降低非唤醒带来的 误唤醒以及后面会导致的问题，需要对检测得到的唤醒词作进一步的确认。本发明采用图 4的方式构建唤醒词确认网络9，唤醒词确认网络和唤醒词检测网络一样，都属于语音识别 中的识别网络，确认网络中只包含Garbage模型，可以使用简单的程序或手工构建。

[0048] 唤醒词确认的主要步骤如下： [0049] a)将唤醒词检测得到唤醒词解码到音素一级，并记录其所有的得分 (Scorephonel, Scoreph0ne2,...，Sc0reph_N)，其中N为唤醒词中总的音素个数。

[0050] b)使用和唤醒词检测同样的特征，在唤醒词确认网络上得到相应的声学得分，并 精确到帧一级（Scoreftamel，Scoreframe2,...，SCore&ameM)，其中M为该句特征总的时长，以帧为 单位。

[0051] c)计算得到唤醒词每个音素的确认得分，计算方式如下： [0052] ^iend CMphouej 二 [Scorephomi - ^ Scorefmmek) / (Kiend - Kistart)(3) [0053] 其中Kistot和Kimd分别为第i个音素的起始时间和结束时间。

[0054] d)计算得到该唤醒词的最终确认得分，计算方式如下所示： 2 CN [0055] CMword=- YjCMphonei (4) 丄 N /=i [0056] e)判断该唤醒词是否为真实的唤醒词，对比该唤醒词的最终确认得分和预先设定 的门限，如果确认得分CMwæd大于门限T则认为该唤醒词为真实的唤醒词，唤醒成功；如果 CMword小于门限T则认为该唤醒词为虚假的唤醒词，重新进行声学特征提取。

[0057] 通过以上工作实现语音唤醒功能，最终将结果反馈给唤醒执行3，执行唤醒操作。

[0058] 本发明工作过程如下： [0059] 第一、部署唤醒词定制模块11，实现训练声学模型8和构建唤醒词检测网络7，该 模型部署到服务器中，用户输入命令词文本，如“茄子茄子”，输出构建唤醒词检测网络7和 声学模型8。

[0060] 第二、通过编程语言实现语音唤醒模块，语音唤醒模块实现主要包含如下逻辑模 块（图2)，语音信号输入接口、语音信号的声学特征提取4、唤醒词检测5、唤醒词确认6和唤 醒执行3命令的输出。集成语音唤醒模块代码到操作系统模块10，开启运行，等待音频模块 13传入的录音数据。

[0035] FIG. 4 of the present invention to build network diagram wake-word confirmation.

DETAILED DESCRIPTION [0036] 1, the present invention starts the speech FIG camera system comprising: an operating system module 10, customization module wakeup word 11, camera module 12, audio module 13 and storage module 14, operating system modules 10 Run Wake voice module.

[0037] As shown in Figure 2, the voice wake-up module includes a voice input, voice wake Algorithm 2 and 3 wake execution. Voice wake Algorithm 2 realized mainly by acoustic feature extraction 4, 5 wake-word detection, Wake words confirm 6, construct wake word detection network 7, 8, and build the acoustic model training wake word Confirm Network 9 is completed, the specific implementation process are: [0038] (1) acoustic model training 8: training the acoustic model is divided into two parts, namely phoneme acoustic model and scrap models (ie Garbage models). Phoneme acoustic models using the traditional speech recognition acoustic model training method, select the appropriate database, use based on the following MLE (Maximum Likelihood Estimation, maximum likelihood estimation) and MPE (Minimum Phone Error, the smallest phoneme error) discriminative training guidelines get. Garbage model-independent voice except for absorbing wake of the word, using the same database and training phoneme model, each phoneme model by calculating the similarity between, each phoneme is divided into 20 categories, each category using the corresponding phoneme all training integrate data using MLE criteria Garbage corresponding training model, so that the model had 20 class Garbage. Garbage phoneme model using training data clustering mixed training, there are two purposes, in the wake of the word detection network to absorb the other voice except wake of the word, in the wake of the word used to calculate confirm Confirm network network score.

[0039] (2) acoustic feature extraction 4: 1 to obtain a voice through the voice input signal input, extracts may have a certain distinction in nature, and is based on the human auditory characteristics of the extracted features, generally used in the selected speech recognition MFCC (Mel-Frequency Cepstrum Coeff icient, Mel Frequency Cepstral) features.

Petitioner Exhibit 1004, Page 447 of 1052

[0061] 第三、进入拍照模块12，开启摄像头，等待操作系统模块10传递过来的启动拍照 指令，与此同时，音频模块13开启录音设备，并实时将录音数据输入到操作系统模块10，语 音唤醒模块在接收操作系统模块10中的录音数据后开始匹配运算。

[0062] 第四、当用户需要启动拍照时，说出音频模块13采集到“茄子茄子”的录音数据 后，语音唤模块法将发出启动指令，操作系统模块10向拍照模块发出拍照指令，并且向音 频模块13发出播报提示音指令，拍照模块12进入倒计时，调整焦点，开始拍照，并将采集到 的图像数据存储在存储模块14。

[0063] 本发明的技术核心是通过软件算法实现，成本低、模块移植方便，适合在各类消费 类电子产品上使用，具有很高的推广价值。

[0064] 本发明说明书中未作详细描述的内容属于本领域专业技术人员公知的现有技术。

[0065] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人 员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应 视为本发明的保护范围。

[0040] (3) Wake-word detection 5: The extracted acoustic features obtained using the acoustic model 8 in the wake of the 7 word detecting network computing acoustic score, if the score optimal path contains word to be detected wake, the detection of wake-up word, enter the next step; otherwise, the operation to re-extract acoustic features. To ensure the normal wake-up word that can be detected, while ineffective voice can be effectively absorbed. Build a network wake-up detector consists of a user-selected words and Garbage wake model composition, as shown in FIG. 2 CN, such a network is also referred to in the speech recognition to identify the network 3, the wake-up detector network structure is very simple, you can simple procedures or the manual build can be. Due to the complexity of the actual use environment, in many cases, received wake-up speech contaminated by noise, then wake speech characteristic corresponding to the acoustic model in the acoustic phoneme score is reduced a lot, and because the model is Garbage phoneme get mixed use more training, which itself is not very precise, acoustic characteristics of finite amplitude score lower on Garbage model, this time it was mistakenly wake absorbed Garbage voice model, the system wake-up rate will be reduced.

[0041] In order to prevent this from happening, when decoding in the wake word detection network, the arc where the decoding of Garbage do score a penalty, namely Penalty, we can not and phoneme acoustic model of fair competition, in order to ensure There is noise pollution wake normal voice can also be detected. Specific penalties amplitude modulation needs to be empirically m iF for different wake-up word. O [0042] wake word detection networks 7 is calculated using the best score path.

[0043] get using the classic formula for the optimal Bayesian scoring path, as follows: [0044] [0045] In the above formula X represents extracted from the input speech acoustic feature vector, W on behalf of the biggest score optimal word sequence.

Conditional probability P (XIW) for the acoustic model score, can be trained acoustic models and phoneme model waste calculated, the prior probability P (W) is the language model score, for this can be understood as different acoustic models added Penalty. P (X) for the whole probability, when the wake-up word acoustic

model and detect network determined that value, so the equation (1) can be written as: [0046] W = ai.g max P {W) P (X \ W) C2) [0047] (4) Wake words confirm 6: Due to the inaccuracy of the acoustic model itself and the complexity of the actual use of the environment, links were detected by word wake wakeup word is not necessarily real wake-up word. In order to reduce non-wake-up waking up and bring back will cause problems, it is necessary to detect the resulting wake word for further confirmation. The present invention is constructed by way of FIG. 4 Confirm the network wakeup word 9, Wake words confirm the network and detect network wake-word, as are speech recognition to identify the network, make sure the network Garbage contains only model, you can use a simple procedure or manual build.

[0048] The main steps are as follows confirmation wake word: [0049] a) word will wake detection decoding to get the wake-up word phoneme-level, and record all of its scoring (Scorephonel, Scoreph0ne2,..., Sc0reph_N), where N to wake up the word in the total number of phonemes.

The same characteristics [0050] b) Use and wake-up word is detected, the wake-up word to confirm the corresponding acoustic score on the network, and frame-accurate level (Scoreftamel, Scoreframe2,..., SCore and ameM), where M is The total length of the sentence characteristics, in units of frames.

[0051] c) is calculated for each phoneme recognition wake word score, calculated as follows: [0052] iend CMphouej two [Scorephomi - Scorefmmek) / (Kiend - Kistart) ? [0053] wherein Kistot and Kimd start time and end time are the i-th phoneme.

[0054] d) the calculated final confirmation of the wake-up word score, calculated as follows: 2 CN [0055] CMword = -YjCMphonei (4) Shang N / = i [0056] e) determine the wake-up word whether real wake-up word, compared to the wake-up word final score and a threshold set in advance, if it is confirmed the score CMwaed greater than the threshold T is that the wake-up word for real wake-up word, wake-up success; if CMword less than the threshold T is considered The wake-up word for word a false wake again acoustic feature extraction.

[0057] Through the above work to achieve voice wake-up function, the final results back to wake

execution 3, the wake-up operation.

[0058] operation of the present invention is as follows: [0059] First, deploy a custom module wakeup word 11, to achieve the acoustic model training and building 8 7 Wake word detection network, deploy the model to the server, the user enters the command word text such as "eggplant eggplant", output wake build word detection network 7 and 8 acoustic models.

[0060] Secondly, the programming language for voice wake-up module, voice wake-up module contains the following main logic module (Fig. 2), the acoustic characteristics of the speech signal input interface, speech signal extracted 4, 5 wake-word detection, wake Words confirm 6 and wake execute 3 commands output. Integrated Voice wake module code to the operating system module 10, turned to run, wait for the incoming audio module 13 recorded data.

[0061] third, into the camera module 12, open the camera, wait for the operating system module 10 is passed over the start photographing instruction, at the same time, an audio module 13 open recording equipment and record data in real time input to the operating system module 10, the speech after receiving the wake-up module operating system module 10 begin recording data matching operation.

[0062] Fourth, when the user needs to start taking pictures, say 13 collected audio module "eggplant eggplant" recording data, voice call modular approach will issue the start command, the operating system module 10 camera to camera module issue instruction, and to broadcast audio module 13 issues instructions tone, camera module 12 countdown, adjust the focus, started taking pictures and collecting image data stored in the memory module 14.

[0063] the core technology of the present invention is realized by software algorithms, low cost, easy to transplant the module is suitable for use on all types of consumer electronic products, promotion of high value.

[0064] The present description of the invention are not described in detail the contents of skilled professional and technical personnel belonging to well-known in the art.

[0065] The above are only preferred embodiments of the present invention, it should be noted that those of ordinary skill in the art, in the present invention without departing from the principle of the premise, you can also make a number of improvements and modifications These improvements and modifications are to be considered the scope of the invention.

Citing: CN101441869 A ; CN102404330 A ; CN102632846 A ; CN102682771 A ; WO0046969 A1 ;

Cited by: CN103714815 A ; CN104811595 A ; CN106020763 A ; CN106020763 B ; CN106098059 A ; CN106297777 A ; US10049662 BB ; WO17202016 A1 ; WO18086033 A1 ;

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103021409 A
(43)申请公布日 2013.04.03

(21)申请号 201210455215.3

(22)申请日 2012.11.13

(71)申请人 安徽科大讯飞信息科技股份有限公司
地址 230088 安徽省合肥市高新开发区望江西路 666 号

(72)发明人 朱国正 马冰 舒银东 潘志兵 周健 张凯 王海坤

(74)专利代理机构 北京科迪生专利代理有限责任公司 11251
代理人 成金玉 杨学明

(51)Int.Cl.
G10L 15/22(2006.01)
G10L 15/06(2013.01)
H04N 5/232(2006.01)

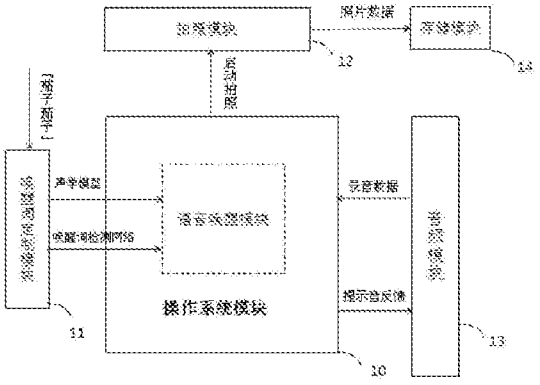
权利要求书 2 页 说明书 5 页 附图 2 页

(54)发明名称

一种语音启动拍照系统

(57)摘要

一种语音启动拍照系统，包括：操作系统模块(10)、唤醒词定制模块(11)、拍照模块(12)、音频模块(13)和存储模块(14)；其中：操作系统模块(10)中运行语音唤醒模块；所述语音唤醒模块包括语音输入(1)、语音唤醒算法(2)和唤醒执行(3)；语音唤醒算法(2)获取语音输入(1)的语音信号，进行语音唤醒处理后，将结果输出给唤醒执行(3)，从而完成唤醒操作；所述语音唤醒算法(2)通过声学特征提取(4)、唤醒词检测(5)、唤醒词确认(6)、构建唤醒词检测网络(7)、训练声学模型(8)和构建唤醒词确认网络(9)来实现。本发明可以根据用户需求进命令词定制，人机交换更人性化启动的准确度更高，抗干扰性更强。



CN 103021409 A

Petitioner Exhibit 1004, Page 452 of 1052

1. 一种语音启动拍照系统，其特征在于包括：操作系统模块(10)、唤醒词定制模块(11)、拍照模块(12)、音频模块(13)和存储模块(14)；其中：

操作系统模块(10)中运行语音唤醒模块；所述语音唤醒模块包括语音输入(1)、语音唤醒算法(2)和唤醒执行(3)；语音唤醒算法(2)获取语音输入(1)的语音信号，进行语音唤醒处理后，将结果输出给唤醒执行(3)，从而完成唤醒操作；所述语音唤醒算法(2)通过声学特征提取(4)、唤醒词检测(5)、唤醒词确认(6)、构建唤醒词检测网络(7)、训练声学模型(8)和构建唤醒词确认网络(9)来实现，具体实现过程如下：

第一步，声学特征提取(4)：通过语音输入(1)获取语音信号输入，提取具有区分性的、并且是基于人耳听觉特性提取的特征，通常选取语音识别中用到的MFCC（Mel-Frequency Cepstrum Coefficient，美尔频率倒谱系数）特征作为声学特征；

第二步，唤醒词检测(5)：将提取得到的声学特征，采用训练的声学模型(8)在唤醒词检测网络(7)上计算声学得分，如果声学得分最优的路径中包含要检测的唤醒词，则确定已检出唤醒词，进入第三步操作，否则回到第一步重新进行提取声学特征(4)；

第三步，唤醒词确认(6)：将提取得到的声学特征，采用训练的声学模型(8)在唤醒词确认网络(9)上进行唤醒词确认，得到最终确认得分；判断该检出的唤醒词是否为真实的唤醒词，即将该唤醒词的最终确认得分和预先设定的门限，如果最终确认得分大于等于门限，则认为该唤醒词是真实的唤醒词，语音唤醒成功，将结果输出给唤醒执行(3)，从而完成语音唤醒操作；如果最终确认得分小于门限，则认为该唤醒词为虚假的唤醒词，重新回到第一步重新进行声学特征提取(4)；

唤醒词定制模块(11)，接收用户输入的唤醒词，输出操作系统模块(10)中语音唤醒模块所需的声学模型(8)和唤醒词检测网络(7)；

拍照模块(12)，具有拍照功能的系统模块，接收操作系统模块(10)的拍照指令，并在拍照后将图像的数字信号发送到存储模块(14)；

音频模块(13)，具有录音采集和音频播报功能，录音功能主要是采集环境中音频数据，并传送到操作系统模块(10)，播音功能主要是接收系统模块(10)发出提示反馈后，播报提示音功能；

存储模块(14)，具有永久存储能力的存储介质，对拍照模块生成的数字图像信息进行存储。

2. 根据权利要求1所述的语音启动拍照系统，其特征在于：所述声学模型(8)的训练分为两部分，分别为音素声学模型和废料模型(即Garbage模型)；音素声学模型采用传统的语音识别中的声学模型训练方法，选取数据库，利用基于MLE（Maximum Likelihood Estimation，最大似然估计）和MPE（Minimum Phone Error，最小音素错误）区分性训练准则下得到；Garbage模型用于吸收除唤醒词之外的无关语音，使用和训练音素模型同样的数据库，通过计算各音素模型之间的相似性，将各音素分为20类，使用每类音素对应的所有训练数据融合起来，采用MLE准则训练对应的Garbage模型，便得到20类Garbage模型。

3. 根据权利要求1所述的语音启动拍照系统，其特征在于：所述唤醒词检测网络(7)的实现采用最优得分路径计算得出，所述最优得分路径的计算公式是：

$$W = \arg\max_{W} P(W)P(X|W)$$

Petitioner Exhibit 1004, Page 453 of 1052

其中 X 代表从输入语音中提取的声学特征向量，W 代表得分最大的最优词序列；条件概率 P(X|W) 为声学模型得分，通过训练好的声学模型(8)计算得到；先验概率 P(W) 为语言模型得分，即为对不同的声学模型所加的 PenaltyP(X)为全概率，当声学模型和唤醒词检测网络确定下来后即是定值。

4. 根据权利要求 1 所述的语音启动拍照系统，其特征在于：所述唤醒词确认网络(9)实现是：

a. 将检出的唤醒词解码到音素一级，并记录所有的得分($Score_{phone1}$, $Score_{phone2}$, …, $Score_{phoneN}$)，其中 N 为唤醒词中总的音素个数，$Score_{phone1}$, $Score_{phone2}$, …, $Score_{phoneN}$ 分别表示该唤醒词中所有音素的是解码得分，其中下标表示音素的 N 个音素的标识；

b. 使用和唤醒词检测同样的特征，得到相应的声学得分，并精确到帧一级($Score_{frame1}$, $Score_{frame2}$, …, $Score_{frameM}$)，其中 M 为该句特征总的时长，以帧为单位；

c. 计算得到唤醒词每个音素的确认得分，计算方式如下：

$$CM_{phonei} = (Score_{phonei} - \sum_{k=K_{istari}}^{K_{iend}} Score_{framek}) / (K_{iend} - K_{istari})$$

其中 K_{istart} 和 K_{iend} 分别为第 i 个音素的起始时间和结束时间；

CM_{phonei} 表示第 i 个音素的确认得分，下标 phonei 表示第 i 个音素，$Score_{phonei}$ 如上面所示第 i 个 phone 的解码得分，$Score_{framek}$ 表示使用唤醒词确认网络解码得到的第 k 帧的得分；

d. 计算得到该唤醒词的最终确认得分，计算方式如下所示：

$$CM_{word} = \frac{1}{N} \sum_{i=1}^{N} CM_{phonei}$$

5. 根据权利要求 1 所述的语音启动拍照系统，其特征在于：所述操作系统模块(10)中的语音唤醒模块能够被移植到被移植到 Windows、Linux、Android、IOS 系统平台。

Petitioner Exhibit 1004, Page 454 of 1052

一种语音启动拍照系统

技术领域

[0001] 本发明涉及一种语音启动拍照系统，可以用于含有拍照功能的消费类电子产品领域。

背景技术

[0002] 随着消费类电子产品快速的发展，各种包含拍照功能的设备使用越来越频繁，人们对拍照的人性化要求越越来越高，语音作为人类最基本的方式，将语音识别技术应用到消费类电子产品中，实现通过自然语音来控制此类产品功能是未来发展的趋势。

发明内容

[0003] 本发明技术解决问题：提供一种语音启动拍照系统，克服了现有按钮式、触控式启动方式带来的不便利，操作方式受限和拍照抖动等不足，具有操作简单更人性化，启动拍照不受限，防抖动等优点。现有的一些语音启动技术，是利用特点语音如口哨，误启动率搞，输入语音命令单一等缺陷，本发明可以根据用户需求进命令词定制，人机交换更人性化启动的准确度更高，抗干扰性更强。

[0004] 本发明是通过以下技术方案实现的：一种语音启动拍照系统，包括：操作系统模块 10、唤醒词定制模块 11、拍照模块 12、音频模块 13 和存储模块 14；其中：

[0005] 操作系统模块 10 中运行语音唤醒模块；所述语音唤醒模块包括语音输入 1、语音唤醒算法 2 和唤醒执行 3；语音唤醒算法 2 获取语音输入 1 的语音信号，进行语音唤醒处理后，将结果输出给唤醒执行 3，从而完成唤醒操作；所述语音唤醒算法 2 通过声学特征提取 4、唤醒词检测 5、唤醒词确认 6、构建唤醒词检测网络 7、训练声学模型 8 和构建唤醒词确认网络 9 来实现，具体实现过程如下：

[0006] 第一步，声学特征提取 4：通过语音输入 1 获取语音信号输入，提取具有区分性的、并且是基于人耳听觉特性提取的特征，通常选取语音识别中用到的 MFCC（Mel-Frequency Cepstrum Coefficient，美尔频率倒谱系数）特征作为声学特征；

[0007] 第二步，唤醒词检测 5：将提取得到的声学特征，采用训练的声学模型 8 在唤醒词检测网络 7 上计算声学得分，如果声学得分最优的路径中包含要检测的唤醒词，则确定已检出唤醒词，进入第三步操作，否则回到第一步重新进行提取声学特征 4；

[0008] 第三步，唤醒词确认 6：将提取得到的声学特征，采用训练的声学模型 8 在唤醒词确认网络 9 上进行唤醒词确认，得到最终确认得分；判断该检出的唤醒词是否为真实的唤醒词，即将该唤醒词的最终确认得分和预先设定的门限，如果最终确认得分大于等于门限，则认为该唤醒词是真实的唤醒词，语音唤醒成功，将结果输出给唤醒执行 3，从而完成语音唤醒操作；如果最终确认得分小于门限，则认为该唤醒词为虚假的唤醒词，重新回到第一步重新进行声学特征提取 4；

[0009] 唤醒词定制模块 11，语音唤醒模块的资源生成模块接收用户输入的唤醒词，输出操作系统模块 10 中语音唤醒模块所需的声学模型 8 和唤醒词检测网络 7；

Petitioner Exhibit 1004, Page 455 of 1052

[0010] 拍照模块 12，具有拍照功能的系统模块，接收操作系统模块 10 的拍照指令，并在拍照后将图像的数字信号发送到存储模块 14；

[0011] 音频模块 13，具有录音采集和音频播报功能，录音功能主要是采集环境中音频数据，并传送到操作系统模块 10，播音功能主要是接收系统模块 10 发出提示反馈后，播报提示音功能；

[0012] 存储模块 14，具有永久存储能力的存储介质，对拍照模块生成的数字图像信息进行存储。

[0013] 所述声学模型 8 的训练分为两部分，分别为音素声学模型和废料模型（即 Garbage 模型）；音素声学模型采用传统的语音识别中的声学模型训练方法，选取数据库，利用基于 MLE（Maximum Likelihood Estimation，最大似然估计）和 MPE（Minimum Phone Error，最小音素错误）区分性训练准则下得到；Garbage 模型用于吸收除唤醒词之外的无关语音，使用和训练音素模型同样的数据库，通过计算各音素模型之间的相似性，将各音素分为 20 类，使用每类音素对应的所有训练数据融合起来，采用 MLE 准则训练对应的 Garbage 模型，便得到 20 类 Garbage 模型。

[0014] 所述唤醒词检测网络 7 的实现采用最优得分路径计算得出，所述最优得分路径的计算公式是：

[0015] $$W = \arg\max_{W} P(W)P(X|W)$$

[0016] 其中 X 代表从输入语音中提取的声学特征向量，W 代表得分最大的最优词序列；条件概率 P(X|W) 为声学模型得分，通过训练好的声学模型 8 计算得到；先验概率 P(W) 为语言模型得分，即为对不同的声学模型所加的 PenaltyP(X) 为全概率，当声学模型和唤醒词检测网络确定下来后即是定值。

[0017] 所述唤醒词确认网络 9 实现是：

[0018] a. 将检出的唤醒词解码到音素一级，并记录所有的得分（$Score_{phone1}$，$Score_{phone2}$，…，$Score_{phoneN}$），其中 N 为唤醒词中总的音素个数，$Score_{phone1}$，$Score_{phone2}$，…，$Score_{phoneN}$ 分别表示该唤醒词中所有音素的是解码得分，其中下标表示音素的 N 个音素的标识；

[0019] b. 使用和唤醒词检测同样的特征，得到相应的声学得分，并精确到帧一级（$Score_{frame1}$，$Score_{frame2}$，…，$Score_{frameM}$），其中 M 为该句特征总的时长，以帧为单位；

[0020] c. 计算得到唤醒词每个音素的确认得分，计算方式如下：

[0021] $$CM_{phonei} = (Score_{phonei} - \sum_{k=K_{istart}}^{K_{iend}} Score_{framek}) / (K_{iend} - K_{istart})$$

[0022] 其中 K_{istart} 和 K_{iend} 分别为第 i 个音素的起始时间和结束时间；

[0023] CM_{phonei} 表示第 i 个音素的确认得分，下标 phonei 表示第 i 个音素，$Score_{phonei}$ 如上面所示第 i 个 phone 的解码得分，$Score_{framek}$ 表示使用唤醒词确认网络解码得到的第 k 帧的得分；

[0024] d. 计算得到该唤醒词的最终确认得分，计算方式如下所示：

[0025] $$CM_{word} = \frac{1}{N}\sum_{i=1}^{N} CM_{phonei}$$

[0026] 所述操作系统模块中的语音唤醒模块能够被移植到被移植到 Windows、Linux、

Petitioner Exhibit 1004, Page 456 of 1052

Android、IOS 系统平台。

[0027] 本发明与现有技术相比的优点在于：

[0028] (1) 本发明通过用户特定的语音唤醒词作为触发源，即使是在嘈杂的环境下，也无需用户利用双手操作，仅通过语音命令快速实现唤醒功能，进行下一步交互操作。

[0029] (2) 本发明带来的价值是，使用本系统后可通过说约定的语音唤醒词开启语音唤醒功能，无需事先暂停音频播放，同时通过实际测试验证，正确识别唤醒率可以达到 90% 以上。

[0030] (3) 本发明实现，成本低，代码移植方便，具有很好的推广应用价值。

[0031] (4) 本发明中的语音唤醒功能均通过软件算法实现，并且算法可以很方便的被移植到 Windows、Linux、Android、IOS 等系统平台。

附图说明

[0032] 图 1 为本发明的结构示意图；

[0033] 图 2 为本发明中的语音唤醒模块实现框图；

[0034] 图 3 为本发明的构建唤醒词检测网络示意图；

[0035] 图 4 为本发明的构建唤醒词确认网络示意图。

具体实施方式

[0036] 如图 1 所示，本发明的语音启动拍照系统包括：操作系统模块 10、唤醒词定制模块 11、拍照模块 12、音频模块 13 和存储模块 14、操作系统模块 10 中运行语音唤醒模块。

[0037] 如图 2 所示，语音唤醒模块包括语音输入 1、语音唤醒算法 2 和唤醒执行 3。语音唤醒算法 2 实现主要由声学特征提取 4、唤醒词检测 5、唤醒词确认 6、构建唤醒词检测网络 7、训练声学模型 8 和构建唤醒词确认网络 9 完成，具体实现过程是：

[0038] (1) 训练声学模型 8：声学模型的训练分为两部分，分别为音素声学模型和废料模型(即 Garbage 模型)。音素声学模型采用传统的语音识别中的声学模型训练方法，选取合适的数据库，利用基于 MLE(Maximum Likelihood Estimation，最大似然估计)和 MPE(Minimum Phone Error，最小音素错误) 区分性训练准则下的得到。Garbage 模型用于吸收除唤醒词之外的无关语音，使用和训练音素模型同样的数据库，通过计算各音素模型之间的相似性，将各音素分为 20 类，使用每类音素对应的所有训练数据融合起来，采用 MLE 准则训练对应的 Garbage 模型，这样即得到 20 类 Garbage 模型。Garbage 模型采用了聚类的音素训练数据混合训练，有两种用途，在唤醒词检测网络中用来吸收除唤醒词之外的其他语音，在唤醒词确认网络中用来计算确认网络的得分。

[0039] (2) 声学特征提取 4：通过语音输入 1 获取语音信号输入，提取可以具有一定区分性的，并且是基于人耳听觉特性提取的特征，一般选取语音识别中用到的 MFCC (Mel-Frequency Cepstrum Coefficient，美尔频率倒谱系数) 特征。

[0040] (3) 唤醒词检测 5：将提取得到的声学特征，使用声学模型 8 在唤醒词检测网络 7 上计算声学得分，如果得分最优的路径中包含要检测的唤醒词，则检出唤醒词，进入下一步操作；否则重新提取声学特征操作。为了保证唤醒词能够被正常的检出，同时无效的语音又能被有效的吸收。唤醒检测网络的构建主要由用户选定的唤醒词和 Garbage 模型组成，如

Petitioner Exhibit 1004, Page 457 of 1052

图 3 所示，这种网络在语音识别中也称为识别网络，由于唤醒检测网路结构非常简单，可以通过简单的程序或则手工构建都可以。由于实际使用环境的复杂性，在很多情况下，接收到的唤醒语音被噪声污染了，此时唤醒语音对应的声学的特征在音素声学模型上的得分就会降低的很多，而由于 Garbage 模型是使用较多音素混合训练得到，其本身不是很精确，声学特征在 Garbage 模型上的得分降低的幅度有限，此时唤醒语音就被 Garbage 模型误吸收，系统唤醒率就会降低。

[0041] 为了防止上述情况的发生，在唤醒词检测网络上解码时，对 Garbage 所在的弧的解码得分做一定的惩罚，即 Penalty，使其不能和音素声学模型公平竞争，以保证有被噪声污染的唤醒语音也能被正常检出。具体的惩罚幅度需要针对不同的唤醒词作经验性的调整。

[0042] 唤醒词检测网络 7 的实现方法是采用最优得分路径计算得出。

[0043] 最优得分路径的获取采用经典的贝叶斯公式，如下所示：

[0044]

$$W = \arg\max_{W} P(W \mid X) = \arg\max_{W} \frac{P(W)P(X \mid W)}{P(X)} \qquad (1)$$

[0045] 上式中 X 代表从输入语音中提取的声学特征向量，W 代表得分最大的最优词序列。条件概率 P(X|W) 为声学模型得分，可以通过训练好的音素声学模型和废料模型计算得到，先验概率 P(W) 为语言模型得分，这里可以理解为对不同的声学模型所加的 Penalty。P(X) 为全概率，当声学模型和唤醒词检测网络确定下来后就是定值，因此公式(1)可写为：

[0046] $$W = \arg\max_{W} P(W)P(X \mid W) \qquad (2)$$

[0047] (4) 唤醒词确认 6：由于声学模型的本身存在不精确性以及实际使用环境的复杂性，通过唤醒词检测环节得到的唤醒词不一定是真实的唤醒词。为了能降低非唤醒带来的误唤醒以及后面会导致的问题，需要对检测得到的唤醒词作进一步的确认。本发明采用图 4 的方式构建唤醒词确认网络 9，唤醒词确认网络和唤醒词检测网络一样，都属于语音识别中的识别网络，确认网络中只包含 Garbage 模型，可以使用简单的程序或手工构建。

[0048] 唤醒词确认的主要步骤如下：

[0049] a) 将唤醒词检测得到唤醒词解码到音素一级，并记录其所有的得分 ($Score_{phone1}, Score_{phone2}, \cdots, Score_{phoneN}$)，其中 N 为唤醒词中总的音素个数。

[0050] b) 使用和唤醒词检测同样的特征，在唤醒词确认网络上得到相应的声学得分，并精确到帧一级 ($Score_{frame1}, Score_{frame2}, \cdots, Score_{frameM}$)，其中 M 为该句特征总的时长，以帧为单位。

[0051] c) 计算得到唤醒词每个音素的确认得分，计算方式如下：

[0052]

$$CM_{phonei} = (Score_{phonei} - \sum_{k=K_{istart}}^{K_{iend}} Score_{framek}) / (K_{iend} - K_{istart}) \qquad (3)$$

[0053] 其中 K_{istart} 和 K_{iend} 分别为第 i 个音素的起始时间和结束时间。

[0054] d) 计算得到该唤醒词的最终确认得分，计算方式如下所示：

Petitioner Exhibit 1004, Page 458 of 1052

[0055] $$CM_{word}=\frac{1}{N}\sum_{i=1}^{N}CM_{phonei} \quad (4)$$

[0056] e）判断该唤醒词是否为真实的唤醒词，对比该唤醒词的最终确认得分和预先设定的门限，如果确认得分 CM_{word} 大于门限 T 则认为该唤醒词为真实的唤醒词，唤醒成功；如果 CM_{word} 小于门限 T 则认为该唤醒词为虚假的唤醒词，重新进行声学特征提取。

[0057] 通过以上工作实现语音唤醒功能，最终将结果反馈给唤醒执行 3，执行唤醒操作。

[0058] 本发明工作过程如下：

[0059] 第一、部署唤醒词定制模块 11，实现训练声学模型 8 和构建唤醒词检测网络 7，该模型部署到服务器中，用户输入命令词文本，如“茄子茄子”，输出构建唤醒词检测网络 7 和声学模型 8。

[0060] 第二、通过编程语言实现语音唤醒模块，语音唤醒模块实现主要包含如下逻辑模块（图 2），语音信号输入接口、语音信号的声学特征提取 4、唤醒词检测 5、唤醒词确认 6 和唤醒执行 3 命令的输出。集成语音唤醒模块代码到操作系统模块 10，开启运行，等待音频模块 13 传入的录音数据。

[0061] 第三、进入拍照模块 12，开启摄像头，等待操作系统模块 10 传递过来的启动拍照指令，与此同时，音频模块 13 开启录音设备，并实时将录音数据输入到操作系统模块 10，语音唤醒模块在接收操作系统模块 10 中的录音数据后开始匹配运算。

[0062] 第四、当用户需要启动拍照时，说出音频模块 13 采集到“茄子茄子”的录音数据后，语音唤模块法将发出启动指令，操作系统模块 10 向拍照模块发出拍照指令，并且向音频模块 13 发出播报提示音指令，拍照模块 12 进入倒计时，调整焦点，开始拍照，并将采集到的图像数据存储在存储模块 14。

[0063] 本发明的技术核心是通过软件算法实现，成本低、模块移植方便，适合在各类消费类电子产品上使用，具有很高的推广价值。

[0064] 本发明说明书中未作详细描述的内容属于本领域专业技术人员公知的现有技术。

[0065] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

Petitioner Exhibit 1004, Page 459 of 1052

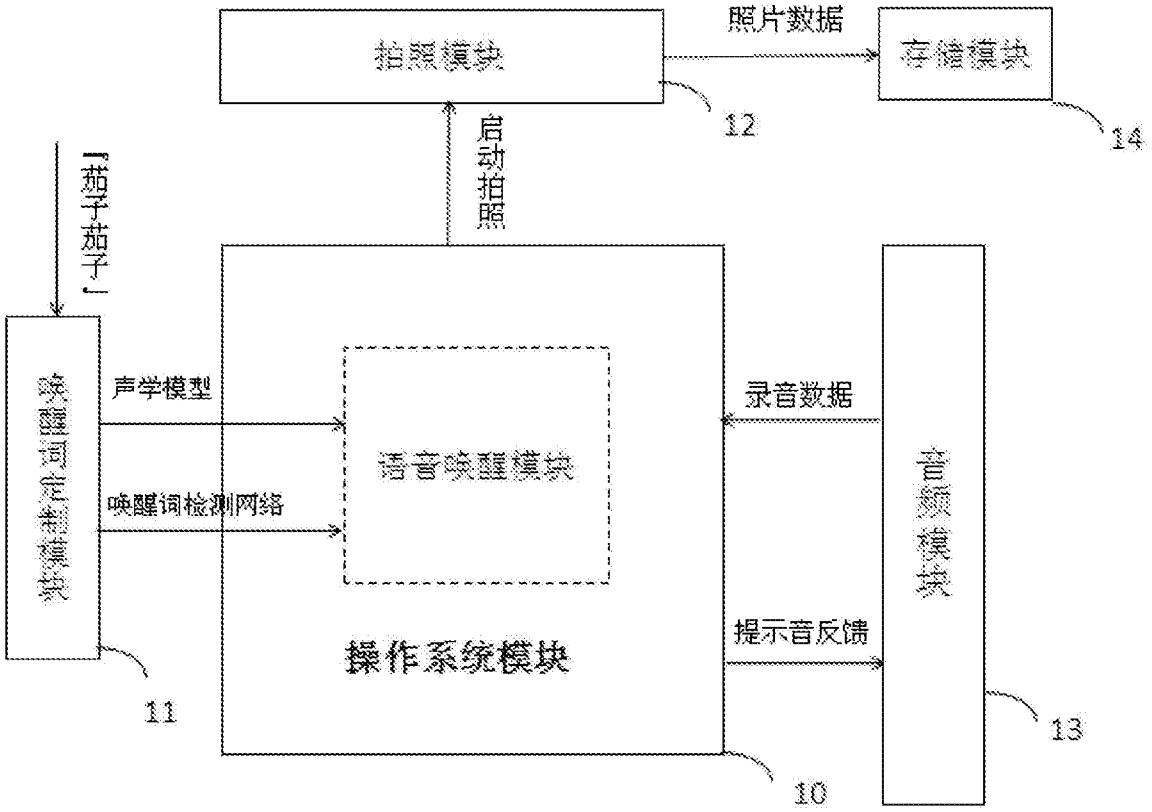


图 1

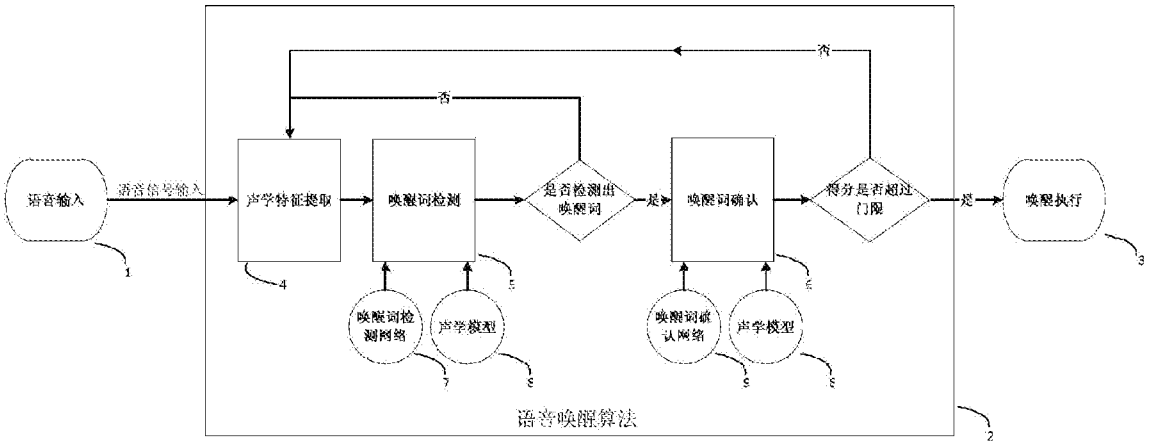


图 2

Petitioner Exhibit 1004, Page 460 of 1052

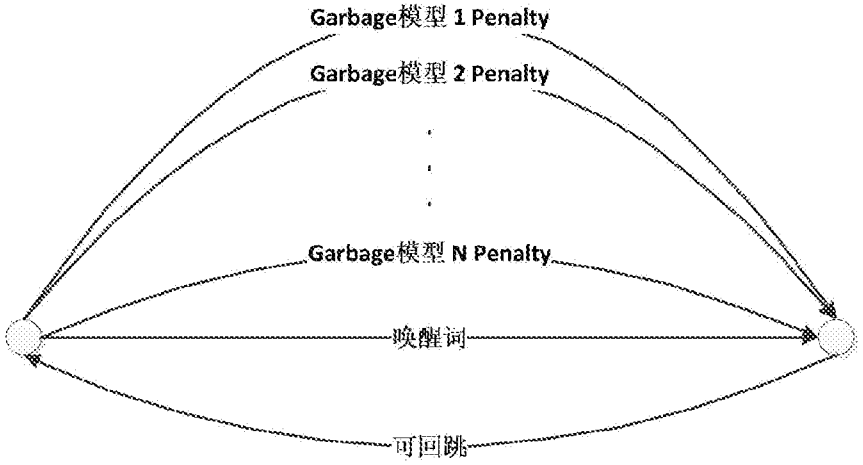


图 3

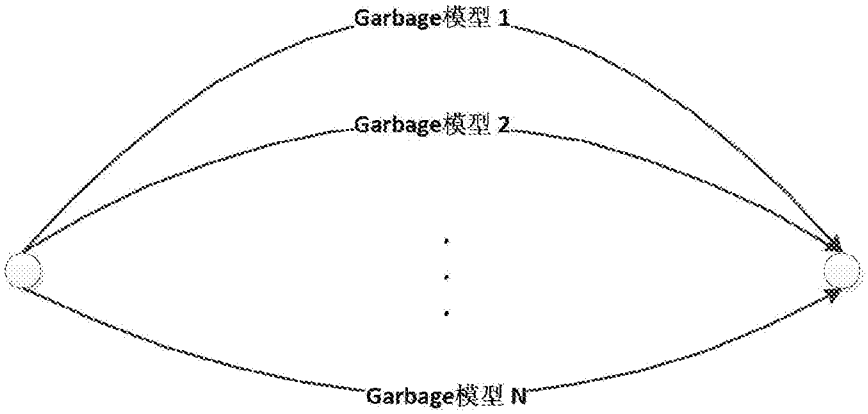


图 4

Petitioner Exhibit 1004, Page 461 of 1052

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 899 955 A1

(12) EUROPEAN PATENT APPLICATION

published in accordance with Art. 153(4) EPC

(43) Date of publication:
29.07.2015 Bulletin 2015/31

(21) Application number: **14814665.7**

(22) Date of filing: **01.08.2014**

(51) Int Cl.:
H04M 1/725 *(2006.01)* ***G10L 17/00*** *(2013.01)*

(86) International application number:
PCT/CN2014/083507

(87) International publication number:
WO 2015/070644 (21.05.2015 Gazette 2015/20)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME

(30) Priority: **15.11.2013 CN 201310574762**

(71) Applicant: **Huawei Device Co., Ltd.**
Longgang District
Shenzhen
Guangdong 518129 (CN)

(72) Inventors:
- **WANG, Xiyong**
  Shenzhen
  Guangdong 518129 (CN)
- **JIANG, Hongrui**
  Shenzhen
  Guangdong 518129 (CN)
- **ZHENG, Weijun**
  Shenzhen
  Guangdong 518129 (CN)
- **WANG, Qing**
  Shenzhen
  Guangdong 518129 (CN)

(74) Representative: **Körber, Martin Hans**
Mitscherlich PartmbB
Patent- und Rechtsanwälte
Sonnenstrasse 33
80331 München (DE)

(54) TERMINAL VOICE CONTROL METHOD, DEVICE, AND TERMINAL

(57) Embodiments of the present invention disclose a terminal voice control method. The method includes: when a voice instruction of a user is received, identifying the voice instruction, and extracting wake-up indication information from the voice instruction; performing voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extracting processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information. According to the terminal voice control method in the embodiments of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

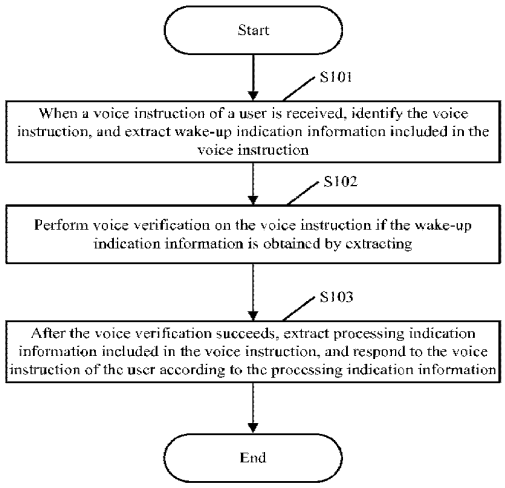


FIG. 1

EP 2 899 955 A1

Printed by Jouve, 75001 PARIS (FR)

Description

[0001] This application claims priority to Chinese Patent Application No. CN 201310574762.8, filed with the Chinese Patent Office on November 15, 2013 and entitled "TERMINAL VOICE CONTROL METHOD AND APARETUS, AND TERMINAL", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present invention relates to the field of terminal technologies, and in particular, to a terminal voice control method and apparatus, and a terminal.

BACKGROUND

[0003] With the development of science and technology, smartphones have become necessities in people's lives. In order to facilitate a user to perform an operation on a smartphone, most smartphones currently have a voice control function. A user can issue a voice instruction to a mobile phone, and the mobile phone makes a response to the voice instruction. However, in the prior art, the foregoing solution is implemented by separately invoking different voice engines for multiple times. For example, a voice wake-up engine and a voiceprint recognition engine are used to identify whether a user is the owner first, and then the mobile phone is woken up, and the mobile phone then inquires about a requirement of the user and directs the user to answer questions step by step. After the user answers the questions, the mobile phone identifies a voice instruction that is issued by the user when the user answers the questions, and executes the voice instruction of the user according to an identification result.

[0004] A defect of the foregoing voice control manner is that: multiple voice instructions are inconsecutive, resulting in that a user needs to be inquired step by step and directed to answer questions; a next step can be started only after the user finishes answering the questions; a process of performing each step is cumbersome, which causes great inconvenience for the user to use a voice control function of a mobile phone. When the mobile phone is standby, no instruction can be directly issued to the mobile phone. The mobile phone needs to be woken up before a further operation can be performed on the mobile phone.

SUMMARY

[0005] According to a terminal voice control method provided in embodiments of the present invention, one voice instruction may be used to control a terminal to complete multiple tasks.

[0006] According to a first aspect, an embodiment of the present invention provides a terminal voice control method, including:

when a voice instruction of a user is received, identifying the voice instruction, and extracting wake-up indication information from the voice instruction;

performing voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and

extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

[0007] With reference to the first aspect, in a first possible implementation manner of the first aspect, the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information includes:

after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, lighting up a terminal screen and unlocking the terminal screen, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information; and

after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

[0008] With reference to the first aspect, in a third possible implementation manner of the first aspect, the method further includes:

if no wake-up indication information is obtained by extracting, continuing to wait for receiving a voice instruction of the user.

[0009] With reference to the first aspect, in a fourth possible implementation manner of the first aspect, the method further includes:

prompting reverification if the voice verification fails, and stopping receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0010] With reference to the first aspect, in a fifth possible implementation manner of the first aspect, the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information includes:

removing the wake-up indication information from

Petitioner Exhibit 1004, Page 463 of 1052

the voice instruction, extracting the processing prompt information in the voice instruction, and converting the processing indication information into a text; and
identifying a meaning of the processing prompt information according to the text, and making a response according to the obtained meaning.

[0011] According to a second aspect, an embodiment of the present invention provides a terminal voice control apparatus, including:

a receiving unit, configured to, when a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction;
a first extracting unit, configured to perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and
a first processing unit, configured to, after the voice verification succeeds, extract processing indication information included in the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0012] With reference to the second aspect, in a first possible implementation manner of the second aspect, the first processing unit includes:

a first executing module, configured to, after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information; and
a second executing module, configured to, after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0013] With reference to the second aspect, in a third possible implementation manner of the second aspect, the apparatus further includes:

a second extracting unit, configured to, if no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of the user.

[0014] With reference to the second aspect, in a fourth possible implementation manner of the second aspect, the apparatus includes:

a second processing unit, configured to prompt reverification when the voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0015] With reference to the second aspect, in a fifth possible implementation manner of the second aspect, the first processing unit includes:

a converting module, configured to remove the wake-up indication information from the voice instruction, extract the processing prompt information in the voice instruction, and convert the processing indication information into a text; and
a responding module, configured to identify a meaning of the processing prompt information according to the text, and make a response according to the obtained meaning.

[0016] According to a third aspect, an embodiment of the present invention further provides a terminal, where the terminal includes an audio monitoring unit and a processor, where:

the audio monitoring unit is configured to pick up a voice instruction issued by a user; and
the processor is configured to, when the voice instruction of the user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction; perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0017] According to the terminal voice control method in the embodiments of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

BRIEF DESCRIPTION OF DRAWINGS

[0018] To describe the technical solutions in the embodiments of the present invention or in the prior art more clearly, the following briefly introduces the accompanying drawings required for describing the embodiments or the prior art. Apparently, the accompanying drawings in the following description show some embodiments of the present invention, and a person of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

FIG. 1 is a flowchart of a terminal voice control meth-

Petitioner Exhibit 1004, Page 464 of 1052

od according to a first embodiment of the present invention;

FIG. 2 is a flowchart of a terminal voice control method according to a second embodiment of the present invention;

FIG. 3 is a flowchart of a terminal voice control method according to a third embodiment of the present invention;

FIG. 4 is a schematic structural diagram of a terminal voice control apparatus according to a fourth embodiment of the present invention;

FIG. 5 is a schematic structural diagram of a terminal voice control apparatus according to a fifth embodiment of the present invention;

FIG. 6 is a schematic structural diagram of a terminal voice control apparatus according to a sixth embodiment of the present invention; and

FIG. 7 is a schematic structural diagram of a terminal according to a seventh embodiment of the present invention.

DESCRIPTION OF EMBODIMENTS

[0019] The following clearly and completely describes the technical solutions in the embodiments of the present invention with reference to the accompanying drawings in the embodiments of the present invention. Apparently, the described embodiments are a part rather than all of the embodiments of the present invention. All other embodiments obtained by persons of ordinary skill in the art based on the embodiments of the present invention without creative efforts shall fall within the protection scope of the present invention.

[0020] Refer to FIG. 1, which is a flowchart of a terminal voice control method according to a first embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The method specifically includes:

[0021] S101. When a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction.

[0022] Specifically, the voice instruction of the user includes the wake-up indication information and processing indication information, where the wake-up indication information is used to wake up a terminal that is in a screen locking state, and the processing indication information is used to instruct the terminal to execute a specific action.

[0023] Identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a word preset by the user, such as hi or hello, so as to extract the wake-up indication information included in the voice instruction of the user.

[0024] S102. Perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting.

[0025] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0026] It should be understood that voiceprint recognition (Voiceprint Recognition) is also referred to as speaker recognition, which refers to identifying, according to a voice feature of a person, who speaks a segment of voice. Generally, voiceprint recognition includes two aspects: speaker identifying and speaker determining, where the former is to determine which one of several persons speaks a segment of voice, and the latter is to determine whether a segment of voice is spoken by a specified person. The voiceprint recognition manner in this embodiment of the present invention is speaker determining, that is, whether a segment of voice is spoken by a preset specified person needs to be identified.

[0027] S103. After the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0028] Specifically, after the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing prompt information is converted into a text; and a meaning of the processing indication information is identified according to the text, and a response is made according to the obtained meaning.

[0029] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0030] Refer to FIG. 2, which is a flowchart of a terminal voice control method according to a second embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The method specifically includes:

[0031] S201. When a voice instruction of a user is received, identify the voice instruction.

[0032] Specifically, identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a preset word such as hi or hello, so as to extract wake-up indication information from the voice instruction of the user.

[0033] S202. Perform voice verification on the voice instruction if wake-up indication information is obtained by extracting.

[0034] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine

Petitioner Exhibit 1004, Page 465 of 1052

whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0035] S203. After the voice verification succeeds, if it is detected that the terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0036] S204. After the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0037] Specifically, if it is detected that the terminal is in a screen locking state, the terminal that is in a screen locking state needs to be woken up first according to the wake-up indication information in the voice instruction, and then another step is performed. If the terminal is not in a screen locking state, the step of lighting up the terminal screen or unlocking the terminal screen does not need to be performed, and a next step is directly performed.

[0038] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0039] Refer to FIG. 3, which is a flowchart of a terminal voice control method according to a third embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The method specifically includes:

[0040] S301. When a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction.

[0041] Specifically, the voice instruction of the user includes the wake-up indication information and processing indication information, where the wake-up indication information is used to wake up a terminal that is in a screen locking state, and the processing indication information is used to instruct the terminal to execute a specific action.

[0042] Identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a preset word such as hi or hello, so as to extract the wake-up indication information included in the voice instruction of the user.

[0043] S302. Perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting.

[0044] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0045] S303. If no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of the user.

[0046] Specifically, if no wake-up indication information is obtained by extracting, the terminal does not perform any processing, and continues to wait for receiving a voice instruction of the user.

[0047] S304. After the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0048] Specifically, after the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing indication information is converted into a text; and a meaning of the processing prompt information is identified according to the text, and a response is made according to the obtained meaning.

[0049] S305. Prompt reverification if the voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0050] Specifically, if the number of times of verification exceeds the preset threshold but the verification still fails, a voice instruction of the user stops being received, and the user can only turn on a terminal screen manually, thereby ensuring security of using the terminal.

[0051] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

[0052] Refer to FIG. 4, which is a schematic structural diagram of a terminal voice control apparatus according to a fourth embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The apparatus 1 specifically includes a receiving unit 10, a first extracting unit 20, and a first processing unit 30.

[0053] The receiving unit 10 is configured to, when a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction.

[0054] Specifically, the voice instruction of the user includes the wake-up indication information and processing indication information, where the wake-up indication information is used to wake up a terminal that is in a screen locking state, and the processing indication information is used to instruct the terminal to execute a specific action.

Petitioner Exhibit 1004, Page 466 of 1052

[0055] Identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a preset wake-up word such as hi or hello, so as to extract the wake-up indication information included in the voice instruction of the user.

[0056] The first extracting unit 20 is configured to perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting.

[0057] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0058] The first processing unit 30 is configured to, after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0059] The first processing unit 30 includes a converting module 31 and a responding module 32.

[0060] The converting module 31 is configured to remove the wake-up indication information from the voice instruction, extract the processing prompt information in the voice instruction, and convert the processing indication information into a text.

[0061] The responding module 32 is configured to identify a meaning of the processing prompt information according to the text, and make a response according to the obtained meaning.

[0062] Specifically, after the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing indication information is converted into a text; and a meaning of the processing prompt information is identified according to the text, and a response is made according to the obtained meaning.

[0063] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0064] Refer to FIG. 5, which is a schematic structural diagram of a terminal voice control apparatus according to a fifth embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The apparatus 2 is an optimization of the apparatus 1 in FIG. 4. A first processing unit 30 of the apparatus 2 includes a first executing module 33 and a second executing module 34.

[0065] The first executing module 33 is configured to, after voice verification succeeds, if it is detected that the terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0066] The second executing module 34 is configured to, after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0067] Specifically, if it is detected that the terminal is in a screen locking state, the step of lighting up the terminal screen and unlocking the terminal screen needs to be performed; if the terminal is not in a screen locking state, the step of lighting up the terminal screen and unlocking the terminal screen does not need to be performed.

[0068] In addition, performing voice verification on the voice instruction is using a voiceprint recognition method to verify whether the user who issues the voice instruction is a preset user, thereby improving security of the terminal.

[0069] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0070] Refer to FIG. 6, which is a schematic structural diagram of a terminal voice control apparatus according to a sixth embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The apparatus 3 is an optimization of the apparatus 1 in FIG. 4. The apparatus 3 further includes a second extracting unit 40 and a second processing unit 50.

[0071] The second extracting unit 40 is configured to, if no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of a user.

[0072] The second processing unit 50 is configured to prompt reverification when voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0073] Specifically, if the number of times of voice verification exceeds the preset threshold but the verification still fails, a voice instruction of the user stops being received, and the user can only turn on a terminal screen manually, thereby ensuring security of using the terminal.

[0074] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

[0075] Refer to FIG. 7, which is a schematic structural diagram of a terminal according to a seventh embodiment

Petitioner Exhibit 1004, Page 467 of 1052

of the present invention. The terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The terminal 4 specifically includes an audio monitoring unit 60 and a processor 70.

[0076] The audio monitoring unit 60 is configured to pick up a voice instruction issued by a user.

[0077] The processor 70 is configured to, when the voice instruction of the user is received, identify the voice instruction and extract wake-up indication information from the voice instruction; perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0078] Specifically, identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a word such as hi or hello preset by the user, so as to extract the wake-up indication information included in the voice instruction of the user.

[0079] Performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0080] After the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing indication information is converted into a text; and a meaning of the processing prompt information is identified according to the text, and a response is made according to the obtained meaning.

[0081] Definitely, in an exemplary implementation manner of this embodiment of the present invention, the audio monitoring unit 60 may be a microphone.

[0082] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

[0083] It should be understood that the first extracting unit and the second extracting unit, the first processing unit and the second processing unit, and the first executing module and the second executing module do not imply a sequence relationship or quantity unit, but are intended to distinguish different modules or units.

[0084] Persons of ordinary skill in the art may understand that, various aspects or possible implementations of various aspects of the present invention may be specifically implemented in a system, method, or computer program product. Moreover, various aspects or possible implementations of various aspects of the present invention may take a form of computer program product, where the computer program product refers to computer readable program code stored in a computer readable medium.

[0085] A computer readable medium may be a computer readable signal medium or computer readable storage medium. The computer readable storage medium includes but not limited to electronic, magnetic, optical, electromagnetic, infrared, or semi-conductive system, device, or apparatus, or any suitable combination thereof, for example, a random access memory (RAM), a read only memory (ROM), an erasable programmable read only memory (EPROM or FLASH), an optical fiber, a compact disc read only memory (CD-ROM).

[0086] A processor in a computer reads computer readable program code stored in the computer readable medium, so that a specific function and action in each step or in a combination of steps in the flowcharts may be executed by the processor; generates an apparatus for implementing a specific function and action in each block or in a combination of blocks in the block diagrams.

[0087] Computer readable program code may be all executed on a user computer alone, or a part thereof may be executed on a user computer as a standalone software package, or a part thereof may be executed on a local computer while the other part is executed on a remote computer, or may be all executed on a remote computer or server. It should also be noted that, in some alternative implementation solutions, each step in the flowcharts or specific functions in each block in the block diagrams may not occur in the illustrated order. For example, two consecutive steps in the illustration which are dependent on an involved function, or two blocks may in fact be executed substantially simultaneously, or these blocks may sometimes be executed in a reverse order.

[0088] It is apparent that persons skilled in the art can make various modifications and variations to the present invention without departing from the spirit and scope of the present invention. The present invention is intended to cover these modifications and variations provided that they fall within the scope of protection defined by the following claims or their equivalent technologies.

Claims

1. A terminal voice control method, comprising:

when a voice instruction of a user is received, identifying the voice instruction, and extracting wake-up indication information from the voice instruction;

performing voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and

extracting, after the voice verification succeeds,

Petitioner Exhibit 1004, Page 468 of 1052

processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

2. The method according to claim 1, wherein the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information comprises:

after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, lighting up a terminal screen and unlocking the terminal screen, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information; and
after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

3. The method according to claim 1, further comprising:

if no wake-up indication information is obtained by extracting, continuing to wait for receiving a voice instruction of the user.

4. The method according to claim 1, further comprising:

prompting reverification if the voice verification fails, and stopping receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

5. The method according to claim 1, wherein the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information comprises:

extracting the processing prompt information in the voice instruction after the wake-up indication information is removed from the voice instruction, and converting the processing indication information into a text; and
identifying a meaning of the processing prompt information according to the text, and making a response according to the obtained meaning.

6. A terminal voice control apparatus, comprising:

a receiving unit, configured to, when a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction;
a first extracting unit, configured to perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and
a first processing unit, configured to, after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

7. The apparatus according to claim 6, wherein the first processing unit comprises:

a first executing module, configured to, after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information; and
a second executing module, configured to, after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

8. The apparatus according to claim 6, further comprising:

a second extracting unit, configured to, if no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of the user.

9. The apparatus according to claim 6, comprising:

a second processing unit, configured to prompt reverification when the voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

10. The apparatus according to claim 6, wherein the first processing unit comprises:

a converting module, configured to extract the processing prompt information in the voice instruction after the wake-up indication information is removed from the voice instruction, and convert the processing indication information in-

Petitioner Exhibit 1004, Page 469 of 1052

to a text; and
a responding module, configured to identify a meaning of the processing prompt information according to the text, and make a response according to the obtained meaning.

11. A terminal, comprising an audio monitoring unit and a processor, wherein:

the audio monitoring unit is configured to pick up a voice instruction issued by a user; and
the processor is configured to, when the voice instruction of the user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction; perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

Petitioner Exhibit 1004, Page 470 of 1052

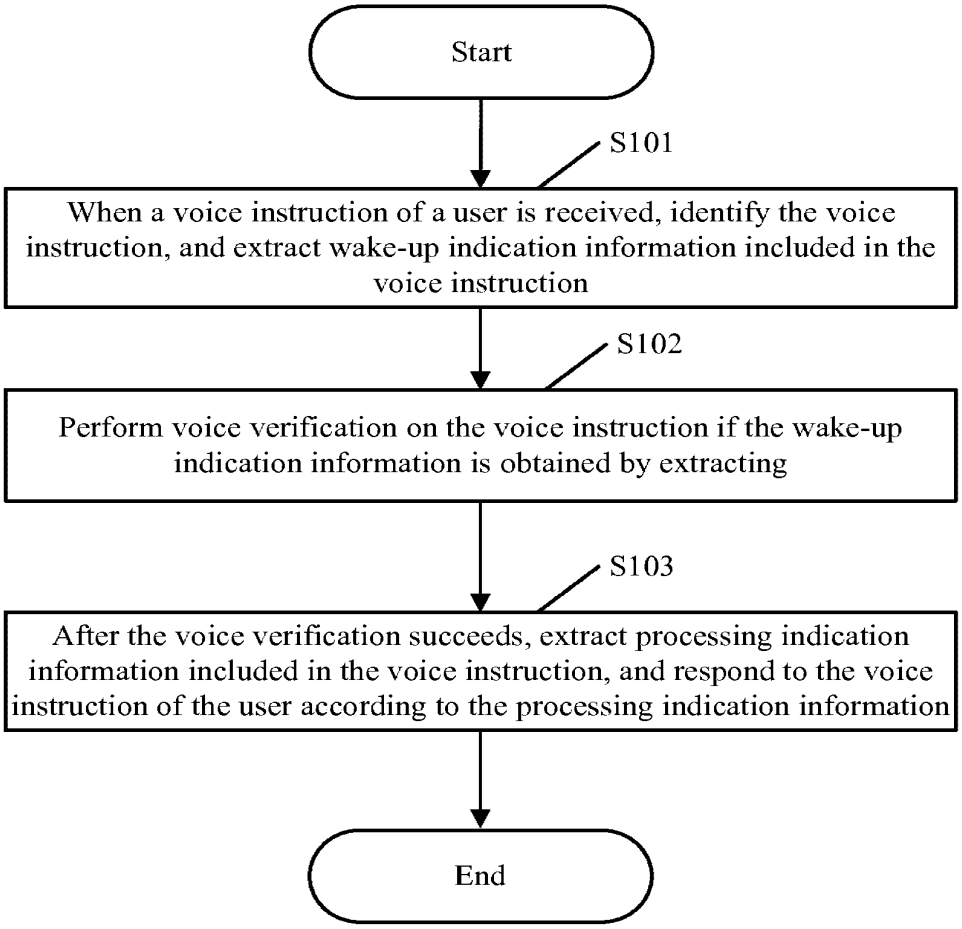


FIG. 1

Petitioner Exhibit 1004, Page 471 of 1052

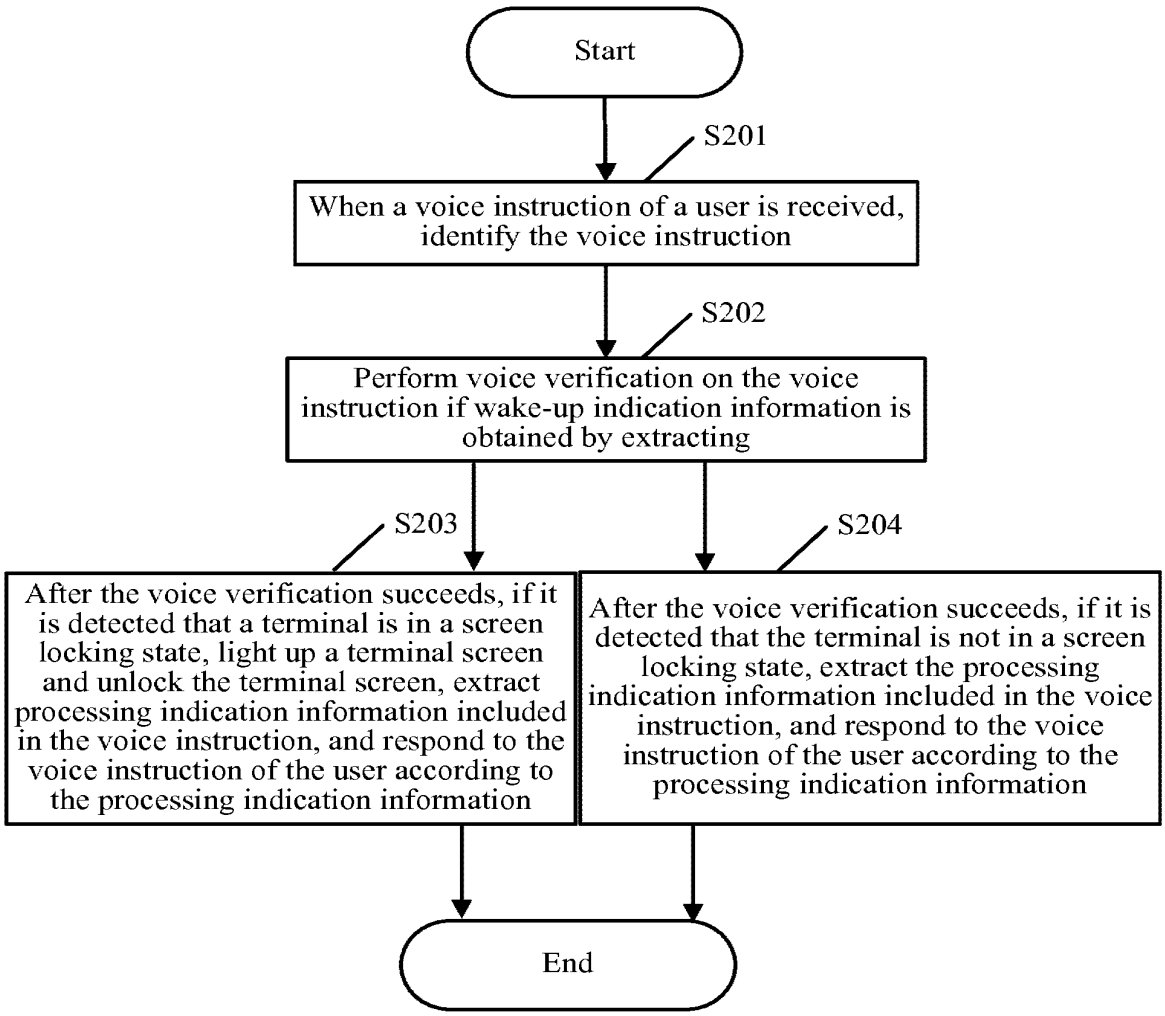


FIG. 2

Petitioner Exhibit 1004, Page 472 of 1052

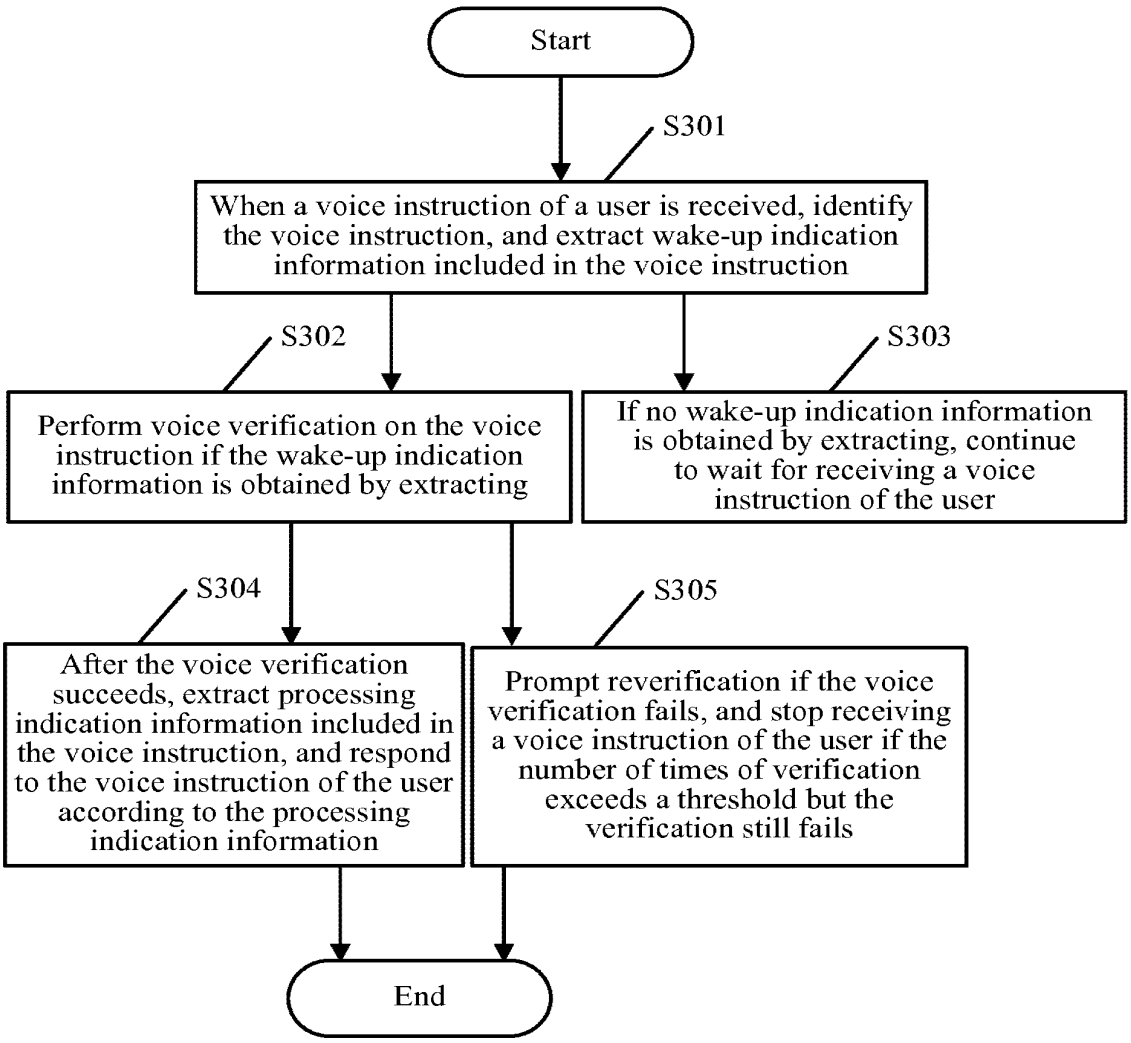


FIG. 3

Petitioner Exhibit 1004, Page 473 of 1052

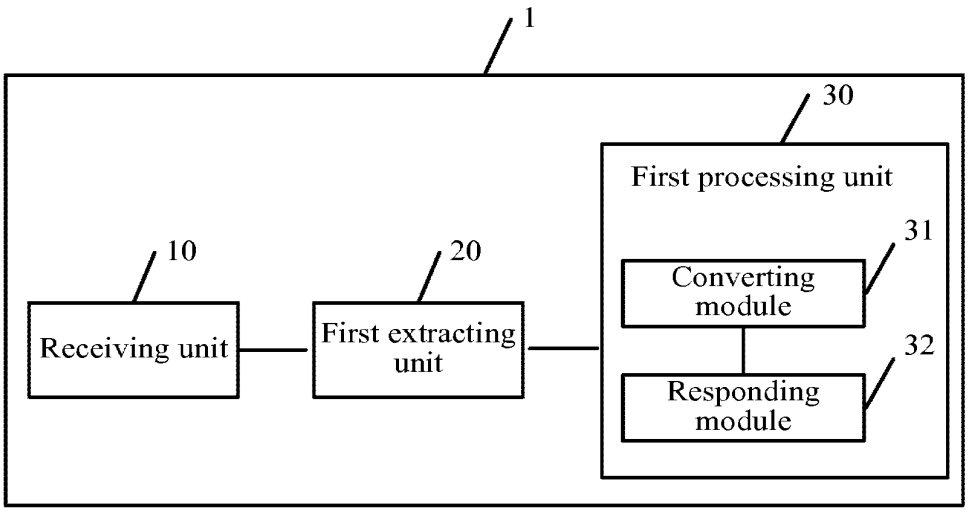


FIG. 4

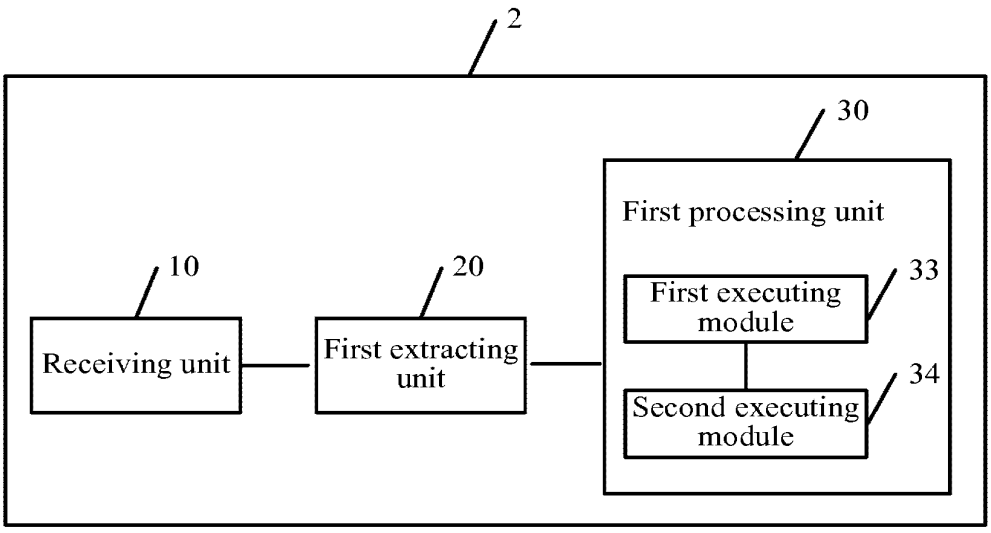


FIG. 5

Petitioner Exhibit 1004, Page 474 of 1052

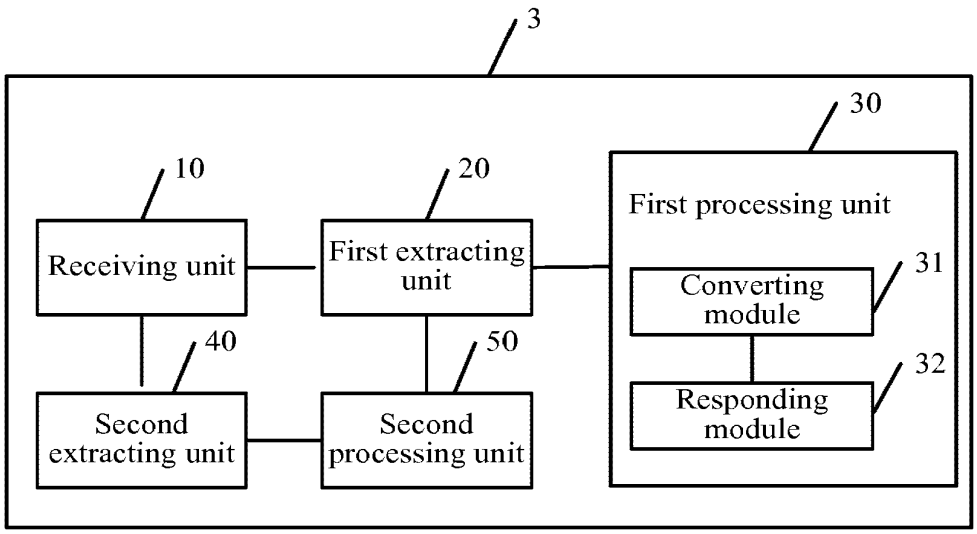


FIG. 6

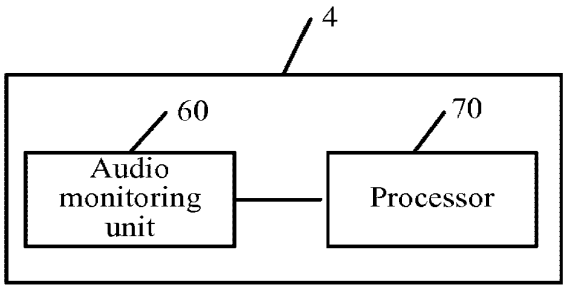


FIG. 7

Petitioner Exhibit 1004, Page 475 of 1052

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/083507

A. CLASSIFICATION OF SUBJECT MATTER

H04M 1/725 (2006.01) i; G10L 7/00 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04M, H04W, G10L, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, CNTXT, CNKI, VEN: handset, communication terminal, mobile telephone, mobile terminal, cellular phone, portable terminal, communication device, voice, sound, control, awak+, instruct+, command, unlock, recogniz+, extract, verify

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103198831 A (VIA TECHNOLOGIES, INC.), 10 July 2013 (10.07.2013), description, paragraphs [0046]-[0076], claims 1-12, and figures 2-5	1, 3-4, 6, 8-9, 11
Y	CN 103198831 A (VIA TECHNOLOGIES, INC.), 10 July 2013 (10.07.2013), description, paragraphs [0046]-[0076], claims 1-12, and figures 2-5	2, 5, 7, 10
Y	CN 102932539 A (SHENZHEN ZTE MOBILE TELECOM CO., LTD.), 13 February 2013 (13.02.2013), description, paragraphs [0016]-[0038], claims 1-7, and figure 1	2, 5, 7, 10
A	CN 102546953 A (GIONEE COMMUNICATION EQUIPMENT CO., LTD.), 04 July 2012 (04.07.2012), the whole document	1-11
A	CN 102148899 A (GUANG DONG OPPO MOBILE TELECOMMUNICATIONS CO., LTD.), 10 August 2011 (10.08.2011), the whole document	1-11
PX	CN 103595869 A (HUAWEI DEVICE CO., LTD.), 19 February 2014 (19.02.2014), claims 1-11	1-11

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 October 2014 (20.10.2014)	**04 November 2014 (04.11.2014)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **LI, Zhenhua** Telephone No.: (86-10) **62087678**

Form PCT/ISA/210 (second sheet) (July 2009)

Petitioner Exhibit 1004, Page 476 of 1052

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2014/083507

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103198831 A	10 July 2013	None	
CN 102932539 A	13 February 2013	None	
CN 102546953 A	04 July 2012	None	
CN 102148899 A	10 August 2011	None	
CN 103595869 A	19 February 2014	None	

Form PCT/ISA/210 (patent family annex) (July 2009)

Petitioner Exhibit 1004, Page 477 of 1052

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- CN 201310574762 **[0001]**

Petitioner Exhibit 1004, Page 478 of 1052

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 May 2014 (01.05.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/066192 A1

(51) International Patent Classification:
G10L 15/24 (2013.01) *G06K 9/62* (2006.01)
G06F 3/01 (2006.01) *G10L 15/22* (2006.01)
G06K 9/00 (2006.01) *A63F 13/00* (2014.01)

(21) International Application Number:
PCT/US2013/065793

(22) International Filing Date:
18 October 2013 (18.10.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/662,293 26 October 2012 (26.10.2012) US

(71) Applicant: **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **KAPUR, Jay**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **TASHEV, Ivan**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **SELTZER, Mike**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **HODGES, Stephen Edward**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

[Continued on next page]

(54) Title: AUGMENTING SPEECH RECOGNITION WITH DEPTH IMAGING

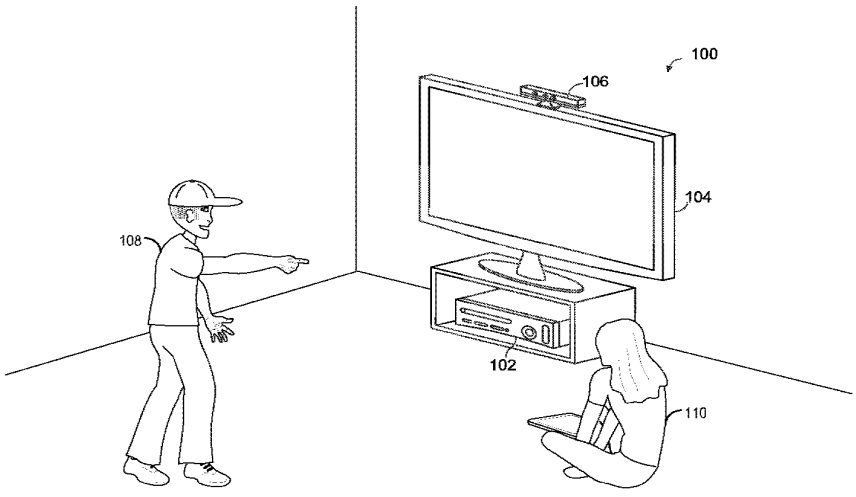


FIG. 1

(57) Abstract: Embodiments related to the use of depth imaging to augment speech recognition are disclosed. For example, one disclosed embodiment provides, on a computing device, a method including receiving depth information of a physical space from a depth camera, receiving audio information from one or more microphones, identifying a set of one or more possible spoken words from the audio information, determining a speech input for the computing device based upon comparing the set of one or more possible spoken words from the audio information and the depth information, and taking an action on the computing device based upon the speech input determined.

WO 2014/066192 A1

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

AUGMENTING SPEECH RECOGNITION WITH DEPTH IMAGING

BACKGROUND

[0001] Computerized speech recognition seeks to identify spoken words from audio information, such as from audio signals received via one or more microphones. However, ambiguities may arise in identifying spoken words in the audio information. Further, the context of the spoken words, for example, whether the spoken words were intended to be a speech input to a computing device, may not be easily determined from such audio information.

SUMMARY

[0002] Embodiments related to the use of depth imaging to augment speech recognition are disclosed. For example, one disclosed embodiment provides, on a computing device, a method including receiving depth information of a physical space from a depth camera, receiving audio information from one or more microphones, identifying a set of one or more possible spoken words from the audio information, determining a speech input for the computing device based upon comparing the set of one or more possible spoken words from the audio information and the depth information, and taking an action on the computing device based upon the speech input determined.

[0003] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Furthermore, the claimed subject matter is not limited to implementations that solve any or all disadvantages noted in any part of this disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] FIG. 1 shows a schematic example of a speech recognition environment according to an embodiment of the disclosure.

[0005] FIG. 2 is a flow chart illustrating a method for recognizing speech according to an embodiment of the disclosure.

[0006] FIG. 3 is a flow chart illustrating a method for recognizing speech according to another embodiment of the disclosure.

[0007] FIG. 4 is a flow chart illustrating a method for recognizing speech according to a further embodiment of the disclosure.

Petitioner Exhibit 1004, Page 481 of 1052

[0008] FIG. 5 schematically shows a non-limiting computing system.

DETAILED DESCRIPTION

[0009] Computerized speech recognition may pose various challenges. For example, pronunciation of individual words, accent, sharpness, tone, imperfections/impediments, and other variables of human speech may differ widely between users. Additionally, reverberation and/or noise and other unwanted sounds (e.g., loudspeakers, vacuum cleaners, etc.) in the room in which the words are spoken may hinder speech recognition. Further, the context in which the recognized words are spoken may impact such factors as whether a recognized speech segment was intended as a speech input.

[0010] Accordingly, embodiments are disclosed that relate to augmenting a speech recognition process with literal and/or contextual information identified in depth information received from a depth camera. For example, in some embodiments, movements of the speaker's mouth, tongue, and/or throat may be identified from the depth information and used to confirm the identity of possible spoken words identified via audio data, identify words not detected in audio data, etc. Additionally, in some embodiments, gestures, postures, etc. performed by the speaker may be identified from the depth information and used to place the identified words into a desired context, such as confirming that the identified spoken words were intended as an input to a computing device. The term "speech recognition" as used herein may include word recognition, speaker recognition (e.g. which of two or more users in an environment is speaking), semantic recognition, emotion recognition, and/or the recognition of any other suitable aspect of speech in a use environment.

[0011] FIG. 1 shows a non-limiting example of a speech recognition environment 100. In particular, FIG. 1 shows a computing system 102 in the form of an entertainment console that may be used to play a variety of different games, play one or more different media types, and/or control or manipulate non-game applications and/or operating systems. FIG. 1 also shows a display device 104 such as a television or a computer monitor, which may be used to present media content, game visuals, non-game computing content, etc., to users.

[0012] The speech recognition environment 100 further includes a capture device 106 in the form of a depth camera that visually monitors or tracks objects and users within an observed scene. Capture device 106 may be operatively connected to the computing system 102 via one or more interfaces. As a non-limiting example, the computing system 102 may include a universal serial bus to which the capture device 106 may be connected. Capture device 106 may be used to recognize, analyze, and/or track one or more human subjects and/or

Petitioner Exhibit 1004, Page 482 of 1052

objects within a physical space, such as user 108. In one non-limiting example, capture device 106 may include an infrared light source to project infrared light onto the physical space and a depth camera configured to receive infrared light. Capture device also may comprise other sensors, including but not limited to two-dimensional image sensor(s) (e.g. a visible light camera such as an RGB image sensor and/or a grayscale sensor) and one or more microphones (e.g. a directional microphone array). While depicted as providing input to an entertainment console, it will be understood that a depth camera may be used to provide input relevant to speech recognition for any suitable computing system, and may be used in non-gaming environments.

[0013[In order to image objects within the physical space, the infrared light source may emit infrared light that is reflected off objects in the physical space and received by the depth camera. Based on the received infrared light, a depth map of the physical space may be constructed. Capture device 106 may output the depth map derived from the infrared light to computing system 102, where it may be used to create a representation of the physical space imaged by the depth camera. The capture device may also be used to recognize objects in the physical space, monitor movement of one or more users, perform gesture recognition, etc. Virtually any depth finding technology may be used without departing from the scope of this disclosure. Example depth finding technologies are discussed in more detail with reference to FIG. 5.

[0014[FIG. 1 also shows a scenario in which capture device 106 tracks user 108 so that the movements of the user may be interpreted by computing system 102. In particular, movements of the mouth, tongue, and/or throat of user 108 may be monitored to determine if the user 108 is speaking. If user 108 is speaking, audio information received by computing system 102 (e.g. via one or more microphones incorporated into capture device 106 and/or located external to capture device 106) may be analyzed to recognize one or more of the words spoken by the user. The mouth, tongue, and/or throat movements also may be used to augment the process of identifying the spoken words, for example by confirming that the identified words were spoken, adding additional identified spoken words, etc.

[0015[Information from the capture device may also be used to determine various contextual elements of the identified spoken words. For example, if additional users are present in the physical space, such as user 110, the user from which the spoken words were received may be distinguished from other users by comparing the spoken words to the mouth/throat/tongue movements of one or more users in the physical space. Further, facial recognition, speaker identification (e.g. based on the user's height, body shape, gait, etc.),

Petitioner Exhibit 1004, Page 483 of 1052

and/or other suitable techniques further may be used to determine the identity of the person speaking. The relative positions and/or orientations of one or more users in a room also may be tracked to help determine whether a speaker is making a speech input. For example, if a user is not facing the capture device when speaking, it may be determined that the user is not speaking to the system. Likewise, where multiple users are visible by the capture device, whether a user is facing the capture device may be used as information to identify which person made a speech input.

[0016[Furthermore, once one or more users have been identified, the one or more users may be tracked (via the capture device, for example). This may help to facilitate the efficient matching of future recognized speech to identified speakers, and therefore to quickly identify which speech recognition model/parameters to use for a particular user (e.g. to tune the speech recognition for that user).

[0017[Further, gestures performed by user 108 identified via information from capture device 106 may be used to identify contextual information related to identified spoken words. For example, if user 108 is speaking with the intent to control computing system 102 via voice commands, user 108 may perform one or more gestures and/or postures, deliberate or otherwise, that may indicate this intent. Examples include, but are not limited to, pointing toward display device 104, looking at computing system 102 or display device 104 while speaking, or performing a specific gesture that is associated with a recognized user input. Thus, by identifying the gesture performed by user 108 as well as identifying the spoken words, a determination of the intent of the user to control the computing device may be made. Likewise, if user 108 is looking at another user, gesturing toward another user, etc., while speaking, an intent to control the computing device may not be inferred in some embodiments.

[0018[Other types of contextual information likewise may be determined from the information received from capture device 106. For example, in some embodiments, an emotional state of user 108 when speaking may be determined by facial and/or body features, postures, gestures, etc., of user 108 from depth information. As yet another example, objects in the imaged physical space may be identified and used to distinguish ambiguous words. For example, compound words such as "quarterback" may be difficult to distinguish from the individual words ("quarter" and "back") that make up the compound word. Therefore, in the case of such ambiguities, depth image data of the physical space may be used to detect objects, actions, etc., that may provide context to help determine the actual word or words spoken. In the specific example of "quarterback," depth image data

Petitioner Exhibit 1004, Page 484 of 1052

may be analyzed to determine the presence of objects and/or other contextual clues to help disambiguate these terms, such as money in a user's hand, football-related objects (e.g. is the user seated in front of the television watching a football game), etc. Such information also may be used in some instances to help disambiguate homonyms, such as "ate" and "eight."

[0019[Computing system 102 also may be configured to communicate with one or more remote computing devices, not shown in FIG. 1. For example, computing system 102 may receive video content directly from a broadcaster, third party media delivery service, or other content provider. Computing system 102 may also communicate with one or more remote services via the Internet or another network, for example in order to analyze the received audio and/or image data, perform the speech recognition, etc. While the embodiment depicted in FIG. 1 shows computing system 102, display device 104, and capture device 106 as separate elements, in some embodiments one or more of the elements may be integrated into a common device.

[0020[FIG. 2 shows a flow diagram depicting an embodiment of a method 200 for recognizing speech of a user. Method 200 may be performed by a computing device configured to receive and process audio and depth information, such as information received from capture device 106.

[0021[At 202, method 200 includes receiving depth information from a depth camera. As explained above, the depth information may be used to construct a depth map of the imaged physical space including one or more users. Additionally, image information from a visible light camera may also be received. At 204, method 200 includes receiving audio information acquired via one or more microphones, which may include directional microphones in some embodiments. At 206, one or more possible spoken words are identified from the audio information. The one or more possible spoken words may be identified by the computing device using any suitable speech recognition processes.

[0022[At 208, method 200 includes determining a speech input for the computing device based on the one or more possible spoken words and the depth information. The speech input may comprise a command that indicates an action to be performed by the computing device, content intended to be displayed on a display device and/or recorded by a computing device, and/or any other suitable speech input.

[0023[The identified possible spoken words and the depth information may be utilized in any suitable manner to determine the speech input. For example, as indicated at 210, movements of the user's mouth, tongue and/or throat may be utilized to determine possible

Petitioner Exhibit 1004, Page 485 of 1052

sounds and/or words spoken by the user. These identified possible sounds/words may then be used to disambiguate any potentially ambiguous possible spoken words from the audio information, and/or to increase a certainty of word identifications, as described in more detail below.

[0024[Similarly, in some embodiments, mouth, tongue and/or throat movements may be used to independently determine a set of possible spoken words. This set of possible spoken words may similarly be compared to the set of possible spoken words determined from the audio information to help disambiguate any uncertainty in the correct identification of words from the audio information, to add any potential missed words to the audio data, etc.

[0025[As mentioned above, the depth information also may be used to identify contextual elements related to the possible speech segments, as indicated at 212. Any suitable contextual elements may be identified. Examples of such contextual elements may include, but are not limited to, an identity of the user, an emotion of the user, a gesture performed by the user, one or more physical objects in the physical space of the user, etc. The contextual elements identified from the depth information may be used to confirm a speech input identified from the audio information, disambiguate any ambiguous possible spoken words (e.g. compound words, homonyms, etc.), place the speech input into a desired context, utilize a directional microphone system to isolate that speaker from others in the environment, tune the speech recognition based on known speech attributes of the identified user, and/or for any other suitable purposes.

[0026[Continuing with FIG. 2, method 200 comprises, at 214, taking an action on the computing device based on upon the speech input. For example, an action indicated by a command speech input may be performed, text content corresponding to the spoken words may be displayed on the display device, etc. Further, in some embodiments, the text content may be tagged with an emotional state, such that words may have a different appearance depending upon the user's detected emotional state when the words were spoken.

[0027[FIG. 3 shows a flow diagram depicting an embodiment of a method 300 for recognizing a command speech input configured to cause a computing device to perform a specified action. Method 300 may be performed by a computing device configured to receive and process audio and depth input. At 302, method 300 includes receiving depth information from a depth camera, and at 304, receiving audio information from one or more microphones. At 306, method 300 comprises identifying one or more possible spoken from the audio information, and at 308, identifying contextual elements from the depth information. Contextual elements may include, but are not limited to, a gesture performed

Petitioner Exhibit 1004, Page 486 of 1052

by the user (e.g. movement of mouth, throat, tongue, body, etc.), as indicated at 310, a physical state of a user (e.g. whether a user is sitting, crouching or standing, whether a user's mouth is open or closed, how far a user is from a display, an orientation of the user's head, etc.), as indicated at 312, and/or an emotional state of the user, as indicated at 314. It will be understood that these contextual elements are described for the purpose of example, and are not intended to be limiting in any manner.

[0028[At 316, method 300 includes comparing the spoken words and the identified contextual elements. The spoken words and the contextual elements may be compared to determine, for example, whether the spoken words are intended as a speech input directing the computing device to perform a specified action based upon the one or more contextual elements identified from the depth information. For example, a particular gesture performed by the user and identified from the depth information may indicate that the spoken words are intended as user input. As more specific example, the user may direct a gesture at a speech recognition system device, such as pointing at the computing device/display/capture device/etc. while speaking, and/or the user may perform a gesture that matches a known gesture associated with a user input.

[0029[Further, an orientation of the user's head may be used to determine if the spoken words are intended as user input. For example, if the user is looking in a particular direction while speaking, such as at toward a speech recognition system device (e.g. a display, computing device, capture device, etc.), it may be determined that the words are intended as a user input to the computing device. Likewise, if the user is looking at another user in the physical space while speaking, it may be indicated that the words are not intended as a user input.

[0030[In a further example, one or more emotions of the user may be determined from the depth data and used to determine if the spoken words are intended as a user input. For example, if the user acting in a commanding and/or directive manner (e.g. deliberate, serious, not facially animated), it may be indicated that the words were intended as user input.

[0031[At 318, method 300 comprises determining from the comparison at 316 whether the spoken words are intended as user input based upon the contextual information. If the words are determined to be intended as speech input, then method 300 comprises, at 320, performing via the computing device the action associated with the speech input. Likewise, if the words are determined not to be intended as a speech input, then method 300 comprises, at 322, not performing an action via the computing device in response to the words.

Petitioner Exhibit 1004, Page 487 of 1052

[0032] FIG. 4 shows a flow diagram depicting an embodiment of a method 400 for identifying spoken words from a combination of audio and depth information. Method 400 may be performed by a computing device configured to receive audio and depth input, such as computing system 102.

[0033] At 402, method 400 comprises receiving depth information from a depth camera, and at 404, receiving audio information from one or more microphone(s). At 406, one or more of the user's mouth, tongue, and throat are located from the depth information. For example, feature extraction may be performed on the depth information to determine where each above-listed facial feature is located.

[0034] At 408, movements of the mouth, tongue, and/or throat may be identified. For example, a degree of opening of the user's mouth, position/shape of the tongue, shape/location of the user's lips, etc., as the user speaks may be tracked to identify the movements.

[0035] At 410, method 400 optionally includes triggering speech recognition to begin responsive to detecting identified movements of the mouth, tongue and/or throat that indicate the user is speaking. In this way, the operation of a resource-intensive speech recognition process may be avoided until identified movements indicate that the user is actually speaking.

[0036] At 412, method 400 comprises identifying a speech input of the user. As explained previously, the speech input may include a command for the computing device to perform an action, or may include input that is to be displayed (e.g. as text) on a display device and/or saved. Identifying the speech input may include for example, identifying one or more possible spoken words from the audio information at 414. The speech input may be identified from the audio data in any suitable manner. Further, as indicated at 416, identifying the speech input may include identifying one or more possible sounds, words, and/or word fragments from the depth information. For example, the mouth, tongue, and/or throat movements of the user may be used to identify sounds, words, etc.

[0037] Identifying the speech input also may include, at 418, comparing the one or more possible spoken words identified from the audio information to the one or more possible spoken words or sounds identified from the depth information. This may help to increase a confidence of possible spoken words identified via the audio data, to help disambiguate possibly ambiguous speech (for example, to identify boundaries between words via hand motion analysis), to identify additional words that were missed in the audio data, and/or may be used in any other suitable manner.

Petitioner Exhibit 1004, Page 488 of 1052

[0038[As a more specific example, movements of the user's mouth, tongue, and/or throat may be analyzed (e.g. by extracting movement data from the depth images and applying one or more classification functions to the movement data) to identify possible words/sounds spoken. Further, in some embodiments, confidence scores may be applied to the possible words/sounds spoken. Then, the determined possible spoken words/sounds determined from the depth information may be compared to the possible spoken words determined from the audio information, which likewise may include confidence score data in some embodiments. From this comparison, a most likely spoken word or words may be identified, e.g. from a highest combined confidence score, or other suitable metric. It will be understood that any suitable mechanism may be used for comparing the possible spoken sounds/words identified via the depth information and the possible spoken words identified via the audio information.

[0039[At 420, method 400 includes taking an action based on the speech input. As described above, any suitable action may be taken. For example, identified speech may be used as a command input to cause the computing device to take an action, may be displayed and/or saved as content, may be used to mark up content based upon a user's determined emotional state when speaking, and/or any other suitable action.

[0040[In some embodiments, the above described methods and processes may be tied to a computing system including one or more computers. In particular, the methods and processes described herein may be implemented as a computer application, computer service, computer API, computer library, and/or other computer program product.

[0041[FIG. 5 schematically shows a non-limiting embodiment of a computing system 500 that can enact one or more of the methods and processes described above. Computing system 500 is one non-limiting example of computing system 102. Computing system 500 is shown in simplified form. It will be understood that virtually any computer architecture may be used without departing from the scope of this disclosure. In different embodiments, computing system 500 may take the form of a mainframe computer, server computer, desktop computer, laptop computer, tablet computer, home-entertainment computer, network computing device, gaming device, mobile computing device, mobile communication device (e.g., smart phone), etc.

[0042[Computing system 500 includes a logic subsystem 502 and a storage subsystem 504. Computing system 500 may optionally include a display subsystem 506, input subsystem 508, communication subsystem 510, and/or other components not shown in FIG. 5.

Petitioner Exhibit 1004, Page 489 of 1052

[0043] Logic subsystem 502 includes one or more physical devices configured to execute instructions. For example, the logic subsystem may be configured to execute instructions that are part of one or more applications, services, programs, routines, libraries, objects, components, data structures, or other logical constructs. Such instructions may be implemented to perform a task, implement a data type, transform the state of one or more components, or otherwise arrive at a desired result.

[0044] The logic subsystem may include one or more processors configured to execute software instructions. Additionally or alternatively, the logic subsystem may include one or more hardware or firmware logic machines configured to execute hardware or firmware instructions. The processors of the logic subsystem may be single-core or multi-core, and the programs executed thereon may be configured for sequential, parallel or distributed processing. The logic subsystem may optionally include individual components that are distributed among two or more devices, which can be remotely located and/or configured for coordinated processing. Aspects of the logic subsystem may be virtualized and executed by remotely accessible, networked computing devices configured in a cloud-computing configuration.

[0045] Storage subsystem 504 includes one or more physical, non-transitory, devices configured to hold data and/or instructions executable by the logic subsystem to implement the methods and processes described herein. When such methods and processes are implemented, the state of storage subsystem 504 may be transformed—e.g., to hold different data.

[0046] Storage subsystem 504 may include removable media and/or built-in devices. Storage subsystem 504 may include optical memory devices (e.g., CD, DVD, HD-DVD, Blu-Ray Disc, etc.), semiconductor memory devices (e.g., RAM, EPROM, EEPROM, etc.) and/or magnetic memory devices (e.g., hard-disk drive, floppy-disk drive, tape drive, MRAM, etc.), among others. Storage subsystem 504 may include volatile, nonvolatile, dynamic, static, read/write, read-only, random-access, sequential-access, location-addressable, file-addressable, and/or content-addressable devices.

[0047] It will be appreciated that storage subsystem 504 includes one or more physical, non-transitory devices. However, in some embodiments, aspects of the instructions described herein may be propagated in a transitory fashion by a pure signal (e.g., an electromagnetic signal, an optical signal, etc.) that is not held by a physical device for a finite duration. Furthermore, data and/or other forms of information pertaining to the present disclosure may be propagated by a pure signal.

Petitioner Exhibit 1004, Page 490 of 1052

[0048[In some embodiments, aspects of logic subsystem 502 and of storage subsystem 504 may be integrated together into one or more hardware-logic components through which the functionally described herein may be enacted. Such hardware-logic components may include field-programmable gate arrays (FPGAs), program- and application-specific integrated circuits (PASIC / ASICs), program- and application-specific standard products (PSSP / ASSPs), system-on-a-chip (SOC) systems, and complex programmable logic devices (CPLDs), for example.

[0049[The term "module" may be used to describe an aspect of computing system 500 implemented to perform a particular function. In some cases, a module may be instantiated via logic subsystem 502 executing instructions held by storage subsystem 504. It will be understood that different modules may be instantiated from the same application, service, code block, object, library, routine, API, function, etc. Likewise, the same module may be instantiated by different applications, services, code blocks, objects, routines, APIs, functions, etc. The term "module" may encompass individual or groups of executable files, data files, libraries, drivers, scripts, database records, etc.

[0050[It will be appreciated that a "service", as used herein, is an application program executable across multiple user sessions. A service may be available to one or more system components, programs, and/or other services. In some implementations, a service may run on one or more server-computing devices.

[0051[When included, display subsystem 506 may be used to present a visual representation of data held by storage subsystem 504. This visual representation may take the form of a graphical user interface (GUI). As the herein described methods and processes change the data held by the storage subsystem, and thus transform the state of the storage subsystem, the state of display subsystem 506 may likewise be transformed to visually represent changes in the underlying data. Display subsystem 506 may include one or more display devices utilizing virtually any type of technology. Such display devices may be combined with logic subsystem 502 and/or storage subsystem 504 in a shared enclosure, or such display devices may be peripheral display devices.

[0052[When included, input subsystem 508 may comprise or interface with one or more user-input devices such as a keyboard, mouse, touch screen, or game controller. In some embodiments, the input subsystem may comprise or interface with selected natural user input (NUI) componentry. Such componentry may be integrated or peripheral, and the transduction and/or processing of input actions may be handled on- or off-board. Example NUI componentry may include one or more microphones for speech and/or voice

Petitioner Exhibit 1004, Page 491 of 1052

recognition; an infrared, color, steroscopic, and/or depth camera for machine vision and/or gesture recognition; a head tracker, eye tracker, accelerometer, and/or gyroscope for motion detection and/or intent recognition; as well as electric-field sensing componentry for assessing brain activity.

[0053] When included, communication subsystem 510 may be configured to communicatively couple computing system 500 with one or more other computing devices. Communication subsystem 510 may include wired and/or wireless communication devices compatible with one or more different communication protocols. As non-limiting examples, the communication subsystem may be configured for communication via a wireless telephone network, or a wired or wireless local- or wide-area network. In some embodiments, the communication subsystem may allow computing system 500 to send and/or receive messages to and/or from other devices via a network such as the Internet.

[0054] Further, computing system 500 may include a skeletal modeling module 512 configured to receive imaging information from a depth camera 520 (described below) and identify and/or interpret one or more postures and gestures performed by a user. Computing system 500 may also include a voice recognition module 514 to identify and/or interpret one or more voice commands or spoken words issued by the user detected via one or more microphones (coupled to computing system 500 or the depth camera). While skeletal modeling module 512 and voice recognition module 514 are depicted as being integrated within computing system 500, in some embodiments, one or both of the modules may instead be included in the depth camera 520.

[0055] Computing system 500 may be operatively coupled to the depth camera 520. Depth camera 520 may include an infrared light 522 and a depth camera 524 (also referred to as an infrared light camera) configured to acquire video of a scene including one or more human subjects. The video may comprise a time-resolved sequence of images of spatial resolution and frame rate suitable for the purposes set forth herein. As described above with reference to FIG. 1, the depth camera and/or a cooperating computing system (e.g., computing system 500) may be configured to process the acquired video to identify one or more postures and/or gestures of the user, determine a location of and track movements of a user's mouth, tongue, and/or throat, and to interpret such postures and/or gestures as device commands configured to control various aspects of computing system 500.

[0056] Depth camera 520 may include a communication module 526 configured to communicatively couple depth camera 520 with one or more other computing devices. Communication module 526 may include wired and/or wireless communication devices

Petitioner Exhibit 1004, Page 492 of 1052

compatible with one or more different communication protocols. In one embodiment, the communication module 526 may include an imaging interface 528 to send imaging information (such as the acquired video) to computing system 500. Additionally or alternatively, the communication module 526 may include a control interface 530 to receive instructions from computing system 500. The control and imaging interfaces may be provided as separate interfaces, or they may be the same interface. In one example, control interface 530 and imaging interface 528 may include a universal serial bus.

[0057[The nature and number of cameras may differ in various depth cameras consistent with the scope of this disclosure. In general, one or more cameras may be configured to provide video from which a time-resolved sequence of three-dimensional depth maps is obtained via downstream processing. As used herein, the term 'depth map' refers to an array of pixels registered to corresponding regions of an imaged scene, with a depth value of each pixel indicating the depth of the surface imaged by that pixel. 'Depth' is defined as a coordinate parallel to the optical axis of the depth camera, which increases with increasing distance from the depth camera.

[0058[In some embodiments, depth camera 520 may include right and left stereoscopic cameras. Time-resolved images from both cameras may be registered to each other and combined to yield depth-resolved video.

[0059[In some embodiments, a "structured light" depth camera may be configured to project a structured infrared illumination comprising numerous, discrete features (e.g., lines or dots). A camera may be configured to image the structured illumination reflected from the scene. Based on the spacings between adjacent features in the various regions of the imaged scene, a depth map of the scene may be constructed.

[0060[In some embodiments, a "time-of-flight" depth camera may include a light source configured to project a pulsed infrared illumination onto a scene. Two cameras may be configured to detect the pulsed illumination reflected from the scene. The cameras may include an electronic shutter synchronized to the pulsed illumination, but the integration times for the cameras may differ, such that a pixel-resolved time-of-flight of the pulsed illumination, from the light source to the scene and then to the cameras, is discernible from the relative amounts of light received in corresponding pixels of the two cameras.

[0061[Depth camera 520 may include a visible light camera 532 (e.g., RGB camera). Time-resolved images from color and depth cameras may be registered to each other and combined to yield depth-resolved color video. Depth camera 520 and/or computing system 500 may further include one or more microphones 534. One or more microphones may

Petitioner Exhibit 1004, Page 493 of 1052

determine directional and/or non-directional sounds coming from users in the physical space and/or other sources. Audio data may be recorded by the one or more microphones 534. Such audio data may be determined in any suitable manner without departing from the scope of this disclosure.

[0062[While depth camera 520 and computing system 500 are depicted in FIG. 5 as being separate devices, in some embodiments depth camera 520 and computing system 500 may be included in a single device. Thus, depth camera 520 may optionally include computing system 500.

[0063[It will be understood that the configurations and/or approaches described herein are exemplary in nature, and that these specific embodiments or examples are not to be considered in a limiting sense, because numerous variations are possible. The specific routines or methods described herein may represent one or more of any number of processing strategies. As such, various acts illustrated and/or described may be performed in the sequence illustrated and/or described, in other sequences, in parallel, or omitted. Likewise, the order of the above-described processes may be changed.

[0064[The subject matter of the present disclosure includes all novel and non-obvious combinations and sub-combinations of the various processes, systems and configurations, and other features, functions, acts, and/or properties disclosed herein, as well as any and all equivalents thereof.

Petitioner Exhibit 1004, Page 494 of 1052

CLAIMS

1. On a computing device, a method for recognizing speech of a user, comprising:

receiving depth information of a physical space from a depth camera;

receiving audio information from one or more microphones;

identifying a set of one or more possible spoken words from the audio information;

determining a speech input for the computing device based upon comparing the set of one or more possible spoken words from the audio information and the depth information; and

taking an action on the computing device based upon the speech input determined.

2. The method of claim 1, further comprising identifying contextual elements in one or more of the depth information from a depth camera, audio information from a directional microphone, and image information from a visible light camera, and comparing the set of one or more possible spoken words from the audio information to the contextual elements to determine the speech input.

3. The method of claim 2, wherein identifying the contextual elements comprises one or more of determining an identity of the user based on one or more of the depth information and information from a visible light camera, determining an emotional state of the user, determining a physical state of the user, determining a gesture performed by the user, and identifying one or more objects in a physical space of the user.

4. The method of claim 1, further comprising identifying a set of one or more possible spoken sounds and/or words from the depth information and comparing the set of one or more possible spoken words identified via the audio information to the set of one or more possible spoken sounds and/or words identified via the depth information to determine the speech input.

5. The method of claim 4, wherein identifying the set of one or more possible spoken sounds and/or words from the depth information further comprises identifying one or more mouth, tongue, and/or throat movements of the user, and identifying the set of one or more possible spoken sounds and/or words based on the movements.

Petitioner Exhibit 1004, Page 495 of 1052

6. The method of claim 1, wherein the speech input comprises a command, and wherein taking the action comprises performing the command.

7. The method of claim 1, further comprising identifying which user of a plurality of users is speaking based on one or more of mouth movements and gaze direction.

8. The method of claim 1, wherein the speech input is content to be stored, and wherein taking the action comprises storing the content.

9. The method of claim 1, wherein the speech input comprises content to be displayed on a display device, and wherein taking the action comprises sending the content to the display device.

10. The method of claim 1, wherein a boundary between possible spoken sounds and/or words is determined based on identified hand movements of the user.

Petitioner Exhibit 1004, Page 496 of 1052

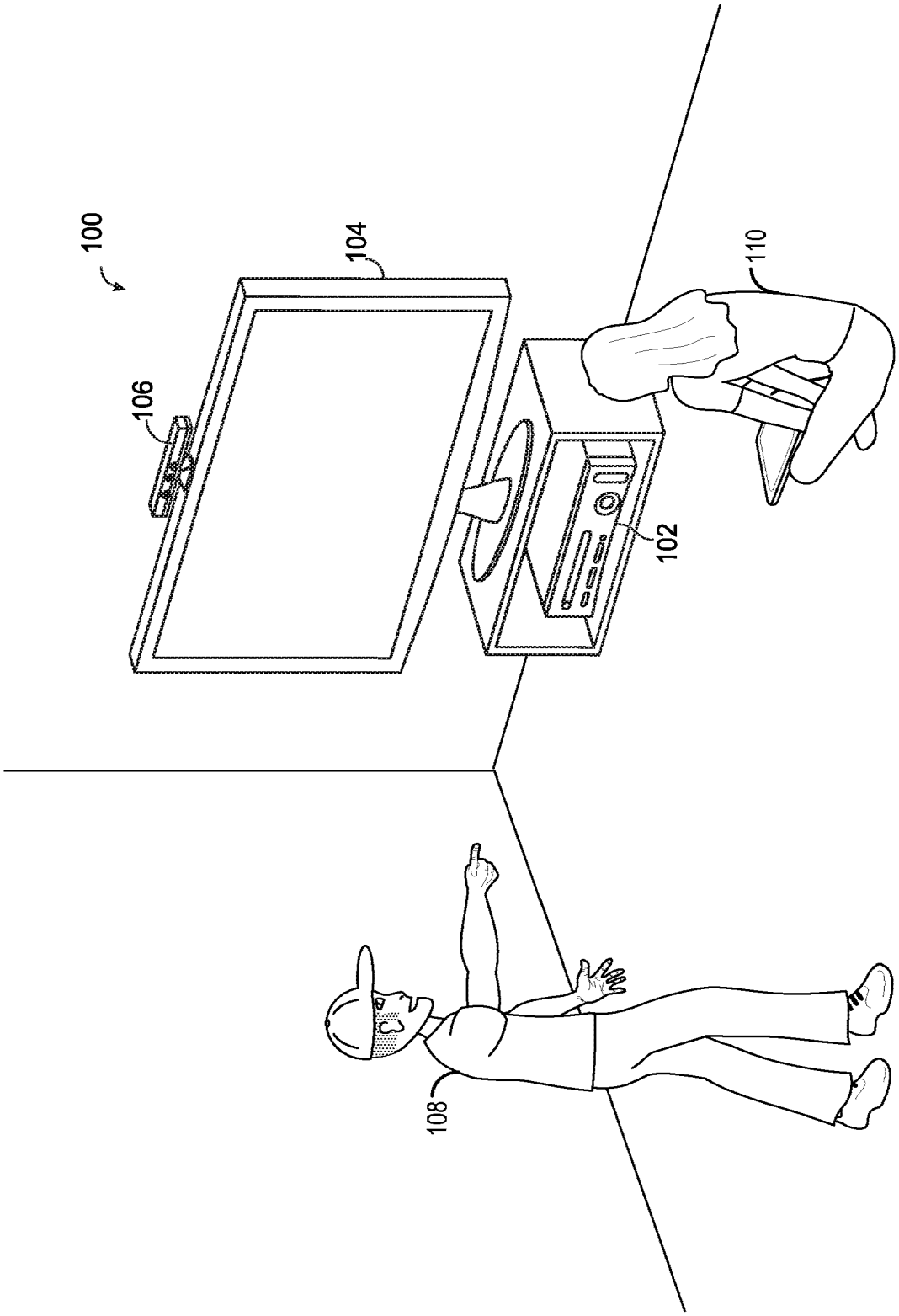


FIG. 1

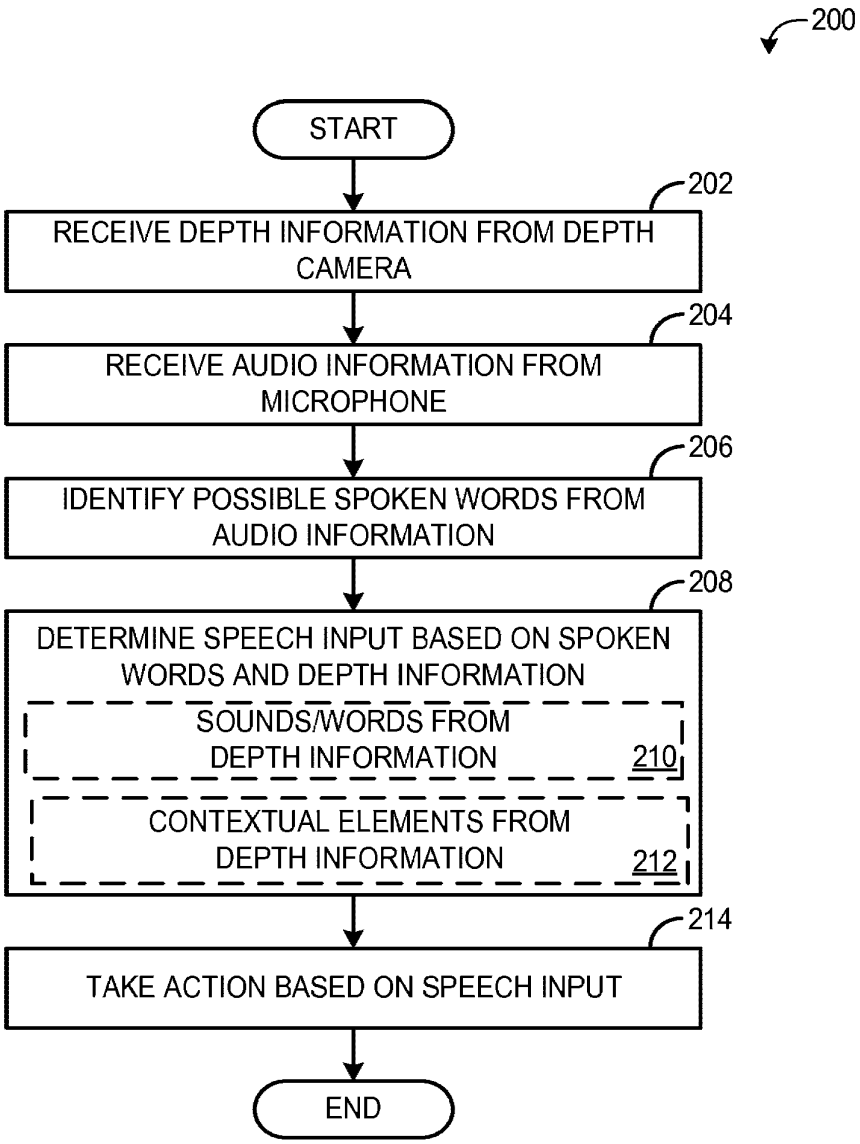

200
START
202
RECEIVE DEPTH INFORMATION FROM DEPTH CAMERA
204
RECEIVE AUDIO INFORMATION FROM MICROPHONE
206
IDENTIFY POSSIBLE SPOKEN WORDS FROM AUDIO INFORMATION
208
DETERMINE SPEECH INPUT BASED ON SPOKEN WORDS AND DEPTH INFORMATION
SOUNDS/WORDS FROM DEPTH INFORMATION 210
CONTEXTUAL ELEMENTS FROM DEPTH INFORMATION 212
214
TAKE ACTION BASED ON SPEECH INPUT
END

FIG. 2

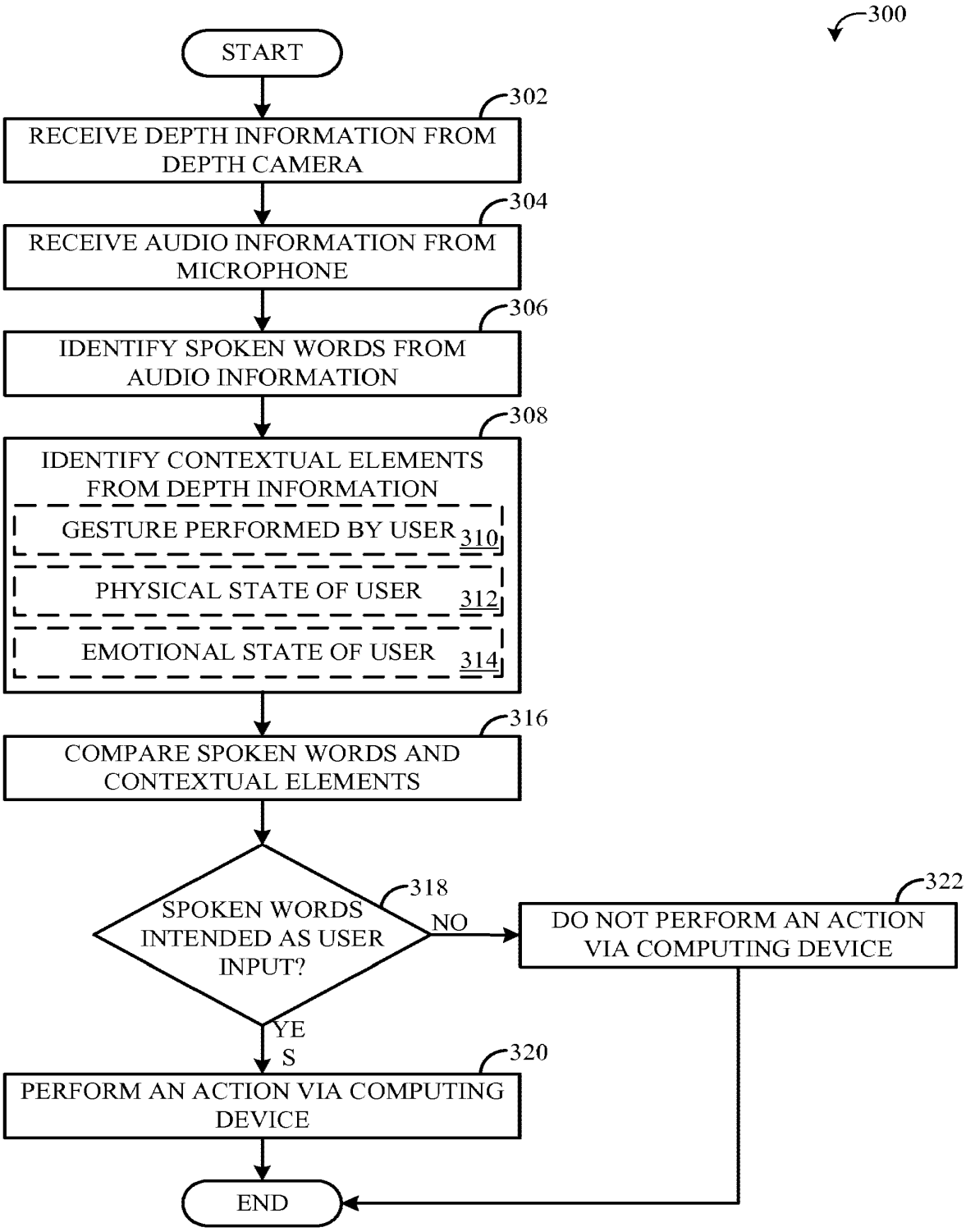

300
START
302
RECEIVE DEPTH INFORMATION FROM DEPTH CAMERA
304
RECEIVE AUDIO INFORMATION FROM MICROPHONE
306
IDENTIFY SPOKEN WORDS FROM AUDIO INFORMATION
308
IDENTIFY CONTEXTUAL ELEMENTS FROM DEPTH INFORMATION
GESTURE PERFORMED BY USER 310
PHYSICAL STATE OF USER 312
EMOTIONAL STATE OF USER 314
316
COMPARE SPOKEN WORDS AND CONTEXTUAL ELEMENTS
318
SPOKEN WORDS INTENDED AS USER INPUT?
NO
322
DO NOT PERFORM AN ACTION VIA COMPUTING DEVICE
YES
320
PERFORM AN ACTION VIA COMPUTING DEVICE
END

FIG. 3

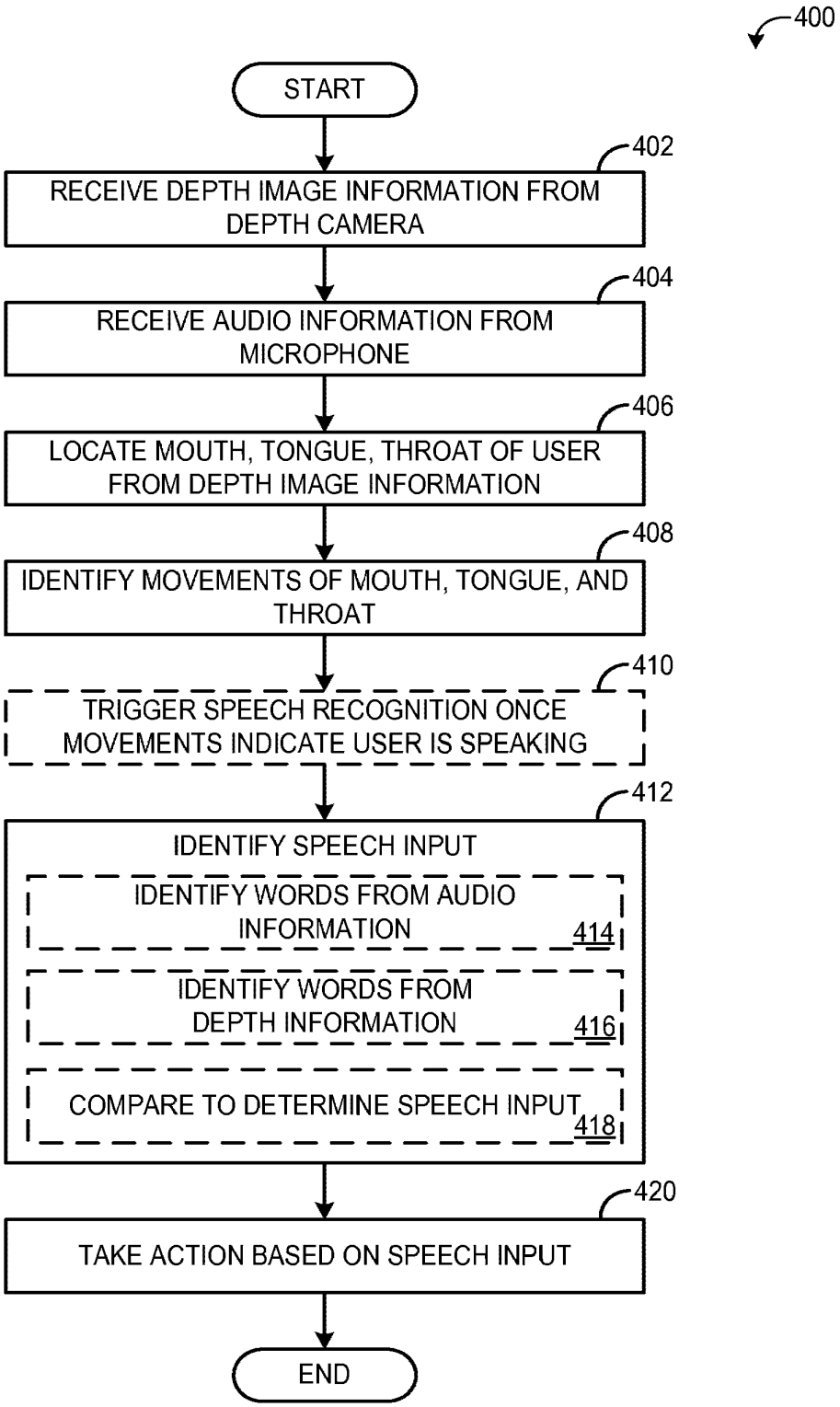

400
START
402
RECEIVE DEPTH IMAGE INFORMATION FROM DEPTH CAMERA
404
RECEIVE AUDIO INFORMATION FROM MICROPHONE
406
LOCATE MOUTH, TONGUE, THROAT OF USER FROM DEPTH IMAGE INFORMATION
408
IDENTIFY MOVEMENTS OF MOUTH, TONGUE, AND THROAT
410
TRIGGER SPEECH RECOGNITION ONCE MOVEMENTS INDICATE USER IS SPEAKING
412
IDENTIFY SPEECH INPUT
IDENTIFY WORDS FROM AUDIO INFORMATION
414
IDENTIFY WORDS FROM DEPTH INFORMATION
416
COMPARE TO DETERMINE SPEECH INPUT
418
420
TAKE ACTION BASED ON SPEECH INPUT
END

FIG. 4

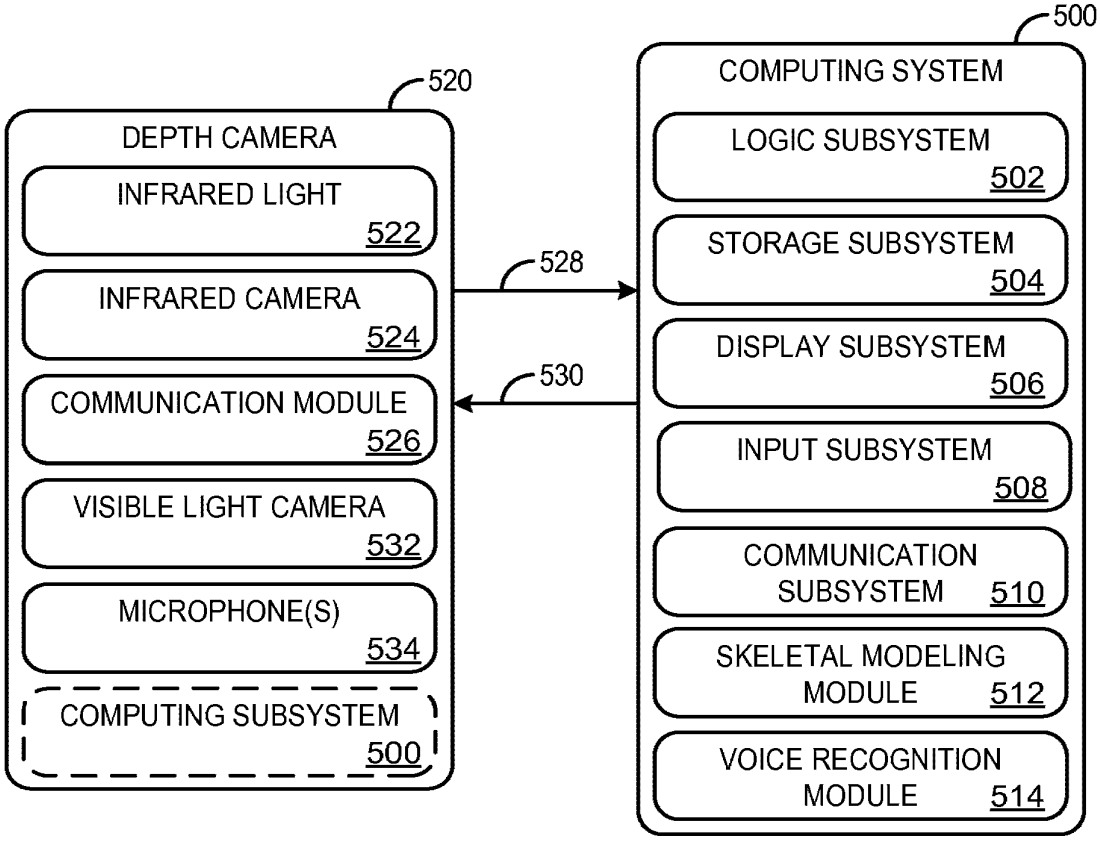

500
COMPUTING SYSTEM
LOGIC SUBSYSTEM 502
STORAGE SUBSYSTEM 504
DISPLAY SUBSYSTEM 506
INPUT SUBSYSTEM 508
COMMUNICATION SUBSYSTEM 510
SKELETAL MODELING MODULE 512
VOICE RECOGNITION MODULE 514
520
DEPTH CAMERA
INFRARED LIGHT 522
INFRARED CAMERA 524
COMMUNICATION MODULE 526
VISIBLE LIGHT CAMERA 532
MICROPHONE(S) 534
COMPUTING SUBSYSTEM 500
528
530

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/065793

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/24 G06F3/01 G06K9/00 G06K9/62
ADD. G10L15/22 A63F13/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A63F G10L G06K G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, COMPENDEX, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 2009/138805 A1 (HILDRETH EVAN [CA]) 28 May 2009 (2009-05-28) paragraphs [0057], [0151] - [0155] figure 2 -----	1-3,6-8 10
X	US 2004/193413 A1 (WILSON ANDREW D [US] ET AL) 30 September 2004 (2004-09-30) paragraphs [0060], [0074], [0084], [0094], [0107], [0108], [0168], [0169] claims 1,5 -----	1-3,9
X	US 2002/116197 A1 (ERTEN GAMZE [US]) 22 August 2002 (2002-08-22) paragraphs [0044], [0046], [0048], [0064] figure 2 -----	1,4,5
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 December 2013	18/12/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Geißler, Christian

1

Form PCT/ISA/210 (second sheet) (April 2005)

Petitioner Exhibit 1004, Page 502 of 1052

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/065793

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/046736 A1 (PRYOR TIMOTHY R [CA] ET AL) 11 March 2004 (2004-03-11) paragraphs [0324], [0327], [0519], [0578] - [0581]	1
X	US 2011/257971 A1 (MORRISON ANDREW R [US]) 20 October 2011 (2011-10-20) paragraphs [0018], [0026] - [0028]	1
Y	US 2010/070268 A1 (SUNG JUN HYUNG [KR]) 18 March 2010 (2010-03-18) paragraphs [0045], [0047], [0052], [0063] claim 8	10

1

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

Petitioner Exhibit 1004, Page 503 of 1052

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2013/065793

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009138805 A1	28-05-2009	CN 101925916 A	22-12-2010
		CN 103369391 A	23-10-2013
		JP 2011504710 A	10-02-2011
		US 2009138805 A1	28-05-2009
		WO 2009067670 A1	28-05-2009
US 2004193413 A1	30-09-2004	US 2004193413 A1	30-09-2004
		US 2010138798 A1	03-06-2010
		US 2010151946 A1	17-06-2010
		US 2013190089 A1	25-07-2013
US 2002116197 A1	22-08-2002	AU 9645901 A	15-04-2002
		US 2002116197 A1	22-08-2002
		WO 0229784 A1	11-04-2002
US 2004046736 A1	11-03-2004	US 6720949 B1	13-04-2004
		US 2004046736 A1	11-03-2004
		US 2006202953 A1	14-09-2006
		US 2008122786 A1	29-05-2008
		US 2013267318 A1	10-10-2013
US 2011257971 A1	20-10-2011	US 2011257971 A1	20-10-2011
		WO 2011130083 A2	20-10-2011
US 2010070268 A1	18-03-2010	JP 2012502325 A	26-01-2012
		KR 20110095236 A	24-08-2011
		US 2010070268 A1	18-03-2010
		WO 2010030129 A2	18-03-2010

Publication number: CN103632668 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20140312
Application number: CN201210299700
Application date: 20120821
Priority: CN201210299700201208 21CN20121299700;
Assignee: BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO ;
Assignee^{std}: BEIJING BAIDU NETCOM SCI AND TEC ;
Inventor^{std}: SU DAN ; WAN GUANGLU ;
International class^{8}: G10L15/06; G10L15/00;
Family members: CN103632668 A CN103632668 B

Title: METHOD AND APPARATUS FOR TRAINING ENGLISH VOICE MODEL BASED ON CHINESE VOICE INFORMATION

Title: 一种基于中文语音信息训练英文语音模型的方法与设备

Abstract:

The invention aims to provide a method and apparatus for training an English voice model based on Chinese voice information. Included are the steps of obtaining voice data inputted by a user through a voice collecting device; determining whether the voice data are English voice data; if so, conducting voice recognition processing for the English voice data based on the Chinese voice information to obtain corresponding pronunciation information to the English voice data; and determining corresponding English data to the English voice data based on the pronunciation information to create or update the English voice model. Compared with the prior art, on the basis of an original Chinese acoustic model, the method and apparatus for training the English voice model based on the Chinese voice information use the similar Chinese voice information and conduct voice recognition processing for the English voice data to obtain the corresponding pronunciation information to the English voice data, thereby achieving the creation or update of the English voice model and effectively controlling the size of the English voice model.

Abstract:

本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备；获取用户通过语音采集装置输入的语音数据；判断该语音数据是否为英文语音数据；当该语音数据为英文语音数据，根据中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息；根据该发音信息，确定该英文语音数据所对应的英文数据，以建立或更新英文语音模型。与现有

The purpose of the present invention is to provide a method and apparatus based Chinese English voice model trained voice information; obtaining user data via voice input voice collecting device; determining whether or not the voice data to voice data in English; when the voice data into English voice data, according to Chinese voice messages, the English voice data for speech recognition processing, in order to obtain the English voice

技术相比，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

data corresponding pronunciation information; based on the pronunciation information, determined that the English voice data corresponding English data, in order to establish or Update English speech model. Compared with the prior art, the invention of the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain the English voice data corresponding pronunciation information to realize the English voice create or update the model, to effectively control the size of the English voice model.

Claims:

1.一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤:a获取用户通过语音采集装置输入的语音数据；b判断所述语音数据是否为英文语音数据；c当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；d根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

2.根据权利要求1所述的方法，其中，所述步骤c包括:-当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

3.根据权利要求1所述的方法，其中，所述步骤c包括:-当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

4.根据权利要求1至3中任一项所述的方法，其中，该方法还包括:-统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；-根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；-根据所述概率值，更新所述英文语音模型。

5.根据权利要求4所述的方法，其中，该方法还包括:-当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；r根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

.

1 for a Chinese voice information speech model training method based on English, wherein the method comprises the steps of: a voice to obtain user data acquisition device through voice input; b judge the speech data Are the English voice data; c when the speech data is English voice data, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the English voice data corresponding pronunciation information; d according to the said pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the pronunciation information and data through the English voice information is mapped to the Chinese.

.

2 The method according to claim 1 or claim 2, wherein said step c comprises: - when the speech data is speech data in English, the word model based on voice information corresponding to the Chinese, English voice data to the voice recognition processing, in order to obtain the English voice data corresponding word pronunciation information.

.

3 method according to claim 1 or claim 2, wherein said step c comprises: - when the speech data is speech data in English, according to the letter model corresponding to the Chinese speech information, English voice data to the voice recognition processing, in order to obtain the

6.根据权利要求5所述的方法，其中，所述步骤r包括:-根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

7.根据权利要求5或6所述的方法，其中，该方法还包括:-根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

8.根据权利要求1至7中任一项所述的方法，其中，该方法还包括:X获取所述用户为目标应用输入的单词语音序列；y根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；Z将与所述英文字符序列相对应的信息提供给所述用户。

9.根据权利要求8所述的方法，其中，所述目标应用为语音输入应用；其中，所述步骤z包括:-将所述英文字符序列作为语音输入候选词条提供给所述用户。

10.根据权利要求8所述的方法，其中，所述目标应用为语音搜索应用；其中，该方法还包括:-将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，所述步骤Z包括:-将所述一个或多个搜索结果提供给所述用户。

11.根据权利要求8至10中任一项所述的方法，其中，所述步骤X包括:-获取所述用户为目标应用输入的字母语音序列；其中，所述步骤y包括:-根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

12.一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括:语音获取装置，用于获取用户通过语音采集装置输入的语音数据；英文判断装置，用于判断所述语音数据是否为英文语音数据；语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置，用于根 据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

13.根据权利要求12所述的模型训练设备，其中，所述语音识别装置用于:-当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

English voice data information corresponding letter sounds.

4 according to the method as claimed in claims 1 3, wherein the method further comprises: - statistical data corresponding to the English, the English voice data in the voice The number of occurrences of the data; - according to the number of occurrences, and the amount of data combined with the voice message, determining the probability value data of English; - based on the probability value, updating the English speech model.

.

5 method according to claim 4, wherein the method further comprises: - when the number of occurrences is greater than a predetermined threshold amount, extracted from the English pronunciation of the English model data with the English voice data; r based on data mapping between the English and the English voice data, create or update data with the corresponding English voice English voice unit.

.

6 The method according to claim 5, wherein said step r comprises: - data according to the mapping relationship between the English and the English voice data, using the method of interpolation calculation, create or update the English voice unit.

7 5 or claim 6, wherein, wherein, the method further comprising: - According to the English voice data and voice data corresponding to the English pronunciation information update The English voice unit.

8 according to the method as claimed in any one of claims 1-7, wherein the method further comprises:. X Get the word speech sequence for the target application user input; y according to The English voice model to determine the sequence of words corresponding to the English voice character sequence; Z will provide the English character sequence information corresponding to the user.

.

9 The method according to claim 8, wherein the target application is speech input application;

14.根据权利要求12所述的模型训练设备，其中，所述语音识别装置用于:-当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

15.根据权利要求12至14中任一项所述的模型训练设备，其中，该设备还包括模型更新装置，用于:-统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；-根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；-根据所述概率值，更新所述英文语音模型。

16.根据权利要求15所述的模型训练设备，其中，该设备还包括:数据提取装置，用于当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；单元训练装置，用于根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

17.根据权利要求16所述的模型训练设备，其中，所述单元训练装置用于:-根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

18.根据权利要求16或17所述的模型训练设备，其中，该设备还包括单元更新装置，用于:-根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

19.根据权利要求12至18中任一项所述的模型训练设备，其中，该设备还包括:序列获取装置，用于获取所述用户为目标应用输入的单词语音序列；序列确定装置，用于根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；信息提供装置，用于将与所述英文字符序列相对应的信息提供给所述用户。

20.根据权利要求19所述的模型训练设备，其中，所述目标应用为语音输入应用；其中，所述信息提供装置用于:-将所述英文字符序列作为语音输入候选词条提供给所述用户。

21.根据权利要求19所述的模型训练设备，其中，所述目标应用为语音搜索应用；其中，该设备还包括:搜索获取装置，用于将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，所述信息提供装置用于:-将所述一个或多个搜索结果提供给所述用户。

wherein, z comprises the step of: - the sequence of characters as the English speech input Candidate entry provided to the user.

.

10 The method according to claim according to claim 8, wherein the target application is a voice search application; wherein the method further comprises: - the sequence of the English characters as the user the query sequence to obtain the query sequence corresponding to one or more search results; wherein Z comprises the steps of: - one or more search results to the user.

.

11 The method of any one of claims 8 to 10, according to claim, wherein X comprises the step of: - obtaining the sequence of the target user for the application of the letter speech input; wherein said step y include: - According to the English voice model to determine the sequence of letters corresponding English voice character sequence.

.

12 A model-based training devices Chinese English speech model training speech information, wherein the apparatus includes: a voice acquisition means for acquiring a voice user input via voice acquisition device transactions; English judging means for judging whether the data of the voice in English speech data; speech recognition means for, when the voice data for the English voice data, voice information based on Chinese, English voice data of the voice recognition processing to obtain with the English voice data corresponding pronunciation information; model training means for the pronunciation information, determine the English voice data corresponding English data, in order to create or update the English voice model, in which the said information relating to the English pronunciation of the Chinese voice information data through mapped.

.

13 in accordance with the model training apparatus according to claim 12, wherein said speech recognition means for: - when the speech data is speech data in English, according to the word model corresponding to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the

22.根据权利要求19至21中任一项所述的模型训练设备，其中，所述序列获取装置用于:-获取所述用户为目标应用输入的字母语音序列；其中，所述序列确定装置用于:-根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

English voice data corresponding word pronunciation information.

.

14 model training apparatus according to claim 12, wherein, wherein the speech recognition means for: - When the voice data into English voice data, based on the letter corresponding to the model Chinese voice message, the English voice data to voice recognition processing, in order to obtain the English voice data information corresponding letter sounds.

.

15 model training apparatus according to claim 12 to 14 in any one of, wherein the apparatus further comprises a model updating means for: - the statistical data corresponding to the English in English, the number of occurrences in the speech data in the speech data; - according to the number of occurrences, and the combined information of the number of voice data, the data determining the probability value in English; - based on the probability value, updating the English speech model.

.

16 model training apparatus according to claim 15, wherein, the apparatus further comprising: data extracting means for a predetermined number of times greater than the threshold value, from the English occurs when the English speech model to extract the data with the English voice data; unit training device for data according to the mapping between the English and the English voice data, create or update data with the corresponding English voice English voice unit.

.

17 in accordance with the model training apparatus according to claim 16, wherein said means for training means: - data according to the mapping relationship between the English and the English voice data, using interpolation method, create or update the English voice unit.

.

18 model training device according to claim 16 or 17, wherein the apparatus further includes means updating means for: - according to the English,

and the English voice data and voice data corresponding pronunciation information, updating the English voice unit.

.

19 model training apparatus according to claim 12 to 18, wherein any one, wherein the apparatus further comprises: a sequence acquisition means for acquiring the user input for the target application word speech sequence; sequence determination means for the English voice model to determine the sequence of words corresponding to the English voice character sequence; information providing means for providing the English character sequence corresponding information to the user.

.

20 in accordance with the model training apparatus according to claim 19, wherein the target application is speech input application; wherein said information providing means for: - the English characters Voice input sequence as a candidate entry provided to the user.

21 in accordance with the model training apparatus according to claim 19, wherein the target application is a voice search application; wherein, the apparatus further comprising: acquisition means searches for the English character sequence as the query sequence to the user, access to the query sequence corresponding to one or more search results; wherein the information providing means for: - the one or more search results to the said user.

.

22 model training apparatus according to claim 19 to 21 in any one of, wherein said sequence acquisition means for: - obtaining the target application user inputs the letter Voice sequence; wherein the means for determining the sequence: - According to the English voice model to determine the sequence of letters corresponding English voice character sequence.

Description:

一种基于中文语音信息训练英文语音模型的方法与设备

A method and apparatus based Chinese English speech training speech information model

技术领域

[0001] 本发明涉及自然语言处理技术领域，尤其涉及一种用于基于中文语音信息训练英文语音模型的技术。

背景技术

[0002] 现有的训练英文语音模型的方式一般利用现有的英语语音数据库，包括语音数据及其标注，在发音单元集合中加入英文发音单元，利用英语数据进行训练。现有技术的问题在于，采用该种方式训练英文语音模型需要一定规模的英语语音数据，且难以对加入模型中的英语词的规模进行控制。

[0003] 因此，如何克服上述困难，在控制规模的同时，实现对英文语音模型的训练，成为本领域技术人员亟需解决的问题之一。

发明内容

[0004] 本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备。

[0005] 根据本发明的一个方面，提供了一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤:

[0006] a获取用户通过语音采集装置输入的语音数据；

[0007] b判断所述语音数据是否为英文语音数据；

[0008] c当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0009] d根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0010] 根据本发明的另一方面，还提供了一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括:

[0011] 语音获取装置，用于获取用户通过语音采集装置输入的语音数据；

[0012] 英文判断装置，用于判断所述语音数据是否为英文语音数据；

FIELD

Invention relates to the field of natural language processing technology, in particular, relates to a voice-based Chinese information technology training English voice model

[0001] The present.

BACKGROUND

[0002] conventional way English voice model training in general use existing English speech database, including data and voice annotation, join English pronunciation unit in pronunciation unit set, use of English training data. Problems of the prior art is that, using this way of training the English voice model requires a certain amount of English voice data, and it is difficult to join model to control the size of the English word.

[0003] Therefore, how to overcome these difficulties, at the same time control the size, to achieve the English speech model training, become one skilled in the need to solve.

Summary

[0004] The present invention is to provide a method and apparatus for Chinese-based voice information training English voice model.

[0005] In accordance with one aspect of the invention, there is provided a voice message for Chinese training English voice model-based approach, wherein the method comprises the following steps:

[0006] a voice to obtain user data acquisition device through voice input;

[0007] b determines whether the data is the voice English voice data;

[0008] c when the speech data is English voice data, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the English voice data corresponding pronunciation information;

[0009] d according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the

[0013] 语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0014] 模型训练装置，用于根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0015] 与现有技术相比，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模；进一步地，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的使用体验。

附图说明

[0016] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本发明的其它特征、目的和优点将会变得更明显:

[0017] 图1示出根据本发明一个方面的用于基于中文语音信息训练英文语音模型的设备不意图；

[0018] 图2示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的设备示意图；

[0019] 图3示出根据本发明另一个方面的用于基于中文语音信息训练英文语音模型的方法流程图；

[0020] 图4示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的方法流程图。

[0021] 附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

[0022] 下面结合附图对本发明作进一步详细描述。

[0023] 图1示出根据本发明一个方面的用于基于中文语音信息训练英文语音模型的设备示意图。模型训练设备1包括语音获取装置101、英文判断装置102、语音识别装置103和模型训练装置104。

[0024] 其中，语音获取装置101获取用户通过语音采集装置输入的语音数据。具体地，用户或大量用

pronunciation information with the English voice information data by the Chinese mapped.

[0010] According to another aspect of the present invention, there is provided a method for Chinese speech information in English speech model training model based training device, wherein the device comprises:

[0011] speech acquisition means for acquiring user voice data acquisition device through voice input;

[0012] English judging means for judging whether or not the voice data is voice data in English;

[0013] speech recognition means for, when the voice data is voice data in English, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain data corresponding to the English voice pronunciation information;

[0014] model training apparatus, according to the pronunciation information for determining the English voice data corresponding English data, to establish or update the English speech model, wherein the pronunciation information by the English data The Chinese voice information is mapped.

[0015] Compared with the prior art, the invention of the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain data corresponding to the English voice pronunciation information to achieve the establishment or update English voice model, effectively control the size of the English voice model; further, the present invention can also be combined with voice recognition that allows users to press two ways to enter the English word speech sequence alphabetical and to enhance the user experience.

Brief Description

[0016] by reading the following detailed description taken with the drawings made of non-limiting example with reference to other features, objects and advantages of the present invention will become more apparent:

[0017] FIG. 1 shows a device-based Chinese English voice model trained voice information is

户通过与语音采集装置的交互，输入了语音数据，语音获取装置101通过调用该语音采集装置所提供的应用程序接口(API)，获取该用户或大量用户输入的语音数据；或者，用户或大量用户通过与语音采集装置的交互，输入了语音数据，并将该语音数据存入语音数据库中，语音获取装置101通过调用该语音数据库所提供的应用程序接口(API)，从该语音数据库中获取用户或大量用户输入的语音数据。

[0025] 本领域技术人员应能理解上述获取语音数据的方式仅为举例，其他现有的或今后可能出现的获取语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0026] 英文判断装置102判断所述语音数据是否为英文语音数据。具体地，英文判断装置102根据语音获取装置101所获取的用户通过语音采集装置输入的语音数据，例如通过声学模型匹配等方式，判断该语音数据是否为英文语音数据。

[0027] 本领域技术人员应能理解上述判断英文语音数据的方式仅为举例，其他现有的或今后可能出现的判断英文语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0028] 语音识别装置103当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息。具体地，当英文判断装置102判断该语音数据为英文语音数据，语音识别装置103根据中文语音信息，如中文声韵母信息，对该英文语音数据进行语音识别处理，例如，当英文判断装置判断该语音数据为英文语音数据时，语音识别装置103采用所有可能的中文声韵母信息去识别该英文语音数据，以确定该英文语音数据的发音信息。

[0029] 例如，用户通过与语音采集装置的交互，输入了语音数据单词“iphone”，语音获取装置101通过调用该语音采集装置所提供的应用程序接口(API)，获取了该语音数据“iphone”；随后，英文判断装置102采用声学模型匹配的方式，判断该语音数据为英文语音数据；随后，语音识别装置103采用所有可能的中文声韵母信息，对该英文语音数据“iphone”进行强制对齐，确定该英文语音数据“iphone”的发音信息为“aa ai4 f engl”。在此，aa、ai4、f、engl等即为中文声韵母信息，其中的数字表示该韵母信息所对应的声调。本领域技术人员应能理解，在此，中文声韵母信息及声调的表示仅为示例，而不应对本发明产生任何限制。

not intended that the present invention according to one aspect;

[0018] Figure 2 illustrates an example of a preferred embodiment of the present invention for the device schematic Chinese English speech voice information model-based training;

[0019] Figure 3 shows a flowchart of a method of training Chinese English speech voice information model-based aspect of the present invention, according to another;

[0020] Figure 4 shows an example of a flowchart of a method for training Chinese English speech voice information model-based accordance with a preferred embodiment of the present invention.

[0021] In the drawings the same or similar reference numerals represent the same or similar parts.

DETAILED DESCRIPTION

[0022] below in conjunction with the accompanying drawings of the present invention will be further described in detail.

[0023] FIG. 1 shows a schematic diagram of the device based on the Chinese model of voice information English voice training with an aspect of the present invention is used in accordance with. Model training device 1 comprises a voice acquisition means 101, judging means 102 in English, the speech recognition apparatus 103 and the model training device 104.

[0024] wherein the voice acquisition means 101 acquires voice data acquisition device user via voice input. Specifically, the user or a large number of users to interact with the voice collecting device, the input voice data, voice acquisition means 101 by calling the voice collection application interface means provided (API), to obtain the user's voice input or a large number of user data; Alternatively, the user or a large number of users to interact with the voice collecting device, the input voice data and voice data stored in the voice database, the speech acquires the application program interface (API) means 101 by calling the speech provided by the database, Get voice data user or a large number of user input from the speech database.

[0030] 优选地，语音识别装置103还可以根据单词模型所对应的中文语音信息、或字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息、或字母发音信息。例如，对于“iphone”，若用户输入语音采集装置的为按单词读的“iphone”，则语音识别装置103确定其对应的采用中文语音信息表示的发音信息为“aa ai4 f engl”;若用户输入语音采集装置的为按字母读的“iphone”，即，“1、p、h、o、n、e”，则语音识别装置103确定其对应的采用中文语音信息表示的发音信息为“aa ai4 p i4ei2 ch oo oul ee enl ii i4，，。

[0031] 本领域技术人员应能理解上述识别发音信息的方式仅为举例，其他现有的或今后可能出现的识别发音信息的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0032] 模型训练装置104根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。具体地，模型训练装置104根据该语音识别装置103所识别获得的发音信息，确定该英文语音数据所对应的英文数据，如根据发音信息“aa ai4 f engl”，确定该发音信息对应的英文数据为“iphone”，再将该发音信息与该英文数据的映射关系存入英文语音模型，以实现对该英文语音模型的建立或更新，其中，该发音信息与该英文数据通过该中文语音信息相映射。

[0033] 在此，英文数据为一具体的英文单词，如“iphone”等，而英文语音数据为该具体的英文单词所对应的语音数据，如“iphone”所对应的语音数据，通过语音获取装置101所获取。

[0034] 本领域技术人员应能理解上述建立或更新英文语音模型的方式仅为举例，其他现有的或今后可能出现的建立或更新英文语音模型的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0035] 在此，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

[0036] 优选地，模型训练设备1的各个装置之间是持续不断工作的。具体地，语音获取装置101获取用户通过语音采集装置输入的语音数据；英文判断装置102判断所述语音数据是否为英文语音数据；

[0025] skilled in the art should be able to understand the way the acquired speech data is only an example, other existing or future way to obtain voice data that may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0026] English judging means 102 judges whether the data is the voice English voice data. In particular, the English voice collecting judging means 102 by means of input speech data according to the voice acquisition unit 101 acquired the user, for example, by matching acoustic models, etc., it is determined whether or not the voice data is voice data in English.

[0027] skilled in the art should be able to understand the judgment of the English way of example only voice data, other available data to determine the English voice mode or may occur in the future as applicable to the present invention, it should also be included Within the scope of the present invention and hereby incorporated herein by reference.

[0028] Speech recognition device 103 when the speech data is English voice data, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the English voice data corresponding pronunciation information. Specifically, when the English judgment means 102 judges that the voice data is English voice data, the voice recognition device 103 according to Chinese voice information, such as Chinese vowel sound information, the English voice data for speech recognition processing, for example, when the judgment means judges that the English When the voice data into English voice data, the voice recognition device 103 uses all possible Chinese vowel sound information to identify the English voice data, voice data to determine if the English pronunciation information.

[0029] For example, the user interacts with the voice collecting device, voice data input word "iphone", a voice interface device 101 to get the application by calling the voice collecting apparatus is provided (API), to obtain the voice Data "iphone"; then judging means 102 English acoustic models using matching methods to determine if the voice data is the English voice data; subsequently, the voice recognition device

语音识别装置103当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置104根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。在此，本领域技术人员应理解“持续”是指模型训练设备1的各装置分别不断地进行语音数据的获取、英文语音数据的判断、英文语音数据的识别处理及英文语音模型的建立或更新，直至该模型训练设备1在较长时间内停止获取语音数据。

[0037] 优选地，语音识别装置103当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。具体地，当该英文判断装置102判断语音获取装置101所获取的语音数据为英文语音数据，语音识别装置103采用单词模型所对应的中文语音信息，如“aaai4 f engl”等，对所述英文语音数据进行语音识别处理，如强制对齐处理，采用所有可能的中文声韵母信息，对该英文语音数据进行强制对齐，以获得与所述英文语音数据相对应的单词发音信息。

[0038] 在此，该单词模型中存储有大量单词所对应的中文语音信息，该单词模型可以位于模型训练设备1中，或者位于与该模型训练设备1通过网络相连接的第三方设备中。

[0039] 优选地，语音识别装置103当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。具体地，当该英文判断装置102判断语音获取装置101所获取的语音数据为英文语音数据，语音识别装置103根据字母模型所对应的中文语音信息，如二十六个英文字母所对应的中文语音信息，如字母“a”对应的中文语音信息“aa eil”、字母“b”对应的中文语音信息“b i4”、字母“c”对应的中文语音信息“s ei4”等，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。

[0040] 优选地，语音识别装置103对于英文判断装置102判断为英文语音数据为语音数据，优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当该语音识别装置103采用单词模型无法识别出该英文语音数据的发音信息时，再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。相似地，该语音识别装置103也可以优先采用字母模型所对应的中

103 uses all possible Chinese vowel sound information, the English voice data "iphone" forced alignment, determine the English voice data "iphone" pronunciation information "aa ai4 f engl". Here, aa, ai4, f, engl and other information that is Chinese vowel sound, which indicates that the vowel digital information corresponding to the tone. The skilled in the art will appreciate that in this, Chinese vowel sound and tonal representation of information are examples only, and not to deal with the present invention, any restrictions.

[0030] Preferably, the speech recognition device 103 may be based on word models corresponding to speech information in Chinese, or Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, in order to obtain the English voice data corresponding word pronunciation information, or to pronounce the letter information. For example, for "iphone", if the user enters a voice collection device for "iphone", the words read by a voice recognition device 103 to determine the corresponding voice information indicating the use of Chinese pronunciation information "aa ai4 f engl"; if the user input voice collecting means is read by the letter "iphone", i.e., "1, p, h, ?, n, e", the speech recognition apparatus 103 determines the corresponding voice information expressed in Chinese pronunciation information "aa ai4 p i4ei2 ch oo oul ee enl ii i4,,.

[0031] skilled in the art should be able to understand the way information is the identification pronunciation of example only, other available information the way for future identification pronunciation or may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0032] Model training apparatus 104 according to the pronunciation information, determine the English voice data corresponding English data, to establish or update the English speech model, wherein the pronunciation information of the data by the English Chinese voice information is mapped. Specifically, the model is based on 103 training device 104 recognizes the voice recognition unit pronunciation information, determined that the English voice data corresponding English data, such as "aa ai4 f

文语音信息，对该英文语音数据进行语音识别处理，当该语音识别装置103采用字母模型无法识别出该英文语音数据的发音信息时，再采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息。在此，优先采用单词模型还是优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，可以是该模型训练设备1预先设定的，也可以应事件而进行调整。

[0041] 例如，对于语音获取装置101所获取的语音数据“iphone”，英文判断装置102判断该语音数据“iphone”为英文语音数据，语音识别装置103优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，但无法识别出该英文语音数据的发音信息；随后，该语音获取装置101再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，获得与该英文语音数据相对应的字母发音信息“aa ai4 p i4 ei2ch oo oul ee enl ii i4”，则可知该语音获取装置101所获取的语音数据为按字母读的“iphone”，即，“1、p、h、o、η、e”。

[0042] 更优选地，语音识别装置103分别采用单词模型所对应的中文语音信息、字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，并分别对语音识别处理所获得的结果进行打分，确定打分较高的发音信息为该英文语音数据的发音信息。

[0043] 在此，由于用户在输入英文单词时有时按单词读有时按字母读，本发明对每个需要识别的英文单词，利用相似的单词模型对应的中文语音信息、字母模型对应的中文语音信息，确定其对应的发音信息，提升了语音识别的准确性。

[0044] 在一优选实施例中，该模型训练设备1还包括模型更新装置(未示出)。该模型更新装置统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；根据所述概率值，更新所述英文语音模型。具体地，当大量用户通过语音采集装置输入语音数据，其中，多个用户所输入的英文语音数据为相同的英文语音数据，语音获取装置101获取该大量用户输入的大量语音数据；英文判断装置102判断该大量语音数据为英文语音数据；随后，语音识别装置103根据中文语音信息，对该大量英文语音数据进行语音识别处理，以获得与该大量英文语音数据相对应的发音信息，在此，由于该多个用户输入的为相同的英文语音数据，因此，该语音识别装置103所获得的关于该多个用户输入的英文语音数据所对

engl", determining the pronunciation information corresponding English data based on pronunciation information the "iphone", then the mapping of the English pronunciation information and data stored in the English speech model, in order to achieve the establishment or update English voice model, wherein the information associated with the English pronunciation of the Chinese voice information data by phase mappings.

[0033] In this case, the English data is a particular English word, such as "iphone" and so on, but the English voice data for specific English word corresponding to the voice data, such as the "iphone" corresponding to the voice data, The acquisition means 101 acquire voice.

[0034] skilled in the art should be understood that the above to create or update the way of example only English voice model, or other existing way to create or update the English voice model may arise in the future as applicable to the present invention, It should also be included within the scope of the present invention and hereby incorporated herein by reference.

[0035] In this case, the present invention is in the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain the English voice data corresponding pronunciation information, to achieve create or update English voice model, effectively control the size of the English voice model.

[0036] Preferably, between each device model training apparatus 1 is continued work. Specifically, the voice acquisition means 101 acquires the user through voice data acquisition device voice input; if the English judgment means 102 judges that the voice data is English voice data; a voice recognition device 103 when the speech data is English voice data, according to Chinese Speech information on the English voice data for voice recognition processing to obtain the data corresponding to the English voice pronunciation information; model training apparatus 104 according to the pronunciation information, determine the English voice data corresponding English data, to English create or update a speech model, wherein the pronunciation information and data through the English voice information is mapped to the Chinese. In this case,

应的发音信息相同或相似；接着，模型训练装置104根据该多个相同或相似的发音信息，确定该多个英文语音数据所对应的相同的英文数据；模型更新装置统计该相同的英文数据所对应的英文语音数据在该语音数据中的出现次数；接着，该模型更新装置再根据该英文语音数据的出现次数，并结合该语音数据的数量信息，确定该英文数据的概率值，如将该出现次数除以该大量语音数据的数量信息，获得该英文数据的概率值；该模型更新装置再将该英文数据及其对应的概率值存入该英文语音模型中，以实现对该英文语音模型的更新。

[0045] 优选地，该模型训练设备1还包括数据提取装置(未示出)和单元训练装置(未示出)。该数据提取装置当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；该单元训练装置根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。具体地，该模型训练设备1可以设置一个预定数量阈值，该预定数量阈值用来判断英文数据所对应的英文语音数据在该语音数据中的出现次数是否超过该值，当该出现次数大于该预定数量阈值时，数据提取装置自该英文语音模型中提取该英文数据与该英文语音数据；该单元训练装置根据该英文数据与该英文语音数据的映射关系，为该英文语音数据新建一个英文语音单元，以实现对该英文语音单元的建立或更新。

[0046] 更优选地，该单元训练装置根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。具体地，该单元训练装置根据该英文数据与该英文语音数据的映射关系，为该英文数据新建一英文语音单元，再例如通过将该中文发音信息所对应的中文语音单元的模型参数复制给该英文语音单元，然后在英文数据上对齐，得到统计量，再与该复制所获得的参数进行插值计算，以实现对该英文语音单元的建立或更新。在此，所述中文语音单元包括为每个中文发音信息，如中文声母或韵母信息所建立的发音单元，如“s”、“ei4”等中文声韵母信息所对应的中文语音单元。

[0047] 更优选地，该模型训练设备1还包括单元更新装置(未示出)。该单元更新装置根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。具体地，在单元训练装置为英文数据新建一英文语音单元之后，单元更新装置不断地根据用户输入的英文语音数据，及由语音识别装置103识别所获得的与该英文语音数据相对应的发音信息，训练该英文语音单元。

the skilled artisan will appreciate that "persistent" means each device model training equipment are constantly acquire voice data and establish judgment English voice data recognition processing and English voice English voice data model or updating until the model training device 1 in a long time I stop getting voice data.

[0037] Preferably, the speech recognition device 103 when the speech data is English voice data, according to the word model corresponding to the Chinese voice message, the English voice data to voice recognition processing to obtain the English and speech data corresponding word pronunciation information. Specifically, when the English voice judging means 102 determines device 101 acquired by the speech data acquired for the English voice data, the voice recognition device 103 using the word model corresponding to the Chinese voice message, such as "aaai4 f engl", etc., on the English voice data for voice recognition processing, such as forced alignment treatment, using all possible Chinese vowel sound information, the mandatory English speech data aligned to obtain the corresponding English voice data word pronunciation information.

[0038] In this case, the word model stored in a large number of words corresponding to the Chinese voice message, the word model in the model training device 1, or in a third party through a network connected with the model training device 1 devices.

[0039] Preferably, the speech recognition device 103 when the speech data is English voice data, based on the letter corresponding to the Chinese model of voice information, data on the English voice for voice recognition processing to obtain the English and Voice data information corresponding letter sounds. Specifically, when the English voice judging means 102 determines device 101 acquired by the speech data acquired for the English voice data, the voice recognition device 103 according to the letter corresponding to the Chinese model of voice information, such as twenty-six the letters corresponding to the Chinese pronunciation information, such as the letter "a" corresponding to the Chinese voice message "aa eil", the letter "b" corresponding Chinese voice message "b i4", the letter "c" corresponding to the Chinese voice messages "s ei4" and so on, the English voice Data for voice recognition

[0048] 在此，本发明对出现此处大于预定阈值的英文语音数据，为该英文语音数据及英文数据新建英文语音单元，并对该英文语音单元不断地进行训练，以进一步提升该英文语音数据的发音信息的准确性。

[0049] 图2示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的设备示意图。该模型训练设备1还包括序列获取装置205、序列确定装置206和信息提供装置207。以下参照图2对该优选实施例进行详细描述:具体地，语音获取装置201获取用户通过语音采集装置输入的语音数据；英文判断装置202判断所述语音数据是否为英文语音数据；语音识别装置203当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置204根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射；序列获取装置205获取所述用户为目标应用输入的单词语音序列；序列确定装置206根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；信息提供装置207将与所述英文字符序列相对应的信息提供给所述用户。其中，语音获取装置201、英文判断装置202、语音识别装置203和模型训练装置204与图1所示对应装置相同或相似，故此处不再赘述，并通过引用的方式包含于此。

[0050] 其中，序列获取装置205获取所述用户为目标应用输入的单词语音序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入单词语音序列，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的单词语音序列。

[0051] 序列确定装置206根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列。具体地，序列确定装置206根据该序列获取装置205所获取的单词语音序列，通过语音识别处理，确定该单词语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了单词语音序列“iphone”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置206根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音

processing, in order to obtain the English voice data information corresponding letter sounds.

[0040] Preferably, the speech recognition apparatus 103 for English English determining means 102 determines that the voice data is voice data, priority to Chinese word models corresponding to speech information, the English voice data for voice recognition processing, when the the speech recognition apparatus using the word model 103 can not recognize the English pronunciation information of voice data, and then using the letter model corresponding to the speech information in Chinese, English speech data to the speech recognition process, to obtain the corresponding English voice data pronunciation information letters. Similarly, the speech recognition device 103 may be preferably used in Chinese letter model corresponding to the voice information, the English voice data for voice recognition processing, when the use of the voice recognition device 103 recognizes that the letters can not model speech data in English pronunciation information When, then using word models corresponding to speech information in Chinese, English speech data to the speech recognition process, to obtain the corresponding English voice data word pronunciation information. Here, the word models using priority or precedence letter model corresponding to the speech information in Chinese, English speech data to the speech recognition processing, the model may be set in advance training device 1 may also be adjusted to be an event.

[0041] For example, the voice data acquired by the speech device 101 acquired "iphone", the English judgment means 102 judges that the speech data "iphone" for the English voice data, the voice recognition device 103 preferentially uses the word model corresponding Chinese voice messages, the English voice data for speech recognition processing, but does not recognize the English pronunciation of voice data information; Subsequently, the voice acquisition means 101 and then using the letter corresponding to the Chinese model of voice information, the English voice data for voice recognition processing, voice data obtained with the English pronunciation of letters corresponding to information "aa ai4 p i4 ei2ch oo oul ee enl ii i4", it is understood that the speech data acquired by the speech device 101 is acquired by letter read "iphone", That is, "1, p, h, ?, eta, e".

信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”。

[0052] 信息提供装置207将与所述英文字符序列相对应的信息提供给所述用户。例如，当该目标应用为语音输入应用，该信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列作为语音输入候选词条提供给该用户；当该目标应用为语音搜索应用；该信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将根据该英文字符序列搜索获得的一个或多个搜索结果提供给该用户。

[0053] 优选地，该目标应用为语音输入应用；其中，该信息提供装置207将所述英文字符序列作为语音输入候选词条提供给所述用户。例如，当该目标应用为语音输入应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置206根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”;进而，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入。

[0054] 优选地，该目标应用为语音搜索应用；其中，该模型训练设备1还包括搜索获取装置(未示出)。该搜索获取装置将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，该信息提供装置207将所述一个或多个搜索结果提供给所述用户。具体地，当该目标应用为语音搜索应用，序列确定装置206根据该英文语音模型，确定与该单词语音序列相对应的英文字符序列；搜索获取装置根据该序列确定装置206所确定的英文字符序列，将该英文字符序列作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该一个或多个搜索结果提供给该用户。

[0055] 例如，当该目标应用为语音搜索应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置206根据该单词

[0042] More preferably, the speech recognition apparatus 103 word models respectively corresponding to the voice information Chinese, Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, voice recognition processing for respectively The results obtained for scoring, scoring higher pronunciation information to determine that the English pronunciation of voice data information.

[0043] In this case, since the user when entering the English word sometimes word by word and by letter read read sometimes, the present invention requires the identification of each English word, use a similar word model corresponding Chinese voice messages, letters, correspondence model The Chinese voice information to determine its corresponding pronunciation information, improve the accuracy of speech recognition.

[0044] In a preferred embodiment, the apparatus further includes a model training model updating means (not shown). The statistical model updating means of said data corresponding to the English English number of occurrences of the voice data in the speech data; according to the number of occurrences, and the amount of data combined with the voice message, the probability of determining the data value of the English; based on the probability value, updating the English voice model. In particular, when a large number of user input via voice speech data acquisition device, wherein a plurality of the English voice data input by the user to the same English voice data, the voice acquisition means 101 acquires the number of users of a large number of input voice data; English judging means 102 Analyzing the bulk voice data into English voice data; then, a voice recognition device 103 according to Chinese voice message, the voice data of a large number of English speech recognition processing, in order to obtain the large number of English speech data corresponding pronunciation information, in this case, since The plurality of user inputs to the same English voice data, therefore, the English voice data on the plurality of user input of the voice recognition device 103 to obtain information corresponding to the same or similar pronunciation; then, based on the model training device 104 a plurality of the same or similar pronunciation information, identify same English voice data of the plurality of data corresponding English; model updating means statistical data of

语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”;进而，搜索获取装置根据该序列确定装置206所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0056] 更优选地，序列获取装置205获取所述用户为目标应用输入的字母语音序列；其中，所述序列确定装置206根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入字母语音序列，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的字母语音序列；随后，序列确定装置206根据该序列获取装置205所获取的字母语音序列，通过语音识别处理，确定该字母语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了字母语音序列“1、p、h、o、n、e”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的字母语音序列“1、p、h、o、n、e”;随后，序列确定装置206根据该字母语音序列“1、p、h、o、n、e”，通过语音识别处理，确定该字母语音序列的发音信息“aa ai4 p i4 ei2 ch oo oul ee enl ii i4”,进而，根据该发音信息,在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列“iphone”。

[0057] 相似地，信息提供装置207将与该英文字符序列相对应的信息提供给所述用户。接上例，当该目标应用为语音输入应用，该信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入；当该目标应用为语音搜索应用；搜索获取装置根据该序列确定装置206所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序

the same English voice data corresponding to the number of occurrences of English in the speech data; then, The model updating means then according to the number of occurrences of the English voice data, combined with the amount of information of the voice data, to determine the probability that the value of the English data, such as the emergence of a large number of voice data divided by the amount of information, access to the English data The probability value; the model data updating means then the English and the corresponding probability values are stored in the English speech model in order to achieve the English voice model update.

[0045] Preferably, the apparatus 1 further comprises a model training data extracting means (not shown) and the unit training device (not shown). The data extracting means when the number of occurrences is greater than a predetermined threshold amount, the English data and extract the data from the English voice English voice model; the unit training apparatus according to the English and the English voice data mapping data relationship, create or update data with the corresponding English voice English voice unit. Specifically, the model training device 1 may set a predetermined threshold amount, the predetermined threshold amount used to determine the English data corresponding English voice data the number of occurrences in the speech data exceeds this value, when the number of occurrences is greater than the predetermined When the threshold number, data extraction means extracts the data associated with the English voice English voice data from the English model; the unit training device mapping relation to the English voice data based on the English data, create a new unit for the English voice English voice data in order to achieve the English phonetic units built or updated.

[0046] More preferably, the unit training device data according to the mapping relation between the English and the English voice data, using interpolation method, create or update the English voice unit. Specifically, the unit training device mappings between data based on the English, and the English voice data, a new English voice unit data for the English, then by the model parameters such as Chinese pronunciation information corresponding Chinese phonetic units copied to the English speech unit, and then on the English data alignment, obtain statistics, and then with the

列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0058] 在此，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的使用体验。

[0059] 图3示出根据本发明另一个方面的用于基于中文语音信息训练英文语音模型的方法流程图。

[0060] 在步骤S301中，模型训练设备1获取用户通过语音采集装置输入的语音数据。具体地，用户或大量用户通过与语音采集装置的交互，输入了语音数据，在步骤S301中，模型训练设备1通过调用该语音采集装置所提供的应用程序接口(API)，获取该用户或大量用户输入的语音数据；或者，用户或大量用户通过与语音采集装置的交互，输入了语音数据，并将该语音数据存入语音数据库中，在步骤S301中，模型训练设备1通过调用该语音数据库所提供的应用程序接口(API)，从该语音数据库中获取用户或大量用户输入的语音数据。

[0061] 本领域技术人员应能理解上述获取语音数据的方式仅为举例，其他现有的或今后可能出现的获取语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0062] 在步骤S302中，模型训练设备1判断所述语音数据是否为英文语音数据。具体地，在步骤S302中，模型训练设备1根据在步骤S301中所获取的用户通过语音采集装置输入的语音数据，例如通过声学模型匹配等方式，判断该语音数据是否为英文语音数据。

[0063] 本领域技术人员应能理解上述判断英文语音数据的方式仅为举例，其他现有的或今后可能出现的判断英文语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0064] 在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息。具体地，当在步骤S302中，模型训练设备1判断该语音数据为英文语音数据，在步骤S303中，模型训练设备1根据中文语音信息，如中文声韵母信息，对该英文语音数据进行语音识别处理，例如，当在步骤S302中，模型训练设备1判断该语音数据为英文语音数据时，在步骤S303中，模型训练设备1采用所有可能的中文

parameters of the copy obtained by interpolating calculations to achieve the establishment or update the English voice unit. Here, the Chinese pronunciation for each speech unit includes Chinese information, such as Chinese initial or final information established pronunciation unit, such as "s", "ei4" and other Chinese vowel sound information corresponding Chinese phonetic units.

[0047] More preferably, the model training apparatus 1 further comprises a unit updating means (not shown). The unit updating means according to the English voice data and voice data corresponding to the English pronunciation information, updating the English voice unit. Specifically, after the unit training device a new English voice unit data is in English, means constantly updating means according to the English voice user input data, and 103 recognized by the voice recognition device obtained with the English voice data corresponding pronunciation information training the English voice unit.

[0048] In this case, the present invention is greater than a predetermined threshold value appears here in English, voice data, voice and data for English and English New English voice unit data, and the English phonetic units constantly training to further enhance The accuracy of the English pronunciation of voice data information.

[0049] Figure 2 shows a schematic diagram of the device based on the Chinese model of voice information in accordance with English voice training example for a preferred embodiment of the present invention. The model training apparatus 1 further comprises a sequence acquisition means 205, determining means 206 and the sequence information providing apparatus 207. Referring now to Figure 2 of this preferred embodiment is described in detail: specifically, the voice acquisition means 201 acquires voice data through the voice collecting user input means; whether the speech data judging means 202 judges English English speech data; speech recognition means 203 When the voice data is voice data in English, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the English voice data corresponding pronunciation information; model training device 204 information based on the pronunciation, determine the English voice data corresponding English data, in order to establish or update the English voice model,

声韵母信息去识别该英文语音数据，以确定该英文语音数据的发音信息。

[0065] 例如，用户通过与语音采集装置的交互，输入了语音数据单词“iphone”，在步骤S301中，模型训练设备1通过调用该语音采集装置所提供的应用程序接口(API)，获取了该语音数据“iphone”;随后，在步骤S302中，模型训练设备1采用声学模型匹配的方式，判断该语音数据为英文语音数据；随后，在步骤S303中，模型训练设备1采用所有可能的中文声韵母信息，对该英文语音数据“iphone”进行强制对齐，确定该英文语音数据“iphone”的发音信息为“aa ai4f engl”。在此，aa、ai4、f、engl等即为中文声韵母信息,其中的数字表示该韵母信息所对应的声调。本领域技术人员应能理解，在此，中文声韵母信息及声调的表示仅为示例，而不应对本发明产生任何限制。

[0066] 优选地，在步骤S303中，模型训练设备1还可以根据单词模型所对应的中文语音信息、或字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息、或字母发音信息。例如，对于“iphone”，若用户输入语音采集装置的为按单词读的“iphone”，则在步骤S303中，模型训练设备1确定其对应的采用中文语音信息表示的发音信息为“aa ai4 fengl”;若用户输入语音采集装置的为按字母读的“iphone”，即，“1、p、h、o、n、e”，则在步骤S303中，模型训练设备1确定其对应的采用中文语音信息表示的发音信息为“aa ai4 pi4 ei2 ch oo oul ee enl iii4”。

[0067] 本领域技术人员应能理解上述识别发音信息的方式仅为举例，其他现有的或今后可能出现的识别发音信息的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0068] 在步骤S304中，模型训练设备1根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。具体地，在步骤S304中，模型训练设备1根据在步骤S303中所识别获得的发音信息，确定该英文语音数据所对应的英文数据，如根据发音信息“aa ai4 fengl”，确定该发音信息对应的英文数据为“iphone”，再将该发音信息与该英文数据的映射关系存入英文语音模型，以实现对该英文语音模型的建立或更新，其中，该发音信息与该英文数据通过该中文语音信息相映射。

wherein the pronunciation information and data through the English voice information is mapped to the Chinese; sequence acquisition means 205 acquires the target users word voice input sequence applications; sequence determination means 206 according to the English voice model to determine the sequence of words corresponding to the English voice character sequence; information providing device 207 will provide the English character sequence information corresponding to the said user. Wherein, the voice acquisition means 201, judging means 202 in English, the speech recognition apparatus corresponding to 203 in FIG. 1 and FIG model training apparatus 204 same as or similar means, it is not mentioned here, and is incorporated herein by reference.

[0050] wherein the sequence acquisition means 205 acquires the user word voice input sequence of the target application. Specifically, for example, by the user to interact with the voice collecting device user device, enter the word voice in the target application, sequence acquisition means 205, for example by calling the user equipment provided by the application program interface (API), etc., to obtain the user Target applications for the word voice input sequence.

[0051] sequence determination means 206 according to the English voice model to determine the sequence of words corresponding to the English speech sequence of characters. Specifically, the sequence determining means 206 to obtain the sequence of words based on the speech unit sequence 205 acquired through voice recognition processing, determines the word sequence of speech pronunciation information, and then, based on the pronunciation information, the match query model in the English speech determining the word sequence corresponding to the English voice character sequence. For example, the user interacts with the voice collecting device user device, enter the word speech sequence "iphone", sequence acquisition means 205, for example by calling the user equipment provides an application program interface (API), etc., to get the user input word speech sequence "iphone"; then, the sequence determining means 206 according to the word speech sequence "iphone", through voice recognition processing, determines the word sequence of speech pronunciation information "aa ai4 f engl", and then, based on the pronunciation

[0069] 在此，英文数据为一具体的英文单词，如“iphone”等，而英文语音数据为该具体的英文单词所对应的语音数据，如“iphone”所对应的语音数据，通过语音获取装置101所获取。

[0070] 本领域技术人员应能理解上述建立或更新英文语音模型的方式仅为举例，其他现有的或今后可能出现的建立或更新英文语音模型的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0071] 在此，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

[0072] 优选地，模型训练设备1的各个步骤之间是持续不断工作的。具体地，在步骤S301中，模型训练设备1获取用户通过语音采集装置输入的语音数据；在步骤S302中，模型训练设备1判断所述语音数据是否为英文语音数据；在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；在步骤S304中，模型训练设备1根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。在此，本领域技术人员应理解“持续”是指模型训练设备1的各步骤分别不断地进行语音数据的获取、英文语音数据的判断、英文语音数据的识别处理及英文语音模型的建立或更新，直至该模型训练设备1在较长时间内停止获取语音数据。

[0073] 优选地，在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。具体地，当在步骤S302中，模型训练设备1判断在步骤S301中所获取的语音数据为英文语音数据，在步骤S303中，模型训练设备1采用单词模型所对应的中文语音信息，如“aa ai4 f engl”等，对所述英文语音数据进行语音识别处理，如强制对齐处理，采用所有可能的中文声韵母信息，对该英文语音数据进行强制对齐，以获得与所述英文语音数据相对应的单词发音信息。

[0074] 在此，该单词模型中存储有大量单词所对应的中文语音信息，该单词模型可以位于模型训练设备1中，或者位于与该模型训练设备1通过网络相连接的第三方设备中。

information, match queries in the English voice model, determined voice the word sequence corresponding English character sequence "iphone".

[0052] information providing device 207 will be provided to the user and the English character sequence corresponding information. For example, when the target application for voice input application, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., the English character sequence as a candidate for entry speech input provided to the user; when the target applications voice search application; the information providing device 207 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will be provided to the user in accordance with one or more of the search results for that search English character sequence obtained.

[0053] Preferably, the target applications for voice input applications; wherein the information providing device 207 as the English voice input candidate character sequence entries available to the user. For example, when the target application for voice input application, users interact with the voice of the user equipment acquisition device, enter the voice input applications, the word speech sequence "iphone", sequence acquisition means 205, for example by calling the user equipment provided Application Program Interface (API), etc., to obtain the sequence of the word voice input by the user "iphone"; then the sequence determination means 206 voice sequence information based on the word "iphone", through voice recognition processing, determine the sequence of the word voice pronunciation "aa ai4 f engl", and then, according to the pronunciation information, the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; In addition, the information providing device 207, such as by calling JSP, ASP or PHP and other dynamic web technology, the English character sequence "iphone" candidate for entry as a voice input provided to the user, for the user to select the input.

[0054] Preferably, the target application for the voice search application; wherein the model training apparatus 1 also includes a search acquiring device (not shown). The search acquisition means the English character sequence

[0075] 优选地，在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。具体地，当在步骤S302中，模型训练设备1判断在步骤S301中所获取的语音数据为英文语音数据，在步骤S303中，模型训练设备1根据字母模型所对应的中文语音信息，如二十六个英文字母所对应的中文语音信息，如字母“a”对应的中文语音信息“aa eil”、字母“b”对应的中文语音信息“b i4”、字母“c”对应的中文语音信息“s ei4”等，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。

[0076] 优选地，在步骤S303中，模型训练设备1对于在步骤S302中判断为英文语音数据为语音数据，优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当在步骤S303中，模型训练设备1采用单词模型无法识别出该英文语音数据的发音信息时，再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。相似地，在步骤S303中，模型训练设备1也可以优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当在步骤S303中，模型训练设备1采用字母模型无法识别出该英文语音数据的发音信息时，再采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息。在此，优先采用单词模型还是优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，可以是该模型训练设备1预先设定的，也可以应事件而进行调整。

[0077] 例如，对于在步骤S301中，模型训练设备1所获取的语音数据“iphone”，在步骤S302中，模型训练设备1判断该语音数据“iphone”为英文语音数据，在步骤S303中，模型训练设备1优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，但无法识别出该英文语音数据的发音信息；随后，在步骤S301中，模型训练设备1再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，获得与该英文语音数据相对应的字母发音信息“aa ai4 p i4 ei2 ch oo oul ee enl ii i4”,则可知在步骤S301中所获取的语音数据为按字母读的“iphone”，即，“1、p、h、o、n、e”。

[0078] 更优选地，在步骤S303中，模型训练设备1分别采用单词模型所对应的中文语音信息、字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，并分别对语音识别处理所获得的结

as the user's query sequence to obtain the query sequence corresponding to one or more search results; wherein the information providing device 207 the one or more search results provided to the user. Specifically, when the target application is a voice search applications, sequence determination means 206 according to the English voice model to determine the word speech sequence corresponding English character sequences; search acquisition means 206 according to the determined English character sequence determines the sequence, the English character sequences as the query sequence for the user, to match the query, get the query sequence corresponding to one or more search results; subsequently, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., the one or more search results to the user.

[0055] For example, when the target application for the voice search application, users interact with the voice collecting device user equipment, applications in voice input Voice input word sequence "iphone", sequence acquisition means 205, for example by calling The equipment provided by the user application program interface (API), etc., to get the word speech sequence of the user input "iphone"; then the sequence determination means 206 according to the word speech sequence "iphone", through voice recognition process, determining the voice pronounce the word sequence information "aa ai4 f engl", and then, according to the pronunciation information, in the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; thus, search Being apparatus according to the sequence determined by determining the sequence of characters in English 206 device, the English character sequence "iphone" as the query sequence for the user, to match the query, get the query sequence corresponding to one or more search results; subsequently, information providing device 207 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will be provided to the English character sequence "iphone" corresponding to one or more search results to the user, for the user to browse.

[0056] More preferably, the sequence acquisition means 205 acquires the target application users voice input letter sequence; wherein, said

果进行打分，确定打分较高的发音信息为该英文语音数据的发音信肩、O

[0079] 在此，由于用户在输入英文单词时有时按单词读有时按字母读，本发明对每个需要识别的英文单词，利用相似的单词模型对应的中文语音信息、字母模型对应的中文语音信息，确定其对应的发音信息，提升了语音识别的准确性。

[0080] 在一优选实施例中，该方法还包括步骤S308(未示出)。在步骤S308中，模型训练设备1统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；根据所述概率值，更新所述英文语音模型。具体地，当大量用户通过语音采集装置输入语音数据，其中，多个用户所输入的英文语音数据为相同的英文语音数据，在步骤S301中，模型训练设备1获取该大量用户输入的大量语音数据；在步骤S302中，模型训练设备1判断该大量语音数据为英文语音数据；随后，在步骤S303中，模型训练设备1根据中文语音信息，对该大量英文语音数据进行语音识别处理，以获得与该大量英文语音数据相对应的发音信息，在此，由于该多个用户输入的为相同的英文语音数据，因此，在步骤S303中，模型训练设备1所获得的关于该多个用户输入的英文语音数据所对应的发音信息相同或相似；接着，在步骤S304中，模型训练设备1根据该多个相同或相似的发音信息，确定该多个英文语音数据所对应的相同的英文数据；在步骤S308中，模型训练设备1统计该相同的英文数据所对应的英文语音数据在该语音数据中的出现次数；接着，在步骤S308中，模型训练设备1再根据该英文语音数据的出现次数，并结合该语音数据的数量信息，确定该英文数据的概率值，如将该出现次数除以该大量语音数据的数量信息，获得该英文数据的概率值；该在步骤S308中，模型训练设备1再将该英文数据及其对应的概率值存入该英文语音模型中，以实现对该英文语音模型的更新。

[0081] 优选地，该方法还包括步骤S309(未示出)和步骤S310(未示出)。在步骤S309中，当所述出现次数大于预定数量阈值，模型训练设备1自所述英文语音模型中提取所述英文数据与所述英文语音数据；在步骤S310中，模型训练设备1根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。具体地，该模型训练设备1可以设置一个预定数量阈值，该预定数量阈值用来判断英文数据所对应的英文语音数据在该语音数据中的出现次数是否超过该值，当该出现次数大于该预定数量阈值时，在步骤S309中，模型训练设备1自该英文语音模型中提取

sequence determining means 206 according to the English speech model, determining the phase sequence of letters speech corresponding English character sequence. Specifically, for example, by the user to interact with the voice collecting device user device, enter the letters speech sequence in the target application, sequence acquisition means 205, for example by calling the user equipment provides an application program interface (API), etc., to obtain the user the letter sequence of the target speech input application; Subsequently, sequence determination means 206 acquires speech sequence of letters apparatus 205 according to the sequence obtained by the speech recognition processing for determining the sequence of letters voice pronunciation information, and then, based on the pronunciation information, match queries in the English speech model, determining the sequence of letters corresponding English voice character sequence. For example, user interaction with the voice collecting device user's device, voice input letter sequence "1, p, h, o, n, e", the sequence to obtain the application by calling the interface device 205 such as provided by the User Equipment (API), etc., to obtain the letters of the user input speech sequence "1, p, h, o, n, e"; then, the sequence determining means 206 based on the letter speech sequence "1, p, h, o, n, e ", through voice recognition processing, it is determined that the sequence of letters voice pronunciation information" aa ai4 p i4 ei2 ch oo oul ee enl ii i4 ", and then, based on the pronunciation information, English voice model in the matching query, determining The letter corresponding to a sequence of English voice character sequence "iphone".

[0057] Similarly, the information providing device 207 will be provided to the user with the English character sequence corresponding information. Connect embodiment, when the target application for voice input application, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., the English character sequence "iphone" as a voice input candidate terms provided to the user, for the user to select the input; if the target application is a voice search application; search acquiring apparatus according to the English character sequence determined by the sequence determination means 206, the English character sequence "iphone" as the query sequence for the user, to match the query, Get in touch with the

该英文数据与该英文语音数据；在步骤S310中，模型训练设备1根据该英文数据与该英文语音数据的映射关系，为该英文语音数据新建一个英文语音单元，以实现对该英文语音单元的建立或更新。

[0082] 更优选地，在步骤S310中，模型训练设备1根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。具体地，在步骤S310中，模型训练设备1根据该英文数据与该英文语音数据的映射关系，为该英文数据新建一英文语音单元，再例如通过将该中文发音信息所对应的中文语音单元的模型参数复制给该英文语音单元，然后在英文数据上对齐，得到统计量，再与该复制所获得的参数进行插值计算，以实现对该英文语音单元的建立或更新。在此，所述中文语音单元包括为每个中文发音信息，如中文声母或韵母信息所建立的发音单元，如“s”、“ei4”等中文声韵母信息所对应的中文语音单元。

[0083] 更优选地，该方法还包括步骤S311 (未示出)。在步骤S311中，模型训练设备1根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。具体地，在步骤S310中，模型训练设备1为英文数据新建一英文语音单元之后，在步骤S311中，模型训练设备1不断地根据用户输入的英文语音数据，及由在步骤S303中识别所获得的与该英文语音数据相对应的发音信息，训练该英文语音单元。

[0084] 在此，本发明对出现此处大于预定阈值的英文语音数据，为该英文语音数据及英文数据新建英文语音单元，并对该英文语音单元不断地进行训练，以进一步提升该英文语音数据的发音信息的准确性。

[0085] 图4示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的方法流程图。以下参照图4对该优选实施例进行详细描述:具体地，在步骤S401中，模型训练设备1获取用户通过语音采集装置输入的语音数据；在步骤S402中，模型训练设备1判断所述语音数据是否为英文语音数据；在步骤S403中，模型训练设备1当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；在步骤S404中，模型训练设备1根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射；在步骤S405中，模型训练设备1获取所述用户为目标应用输入的单词语音序列；在步骤S406中，模型训练设备1根据所述英文语音模型，确定与所述单词语音序列相对应的英文

query sequence corresponding to one or more search results; subsequently, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., will work with the English character sequence "iphone" corresponding to one or more Results provided to the user, for the user to browse.

[0058] In this case, the present invention can also be combined with voice recognition that allows users to press letters and press Enter word English speech sequence in two ways to improve the user experience.

[0059] Figure 3 shows a flowchart of a method Chinese voice information model-based training English voice to another aspect of the present invention is used in accordance with.

[0060] In step S301, the model training device 1 acquires the user voice data input through the voice collecting apparatus. Specifically, the user or a large number of users to interact with the voice collecting device, the input voice data, in step S301, the model training device 1 by calling the application program interface that provides voice collecting apparatus (API), to obtain the user or a large number of voice data input by the user; or, users or large number of users through an interactive voice acquisition device, the input voice data and voice data stored in the voice database, in step S301, the model training device 1 by calling the voice database application program interface provided by (API), to obtain a large number of users or user voice data input from the speech database.

[0061] skilled in the art should be able to understand the way the acquired speech data is only an example, other existing or future way to obtain voice data that may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0062] In step S302, the speech model training apparatus 1 determines whether the data is the English voice data. Specifically, in step S302, the model based on the training apparatus 1 in step S301 the user acquired by the speech data input voice collecting means, e.g., by matching acoustic models, etc., it is determined whether the speech data is speech data in English.

字符序列；在步骤S407中，模型训练设备1将与所述英文字符序列相对应的信息提供给所述用户。其中，步骤S401-S404与图3所示对应步骤相同或相似，故此处不再赘述，并通过引用的方式包含于此。

[0086] 在步骤S405中，模型训练设备1获取所述用户为目标应用输入的单词语音序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入单词语音序列，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的单词语音序列。

[0087] 在步骤S406中，模型训练设备1根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列。具体地，在步骤S406中，模型训练设备1根据在步骤S405中所获取的单词语音序列，通过语音识别处理，确定该单词语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，在步骤S406中，模型训练设备1根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aaai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”。

[0088] 在步骤S407中，模型训练设备1将与所述英文字符序列相对应的信息提供给所述用户。例如，当该目标应用为语音输入应用，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列作为语音输入候选词条提供给该用户；当该目标应用为语音搜索应用；在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将根据该英文字符序列搜索获得的一个或多个搜索结果提供给该用户。

[0089] 优选地，该目标应用为语音输入应用；其中，在步骤S407中，模型训练设备1将所述英文字符序列作为语音输入候选词条提供给所述用户。例如，当该目标应用为语音输入应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”;随后，在步骤S406中，模型训练

[0063] skilled in the art should be able to understand the judgment of the English way of example only voice data, other available data to determine the English voice mode or may occur in the future as applicable to the present invention, it should also be included Within the scope of the present invention and hereby incorporated herein by reference.

[0064] In step S303, the model training device 1 when the speech data is English voice data, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the data with English voice corresponding pronunciation information. In particular, when in step S302, the model training device 1 determines the speech data into English voice data, in step S303, the model training apparatus according to Chinese voice information, such as Chinese vowel sound information, the English voice data for voice recognition treatment, for example, when in step S302, the model training device 1 determines the speech data into English voice data, in step S303, the model training device 1 uses all possible Chinese vowel sound information to identify the English voice data to determine The English pronunciation of voice data information.

[0065] For example, the user interacts with the voice collecting device, voice data input word "iphone", in step S301, the model training device 1 by calling the voice collecting apparatus provided application programming interface (API) obtain the speech data "iphone"; Subsequently, in step S302, the model uses an acoustic model training device 1 matching methods to determine if the voice data is the English voice data; Subsequently, in step S303, the model uses all the training equipment 1 Possible Chinese vowel sound information, the English voice data "iphone" to force alignment, determine the English voice data "iphone" pronunciation information "aa ai4f engl". Here, aa, ai4, f, engl and other information that is Chinese vowel sound, which indicates that the vowel digital information corresponding to the tone. The skilled in the art will appreciate that in this, Chinese vowel sound and tonal representation of information are examples only, and not to deal with the present invention, any restrictions.

[0066] Preferably, in the step S303, the model training device 1 may also be based on a word model corresponding to the speech information in

设备1根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“ iphone”;进而，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入。

[0090] 优选地，该目标应用为语音搜索应用；其中，该方法还包括步骤S412 (未示出)。在步骤S412中，模型训练设备1将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，在步骤S407中，模型训练设备1将所述一个或多个搜索结果提供给所述用户。具体地，当该目标应用为语音搜索应用，在步骤S406中，模型训练设备1根据该英文语音模型，确定与该单词语音序列相对应的英文字符序列；在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该一个或多个搜索结果提供给该用户。

[0091] 例如，当该目标应用为语音搜索应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，在步骤S406中，模型训练设备1根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”;进而，在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0092] 更优选地，在步骤S405中，模型训练设备1获取所述用户为目标应用输入的字母语音序列；其中，在步骤S406中，模型训练设备1根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。具体地，用户例如通过与用户设备的语

Chinese, or Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, to get the English word voice data corresponding pronunciation information, or to pronounce the letter information. For example, for "iphone", if the user input voice collecting device is read by the word "iphone", then in step S303, the model training device 1 to determine the corresponding voice information indicating the use of Chinese pronunciation information "aa ai4 fengl "; if the user input device for voice collecting alphabetical read" iphone ", namely," 1, p, h, ?, n, e ", then in step S303, the model training device 1 to determine the corresponding adopt Chinese Voice messages indicated pronunciation information "aa ai4 pi4 ei2 ch oo oul ee enl iii4".

[0067] skilled in the art should be able to understand the way information is the identification pronunciation of example only, other available information the way for future identification pronunciation or may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0068] In step S304, the model training apparatus according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the pronunciation information and the English data by the Chinese voice information is mapped. Specifically, the English data corresponding English voice data in step S304, the model training device 1 in step S303, based on the identified information received pronunciation, identified as "aa ai4 fengl", determined in accordance with the pronunciation information corresponding pronunciation information English data "iphone", then the mapping of the English pronunciation information and data stored in the English speech model, in order to achieve the establishment or update English voice model, wherein the information associated with the English pronunciation of the Chinese data voice information is mapped.

[0069] In this case, the English data is a particular English word, such as "iphone" and so on, but the English voice data for specific English word corresponding to the voice data, such as the

音采集装置的交互，在目标应用中输入字母语音序列，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的字母语音序列；随后，在步骤S406中，模型训练设备1根据在步骤S405中所获取的字母语音序列，通过语音识别处理，确定该字母语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了字母语音序列“1、P、h、o、n、e”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的字母语音序列“1、P、h、0、n、e”；随后，在步骤S406中，模型训练设备1根据该字母语音序列“1、P、h、o、n、e”，通过语音识别处理，确定该字母语音序列的发音信息“aa ai4 pi4 ei2 ch oo oul ee enl ii i4”,进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列“iphone”。

[0093] 相似地，在步骤S407中，模型训练设备1将与该英文字符序列相对应的信息提供给所述用户。接上例，当该目标应用为语音输入应用，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入；当该目标应用为语音搜索应用；在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0094] 在此，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的使用体验

[0095] 本发明的软件程序可以通过处理器执行以实现上文所述步骤或功能。同样地，本发明的软件程序(包括相关的数据结构)可以被存储到计算机可读记录介质中，例如，RAM存储器，磁或光驱动器或软磁盘及类似设备。另外，本发明的一些步骤或功能可采用硬件来实现，例如，作为与处理器配合从而执行各个功能或步骤的电路。

[0096]另外，本发明的一部分可被应用为计算机程序产品，例如计算机程序指令，当其被计算机执行时，通过该计算机的操作，可以调用或提供根据本

"iphone" corresponding to the voice data, The acquisition means 101 acquire voice.

[0070] skilled in the art should be understood that the above to create or update the way of example only English voice model, or other existing way to create or update the English voice model may arise in the future as applicable to the present invention, It should also be included within the scope of the present invention and hereby incorporated herein by reference.

[0071] In this case, the present invention is in the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain the English voice data corresponding pronunciation information, to achieve create or update English voice model, effectively control the size of the English voice model.

[0072] Preferably, between each step model training apparatus 1 is continued work. Specifically, in step S301, the model training device 1 acquires the user via voice input speech data acquisition device; in step S302, the speech model training apparatus 1 determines whether the data is the English voice data; in step S303, the model train device 1 when the speech data is English voice data, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the English voice data corresponding pronunciation information; in step S304, the Model Training Equipment 1 according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the pronunciation information and data through the English voice information is mapped to the Chinese. In this case, the skilled artisan will appreciate that "persistent" refers to the steps of model training device 1 constantly acquire voice data respectively, set up to judge the English voice data recognition processing and English voice English voice data model or updating until the model training device 1 in a long time I stop getting voice data.

[0073] Preferably, in step S303, the model training device 1 when the speech data is English voice data, according to the word model corresponding to the Chinese voice message, the English voice data to voice recognition processing, in order to

发明的方法和/或技术方案。而调用本发明的方法的程序指令，可能被存储在固定的或可移动的记录介质中，和/或通过广播或其他信号承载媒体中的数据流而被传输，和/或被存储在根据所述程序指令运行的计算机设备的工作存储器中。在此，根据本发明的一个实施例包括一个装置，该装置包括用于存储计算机程序指令的存储器和用于执行程序指令的处理器，其中，当该计算机程序指令被该处理器执行时，触发该装置运行基于前述根据本发明的多个实施例的方法和/或技术方案。

[0097] 对于本领域技术人员而言，显然本发明不限于上述示范性实施例的细节，而且在不背离本发明的精神或基本特征的情况下，能够以其他的具体形式实现本发明。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本发明的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此夕卜，显然“包括”一词不排除其他单元或步骤，单数不排除复数。装置权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

obtaining voice data corresponding to the English word pronunciation information. Specifically, when the step S302, the voice data model training apparatus determines the acquired in step S301 English voice data, in step S303, the model uses the word model training device 1 corresponding to the Chinese voice message, such as "aa ai4 f engl "and so on, the English voice data for speech recognition processing, such as mandatory alignment process, using all possible Chinese vowel sound information, the English voice data to force alignment for the English voice data corresponding to word pronunciation information.

[0074] In this case, the word model stored in a large number of words corresponding to the Chinese voice message, the word model in the model training device 1, or in a third party through a network connected with the model training device 1 devices.

[0075] Preferably, in step S303, the model training device 1 when the speech data is English voice data, based on the letter corresponding to the Chinese model of voice information, data on the English voice for voice recognition processing, in order to obtaining voice data corresponding to the English pronunciation of letters information. Specifically, when the step S302, the voice data model training apparatus determines the acquired in step S301 English voice data, in step S303, the model training device 1 according to the letter corresponding to the Chinese model of voice information, such as the twenty six letters of the alphabet that corresponds to Chinese voice information, such as the letter "a" corresponding to the Chinese voice message "aa eil", the letter "b" corresponding Chinese voice message "b i4", the letter "c" corresponding to the Chinese voice message " s ei4 "and so on, the English voice data for speech recognition processing, in order to obtain the English voice data information corresponding letter sounds.

[0076] Preferably, in step S303, the model training device 1 is determined in step S302 for the English voice data as voice data, give priority to the word model corresponding to the Chinese voice message, the English voice data for voice When the recognition process, when in step S303, the model uses the word model training apparatus can not recognize the English pronunciation of voice data information, and then using the letter corresponding to the Chinese model of voice

information, the English voice data for speech recognition processing, in order to get the English voice data information corresponding to the pronunciation of letters. Similarly, in step S303, the model training device 1 may be given priority letter model corresponding to the Chinese voice message, the English voice data for speech recognition processing, when in step S303, the model training apparatus does not recognize the use of letter model When the pronunciation of the English voice data information, and then using the word model corresponding to the Chinese voice message, the English voice data for speech recognition processing, in order to obtain the English voice data corresponding word pronunciation information. Here, the word models using priority or precedence letter model corresponding to the speech information in Chinese, English speech data to the speech recognition processing, the model may be set in advance training device 1 may also be adjusted to be an event.

[0077] For example, in step S301, the model train equipment acquired 1 speech data "iphone", in step S302, the model training device 1 determines the speech data "iphone" for the English voice data, in step S303, the model training device 1 priority model using the corresponding Chinese word speech information, the English voice data for speech recognition processing, but does not recognize the English pronunciation of voice data information; then, in step S301, the model training equipment 1 and then using the letter corresponding to the Chinese model of voice information, the English voice for voice recognition processing data obtained with the English voice data information corresponding to pronounce the letter "aa ai4 p i4 ei2 ch oo oul ee enl ii i4", then Voice data showed acquired in step S301 is alphabetical read "iphone", namely, "1, p, h, ?, n, e".

[0078] More preferably, in step S303, the model training device 1 word models respectively corresponding to the voice information Chinese, Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, and Speech recognition processing results were obtained for scoring, scoring higher pronunciation information to determine the pronunciation for the English letter shoulder voice data, ?

[0079] In this case, due to user input English word sometimes word by word reading Sometimes alphabetical read, the present invention requires the identification of each English word, use a similar word model corresponding Chinese voice messages, letters, Chinese voice information corresponding to the model to determine the corresponding pronunciation information, improve the accuracy of speech recognition.

[0080] In a preferred embodiment, the method further comprises the step S308 (not shown). In step S308, a statistical model of the English training device data corresponding to the number of occurrences English voice data in the voice data; based on the number of occurrences, combined with the amount of information of the voice data, to determine the English the probability value of the data; based on the probability value, updating the English voice model. Specifically, when a large number of users via voice input speech data acquisition device, wherein a plurality of the English voice data input by the user to the same English voice data, in step S301, the model of the large number of training device 1 acquires a large number of voice data input by the user; in step S302, the model training device 1 judges that a lot of voice data into English voice data; Subsequently, in step S303, the model training apparatus according to Chinese voice information, a large number of English speech data to the speech recognition process to obtain The large number of English speech data corresponding pronunciation information, in this case, since the plurality of user input to the same English voice data, and therefore, in step S303, the model obtained training device 1 is about the plurality of user input in English voice data corresponding to the same or similar pronunciation information; then, in step S304, the model training device 1 according to the plurality of the same or similar pronunciation information, identify same English data of the plurality of data corresponding to the English voice; in step S308, a statistical model training equipment that same data in English English voice data corresponding to the number of occurrences in the speech data; then, in step S308, the model number of occurrences of re-training device 1 according to the English voice data, and combined with the number of messages in the voice data to determine the probability of the value of the English data, such as the emergence of a large number of voice data divided by the amount of information, the probability of obtaining

the value of the data in English; this in step S308, the model training device 1 again The English data and its corresponding probability values are stored in the voice model in English, in order to achieve the English voice model update.

[0081] Preferably, the method further comprises the step S309 (not shown) and step S310 (not shown). In step S309, when the number of occurrences is greater than a predetermined threshold number, model in English speech model training equipment to extract a data from said in English with the English voice data; in step S310, the model based on the training apparatus data mapping between the English and the English voice data, create or update data with the corresponding English voice English voice unit. Specifically, the model training device 1 may set a predetermined threshold amount, the predetermined threshold amount used to determine the English data corresponding English voice data the number of occurrences in the speech data exceeds this value, when the number of occurrences is greater than the predetermined When the threshold number, in step S309, the model training device 1 extracts the English data associated with the English voice data from the English voice model; step S310 in the model training device 1 mapping relationship with the English voice data based on the English data, create a new unit for the English voice English voice data, in order to achieve the English phonetic units built or updated.

[0082] More preferably, in step S310, the model training device 1 according to data mapping between the English and the English voice data, using interpolation method, create or update the English voice unit. Specifically, in step S310, the model training device 1 according to the English data mapping relation to the English voice data, a new English voice unit data for the English, and then, for example by the Chinese pronunciation information corresponding Chinese phonetic units copy the model parameters to the English speech unit, and then align the English data, obtain statistics, and then with the parameters of the copy obtained by interpolating calculations to achieve the establishment or update the English voice unit. Here, the Chinese pronunciation for each speech unit includes Chinese information, such as Chinese initial or final information established pronunciation unit, such as "s", "ei4"

and other Chinese vowel sound information corresponding Chinese phonetic units.

[0083] More preferably, the method further comprises the step S311 (not shown). In step S311, the model training apparatus according to the English voice data and voice data corresponding to the English pronunciation information, updating the English voice unit. After Specifically, in step S310, the model training device 1 is a new English voice data in English units, in step S311, the model continuously training device 1 according to the English voice user input data, and in step S303 is identified by the acquired English voice data with the corresponding pronunciation information, training the English voice unit.

[0084] In this case, the present invention is greater than a predetermined threshold value appears here in English, voice data, voice and data for English and English New English voice unit data, and the English phonetic units constantly training to further enhance The accuracy of the English pronunciation of voice data information.

[0085] Figure 4 shows an example of a flowchart of a method for training Chinese English speech voice information model-based accordance with a preferred embodiment of the present invention. Referring now to Figure 4 of the preferred embodiment described in detail: specifically, in step S401, the model training device 1 acquires voice data through the voice collecting user input means; in the step S402, the speech model training apparatus 1 determines the data Are the English voice data; step S403, the model training apparatus when the voice data is voice data in English, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the data with English voice corresponding pronunciation information; in step S404, the model training apparatus according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the information relating to the English pronunciation data by the Chinese voice information is mapped; in step S405, the model training device 1 acquires the user word voice input sequence of the target application; in step S406, the model training apparatus according to the English voice model to determine the The word speech sequence corresponding to a

sequence of characters in English; in step S407, the model will provide a training device with the English character sequence information corresponding to the user. Wherein the step of S401-S404 of FIG. 3 correspond to the same or similar steps, it is not mentioned here, and incorporated herein by reference.

[0086] In step S405, the model training device 1 acquires the user application for the target word voice input sequence. Specifically, the user, for example user equipment through interactive voice collecting device, enter the word speech sequence in the target application, in step S405, a model training equipment such as an application by calling the user equipment provided interface (API), etc. way, to get the user to the word voice input sequence of the target application.

[0087] In step S406, the model training apparatus according to the English voice model to determine the sequence of words corresponding to the English speech sequence of characters. Specifically, in step S406, the model training device 1 according to the word speech sequence in step S405 acquired, through voice recognition processing, determine the sequence of the pronunciation of the word voice information, and thus, according to the pronunciation information, in the English speech model In matching the query to determine the word speech sequence corresponding English character sequence. For example, a user with the user interaction device audio capture device, enter the word speech sequence "iphone", in step S405, a model training equipment such as an application program interface (API) calls by other means provided by the user equipment, Get the word speech sequence of the user input "iphone"; Subsequently, in step S406, the model training device 1 according to the word speech sequence "iphone", through voice recognition processing, determine the sequence of the word voice pronunciation information "aaai4 f engl "Furthermore, according to the pronunciation information, matching the query in this English voice model, determined voice the word sequence corresponding English character sequence" iphone ".

[0088] In step S407, the model training device 1 will be provided to the user and the English character sequence corresponding information. For example, when the target application to use

voice input, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., the English character sequence as a candidate for entry speech input provided to the user; when The target application for the voice search application; in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will be provided to the user in accordance with one or more of the search results for that search English character sequence obtained.

[0089] Preferably, the target applications for voice input applications; wherein, in step S407, the model training device 1 as the English voice input candidate character sequence entries available to the user. For example, when the target application for voice input application, the user through an interactive voice collecting device user devices with voice input application input speech sequence word "iphone", in step S405, the model training device 1 for example, by calling the user equipment provided by the application program interface (API), etc., to get the word speech sequence of the user input "iphone"; Subsequently, in step S406, the model training device 1 according to the word speech sequence "iphone", through voice recognition process to determine the sequence of the word voice pronunciation information "aa ai4 f engl", and then, according to the pronunciation information, in the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; Further, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., the English character sequence "iphone" as a voice input candidate terms provided to the user, for the user to select the input.

[0090] Preferably, the target application for the voice search application; wherein the method further comprises the step S412 (not shown). In step S412, a model training apparatus the English character sequence as the query sequence to the user, access to the query sequence corresponding to one or more search results; wherein, in step S407, the model training device 1 the one or more search results to the user. Specifically, when the target application is a voice search applications, in step S406, the model training device 1 according to the English voice model to determine the word

speech sequence corresponding English character sequences; in step S412, the model training device 1 according to English character sequence determined in step S406, the English character sequence as a query sequence to the user to match the query, get the query sequence corresponding to one or more search results; Subsequently, in step S407, the model Training Equipment 1 by calling dynamic web technology such as JSP, ASP or PHP, etc., the one or more search results to the user.

[0091] For example, when the target application is interactive voice acquisition device voice search applications, users and user devices, enter the word in the speech input speech sequence applications "iphone", in step S405, the model train for example, by calling the device a user equipment provided by the application program interface (API), etc., to get the word speech sequence of the user input "iphone"; Subsequently, in step S406, the model training device 1 according to the word speech sequence " iphone ", through voice recognition processing, determine the sequence of the word voice pronunciation information" aa ai4 f engl ", and then, according to the pronunciation information, the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; and then, in step S412, the model training device 1 according to the English character sequence determined in step S406, the English character sequence "iphone" as the query sequence for the user, to match the query, access and A Subsequently, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will work with the English character sequence "iphone" corresponding; the query sequence corresponding to one or more search results or more search results to the user, for the user to browse.

[0092] More preferably, in step S405, the model training device 1 acquires the user-letter sequence of the target application voice input; wherein, in step S406, the model training apparatus according to the English speech model, determining the sequence of letters corresponding English voice character sequence. Specifically, the user, for example user equipment through interactive voice collecting device, enter the letters speech sequence in the target application, in step S405, a model training equipment such as an application

by calling the user equipment provided interface (API), etc. way, to get the user to the letter sequence of the target application voice input; Subsequently, in step S406, the model training device 1 according to the letter sequence in step S405 voice acquired by the voice recognition processing, determine the sequence of voice pronounce the letter information, and thus, according to the pronunciation information, the match query in the English speech model, determining the sequence of letters corresponding English voice character sequence. For example, user interaction with the voice collecting device user's device, voice input letter sequence "1, P, h, ?, n, e", in step S405, the model training device 1 for example by calling in the user equipment is provided The application program interface (API) and other ways to get the user input speech sequence of letters "1, P, h, 0, n, e"; Subsequently, in step S406, the model training device 1 according to the letter speech sequence " 1, P, h, o, n, e ", through voice recognition processing, it is determined that the sequence of letters voice pronunciation information" aa ai4 pi4 ei2 ch oo oul ee enl ii i4 ", and then, based on the pronunciation information in the English voice model to match the query to determine the sequence of letters corresponding English voice character sequence "iphone".

[0093] Similarly, in step S407, the model training device 1 will be provided to the user with the English character sequence corresponding information. Connected cases, while providing the target application for voice input applications, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., the English character sequence "iphone" as a voice input candidate entry to the user, for the user to select the input; if the target application is a voice search application; in step S412, the model training device 1 according to the English character sequence determined in step S406, the English character sequence "iphone" as The user's query sequence, to match the query, get the query sequence corresponding to one or more search results; Subsequently, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will the English character sequence "iphone" corresponding to one or more search results to the user, for the user to browse.

[0094] In this case, the present invention can also be combined with voice recognition that allows users to press letters and press Enter word English speech sequence in two ways to improve the user experience

[0095] software program of the present invention may be executed to implement the above steps or functions by a processor. Similarly, the software program of the present invention (including associated data structures) can be stored in the computer-readable recording medium, for example, RAM memory, magnetic or optical drive or floppy disk and similar devices. In addition, some steps or functions of the present invention may be implemented in hardware, such as the processor with the circuit to perform various functions or steps.

[0096] In addition, the present invention can be applied as part of a computer program product, such as computer program instructions which, when executed by a computer, by operating the computer, you can call or provide the method according to the invention and / or Technical solutions. The method of the present invention is calling program instructions may be stored in fixed or removable recording medium, and / or are transmitted by radio or other signal bearing medium, data flow, and / or stored in accordance with the working memory of said computer program instructions to run the device. Here, according to one embodiment of the present invention comprises an apparatus embodiment, the apparatus includes a memory for storing computer program instructions and a processor for executing program instructions, wherein, when the computer program instructions being executed by the processor, trigger The device is running according to the aforementioned embodiments of the present invention a plurality-based methods and / or technology programs.

The case of

[0097] skilled artisan, the present invention is clearly not limited to the details of the above-described exemplary embodiment, but without departing from the spirit or essential characteristics of the present invention, it is possible to achieve this in other specific forms invention. Therefore, whether from what point of view, should be considered as an exemplary embodiment, but not limiting, the scope of the

present invention is defined by the appended claims rather than the foregoing description define, therefore intended to fall claim All changes which come within the meaning and range of equivalents of the elements within the present invention covers. The claims should not be seen as the right to restrict any reference signs covered requirements. Bu this evening, apparently "comprising" does not exclude other elements or steps, singular does not exclude a plurality. A plurality of units or devices recited in the claims means may be formed of a single unit or be implemented by means of software or hardware. First, second, etc. words used to indicate the name, but does not indicate any particular order.

Citing: CN101593173 ; CN101593173 A ; CN101650943 ; CN101650943 A ; CN101826325 ; CN101826325 A ; CN102193643 ; CN102193643 A ; JP2006092994 ; JP2006092994 A2 ;

Cited by: CN104951100 A ; CN105070289 A ; CN105244027 A ; WO18086033 A1 ;

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103632668 A
(43)申请公布日 2014.03.12

(21)申请号 201210299700.6

(22)申请日 2012.08.21

(71)申请人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街 10 号百度大厦 2 层

(72)发明人 苏丹 万广鲁

(74)专利代理机构 北京汉昊知识产权代理事务所(普通合伙) 11370
代理人 罗朋 周建华

(51)Int.Cl.
G10L 15/06(2013.01)
G10L 15/00(2013.01)

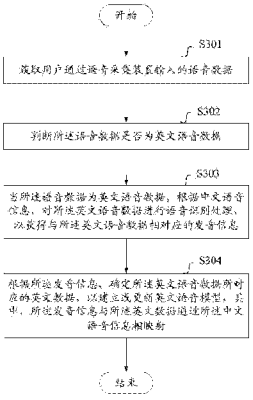
权利要求书3页 说明书14页 附图4页

(54)发明名称

一种基于中文语音信息训练英文语音模型的方法与设备

(57)摘要

本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备;获取用户通过语音采集装置输入的语音数据;判断该语音数据是否为英文语音数据;当该语音数据为英文语音数据,根据中文语音信息,对该英文语音数据进行语音识别处理,以获得与该英文语音数据相对应的发音信息;根据该发音信息,确定该英文语音数据所对应的英文数据,以建立或更新英文语音模型。与现有技术相比,本发明在原有中文声学模型基础上,利用相似的中文语音信息,对英文语音数据进行语音识别处理,以获得与该英文语音数据相对应的发音信息,实现对英文语音模型的建立或更新,有效地控制了英文语音模型的规模。



CN 103632668 A

Petitioner Exhibit 1004, Page 541 of 1052

1. 一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤：

a 获取用户通过语音采集装置输入的语音数据；

b 判断所述语音数据是否为英文语音数据；

c 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

d 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

2. 根据权利要求 1 所述的方法，其中，所述步骤 c 包括：

- 当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

3. 根据权利要求 1 所述的方法，其中，所述步骤 c 包括：

- 当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

4. 根据权利要求 1 至 3 中任一项所述的方法，其中，该方法还包括：

- 统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；

- 根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；

- 根据所述概率值，更新所述英文语音模型。

5. 根据权利要求 4 所述的方法，其中，该方法还包括：

- 当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；

r 根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

6. 根据权利要求 5 所述的方法，其中，所述步骤 r 包括：

- 根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

7. 根据权利要求 5 或 6 所述的方法，其中，该方法还包括：

- 根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

8. 根据权利要求 1 至 7 中任一项所述的方法，其中，该方法还包括：

x 获取所述用户为目标应用输入的单词语音序列；

y 根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；

z 将与所述英文字符序列相对应的信息提供给所述用户。

9. 根据权利要求 8 所述的方法，其中，所述目标应用为语音输入应用；

其中，所述步骤 z 包括：

- 将所述英文字符序列作为语音输入候选词条提供给所述用户。

10. 根据权利要求 8 所述的方法，其中，所述目标应用为语音搜索应用；

其中，该方法还包括：

- 将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个

Petitioner Exhibit 1004, Page 542 of 1052

或多个搜索结果；

其中，所述步骤 z 包括：

- 将所述一个或多个搜索结果提供给所述用户。

11. 根据权利要求 8 至 10 中任一项所述的方法，其中，所述步骤 x 包括：

- 获取所述用户为目标应用输入的字母语音序列；

其中，所述步骤 y 包括：

- 根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

12. 一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括：

语音获取装置，用于获取用户通过语音采集装置输入的语音数据；

英文判断装置，用于判断所述语音数据是否为英文语音数据；

语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

模型训练装置，用于根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

13. 根据权利要求 12 所述的模型训练设备，其中，所述语音识别装置用于：

- 当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

14. 根据权利要求 12 所述的模型训练设备，其中，所述语音识别装置用于：

- 当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

15. 根据权利要求 12 至 14 中任一项所述的模型训练设备，其中，该设备还包括模型更新装置，用于：

- 统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；

- 根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；

- 根据所述概率值，更新所述英文语音模型。

16. 根据权利要求 15 所述的模型训练设备，其中，该设备还包括：

数据提取装置，用于当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；

单元训练装置，用于根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

17. 根据权利要求 16 所述的模型训练设备，其中，所述单元训练装置用于：

- 根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

18. 根据权利要求 16 或 17 所述的模型训练设备，其中，该设备还包括单元更新装置，用于：

- 根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

Petitioner Exhibit 1004, Page 543 of 1052

19. 根据权利要求 12 至 18 中任一项所述的模型训练设备，其中，该设备还包括：

序列获取装置，用于获取所述用户为目标应用输入的单词语音序列；

序列确定装置，用于根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；

信息提供装置，用于将与所述英文字符序列相对应的信息提供给所述用户。

20. 根据权利要求 19 所述的模型训练设备，其中，所述目标应用为语音输入应用；

其中，所述信息提供装置用于：

- 将所述英文字符序列作为语音输入候选词条提供给所述用户。

21. 根据权利要求 19 所述的模型训练设备，其中，所述目标应用为语音搜索应用；

其中，该设备还包括：

搜索获取装置，用于将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；

其中，所述信息提供装置用于：

- 将所述一个或多个搜索结果提供给所述用户。

22. 根据权利要求 19 至 21 中任一项所述的模型训练设备，其中，所述序列获取装置用于：

- 获取所述用户为目标应用输入的字母语音序列；

其中，所述序列确定装置用于：

- 根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

Petitioner Exhibit 1004, Page 544 of 1052

Petitioner Exhibit 1004, Page 545 of 1052

一种基于中文语音信息训练英文语音模型的方法与设备

技术领域

[0001] 本发明涉及自然语言处理技术领域，尤其涉及一种用于基于中文语音信息训练英文语音模型的技术。

背景技术

[0002] 现有的训练英文语音模型的方式一般利用现有的英语语音数据库，包括语音数据及其标注，在发音单元集合中加入英文发音单元，利用英语数据进行训练。现有技术的问题在于，采用该种方式训练英文语音模型需要一定规模的英语语音数据，且难以对加入模型中的英语词的规模进行控制。

[0003] 因此，如何克服上述困难，在控制规模的同时，实现对英文语音模型的训练，成为本领域技术人员亟需解决的问题之一。

发明内容

[0004] 本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备。

[0005] 根据本发明的一个方面，提供了一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤：

[0006] a 获取用户通过语音采集装置输入的语音数据；

[0007] b 判断所述语音数据是否为英文语音数据；

[0008] c 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0009] d 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0010] 根据本发明的另一方面，还提供了一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括：

[0011] 语音获取装置，用于获取用户通过语音采集装置输入的语音数据；

[0012] 英文判断装置，用于判断所述语音数据是否为英文语音数据；

[0013] 语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0014] 模型训练装置，用于根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0015] 与现有技术相比，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模；进一步地，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的

使用体验。

附图说明

[0016] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本发明的其它特征、目的和优点将会变得更明显：

[0017] 图 1 示出根据本发明一个方面的用于基于中文语音信息训练英文语音模型的设备示意图；

[0018] 图 2 示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的设备示意图；

[0019] 图 3 示出根据本发明另一个方面的用于基于中文语音信息训练英文语音模型的方法流程图；

[0020] 图 4 示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的方法流程图。

[0021] 附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

[0022] 下面结合附图对本发明作进一步详细描述。

[0023] 图 1 示出根据本发明一个方面的用于基于中文语音信息训练英文语音模型的设备示意图。模型训练设备 1 包括语音获取装置 101、英文判断装置 102、语音识别装置 103 和模型训练装置 104。

[0024] 其中，语音获取装置 101 获取用户通过语音采集装置输入的语音数据。具体地，用户或大量用户通过与语音采集装置的交互，输入了语音数据，语音获取装置 101 通过调用该语音采集装置所提供的应用程序接口 (API)，获取该用户或大量用户输入的语音数据；或者，用户或大量用户通过与语音采集装置的交互，输入了语音数据，并将该语音数据存入语音数据库中，语音获取装置 101 通过调用该语音数据库所提供的应用程序接口 (API)，从该语音数据库中获取用户或大量用户输入的语音数据。

[0025] 本领域技术人员应能理解上述获取语音数据的方式仅为举例，其他现有的或今后可能出现的获取语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0026] 英文判断装置 102 判断所述语音数据是否为英文语音数据。具体地，英文判断装置 102 根据语音获取装置 101 所获取的用户通过语音采集装置输入的语音数据，例如通过声学模型匹配等方式，判断该语音数据是否为英文语音数据。

[0027] 本领域技术人员应能理解上述判断英文语音数据的方式仅为举例，其他现有的或今后可能出现的判断英文语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0028] 语音识别装置 103 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息。具体地，当英文判断装置 102 判断该语音数据为英文语音数据，语音识别装置 103 根据中文语音信息，如中文声韵母信息，对该英文语音数据进行语音识别处理，例如，当英文判断装置判断

Petitioner Exhibit 1004, Page 546 of 1052

该语音数据为英文语音数据时，语音识别装置 103 采用所有可能的中文声韵母信息去识别该英文语音数据，以确定该英文语音数据的发音信息。

[0029] 例如，用户通过与语音采集装置的交互，输入了语音数据单词“iphone”，语音获取装置 101 通过调用该语音采集装置所提供的应用程序接口 (API)，获取了该语音数据“iphone”；随后，英文判断装置 102 采用声学模型匹配的方式，判断该语音数据为英文语音数据；随后，语音识别装置 103 采用所有可能的中文声韵母信息，对该英文语音数据“iphone”进行强制对齐，确定该英文语音数据“iphone”的发音信息为“aa ai4 f engl”。在此，aa、ai4、f、engl 等即为中文声韵母信息，其中的数字表示该韵母信息所对应的声调。本领域技术人员应能理解，在此，中文声韵母信息及声调的表示仅为示例，而不应对本发明产生任何限制。

[0030] 优选地，语音识别装置 103 还可以根据单词模型所对应的中文语音信息、或字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息、或字母发音信息。例如，对于“iphone”，若用户输入语音采集装置的为按单词读的“iphone”，则语音识别装置 103 确定其对应的采用中文语音信息表示的发音信息为“aa ai4 f engl”；若用户输入语音采集装置的为按字母读的“iphone”，即，“i、p、h、o、n、e”，则语音识别装置 103 确定其对应的采用中文语音信息表示的发音信息为“aa ai4 p i4ei2 ch oo ou1 ee en1 ii i4”。

[0031] 本领域技术人员应能理解上述识别发音信息的方式仅为举例，其他现有的或今后可能出现的识别发音信息的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0032] 模型训练装置 104 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。具体地，模型训练装置 104 根据该语音识别装置 103 所识别获得的发音信息，确定该英文语音数据所对应的英文数据，如根据发音信息“aa ai4 f engl”，确定该发音信息对应的英文数据为“iphone”，再将该发音信息与该英文数据的映射关系存入英文语音模型，以实现对该英文语音模型的建立或更新，其中，该发音信息与该英文数据通过该中文语音信息相映射。

[0033] 在此，英文数据为一具体的英文单词，如“iphone”等，而英文语音数据为该具体的英文单词所对应的语音数据，如“iphone”所对应的语音数据，通过语音获取装置 101 所获取。

[0034] 本领域技术人员应能理解上述建立或更新英文语音模型的方式仅为举例，其他现有的或今后可能出现的建立或更新英文语音模型的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0035] 在此，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

[0036] 优选地，模型训练设备 1 的各个装置之间是持续不断工作的。具体地，语音获取装置 101 获取用户通过语音采集装置输入的语音数据；英文判断装置 102 判断所述语音数据是否为英文语音数据；语音识别装置 103 当所述语音数据为英文语音数据，根据中文语音

Petitioner Exhibit 1004, Page 547 of 1052

Petitioner Exhibit 1004, Page 548 of 1052

信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置 104 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。在此，本领域技术人员应理解“持续”是指模型训练设备 1 的各装置分别不断地进行语音数据的获取、英文语音数据的判断、英文语音数据的识别处理及英文语音模型的建立或更新，直至该模型训练设备 1 在较长时间内停止获取语音数据。

[0037] 优选地，语音识别装置 103 当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。具体地，当该英文判断装置 102 判断语音获取装置 101 所获取的语音数据为英文语音数据，语音识别装置 103 采用单词模型所对应的中文语音信息，如“aa ai4 f engl”等，对所述英文语音数据进行语音识别处理，如强制对齐处理，采用所有可能的中文声韵母信息，对该英文语音数据进行强制对齐，以获得与所述英文语音数据相对应的单词发音信息。

[0038] 在此，该单词模型中存储有大量单词所对应的中文语音信息，该单词模型可以位于模型训练设备 1 中，或者位于与该模型训练设备 1 通过网络相连接的第三方设备中。

[0039] 优选地，语音识别装置 103 当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。具体地，当该英文判断装置 102 判断语音获取装置 101 所获取的语音数据为英文语音数据，语音识别装置 103 根据字母模型所对应的中文语音信息，如二十六个英文字母所对应的中文语音信息，如字母“a”对应的中文语音信息“aa ei1”、字母“b”对应的中文语音信息“b i4”、字母“c”对应的中文语音信息“s ei4”等，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。

[0040] 优选地，语音识别装置 103 对于英文判断装置 102 判断为英文语音数据为语音数据，优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当该语音识别装置 103 采用单词模型无法识别出该英文语音数据的发音信息时，再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。相似地，该语音识别装置 103 也可以优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当该语音识别装置 103 采用字母模型无法识别出该英文语音数据的发音信息时，再采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息。在此，优先采用单词模型还是优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，可以是该模型训练设备 1 预先设定的，也可以应事件而进行调整。

[0041] 例如，对于语音获取装置 101 所获取的语音数据“iphone”，英文判断装置 102 判断该语音数据“iphone”为英文语音数据，语音识别装置 103 优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，但无法识别出该英文语音数据的发音信息；随后，该语音获取装置 101 再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，获得与该英文语音数据相对应的字母发音信息“aa ai4 p i4 ei2 ch oo ou1 ee en1 ii i4”，则可知该语音获取装置 101 所获取的语音数据为按字母读的“iphone”，即，“i、p、h、o、n、e”。

Petitioner Exhibit 1004, Page 549 of 1052

[0042] 更优选地，语音识别装置 103 分别采用单词模型所对应的中文语音信息、字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，并分别对语音识别处理所获得的结果进行打分，确定打分较高的发音信息为该英文语音数据的发音信息。

[0043] 在此，由于用户在输入英文单词时有时按单词读有时按字母读，本发明对每个需要识别的英文单词，利用相似的单词模型对应的中文语音信息、字母模型对应的中文语音信息，确定其对应的发音信息，提升了语音识别的准确性。

[0044] 在一优选实施例中，该模型训练设备 1 还包括模型更新装置（未示出）。该模型更新装置统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；根据所述概率值，更新所述英文语音模型。具体地，当大量用户通过语音采集装置输入语音数据，其中，多个用户所输入的英文语音数据为相同的英文语音数据，语音获取装置 101 获取该大量用户输入的大量语音数据；英文判断装置 102 判断该大量语音数据为英文语音数据；随后，语音识别装置 103 根据中文语音信息，对该大量英文语音数据进行语音识别处理，以获得与该大量英文语音数据相对应的发音信息，在此，由于该多个用户输入的为相同的英文语音数据，因此，该语音识别装置 103 所获得的关于该多个用户输入的英文语音数据所对应的发音信息相同或相似；接着，模型训练装置 104 根据该多个相同或相似的发音信息，确定该多个英文语音数据所对应的相同的英文数据；模型更新装置统计该相同的英文数据所对应的英文语音数据在该语音数据中的出现次数；接着，该模型更新装置再根据该英文语音数据的出现次数，并结合该语音数据的数量信息，确定该英文数据的概率值，如将该出现次数除以该大量语音数据的数量信息，获得该英文数据的概率值；该模型更新装置再将该英文数据及其对应的概率值存入该英文语音模型中，以实现对该英文语音模型的更新。

[0045] 优选地，该模型训练设备 1 还包括数据提取装置（未示出）和单元训练装置（未示出）。该数据提取装置当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；该单元训练装置根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。具体地，该模型训练设备 1 可以设置一个预定数量阈值，该预定数量阈值用来判断英文数据所对应的英文语音数据在该语音数据中的出现次数是否超过该值，当该出现次数大于该预定数量阈值时，数据提取装置自该英文语音模型中提取该英文数据与该英文语音数据；该单元训练装置根据该英文数据与该英文语音数据的映射关系，为该英文语音数据新建一个英文语音单元，以实现对该英文语音单元的建立或更新。

[0046] 更优选地，该单元训练装置根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。具体地，该单元训练装置根据该英文数据与该英文语音数据的映射关系，为该英文数据新建一英文语音单元，再例如通过将该中文发音信息所对应的中文语音单元的模型参数复制给该英文语音单元，然后在英文数据上对齐，得到统计量，再与该复制所获得的参数进行插值计算，以实现对该英文语音单元的建立或更新。在此，所述中文语音单元包括为每个中文发音信息，如中文声母或韵母信息所建立的发音单元，如“s”、“ei4”等中文声韵母信息所对应的中文语音单元。

[0047] 更优选地，该模型训练设备 1 还包括单元更新装置（未示出）。该单元更新装置根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。具

体地，在单元训练装置为英文数据新建一英文语音单元之后，单元更新装置不断地根据用户输入的英文语音数据，及由语音识别装置 103 识别所获得的与该英文语音数据相对应的发音信息，训练该英文语音单元。

[0048] 在此，本发明对出现此处大于预定阈值的英文语音数据，为该英文语音数据及英文数据新建英文语音单元，并对该英文语音单元不断地进行训练，以进一步提升该英文语音数据的发音信息的准确性。

[0049] 图 2 示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的设备示意图。该模型训练设备 1 还包括序列获取装置 205、序列确定装置 206 和信息提供装置 207。以下参照图 2 对该优选实施例进行详细描述：具体地，语音获取装置 201 获取用户通过语音采集装置输入的语音数据；英文判断装置 202 判断所述语音数据是否为英文语音数据；语音识别装置 203 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置 204 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射；序列获取装置 205 获取所述用户为目标应用输入的单词语音序列；序列确定装置 206 根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；信息提供装置 207 将与所述英文字符序列相对应的信息提供给所述用户。其中，语音获取装置 201、英文判断装置 202、语音识别装置 203 和模型训练装置 204 与图 1 所示对应装置相同或相似，故此处不再赘述，并通过引用的方式包含于此。

[0050] 其中，序列获取装置 205 获取所述用户为目标应用输入的单词语音序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入单词语音序列，序列获取装置 205 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取该用户为目标应用所输入的单词语音序列。

[0051] 序列确定装置 206 根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列。具体地，序列确定装置 206 根据该序列获取装置 205 所获取的单词语音序列，通过语音识别处理，确定该单词语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了单词语音序列“iphone”，序列获取装置 205 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置 206 根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”。

[0052] 信息提供装置 207 将与所述英文字符序列相对应的信息提供给所述用户。例如，当该目标应用为语音输入应用，该信息提供装置 207 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将该英文字符序列作为语音输入候选词条提供给该用户；当该目标应用为语音搜索应用；该信息提供装置 207 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将根据该英文字符序列搜索获得的一个或多个搜索结果提供给该用户。

[0053] 优选地，该目标应用为语音输入应用；其中，该信息提供装置 207 将所述英文字符序列作为语音输入候选词条提供给所述用户。例如，当该目标应用为语音输入应用，用户通

Petitioner Exhibit 1004, Page 550 of 1052

过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，序列获取装置 205 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置 206 根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f eng1”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”；进而，信息提供装置 207 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入。

[0054] 优选地，该目标应用为语音搜索应用；其中，该模型训练设备 1 还包括搜索获取装置（未示出）。该搜索获取装置将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，该信息提供装置 207 将所述一个或多个搜索结果提供给所述用户。具体地，当该目标应用为语音搜索应用，序列确定装置 206 根据该英文语音模型，确定与该单词语音序列相对应的英文字符序列；搜索获取装置根据该序列确定装置 206 所确定的英文字符序列，将该英文字符序列作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置 207 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将该一个或多个搜索结果提供给该用户。

[0055] 例如，当该目标应用为语音搜索应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，序列获取装置 205 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置 206 根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f eng1”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”；进而，搜索获取装置根据该序列确定装置 206 所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置 207 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0056] 更优选地，序列获取装置 205 获取所述用户为目标应用输入的字母语音序列；其中，所述序列确定装置 206 根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入字母语音序列，序列获取装置 205 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取该用户为目标应用所输入的字母语音序列；随后，序列确定装置 206 根据该序列获取装置 205 所获取的字母语音序列，通过语音识别处理，确定该字母语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了字母语音序列“i、p、h、o、n、e”，序列获取装置 205 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取了该用户输入的字母语音序列“i、p、h、o、n、e”；随后，序列确定装置 206 根据该字母语音序列“i、p、h、o、n、e”，通过语音识别处理，确定该字母语音序列的发音信息“aa ai4 p i4 ei2 ch oo ou1 ee en1 ii i4”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列“iphone”。

Petitioner Exhibit 1004, Page 551 of 1052

[0057] 相似地，信息提供装置 207 将与该英文字符序列相对应的信息提供给所述用户。接上例，当该目标应用为语音输入应用，该信息提供装置 207 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入；当该目标应用为语音搜索应用；搜索获取装置根据该序列确定装置 206 所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置 207 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0058] 在此，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的使用体验。

[0059] 图 3 示出根据本发明另一个方面的用于基于中文语音信息训练英文语音模型的方法流程图。

[0060] 在步骤 S301 中，模型训练设备 1 获取用户通过语音采集装置输入的语音数据。具体地，用户或大量用户通过与语音采集装置的交互，输入了语音数据，在步骤 S301 中，模型训练设备 1 通过调用该语音采集装置所提供的应用程序接口 (API)，获取该用户或大量用户输入的语音数据；或者，用户或大量用户通过与语音采集装置的交互，输入了语音数据，并将该语音数据存入语音数据库中，在步骤 S301 中，模型训练设备 1 通过调用该语音数据库所提供的应用程序接口 (API)，从该语音数据库中获取用户或大量用户输入的语音数据。

[0061] 本领域技术人员应能理解上述获取语音数据的方式仅为举例，其他现有的或今后可能出现的获取语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0062] 在步骤 S302 中，模型训练设备 1 判断所述语音数据是否为英文语音数据。具体地，在步骤 S302 中，模型训练设备 1 根据在步骤 S301 中所获取的用户通过语音采集装置输入的语音数据，例如通过声学模型匹配等方式，判断该语音数据是否为英文语音数据。

[0063] 本领域技术人员应能理解上述判断英文语音数据的方式仅为举例，其他现有的或今后可能出现的判断英文语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0064] 在步骤 S303 中，模型训练设备 1 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息。具体地，当在步骤 S302 中，模型训练设备 1 判断该语音数据为英文语音数据，在步骤 S303 中，模型训练设备 1 根据中文语音信息，如中文声韵母信息，对该英文语音数据进行语音识别处理，例如，当在步骤 S302 中，模型训练设备 1 判断该语音数据为英文语音数据时，在步骤 S303 中，模型训练设备 1 采用所有可能的中文声韵母信息去识别该英文语音数据，以确定该英文语音数据的发音信息。

[0065] 例如，用户通过与语音采集装置的交互，输入了语音数据单词“iphone”，在步骤 S301 中，模型训练设备 1 通过调用该语音采集装置所提供的应用程序接口 (API)，获取了该语音数据“iphone”；随后，在步骤 S302 中，模型训练设备 1 采用声学模型匹配的方式，判断该语音数据为英文语音数据；随后，在步骤 S303 中，模型训练设备 1 采用所有可能的中文声韵母信息，对该英文语音数据“iphone”进行强制对齐，确定该英文语音数据“iphone”的发

Petitioner Exhibit 1004, Page 552 of 1052

音信息为“aa ai4f eng1”。在此，aa、ai4、f、eng1 等即为中文声韵母信息，其中的数字表示该韵母信息所对应的声调。本领域技术人员应能理解，在此，中文声韵母信息及声调的表示仅为示例，而不应对本发明产生任何限制。

[0066] 优选地，在步骤 S303 中，模型训练设备 1 还可以根据单词模型所对应的中文语音信息、或字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息、或字母发音信息。例如，对于“iphone”，若用户输入语音采集装置的为按单词读的“iphone”，则在步骤 S303 中，模型训练设备 1 确定其对应的采用中文语音信息表示的发音信息为“aa ai4 feng1”；若用户输入语音采集装置的为按字母读的“iphone”，即，“i、p、h、o、n、e”，则在步骤 S303 中，模型训练设备 1 确定其对应的采用中文语音信息表示的发音信息为“aa ai4 pi4 ei2 ch oo ou1 ee en1 iii4”。

[0067] 本领域技术人员应能理解上述识别发音信息的方式仅为举例，其他现有的或今后可能出现的识别发音信息的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0068] 在步骤 S304 中，模型训练设备 1 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。具体地，在步骤 S304 中，模型训练设备 1 根据在步骤 S303 中所识别获得的发音信息，确定该英文语音数据所对应的英文数据，如根据发音信息“aa ai4 f eng1”，确定该发音信息对应的英文数据为“iphone”，再将该发音信息与该英文数据的映射关系存入英文语音模型，以实现对该英文语音模型的建立或更新，其中，该发音信息与该英文数据通过该中文语音信息相映射。

[0069] 在此，英文数据为一具体的英文单词，如“iphone”等，而英文语音数据为该具体的英文单词所对应的语音数据，如“iphone”所对应的语音数据，通过语音获取装置 101 所获取。

[0070] 本领域技术人员应能理解上述建立或更新英文语音模型的方式仅为举例，其他现有的或今后可能出现的建立或更新英文语音模型的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0071] 在此，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

[0072] 优选地，模型训练设备 1 的各个步骤之间是持续不断工作的。具体地，在步骤 S301 中，模型训练设备 1 获取用户通过语音采集装置输入的语音数据；在步骤 S302 中，模型训练设备 1 判断所述语音数据是否为英文语音数据；在步骤 S303 中，模型训练设备 1 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；在步骤 S304 中，模型训练设备 1 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。在此，本领域技术人员应理解“持续”是指模型训练设备 1 的各步骤分别不断地进行语音数据的获取、英文语音数据的判断、英文语音数据的识别处理及英文语音模型的建立或更新，直至该模型训练设备 1 在较长时间内停止获取语音数据。

Petitioner Exhibit 1004, Page 553 of 1052

[0073] 优选地，在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。具体地，当在步骤S302中，模型训练设备1判断在步骤S301中所获取的语音数据为英文语音数据，在步骤S303中，模型训练设备1采用单词模型所对应的中文语音信息，如“aa ai4 f engl”等，对所述英文语音数据进行语音识别处理，如强制对齐处理，采用所有可能的中文声韵母信息，对该英文语音数据进行强制对齐，以获得与所述英文语音数据相对应的单词发音信息。

[0074] 在此，该单词模型中存储有大量单词所对应的中文语音信息，该单词模型可以位于模型训练设备1中，或者位于与该模型训练设备1通过网络相连接的第三方设备中。

[0075] 优选地，在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。具体地，当在步骤S302中，模型训练设备1判断在步骤S301中所获取的语音数据为英文语音数据，在步骤S303中，模型训练设备1根据字母模型所对应的中文语音信息，如二十六个英文字母所对应的中文语音信息，如字母“a”对应的中文语音信息“aa ei1”、字母“b”对应的中文语音信息“b i4”、字母“c”对应的中文语音信息“s ei4”等，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。

[0076] 优选地，在步骤S303中，模型训练设备1对于在步骤S302中判断为英文语音数据为语音数据，优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当在步骤S303中，模型训练设备1采用单词模型无法识别出该英文语音数据的发音信息时，再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。相似地，在步骤S303中，模型训练设备1也可以优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当在步骤S303中，模型训练设备1采用字母模型无法识别出该英文语音数据的发音信息时，再采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息。在此，优先采用单词模型还是优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，可以是该模型训练设备1预先设定的，也可以应事件而进行调整。

[0077] 例如，对于在步骤S301中，模型训练设备1所获取的语音数据“iphone”，在步骤S302中，模型训练设备1判断该语音数据“iphone”为英文语音数据，在步骤S303中，模型训练设备1优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，但无法识别出该英文语音数据的发音信息；随后，在步骤S301中，模型训练设备1再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，获得与该英文语音数据相对应的字母发音信息“aa ai4 p i4 ei2 ch oo ou1 ee en1 ii i4”，则可知在步骤S301中所获取的语音数据为按字母读的“iphone”，即，“i、p、h、o、n、e”。

[0078] 更优选地，在步骤S303中，模型训练设备1分别采用单词模型所对应的中文语音信息、字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，并分别对语音识别处理所获得的结果进行打分，确定打分较高的发音信息为该英文语音数据的发音信息。

Petitioner Exhibit 1004, Page 554 of 1052

[0079] 在此，由于用户在输入英文单词时有时按单词读有时按字母读，本发明对每个需要识别的英文单词，利用相似的单词模型对应的中文语音信息、字母模型对应的中文语音信息，确定其对应的发音信息，提升了语音识别的准确性。

[0080] 在一优选实施例中，该方法还包括步骤 S308（未示出）。在步骤 S308 中，模型训练设备 1 统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；根据所述概率值，更新所述英文语音模型。具体地，当大量用户通过语音采集装置输入语音数据，其中，多个用户所输入的英文语音数据为相同的英文语音数据，在步骤 S301 中，模型训练设备 1 获取该大量用户输入的大量语音数据；在步骤 S302 中，模型训练设备 1 判断该大量语音数据为英文语音数据；随后，在步骤 S303 中，模型训练设备 1 根据中文语音信息，对该大量英文语音数据进行语音识别处理，以获得与该大量英文语音数据相对应的发音信息，在此，由于该多个用户输入的为相同的英文语音数据，因此，在步骤 S303 中，模型训练设备 1 所获得的关于该多个用户输入的英文语音数据所对应的发音信息相同或相似；接着，在步骤 S304 中，模型训练设备 1 根据该多个相同或相似的发音信息，确定该多个英文语音数据所对应的相同的英文数据；在步骤 S308 中，模型训练设备 1 统计该相同的英文数据所对应的英文语音数据在该语音数据中的出现次数；接着，在步骤S308中，模型训练设备1再根据该英文语音数据的出现次数，并结合该语音数据的数量信息，确定该英文数据的概率值，如将该出现次数除以该大量语音数据的数量信息，获得该英文数据的概率值；该在步骤 S308 中，模型训练设备 1 再将该英文数据及其对应的概率值存入该英文语音模型中，以实现对该英文语音模型的更新。

[0081] 优选地，该方法还包括步骤 S309（未示出）和步骤 S310（未示出）。在步骤 S309 中，当所述出现次数大于预定数量阈值，模型训练设备 1 自所述英文语音模型中提取所述英文数据与所述英文语音数据；在步骤 S310 中，模型训练设备 1 根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。具体地，该模型训练设备 1 可以设置一个预定数量阈值，该预定数量阈值用来判断英文数据所对应的英文语音数据在该语音数据中的出现次数是否超过该值，当该出现次数大于该预定数量阈值时，在步骤 S309 中，模型训练设备 1 自该英文语音模型中提取该英文数据与该英文语音数据；在步骤 S310 中，模型训练设备 1 根据该英文数据与该英文语音数据的映射关系，为该英文语音数据新建一个英文语音单元，以实现对该英文语音单元的建立或更新。

[0082] 更优选地，在步骤 S310 中，模型训练设备 1 根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。具体地，在步骤 S310 中，模型训练设备 1 根据该英文数据与该英文语音数据的映射关系，为该英文数据新建一英文语音单元，再例如通过将该中文发音信息所对应的中文语音单元的模型参数复制给该英文语音单元，然后在英文数据上对齐，得到统计量，再与该复制所获得的参数进行插值计算，以实现对该英文语音单元的建立或更新。在此，所述中文语音单元包括为每个中文发音信息，如中文声母或韵母信息所建立的发音单元，如“s”、“ei4”等中文声韵母信息所对应的中文语音单元。

[0083] 更优选地，该方法还包括步骤 S311（未示出）。在步骤 S311 中，模型训练设备 1 根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。具

Petitioner Exhibit 1004, Page 555 of 1052

Petitioner Exhibit 1004, Page 556 of 1052

体地，在步骤 S310 中，模型训练设备 1 为英文数据新建一英文语音单元之后，在步骤 S311 中，模型训练设备 1 不断地根据用户输入的英文语音数据，及由在步骤 S303 中识别所获得的与该英文语音数据相对应的发音信息，训练该英文语音单元。

[0084] 在此，本发明对出现此处大于预定阈值的英文语音数据，为该英文语音数据及英文数据新建英文语音单元，并对该英文语音单元不断地进行训练，以进一步提升该英文语音数据的发音信息的准确性。

[0085] 图 4 示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的方法流程图。以下参照图 4 对该优选实施例进行详细描述：具体地，在步骤 S401 中，模型训练设备 1 获取用户通过语音采集装置输入的语音数据；在步骤 S402 中，模型训练设备 1 判断所述语音数据是否为英文语音数据；在步骤 S403 中，模型训练设备 1 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；在步骤 S404 中，模型训练设备 1 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射；在步骤 S405 中，模型训练设备 1 获取所述用户为目标应用输入的单词语音序列；在步骤 S406 中，模型训练设备 1 根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；在步骤 S407 中，模型训练设备 1 将与所述英文字符序列相对应的信息提供给所述用户。其中，步骤 S401-S404 与图 3 所示对应步骤相同或相似，故此处不再赘述，并通过引用的方式包含于此。

[0086] 在步骤 S405 中，模型训练设备 1 获取所述用户为目标应用输入的单词语音序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入单词语音序列，在步骤 S405 中，模型训练设备 1 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取该用户为目标应用所输入的单词语音序列。

[0087] 在步骤 S406 中，模型训练设备 1 根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列。具体地，在步骤 S406 中，模型训练设备 1 根据在步骤 S405 中所获取的单词语音序列，通过语音识别处理，确定该单词语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了单词语音序列“iphone”，在步骤 S405 中，模型训练设备 1 例如通过调用该用户设备所提供的应用程序接口 (API) 等方式，获取了该用户输入的单词语音序列“iphone”；随后，在步骤 S406 中，模型训练设备 1 根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f eng1”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”。

[0088] 在步骤 S407 中，模型训练设备 1 将与所述英文字符序列相对应的信息提供给所述用户。例如，当该目标应用为语音输入应用，在步骤 S407 中，模型训练设备 1 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将该英文字符序列作为语音输入候选词条提供给该用户；当该目标应用为语音搜索应用；在步骤 S407 中，模型训练设备 1 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术，将根据该英文字符序列搜索获得的一个或多个搜索结果提供给该用户。

[0089] 优选地，该目标应用为语音输入应用；其中，在步骤 S407 中，模型训练设备 1 将所

述英文字符序列作为语音输入候选词条提供给所述用户。例如，当该目标应用为语音输入应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，在步骤S406中，模型训练设备1根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f eng1”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”；进而，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入。

[0090] 优选地，该目标应用为语音搜索应用；其中，该方法还包括步骤S412(未示出)。在步骤S412中，模型训练设备1将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，在步骤S407中，模型训练设备1将所述一个或多个搜索结果提供给所述用户。具体地，当该目标应用为语音搜索应用，在步骤S406中，模型训练设备1根据该英文语音模型，确定与该单词语音序列相对应的英文字符序列；在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该一个或多个搜索结果提供给该用户。

[0091] 例如，当该目标应用为语音搜索应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，在步骤S406中，模型训练设备1根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f eng1”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”；进而，在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0092] 更优选地，在步骤S405中，模型训练设备1获取所述用户为目标应用输入的字母语音序列；其中，在步骤S406中，模型训练设备1根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入字母语音序列，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的字母语音序列；随后，在步骤S406中，模型训练设备1根据在步骤S405中所获取的字母语音序列，通过语音识别处理，确定该字母语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了字母语音序列“i、p、h、o、n、e”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户

Petitioner Exhibit 1004, Page 557 of 1052

输入的字母语音序列“i、p、h、o、n、e”;随后,在步骤 S406 中,模型训练设备 1 根据该字母语音序列“i、p、h、o、n、e”,通过语音识别处理,确定该字母语音序列的发音信息“aa ai4 p i4 ei2 ch oo ou1 ee en1 ii i4”,进而,根据该发音信息,在该英文语音模型中进行匹配查询,确定与该字母语音序列相对应的英文字符序列“iphone”。

[0093] 相似地,在步骤 S407 中,模型训练设备 1 将与该英文字符序列相对应的信息提供给所述用户。接上例,当该目标应用为语音输入应用,在步骤 S407 中,模型训练设备 1 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术,将该英文字符序列“iphone”作为语音输入候选词条提供给该用户,以供该用户选择输入 ;当该目标应用为语音搜索应用 ;在步骤 S412 中,模型训练设备 1 根据在步骤 S406 中所确定的英文字符序列,将该英文字符序列“iphone”作为该用户的查询序列,进行匹配查询,获取与该查询序列相对应的一个或多个搜索结果 ;随后,在步骤 S407 中,模型训练设备 1 通过调用诸如 JSP、ASP 或 PHP 等动态网页技术,将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户,供该用户浏览。

[0094] 在此,本发明还可与语音识别相结合,支持用户以按字母及按单词两种方式输入英文语音序列,提升了用户的使用体验

[0095] 本发明的软件程序可以通过处理器执行以实现上文所述步骤或功能。同样地,本发明的软件程序(包括相关的数据结构)可以被存储到计算机可读记录介质中,例如,RAM 存储器,磁或光驱动器或软磁盘及类似设备。另外,本发明的一些步骤或功能可采用硬件来实现,例如,作为与处理器配合从而执行各个功能或步骤的电路。

[0096] 另外,本发明的一部分可被应用为计算机程序产品,例如计算机程序指令,当其被计算机执行时,通过该计算机的操作,可以调用或提供根据本发明的方法和 / 或技术方案。而调用本发明的方法的程序指令,可能被存储在固定的或可移动的记录介质中,和 / 或通过广播或其他信号承载媒体中的数据流而被传输,和 / 或被存储在根据所述程序指令运行的计算机设备的工作存储器中。在此,根据本发明的一个实施例包括一个装置,该装置包括用于存储计算机程序指令的存储器和用于执行程序指令的处理器,其中,当该计算机程序指令被该处理器执行时,触发该装置运行基于前述根据本发明的多个实施例的方法和 / 或技术方案。

[0097] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。装置权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一,第二等词语用来表示名称,而并不表示任何特定的顺序。

Petitioner Exhibit 1004, Page 558 of 1052

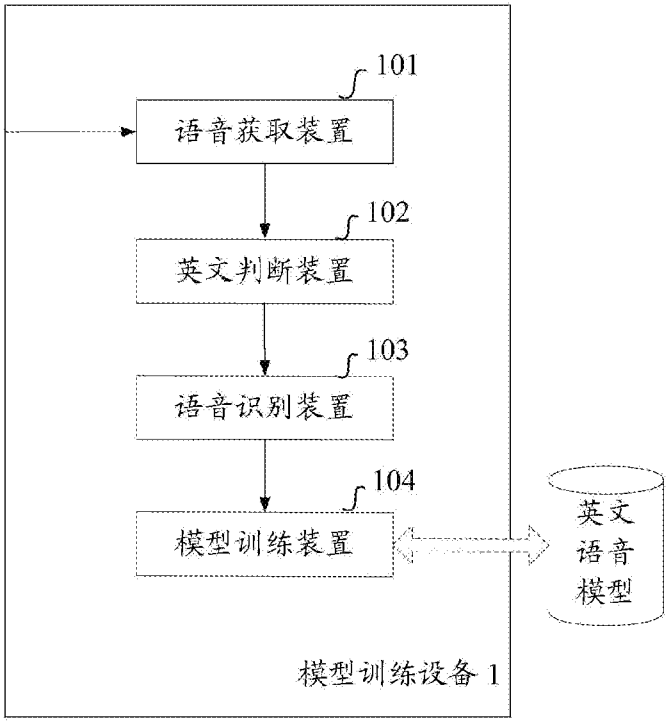


图 1

Petitioner Exhibit 1004, Page 559 of 1052

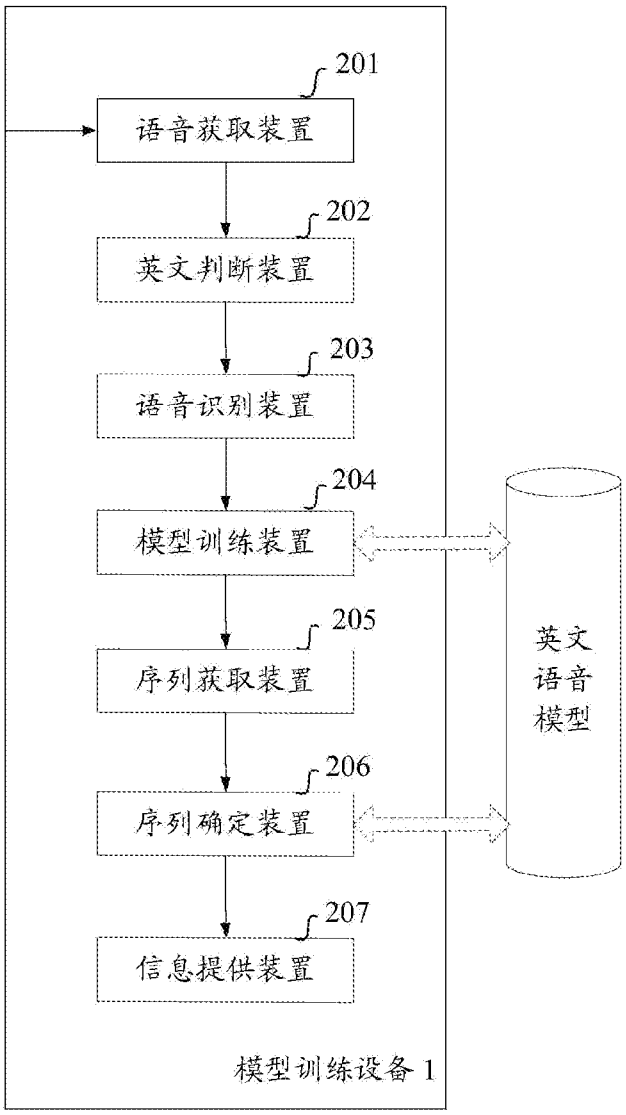


图 2

Petitioner Exhibit 1004, Page 560 of 1052

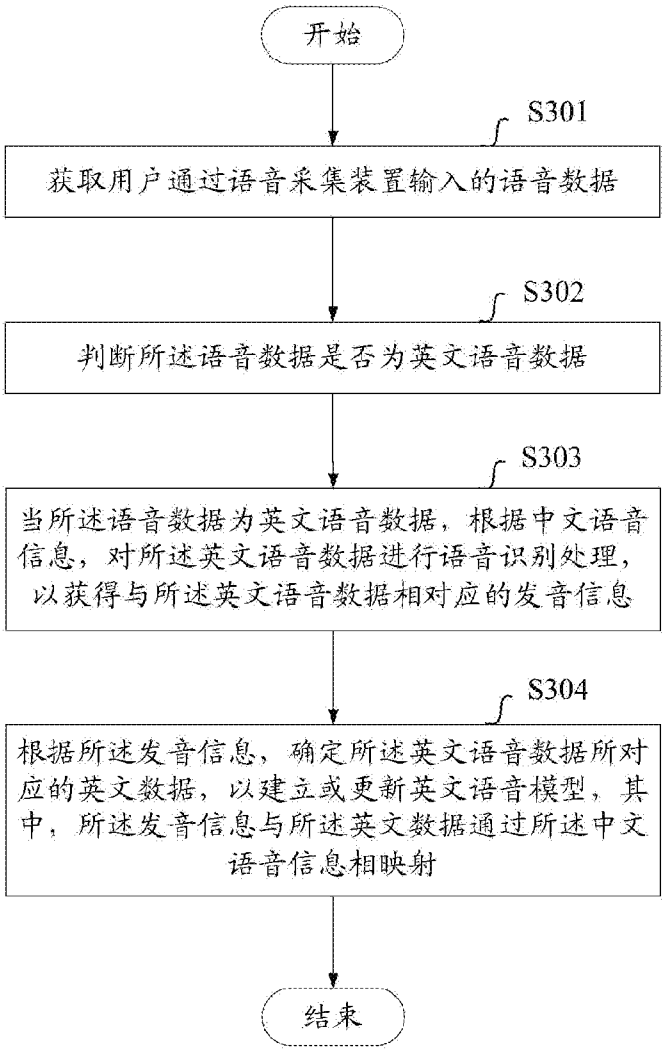


图 3

Petitioner Exhibit 1004, Page 561 of 1052

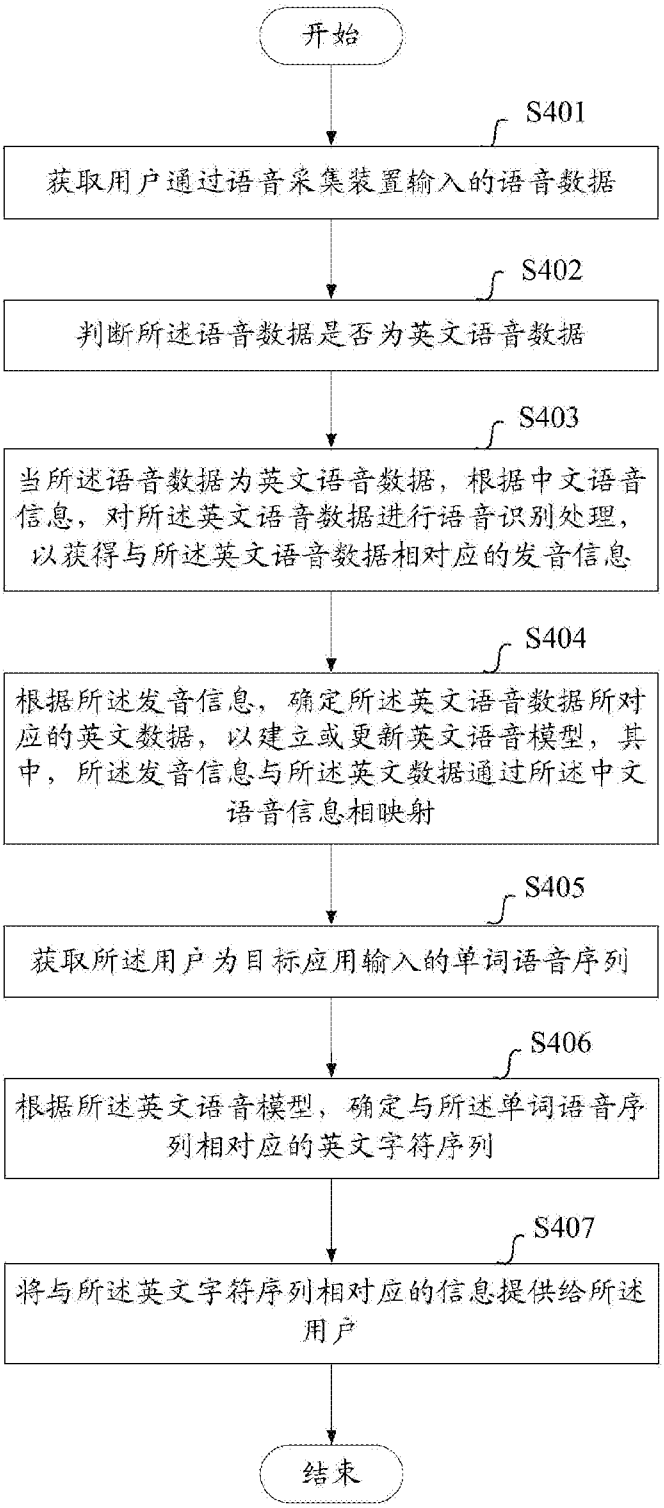


图 4

Petitioner Exhibit 1004, Page 562 of 1052

CORRECTED VERSION

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference **IEE160077PCT**	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. **PCT/CN2016/105343**	International filing date *(day/month/year)* **10 November 2016**	(Earliest) Priority Date *(day/month/year)*
Applicant **NUANCE SOFTWARE TECHNOLOGY (BEIJING) CO., LTD.**		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of **5** sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of:

   ☑ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

   b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

   c. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☐ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

   ☑ the text is approved as submitted by the applicant.

   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   ☐ the text is approved as submitted by the applicant.

   ☑ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

   a. the figure of the **drawings** to be published with the abstract is Figure No. **2**

   ☑ as suggested by the applicant.

   ☐ as selected by this Authority, because the applicant failed to suggest a figure.

   ☐ as selected by this Authority, because this figure better characterizes the invention.

   b. ☐ none of the figures is to be published with the abstract.

Box No. IV Text of the abstract (Continuation of item 5 of the first sheet)

A user device configured to perform wake-up word detection in a target language. The user device comprises at least one microphone(430) configured to obtain acoustic information from the environment of the user device, at least one computer readable medium(435) storing an acoustic model(150) trained on a corpus of training data(105) in a source language different than the target language, and storing a first sequence of speech units obtained by providing acoustic features(110) derived from audio comprising the user speaking a wake-up word in the target language to the acoustic model(150), and at least one processor(415,425) coupled to the at least one computer readable medium(435) and programmed to perform receiving, from the at least one microphone(430), acoustic input from the user speaking in the target language while the user device is operating in a low-power mode, applying acoustic features derived from the acoustic input to the acoustic model(150) to obtain a second sequence of speech units corresponding to the acoustic input, determining if the user spoke the wake-up word at least in part by comparing the first sequence of speech units to the second sequence of speech units, and exiting the low-power mode if it is determined that the user spoke the wake-up word.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/105343

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G06F 1/32(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15/-; G06F 1/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: sound+,voice?,speech+,audio?,acoustic+,model?,phoneme?,wak+,up,idle+,sleep+,stand+,by,low,power,activat+,trigger+,source?,target?,language?,native,nonnative

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2016189706 A1 (BROADCOM CORP.) 30 June 2016 (2016-06-30) description paragraphs [0030]-[0032], [0049]-[0070], [0086]-[0092], figures 2, 4, 5, 8	1-20
Y	CN 101650943 A (CHINESE ACAD INST. ACOUSTICS, ET AL.) 17 February 2010 (2010-02-17) description page 5 line 3 -page 6 line 9 , figures 1-3	1-20
A	CN 103021409 A (ANHUI USTC IFLYTEK CO., LTD.) 03 April 2013 (2013-04-03) the whole document	1-20
A	CN 105575395 A (ZTE CORP.) 11 May 2016 (2016-05-11) the whole document	1-20
A	US 2014365225 A1 (DSP GROUP) 11 December 2014 (2014-12-11) the whole document	1-20
A	CN 106098059 A (SHANGHAI JIAO TONG UNIVERSITY) 09 November 2016 (2016-11-09) the whole document	1-20

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 July 2017	**21 September 2017**
Name and mailing address of the ISA/CN **STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer **KANG,Dandan** Telephone No. **(86-10)61648170**

Form PCT/ISA/210 (second sheet) (July 2009)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103632668 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 12 March 2014 (2014-03-12) the whole document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/105343

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2016189706	A1	30 June 2016	None			
CN	101650943	A	17 February 2010	None			
CN	103021409	A	03 April 2013	CN	103021409	B	24 February 2016
CN	105575395	A	11 May 2016	None			
US	2014365225	A1	11 December 2014	None			
CN	106098059	A	09 November 2016	None			
CN	103632668	A	12 March 2014	None			

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
100037
China 8th Floor, Vantone New World Plaza, 2
Fuchengmenwai Street, Xicheng District, Beijing

CCPIT PATENT AND TRADEMARK LAW OFFICE

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	**04 August 2017**
Applicant's or agent's file reference **IEE160077PCT**	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/CN2016/105343	**10 November 2016**	

International Patent Classification (IPC) or both national classification and IPC
G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G06F 1/32(2006.01)i

Applicant
NUANCE SOFTWARE TECHNOLOGY (BEIJING) CO., LTD.

1. This opinion contains indications relating to the following items:

- [x] Box No. I Basis of the opinion
- [ ] Box No. II Priority
- [ ] Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- [ ] Box No. IV Lack of unity of invention
- [x] Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- [ ] Box No. VI Certain documents cited
- [ ] Box No. VII Certain defects in the international application
- [ ] Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/	Date of completion of this opinion	Authorized officer
STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA **China 6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088**	**31 July 2017**	**KANG,Dandan**
Facsimile No. **(86—10) 62019451**		Telephone No. **(86-10)61648170**

Form PCT/ISA/237 (Cover sheet) (July 2011)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [✓] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. [ ] This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing filed or furnished:
   a. (means)
      - [ ] on paper
      - [ ] in electronic form
   b. (time)
      - [ ] in the international application as filed
      - [ ] together with the international application in electronic form
      - [ ] subsequently to this Authority for the purposes of search

4. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-20	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-20	NO
Industrial applicability (IA)	Claims	1-20	YES
	Claims	None	NO

2. Citations and explanations :

[1] D1: US 2016189706 A1

[2] D2: CN 101650943 A

[3] D1 discloses a wake up phrase detection method and an electronic device(see description paragraphs [0030]-[0032], [0049]-[0070],[0086]-[0092], figures 2-5,8): electronic device may comprise a smart phone; FIG.4 depicts a training system 400, a user is prompted to utter their custom wake-up phrase(WUP), the utterances are captured as user training sequences 402(corresponding to "receiving acoustic input of a user speaking a wake-up word"); user training sequences 402 is provided to VAD 404, then to feature extraction component 406 to derive the feature vectors 408(corresponding to "acoustic features derived from the acoustic input"); the extracted features 408 are decoded by concatenated phoneme ASR engine 410 using phoneme model DB 414 which is implemented in a memory or storage medium and store phoneme models 206, and produce a phoneme transcription of the WUP, a Hidden Markov Model(HMM) is trained offline to generate a phoneme model 206(corresponding to "acoustic model"); based on the results produced by ASR 410, phoneme model concatenation component 416 determines the best phonemes models and concatenates them to produce an initial WUP model consist of the concatenated phoneme models(corresponding to "a sequence of speech units corresponding to the wake-up word spoken by the user" and "a first sequence of speech units" in claim 9 or "a second sequence of speech units" in claim 16); the WUP model used for WUP detector is stored in WUP model DB 418 which is implemented in a memory or storage medium ; FIG.5 depicts a user-specific wake up phrase detection system 500, a low-power processor is included and operates in a stand-by mode of the electronic device, receiving a wake-up phrase 514 as an input by a user(corresponding to "receiving acoustic input from the user speaking while the user device is operating in a low-power mode"), WUP 514 is provided to VAD 504 and then to feature extractor and WUP detector(WUPD) component 506 to detect the presence of the WUP in the captured audio by comparing the feature extracted to the WUP model stored in a non-volatile RAM; the presence of the WUP triggers the device to wake up, or come out of a stand-by mode or a sleep mode into a normal mode of operation(corresponding to "exiting the low-power mode if it is determined that the user spoke the wake-up word").

[4] D2 discloses a non-native speech recognition system and method (see description page 5 line 3 – page 6 line 9 , figures 1-3): a native language(corresponding to "the source language") model is provided for the native state decoding module, the non-native language(corresponding to "the target language") speech is decoded by the native state decoding module using the native language model, thus obtaining a native state decode sequence "b aa zh " (see figure 3).

[5] **Novelty and Inventive step**

[6] Claim 1 differs from D1 in: (1) the acoustic input of a user is spoken in a target language, and the acoustic model is trained in a source language different than the target language.

[7] Claim 5 differs from D1 in the above (1), and (2) at least one microphone configured to obtain acoustic information from the environment of the user device.

[8] Claim 9 differs from D1 in the above (1), (2) and (3) applying acoustic features derived from the acoustic input to the acoustic model to obtain a second sequence of speech units; determining if the user spoke the wake-up word at least in part by comparing the first sequence of speech units to the second sequence of speech units.

Petitioner Exhibit 1004, Page 570 of 1052

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

[9] Claim 16 differs from D1 in the above (1) and (3).

[10] Therefore, claims 1-20 are novel (PCT Article 33(2)).

[11] The feature (1) has already been disclosed by D2. The person skilled in the art would easily acquire the technical feature of D2 to decode the acoustic input.

[12] The feature (2) of obtaining acoustic information by microphone is general solution in the art.

[13] As to the feature (3), it is common practice for the person skilled in the art to determine the wake-up word by comparing the first sequence of speech units to the second sequence of speech units which is obtained by applying acoustic features derived from the acoustic input to the acoustic model instead of the WUP model.

[14] Therefore, claims 1,5,9,16 do not involve an inventive step (PCT Article 33 (3)).

[15] Dependent claims 2-4,6-8,10-15,17-20 do not involve an inventive step (PCT Article 33 (3)).

[16] As to claims 2,6,10,17,4,8,13,18, it is common practice to adapt the acoustic model using the acoustic input and the sequence, and associate at least one task with the wake-up word and initiate performance of the task based on the wake-up detection.

[17] As to claims 3,7,11,12,14,15,19,20, those additional features have been disclosed by D1.

[18] **Industrial applicability**

[19] Claims 1-20 can find industrial applicability in the technical field of wake-up detecting (PCT Article 33 (4)).

Petitioner Exhibit 1004, Page 571 of 1052

Publication number: CN101650943 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20100217
Application number: CN200810239892
Application date: 20081219
Priority: CN200810239892 20081219 CN20081239892;
Assignee: OF ACOUSTICS CHINESE ACADEMY OF SCIENCES INST ;
Assignee[std]: CHINESE ACAD INST ACOUSTICS ;
Inventor: YAN YONGHONG ; ZHANG QINGQING ; PAN JIELIN ;
Inventor[std]: QINGQING ZHANG ; JIELIN PAN ; YONGHONG YAN ;
International class[8]: G10L15/065; G10L17/06;
Family members: CN101650943 A

Title: NON-NATIVE SPEECH RECOGNITION SYSTEM AND METHOD THEREOF

Title: 一种非母语语音识别系统及方法

Abstract:

The invention relates to a non-native speech recognition system based on mixed model state correction and a method thereof. The non-native speech recognition system comprises a non-native speech interface, a native model module, a non-native model module, a native state decoding module, a non-native state forced alignment module, a native-non-native state similarity matrix computation module, a native-non-native state mapping table computation module and a non-native state correction model decoding module. In the system and the method thereof, a non-native acoustic model is corrected at statelevels based on a native acoustic model of a speaker and state mapping among different models, thus obtaining a model that better meets non-native pronunciation characteristics. The system and the method thereof have the advantages of obvious improvement of recognition performance compared with that of a recognition system which is not corrected by the method on the premise of only the native training data without increase of any non-native speech training data, no obvious speed fall of the recognition speech of the system and very high practicability.

Abstract:

本发明涉及一种基于混合模型状态修正的非母语语音识别系统及方法。该系统包括：非母语语音接口、母语模型模块、非母语模型模块、母语状态解码模块、非母语状态强制对齐模块、母语与非母语状态相似度矩阵计算模块、母语与非母语状态映射表计算模块及非母语状态修正模型解码模块，该系统及方法通过不同模型间的状态映射，利用说话人

The present invention relates to a non-native speech recognition system and method based on hybrid model state amended. The system comprises: a non-native speech interfaces, up model module, non-native model module, decoding module native state, non-native status forcing alignment module, native and non-native

母语的声学模型在状态级别上对非母语声学模型进行修正，从而得到更为符合非母语发音特点的模型。该系统及方法有如下优点：在不增加任何非母语语音训练数据，仅依靠说话人母语训练数据的前提下，相对未采用该方法修正过的识别系统的识别性能有明显提高；同时系统识别语音的速度并未明显降低，具有很高的实用性。

state similarity matrix calculation module, native and non-native status mapping table calculation module and non-native state correction model decoding module, the system and method of mapping between the state of different models, the use of mother tongue speaker acoustic models on a state level for non-native acoustic model is modified to give more in line with the characteristics of non-native pronunciation model. The system and method has the following advantages: without adding any non-native speech training data, relying only native speaker training data are available, the relative performance of the method is not used to identify corrected identification system has significantly improved; the same system Speech Recognition It did not significantly reduce the speed, with high availability.

Claims:

1、一种非母语语音识别系统，其特征在于，所述系统包括: 一非母语语音接口，用于采集非母语语音数据，并将该非母语语音数据送入非 母语状态强制对齐模块和母语状态解码模块; 一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提 供母语声学模型; 一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解 码模块提供非母语声学模型; 一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码， 得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分 段信息送入母语与非母语状态相似度矩阵计算模块; 一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数据进行 强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非 母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块; 一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和 非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重 合时间大于预先设定的门限值时，认为这两个状态出现一次"同现"，统计出所有的 "同现"情况，并计算得到非母语语音状态对应于母语语音状态的相似度矩阵，并 将该相似度矩阵信息送入母语与非母语状态映射表计算模块; 一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射 表; 及 一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射 表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母 语声学模型;最终利用该修正后的非母语声学模型完成非母语语音识别。

1, a non-native speech recognition system, wherein the system comprises: a non-native speech interface for data collection of non-native speech, and the non-native voice data into non-Native Status force alignment module and decoding module up state; a native model module to provide up to up status acoustic model decoding module and non-native state correction model decoding module; a non-native model module, used to force alignment to non-native state modules and non-native state correction model decoding modules provide non-native acoustic models; a native state decoding module for up acoustic model standard non-native voice data is decoded to give up the speech state-level segment information, namely the status of native speech decoding information and state-level native speech segment information is fed up with the non-native state similarity matrix calculation module; a non-native status forcing alignment module for non-native acoustic models will be compulsory for non-native speech data aligned to give non- native speech state-level segment information, namely non-native speech state reference information and non-native speech state level segment information is fed up with the non-native state similarity matrix calculation module; a native and non-native state similarity matrix calculation module for native speech state-level segment information and non-native speech state-level segment information on the time alignment, when the time is up speech state coincides with the non-native speech state is greater than a

2、根据权利要求1所述的非母语语音识别系统，其特征在于，所述非母语语音 状态对应于母语语音状态的相似度矩阵通过以下公式得到: lJ N ^count(t_ n=^ -其中，M，N分别为中文和英文的状态数量;A_(M,N)为相似度矩阵;A"为矩 阵的第严行尸列元素»tj为非母语语音状态，^i为母语语音状态;A;;。A，，^(M,N)J ^=1 M, ^=l N^ C_ (^ l_)是母语语音状态_和非母语语音状态^j之 间的"同现"出现次数。

书 求 要 3 禾 权 3、根据权利要求1所述的非母语语音识别系统，其特征在于，所述状态映射表 通过以下方式得到: 如果A。矩阵中n_行尸列元素是尸列的第n大元素，则状态叫是状态尸的第 n候选修正状态;每个非母语语言的状态都能在矩阵中找到母语语言的n个候选修正 状态，其中，n<Mc 4、根据权利要求1所述的非母语语音识别系统，其特征在于，所述非母语状态 修正模型解码模块在语音识别解码过程中，对于观察值0。，用n个候选状态修正以 后的非母语状态尸的观察概率巧"_·)变为: p_o_ = op^"^o^)to_o)Z^ory_1^(_)，。" 其中，0表示非母语状态的修正权重，勺是相应第叫个候选状态，n是候选数; 巧"^P·)"和p，"(ot)，""分别表示非母语状态尸和母语状态竹在观察值0·下的原始观 察概率。

5、一种非母语语音识别方法，包括如下步骤: (1)非母语语音接口采集一定数量的非母语语音数据，用于得到模型状态映射 表; (2)非母语状态强制对齐模块用非母语声学模型对非母语语音数据进行强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息; (3)母语状态解码模块用标准母语声学模型对非母语语音数据进行解码，得到 母语语音状态级分段信息，即母语语音状态解码信息; (4)母语与非母语状态相似度矩阵计算模块将得到的非母语语音状态级分段信 息和母语语音状态级分段信息在时间上对齐，当某两个状态的重合时间大于预先设 定的门限值时，这两个状态出现一次"同现"; (5)母语与非母语状态相似度矩阵计算模块统计出所有的"同现"情况并计算 得到非母语语音状态对应于母语语音状态的相似度矩阵; (6)母语与非母语状态映射表计算模块根据该相似度矩阵得到状态映射表; (7)非母语状态修正模型解码模块根据得到的状态映射表，在语音识别的解码 过程中，用映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得 到修正后的非母语声学模型。

6、根据权利要求5所述的非母语语音识别方法，其特征在于，所述非母语语音 状态对应于母语语音状态的相似度矩阵通过以下公式得到: 书 求 要 3 禾 权 A¨=_unt(t_SJ ·JN 散。u"·(、。_) n=^ 一 其中，M，N分别为中文和英文的状态数量;A^，^(M,N)为相似度矩阵;A"为矩 阵的第产行尸列元素;tj为非母语

predetermined threshold value, that the two states once "with the current", the statistics of all "with the current" situation, and calculate non-native speech state corresponds to the state of native speech similarity matrix, and the similarity matrix information into the native and non-native state map calculation module; a native and non-native status mapping table calculation module for obtaining the status of a mapping table according to the similarity matrix calculation; and a non-native state correction model decoding module for decoding the speech recognition process, with the state up acoustic model state map found amend non-native acoustic models corresponding state, to be non-native acoustic models after correction; the final use of non-native acoustic models after the completion of the correction of non-native speech recognition.

2, according to non-native speech recognition system according to claim 1, wherein said non-native speech state corresponds to the state of native speech similarity matrix obtained by the following formula: lJ N count (t_ n = - where, M, N, respectively the number of state Chinese and English; A_ (M, N) for the similarity matrix; A "is the first serious Xingshi matrix column elements tj non-native voice state, i-speaking voice status;. A;; A,, (M, N) J = 1 M, = l N C_ (l_) is native and non-native speech state _ voice status j between "co-occurrence" occurrences.

3 Wo books seeking the right to play according to claim non-native speech recognition system of claim 1, characterized in that said state mapping table in the following manner: If A. N_ Xingshi matrix column elements are the major elements of the corpse n columns, then the state is called the state of the corpse of the first amendment n candidate status; the status of each non-native language can be found n candidate correction native language in the matrix state, wherein, n and lt; Mc 4, according to non-native speech recognition system according to claim 1, characterized in that the non-native state correction model decoding module decoding the speech recognition process, the observed value of zero. With n candidate status to amend non-native state after the corpse of observation probabilities clever "_) becomes: p_o_ = op " o) to_o) Z ory_1 (_),. "Where 0 represents the right to amend non-native state of heavy, spoon the corresponding section called candidate status, n is the number of candidates; clever" P) "and p,"

语音状态，_为母语语音状态;A。A，，t(^,N)， r=l M, ^=1 N^ C^un^ (_>)是母语语音状态_和非母语语音状态_之 ^间的"同现"出现次数。

7、根据权利要求5所述的非母语语音识别方法，其特征在于，所述状态映射表 通过以下万式得到: 如果A"矩阵中叫行川列元素是尸列的第n大元素，则状态n"是状态尸的第 n候选修正状态;每个非母语语言的状态都能在矩阵中找到母语语言的n个候选修正 状态，其中，n<Mc 8、根据权利要求5所述的非母语语音识别万法，其特征在于，所述非母语状态 修正模型解码模块在语音识别解码过程中，对于观察值0。，用n个候选状态修正以 后的非母语状态尸的观察概率巧"^0。)变为: p_) = ^p ^" (o^)^a_O_o)Z^o4y_)，"" 其中，0表示非母语状态的修正权重，%是相应第叫个候选状态，n是候选数; 巴"(^·)"和p，"(^。)"分别表示非母语状态尸和母语状态叫在观察值0·下的原始观 察概率。

(ot), "" denote non-native status corpse and Bamboo native state under observation 0.5 original observation probabilities.

5, a non-native speech recognition method comprising the steps of: (1) non-native speech interface collecting a certain number of non-native voice data, for obtaining a model state map; (2) non-native status force alignment block with non-native acoustic model for non-native speech data alignment forced to give non-native speech state-level segment information, namely the status of non-native speech reference information; (3) the native state standard decoding module up Acoustics model of non-native speech data is decoded to give up the speech state-level segment information, namely the state of native speech decoding information; and (4) native and non-native state similarity matrix calculation module and the resulting non-native speech state-level segment information and native speech state-level segment information in time alignment, a time when the two states coincide greater than a preset threshold, these two states appear a "co-occurrence"; (5) Native and non-native status similar matrix calculation module statistics of all the "co-occurrence" the situation and calculate the non-native speech state corresponds to the state of native speech similarity matrix; (6) native and non-native state map state maps obtained computing module based on the similarity matrix Table; (7) non-native state correction model decoding module based on the obtained state map, in the decoding process of speech recognition, the acoustic model with native state mapping table to find the corresponding non-modified up acoustic model state, to be corrected non-native acoustic models.

6, according to non-native speech recognition method as claimed in claim 5, characterized in that the non-native speech state corresponds to the state of native speech similarity matrix obtained by the following equation: Book 3 Wo seek playing right A&#xfffd; = _unt (t_SJ &#xfffd; JN San.u "&#xfffd; (,._) n = one where, M, N, respectively Chinese and English, the number of state; A, (M, N) the similarity matrix; A "is the first production Xingshi column matrix element; tj state non-native speech, native speech _ to state; A.A,, t (, N), r = l M, = 1 "The same is now" is the native speech state and non-native speech state _ _ between the number of occurrences of

N C un (_ and; gt).

7, according to non-native speech recognition method as claimed in claim 5, characterized in that said state map obtained by the following square formula: if A "column of the matrix element is called row Nakagawa corpse column The n-th largest element, the state n "is the state of the corpse of the first amendment n candidate status; the status of each non-native language can be found n the status of candidate native language of the amendment in the matrix, where, n and lt; Mc 8 The non-native speech recognition methodology set forth in claim 5, wherein said non-native state correction model decoding module decodes the speech recognition process, the observed value of 0., With non-native state n candidate status after correcting corpse observation probabilities clever " 0) becomes:. P_) = p " (o) a_O_o) Z o4y_), "" where, 0 Fixed right non-native state of weight, percent is the corresponding section called candidate status, n is the number of candidates; Pakistan "(*)" and p, "(.)" denote non-native status corpse and native state is called the observation The original observation probabilities under 0.5.

Description:

_1_ 5 tFL明书 一种非母语语音识别系统及方法 技术领域 本发明涉及语音识别系统及方法，特别涉及基于混合模型状态修正的非母语语 音识别系统及方法。

背景技术 随着现代社会信息的全球化，非母语语音识别成为语音识别技术领域研究热点 之一。对非母语的语音识别相对母语的语音识别而言，识别性能将会大幅度降低， 特别是对于有着严重口音的说话人。如何在保证语音识别系统在仅有少量非母语训练数据的前提下，对带有不同程度口音的非母语语音识别性能的提高是非母语语音 识别研究的重点。

文献(Bohn.O_S., F^ege,J·E., nThe production Of new and Similar Vowels by adult German learners Of English."」[MM] Stud. Second Lang. Acquis. 14, 131-)指出， 非母语说话人在进行目标语言发音时，可能会用说话人自身的母语发音进行替代，或者产生出有说话人母语发音特点和目标语言发音特点相结合的发音。这个结论启 发我们，说话人自身的母语语言训练数据，可能对说话人非母语语音识别有帮助， 特别是针对那些发音口音较重的语音数据。

_1_ 5 tFL Ming book a non-native speech recognition system and method TECHNICAL FIELD The present invention relates to a speech recognition system and method, and more particularly to non-native speech recognition system and method based on hybrid model state amended.

BACKGROUND With globalization, non-native speech recognition has become one of modern society information technology research focus in the field of speech recognition. Non-native speech recognition relative native speech recognition, the recognition performance will be significantly reduced, especially for people with serious accent speak. How to ensure that the voice recognition system in only a small number of non-native language training under the premise of the data, speech recognition performance for non-native accents with varying degrees of improvement are non-native speech recognition research.

Document (Bohn.O_S., F ege, J ◆ E., NThe production Of new and Similar Vowels by adult German learners Of English. "" [MM] Stud. Second

当前，说话人自适应技术(如MAP^ ML^R)已广泛的应用于处理非母语语音 识别中(Z. Wang,T. Schultz.A. Waibel, wComparison Of acoustic model adaptation techniques On non-native Speech" 「C^, Proc^ ICASSP 2003.)。这类方法主要是通过对少 量非母语语音数据进行自适应，使得母语发音声学模型能够在一定程度上逼近非母 语的发音特点。在这些万法中，自适应数据和测试数据的相似性是决定识别性能好 坏的关键因素。尽管自适应技术在提高非母语语音识别性能上有贡献，但通过自适 应以后的模型在非母语语音识别中的性能相比母语语音模型在母语语音识别的性 能，仍然较低。文献(J. Humlphries, P. Woodland, and D. Pearce. "Using accent-specific pronunciation modeling for robust Speech recognition." [C] In Proc. ICSLP '96, pages， 2324-2327, Philadelphia, PA, October 1996·)研究了自适应算法的局限性，指出导致非 母语语音识别率低的主要原因来自于说话人母语语系中无法涵盖的非母语语言发 音。如何便非母语语音声学模型较好的模拟出这类语言发音是非母语语音识别研究 工作的重点。

书 明 说 发明内容 为了克服上述现有技术的不足，本发明提供一种基于状态修正的非母语语音识 别系统及万法，该系统及万法通过不同模型间的状态映射，利用说话人母语的声学 模型在状态级别上对非母语声学模型进行修正，从而得到更为符合非母语发音特点 的模型。

为了实现上述目的，本发明提出的一种非母语语音识别系统，包括: 一非母语语音接口，用于采集非母语语音数据，并将该非母语语音数据送入非 母语状态强制对齐模块和母语状态解码模块。

一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提 供母语声学模型。

一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解 码模块提供非母语声学模型。

一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码， 得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分 段信息送入母语与非母语状态相似度矩阵计算模块。

一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数据进行 强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非 母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

Lang. Acquis. 14, 131-) He notes that non-native speaker pronunciation language during target may be replaced with the speaker's own Pronounced, or have speakers produce Pronounced features and characteristics of the combination of the target language pronunciation pronunciation. This conclusion Kai made us, the speaker's own native language training data may be non-native speaker voice recognition helpful, especially for those heavier voice data pronunciation accent.

Currently, the speaker adaptation techniques (such as MAP ML R) has been widely used in the treatment of non-native speech recognition (Z. Wang, T. Schultz.A. Waibel, wComparison Of acoustic model adaptation techniques On non - native Speech "" C, Proc ICASSP 2003.). Such methods mainly by a small number of non-native speech data adaptive, so Pronounced acoustic model to a certain extent on the approximation of non-native pronunciation features. In The methodology, the similarity is to determine the key factors to identify good or bad performance adaptive data and test data. Although adaptive technology to improve performance on non-native speech recognition contribute, but after an adaptive model in a non-native speech native speech recognition performance model in native speech recognition performance, is still low compared to the literature (J. Humlphries, P. Woodland, and D. Pearce. "Using accent-specific pronunciation modeling for robust Speech recognition." [C] In Proc. ICSLP '96, pages, 2324-2327, Philadelphia, PA, October 1996) adaptive algorithm study limitations, noting that the main cause of the low rate of non-native speech recognition from native language speakers can not Pronounced non-covered language. How would a non-native speech acoustic model can simulate the focus of such a non-native language pronunciation of speech recognition research work.

Book confessed SUMMARY To overcome the shortcomings of the prior art, the present invention provides a state-based correction of non-native speech recognition system and methodology, the system and methodology by state mapping between different models, the use of the speaker up acoustic models in the state level of non-native acoustic model is modified, resulting model is more in line with the characteristics of non-native pronunciation.

一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和 非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重 合时间大于预先设定的门限值时，认为这两个状态出现一次"同现"，统计出所有的 "同现"情况，并计算得到非母语语音状态对应于母语语音状态的相似度矩阵，并 将该相似度矩阵信息送入母语与非母语状态映射表计算模块。

一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射 表。及 一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射 表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母 语声学模型;最终利用该修正后的非母语声学模型完成非母语语音识别。

其中，所述非母语语音状态对应于母语语音状态的相似度矩阵通过以下公式得 到: 书 明 说 散_t(t_Si) 其中，M,N分别为中文和英文的状态数量;A，，T(lv4,?N_为相似度矩阵;A"为矩 阵的第严行尸列元素;tj为非母语语音状态，_为母语语音状态;A"。A，，t(M,_9 i=\ M , _ A_ C^un^ (_是母语语音状态s飞口非母语语音状态^j之 ^间的"同现"出现次数。

其中，所述状态映射表通过以下方式得到: 如果A"矩阵中叫行芦列元素是尸列的第n大元素，则状态叫是状态尸的第 n候选修正状态;例如:如果A。矩阵中产行卢列元素是矩阵中尸列的最大元素， 这意味着来自相应语言的状态严和状态卢最为相似，状态严是状态芦的第一候选 修正状态。如果伊行尸列元素是芦列的第二大元素，那么状态伊是状态户的第二 候选修正状态。如此类推，每个非母语语言的状态邵能在矩阵中找到母语语言的n 个候选修正状态(n<M)o 其中，所述非母语状态修正模型解码模块在语音识别解码过程中，对于观察值0。，用n个候选状态修正以后的非母语状态尸的观察概率p，"_。)变为: p.14(0^) = opJn4^0^)6a_)_o^y_)，。u 其中，0表示非母语状态的修正权重，%是相应第川个候选状态，n是候选数; p，"^0。)"和p，"^0。)，。u分别表示非母语状态尸和母语状态叫在观察值0·下的原始观 察概率。

本发明提供的一种非母语语音识别方法，包括如下步骤: (1)非母语语音接口采集一定数量的非母语语音数据，用于得到模型状态映射 表。

(2)非母语状态强制对齐模块用非母语声学模型对非母语语音数掘进行强制对 齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息。

To achieve the above object, the present invention provides a non-native speech recognition system, comprising: a non-native speech interface for data collection of non-native speech, and the non-native voice data into the non-native status force alignment module and decoding module native state.

A native model module for native and non-native status decoding module decoding state correction model module provides up acoustic model for.

A non-native language model module for non-native status force alignment module and non-native state correction model decoding modules provide non-native acoustic models. a native state decoding module for up according to the standard acoustic model for non-native voice data is decoded to give up the speech state-level segment information, namely the state of native speech decoding information, and native speech state-level segment information sent the native and non-native state similarity matrix calculation module.

A non-native status forcing alignment module for non-native acoustic models to non-native speech data aligned native speech forced to give non-native speech state-level segment information that is non-native speech state reference information and non- state-level segment information is fed up with the non-native state similarity matrix calculation module.

A native and non-native state similarity matrix calculation module for native speech state-level segment information and non-native speech state-level segment information on the time alignment, coincidence when the native and non-native speech speech state status When the time is greater than a predetermined threshold, I consider these two states appear a "co-occurrence", the statistics of all "with the current" situation, and calculate non-native speech state corresponds to the state of native speech similarity matrix, and the similarity matrix information into the native and non-native status mapping table calculation module.

A native and non-native status mapping table calculation module for obtaining the status of a mapping table according to the similarity matrix calculation. And a non-native state correction model decoding module for decoding the speech recognition process, the acoustic model with

(3)母语状态解码模块用标准母语声学模型对非母语语音数据进行解码，得到 母语语音状态级分段信息，即母语语音状态解码信息。

(4)母语与非母语状态相似度矩阵计算模块将得到的非母语语音状态级分段信 息和母语语音状态级分段信息在时间上对齐，当某两个状态的重合时间大于预先设 定的门限值时，这两个状态出现一次"同现"。

(5)母语与非母语状态相似度矩阵计算模块统计出所有的"同现"情况并计算 书 明 说 得到非母语语音状态对应于母语语音状态的相似度矩阵。

(6)母语与非母语状态映射表计算模块根据该相似度矩阵得到状态映射表。

(7)非母语状态修正模型解码模块根据得到的状态映射表，在语音识别的解码 过程中，用映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得 到修正后的非母语声学模型。

本发明的优点在于: (1)本发明的非母语语音识别系统及方法使非母语声学模型更适应带有母语口 音的非母语语音特点。

(2)本发明的非母语语音识别系统及方法，采用不同模型间的状态映射，利用 说话人母语的声学模型在状态级别上对非母语声学模型进行修正，解码器根据得到 的状态映射表实现非母语语音识别。混合模型状态修正的非母语语音识别系统对非 母语语音的识别率与未采用该方法修正过的识别系统相比较;并且在不增加任何非 母语语音训练数据，仅依靠标准母语训练数据的前提下，对带有母语口音的非母语语音的识别情况相对未采用该方法修正过的识别系统有了明显提高;同时系统识别 语音的速度并未明显下降。

(3)本发明的非母语语音识别系统及方法实现了利用母语声学模型对非母语声 学模型的修正，提高系统对非母语语音的识别性能。相比语音自适应技术，基于混 合模型状态的修正方法不需要增加更多的非母语训练数据。由于用于修正的母语声 学模型的规模可以控制在很小的范围，修正后的模型在计算速度上也不会有明显增 加。我们进行了大量实网数据的测试，结果显示在识别性能方面，由于该系统在状 态级别上的映射信息是基于解码器输出的状态序列情况进行统计得到的，这种准则 较为直接，基本能真实的反映出两种语言状态间的相似程度。与未使用该方法的系 统相比，基于混合模型状态修正的非母语语音识别系统对非母语语音的识别错误率 相对下降可达到5-10_ 附图说明 图1本发明的基于混

mother tongue Status mapping table to find the corresponding non-modified up acoustic model state, to be non-native acoustic models corrected; final non-native acoustic models using the revised completion non-native speech recognition. wherein the non-native speech state corresponds to the state of native speech similarity matrix obtained by the following equation: Book confessed San _t (t_Si) where, M, N, respectively, for the Chinese and English, the number of state; A,, T (lv4, N_ is similarity matrix;? A "is the first serious Xingshi column matrix element; tj state non-native speech, native speech _ to state; A".A,, t (M, _9 i = \ M, _ A_ C un (_ is opening up speech state s fly non-native speech state "co-occurrence" j between the number of occurrences. wherein the state mapping table in the following manner: If A "matrix called reed row n column elements are the major elements of the corpse column, the state is called the state of the corpse of the first amendment n candidate status; for example: if middle row matrix A. Lu column element is the largest element in the matrix column corpse, which means the state from the corresponding language Lu Yan and the state most similar state is the state of Lu Yan corrected first candidate status. If the Iraqi Xingshi column element Lo is the second largest element column, then the state of Iraq is the second candidate status households revision status and so forth, the status of each non-native language can find n Shao revision status of candidate native language (n and amp in a matrix;. lt; M) o wherein the non-native state correction model decoding module in speech recognition decoding process, for observations 0.5, with n candidate status of non-native state after correction corpse observation probability p, "_.) Becomes: p.14 (0) = opJn4 0) 6a _) _ o y_),. u where 0 represents the right to amend non-native state of weight, percent of the corresponding first Chuan candidate status, n is the number of candidates;. p, ". 0)" and p, " 0),. u denote non-native state Pronounced dead and called on the state to observe a value of 0 [illegible] Under the original observation probabilities.

The present invention provides a non-native speech recognition method comprising the steps of: (1) non-native speech interface collecting a certain number of non-native voice data, for obtaining a model state map.

(2) non-native status force alignment block with non-native language to non-native speech

合模型状态修正算法的非母语语音识别系统及方法的基本 原理框图; 图2本发明的非母语语音识别系统的流程框图; 图3本发明的具体实施例母语语音状态与非母语语音状态之间的"同现"说明 图。

书 明 说 具体实施方式 下面结合附图对本发明的具体实施方式做进一步详细描述: 图1是基于混合模型状态修正算法的非母语语音识别系统的基本原理框图，它 描述了混合模型状态修正算法的核心组成成分，主要由几个部分构成:非母语语音、 母语/非母语声学模型、强制对齐、解码、状态映射表生成以及非母语状态修正模型 的解码。图2是基于混合模型状态修正算法的非母语语音识别系统的具体实施流程 框图。

本发明涉及的混合模型状态修正算法的非母语语音识别的核心技术在于状态映 射表的获取(如图2所示的模块1到7)。混合模型状态修正算法是一种新型的基于 相似度矩阵的状态映射算法，通过对同现次数的统计，得到两种语言状态间的对应 关系，利用该关系确定用于修正的状态对。

本发明的系统包括:一非母语语音接口，用于采集非母语语音数据，并将该非 母语语音数据送入非母语状态强制对齐模块和母语状态解码模块。

一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提 供母语声学模型。

一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解 码模块提供非母语声学模型。

一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码， 得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分 段信息送入母语与非母语状态相似度矩阵计算模块。

一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数掘进行 强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非 母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和 非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重 合时间大于预先设定的门限值时，认为这两个状态出现一次"同现"，统计出所有的 "同现"情况，并计算得到非母语语音状态对应于母语语音状

acoustic model number forced to dig together to give non-native speech state-level segment information that is non-native speech state reference information.

(3) up state decoding module up with a standard acoustic model for non-native speech data is decoded to give up the speech state-level segment information, namely the status of native speech decoding information.

(4) Native and non-native state similarity matrix calculation module and the resulting non-native speech state-level segment information and native speech state-level segment information on the time alignment, when a coincidence of time greater than a predetermined two states When the set threshold, these two states appear once "with now."

(5) Native and non-native state similarity matrix calculation module statistics of all the "co-occurrence" Ming said the situation and calculations to obtain a non-native speech state corresponds to the state of the similarity matrix native speech.

(6) native and non-native state map state map obtained calculation module based on the similarity matrix.

(7) non-native state correction model decoding module based on the obtained state map, in the decoding process of speech recognition, the acoustic model with native state mapping table to find the corresponding non-modified up acoustic model state, to be amended The non-native acoustic models.

Advantages

the present invention is: (1) non-native speech recognition system and method of the present invention enables non-native acoustic model is more adapted to the characteristics of the non-native speech with a native accent.

(2) non-native speech recognition system and method of the present invention, the use of state mapping between different models, the use of mother tongue speaker acoustic models on a state level obtained according to the state of non-native acoustic model is modified, the decoder mapping table non-native speech recognition. Hybrid model state revised non-native speech recognition

态的相似度矩阵，并 将该相似度矩阵信息送入母语与非母语状态映射衷计算模块。

一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射 表。及 一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射 表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母 书明 说 语声学模型;最终利用该修正后的非母语声学模型完成非母语语音识别。

本发明的具体计算流程如下: (如图2) 第一步，非母语状态参考答案:选择一定量的非母语语音数据(非母语语音接 口1)，这些数据用于生成母语对非母语的相似度矩阵。用非母语声学模型(非母语 模型模块2)对这些数据进行强制对齐(^orced-alignmlen^)，得到非母语状态序列， 同时记录下每个状态的时间信息(非母语状态强制对齐模块4). 第二步，母语状态识别结果:用母语声学模型(母语模型模块3)对第一步中的 非母语语音数据进行解码(Decode)，得到母语状态序列，同时记录下每个状态的时 间信息(母语状态解码模块5). 第三步，"同现"准则:通过第一步和第二步，在同一批非母语语音数据上，分 别得到了非母语和母语的状态序列以及相应的时间段信息，根据这两个状态序列在 时间轴上的位置关系，可以定义两个状态的"同现"现象。在计算"同现"前，首 先定义一个"同现"矩阵，该矩阵为(母语语言状态个数*非母语语言状态个数)的 矩阵，每个元素位置上记录了对应行和列的状态之间的"同现"次数。在本发明所 述方法的实验中，定义当两个状态在时间段上重叠的时间占到非母语语言状态持续 时间一半时，算作一次"同现"。当非母语语言状态产和母语语言状态尸出现一次 "同现"时，在对应同现矩阵的严行尸列的位置上加1记录。如图3所示，母语语 音状态"aa native"和非母语语音状态"ae nonnat^ve"之间出现的一次"同现"。

一一 第四步，状态相似度矩阵计算(母语与非母语状态相似度矩阵计算模块6):若 设定M，N分别为母语语言和非母语语言状态个数，A，，t(^,_为从同现矩阵派生出来 的相似度矩阵，矩阵元素记录了母语语言和非母语语言状态之间的相似度情况。A" 为矩阵严行尸列的元素，设定^j为非母语语言状态，_为母语语言状态，两者之间 的相似度计算为: _C^unt^t_SJ AiJ N 散counntOOJ^SJ ""(1) 其中，_T(M,N)^ _ M^ ^=1 N^ 第五步，得到相似度矩阵后，根据该矩阵得到状态映射表(母语与非母语状态映射表计算模块7)。如果A。矩阵中产行尸列元素是矩阵中尸列的最大元素，这 意味着来自相应语言的状态严和状态尸最为相似，状态产是状态尸的第一候选修正 状态。如果伊行尸列元素是尸列的第二大元素，那么状态叫是状态尸的第二候选 修正状

system for non-native speech recognition rate and end the use of the revised methodology of comparing identification system; and without adding any non-native speech training data, the training data only rely on standard up premise under non-native speech recognition with a native accent situation is relatively late adoption of the revised methodology recognition system has been significantly improved; and the system did not recognize the voice of the speed decreased.

(3) non-native speech recognition system and the methodology of the present invention achieves up acoustic model using non-native amendment acoustic models for improved identification performance of non-native speech. Compared voice adaptive technology, do not need to add more non-native hybrid model training data correction method based on the state. Because of the size used up acoustic model correction can be controlled in a very small range, model corrected in the calculation would not be significant speed increase. We conducted extensive testing real network data, the results are displayed in the identification of performance, because the system mapping information on the state level is based on the state of the decoding output sequence statistics obtained, such guidelines are more direct, basically true It reflects the degree of similarity between the two languages status. Compared with the system at the end of the use of the methodology based on the modified hybrid model state non-native speech recognition system error rate for non-native speech relative decline may reach 5-10_ drawings illustrating an invention based on hybrid model state correction The basic block diagram of the non-native speech recognition system and method of the algorithm; flow block diagram of a non-native speech recognition system 2 of the present invention; "with the current status of a particular embodiment up voice and non-native speech state Figure 3 between invention "Fig.

Book confessed specific embodiments with reference to the following specific embodiments of the present invention will be described in further detail: Figure 1 is a hybrid model based on the state of the basic block diagram of the non-native speech recognition system correction algorithm, which describes a hybrid model state correction core algorithm components, mainly composed of several components: a non-native speech, native / non-native acoustic models, forcing alignment,

态。如此类推，每个非母语语言的状态都能在矩阵中找到母语语言的n个候 书 明 说 选修正状态(n<M)Q 到此，我们基于混合模型状态修正算法，得到了母语语言和非母语语言状态^间 的映射信息。

第六步，接下来根据得到的状态间的映射关系，在解码过程中用母语声学模型 对非母语声学模型进行修正(非母语状态修正模型解码模块8)。根据第五步得到的 状态映射表，在语音识别的解码过程中，对于观察值0·，用n个候选状态修正以后 的非母语状态尸的观察概率p，"(P。)变为: p^W_o_ = op^4_o^)_)^o^p，"_)，M ""(2) 这里^表示非母语状态的修正权重，勺是相应第沪个候选状态，n是候选数。

p，"(P。)"和印^0·)，"u分别表示非母语状态尸和母语状态川在观察值0。下的原始观 察概率。

本发明的特点: (1)使非母语声学模型更适应带有母语口音的非母语语音特点 (2)提出一种新型基于混合模型状态修正的非母语语音识别方法，(3)在仅依靠标 准母语声学模型的前提下，提高系统对非母语语音的识别性能。

在进行基于给定语法的带有中文口音的英文实网数据测试中显示，基于混合模 型状态修正算法的非母语语音识别系统在带有中文口音的英文测试集上的错误率相 对末使用该万法的非母语语音识别系统下降5_10% (相对值)^计算速度相对于末 使用该万法的非母语语音识别系统降低20%-25% (相对值)。

decoding, decoding status map generation and non-native state correction model. Figure 2 is a block diagram of a specific implementation process of the mixed model state correction algorithm non-native speech recognition system.

The present invention relates to a hybrid model state correction algorithm of non-native speech recognition is the core technology of obtaining the status of mapping table (shown in Figure 2 modules 1-7). Hybrid model state correction algorithm is a new state based on similarity matrix mapping algorithm, the number of co-occurrence statistics, obtained correspondence between the two languages between states, using the state of the relationship is determined for correcting the right. system of the present invention comprises: a non-native speech interface for data collection of non-native speech, and the non-native voice data into the non-native status force alignment module and decoding module native state.

A native model module for native and non-native status decoding module decoding state correction model module provides up acoustic model for. a non-native language model module for non-native status force alignment module and non-native state correction model decoding modules provide non-native acoustic models. a native state decoding module for up according to the standard acoustic model for non-native voice data is decoded to give up the speech state-level segment information, namely the state of native speech decoding information, and native speech state-level segment information sent the native and non-native state similarity matrix calculation module. a non-native status forcing alignment module for non-native speakers of non-native speech acoustic model number dig forced alignment to obtain non-native speech state-level segment information that is non-native speech state reference information and non-native speech state-level segment information is fed up with the non-native state similarity matrix calculation module.

A native and non-native state similarity matrix calculation module for native speech state-level segment information and non-native speech state-level segment information on the time alignment, coincidence when the native and non-native speech speech state status When the time is greater than a predetermined threshold, I consider these two states appear a "co-occurrence", the

statistics of all "with the current" situation, and calculate non-native speech state corresponds to the state of native speech similarity matrix, and the similarity matrix information into the native and non-native state map heartfelt calculation module.

A native and non-native status mapping table calculation module for obtaining the status of a mapping table according to the similarity matrix calculation. And a non-native state correction model decoding module for decoding the speech recognition process, the acoustic model with mother tongue Status mapping table to find the corresponding non-modified up acoustic model state, non-female vocal school books say so get corrected end-use non-native acoustic models after the completion of the correction of non-native speech recognition; model.

The calculation process of the present invention are as follows: (Figure 2) First, non-native status Answer: choose a certain amount of non-native voice data (non-native voice interface 1), these data are used to generate non-native speakers Native similarity matrix. Acoustic model with a non-native (non-native model module 2) enforce these data alignment (orced-alignmlen), to obtain the status of non-native sequences, and record the time each state information (non-native status force alignment block 4) time information up acoustic model (up model Module 3) on the first step of non-native speech data decoding (Decode), to give up state sequences, and record each state: the second step up state recognition results. (Native state decoding module 5) The third step, "with the current" criteria: the first step and the second step, in the same number of non-native voice data, were obtained by non-native and native state sequence and the corresponding time segment information based on the positional relationship between the two states on the timeline sequence, you can define two states "with the current" phenomenon. Before computing "co-occurrence", the first definition of a "co-occurrence" matrix that is (state the number of native language * non-native language state number) of the matrix, each element of the position on the recording of the corresponding row and column of the state "The same is now" between the number. In the experiments of the present invention, the method, the definition of when two states in overlapping time period accounts for non-native language state duration half, counted as one "with the current." When non-

native language and native language status state capacity corpse appears a "co-occurrence", add 1 recorded in the corresponding co-occurrence matrix strict Xingshi column position. As shown in Figure 3, appearing between the "ae nonnat ve" Native Voice Status "aa native" state and non-native speech a "co-occurrence." eleven fourth step, the state of the similarity matrix calculation (native and non-native state similarity matrix calculation module 6): If set M, N, respectively native language and non-native language state number, A,, t (_ is derived from the co-occurrence matrix out of the similarity matrix, the matrix elements of the similarity of the case record and non-native language native language between states.A "the elements of the matrix column Xingshi strict set j is non-native language status, native language _ for the state, the similarity between the two is calculated as: _ C unt t_SJ AiJ N scattered counntOOJ SJ "" '(1) where, _T (M, N) _ M = 1 N fifth step, after obtained similarity matrix, the matrix is obtained based on the state map (native and non-native status mapping table calculation module 7). If the matrix A. Xingshi middle column element is a column matrix corpse The largest element, this means that the appropriate language from the strict state and the state of the corpse is most similar to the state produced the first candidate for correction Status corpse. If Iraq Xingshi column element is the second largest element corpse column, then the state is called state of the corpse of the second candidate revision status and so on, each state can be found in non-native language book n alternatives confessed amended state election native language in a matrix (n and lt; M) Q At this point, we Hybrid mapping information model state correction algorithm to obtain the native language and non-native language between state.

The sixth step, the next state based on the mapping relations between the mother tongue in the decoding process using the acoustic model for non-native acoustic model is modified (non-native state correction model decoding module 8). (. P) obtained according to the state of the fifth step mapping table, in the decoding process of speech recognition, the observed value of 0.5, with a non-native state n after correction candidate state body observation probability p, "becomes: p W_o_ = op 4_o) _) o p, "_), M "" (2) where represents the right to amend non-native state of heavy, spoon the corresponding first Shanghai candidate status, n is the number of candidates. p, "(P.)" And India 0 &#65533;),

"u denote the state of the corpse and the non-native Sichuan native state in the original concept of the value of 0.1 under police observation probability.

The present invention features: (1) a non-native language acoustic model is more adapted to the characteristics of the non-native speech with a native accent (2) A novel non-native speech recognition method based on hybrid model state as amended, (3) only rely on the premise of standard native acoustic models for improved identification performance of non-native speech. making English grammar based on a given real network data test with Chinese accent shows that correction algorithm based on hybrid model state of non-native speech recognition system error rate is set on an English test with a Chinese accent opposite end non-native speech recognition system using this methodology 5_10 percent decrease (relative value) calculation speed relative to the end of the use of non-native speech recognition system that methodology reduced by 20 percent -25 percent (relative value).

Cited by: CN102254553 A ; CN102254553 B ; CN103632668 A ; CN103632668 B ; US10224023 BB ; WO18086033 A1 ;

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 15/06 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200810239892.5

[43] 公开日 2010年2月17日

[11] 公开号 CN 101650943A

[22] 申请日 2008.12.19

[21] 申请号 200810239892.5

[71] 申请人 中国科学院声学研究所
地址 100190北京市海淀区北四环西路21号中国科学院声学研究所
共同申请人 北京中科信利技术有限公司

[72] 发明人 颜永红 潘接林 张晴晴

[74] 专利代理机构 北京法思腾知识产权代理有限公司
代理人 杨小蓉

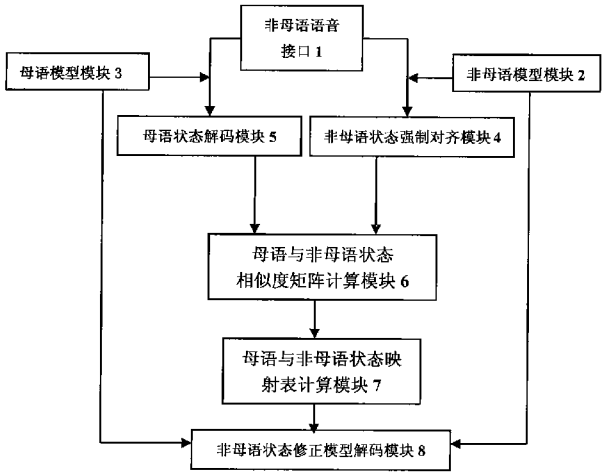
权利要求书 3 页 说明书 7 页 附图 2 页

[54] 发明名称

一种非母语语音识别系统及方法

[57] 摘要

本发明涉及一种基于混合模型状态修正的非母语语音识别系统及方法。该系统包括：非母语语音接口、母语模型模块、非母语模型模块、母语状态解码模块、非母语状态强制对齐模块、母语与非母语状态相似度矩阵计算模块、母语与非母语状态映射表计算模块及非母语状态修正模型解码模块，该系统及方法通过不同模型间的状态映射，利用说话人母语的声学模型在状态级别上对非母语声学模型进行修正，从而得到更为符合非母语发音特点的模型。该系统及方法有如下优点：在不增加任何非母语语音训练数据，仅依靠说话人母语训练数据的前提下，相对未采用该方法修正过的识别系统的识别性能有明显提高；同时系统识别语音的速度并未明显降低，具有很高的实用性。



1、一种非母语语音识别系统，其特征在于，所述系统包括：

一非母语语音接口，用于采集非母语语音数据，并将该非母语语音数据送入非母语状态强制对齐模块和母语状态解码模块；

一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提供母语声学模型；

一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解码模块提供非母语声学模型；

一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码，得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块；

一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数据进行强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块；

一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重合时间大于预先设定的门限值时，认为这两个状态出现一次“同现”，统计出所有的“同现”情况，并计算得到非母语语音状态对应于母语语音状态的相似度矩阵，并将该相似度矩阵信息送入母语与非母语状态映射表计算模块；

一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射表；及

一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母语声学模型；最终利用该修正后的非母语声学模型完成非母语语音识别。

2、根据权利要求1所述的非母语语音识别系统，其特征在于，所述非母语语音状态对应于母语语音状态的相似度矩阵通过以下公式得到：

$$A_{i,j}=\frac{\mathrm{count}(t_j|s_i)}{\sum_{n=1}^{T}\mathrm{count}(t_n|s_i)}$$

其中，M，N分别为中文和英文的状态数量；$A_{S,T}(M,N)$为相似度矩阵；$A_{i,j}$为矩阵的第i^{th}行j^{th}列元素；t_j为非母语语音状态，s_i为母语语音状态；$A_{i,j}\in A_{S,T}(M,N)$，$i=1\cdots\cdots M$，$j=1\cdots\cdots N$；$\mathrm{count}(t_j|s_i)$是母语语音状态s_i和非母语语音状态t_j之间的“同现”出现次数。

Petitioner Exhibit 1004, Page 587 of 1052

3、根据权利要求1所述的非母语语音识别系统，其特征在于，所述状态映射表通过以下方式得到：

如果$A_{i,j}$矩阵中n^{th}行j^{th}列元素是j^{th}列的第n大元素，则状态n^{th}是状态j^{th}的第n候选修正状态；每个非母语语言的状态都能在矩阵中找到母语语言的n个候选修正状态，其中，n<M。

4、根据权利要求1所述的非母语语音识别系统，其特征在于，所述非母语状态修正模型解码模块在语音识别解码过程中，对于观察值o_t，用n个候选状态修正以后的非母语状态j^{th}的观察概率$p_{j^{th}}(o_t)$变为：

$$p_{j^{th}}(o_t)=\alpha p_{j^{th}}(o_t)_{tar}+(1-\alpha)\sum_{l=1}^{n}a_{lj}p_{l^{th}}(o_t)_{sou}$$

其中，α表示非母语状态的修正权重，a_{lj}是相应第l^{th}个候选状态，n是候选数；$p_{j^{th}}(o_t)_{tar}$和$p_{l^{th}}(o_t)_{sou}$分别表示非母语状态j^{th}和母语状态l^{th}在观察值o_t下的原始观
察概率。

5、一种非母语语音识别方法，包括如下步骤：

(1)非母语语音接口采集一定数量的非母语语音数据，用于得到模型状态映射表；

(2)非母语状态强制对齐模块用非母语声学模型对非母语语音数据进行强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息；

(3)母语状态解码模块用标准母语声学模型对非母语语音数据进行解码，得到母语语音状态级分段信息，即母语语音状态解码信息；

(4)母语与非母语状态相似度矩阵计算模块将得到的非母语语音状态级分段信息和母语语音状态级分段信息在时间上对齐，当某两个状态的重合时间大于预先设定的门限值时，这两个状态出现一次“同现”；

(5)母语与非母语状态相似度矩阵计算模块统计出所有的“同现”情况并计算得到非母语语音状态对应于母语语音状态的相似度矩阵；

(6)母语与非母语状态映射表计算模块根据该相似度矩阵得到状态映射表；

(7)非母语状态修正模型解码模块根据得到的状态映射表，在语音识别的解码过程中，用映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母语声学模型。

6、根据权利要求5所述的非母语语音识别方法，其特征在于，所述非母语语音状态对应于母语语音状态的相似度矩阵通过以下公式得到：

Petitioner Exhibit 1004, Page 588 of 1052

$$A_{i,j}=\frac{\mathrm{count}(t_j|s_i)}{\sum_{n=1}^{M}\mathrm{count}(t_n|s_i)}$$

其中，M，N分别为中文和英文的状态数量；$A_{S,T}(M,N)$为相似度矩阵；$A_{i,j}$为矩阵的第i^{th}行j^{th}列元素；t_j为非母语语音状态，s_i为母语语音状态；$A_{i,j}\in A_{S,T}(M,N)$，$i=1\cdots\cdots M$，$j=1\cdots\cdots N$；$\mathrm{count}(t_j|s_i)$是母语语音状态s_i和非母语语音状态t_j之间的“同现”出现次数。

7、根据权利要求5所述的非母语语音识别方法，其特征在于，所述状态映射表通过以下方式得到：

如果$A_{i,j}$矩阵中n^{th}行j^{th}列元素是j^{th}列的第n大元素，则状态n^{th}是状态j^{th}的第n候选修正状态；每个非母语语言的状态都能在矩阵中找到母语语言的n个候选修正状态，其中，$n<M$。

8、根据权利要求5所述的非母语语音识别方法，其特征在于，所述非母语状态修正模型解码模块在语音识别解码过程中，对于观察值o_t，用n个候选状态修正以后的非母语状态j^{th}的观察概率$p_{j^{th}}(o_t)$变为：

$$p_{j^{th}}(o_t)=\alpha p_{j^{th}}(o_t)_{tar}+(1-\alpha)\sum_{l=1}^{n}a_{lj}p_{l^{th}}(o_t)_{sou}$$

其中，α表示非母语状态的修正权重，a_{lj}是相应第l^{th}个候选状态，n是候选数；$p_{j^{th}}(o_t)_{tar}$和$p_{l^{th}}(o_t)_{sou}$分别表示非母语状态j^{th}和母语状态l^{th}在观察值o_t下的原始观

察概率。

Petitioner Exhibit 1004, Page 589 of 1052

一种非母语语音识别系统及方法

技术领域

本发明涉及语音识别系统及方法，特别涉及基于混合模型状态修正的非母语语音识别系统及方法。

背景技术

随着现代社会信息的全球化，非母语语音识别成为语音识别技术领域研究热点之一。对非母语的语音识别相对母语的语音识别而言，识别性能将会大幅度降低，特别是对于有着严重口音的说话人。如何在保证语音识别系统在仅有少量非母语训练数据的前提下，对带有不同程度口音的非母语语音识别性能的提高是非母语语音识别研究的重点。

文献(Bohn，O.-S.，Flege，J.E.，"The production of new and similar vowels by adult German learners of English."[M]Stud.Second Lang.Acquis.14，131-158，1992.)指出，非母语说话人在进行目标语言发音时，可能会用说话人自身的母语发音进行替代，或者产生出有说话人母语发音特点和目标语言发音特点相结合的发音。这个结论启发我们，说话人自身的母语语言训练数据，可能对说话人非母语语音识别有帮助，特别是针对那些发音口音较重的语音数据。

当前，说话人自适应技术(如MAP、MLLR)已广泛的应用于处理非母语语音识别中(Z.Wang，T.Schultz，A.Waibel，"Comparison of acoustic model adaptation techniques on non-native speech"[C]，Proc.ICASSP 2003.)。这类方法主要是通过对少量非母语语音数据进行自适应，使得母语发音声学模型能够在一定程度上逼近非母语的发音特点。在这些方法中，自适应数据和测试数据的相似性是决定识别性能好坏的关键因素。尽管自适应技术在提高非母语语音识别性能上有贡献，但通过自适应以后的模型在非母语语音识别中的性能相比母语语音模型在母语语音识别的性能，仍然较低。文献(J.Humphries，P.Woodland，and D.Pearce."Using accent-specific pronunciation modeling for robust speech recognition."[C]In Proc.ICSLP'96，pages 2324-2327，Philadelphia，PA，October 1996.)研究了自适应算法的局限性，指出导致非母语语音识别率低的主要原因来自于说话人母语语系中无法涵盖的非母语语言发音。如何使非母语语音声学模型较好的模拟出这类语言发音是非母语语音识别研究工作的重点。

Petitioner Exhibit 1004, Page 590 of 1052

发明内容

为了克服上述现有技术的不足，本发明提供一种基于状态修正的非母语语音识别系统及方法，该系统及方法通过不同模型间的状态映射，利用说话人母语的声学模型在状态级别上对非母语声学模型进行修正，从而得到更为符合非母语发音特点的模型。

为了实现上述目的，本发明提出的一种非母语语音识别系统，包括：

一非母语语音接口，用于采集非母语语音数据，并将该非母语语音数据送入非母语状态强制对齐模块和母语状态解码模块。

一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提供母语声学模型。

一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解码模块提供非母语声学模型。

一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码，得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数据进行强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重合时间大于预先设定的门限值时，认为这两个状态出现一次“同现”，统计出所有的“同现”情况，并计算得到非母语语音状态对应于母语语音状态的相似度矩阵，并将该相似度矩阵信息送入母语与非母语状态映射表计算模块。

一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射表。及

一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母语声学模型；最终利用该修正后的非母语声学模型完成非母语语音识别。

其中，所述非母语语音状态对应于母语语音状态的相似度矩阵通过以下公式得到：

Petitioner Exhibit 1004, Page 591 of 1052

$$A_{i,j}=\frac{count(t_j|s_i)}{\sum_{n=1}^{N} count(t_n|s_i)}$$

其中，M，N分别为中文和英文的状态数量；$A_{S,T}(M,N)$为相似度矩阵；$A_{i,j}$为矩阵的第i^{th}行j^{th}列元素；t_j为非母语语音状态，s_i为母语语音状态；$A_{i,j}\in A_{S,T}(M,N)$，$i=1\cdots\cdots M$，$j=1\cdots\cdots N$；$count(t_j|s_i)$是母语语音状态s_i和非母语语音状态t_j之间的“同现”出现次数。

其中，所述状态映射表通过以下方式得到：

如果$A_{i,j}$矩阵中n^{th}行j^{th}列元素是j^{th}列的第n大元素，则状态n^{th}是状态j^{th}的第n候选修正状态；例如：如果$A_{i,j}$矩阵中i^{th}行j^{th}列元素是矩阵中j^{th}列的最大元素，这意味着来自相应语言的状态i^{th}和状态j^{th}最为相似，状态i^{th}是状态j^{th}的第一候选修正状态。如果k^{th}行j^{th}列元素是j^{th}列的第二大元素，那么状态k^{th}是状态j^{th}的第二候选修正状态。如此类推，每个非母语语言的状态都能在矩阵中找到母语语言的n个候选修正状态(n<M)。

其中，所述非母语状态修正模型解码模块在语音识别解码过程中，对于观察值o_t，用n个候选状态修正以后的非母语状态j^{th}的观察概率$p_{j^{th}}(o_t)$变为：

$$p_{j^{th}}(o_t)=\alpha p_{j^{th}}(o_t)_{tar}+(1-\alpha)\sum_{i=1}^{n}a_{1j}p_{i^{th}}(o_t)_{sou}$$

其中，α表示非母语状态的修正权重，a_{1j}是相应第1^{th}个候选状态，n是候选数；$p_{j^{th}}(o_t)_{tar}$和$p_{l^{th}}(o_t)_{sou}$分别表示非母语状态j^{th}和母语状态1^{th}在观察值o_t下的原始观

察概率。

本发明提供的一种非母语语音识别方法，包括如下步骤：

(1)非母语语音接口采集一定数量的非母语语音数据，用于得到模型状态映射表。

(2)非母语状态强制对齐模块用非母语声学模型对非母语语音数据进行强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息。

(3)母语状态解码模块用标准母语声学模型对非母语语音数据进行解码，得到母语语音状态级分段信息，即母语语音状态解码信息。

(4)母语与非母语状态相似度矩阵计算模块将得到的非母语语音状态级分段信息和母语语音状态级分段信息在时间上对齐，当某两个状态的重合时间大于预先设定的门限值时，这两个状态出现一次“同现”。

(5)母语与非母语状态相似度矩阵计算模块统计出所有的“同现”情况并计算

Petitioner Exhibit 1004, Page 592 of 1052

得到非母语语音状态对应于母语语音状态的相似度矩阵。

(6)母语与非母语状态映射表计算模块根据该相似度矩阵得到状态映射表。

(7)非母语状态修正模型解码模块根据得到的状态映射表，在语音识别的解码过程中，用映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母语声学模型。

本发明的优点在于：

(1)本发明的非母语语音识别系统及方法使非母语声学模型更适应带有母语口音的非母语语音特点。

(2)本发明的非母语语音识别系统及方法，采用不同模型间的状态映射，利用说话人母语的声学模型在状态级别上对非母语声学模型进行修正，解码器根据得到的状态映射表实现非母语语音识别。混合模型状态修正的非母语语音识别系统对非母语语音的识别率与未采用该方法修正过的识别系统相比较；并且在不增加任何非母语语音训练数据，仅依靠标准母语训练数据的前提下，对带有母语口音的非母语语音的识别情况相对未采用该方法修正过的识别系统有了明显提高；同时系统识别语音的速度并未明显下降。

(3)本发明的非母语语音识别系统及方法实现了利用母语声学模型对非母语声学模型的修正，提高系统对非母语语音的识别性能。相比语音自适应技术，基于混合模型状态的修正方法不需要增加更多的非母语训练数据。由于用于修正的母语声学模型的规模可以控制在很小的范围，修正后的模型在计算速度上也不会有明显增加。我们进行了大量实网数据的测试，结果显示在识别性能方面，由于该系统在状态级别上的映射信息是基于解码器输出的状态序列情况进行统计得到的，这种准则较为直接，基本能真实的反映出两种语言状态间的相似程度。与未使用该方法的系统相比，基于混合模型状态修正的非母语语音识别系统对非母语语音的识别错误率相对下降可达到5-10％。

附图说明

图1本发明的基于混合模型状态修正算法的非母语语音识别系统及方法的基本原理框图；

图2本发明的非母语语音识别系统的流程框图；

图3本发明的具体实施例母语语音状态与非母语语音状态之间的“同现”说明图。

Petitioner Exhibit 1004, Page 593 of 1052

具体实施方式

下面结合附图对本发明的具体实施方式做进一步详细描述：

图1是基于混合模型状态修正算法的非母语语音识别系统的基本原理框图，它描述了混合模型状态修正算法的核心组成成分，主要由几个部分构成：非母语语音、母语/非母语声学模型、强制对齐、解码、状态映射表生成以及非母语状态修正模型的解码。图2是基于混合模型状态修正算法的非母语语音识别系统的具体实施流程框图。

本发明涉及的混合模型状态修正算法的非母语语音识别的核心技术在于状态映射表的获取（如图2所示的模块1到7）。混合模型状态修正算法是一种新型的基于相似度矩阵的状态映射算法，通过对同现次数的统计，得到两种语言状态间的对应关系，利用该关系确定用于修正的状态对。

本发明的系统包括：一非母语语音接口，用于采集非母语语音数据，并将该非母语语音数据送入非母语状态强制对齐模块和母语状态解码模块。

一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提供母语声学模型。

一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解码模块提供非母语声学模型。

一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码，得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数据进行强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重合时间大于预先设定的门限值时，认为这两个状态出现一次“同现”，统计出所有的“同现”情况，并计算得到非母语语音状态对应于母语语音状态的相似度矩阵，并将该相似度矩阵信息送入母语与非母语状态映射表计算模块。

一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射表。及

一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母

Petitioner Exhibit 1004, Page 594 of 1052

语声学模型；最终利用该修正后的非母语声学模型完成非母语语音识别。

本发明的具体计算流程如下：(如图2)

第一步，非母语状态参考答案：选择一定量的非母语语音数据(非母语语音接口1)，这些数据用于生成母语对非母语的相似度矩阵。用非母语声学模型(非母语模型模块2)对这些数据进行强制对齐(Forced-alignment)，得到非母语状态序列，同时记录下每个状态的时间信息(非母语状态强制对齐模块4)。

第二步，母语状态识别结果：用母语声学模型(母语模型模块3)对第一步中的非母语语音数据进行解码(Decode)，得到母语状态序列，同时记录下每个状态的时间信息(母语状态解码模块5)。

第三步，“同现”准则：通过第一步和第二步，在同一批非母语语音数据上，分别得到了非母语和母语的状态序列以及相应的时间段信息，根据这两个状态序列在时间轴上的位置关系，可以定义两个状态的“同现”现象。在计算“同现”前，首先定义一个“同现”矩阵，该矩阵为(母语语言状态个数*非母语语言状态个数)的矩阵，每个元素位置上记录了对应行和列的状态之间的“同现”次数。在本发明所述方法的实验中，定义当两个状态在时间段上重叠的时间占到非母语语言状态持续时间一半时，算作一次“同现”。当非母语语言状态i^{th}和母语语言状态j^{th}出现一次“同现”时，在对应同现矩阵的i^{th}行j^{th}列的位置上加1记录。如图3所示，母语语音状态“aa_native”和非母语语音状态“ae_nonnative”之间出现的一次“同现”。

第四步，状态相似度矩阵计算(母语与非母语状态相似度矩阵计算模块6)：若设定M，N分别为母语语言和非母语语言状态个数，$A_{S,T}(M, N)$为从同现矩阵派生出来的相似度矩阵，矩阵元素记录了母语语言和非母语语言状态之间的相似度情况。$A_{i,j}$为矩阵i^{th}行j^{th}列的元素，设定t_j为非母语语言状态，s_i为母语语言状态，两者之间的相似度计算为：

$$A_{ij} = \frac{count(t_j, s_i)}{\sum_{n=1}^{M} count(t_n, s_i)} \quad \cdots (1)$$

其中，$A_{i,j} \in A_{S,T}(M, N)$，$i = 1 \cdots\cdots M$，$j = 1 \cdots\cdots N$。

第五步，得到相似度矩阵后，根据该矩阵得到状态映射表(母语与非母语状态映射表计算模块7)。如果$A_{i,j}$矩阵中i^{th}行j^{th}列元素是矩阵中j^{th}列的最大元素，这意味着来自相应语言的状态i^{th}和状态j^{th}最为相似，状态i^{th}是状态j^{th}的第一候选修正状态。如果k^{th}行j^{th}列元素是j^{th}列的第二大元素，那么状态k^{th}是状态j^{th}的第二候选修正状态。如此类推，每个非母语语言的状态都能在矩阵中找到母语语言的n个候

Petitioner Exhibit 1004, Page 595 of 1052

选修正状态(n<M)。

到此，我们基于混合模型状态修正算法，得到了母语语言和非母语语言状态间的映射信息。

第六步，接下来根据得到的状态间的映射关系，在解码过程中用母语声学模型对非母语声学模型进行修正(非母语状态修正模型解码模块8)。根据第五步得到的状态映射表，在语音识别的解码过程中，对于观察值o_t，用n个候选状态修正以后的非母语状态j^{th}的观察概率$p_{j^{th}}(o_t)$变为：

$$p_{j^{th}}(o_t)=\alpha p_{j^{th}}(o_t)_{tar}+(1-\alpha)\sum_{l=1}^{n}a_{lj}p_{l^{th}}(o_t)_{sou} \quad ---(2)$$

这里α表示非母语状态的修正权重，a_{lj}是相应第l^{th}个候选状态，n是候选数。$p_{j^{th}}(o_t)_{tar}$和$p_{l^{th}}(o_t)_{sou}$分别表示非母语状态j^{th}和母语状态l^{th}在观察值o_t下的原始观

察概率。

本发明的特点：(1)使非母语声学模型更适应带有母语口音的非母语语音特点(2)提出一种新型基于混合模型状态修正的非母语语音识别方法，(3)在仅依靠标准母语声学模型的前提下，提高系统对非母语语音的识别性能。

在进行基于给定语法的带有中文口音的英文实网数据测试中显示，基于混合模型状态修正算法的非母语语音识别系统在带有中文口音的英文测试集上的错误率相对未使用该方法的非母语语音识别系统下降5%-10%(相对值)；计算速度相对于未使用该方法的非母语语音识别系统降低20%-25%(相对值)。

Petitioner Exhibit 1004, Page 596 of 1052

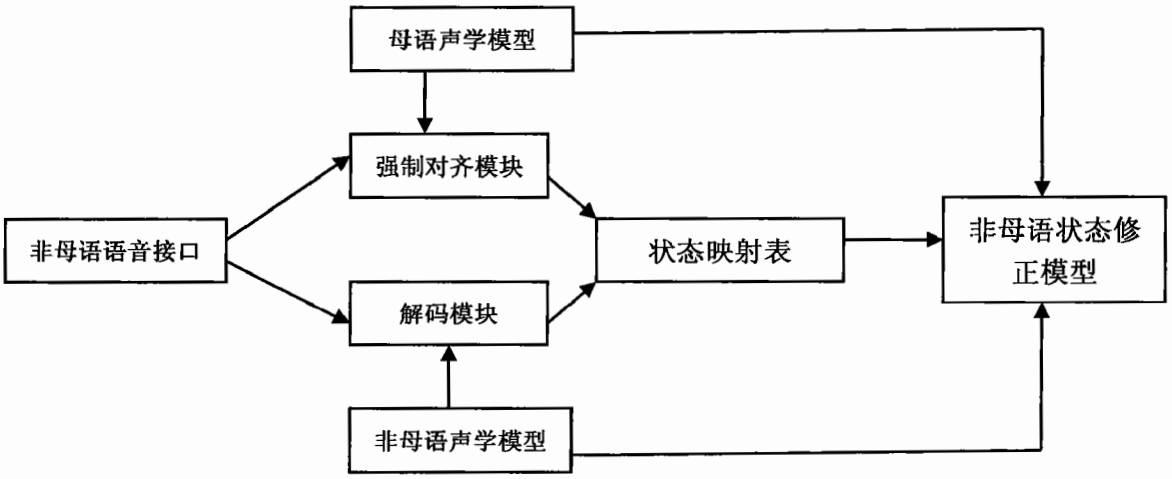


图1

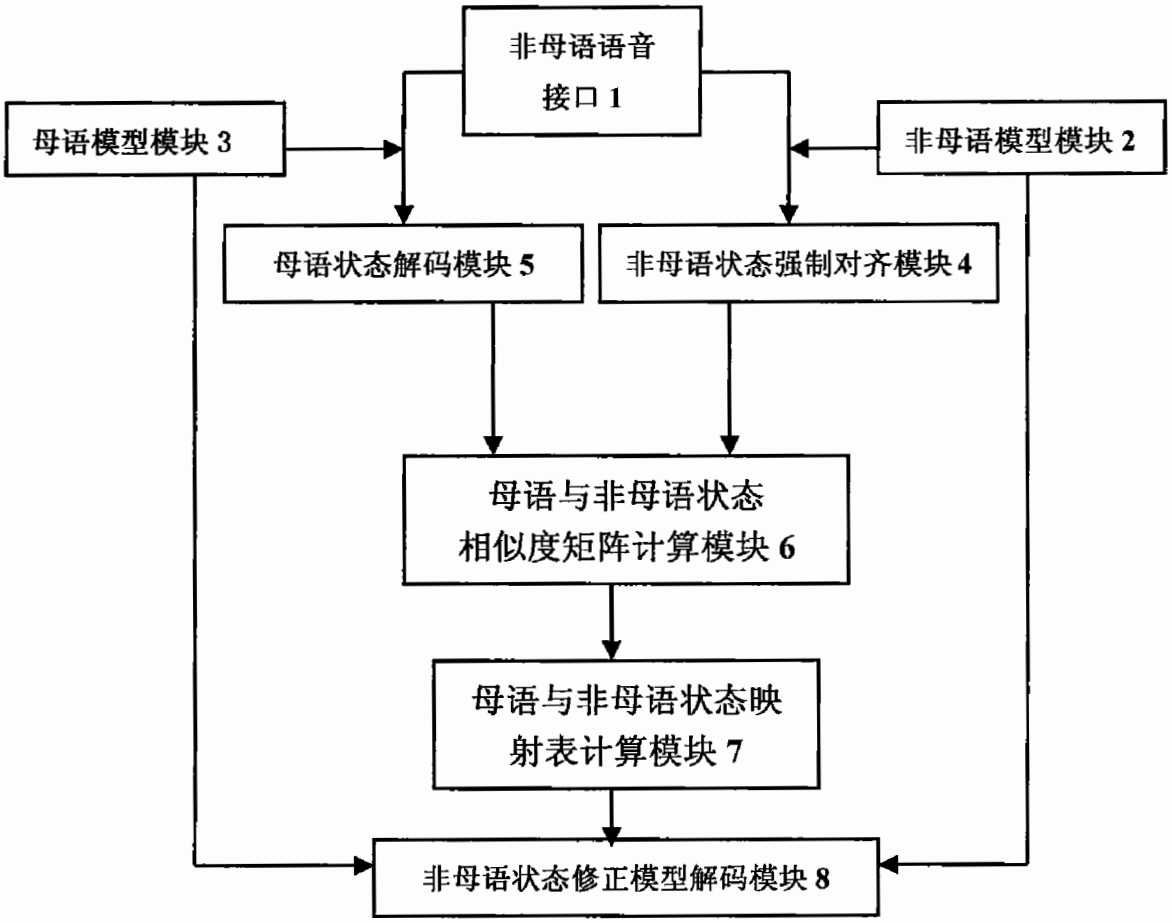


图2

Petitioner Exhibit 1004, Page 597 of 1052

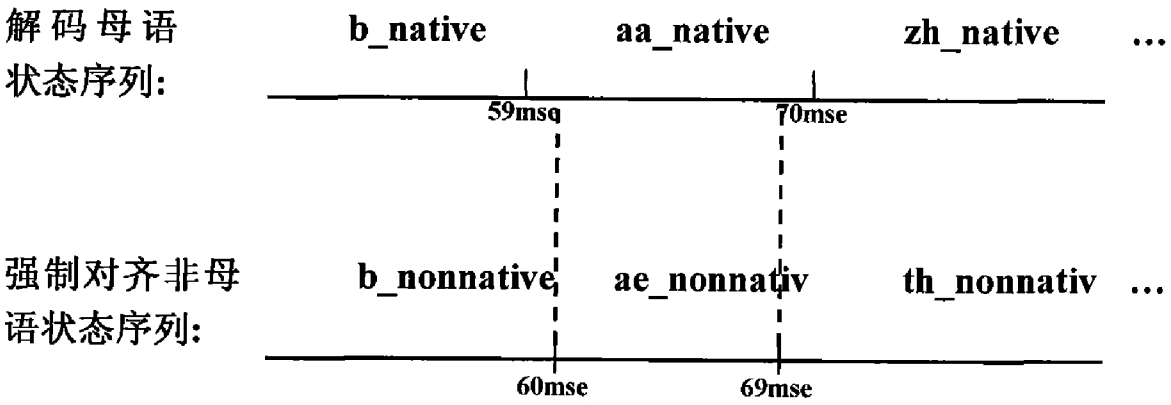


图3

Petitioner Exhibit 1004, Page 598 of 1052

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
N0484.71178		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US2016/037495	15 June 2016 (15-06-2016)	
Applicant		
NUANCE COMMUNICATIONS, INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**
   a. With regard to the **language,** the international search was carried out on the basis of:
      [X] the international application in the language in which it was filed
      [ ] a translation of the international application into _____, which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b))
   b. [ ] This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).
   c. [ ] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. [ ] **Certain claims were found unsearchable** (See Box No. II)

3. [ ] **Unity of invention is lacking** (see Box No III)

4. With regard to the **title,**
   [X] the text is approved as submitted by the applicant
   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**
   [X] the text is approved as submitted by the applicant
   [ ] the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority

6. With regard to the **drawings,**
   a. the figure of the **drawings** to be published with the abstract is Figure No. 1
      [X] as suggested by the applicant
      [ ] as selected by this Authority, because the applicant failed to suggest a figure
      [ ] as selected by this Authority, because this figure better characterizes the invention
   b. [ ] none of the figures is to be published with the abstract

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/037495

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/22
ADD. G10L15/08 G10L17/22

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104 575 504 A (UNIV SHANGHAI) 29 April 2015 (2015-04-29) paragraph [0043] - paragraph [0049]; figure 3	1-53
A	EP 2 899 955 A1 (HUAWEI DEVICE CO LTD [CN]) 29 July 2015 (2015-07-29) paragraph [0020] - paragraph [0029]; figure 1	1-53
A	US 2014/249817 A1 (HART GREGORY MICHAEL [US] ET AL) 4 September 2014 (2014-09-04) paragraph [0025] - paragraph [0026]; figure 1	1-53

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 November 2016	05/12/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Zimmermann, Elko

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/037495

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CN 104575504 A	29-04-2015	NONE	
EP 2899955 A1	29-07-2015	CN 103595869 A	19-02-2014
		EP 2899955 A1	29-07-2015
		JP 2016502829 A	28-01-2016
		US 2015141079 A1	21-05-2015
		WO 2015070644 A1	21-05-2015
US 2014249817 A1	04-09-2014	EP 2965314 A1	13-01-2016
		US 2014249817 A1	04-09-2014
		WO 2014137751 A1	12-09-2014

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/037495	15.06.2016	

International Patent Classification (IPC) or both national classification and IPC
INV. G10L15/22 ADD. G10L15/08 G10L17/22

Applicant
NUANCE COMMUNICATIONS, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	see form PCT/ISA/210	Zimmermann, Elko Telephone No. +49 89 2399-0

Form PCT/ISA/237 (Cover Sheet) (January 2015)

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-53
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-53
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1-53
	No:	Claims	

2. Citations and explanations

see separate sheet

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

CN 104 575 504 A is regarded as being the prior art closest to the subject-matter of claim 1, and discloses a system wherein an acoustic signal is processed in parallel by a speech recognizer and a voiceprint identification unit in order to determine whether a wake-up command was received from a registered user, see figure 3 and paragraphs [0050] to [0056] of the computer-generated translation annexed to this opinion, which correspond to paragraphs [0043] to [0049] of CN 104 575 504 A.

The subject-matter of claim 1 differs from this known system in that the acoustic signal is interpreted using information indicative of the speaker's identity and is therefore new (Article 33(2) PCT).

The problem to be solved by the present invention may be regarded as improving the robustness of the wake-up word recognition.

The solution to this problem proposed in claim 1 of the present application is considered as involving an inventive step (Article 33(3) PCT) for the following reasons: By using information indicative of the speaker's identity components of the speech recognizer like vocabulary or acoustic model can be adapted in order to improve the speech recognizer's ability to accurately determine whether a speaker uttered any one of the wake-up words. In the prior art at hand there is no hint at interpreting the acoustic signal using information indicative of the speaker's identity, and it is not considered to be obvious in view of the parallel processing illustrated in figure 3 of CN 104 575 504 A.

Claims 20 and 37 correspond to claim 1, and claims 2-19, 21-36 and 38-53 are dependent claims and as such also meet the requirements of the PCT with respect to novelty and inventive step.

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

Form PTO - 1449

Copies of Cited References

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature			
Printed name	Andrew J. Tibbetts		
Date		Reg. No.	65,139

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: May 22, 2019 Electronic Signature for MacAulay S. Rush: /MacAulay S. Rush/

7243456.1

PATENT COOPERATION TREATY

ADVANCE E-MAIL

PCT

From the INTERNATIONAL BUREAU

NOTIFICATION CONCERNING TRANSMITTAL OF COPY OF INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY (CHAPTER I OF THE PATENT COOPERATION TREATY)
(PCT Rule 44bis.1(c))

To:

RUDOY, Daniel, G.
Wolf, Greenfield & Sacks. P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)*
27 December 2018 (27.12.2018)

Applicant's or agent's file reference
N0484.71178

IMPORTANT NOTICE

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/037495	15 June 2016 (15.06.2016)	

Applicant
NUANCE COMMUNICATIONS, INC.

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Kihwan Moon

e-mail: pct.team1@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference N0484.71178	**FOR FURTHER ACTION** See item 4 below	
International application No. PCT/US2016/037495	International filing date *(day/month/year)* 15 June 2016 (15.06.2016)	Priority date *(day/month/year)*
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant NUANCE COMMUNICATIONS, INC.		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☐	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 18 December 2018 (18.12.2018)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Kihwan Moon e-mail: pct.team1@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/037495	15.06.2016	

International Patent Classification (IPC) or both national classification and IPC
INV. G10L15/22 ADD. G10L15/08 G10L17/22

Applicant
NUANCE COMMUNICATIONS, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	see form PCT/ISA/210	Zimmermann, Elko Telephone No. +49 89 2399-0

Form PCT/ISA/237 (Cover Sheet) (January 2015)

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-53
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-53
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1-53
	No:	Claims	

2. Citations and explanations

see separate sheet

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

CN 104 575 504 A is regarded as being the prior art closest to the subject-matter of claim 1, and discloses a system wherein an acoustic signal is processed in parallel by a speech recognizer and a voiceprint identification unit in order to determine whether a wake-up command was received from a registered user, see figure 3 and paragraphs [0050] to [0056] of the computer-generated translation annexed to this opinion, which correspond to paragraphs [0043] to [0049] of CN 104 575 504 A.

The subject-matter of claim 1 differs from this known system in that the acoustic signal is interpreted using information indicative of the speaker's identity and is therefore new (Article 33(2) PCT).

The problem to be solved by the present invention may be regarded as improving the robustness of the wake-up word recognition.

The solution to this problem proposed in claim 1 of the present application is considered as involving an inventive step (Article 33(3) PCT) for the following reasons: By using information indicative of the speaker's identity components of the speech recognizer like vocabulary or acoustic model can be adapted in order to improve the speech recognizer's ability to accurately determine whether a speaker uttered any one of the wake-up words. In the prior art at hand there is no hint at interpreting the acoustic signal using information indicative of the speaker's identity, and it is not considered to be obvious in view of the parallel processing illustrated in figure 3 of CN 104 575 504 A.

Claims 20 and 37 correspond to claim 1, and claims 2-19, 21-36 and 38-53 are dependent claims and as such also meet the requirements of the PCT with respect to novelty and inventive step.

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong, III	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		9,940,936	B2	Sejnoha et al.	04-10-2018
		2014-0249817	A1	Hart et al.	09-04-2014
		2014-0365225	A1	Haiut	12-11-2014
		2016-0189706	A1	Zopf et al.	06-30-2016

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		CN	101650943	A	Chinese Acad Inst Acoustics	02-17-2010	Y
		CN	103021409	A	Anhui Ustc Iflytek Co	04-03-2013	Y
		CN	103632668	A	Beijing Baidu Netcom Sci and Tec.	03-12-2014	Y
		CN	104575504	A	Univ Shanghai	04-29-2015	Y
		CN	105575395	A	Zte corp	05-11-2016	Y
		CN	106098059	A	Univ Shanghai Jiaotong	11-09-2016	Y
		EP	2 899 955	A1	Huawei Device	07-29-2015	
		WO	2014/066192	A1	Microsoft Corp.	05-01-2014	

OTHER ART — NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		International Search Report and Written Opinion for International Application No. PCT/CN2016/105343 mailed September 21, 2017 (N0484.71177WO00).	
		International Search Report and Written Opinion for International Application No. PCT/US2016/037495 mailed December 05, 2016 (N0484.71178WO00).	
		International Preliminary Report on Patentability for International Application No. PCT/US2016/037495 mailed December 27, 2018 (N0484.71178WO00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7243457.1

DOCKET NO.: N0484.71090US04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: May 22, 2019 Electronic Signature for MacAulay S. Rush: /MacAulay S. Rush/

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed before the mailing of a first Office Action after the filing of a Request for Continued Examination under 37 C.F.R. § 1.114.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/308,849	12-11-2018	Pfeffinger et al.	N0484.71178US01
*16/348,689	05-09-2019	Ren et al.	N0484.71177US00

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR .

The above-identified co-pending applications may include subject matter that is technically-related to subject matter of the present application and may contain claims that recite elements similar to those recited in claims of the present application. In handling prosecution of the above-identified applications, the Examiners for these other applications presumably will search for relevant prior art and, in some circumstances, may take positions about the prior art and/or the interpretation of the claims that may be material to the examination of the present application.

Accordingly, the Examiner in the present application is requested to review the file histories of the above-identified co-pending applications to see whether there is any information that the Examiner believes may be relevant to the prosecution of the present application. It is assumed that the Examiner has access to the file histories of the above-identified applications. However, if the Examiner would like copies of any or all of the information included in the file histories of the above-identified applications, the Examiner is asked to contact the undersigned representative.

The undersigned would like to bring to the Examiner's attention the enclosed search reports or other communications from corresponding or related International or Foreign National Applications:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.

Petitioner Exhibit 1004, Page 615 of 1052

PCT/US2016/037495	**12-05-2016**	**International Search Report and Written Opinion**	**N0484.71178WO00**
PCT/CN2016/105343	**09-21-2017**	**International Search Report and Written Opinion**	**N0484.71177WO00**
PCT/US2016/037495	**12-27-2018**	**International Preliminary Report on Patentability**	**N0484.71178WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

Petitioner Exhibit 1004, Page 616 of 1052

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Respectfully submitted,

/Andrew J. Tibbetts/

By: _____

Andrew J. Tibbetts, Reg. No. 65,139
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: N0484.71090US04
Date: May 22, 2019

Electronic Acknowledgement Receipt

EFS ID:	36086256
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Andrew J. Tibbetts/MacAulay Rush
Filer Authorized By:	Andrew J. Tibbetts
Attorney Docket Number:	N0484.71090US04
Receipt Date:	22-MAY-2019
Filing Date:	16-AUG-2016
Time Stamp:	14:52:11
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	N048471090US04-TRN-AJT.pdf	76221 f23a2e63ec3100f2b601a33527751cf730cb34f1	no	1

Warnings:

Information:					
2		N048471090US04-IDS-AJT.pdf	147208 40f7f11f0bfa2a4d363e46729af2550c81a30737	yes	5
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Transmittal Letter		1	4	
	Information Disclosure Statement (IDS) Form (SB08)		5	5	
Warnings:					
Information:					
3	Foreign Reference	CN101650943_pdf.pdf	446560 161e5f78f8b1f253c53924532a5ed43573eb1d30	no	27
Warnings:					
Information:					
4	Foreign Reference	CN103021409_pdf.pdf	1809820 a4fd72fa78b0a58c22f0efa782fa1f97e590fe19	no	23
Warnings:					
Information:					
5	Foreign Reference	CN103632668_pdf.pdf	1526633 c95e56301e56b2dff2e74e2a14e5f43abfe39831	no	58
Warnings:					
Information:					
6	Foreign Reference	CN104575504_pdf.pdf	846845 ae8a77c28ce3186c9abca2777b9832b1f8c988ed	no	30
Warnings:					
Information:					
7	Foreign Reference	CN105575395_pdf.pdf	2748082 31fa4c0d963427b141e6ff0b63a88cdaf709d239	no	52
Warnings:					
Information:					

8	Foreign Reference	CN106098059_pdf.pdf	922481 089ee57084b2d748ac5be13855b5e3c675eec904	no	24
Warnings:					
Information:					
9	Foreign Reference	EP2899955.pdf	301349 6db924a5d95d09a049b9795c987a1ea56ff194e3	no	17
Warnings:					
Information:					
10	Foreign Reference	WO14066192.pdf	258117 400ca0a10e0462f415754853a6bc1f729aa89ca3	no	26
Warnings:					
Information:					
11	Other Reference-Patent/App/Search documents	N048471177WO00_ISRWO_Mailed_09-21-2017.pdf	355918 f95cdf82350592183c72d020d22e5ce76e874819	no	9
Warnings:					
Information:					
12	Other Reference-Patent/App/Search documents	N048471178WO00_IPRP_Maild_12-27-2018.pdf	193960 3350bf4bdec543970a0ba4ed58112b9df530b6f9	no	6
Warnings:					
Information:					
13	Other Reference-Patent/App/Search documents	N048471178WO00_ISRWO_Mailed_12-05-2016.pdf	231787 fe8e860e798795f77354fedb3935c5970719174b	no	7
Warnings:					
Information:					
		Total Files Size (in bytes):	9864981		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Publication number: CN106098059 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20161109
Application number: CN201610462976
Application date: 20160623
Priority: CN201610462976201606 23CN201614629 76;
Assignee: ?????? ; SHANGHAI JIAO TONG UNIV ;
Assignee[std]: UNIV SHANGHAI JIAOTONG ;
Inventor: ??? ; ??? ; ??? ; ?? ; ??? ;
Inventor[std]: CHANG XUANKAI ; YU KAI ; QIAN YANMIN ; CHEN ZHEHUAI ; ZHUANG YIMENG ;
International class[8]: G10L15/02; G10L15/06; G10L15/22;
Cooperative Patent class: G10L15/02 ; G10L15/063 ; G10L15/22 ; G10L2015/025 ; G10L2015/0631 ; G10L2015/223 ;
Family members: CN106098059 A

Title: CUSTOMIZABLE VOICE WAKE-UP METHOD AND SYSTEM ????????????

Title: 可定制语音唤醒方法及系统

Abstract:

The invention discloses a customizable voice wake-up method and a system. Through using a long and short memory network and a connection time sequence classification model, phoneme information of the voice information is modeled, the model is trained, the model after training is adopted for testing, and a possible phoneme sequence most similar to a customized wake-up word is searched in a generated Lattice network structure and serves as a judgment basis. The feature of CTC model output posteriori probability sparsity is used for high-efficiency searching, and the technique of calculating the confidence of the wake-up word is completed. On one hand, high wake-up performance can be obtained, that is, high accuracy and low error wake-up can be obtained; and on the other hand, the calculation resources of the application system are relatively little consumed.

?????????????,??,????????????????? Lattice???????????????????????????????????? CTC??????????????????????,????????????????????????????????????,?????,????,?????????????????????

Abstract:

一种可定制语音唤醒方法及系统，通过使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进行建模并对模型进行训练，采用训练后模型进行测试并在生成的Lattice网络结构上搜索与定制的唤醒

A customizable voice wake methods and systems, by using short-term memory network and connection time series classification model for the modeling of phoneme information voice messages and training

词最相似的可能音素序列来作为判断依据。本发明利用CTC模型输出后验概率稀疏的特点进行高效搜索，从而完成对唤醒词置信度计算的技术。本发明一方面可以得到较高的唤醒性，即高准确率，低误唤醒，另一方面对应用系统的计算资源消耗相对较少。

model, training model is used to test and build on a Lattice network structure search and customized wake the most similar word of possible phoneme sequences as a judge. By CTC model outputs of the invention a posteriori probability characteristics for efficient search of sparse, thus completing the wake the word confidence measure technology. The invention can obtain a higher sexual arousal, high accuracy and low false alarm wake up, on the application of the system of consumption of computing resources is relatively small.

Claims:

1.一种可定制语音唤醒方法，其特征在于，包括以下步骤：

步骤1)使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进行建 模；

步骤2)对模型进行训练：将事先收集并标注好的音频数据首先要经过传统的信号处理 方法做预处理并提取可供模型训练的声学特征，模型将特征数据作为输入，将标注的音素 信息作为输出，在海量数据下通过深度学习的方法完成模型各参数量的训练，得到可使用 的长短时记忆网络和连接时序分类模型；

步骤3)采用训练后模型进行测试：对测试语音做同样的预处理和特征提取，并输入模 型，模型将输出每一帧所有建模单元，即音素可能出现的后验概率；

步骤4)唤醒词搜索：通过在生成的Lattice网络结构上搜索与定制的唤醒词最相似的 可能音素序列来作为判断依据。

2.根据权利要求1所述的可定制语音唤醒方法，其特征是，所述的最相似的可能音素序 列H_{max}，通过判断音素序列在语音中的出现概率，即出现概率越高且越与唤醒词音素序列相 似的序列越可能是唤醒词的方式得到，具体为：

$$H_{max} = \underset{H}{\operatorname{argmax}} P(H|T) = \underset{H}{\operatorname{argmax}} \frac{P(T|H)P(H)}{P(T)} = \underset{H}{\operatorname{argmax}} P(T|H)P(H)$$

其中：P(T)为目标音素序列T，$T=\{t_1,t_2,\cdots,t_n\}$的观察概率，t_i表示T中的第i个音素，i为1～n；

目标音素序列T位于CTC音素网格结构中的所有音素序列的概率为：

$P(T|L_H) \propto P(L_H|T)P(T) \approx P(H_{max}|T)P(T)$，其中：$L_H$表示CTC音素网格结构中的所有音素 序列，而可能音素序列H_{max}即为T已知时L_H中的最高概率；

P(H)为音素序列H，$H=\{n_{ij1},n_{(i+1)j2},\cdots,n_{(i+m-1)jm}\}$的观察概率，$n_{ij}$为lattice网格 结构中第i栏第j列的音素，音素序列H也可以表示为$H=\{h_1,h_2,\cdots,h_m\}$，其中：$h_k = n_{(i+k-1)jk}$；

1. a customizable voice awakens, which is characterized, comprising the following steps:

Step 1) use short-term memory network and connection time series classification model for the modeling of phoneme information voice messages;

Steps 2) on model for training: will prior collection and mark good of audio data first to after traditional of signal processing method do pretreatment and extraction can for model training of acoustic features, model will features data as entered, will mark of phonemes information as output, in mass data Xia through depth learning of method completed model the parameter number of training, get can using of length Shi memory network and connection timing classification model;

Step 3) training model to test: to test voice doing the same preprocessing and feature extracting and enter the model, the model output each frame of all model units, that phoneme posterior probabilities that may occur;

Step 4) wake the word search: Lattice network structure generated by searches and customized wake the most similar word of possible

P(H)通过unigram假设，即累积乘音素序列中每个音素的后验概率获得，具体为：

$$P(H)=\prod_{i=1}^{m}P(h_i|H_{1:i-1})\approx\prod_{i=1}^{m}P(h_i)$$

；

P(T/H)为目标音素序列T和音素序列H之间的相似程 度，

$$P(T|H)\triangleq\prod_{i=1}^{MED(T,H)}P(op_i|R=T,E=H)$$

，

即目标音素序列和假设序列之间的每个编辑 操作的概率之积作为衡量音素序列相似性的指标，MED(T,H)表示目标音素序列T和音素序 列H的最少编辑操作次数，$P(op_i|R=T,E=H)$表示当参考音素序列R为T时猜测音素序列E为 H时，序列E和R之间的第i次编辑操作op_i的概率。

3.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的编辑操作是指：插入、 删除、替换操作，其概率，即$P(insert(e_i))$、$P(delete(r_i))$和$P(r_i/e_i)$直接由先验知识得 出，r_i和e_i分别取自参考音素序列R和猜测音素序列E。

4.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤1包括：

1.1)确定模型结构：根据应用场景设备的计算能力预估模型复杂度，首先设定参数量 的上限，然后在此限定下指定网络结构；

1.2)初始化长短时记忆网络：使用随机初始化的参数或model transferring方式初始 化。

5.根据权利要求4所述的可定制语音唤醒方法，其特征是，所述的初始化长短时记忆网 络，使用cross entropy准则预训练一个标准的声学长短时记忆网络，然后通过复制参数的 方式初始化模型。

6.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤2包括：

2.1)对训练数据提取fbank，即Filter-bank声学特征；

2.2)使用随机梯度下降方式完成模型的训练，训练参数需根据模型结构以及训练数据 大小设定。

7.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤3包括：

3.1)对测试数据提取fbank声学特征，特征提取过程要求与训练数据特征提取一致；

phoneme sequences as a judge.

2. According to claim 1 can be customized voice-wake method, which is characterized, the closest possible phoneme sequences described h_{Max}, by determining the probability in a sequence of phonemes in speech, the higher the probability of occurrence and more and wake the word phoneme sequences more similar sequences may be wake the word manner, in particular:

[IMAGE]

: P (t) is the target phoneme sequence T,T= {t_1, t_2, ���, t_n} the observed probability, t_i t l in phonemes, l 1~n;

T target phoneme sequences in the CTC phoneme lattice structure probabilities of all sequences of phonemes are:

P(T| L_H)?P(L_H| T)P(T)�P(H_{Max}| T) p (t), of which: l_h all phoneme sequences in the CTC phoneme lattice structure, possible phoneme sequences h_{Max} is known that t l_h, the highest probability;

P (h) for phonemes sequence H,H={n_{ij1}, $n_{(i\ 1)\ j2}$, ���, $n_{(i\ m-1)\ JM}$} of observation probability, n_{IJ} for lattice network grid structure in the subsection I bar subsection j column of phonemes, phonemes sequence h also can said for H={h_1, h_2, ���, H_m} where: $h_k = n_{(i\ k-1)\ JK}$;

P (h) through unigram assumptions, cumulative phoneme sequences, each phoneme posterior probabilities obtained, in particular to:

3.2)将提取的特征逐帧输入训练得到的模型，直接计算出每一帧的后验概率。

8.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤4包括：

4.1)对每一句测试语音生成lattice网络结构，扫描每一帧空白的后验概率，当其后验 低于预设值时判定该帧为一个尖峰，将一句话中所有尖峰后，将时间上连续的尖峰合并为 一个尖峰，对于每个尖峰，选择出这一帧上后验概率超过预设值的音素以组成lattice网络 结构中的一列；在构造出所有列后，两两连接lattice中相邻两列中的每个节点音素，得到 所需lattice网络结构；

4.2)在生成的lattice网络结构上执行搜索与定制，找到与目标音素序列最相似的音 素序列；

4.3)计算找到的音素序列的观察概率与相似程度的乘积，并与设置的阈值做比较，当 大于阈值，则判断测试语音包含唤醒词，否则不唤醒。

9.一种实现权利要求1～8中任一所述方法的可定制语音唤醒系统，其特征在于，包括： 声学特征提取模块、记忆网络模块、分类模型模块、唤醒词搜索模块、决策模块和阈值估计 模块，其中：声学特征提取模块与记忆网络模块相连并输出待测语音的声学特征信息，记忆 网络模块与分类模型模块相连并传输音素后验信息，分类模型模块与唤醒词搜索模块相连 并传输最大相似音素序列，唤醒词搜索模块与决策模块相连并根据收到的唤醒词输出待测 语音可能解，决策模块根据来自阈值估计模块的判断阈值以及待测语音可能解得出判断结 果；

所述的唤醒词搜索模块采用基于连接时序分类模型的Lattice网络结构，其中包含了 所有可能的语音识别结果及其概率。

P (T/H) for the target phoneme sequence similarity between the t and h phoneme sequences,

The target phoneme sequences and hypothetical sequence of product of the probabilities of each edit operation as an indicator of the phoneme sequence similarity, MED (T,H) representing the target phoneme sequences t and h phoneme sequences of a minimum number of edits, p (op $_l$ | R=T,E=H) represents the phoneme sequences reference r to t guess phoneme sequences e h, series e and r l between editing operation op $_l$ probability.

3. According to right requirements 2 by said of can custom voice wake method, its features is, by said of edit operation is refers to: Insert, and delete, and replaced operation, its probability, that p (Insert (e $_l$)), and p (delete (r $_l$)) and p (r $_l$ /e $_l$) directly by prior knowledge obtained, r $_l$ And e $_l$ taken from reference sequence of phonemes respectively r and guess the phoneme sequence e.

4. According to claim 2 customizable voice awakens, which is, as described in step 1, including:

1.1) determines the structure of the model: a predictive model of the computing power of the device depending on the application scenario complexity and quantity ceiling set by first, and then under this limit specifies the network structure;

1.2) memory to initialize the length: using initialization parameters or model of transferring method initialize.

5. According to claim 4 customizable voice awakens, which is characterized, the length of initialization of memory, using cross entropy criterion when training the length of a standard acoustic memory and initialized by copying parameters model.

6. According to claim 2 customizable voice awakens, which is, as described in step 2, including:

2.1) the training data extraction fbank, namely Filter LFAT acoustic characteristics of the Bank;

2.2) model is accomplished using stochastic gradient descent training training parameter based on model structure and the training data set size.

7. According to claim 2 customizable voice awakens, which is, as described in step 3, including:

3.1) to extract test data fbank acoustic features, the feature extraction process requires consistent with the training data extraction;

3.2) extraction characteristics of frame-by-frame to enter training to get the model directly compute the posterior probability of each frame.

8. According to claim 2 customizable voice awakens, which is, as described in step 4 include:

4.1) on each sentence test voice generated lattice network structure, scan each frame blank of Hou inspection probability, dang subsequently inspection below preset value Shi judge the frame for a pointed peak, will word in the all

pointed peak Hou, will time Shang continuous of pointed peak merged for a pointed peak, for each pointed peak, select out this a frame Shang Hou inspection probability over preset value of sound are renowned for their composition lattice network structure in the of a column; in structure out all column Hou, 22 connecting two adjacent columns of each node in the lattice phonemes, given the lattice network structure;

4.2) generated by performing a search on a lattice network structure and customization, the target phoneme sequences are found the most similar sequences of phonemes;

4.3) calculation found observing the product of the probability and the degree of similarity of the phoneme sequence, and compare with the set threshold value, when greater than a threshold value, then test with wake Word speech, you do not wake up.

9. a achieved right requirements 1~8 in the any a by said method of can custom voice wake system, its features is, including: acoustic features extraction module, and memory network module, and classification model module, and wake word search module, and decision module and threshold value estimated module, which: acoustic features extraction module and memory network module connected and output stay measuring voice of acoustic features information, memory network module and classification model module connected and transmission phonemes Hou inspection information, Classification model of module and wake up word search module and transfer the most similar

phoneme sequences, wake up word search module connected with decision-making module and select the wake the word output sound to be measured according to possible solutions, decision-making module based on threshold estimation module to determine thresholds and sound judgment findings of possible solutions to be tested;

The wake of Word search module based on connection time series classification model of Lattice structure, which contains all possible speech recognition results and probabilities.

Description:

可定制语音唤醒方法及系统

技术领域

本发明涉及的是一种计算机辅助控制领域的技术，具体是一种基于长短时记忆网 络

(LSTM)和连接时序分类模型(CTC)的可定制语音唤醒方法及系统。

背景技术

近年来，随着信息技术的发展，包括语音在内多媒体信息技术越来越成为研究的 关注热点。语言唤醒技术是语音识别中的一个重要领域，并被广泛应用于语音命令控制系 统中。一个可定制的语音唤醒(Voice Wake-up)系统的任务是从一段连续的语音中自动发 现并定位一些事先指定的命令词(唤醒词)。可定制的特性体现在唤醒词检测模型不依赖于 用户指定的唤醒词，从而实现无需修改模型即可方便更改用户唤醒词。相反，非可定制的唤 醒技术与指定的唤醒词有关，唤醒词固定，无法轻易变更唤醒词。语音唤醒与连续语音识别 关系密切，但语音唤醒技术并不要求将完整的语音句子全部识别出来，而是仅关心用户指 定的关键信息。因此，唤醒技术降低了对识别系统的要求。与传统的文本态文档比较，语音 数据作为一种对声音的编码形式，使得语音信息成为一种对于计算机来说更加难以直接检 索和提取有效信息的数据形态。此外由于多种潜在的因素(如背景噪声、说话人口音等)，也 使得开发出一套有效的语音唤醒系统变得更加复杂和困难。主要的语音唤醒技术包括早期 的动态时间规整法，如今的基于隐马尔科夫模型的方法，以及基于深度学习的方法。

发明内容

本发明针对现有技术中唤醒词无法定制且依赖预设语言模型的不足，提出一种可 定制语音唤醒方法及系统，利用CTC模型输出后验概率稀疏的特点进行高效搜索，从而完成 对唤醒词置信度计算的技术。本发明一方面可以

Customizable voice wake method and system

technology the invention relates to the field of computer-aided control technology, which is based on the short-term memory network

(LSTM) time series classification model and connection (CTC) and customizable voice wake methods and systems. background technology in recent years, with the development of information technologies, multimedia and information technology, including voice more and more become the focus of study. Language wake on LAN technology is an important area of speech recognition, and is widely used in voice command-and-control systems. A customizable wake voice (Voice Wake LFAT up) system is a continuous voice commands that automatically find and locate some pre-specified in words (wake up). Customizable characteristics reflected in the wake of the word wake model

得到较高的唤醒性能(准确率、召回率)，高 准确率，低误唤醒，另一方面对应用系统的计算资源消耗相对较少。

本发明是通过以下技术方案实现的：

本发明涉及一种可定制语音唤醒方法，包括以下步骤：

步骤1)使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进 行建模；

步骤2)对模型进行训练：将事先收集并标注好的音频数据首先要经过传统的信号 处理方法做预处理并提取可供模型训练的声学特征，模型将特征数据作为输入，将标注的 音素信息作为输出，在海量数据下通过深度学习的方法完成模型各参数量的训练，得到可 使用的长短时记忆网络和连接时序分类模型；

步骤3)采用训练后模型进行测试：对测试语音做同样的预处理和特征提取，并输 入模型，模型将输出每一帧所有建模单元，即音素可能出现的后验概率；

步骤4)唤醒词搜索：通过在生成的Lattice网络结构上搜索与定制的唤醒词最相 似的可能音素序列来作为判断依据。

所述的最相似的可能音素序列H_{max}，通过判断音素序列在语音中的出现概率，即出 现概率越高且越与唤醒词音素序列相似的序列越可能是唤醒词的方式得到，具体为：

$$H_{max} = \underset{H}{\operatorname{argmax}} P(H|T) = \underset{H}{\operatorname{argmax}} \frac{P(T|H)P(H)}{P(T)} = \underset{H}{\operatorname{argmax}} P(T|H)P(H)$$

，

其中：P(T)为目标音素序列 T，$T=\{t_1,t_2,\cdots,t_n\}$的观察概率，t_i表示T中的第i个音素，i为1～n；

目标音素序列T位于CTC音素网格结构中的所有音素序列的概率为：

$P(T|L_H) \propto P(L_H|T)P(T) \approx P(H_{max}|T)P(T)$，其中：$L_H$表示CTC音素网格结构中的所有 音素序列，而可能音素序列H_{max}即为T已知时L_H中的最高概率。

P(H)为音素序列H，$H=\{n_{ij1},n_{(i+1)j2},\cdots,n_{(i+m-1)jm}\}$的观察概率，$n_{ij}$为lattice网格结构中第i栏第j列的音素，音素序列H也可以表示为$H=\{h_1,h_2,\cdots,h_m\}$，其中：$h_k = n_{(i+k-1)jk}$；

P(H)通过unigram假设，即累积乘音素序列中每个音素的后验概率获得，具体为：

$$P(H) = \prod_{i=1}^{m} P(h_i|H_{1:i-1}) \approx \prod_{i=1}^{m} P(h_i)$$

；

does not rely on the user to specify words, which do not need to modify the model to easily change the user wake up Word. In contrast, non-customizable wake wake technology specified words related to wake the word fixed wake cannot easily change the word. Voice awakens and closely related to the continuous speech recognition, voice wake on LAN technology does not require complete voice sentences all identified, but only concerned with user-specified key information. Therefore, the wake on LAN technology reduces the requirements for recognition system. Compared with the conventional text mode documents, voice data as a form of coding of sound makes voice messages became more difficult to direct a computer to retrieve and extract data information form. In addition various potential factors (such as background noise, the speaker's accent), and develop an effective voice to wake up systems become more complex and difficult. Main voice of wake on LAN technology, including an early method of dynamic time warping, now based on hidden Markov model methods, as well as the deep learning method. contents of invention wake up word in the existing technology of the invention cannot be customized and depends on the preset language model is insufficient, presents a customizable voice wake method and system using CTC model outputs a posteriori sparse efficient search features to complete technology to wake the word confidence measure. The invention can obtain a higher wake performance (precision and recall), high accuracy and low false alarm wake up, on the application of the system of consumption of

P(T/H)为目标音素序列T和音素序列H之间的相 似程度，

$$P(T|H) \triangleq \prod_{i=1}^{MED(T,H)} P(op_i|R=T,E=H)$$

，

即目标音素序列和假设序列之间的每个 编辑操作的概率之积作为衡量音素序列相似性的指标，MED(T,H)表示目标音素序列T和音 素序列H的最少编辑操作次数，$P(op_i|R=T，E=H)$表示当参考音素序列R为T时猜测音素序列E为H时，序列E和R之间的第i次编辑操作op_i的概率。

所述的编辑操作是指：插入、删除、替换操作，其概率，即P(insert(e_i))、P(delete (r_i))和P(r_i/e_i)直接由先验知识得出，r_i和e_i分别取自参考音素序列R和猜测音素序列E。

本发明涉及一种实现上述方法的可定制语音唤醒系统，包括：声学特征提取模块、 记忆网络模块、分类模型模块、唤醒词搜索模块、决策模块和阈值估计模块，其中：声学特征 提取模块与记忆网络模块相连并输出待测语音的声学特征信息，记忆网络模块与分类模型 模块相连并传输音素后验信息，分类模型模块与唤醒词搜索模块相连并传输最大相似音素 序列，唤醒词搜索模块与决策模块相连并根据收到的唤醒词输出待测语音可能解，决策模 块根据来自阈值估计模块的判断阈值以及待测语音可能解得出判断结果。

所述的唤醒词搜索模块采用基于连接时序分类(Connectionist Temporal Classification,CTC)模型的Lattice网络结构，其中包含了所有可能的语音识别结果及其 概率。

技术效果

本发明相比较传统方法主要有以下不同：

系统构建各环节	传统方法	本发明方法	优点比较
声学建模 Acoustic Modeling	隐马尔科夫模型	连接时序分类模型	更准确的声学建模
置信度打分 Confidence scoring	声学解码	基于编辑距离的动态搜索	模型搜索计算量和次数减少

附图说明

图1为本发明系统结构示意图；

图2为神经网络结构示意图；

图中给出了不同手机对应的后验概率，下半部分的网格中的实线表示潜在路径， 虚线表示所有有效连接。

具体实施方式

本实施例包括：声学特征提取模块、记忆网络模块、分类模型模块、唤醒词搜索模 块、决策模块和阈值估计模块，其中：声学特征提取模块与记忆

computing resources is relatively small. the invention is realized through the following technology: the invention relates to a customizable voice wake method, comprising the following steps: step 1) use short-term memory network and connection time series classification model for the modeling of phoneme information voice messages; steps 2) on model for training: will prior collection and mark good of audio data first to after traditional of signal processing method do pretreatment and extraction can for model training of acoustic features, model will features data as entered, will mark of phonemes information as output, in mass data Xia through depth learning of method completed model the parameter number of training, get can using of length Shi memory network and connection timing classification model; step 3) training model to test: to test voice doing the same preprocessing and feature extracting and enter the model, the model output each frame of all model units, that phoneme posterior probabilities that may occur; step 4) wake the word search: Lattice network structure generated by searches and customized wake the most similar word of possible phoneme sequences as a judge. the closest possible phoneme sequences described h_{Max}, by determining the probability in a sequence of phonemes in speech, the higher the probability of occurrence and more and wake the word phoneme sequences more similar sequences may be wake the word manner, in particular:

[IMAGE]

: P (t) is the target phoneme sequence T,T={t_1, t_2, ���, t

Petitioner Exhibit 1004, Page 630 of 1052

网络模块相连并输出待测语 音的声学特征信息，记忆网络模块与分类模型模块相连并传输音素后验信息，分类模型模 块与唤醒词搜索模块相连并传输最大相似音素序列，唤醒词搜索模块与决策模块相连并根 据收到的唤醒词输出待测语音可能解，决策模块根据来自阈值估计模块的判断阈值以及待 测语音可能解得出判断结果。

本实施例涉及上述系统的语音唤醒方法，具体包括以下步骤：

步骤1)使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进 行建模，具体步骤包括：

1.1)确定模型结构：根据应用场景设备的计算能力预估模型复杂度，由于计算复 杂度与模型的参数量正相关，因此首先要设定参数量的上限，例如不超过5.5M大小。然后在 此限定下指定网络结构，例如使用3隐层网络，每次256个节点，投影为96个节点。

1.2)初始化长短时记忆网络：使用随机初始化的参数，也可以通过model transferring方式初始化，由于随机初始化会造成之后CTC训练困难的问题，这里推荐使用 cross entropy准则预训练一个标准的声学长短时记忆网络，然后通过复制参数的方式初 始化模型。

步骤2)对模型进行训练：将事先收集并标注好的音频数据首先要经过传统的信号 处理方法做预处理并提取可供模型训练的声学特征，模型将特征数据作为输入，将标注的 音素信息作为输出，在海量数据下通过深度学习的方法完成模型各参数量的训练，得到可 使用的长短时记忆网络和连接时序分类模型，具体步骤包括：

2.1)对训练数据提取fbank，即Filter-bank声学特征。

2.2)使用随机梯度下降方式完成模型的训练，训练参数需根据模型结构以及训练 数据大小设定，例如对于上述例子中的模型，可采用0.00001的学习率，动量值为0.9，batch 大小256等等。

步骤3)采用训练后模型进行测试：对测试语音做同样的预处理和特征提取，并输 入模型，模型将输出每一帧所有建模单元可能出现的后验概率，具体步骤包括：

3.1)对测试数据提取fbank声学特征，特征提取过程要求与训练数据特征提取一 致。

3.2)将提取的特征逐帧输入训练得到的模型，直接计算出每一帧的后验概率。

步骤4)唤醒词搜索：通过在生成的lattice网络结构上搜索与定制的唤醒词最相 似的可能音素序列来作为判断依据，具体步骤包括：

4.1)对每一句测试语音生成lattice网络结构，扫描每一帧"空白"的后验概率，当 其后验低于预设值，例如0.8，那么认为这帧为一个尖峰。在找出一句话的所有尖峰后，将时 间上连续的尖峰合并为一个尖峰，对于每个尖峰，选择出这一帧上后验概率较大的音素，例 如后验概率大于0.005的音素，组成lattice网络结构中的一列；在构造出所有列后，两两连 接lattice中相邻两列中的每个节点音素，得到所需lattice网络结构。

4.2)根据前面提到的搜索算法公式，在生成的lattice网络结构上执行搜索算法， 找到与目标音素序列最相似的音素序列。

$_{n}$ } the observed probability, t $_{l}$ t l in phonemes, l 1~n; t target phoneme sequences in the CTC phoneme lattice structure probabilities of all sequences of phonemes are:

P(T| L_H)?P(L_H| T)P(T)$\hat{}$P(H_{max}| T) p (t), of which: l $_{h}$ all phoneme sequences in the CTC phoneme lattice structure, possible phoneme sequences h $_{Max}$ is known that t l $_{h}$, the highest probability. p (h) for phonemes sequence H,H={n $_{ij1}$, n $_{(i\ 1)\ j2}$, $\hat{}$ $\hat{}$ $\hat{}$, n $_{(i\ m\text{-}1)\ JM}$ } of observation probability, n $_{lJ}$ for lattice network grid structure in the subsection l bar subsection j column of phonemes, phonemes sequence h also can said for H={h $_{1}$, h $_{2}$, $\hat{}$ $\hat{}$ $\hat{}$, H $_{m}$ } where: h $_{k}$ = n $_{(i\ k\text{-}1)\ JK}$; p (h) through unigram assumptions, cumulative phoneme sequences, each phoneme posterior probabilities obtained, in particular to:

[IMAGE]

P (T/H) for the target phoneme sequence similarity between the t and h phoneme sequences,

The target phoneme sequences and hypothetical sequence of product of the probabilities of each edit operation as an indicator of the phoneme sequence similarity, MED (T,H) representing the target phoneme sequences t and h phoneme sequences of a minimum number of edits, p (op $_{l}$ | R=T,E=H) represents the phoneme sequences reference r to t guess phoneme sequences e h, series e and r l between editing operation op $_{l}$ probability. by said of edit operation is refers to: Insert, and delete, and replaced

4.3)计算找到的音素序列的观察概率与相似程度的乘积，并与设置的阈值做比 较，若大于阈值，则判断测试语音包含唤醒词，否则不唤醒。

本实施例与现有技术的对比结果如下表：

	EER	FOM	Parameters
GMM-HMM (40mix)	6.4	71.8	5.4M
DNN-HMM (4x512)	5.1	75.5	2.0M
DNN-HMM (7x512)	5.1	76.0	2.8M
LSTM-CTC KWS	3.6	85.2	813K

其中LSTM-CTC KWS为本发明提出的方法。表格中展示了本方法和目前处理可定制 唤醒词的主流方法HMM的对比结果，性能指标EER为平均相等错误率，FOM是错误唤醒在0至 10范围内唤醒率的平均值，EER越小越好，FOM越大越好。同时表格也说明了每个模型的参数 量。实验中使用标准的WSJ0数据集，使用了50个唤醒词做测试。可以看到，本发明提出的方 法明显优于传统的GMM-HMM、DNN-HMM方法，且参数量更少。

综上所述，与现有技术相比本发明基于唤醒词置信度与阈值的比较以判别是否唤 醒。由于可定制唤醒系统不限定唤醒词，对于不同唤醒词所使用的阈值也很难统一，所以这 里提出了一种针对不同唤醒词自动估计阈值的方法，一定程度上解决阈值难以统一的问 题，从而提升系统唤醒的准确性；此外，本发明基于CTC的lattice网格规模更小，同时尽可 能保留了有用信息。在此基础上提出的搜索算法，利用了音素序列的观察概率和与目标序列相似程度两方面信息，通过动态规划实现高效搜索唤醒词的目标，算法的时间和空间复 杂度较低，但准确性很高。

上述具体实施可由本领域技术人员在不背离本发明原理和宗旨的前提下以不同 的方式对其进行局部调整，本发明的保护范围以权利要求书为准且不由上述具体实施所 限，在其范围内的各个实现方案均受本发明之约束。

operation, its probability, that p (Insert (e_i)), and p (delete (r_i)) and p (r_i/e_i) directly by prior knowledge obtained, r_i and e_i respectively from reference phonemes sequence r and guess phonemes sequence e. this invention involved a achieved above method of can custom voice wake system, including: acoustic features extraction module, and memory network module, and classification model module, and wake word search module, and decision module and threshold value estimated module, which: acoustic features extraction module and memory network module connected and output stay measuring voice of acoustic features information, memory network module and classification model module connected and transmission phonemes Hou inspection information, classification model module and wake word search module connected and transmission maximum similar phonemes sequence, Wake up word search module connected with decision-making module and select the wake the word output sound to be measured according to possible solutions, decision-making module based on threshold estimation module to determine thresholds and results judged the voice solution to be measured. the wake of Word search module based on connection time series classification (Connectionist Temporal Classification,CTC) model of the Lattice network structure, which contains all possible speech recognition results and probabilities. technical results compared to the traditional method of the invention has the following differences:

[IMAGE]

Description

Figure 1 is a sketch of the invention system structure;

Figure 2 is a neural network structure diagram; corresponds to the figure given in the different cell phones a posteriori, the grid in the lower part of the solid line represents a potential path, dotted line represents all active connections. the specific way of implementing this implementation cases including: acoustic features extraction module, and memory network module, and classification model module, and wake word search module, and decision module and threshold value estimated module, which: acoustic features extraction module and memory network module connected and output stay measuring voice of acoustic features information, memory network module and classification model module connected and transmission phonemes Hou inspection information, classification model module and wake word search module connected and transmission maximum similar phonemes sequence, wake word search module and decision module connected and root According to awaken received word output voice possible solutions to be tested, decision-making module based on threshold estimation module to determine thresholds and results judged the voice solution to be measured. this voice-wake method embodiment relates to the system, including the following steps: step 1) use short-term memory network and connection time series classification model for the modeling of phoneme information voice messages, specific steps, including:

1.1) determines the structure of the model: depending on the application scenario the computing power of the device model complexity, due to the complexity and the parameters of the model are related, so the first thing to set the upper limit of the number of, for example, does not exceed 5.5M in size. Then under this limit specifies the network structure, for example, use 3 hidden networks, each 256 nodes, projecting to 96 nodes.

1.2) memory to initialize the length: uses random initialization parameter, or by model transferring initialized, random CTC training after the initial cause difficult problems, it is recommended using the cross entropy rule when training the length of a standard acoustic memory and initialized by copying parameters model. steps 2) on model for training: will prior collection and mark good of audio data first to after traditional of signal processing method do pretreatment and extraction can for model training of acoustic features, model will features data as entered, will mark of phonemes information as output, in mass data Xia through depth learning of method completed model the parameter number of training, get can using of length Shi memory network and connection timing classification model, specific steps including:

2.1) the training data extraction fbank, namely Filter LFAT Bank features.

2.2) model of training is accomplished using stochastic gradient descent training parameter based on model structure and the training data set, for example, for the model in the above example, use learning rates of 0.00001,

momentum values for 0.9,batch size 256, and so on.

Step 3) training model to test: to test voice doing the same preprocessing and feature extracting and enter the model, the model output each frame all the posterior probability of the model element may occur, specific steps, including:

3.1) to extract test data fbank acoustic features, the feature extraction process requires consistent with the training data extraction.

3.2) extraction characteristics of frame-by-frame to enter training to get the model directly compute the posterior probability of each frame. step 4) wake the word search: on a lattice network structure generated by search and customized wake most similar sequence of phonemes to words used as a basis, specific steps, including:

4.1) to test voice generate a lattice network structure for each sentence, scanning every frame of "blank" a posteriori, when testing is lower than the preset value, such as 0.8, so think this frame for a spike. In find a sentence words of all pointed peak Hou, will time Shang continuous of pointed peak merged for a pointed peak, for each pointed peak, select out this a frame Shang Hou inspection probability larger of phonemes, for example Hou inspection probability is greater than 0.005 of phonemes, composition lattice network structure in the of a column; in structure out all column Hou, 22 connection lattice in the adjacent two column in the of each node phonemes, get by needed lattice network structure.

4.2) according to the formula mentioned search algorithm and search algorithm on a generate a lattice network structure, the target phoneme sequences are found the most similar sequences of phonemes.

4.3) calculation found observing the product of the probability and the degree of similarity of the phoneme sequence, and compare with the set threshold value, if it is greater than the threshold, then test with wake Word speech, you do not wake up. these cases compared with existing technology results in the following table:

[IMAGE]

LSTM LFAT CTC KWS for the method of the present invention. Table shows the method and customizable wake up Word mainstream methods comparison of HMM, EER for the average equal error rate performance indicators, FOM is wrong to wake up wake up rate in the range of 0 to 10, on average, EER, the smaller the better, FOM the bigger the better. While the table also shows that the amount of parameters for each model. Standard WSJ0 data set used in the experiment, using wake 50 words for testing. As you can see, the invention proposed method outperforms the GMM LFAT HMM, DNN LFAT HMM method and parameters less. to sum up, compared with the prior art invention based on wake the word confidence compared with threshold to distinguish whether the wake. Due to can custom wake system not qualified wake Word, for different wake Word by using of threshold value also hard unified, so here proposed has a for different wake Word automatically estimated threshold value of method, must degree Shang solution

threshold value to unified of problem, to upgrade system wake of accuracy; also, this invention based on CTC of lattice network grid scale more small, while as retained has useful information. Based on search algorithms using phoneme sequence observed probabilities and the target sequence similarity information in two ways, effective search wake up Word's goals through dynamic programming, algorithms and complexity at a lower, but accuracy is very high. this implementation can be invented by the technicians in this field do not deviate from the principle and purpose of these rules in a different manner to its local adjustment, the scope of protection of the invention in the claims and are not limited by the implementation of the specific, in the context of its implementation are subject to the constraints of the present invention.

Citing: CN103021409 A ; CN103095911 A ; CN103956164 A ; CN104538031 A ; CN104620314 A ; CN104732978 A ; CN105551483 A ; US2015340034 AA ; US9263036 BA ;

Cited by: CN107358951 A ; WO18086033 A1 ; WO18153200 A1 ; WO18161763 A1 ; WO19001428 A1 ;

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106098059 A
(43)申请公布日 2016.11.09

(21)申请号 201610462976.X

(22)申请日 2016.06.23

(71)申请人 上海交通大学
地址 200240 上海市闵行区东川路800号

(72)发明人 俞凯 钱彦旻 庄毅萌 陈哲怀
常烜恺

(74)专利代理机构 上海交达专利事务所 31201
代理人 王毓理 王锡麟

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2013.01)
G10L 15/22(2006.01)

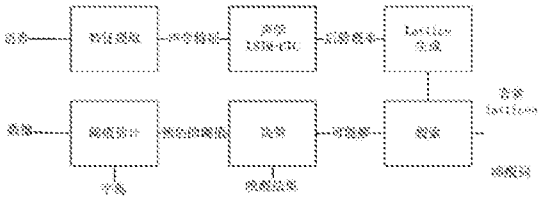
权利要求书2页 说明书4页 附图1页

(54)发明名称

可定制语音唤醒方法及系统

(57)摘要

一种可定制语音唤醒方法及系统，通过使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进行建模并对模型进行训练，采用训练后模型进行测试并在生成的Lattice网络结构上搜索与定制的唤醒词最相似的可能音素序列来作为判断依据。本发明利用CTC模型输出后验概率稀疏的特点进行高效搜索，从而完成对唤醒词置信度计算的技术。本发明一方面可以得到较高的唤醒性，即高准确率，低误唤醒，另一方面对应用系统的计算资源消耗相对较少。



CN 106098059 A

1.一种可定制语音唤醒方法，其特征在于，包括以下步骤：

步骤1)使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进行建模；

步骤2)对模型进行训练：将事先收集并标注好的音频数据首先要经过传统的信号处理方法做预处理并提取可供模型训练的声学特征，模型将特征数据作为输入，将标注的音素信息作为输出，在海量数据下通过深度学习的方法完成模型各参数量的训练，得到可使用的长短时记忆网络和连接时序分类模型；

步骤3)采用训练后模型进行测试：对测试语音做同样的预处理和特征提取，并输入模型，模型将输出每一帧所有建模单元，即音素可能出现的后验概率；

步骤4)唤醒词搜索：通过在生成的Lattice网络结构上搜索与定制的唤醒词最相似的可能音素序列来作为判断依据。

2.根据权利要求1所述的可定制语音唤醒方法，其特征是，所述的最相似的可能音素序列H_{max}，通过判断音素序列在语音中的出现概率，即出现概率越高且越与唤醒词音素序列相似的序列越可能是唤醒词的方式得到，具体为：

$$H_{max} = \arg\max_{H} P(H|T) = \arg\max_{H} \frac{P(T|H)P(H)}{P(T)} = \arg\max_{H} P(T|H)P(H)$$

其中：P(T)为目标音素序列T，T＝$\{t_1, t_2, \cdots, t_n\}$的观察概率，$t_i$表示T中的第i个音素，i为1～n；

目标音素序列T位于CTC音素网格结构中的所有音素序列的概率为：

$P(T|L_H) \propto P(L_H|T)P(T) \approx P(H_{max}|T)P(T)$，其中：$L_H$表示CTC音素网格结构中的所有音素序列，而可能音素序列H_{max}即为T已知时L_H中的最高概率；

P(H)为音素序列H，H＝$\{n_{ij1}, n_{(i+1)j2}, \cdots, n_{(i+m-1)jm}\}$的观察概率，$n_{ij}$为lattice网格结构中第i栏第j列的音素，音素序列H也可以表示为H＝$\{h_1, h_2, \cdots, h_m\}$，其中：$h_k$＝$n_{(i+k-1)jk}$；

P(H)通过unigram假设，即累积乘音素序列中每个音素的后验概率获得，具体为：$P(H) = \prod_{i=1}^{m} P(h_i|H_{1:i-1}) \approx \prod_{i=1}^{m} P(h_i)$；P(T/H)为目标音素序列T和音素序列H之间的相似程度，$P(T|H) \triangleq \prod_{i=1}^{MED(T,H)} P(op_i|R=T, E=H)$，即目标音素序列和假设序列之间的每个编辑操作的概率之积作为衡量音素序列相似性的指标，MED(T,H)表示目标音素序列T和音素序列H的最少编辑操作次数，$P(op_i|R=T, E=H)$表示当参考音素序列R为T时猜测音素序列E为H时，序列E和R之间的第i次编辑操作op_i的概率。

3.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的编辑操作是指：插入、删除、替换操作，其概率，即$P(insert(e_i))$、$P(delete(r_i))$和$P(r_i/e_i)$直接由先验知识得出，r_i和e_i分别取自参考音素序列R和猜测音素序列E。

4.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤1包括：

1.1)确定模型结构：根据应用场景设备的计算能力预估模型复杂度，首先设定参数量

Petitioner Exhibit 1004, Page 639 of 1052

的上限，然后在此限定下指定网络结构；

1.2)初始化长短时记忆网络：使用随机初始化的参数或model transferring方式初始化。

5.根据权利要求4所述的可定制语音唤醒方法，其特征是，所述的初始化长短时记忆网络，使用cross entropy准则预训练一个标准的声学长短时记忆网络，然后通过复制参数的方式初始化模型。

6.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤2包括：

2.1)对训练数据提取fbank，即Filter-bank声学特征；

2.2)使用随机梯度下降方式完成模型的训练，训练参数需根据模型结构以及训练数据大小设定。

7.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤3包括：

3.1)对测试数据提取fbank声学特征，特征提取过程要求与训练数据特征提取一致；

3.2)将提取的特征逐帧输入训练得到的模型，直接计算出每一帧的后验概率。

8.根据权利要求2所述的可定制语音唤醒方法，其特征是，所述的步骤4包括：

4.1)对每一句测试语音生成lattice网络结构，扫描每一帧空白的后验概率，当其后验低于预设值时判定该帧为一个尖峰，将一句话中所有尖峰后，将时间上连续的尖峰合并为一个尖峰，对于每个尖峰，选择出这一帧上后验概率超过预设值的音素以组成lattice网络结构中的一列；在构造出所有列后，两两连接lattice中相邻两列中的每个节点音素，得到所需lattice网络结构；

4.2)在生成的lattice网络结构上执行搜索与定制，找到与目标音素序列最相似的音素序列；

4.3)计算找到的音素序列的观察概率与相似程度的乘积，并与设置的阈值做比较，当大于阈值，则判断测试语音包含唤醒词，否则不唤醒。

9.一种实现权利要求1～8中任一所述方法的可定制语音唤醒系统，其特征在于，包括：声学特征提取模块、记忆网络模块、分类模型模块、唤醒词搜索模块、决策模块和阈值估计模块，其中：声学特征提取模块与记忆网络模块相连并输出待测语音的声学特征信息，记忆网络模块与分类模型模块相连并传输音素后验信息，分类模型模块与唤醒词搜索模块相连并传输最大相似音素序列，唤醒词搜索模块与决策模块相连并根据收到的唤醒词输出待测语音可能解，决策模块根据来自阈值估计模块的判断阈值以及待测语音可能解得出判断结果；

所述的唤醒词搜索模块采用基于连接时序分类模型的Lattice网络结构，其中包含了所有可能的语音识别结果及其概率。

Petitioner Exhibit 1004, Page 640 of 1052

可定制语音唤醒方法及系统

技术领域

[0001] 本发明涉及的是一种计算机辅助控制领域的技术，具体是一种基于长短时记忆网络

(LSTM)和连接时序分类模型(CTC)的可定制语音唤醒方法及系统。

背景技术

[0002] 近年来，随着信息技术的发展，包括语音在内多媒体信息技术越来越成为研究的关注热点。语言唤醒技术是语音识别中的一个重要领域，并被广泛应用于语音命令控制系统中。一个可定制的语音唤醒(Voice Wake-up)系统的任务是从一段连续的语音中自动发现并定位一些事先指定的命令词(唤醒词)。可定制的特性体现在唤醒词检测模型不依赖于用户指定的唤醒词，从而实现无需修改模型即可方便更改用户唤醒词。相反，非可定制的唤醒技术与指定的唤醒词有关，唤醒词固定，无法轻易变更唤醒词。语音唤醒与连续语音识别关系密切，但语音唤醒技术并不要求将完整的语音句子全部识别出来，而是仅关心用户指定的关键信息。因此，唤醒技术降低了对识别系统的要求。与传统的文本态文档比较，语音数据作为一种对声音的编码形式，使得语音信息成为一种对于计算机来说更加难以直接检索和提取有效信息的数据形态。此外由于多种潜在的因素(如背景噪声、说话人口音等)，也使得开发出一套有效的语音唤醒系统变得更加复杂和困难。主要的语音唤醒技术包括早期的动态时间规整法，如今的基于隐马尔科夫模型的方法，以及基于深度学习的方法。

发明内容

[0003] 本发明针对现有技术中唤醒词无法定制且依赖预设语言模型的不足，提出一种可定制语音唤醒方法及系统，利用CTC模型输出后验概率稀疏的特点进行高效搜索，从而完成对唤醒词置信度计算的技术。本发明一方面可以得到较高的唤醒性能(准确率、召回率)，高准确率，低误唤醒，另一方面对应用系统的计算资源消耗相对较少。

[0004] 本发明是通过以下技术方案实现的：

[0005] 本发明涉及一种可定制语音唤醒方法，包括以下步骤：

[0006] 步骤1)使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进行建模；

[0007] 步骤2)对模型进行训练：将事先收集并标注好的音频数据首先要经过传统的信号处理方法做预处理并提取可供模型训练的声学特征，模型将特征数据作为输入，将标注的音素信息作为输出，在海量数据下通过深度学习的方法完成模型各参数量的训练，得到可使用的长短时记忆网络和连接时序分类模型；

[0008] 步骤3)采用训练后模型进行测试：对测试语音做同样的预处理和特征提取，并输入模型，模型将输出每一帧所有建模单元，即音素可能出现的后验概率；

[0009] 步骤4)唤醒词搜索：通过在生成的Lattice网络结构上搜索与定制的唤醒词最相似的可能音素序列来作为判断依据。

Petitioner Exhibit 1004, Page 641 of 1052

[0010] 所述的最相似的可能音素序列H_{max}，通过判断音素序列在语音中的出现概率，即出现概率越高且越与唤醒词音素序列相似的序列越可能是唤醒词的方式得到，具体为：

[0011] $$H_{max} = \arg\max_H P(H|T) = \arg\max_H \frac{P(T|H)P(H)}{P(T)} = \arg\max_H P(T|H)P(H)$$ 其中：P(T)为目标音素序列T，$T=\{t_1, t_2, \cdots, t_n\}$的观察概率，$t_i$表示T中的第i个音素，i为1～n；

[0012] 目标音素序列T位于CTC音素网格结构中的所有音素序列的概率为：

[0013] $P(T|L_H) \propto P(L_H|T)P(T) \approx P(H_{max}|T)P(T)$，其中：$L_H$表示CTC音素网格结构中的所有音素序列，而可能音素序列H_{max}即为T已知时L_H中的最高概率。

[0014] P(H)为音素序列H，$H=\{n_{ij1}, n_{(i+1)j2}, \cdots, n_{(i+m-1)jm}\}$的观察概率，$n_{ij}$为lattice网格结构中第i栏第j列的音素，音素序列H也可以表示为$H=\{h_1, h_2, \cdots, h_m\}$，其中：$h_k = n_{(i+k-1)jk}$；

[0015] P(H)通过unigram假设，即累积乘音素序列中每个音素的后验概率获得，具体为：$P(H) = \prod_{i=1}^{m} P(h_i|H_{1:i-1}) \approx \prod_{i=1}^{m} P(h_i)$ P(T/H)为目标音素序列T和音素序列H之间的相似程度，$P(T|H) \triangleq \prod_{i=1}^{MED(T,H)} P(op_i|R=T, E=H)$ 即目标音素序列和假设序列之间的每个编辑操作的概率之积作为衡量音素序列相似性的指标，MED(T,H)表示目标音素序列T和音素序列H的最少编辑操作次数，$P(op_i|R=T, E=H)$表示当参考音素序列R为T时猜测音素序列E为H时，序列E和R之间的第i次编辑操作op_i的概率。

[0016] 所述的编辑操作是指：插入、删除、替换操作，其概率，即P(insert(e_i))、P(delete(r_i))和$P(r_i/e_i)$直接由先验知识得出，r_i和e_i分别取自参考音素序列R和猜测音素序列E。

[0017] 本发明涉及一种实现上述方法的可定制语音唤醒系统，包括：声学特征提取模块、记忆网络模块、分类模型模块、唤醒词搜索模块、决策模块和阈值估计模块，其中：声学特征提取模块与记忆网络模块相连并输出待测语音的声学特征信息，记忆网络模块与分类模型模块相连并传输音素后验信息，分类模型模块与唤醒词搜索模块相连并传输最大相似音素序列，唤醒词搜索模块与决策模块相连并根据收到的唤醒词输出待测语音可能解，决策模块根据来自阈值估计模块的判断阈值以及待测语音可能解得出判断结果。

[0018] 所述的唤醒词搜索模块采用基于连接时序分类(Connectionist Temporal Classification,CTC)模型的Lattice网络结构，其中包含了所有可能的语音识别结果及其概率。

技术效果

[0019] 本发明相比较传统方法主要有以下不同：

Petitioner Exhibit 1004, Page 642 of 1052

系统构建各环节	传统方法	本发明方法	优点比较
声学建模 Acoustic Modeling	隐马尔科夫模型	连接时序分类模型	更准确的声学建模
置信度打分 Confidence scoring	声学解码	基于编辑距离的动态搜索	模型搜索计算量和次数减少

附图说明

[0020] 图1为本发明系统结构示意图；

[0021] 图2为神经网络结构示意图；

[0022] 图中给出了不同手机对应的后验概率，下半部分的网格中的实线表示潜在路径，虚线表示所有有效连接。

具体实施方式

[0023] 本实施例包括：声学特征提取模块、记忆网络模块、分类模型模块、唤醒词搜索模块、决策模块和阈值估计模块，其中：声学特征提取模块与记忆网络模块相连并输出待测语音的声学特征信息，记忆网络模块与分类模型模块相连并传输音素后验信息，分类模型模块与唤醒词搜索模块相连并传输最大相似音素序列，唤醒词搜索模块与决策模块相连并根据收到的唤醒词输出待测语音可能解，决策模块根据来自阈值估计模块的判断阈值以及待测语音可能解得出判断结果。

[0024] 本实施例涉及上述系统的语音唤醒方法，具体包括以下步骤：

[0025] 步骤1)使用基于长短时记忆网络和连接时序分类模型对语音信息的音素信息进行建模，具体步骤包括：

[0026] 1.1)确定模型结构：根据应用场景设备的计算能力预估模型复杂度，由于计算复杂度与模型的参数量正相关，因此首先要设定参数量的上限，例如不超过5.5M大小。然后在此限定下指定网络结构，例如使用3隐层网络，每次256个节点，投影为96个节点。

[0027] 1.2)初始化长短时记忆网络：使用随机初始化的参数，也可以通过model transferring方式初始化，由于随机初始化会造成之后CTC训练困难的问题，这里推荐使用cross entropy准则预训练一个标准的声学长短时记忆网络，然后通过复制参数的方式初始化模型。

[0028] 步骤2)对模型进行训练：将事先收集并标注好的音频数据首先要经过传统的信号处理方法做预处理并提取可供模型训练的声学特征，模型将特征数据作为输入，将标注的音素信息作为输出，在海量数据下通过深度学习的方法完成模型各参数量的训练，得到可使用的长短时记忆网络和连接时序分类模型，具体步骤包括：

[0029] 2.1)对训练数据提取fbank，即Filter-bank声学特征。

[0030] 2.2)使用随机梯度下降方式完成模型的训练，训练参数需根据模型结构以及训练数据大小设定，例如对于上述例子中的模型，可采用0.00001的学习率，动量值为0.9，batch大小256等等。

[0031] 步骤3)采用训练后模型进行测试：对测试语音做同样的预处理和特征提取，并输

Petitioner Exhibit 1004, Page 643 of 1052

入模型，模型将输出每一帧所有建模单元可能出现的后验概率，具体步骤包括：

[0032] 3.1)对测试数据提取fbank声学特征，特征提取过程要求与训练数据特征提取一致。

[0033] 3.2)将提取的特征逐帧输入训练得到的模型，直接计算出每一帧的后验概率。

[0034] 步骤4)唤醒词搜索：通过在生成的lattice网络结构上搜索与定制的唤醒词最相似的可能音素序列来作为判断依据，具体步骤包括：

[0035] 4.1)对每一句测试语音生成lattice网络结构，扫描每一帧“空白”的后验概率，当其后验低于预设值，例如0.8，那么认为这帧为一个尖峰。在找出一句话的所有尖峰后，将时间上连续的尖峰合并为一个尖峰，对于每个尖峰，选择出这一帧上后验概率较大的音素，例如后验概率大于0.005的音素，组成lattice网络结构中的一列；在构造出所有列后，两两连接lattice中相邻两列中的每个节点音素，得到所需lattice网络结构。

[0036] 4.2)根据前面提到的搜索算法公式，在生成的lattice网络结构上执行搜索算法，找到与目标音素序列最相似的音素序列。

[0037] 4.3)计算找到的音素序列的观察概率与相似程度的乘积，并与设置的阈值做比较，若大于阈值，则判断测试语音包含唤醒词，否则不唤醒。

[0038] 本实施例与现有技术的对比结果如下表：

	EER	FOM	Parameters
GMM-HMM (40mix)	6.4	71.8	5.4M
DNN-HMM (4x512)	5.1	75.5	2.0M
DNN-HMM (7x512)	5.1	76.0	2.8M
LSTM-CTC KWS	3.6	85.2	813K

[0039] 其中LSTM-CTC KWS为本发明提出的方法。表格中展示了本方法和目前处理可定制唤醒词的主流方法HMM的对比结果，性能指标EER为平均相等错误率，FOM是错误唤醒在0至10范围内唤醒率的平均值，EER越小越好，FOM越大越好。同时表格也说明了每个模型的参数量。实验中使用标准的WSJ0数据集，使用了50个唤醒词做测试。可以看到，本发明提出的方法明显优于传统的GMM-HMM、DNN-HMM方法，且参数量更少。

[0040] 综上所述，与现有技术相比本发明基于唤醒词置信度与阈值的比较以判别是否唤醒。由于可定制唤醒系统不限定唤醒词，对于不同唤醒词所使用的阈值也很难统一，所以这里提出了一种针对不同唤醒词自动估计阈值的方法，一定程度上解决阈值难以统一的问题，从而提升系统唤醒的准确性；此外，本发明基于CTC的lattice网格规模更小，同时尽可能保留了有用信息。在此基础上提出的搜索算法，利用了音素序列的观察概率和与目标序列相似程度两方面信息，通过动态规划实现高效搜索唤醒词的目标，算法的时间和空间复杂度较低，但准确性很高。

[0041] 上述具体实施可由本领域技术人员在不背离本发明原理和宗旨的前提下以不同的方式对其进行局部调整，本发明的保护范围以权利要求书为准且不由上述具体实施所限，在其范围内的各个实现方案均受本发明之约束。

Petitioner Exhibit 1004, Page 644 of 1052

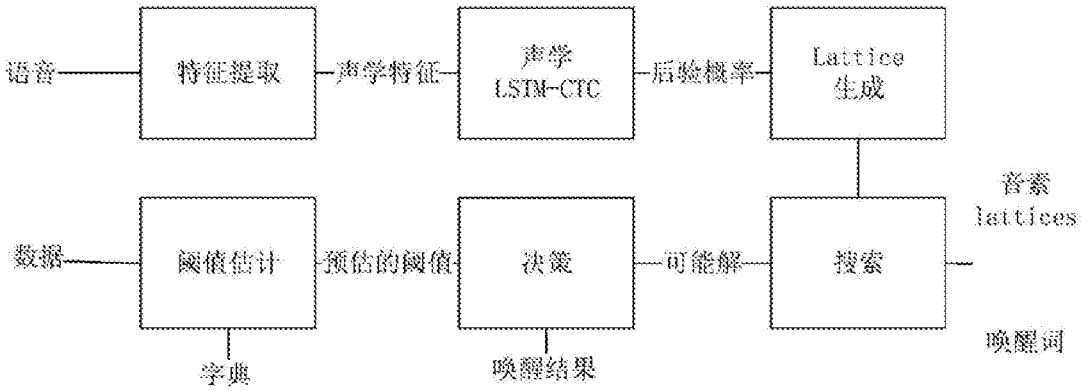


图1

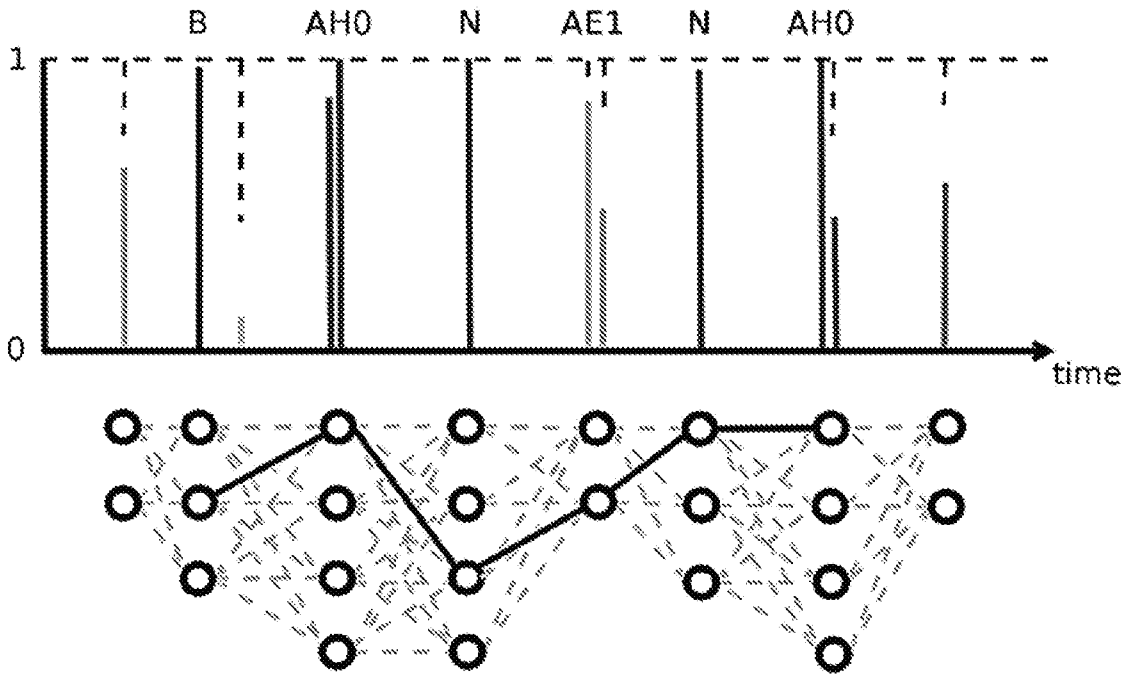


图2

Petitioner Exhibit 1004, Page 645 of 1052

Publication number: CN104575504 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20150429
Application number: CN201410840544
Application date: 20141224
Priority: CN201410840544201412 24CN201410840544;
Assignee: SHANGHAI NORMAL UNIV ;
Assignee[std]: UNIV SHANGHAI ;
Inventor[std]: YE HONG ; WEI HAORAN ; LONG YANHUA ;
International class[8]: G10L17/04; G10L15/26; H04N21/439; H04N21/441;
Family members: CN104575504 A

Title: METHOD FOR PERSONALIZED TELEVISION VOICE WAKE-UP BY VOICEPRINT AND VOICE IDENTIFICATION

Title: 采用声纹和语音识别进行个性化电视语音唤醒的方法

Abstract:

The invention discloses a method for personalized television voice wake-up by voiceprint and voice identification, particularly a method for performing identity confirmation on a television user through voiceprint identification and controlling a television to perform personalized voice wake-up through confirmed identity and a voice identification result of user voice, and relates to voiceprint identification and voice identification technologies. A composition system comprises a voice control system (1), an information storage unit (2) and a television main controller (3) which are connected through electric signals. The method has the characteristics of short training time, very high voiceprint and voice identification speed and high identification rate. Voiceprint and voice identification can be finished by only offline training and testing, identification results do not need to be sent to a cloud server, use is convenient, and the safety of family information is guaranteed. The method also can be applied to user-personalized automatic voice channel change of the television, can be transplanted to a common high-speed DSP (digital signal processor) or chip for operation, and can be widely applied to the related fields of smart homes.

Abstract:

本发明采用声纹和语音识别进行个性化电视语音唤醒的方法，具体指一种通过声纹识别进行电视机使用用户的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒，涉及声纹识别和语音识别技术。本发明构成系统包括语音控制系统(1)，信息存储单元(2)和电视机主控制器(3)以电信号方式连接组成。本发明具

The present invention uses a voiceprint and voice recognition personalize Television Voice wake method, specifically refers to a voiceprint identification performed by the TV using the user's identification, and identification by the user's voice and speech recognition results to control the TV voice wake personalize involving voiceprint recognition and speech recognition technology.

有训练时间短，声纹和语音识别速度都很快，且识别率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。还可应用于电视机的用户个性化自动语音换台，同时可被移植到通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

Constitute a system of the present invention includes a voice control system (1), an information storage unit (2) and TV host controller (3) connected to the electrical components. The present invention has the training time is short, voiceprint and voice recognition speed very quickly, and recognition rate characteristics. Voiceprint and voice recognition are just off-line training and testing can be completed without the need to send to the cloud server, easy to use and ensure the safety of family information. Users can also be used in televisions personalized automated voice for Taiwan, and can be ported to run on general-purpose high-speed processor or DSP chips, widely used in smart home related fields.

Claims:

1.一种采用声纹和语音识别进行个性化电视语音唤醒的方法，通 过采用声纹识别和语音识别系统，对用户的身份进行监测，实现个性 化电视语音唤醒，其特征在于，包括如下步骤：

A.个性化信息注册

采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别 出用户最喜爱的电视频道信息；

B.信息存储

将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电 视机信息存储单元；

C.语音唤醒

用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可 根据声纹识别和语音识别系统对唤醒词的声纹和语音识别结果从电 视机信息存储单元中调出用户的个性化信息，自动完成打开电视并自 动切换到用户最喜爱的电视频道。

2.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤A、步骤B，在用户作第一次 操作时完成，以后只需操作步骤C，即可完成其个性化语音电视的唤醒；

当用户对电视频道的喜好发生改变时，通过重新执行步骤A，进 行用户个性化信息注册，对信息存储单元中的电视频道信息进行更新， 此时无需进行声纹建模，仅需进行声纹匹配和语音识别。

1. One kind of speech recognition using the voiceprint and personalize the TV voice wake method, by using the voiceprint recognition and speech recognition systems, to monitor the user's identity, personalized television voice awakened, characterized by comprising the steps of:

A. Personalized Information Register

Using the voiceprint identification system to establish the user's voiceprint model, using voice recognition system identifies the user's favorite television channel information;

B. Information storage

The user voiceprint model and favorite TV channels and stores information related to the TV information storage unit;

C. Voice Wake Up

Users only need to tell the television voice collecting unit pre-set wake-up word, you can wake up the words of the voiceprint and voice recognition results from the television tune information storage unit based on the voice recognition and speech recognition systems the user's personalized information, auto-complete turn on the TV and automatically switch to the user's favorite TV channels.

2. Voiceprint as claimed in using voice recognition and speech personalized TV Wake method according to claim 1, characterized in that said step A, step B, the user for the first When once the

3.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法的系统，其特征在于，所述系统包括用于用户个性化信 息注册、唤醒词识别和声纹匹配的语音控制系统(1)，用于接收并存 储从语音控制系统(1)中传递来的已训练好的用户声纹模型和已识 别出的电视频道信息的信息存储单元(2)和接收从语音控制系统(1) 和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操 作的电视机主控制器(3)组成，以电信号方式连接，其中：

所述语音控制系统(1)由语音采集单元(11)，语音识别单元(12) 和声纹识别单元(13)组成，其中：

所述语音采集单元(11)用于完成用户语音信号的采集和处理；

所述语音识别单元(12)用于识别用户所说的电视频道信息和唤 醒词；

所述声纹识别单元(13)用于从用户语音中提取出代表其身份信 息的声纹信息进行声纹建模和声纹匹配。

4.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤A个性化信息注册，具体实现 步骤为：

A1.开启语音控制系统(1)，使其处于"用户注册"状态；

A2.若用户为首次进行个性化信息注册，执行该步骤，否则转向 步骤A6开始执行，用户对着电视机的语音采集单元(11)说话，说 话内容为任意文本，持续时间至少30秒；

A3.语音采集单元(11)采集输入的说话人语音信号，并对其进 行A/D转换，滤波处理；

A4.声纹识别单元(13)对经步骤A3处理后的语音信号作预处理， 包括加窗分帧、预加重、端点检测，提取声纹特征参数，然后调用信 息存储单元(2)中已有的声纹模型进行声纹匹配，若匹配成功，转 向步骤A7；反之，进入下一步操作；

A5.调用声纹识别单元(13)对从步骤A4中提取的特征参数进行 声纹模型的训练，将训练好的声纹模型存储到信息存储单元(2)；

A6.用户对着电视机的语音采集单元(11)说出自己最喜欢的电 视频道信息，如频道代码"3"代表湖南卫视，则说出数字"3"，转 向步骤A3重新开始；

operation is complete, just after the operation step C, and to complete the wake of its personalized voice TV;

When the user changes the TV channel preferences, by re-executing step A, user personalization information registration information storage unit to update the information television channel, this time without the need for a voiceprint model, only conduct voiceprint matching and speech recognition.

3. Voiceprint as claimed in using voice recognition and speech personalized TV system wake-up method according to claim 1, characterized in that the system comprises a user personalization information Sign up, wake up word recognition voiceprint matching voice control system (1) for receiving and storing transmitted from the voice control system (1) to have trained user voiceprint model and have identified television channel information information storage unit (2) and receiving control information transmitted from the voice control system (1) and the information storage unit, responsible for the implementation of personalized television voice Wake TV host controller operations (3), with an electrical signal connected, where:

The voice control system (1) by the voice acquisition unit (11), the voice recognition unit (12) voiceprint identification unit (13), wherein:

The voice acquisition unit (11) for the completion of the acquisition and processing user speech signal;

The voice recognition unit (12) for identifying a user of said television information and wake-up word; the voiceprint identification unit (13) is used to extract from a speech on behalf of its user identity information voiceprint information voiceprint model voiceprint match.

4. Claim and speech recognition using the voiceprint of claim 1, wherein the personalized television voice awakened, characterized in that said step A personalized information registration, concrete implementation steps to:

A1 open the voice control system (1), it is in "User Registration" state;.

A2. If the user is registered for the first time personalized information, perform this step,

A7.语音识别单元(12)对语音采集单元(11)采集到的语音进 行识别，将识别出的电视频道代码存储到信息存储单元(2)中并与 声纹匹配成功的声纹模型进行关联；

A8.退出个性化信息注册。

5.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤C语音唤醒，具体实现步骤为：

C1.当语音控制系统(1)处于"语音唤醒"开启状态时，用户对 着电视机语音采集单元(11)说出唤醒词；

C2.语音采集单元(11)采集输入的说话人语音信号，并对其进 行A/D转换，滤波处理；

C3.语音识别单元(12)对上述处理过的语音进行识别，将识别 出的内容与预先设定的唤醒词进行匹配，以判断采集到的输入语音信 号是否为电视机的语音唤醒指令，若匹配成功则将唤醒操作指令传递给电视机主控制器(3)，否则退出唤醒系统；

C4.声纹识别单元(13)将处理过的语音与信息存储单元(2)中 的声纹模型进行声纹匹配，若匹配成功，则将与用户声纹模型关联的 电视频道信息和唤醒操作指令输出给电视机主控制器(3)，否则退出 唤醒系统；

C5.电视机主控制器(3)在成功收到语音控制系统(1)传递来 的电视频道信息和唤醒操作指令后，执行电视开机操作并自动切换到 对应用户喜爱的电视频道上。

6.如权利要求4所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤A5中的声纹模型的训练方法 为总变化因子建模技术。

7.如权利要求6所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述总变化因子建模技术，给定一段用 户的语音，与说话人及信道相关的高斯混合模型GMM的均值超向量S 表示如下：

$$S=m+Tv+\varepsilon$$

其中，m为通用背景模型的均值超向量，T矩阵为总变化空间， ε为残差，v矢量是与当前目标说话人(用户)及信道相关的总变化因 子，它代表着当前说话人的身份信息，即用户的声纹模型。

otherwise to step A6 started, users facing the TV voice acquisition unit (11) to speak, speak the contents of any text, the duration of at least 30 sec;

. Speaker voice signal

A3 voice acquisition unit (11) gathers the input, and its A / D conversion, filter processing;

A4. Voiceprint identification unit (13) after step A3 by pretreatment of the speech signal processing, including windowing framing, pre-emphasis, endpoint detection, feature extraction voiceprint parameters, and then call the information storage unit (2) already voiceprint model voiceprint matches, if the match is successful, go to step A7; on the contrary, enter the next step;

.

A5 call voiceprint identification unit (13) performs voiceprint model from step A4 feature parameters extracted training, the trained model voiceprint information stored in the storage unit (2);

A6. Users facing the TV voice acquisition unit (11) Name your favorite TV channel information such as channel codes "3" represents Hunan Satellite TV, they say the number "3" to step A3 resume;

A7. Voice recognition unit (12) for voice acquisition unit (11) collected speech recognition, it will recognize the TV channel code is stored in the information storage unit (2) and with a voiceprint matching the success of voiceprint Model associate;

A8. Exit personalized information registration.

.

5 according to claim 1 using a voiceprint and voice recognition personalized TV voice awakened, characterized in that said step C wake voice, specific implementation steps:

C1 when the voice control system (1) in the "Voice wake up" turned on, the user facing the TV voice acquisition unit (11) Name wake word;.

. Speaker voice signal

C2 voice acquisition unit (11) gathers the input, and its A / D conversion, filter processing;

8.如权利要求7所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述ε和v均服从高斯分布：

ε～N(0,I)

v～N(0,Σ)

其中，Σ为对角协方差矩阵，为得到v，需要事先采用与电视机 语音唤醒场景近似的语音数据集合，利用期望最大化算法训练得到通 用背景模型，得出总变化空间和残差向量；对于任意电视用户的语音，就可利用已训练好的T，ε和m以及由用户语音构建的高斯混合模型 S，求出每个用户的总变化因子，即声纹模型。

9.如权利要求4，5所述的采用声纹和语音识别进行个性化电视 语音唤醒的方法，其特征在于，所述声纹匹配，使用的是基于总变化 因子的声纹识别系统，采用两个总变化因子v_1和v_2之间的余弦距离 来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

d (v 1, v 2) = < v 1, v 2 > | | v 1 | | · | | v 2 | |

如果d值大于预先设定的门限值则认为这两段语音来自同一个 说话人，声纹匹配成功，否则匹配失败。

10.如权利要求3所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法的系统，其特征在于，所述语音识别单元(12)声学模 型训练时，隐马尔可夫模型的状态数取5，每个状态使用一个含8个 混合高斯成分的高斯混合模型GMM表示。

11.如权利要求3所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法的系统，其特征在于，所述语音识别单元(12)的语音 解码，采用将声学、语言模型和发音字典进行有效的整合，对输入的 语音信号PLP特征参数在词网络中进行搜索和匹配，从而找到统计意 义上最匹配的词串作为最终识别结果的基于加权有限状态转换的 Viterbi解码。

C3. Voice recognition unit (12) on said processed speech recognition, the recognized content with a preset wake-up word matching, to determine whether the input speech signal collected wake of TV voice instruction, if the match is successful then the wake-up command is passed to the TV host controller (3), otherwise exit wake;

C4. Voiceprint identification unit (13) and the processed voice information storage unit (2) voiceprint model voiceprint matches, if the match is successful, then the TV channel information associated with the user's voiceprint model and wake-up operation command output to a television host controller (3), otherwise exit the wake;

C5. TV host controller (3) After the successful receipt of the voice control system (1) is passed to the TV channel information and wake operating instructions, perform TV start operation and automatically switch to the corresponding user's favorite TV channels.

6. As the use of voiceprint and voice recognition voice personalized TV Wake method according to claim 4, wherein said step of training methods A5 voiceprint Model total change factor modeling techniques.

7. Claim using the voiceprint and voice recognition according to claim 6 personalized TV voice awakened, characterized in that the total change factor modeling techniques given period the user's voice, the mean and the speaker and channel associated Gaussian mixture model GMM ultra-vector S is expressed as follows:

S = m Tv e where, m is the mean supervectors universal background model, T matrix for the total change in space, e is the residual, v vector is the current target speaker (the user) and total change factors associated channel, it represents the current the speaker's identity information, that user voiceprint model.

.

8 claim using the voiceprint and voice recognition personalize Television Voice wake method according to claim 7, wherein said e and v are Gaussian:

e ~ N (0, I) v ~ N (0, S) where, S is a diagonal covariance matrix, to obtain v, require prior recourse to wake the TV voice scene approximate

voice data collection, use EM algorithm to train to get general background model, gave a total change of space and residual difference vector; for any television user's voice, you can use the trained T, e and S m and Gaussian mixture model built by the user's voice, determine the total change factor for each user, that voiceprint model.

9. The as claimed in claim 4, 5 using voiceprint and voice recognition personalize the TV voice awakened, characterized in that said voiceprint matching, is based on the use of voiceprint recognition system total change factors, the total change in the use of two factor v_1 and v_2 to measure between two speech contained speaker sound the degree of matching pattern information, the formula is as follows: d (v 1, v 2) = and lt; v 1, v 2 and gt; | | v 1 | | ◆ | | v 2 | |

If d is greater than a predetermined threshold is considered two paragraphs from the same speaker's voice, voiceprint matches, otherwise the match fails.

10. Voiceprint as claimed in using voice recognition and speech personalized wake television system according to the method of (12) acoustic models, characterized in that, the voice recognition unit training, state the number of HMM take 5, each state using a Gaussian mixture model GMM Gaussian mixture containing eight component representation.

11. Voiceprint as claimed in using voice recognition and speech personalized TV system wake-up method according to claim 3, characterized in that said speech voice recognition unit (12) decoding, using acoustic, language models and pronunciation dictionaries for effective integration of the voice signal input characteristic parameters PLP search and match the word network, in order to find the statistical sense of the word that best matches the string as a final recognition result based on the weighted Finite state transitions Viterbi decoding.

Description:

采用声纹和语音识别进行个性化电视语音唤醒的方法

Using the voiceprint and voice recognition personalize Television Voice Wake methods

技术领域

Technical Field

本发明涉及声纹识别和语音识别技术，通过声纹识别进行电视机使用用户 的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒。

背景技术

近年来，随着智能语音处理技术、互联网和云计算技术的快速发展，目前 市面上出现的智能电视大部分都开始支持简单的语音操作，如进行音量调节，资源搜索等，同时电视语音助手等应用也开始出现。但由于中国各地的方言种 类多、同音字多、各地方言的语音语调的不同、周围环境噪声等因素造成语音 识别的准确率很低，电视无法执行用户语音命令的现象时常出现。因此，对消 费者来说，目前的语音电视在一定程度上并不能满足其各项需求。同时，这些 人机语音交互功能大多需要通过云端计算来完成，存在一定的安全隐患。

现在几乎每个家庭都拥有一台或几台电视机，但一台电视通常会被几位家庭 成员使用，特别是放置在客厅的电视，而每位家庭成员对电视节目频道的喜好 却不尽相同，这时在打开电视后通常需要用遥控器对电视节目频道进行遍历切 换以找到自己最喜欢的频道，非常不方便且繁琐。

发明内容

为解决现有技术的不足，本发明提供一种采用语音进行个性化电视唤醒的 实现方法，能够对用户语音采用声纹识别和语音识别，实现对用户的身份进行监测并使用非常简单的唤醒词就能唤醒电视打开其最喜欢的电视频道。

众所周知，声纹识别又称说话人识别，是生物识别技术的一种，声纹识别 和语音识别均属于智能语音处理技术的范畴。具体是将说话人语音信号中的声纹特征参数与训练好的声纹模型进行模式匹配从而自动进行说话人身份确认或 辨认的技术。语音识别是通过模式识别和语言理解将语音信号中包含的文字信 息提取出来的技术，主要包括特征提取、声学模型和语言模型训练、解码技术 等。

本发明具体实现步骤如下：

A.个性化信息注册

采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别出用户最 喜爱的电视频道信息。

B.信息存储

The present invention relates to a voiceprint recognition and speech recognition technology, by voiceprint identification were TV user's identification, and through identification of a user's speech and the speech recognition result to control the TV personalize voice wake.

BACKGROUND

In recent years, with the rapid development of intelligent voice processing technology, the Internet and cloud computing, smart TV currently on the market that appears most are beginning to support a simple voice operations, such as adjusting the volume, resource searches, etc., while TV voice assistant applications have begun to appear. However, due to the multiple dialects vary throughout China, homophone many different factors such as ambient noise around the dialect pronunciation and intonation of speech recognition accuracy caused by low TV does not execute user voice commands phenomenon often occurs. So, for consumers, the current television voice to some extent and can not meet its needs. At the same time, most of these machine voice interaction capabilities needed to accomplish through cloud computing, there are some security risks.

Now almost every family has one or several TV sets, but a TV is usually used by several family members, especially placed in the living room TV, and each family member on the television program channel preferences but not the same, then after the turn on the TV remote control is typically required to traverse switching television channels to find their favorite channel, very inconvenient and cumbersome.

Summary

To solve the problem of the prior art, the present invention provides a method of using speech realization of personalized TV Wake, it is possible for the user using the voiceprint voice recognition and speech recognition, to achieve the user's identity and to monitor the use of very simple Words can wake wakeup open their favorite TV channels.

As we all know, voiceprint identification, also known as speaker recognition is a biometric technology, voiceprint recognition and speech recognition are within the scope of intelligent voice processing technology. Specifically the speaker's

将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电视机信息存 储单元。

C.语音唤醒

用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可根据声纹 识别和语音识别系统对唤醒词的声纹和语音识别结果从电视机信息存储单元中调出用户的个性化信息，自动完成打开电视并自动切换到用户最喜爱的电视频 道。

以上所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特点 是，所述步骤A、步骤B，在用户作第一次操作时完成，以后只需操作步骤C， 即可完成其个性化语音电视的唤醒。

当用户对电视频道的喜好发生改变时，通过重新执行步骤A，进行用户个性 化信息注册，对信息存储单元中的电视频道信息进行更新，此时无需进行声纹 建模，仅需进行声纹匹配和语音识别。

由于本发明使用了声纹识别技术，很容易对不同用户建立起其对应的声纹 密钥(声纹模型)，通过不同用户声纹密钥的唯一性实现对其喜爱的电视频道信息进行关联和保护，进而实现具有用户个性化的电视语音唤醒功能。

本发明与现有技术相比，具有以下特点：

第一、用户只需张嘴说简单的唤醒词，如"hello电视"就能自动打开电视 并立即切换到自己最喜欢的电视台节目频道，无需再手动开机遥控，给用户带来智能、时尚的应用体验。

第二、个性化注册时的声纹识别性能与所说的文本和使用的语言无关，方 便使用。

第三、用声纹作为密码，简单的电视频道代码和唤醒词语音作为交互方式， 因此语音识别系统仅需实现简单的关键词识别即可，识别速度快，且声纹识别 和语音识别仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便 且保证了家庭信息的安全性。

第四、可设置多位用户的个性化语音开机信息，除唤醒功能外，同时也可 用于个性化自动语音换台。

第五、本发明的算法实现模块具有很好的可移植性和可扩展性，可被移植 到有线电视的机顶盒、通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

附图说明

voiceprint speech signal characteristic parameters and trained voiceprint model for pattern matching to automatically perform speaker identification or identification technology. Speech recognition is through pattern recognition and language information contained in the understanding of the speech signal will be extracted from the technology, including feature extraction, acoustic model and language model training, decoding technology.

The present invention specific implementation steps are as follows:

A. Personalized Information Register using the voiceprint identification system to establish the user's voiceprint model, using voice recognition system identifies the user's favorite TV channel information.

B. Information storage

The user voiceprint model and favorite TV channels and stores information related to the TV information storage unit.

C. Voice Wake Up

Users only need to tell the television voice collecting unit pre-set wake-up word, you can wake up the words of the voiceprint and voice recognition results from the television tune information storage unit based on the voice recognition and speech recognition systems the user's personalized information, auto-complete turn on the TV and automatically switch to the user's favorite TV channels. voiceprint and voice recognition using personalized television speech wake method described above, characterized in that said step A, step B, at the time of completion of the user operation for the first time, just after the steps C to complete their personalized voice TVs wake.

When the user changes the TV channel preferences, by re-executing step A, user personalization information registration information storage unit to update the information television channel, this time without the need for a voiceprint model, only conduct voiceprint matching and speech recognition.

Since the present invention uses a voiceprint recognition technology, it is easy to establish the corresponding voiceprint key (voiceprint model) for

图1为本发明采用声纹和语音识别进行个性化电视语音唤醒的方法 系统组成框图；

图2为本发明实施例用户个性化信息注册系统流程图；

图3为本发明实施例个性化语音唤醒电视的流程图。

具体实施方式

以下结合附图和实施例对本发明作进一步的描述

本发明构建了一套采用声纹和语音识别进行个性化电视语音唤醒的系统 (如图1所示)，该系统包括用于用户个性化信息注册、唤醒词识别和声纹匹配 的语音控制系统1，用于接收并存储从语音控制系统1中传递来的已训练好的用 户声纹模型和已识别出的电视频道信息的信息存储单元2和接收从语音控制系 统1和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操作的 电视机主控制器3组成，以电信号方式连接。其中：

所述语音控制系统1由语音采集单元11，语音识别单元12和声纹识别单元 13组成，其中：

所述语音采集单元11用于完成用户语音信号的采集和处理；

所述语音识别单元12用于识别用户所说的电视频道信息和唤醒词；

所述声纹识别单元13用于从用户语音中提取出代表其身份信息的声纹信息 进行声纹建模和声纹匹配。

本发明的语音控制系统有二个状态按钮，即"用户注册"和"语音唤醒"。 当用户需要进行个性化信息注册时开启"用户注册"，平时均处于"语音唤醒" 状态，这样有利于节省电能。

下面结合附图分别介绍用户个性化信息注册过程和自动语音唤醒的具体实 施方式

用户首次进行个性化信息注册时(如图2所示)实施例，需对电视机说话 二次，第一次为任意文本的语音，用于声纹建模，第二次为用户自己最喜欢的 电视频道信息，用于将电视频道信息与用户声纹模型进行关联。具体步骤如下：

A1.开启语音控制系1，使其处于"用户注册"状态；

A2.若用户为首次进行个性化信息注册，执行该步骤，否则转向步骤A6开 始执行，用户对着电视机

different users of different users through a unique voiceprint key to their favorite TV channels correlate information and protection, so as to realize the TV voice with user personalization wake-up function.

Compared with the prior art, it has the following characteristics:

First, the user simply opened his mouth to say a simple wake-up words, such as "hello TV" will automatically turn on the TV and immediately switch to your favorite television program channels, without the need to manually turn the remote control, giving users a smart, fashion experience.

Second, voiceprint recognition performance and said the language used in the text and personalized registration has nothing to do, easy to use.

Third, using the voiceprint as a password, a simple television channel code and wake-word speech as an interactive way, so speech recognition system just to implement simple keyword recognition, recognition speed, and voiceprint recognition and speech Only trained and tested off-line identification can be completed without the need to send to the cloud server, easy to use and ensure the safety of family information.

Fourth, you can set personalized voice multiple users boot information, in addition to wake-up function, but also can be used to personalize automated voice for Taiwan.

Fifth, the algorithm of the present invention achieves module has good portability and scalability, can be ported to run on cable television set-top boxes, high-speed general-purpose processor or DSP chips, widely used in smart home related fields.

Brief Description

1 of the present invention using the voiceprint and voice recognition voice wake personalized TV system block diagram of the method;

2 is a schematic invention registered user personalization information systems implementation flow chart;

的语音采集单元11说话，说话内容为任意文本，持续 时间至少30秒；

A3.语音采集单元11采集输入的说话人语音信号，并对其进行A/D转换， 滤波处理；

A4.声纹识别单元13对经步骤A3处理后的语音信号作预处理，包括加窗分 帧、预加重、端点检测，提取声纹特征参数，然后调用信息存储单元2中已有的声纹模型进行声纹匹配，若匹配成功，转向步骤A7；反之，进入下一步操作；

A5.调用声纹识别单元13对从步骤A4中提取的特征参数进行声纹模型的训 练，将训练好的声纹模型存储到信息存储单2；

A6.用户对着电视机的语音采集单元11说出自己最喜欢的电视频道信息， 如频道代码"3"代表湖南卫视，则说出数字"3"，转向步骤A3重新开始；

A7.语音识别单元12对语音采集单元11采集到的语音进行识别，将识别出 的电视频道代码存储到信息存储单元2中并与声纹匹配成功的声纹模型进行关联；

A8.退出个性化信息注册。

当用户的电视频道喜好发生改变时，可通过重新进行个性化信息注册来更 新信息存储单元2中的电视频道信息，此时只需对着电视机的语音采集单元11说出电视频道信息即可，语音控制系统1会自动从上述的步骤A6开始执行，无 需对用户声纹重新建模。

自动进行个性化语音唤醒电视(如图3所示)实施例的具体步骤如下：

C1.当语音控制系统1处于"语音唤醒"开启状态时，用户对着电视机语音 采集单元11说出唤醒词；

C2.语音采集单元11采集输入的说话人语音信号，并对其进行A/D转换， 滤波处理；

C3.语音识别单元12对上述处理过的语音进行识别，将识别出的内容与预 先设定的唤醒词进行匹配，以判断采集到的输入语音信号是否为电视机的语音 唤醒指令，若匹配成功则将唤醒操作指令传递给电视机主控制器3，否则退出唤 醒系统；

C4.声纹识别单元13将处理过的语音与信息存储单元2中的声纹模型进行 声纹匹配，若匹配成功，则将与用户声纹模型关联的电视频道信息和唤醒操作

Figure 3 is a flowchart illustrating the invention personalized voice wake-TV implementation.

DETAILED DESCRIPTION

The following accompanying drawings and the embodiment of the present invention will be further described

The present invention is constructed using a voiceprint and voice recognition voice wake personalized TV system (Figure 1), which includes personalized information for users registered, Wake word recognition voiceprint matches Voice control system for receiving and storing passed from voice control system 1 to a good user voiceprint model and has identified information television channel information storage unit 2 has been training and receive voice control system 1 and the information store from control information transmission unit, responsible for the implementation of personalized television voice wake-up consisting of a TV host controller 3, connected to an electrical signal. Where:

The voice control system by the voice acquisition unit 11, a voice recognition unit 12 voiceprint identification unit 13, of which:

The voice collecting unit 11 for the completion of the acquisition and processing user speech signal;

The voice recognition unit 12 for identifying a user of said television information and wake-up word;

The voiceprint identifying unit 13 for extracting from a speech on behalf of its user identity information voiceprint information modeling voiceprint matches the voiceprint. voice control system of the present invention has two state button, the "Registration" and "Voice wake up." When the user needs to open the "User Registration", usually are in the "Voice wake-up" state, which will help save energy when personalized information registration. are described below in conjunction with the accompanying drawings of specific embodiments of the registration process the user personalized information and automated voice wake of

When users first time personalized information registered (Figure 2) Example, the need for television to speak twice, the first time any text of the speech, the voiceprint model for the second time for the user his favorite TV channel

指令输出给电视机主控制器3，否则退出唤醒系统；

C5.电视机主控制器3在成功收到语音控制系统1传递来的电视频道信息和 唤醒操作指令后，执行电视开机操作并自动切换到对应用户喜爱的电视频道上。

上述提到的预先设定的特定语音唤醒词可为任意设定的词组，如"打开电 视机"、"开电视"、"打开电视"、"hello电视"、"hello TV"等等。语音采集 单元11可为麦克风或其他语音采集设备，用于感知周边的语音信号。

本发明涉及到的声纹识别单元13所采用的识别算法为基于总变化因子(又 称为i-Vector)的说话人确认方法，其具体构建步骤如下：

一、语音信号的预处理

语音信号的预处理包括对语音信号进行预加重、加窗分帧和端点检测。对 于经过A/D转换后的数字语音信号，预加重使用如下的一阶数字滤波器来提升语音信号的高频分量：

$H(z)=1-0.97z^{-1}$

然后采用如下式的汉明窗窗函数对语音信号进行平滑，并在加窗过程中进行语 音信号的分帧操作，其中N为窗的长度。

$W(n)=0.54+0.46\cos[2pin/(N-1)]$

对于采样率为8000Hz的语音，分帧时使用的帧长为25毫秒，帧移为10毫 秒。本发明采用的是基于短时能量的端点检测法对语音信号进行端点检测，从 而判断语音信号的起始点和终止点，设x(·)为通过预加重后的语音信号，则表示 从语音信号的第n个点开始加窗时的短时能量E_n计算公式为：

E n = Sigma m = - ∞ ∞ [x (m) W (n - m) 2 = x 2 (n) * W 2 (n)

二、声纹特征参数的提取

声纹特征参数的提取即从说话人的语音信号中提取出能反映其身份个性的 参数。本发明使用的是39维的感知线性预测系数(Perception Linear Predictive Coefficients，PLP)，具体提取过程为：首先对上述步骤一预处理 过后的语音做短时快速傅里叶变换，然后使用24阶子带的梅尔三角滤波器组进 行临界频率分析，接着对得到的语音能量谱进行等响度曲线预加重用以模拟人 耳的听觉感知系统对频率的敏感

information for the TV channel information and user voiceprint model association. Specific steps are as follows:

.

A1 open voice control system 1, it is in "User Registration" status;

.

A2 If the user is registered for the first time personalized information, perform this step, otherwise to step A6 started, users facing the TV voice collecting unit 11 to speak, speak any text content, duration of at least 30 seconds;

A3 11 gathering input voice Speaker voice signal acquisition unit, and its A / D conversion, filter processing;.

A4. Voiceprint identification unit after 13 pairs step A3 by pretreatment of the speech signal processing, including windowing framing, pre-emphasis, endpoint detection, feature extraction voiceprint parameters, and then call the information storage unit 2 has voiceprint model voiceprint matches, if the match is successful, go to step A7; on the contrary, enter the next step;

.

A5 call voiceprint identification unit 13 from step A4 extracted characteristic parameters training voiceprint model, the model will be trained voiceprint information stored in the storage unit 2;

.

A6 user facing the TV voice collecting unit 11 say their favorite TV channel information such as channel codes "3" represents Hunan Satellite TV, they say the number "3" to step A3 restart;

.

A7 voice recognition unit 12 speech acquisition unit 11 acquired speech recognition, will recognize the TV channel code is stored in the information storage unit 2 and the voiceprint matches the voiceprint model successful association;

A8. Exit personalized information registration.

When the user's preferences change TV channels, information storage unit can be updated by re-

度，即将能量谱用如下式中的等响度曲线进行 加权。

E (omega) = (omega 2 + 56.8 x 10 6) omega 4 (omega 2 + 6.3 x 10 6) 2 (omega 2 + 0.38 x 10 9) (omega 6 + 9.58 x 10 26)

然后使用指数为1/15的幂函数对能量进行压缩，接下来做离散傅里叶反变 换，用dubin算法计算M阶全极点模型，最后对得到的模型系数进一步转换为倒谱系数得到PLP特征参数。根据端点检测结果对采集到的信号中为语音信号 之外的信号，如噪声或静音帧对应的特征参数直接丢弃。

三、声纹模型训练

本发明中的用户声纹模型训练采用的是总变化因子建模技术，该技术的目 标在于对语音信号中含有的说话人身份信息和周围环境噪声、麦克风传输等信道信息变化空间进行同时建模。给定一段用户的语音，经过上述步骤一和二的 处理后，其与说话人及信道相关的高斯混合模型GMM的均值超向量S表示如下：

$$S=m+Tv+\varepsilon$$

其中m为通用背景模型的均值超向量，T矩阵为总变化空间，ε表示残 差，v矢量是与当前目标说话人(用户)及信道相关的总变化因子，它是一个低 维矢量，代表着当前说话人的身份信息，即用户的声纹模型。ε和v均服从高 斯分布：

$$\varepsilon \sim N(0, I)\ v \sim N(0, \sum)$$

其中Σ为对角协方差矩阵。对于本发明而言，为了得到v，需要事先采用 与电视机语音唤醒场景近似的语音数据集合，利用期望最大化算法(EM)训练得到通用背景模型，得出总变化空间和残差向量。对于任意电视用户的语音，就 可利用已训练好的T，ε和m以及由用户语音构建的高斯混合模型S，求出每 个用户的总变化因子，即声纹模型。

四、声纹匹配

本发明使用的基于总变化因子的声纹识别系统中，采用两个总变化因子v_1和v_2之间的余弦距离来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

$$d(v1, v2) = < v1, v2 > | | v1 | | \cdot | | v2 | |$$

如果d值大于预先设定的门限值则认为这两段语音来自同一个说话人，声纹匹 配成功，否则匹配失

personalized information registered 2 television channel information, then just in front of the TV voice collecting unit 11 television channels say Information can be, the voice control system automatically A6 started from the above step, without the need for user voiceprint remodel. automatically personalized voice wake television specific procedure (Figure 3) to implement the following:

C1 when the voice control system is in the "Voice wake up" turned on, the user facing the TV voice collecting unit 11 wake-up word spoken;.

C2 11 gathering input voice Speaker voice signal acquisition unit, and its A / D conversion, filter processing;.

C3. Speech recognition unit 12 through the above-described processing speech recognition, the recognized content with a preset wake-word matches, collected to determine whether the input speech signal wakeup command is TV voice, If the match is passed to the TV main controller 3, or exit the wake success will wake operating instructions;

C4. Voiceprint identification unit 13 processed voice information storage unit 2 voiceprint model voiceprint matches, if the match is successful, then the model associated with the user's voiceprint information television channel and wake operating instructions output to a television host controller 3, or withdraw wake;

C5. TV host controller 3 successfully receive voice control system transfer to the TV channel information and operation instruction after wake-up, perform TV start operation and automatically switch to the corresponding user's favorite TV channels. above-mentioned predetermined specific voice can be set to wake word phrases, such as "Turn on the TV," "On TV," "turn on the TV", "hello television", "hello TV", etc. and so on. The voice collecting unit 11 may collect equipment for the microphone or other voice, the voice signal for sensing peripheral.

Recognition algorithm of the present invention relates to the voiceprint identification unit 13 uses for the speaker based on the total change factor (also known as i-Vector) confirmation method to build concrete steps are as follows:

First, the pre-processing of the speech signal

败。有研究结果显示，通过概率线性判别分析规整后的总 变化因子的鉴别能力更强，可进一步提升声纹识别的性能，但由于本发明主要 适用于家庭的电视唤醒，家庭成员数目并不多，一般为1到7之间，因此上述 基于总变化因子的声纹识别算法已足以保证本发明在声纹识别方面的准确率。

下面介绍本发明中涉及到的语音识别单元所采用的语音识别算法

由于本发明主要用于电视唤醒和频道切换，用户使用时涉及到的唤醒词和 用于表示电视频道信息的词语数量都非常少，因此语音识别单元中采用的语音识别算法只需一套简单的基于GMM-HMM(基于高斯混合模型的隐马尔可夫模型) 识别系统即可，且训练和识别过程均采用离线方法，这样能够大大提高语音识 别的速度。构建一套GMM-HMM语音识别系统主要包括声学特征参数的提取，声 学模型和语言模型训练，语音解码过程。具体实现步骤如下：

一、声学特征参数的提取

由于感知线性预测系数中同时含有说话人的身份信息和语义信息，因此上 述介绍的声纹识别算法中的PLP参数提取方法同样用于语音识别，但PLP参数的维数为52维，同时为了增强声学特征的区分能力，需要将52维的PLP参数 进行异方差线性判别分析(Heteroscedastic Linear Discriminant Analysis, HLDA)后再用于声学模型建模和解码。

二、声学模型训练

声学模型的训练采用最大似然估计准则，需要用到期望最大化算法EM和 Baum-Welch算法，前者用来处理隐马尔可夫模型中由状态序列隐变量带来的不 完全数据下的训练问题，而后者用于从训练数据中高效的累积统计量，从而有 效抽取模型参数更新所需要的信息。对于本发明的应用，在进行声学模型训练 时，隐马尔可夫模型的状态数取5，每个状态使用一个含8个混合高斯成分的 GMM模型表示，第j状态的高斯混合密度表达式如下：

b j (x) = Sigma m = 1 M c jm p jm (x)

其中 Sigma m = 1 M c jm = 1, c jm greater than or equal to 0

p jm (x) = 1 (2 pi) D / 2 | Sigma jm | 1 / 2 exp { - 1 2 (x - mu jm) T Sigma jm - 1 (x - mu jm) }

M＝8表示高斯混合数目，x为任意语音帧的特征参数，D为特征的维数，c_{jm}， micro $_{jm}$和Σ_{jm}为第j状

Pretreatment

speech signal comprises speech signal pre-emphasis, plus the window framing and endpoint detection. For after A / D converted digital voice signal pre-emphasis using the following first-order digital filter to enhance the high-frequency component of the speech signal:

H (z) = 1-0.97z $^{-1}$ and then using the following formula Hamming window window function smoothed speech signal, and framing operations in the speech signal windowing process, where N is the length of the window.

W (n) = 0.54 0.46cos [2pin / (N-1)]

For a sampling rate of 8000Hz voice, frame length framing the use of 25 milliseconds, a frame shift of 10 ms. The present invention is based on the use of short-term energy endpoint detection of speech signal endpoint detection, to determine the start and end point of the speech signal, set x (�) through the pre-emphasis of the speech signal, then from the voice signal The n-th point short-term energy E_n windowed time is calculated as:

E n = Sigma m = - 8 8 [x (m) W (n - m) 2 = x 2 (n) * W 2 (n)

Second, extract voiceprint parameters extract voiceprint characteristic parameter that is extracted from the speaker's voice signal can reflect the personality of its identity parameters. The present invention is a 39-dimensional perception of linear prediction coefficient (Perception Linear Predictive Coefficients, PLP), the specific extraction process: first speech after the above steps do a short pretreatment fast Fourier transform, and then use the 24-step son Mel triangle filter bank with the frequency of critical analysis, and then to get the energy spectrum of the speech loudness curve pre-emphasis to simulate the auditory perception system sensitivity frequency energy spectrum coming by the following formula, etc. loudness weighting curve.

E (omega) = (omega 2 56.8 x 10 6) omega 4 (omega 2 6.3 x 10 6) 2 (omega 2 0.38 x 10 9) (omega 6 9.58 x 10 26)

Then use the index as a power function of the energy of 1/15 compression, inverse discrete Fourier transform to do next, calculate M-order all-

态的GMM模型中第m个高斯分量的权重，均值向量和协方差 矩阵。

三、语言模型训练

目前主流语音识别所用到的语言模型绝大部分是n-元(n-gram)统计语言模 型，是将词串的生成过程表示为一个词间跳转的n-1阶马尔可夫过程。训练和评价n-gram语言模型的指标为下式的模型复杂度(Perplexity)：

H = - min K - ∞ 1 K log 2 (P (w 1,..., w K)) ≈ - 1 K Sigma k = 1 K log 2 (P (w k | w k - 1, w k - 2,..., w k - N +!))

其中w_1,...,w_K为含有K个词的词序列，N为语言模型的元数，本发明中取 值为2。从理论上来看，语言模型对词串的期望复杂度应尽可能小，这表示语言模型在对给定的历史词串进行当前词预测时，能拥有更高的确信度。实际上语 言模型的训练准则就是使其对训练集合中句子的复杂度最小化。

四、语音解码

在给定声学模型和语言模型的情况下，解码器的好坏与语音识别的速度直 接相关，从而直接决定了用户的体验。虽然本发明用来唤醒电视所用的词语有限，由它们构成的解码网络并不大，但本发明仍然推荐采用基于加权有限状态 转换(WFST)的Viterbi解码技术，它可以将声学、语言模型和发音字典进行有 效的整合，对输入的语音信号PLP特征参数在词网络中进行搜索和匹配，从而 找到统计意义上最匹配的词串作为最终的识别结果。

上述语音识别单元中的声学模型和语言模型均可采用目前主流的开源工具 HTK或者Kaldi提前训练好并存储起来，在进行语音唤醒时，仅需调用存储好的 声学和语言模型，离线对采集到的用户语音进行PLP特征参数提取和解码即可。 实验证明，本发明方法具有训练时间短，声纹和语音识别速度都很快，且识别 率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云 端服务器，使用方便且保证了家庭信息的安全性。另外，本发明在电视关机时， 电视整体处于掉电状态，即电视机主控制器、语音识别单元、声纹识别单元、 信息存储单元均断电，只有语音采集单元实时采集用户的语音信号消耗电能， 而该模块耗电非常少，大大节省了电能消耗。

综上所述，本发明提供的基于声纹识别和语音识别进行个性化电视语音唤 醒的实现方法，能够在电视处于关机状态时，用户直接说类似于"hello电视" 这

pole model with dubin algorithm. Finally, the model coefficients are further converted to Cepstral PLP obtained characteristic parameters. According to the results of the collected endpoint detection signal is a signal to voice signal, such as noise or silence frame corresponding to the characteristic parameter discarded.

Third, voiceprint model training

The present invention is a user voiceprint model training uses a total change factor modeling techniques, the goal of this technology is to change the channel information space speaker identity information and ambient noise, the microphone to transmit voice signals contained etc. At the same time modeling. To fix a user's voice, after the processes of steps one and two, with the speaker and the average channel correlation Gaussian mixture model GMM ultra-vector S is expressed as follows:

S = m Tv e where m is the mean supervectors universal background model, T matrix for the total change in space, e represents the residual, v vector is the current target speaker (the user) and total change factors associated channel, which is a low-dimensional vector, representing the current speaker's identity information that the user's voiceprint model. e and v are subject to Gaussian distribution:

e ~ N (0, I) v ~ N (0, S) where S is a diagonal covariance matrix. For the purposes of the present invention, in order to obtain v, requires prior wake scene using approximate voice data collection and TV voice, use Expectation Maximization algorithm (EM) training to give general background model, gave a total change of space and residual vector. For any television user's voice, you can use the trained T, e and m and Gaussian mixture model S constructed by the user's voice, determine the total change factor for each user, that voiceprint model.

Fourth, voiceprint matching

The present invention is based on the use of the voice recognition system total change factors, the overall change in the use of two factor v_1 and v_2 to measure between the degree of matching voice contained two speaker voiceprint information, the formula is as follows: d (v 1, v 2) = <v 1, v 2> | | v 1 | | | | v 2 | |

样的唤醒词，电视机就可以实现直接开机并立即显示自己最喜爱的电视频道。 本发明的实现方法还可应用于电视机的用户个性化自动语音换台，同时可被移 植到通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

If d is greater than a predetermined threshold is considered two paragraphs from the same speaker's voice, voiceprint matches, otherwise the match fails. The results of research show that the probability of linear discriminant analysis by discriminating factor in the total change after more structured, can further improve the performance of voiceprint identification, but the present invention is primarily for home TV Wake, the number of family members is not much, generally between 1-7, so the above voiceprint recognition algorithm based on total variation factors are sufficient to ensure the accuracy of the present invention in terms of voiceprint identification.

Speech recognition algorithm

The following describes the present invention relates to a voice recognition unit used

Since the present invention is mainly used for wake-up and TV channel switching, when it comes to the user and a wake-up word indicates the number of words in the information channels are very small, so the speech recognition algorithm used in the speech recognition unit just one Based on a simple set of GMM-HMM (Gaussian mixture model based on hidden Markov model) identification system can be, and training and recognition processes are off-line method, which can greatly improve the speech recognition rate. Construction of a GMM-HMM speech recognition system includes an acoustic characteristic parameter extraction, acoustic model and language model training, speech decoding process. Specific steps are as follows:

First, extract the acoustic feature parameters

Since the perception of identity information and semantic information containing both linear prediction coefficient speaker, so the sound pattern recognition algorithm to the introduction of the PLP parameter extraction method is also used for voice recognition, but the dimension of the PLP for 52-dimensional parameter, Meanwhile, in order to enhance the ability to distinguish between acoustic features you need to be 52-dimensional parameter Heteroscedasticity PLP linear discriminant analysis (Heteroscedastic Linear Discriminant Analysis, HLDA) then used to carry out the acoustic modeling and decoding.

Second, the acoustic model training

Training

acoustic model using the maximum likelihood estimation, need to use EM expectation maximization algorithm and Baum-Welch algorithm, the former is used to dealing with incomplete data hidden Markov model brought down by the state series of hidden variables training issues and the latter for the efficient accumulation of statistics from the training data, so as to effectively extract model parameters to update the information they need. For the application of the present invention, during the acoustic model training, state the number of HMM take 5, each state GMM model using a Gaussian mixture containing eight components, said Gaussian mixture density of the j state following expression: b j (x) = Sigma m = 1 M c jm p jm (x) where Sigma m = 1 M c jm = 1, c jm greater than or equal to 0 p jm (x) = 1 (2 pi) D / 2 | Sigma jm | 1/2 exp {- 1 2 (x - mu jm) T Sigma jm - 1 (x - mu jm)}

M = 8 indicates the number of Gaussian mixture, x is the characteristic parameter to any speech frame, D is the dimension of features, c_{jm}, micro $_{jm}$ and S_{jm} j state for the GMM model in the m-th power of the Gaussian component weight, mean vector and covariance matrix.

Third, the language model training mainstream speech recognition language models currently used by the vast majority of n- yuan (n-gram) language model, the word sequence generation process is expressed as one word jumps between n-1-order Markov process. Training and Evaluation n-gram language model indicators for the formula complexity of the model (Perplexity):

H = - min K - 8 1 K log 2 (P (w 1,..., w K)) ◆ - 1 K Sigma k = 1 K log 2 (P (wk | wk - 1, wk - 2,..., wk -! N)) where w_1 ,..., w_K as word sequence contains K-word, N is metadata language model, the present invention, a value of 2. From the theoretical point of view, the complexity of the language model to the desired word strings should be as small as possible, which means that the language model in the history of a given word string current word prediction, to have a higher degree of certainty. In fact, the language model training guidelines is to make it to the training set sentence complexity is minimized.

Fourth, the speech decoding

In the case of a given acoustic model and language model, the quality and speed of speech recognition is directly related to the decoder, which directly determines the user's experience. Although the present invention is used to wake up terms used in television is limited, they constitute a decoding network is not large, but the present invention is still recommended based on weighted finite state transition (WFST) of Viterbi decoding technology, it can be acoustic, language model and pronunciation Dictionary for effective integration of the voice signal input characteristic parameters PLP search and matching network in words, in order to find the statistical sense of the word that best matches the string as the final recognition result. the speech recognition unit of acoustic models and language models can be with the current mainstream open source tools HTK or Kaldi good training and stored in advance, during a voice wake, just call the stored good acoustic and language models offline the collected user speech PLP feature extraction and decoding can be. Experiments show that the method of the present invention has the training time is short, voiceprint and voice recognition speed very quickly, and the recognition rate characteristics. Voiceprint and voice recognition are just off-line training and testing can be completed without the need to send to the cloud server, easy to use and ensure the safety of family information. Further, the present invention when the TV off, the TV is powered down overall, the TV host controller, a voice recognition unit, voiceprint identification unit, an information storage unit are off, only the voice collecting unit collecting the user's voice signal in real-time consumption energy, and the module power consumption is very low, saving power consumption.

In summary, the implementation personalize Television Voice wake voiceprint recognition and speech recognition based on the present invention provides, it is possible when the TV is turned off, users just say like "hello television" this wake-up word TV can be turned on and direct display your favorite TV channels immediately. Implementation method of the present invention can also be used in TV user personalization automatic speech for Taiwan, and can be ported to run on general-purpose high-speed processor or DSP chips, widely used in smart home related fields.

Citing: CN103442290 A ; CN103595869 A ; CN103632669 A ; CN103646646 A ;

Cited by: CN104821168 A ; CN104821168 B ; CN105049882 A ; CN105049882 B ; CN105137768 A ; CN105141768 A ; CN105206271 A ; CN105355195 A ; CN105374355 A ; CN105609103 A ; CN105719650 A ; CN105791935 A ; CN105915988 A ; CN105959806 A ; CN105979376 A ; CN106098068 A ; CN107015781 A ; CN108320753 A ; US9799335 BB ; WO17092189 A1 ; WO17217978 A1 ; WO19051668 A1 ;

Non patent citations: NAJIM DEHAK, ETC: "Support Vector Machines versus Fast Scoring in the Low-Dimensional Total Variability Space for Speaker Verification",

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104575504 A
(43)申请公布日 2015.04.29

(21)申请号 201410840544.9

(22)申请日 2014.12.24

(71)申请人 上海师范大学
地址 200234 上海市徐汇区桂林路100号

(72)发明人 龙艳花 叶宏 魏浩然

(74)专利代理机构 上海申新律师事务所 31272
代理人 周云

(51)Int.Cl.
G10L 17/04(2013.01)
G10L 15/26(2006.01)
H04N 21/441(2011.01)
H04N 21/439(2011.01)

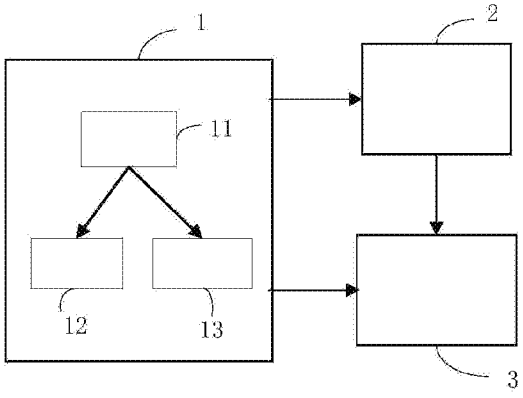
权利要求书3页 说明书7页 附图1页

(54)发明名称

采用声纹和语音识别进行个性化电视语音唤醒的方法

(57)摘要

本发明采用声纹和语音识别进行个性化电视语音唤醒的方法，具体指一种通过声纹识别进行电视机使用用户的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒，涉及声纹识别和语音识别技术。本发明构成系统包括语音控制系统(1)，信息存储单元(2)和电视机主控制器(3)以电信号方式连接组成。本发明具有训练时间短，声纹和语音识别速度都很快，且识别率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。还可应用于电视机的用户个性化自动语音换台，同时可被移植到通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。



CN 104575504 A

1.一种采用声纹和语音识别进行个性化电视语音唤醒的方法，通过采用声纹识别和语音识别系统，对用户的身份进行监测，实现个性化电视语音唤醒，其特征在于，包括如下步骤：

A. 个性化信息注册

采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别出用户最喜爱的电视频道信息；

B. 信息存储

将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电视机信息存储单元；

C. 语音唤醒

用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可根据声纹识别和语音识别系统对唤醒词的声纹和语音识别结果从电视机信息存储单元中调出用户的个性化信息，自动完成打开电视并自动切换到用户最喜爱的电视频道。

2.如权利要求1所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述步骤A、步骤B，在用户作第一次操作时完成，以后只需操作步骤C，即可完成其个性化语音电视的唤醒；

当用户对电视频道的喜好发生改变时，通过重新执行步骤A，进行用户个性化信息注册，对信息存储单元中的电视频道信息进行更新，此时无需进行声纹建模，仅需进行声纹匹配和语音识别。

3.如权利要求1所述的采用声纹和语音识别进行个性化电视语音唤醒的方法的系统，其特征在于，所述系统包括用于用户个性化信息注册、唤醒词识别和声纹匹配的语音控制系统(1)，用于接收并存储从语音控制系统(1)中传递来的已训练好的用户声纹模型和已识别出的电视频道信息的信息存储单元(2)和接收从语音控制系统(1)和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操作的电视机主控制器(3)组成，以电信号方式连接，其中：

所述语音控制系统(1)由语音采集单元(11)，语音识别单元(12)和声纹识别单元(13)组成，其中：

所述语音采集单元(11)用于完成用户语音信号的采集和处理；

所述语音识别单元(12)用于识别用户所说的电视频道信息和唤醒词；

所述声纹识别单元(13)用于从用户语音中提取出代表其身份信息的声纹信息进行声纹建模和声纹匹配。

4.如权利要求1所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述步骤A个性化信息注册，具体实现步骤为：

A1. 开启语音控制系统(1)，使其处于“用户注册”状态；

A2. 若用户为首次进行个性化信息注册，执行该步骤，否则转向步骤A6开始执行，用户对着电视机的语音采集单元(11)说话，说话内容为任意文本，持续时间至少30秒；

A3. 语音采集单元(11)采集输入的说话人语音信号，并对其进行A/D转换，滤波处理；

A4. 声纹识别单元(13)对经步骤A3处理后的语音信号作预处理，包括加窗分帧、预加重、端点检测，提取声纹特征参数，然后调用信息存储单元(2)中已有的声纹模型进行声纹匹配，若匹配成功，转向步骤A7；反之，进入下一步操作；

Petitioner Exhibit 1004, Page 665 of 1052

A5. 调用声纹识别单元 (13) 对从步骤A4中提取的特征参数进行声纹模型的训练，将训练好的声纹模型存储到信息存储单元 (2)；

A6. 用户对着电视机的语音采集单元 (11) 说出自己最喜欢的电视频道信息，如频道代码“3”代表湖南卫视，则说出数字“3”，转向步骤 A3 重新开始；

A7. 语音识别单元 (12) 对语音采集单元 (11) 采集到的语音进行识别，将识别出的电视频道代码存储到信息存储单元 (2) 中并与声纹匹配成功的声纹模型进行关联；

A8. 退出个性化信息注册。

5. 如权利要求 1 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述步骤 C 语音唤醒，具体实现步骤为：

C1. 当语音控制系统 (1) 处于“语音唤醒”开启状态时，用户对着电视机语音采集单元 (11) 说出唤醒词；

C2. 语音采集单元 (11) 采集输入的说话人语音信号，并对其进行 A/D 转换，滤波处理；

C3. 语音识别单元 (12) 对上述处理过的语音进行识别，将识别出的内容与预先设定的唤醒词进行匹配，以判断采集到的输入语音信号是否为电视机的语音唤醒指令，若匹配成功则将唤醒操作指令传递给电视机主控制器 (3)，否则退出唤醒系统；

C4. 声纹识别单元 (13) 将处理过的语音与信息存储单元 (2) 中的声纹模型进行声纹匹配，若匹配成功，则将与用户声纹模型关联的电视频道信息和唤醒操作指令输出给电视机主控制器 (3)，否则退出唤醒系统；

C5. 电视机主控制器 (3) 在成功收到语音控制系统 (1) 传递来的电视频道信息和唤醒操作指令后，执行电视开机操作并自动切换到对应用户喜爱的电视频道上。

6. 如权利要求 4 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述步骤 A5 中的声纹模型的训练方法为总变化因子建模技术。

7. 如权利要求 6 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述总变化因子建模技术，给定一段用户的语音，与说话人及信道相关的高斯混合模型 GMM 的均值超向量 S 表示如下：

$$S = m + Tv + \varepsilon$$

其中，m 为通用背景模型的均值超向量，T 矩阵为总变化空间，ε 为残差，v 矢量是与当前目标说话人（用户）及信道相关的总变化因子，它代表着当前说话人的身份信息，即用户的声纹模型。

8. 如权利要求 7 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述 ε 和 v 均服从高斯分布：

$$\varepsilon \sim N(0, I)$$

$$v \sim N(0, \Sigma)$$

其中，Σ 为对角协方差矩阵，为得到 v，需要事先采用与电视机语音唤醒场景近似的语音数据集合，利用期望最大化算法训练得到通用背景模型，得出总变化空间和残差向量；对于任意电视用户的语音，就可利用已训练好的 T，ε 和 m 以及由用户语音构建的高斯混合模型 S，求出每个用户的总变化因子，即声纹模型。

9. 如权利要求 4，5 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述声纹匹配，使用的是基于总变化因子的声纹识别系统，采用两个总变化因子 v_1

Petitioner Exhibit 1004, Page 666 of 1052

和 v_2之间的余弦距离来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

$$d(\mathbf{v}_1,\mathbf{v}_2)=\frac{\langle \mathbf{v}_1,\mathbf{v}_2\rangle}{\|\mathbf{v}_1\|\cdot\|\mathbf{v}_2\|}$$

如果 d 值大于预先设定的门限值则认为这两段语音来自同一个说话人，声纹匹配成功，否则匹配失败。

10. 如权利要求 3 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法的系统，其特征在于，所述语音识别单元 (12) 声学模型训练时，隐马尔可夫模型的状态数取 5，每个状态使用一个含 8 个混合高斯成分的高斯混合模型 GMM 表示。

11. 如权利要求 3 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法的系统，其特征在于，所述语音识别单元 (12) 的语音解码，采用将声学、语言模型和发音字典进行有效的整合，对输入的语音信号 PLP 特征参数在词网络中进行搜索和匹配，从而找到统计意义上最匹配的词串作为最终识别结果的基于加权有限状态转换的 Viterbi 解码。

Petitioner Exhibit 1004, Page 667 of 1052

采用声纹和语音识别进行个性化电视语音唤醒的方法

技术领域

[0001] 本发明涉及声纹识别和语音识别技术，通过声纹识别进行电视机使用用户的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒。

背景技术

[0002] 近年来，随着智能语音处理技术、互联网和云计算技术的快速发展，目前市面上出现的智能电视大部分都开始支持简单的语音操作，如进行音量调节，资源搜索等，同时电视语音助手等应用也开始出现。但由于中国各地的方言种类多、同音字多、各地方言的语音语调的不同、周围环境噪声等因素造成语音识别的准确率很低，电视无法执行用户语音命令的现象时常出现。因此，对消费者来说，目前的语音电视在一定程度上并不能满足其各项需求。同时，这些人机语音交互功能大多需要通过云端计算来完成，存在一定的安全隐患。

[0003] 现在几乎每个家庭都拥有一台或几台电视机，但一台电视通常会被几位家庭成员使用，特别是放置在客厅的电视，而每位家庭成员对电视节目频道的喜好却不尽相同，这时在打开电视后通常需要用遥控器对电视节目频道进行遍历切换以找到自己最喜欢的频道，非常不方便且繁琐。

发明内容

[0004] 为解决现有技术的不足，本发明提供一种采用语音进行个性化电视唤醒的实现方法，能够对用户语音采用声纹识别和语音识别，实现对用户的身份进行监测并使用非常简单的唤醒词就能唤醒电视打开其最喜欢的电视频道。

[0005] 众所周知，声纹识别又称说话人识别，是生物识别技术的一种，声纹识别和语音识别均属于智能语音处理技术的范畴。具体是将说话人语音信号中的声纹特征参数与训练好的声纹模型进行模式匹配从而自动进行说话人身份确认或辨认的技术。语音识别是通过模式识别和语言理解将语音信号中包含的文字信息提取出来的技术，主要包括特征提取、声学模型和语言模型训练、解码技术等。

[0006] 本发明具体实现步骤如下：

[0007] A. 个性化信息注册

[0008] 采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别出用户最喜爱的电视频道信息。

[0009] B. 信息存储

[0010] 将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电视机信息存储单元。

[0011] C. 语音唤醒

[0012] 用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可根据声纹识别和语音识别系统对唤醒词的声纹和语音识别结果从电视机信息存储单元中调出用户的个性

Petitioner Exhibit 1004, Page 668 of 1052

化信息，自动完成打开电视并自动切换到用户最喜爱的电视频道。

[0013] 以上所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特点是，所述步骤 A、步骤 B，在用户作第一次操作时完成，以后只需操作步骤 C，即可完成其个性化语音电视的唤醒。

[0014] 当用户对电视频道的喜好发生改变时，通过重新执行步骤 A，进行用户个性化信息注册，对信息存储单元中的电视频道信息进行更新，此时无需进行声纹建模，仅需进行声纹匹配和语音识别。

[0015] 由于本发明使用了声纹识别技术，很容易对不同用户建立起其对应的声纹密钥（声纹模型），通过不同用户声纹密钥的唯一性实现对其喜爱的电视频道信息进行关联和保护，进而实现具有用户个性化的电视语音唤醒功能。

[0016] 本发明与现有技术相比，具有以下特点：

[0017] 第一、用户只需张嘴说简单的唤醒词，如“hello 电视”就能自动打开电视并立即切换到自己最喜欢的电视台节目频道，无需再手动开机遥控，给用户带来智能、时尚的应用体验。

[0018] 第二、个性化注册时的声纹识别性能与所说的文本和使用的语言无关，方便使用。

[0019] 第三、用声纹作为密码，简单的电视频道代码和唤醒词语音作为交互方式，因此语音识别系统仅需实现简单的关键词识别即可，识别速度快，且声纹识别和语音识别仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。

[0020] 第四、可设置多位用户的个性化语音开机信息，除唤醒功能外，同时也可用于个性化自动语音换台。

[0021] 第五、本发明的算法实现模块具有很好的可移植性和可扩展性，可被移植到有线电视的机顶盒、通用的 DSP 高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

附图说明

[0022] 图 1 为本发明采用声纹和语音识别进行个性化电视语音唤醒的方法系统组成框图；

[0023] 图 2 为本发明实施例用户个性化信息注册系统流程图；

[0024] 图 3 为本发明实施例个性化语音唤醒电视的流程图。

具体实施方式

[0025] 以下结合附图和实施例对本发明作进一步的描述

[0026] 本发明构建了一套采用声纹和语音识别进行个性化电视语音唤醒的系统（如图 1 所示），该系统包括用于用户个性化信息注册、唤醒词识别和声纹匹配的语音控制系统 1，用于接收并存储从语音控制系统 1 中传递来的已训练好的用户声纹模型和已识别出的电视频道信息的信息存储单元 2 和接收从语音控制系统 1 和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操作的电视机主控制器 3 组成，以电信号方式连接。其中：

[0027] 所述语音控制系统 1 由语音采集单元 11，语音识别单元 12 和声纹识别单元 13 组成，其中：

Petitioner Exhibit 1004, Page 669 of 1052

[0028] 所述语音采集单元 11 用于完成用户语音信号的采集和处理；

[0029] 所述语音识别单元 12 用于识别用户所说的电视频道信息和唤醒词；

[0030] 所述声纹识别单元 13 用于从用户语音中提取出代表其身份信息的声纹信息进行声纹建模和声纹匹配。

[0031] 本发明的语音控制系统有二个状态按钮，即“用户注册”和“语音唤醒”。当用户需要进行个性化信息注册时开启“用户注册”，平时均处于“语音唤醒”状态，这样有利于节省电能。

[0032] 下面结合附图分别介绍用户个性化信息注册过程和自动语音唤醒的具体实施方式

[0033] 用户首次进行个性化信息注册时（如图 2 所示）实施例，需对电视机说话二次，第一次为任意文本的语音，用于声纹建模，第二次为用户自己最喜欢的电视频道信息，用于将电视频道信息与用户声纹模型进行关联。具体步骤如下：

[0034] A1. 开启语音控制系 1，使其处于“用户注册”状态；

[0035] A2. 若用户为首次进行个性化信息注册，执行该步骤，否则转向步骤 A6 开始执行，用户对着电视机的语音采集单元 11 说话，说话内容为任意文本，持续时间至少 30 秒；

[0036] A3. 语音采集单元 11 采集输入的说话人语音信号，并对其进行 A/D 转换，滤波处理；

[0037] A4. 声纹识别单元 13 对经步骤 A3 处理后的语音信号作预处理，包括加窗分帧、预加重、端点检测，提取声纹特征参数，然后调用信息存储单元 2 中已有的声纹模型进行声纹匹配，若匹配成功，转向步骤 A7；反之，进入下一步操作；

[0038] A5. 调用声纹识别单元 13 对从步骤 A4 中提取的特征参数进行声纹模型的训练，将训练好的声纹模型存储到信息存储单 2；

[0039] A6. 用户对着电视机的语音采集单元 11 说出自己最喜欢的电视频道信息，如频道代码“3”代表湖南卫视，则说出数字“3”，转向步骤 A3 重新开始；

[0040] A7. 语音识别单元 12 对语音采集单元 11 采集到的语音进行识别，将识别出的电视频道代码存储到信息存储单元 2 中并与声纹匹配成功的声纹模型进行关联；

[0041] A8. 退出个性化信息注册。

[0042] 当用户的电视频道喜好发生改变时，可通过重新进行个性化信息注册来更新信息存储单元 2 中的电视频道信息，此时只需对着电视机的语音采集单元 11 说出电视频道信息即可，语音控制系统 1 会自动从上述的步骤 A6 开始执行，无需对用户声纹重新建模。

[0043] 自动进行个性化语音唤醒电视（如图 3 所示）实施例的具体步骤如下：

[0044] C1. 当语音控制系统 1 处于“语音唤醒”开启状态时，用户对着电视机语音采集单元 11 说出唤醒词；

[0045] C2. 语音采集单元 11 采集输入的说话人语音信号，并对其进行 A/D 转换，滤波处理；

[0046] C3. 语音识别单元 12 对上述处理过的语音进行识别，将识别出的内容与预先设定的唤醒词进行匹配，以判断采集到的输入语音信号是否为电视机的语音唤醒指令，若匹配成功则将唤醒操作指令传递给电视机主控制器 3，否则退出唤醒系统；

[0047] C4. 声纹识别单元 13 将处理过的语音与信息存储单元 2 中的声纹模型进行声纹匹

Petitioner Exhibit 1004, Page 670 of 1052

配，若匹配成功，则将与用户声纹模型关联的电视频道信息和唤醒操作指令输出给电视机主控制器 3，否则退出唤醒系统；

[0048] C5. 电视机主控制器 3 在成功收到语音控制系统 1 传递来的电视频道信息和唤醒操作指令后，执行电视开机操作并自动切换到对应用户喜爱的电视频道上。

[0049] 上述提到的预先设定的特定语音唤醒词可为任意设定的词组，如“打开电视机”、“开电视”、“打开电视”、“hello 电视”、“hello TV”等等。语音采集单元 11 可为麦克风或其他语音采集设备，用于感知周边的语音信号。

[0050] 本发明涉及到的声纹识别单元 13 所采用的识别算法为基于总变化因子（又称为 i-Vector）的说话人确认方法，其具体构建步骤如下：

[0051] 一、语音信号的预处理

[0052] 语音信号的预处理包括对语音信号进行预加重、加窗分帧和端点检测。对于经过 A/D 转换后的数字语音信号，预加重使用如下的一阶数字滤波器来提升语音信号的高频分量：

[0053] $H(z) = 1-0.97z^{-1}$

[0054] 然后采用如下式的汉明窗窗函数对语音信号进行平滑，并在加窗过程中进行语音信号的分帧操作，其中 N 为窗的长度。

[0055] $W(n) = 0.54+0.46\cos[2\pi n/(N-1)]$

[0056] 对于采样率为 8000Hz 的语音，分帧时使用的帧长为 25 毫秒，帧移为 10 毫秒。本发明采用的是基于短时能量的端点检测法对语音信号进行端点检测，从而判断语音信号的起始点和终止点，设 x(•) 为通过预加重后的语音信号，则表示从语音信号的第 n 个点开始加窗时的短时能量 E_n 计算公式为：

[0057] $$E_n = \sum_{m=-\infty}^{\infty}\left[x(m)W(n-m)\right]^2 = x^2(n) * W^2(n)$$

[0058] 二、声纹特征参数的提取

[0059] 声纹特征参数的提取即从说话人的语音信号中提取出能反映其身份个性的参数。本发明使用的是 39 维的感知线性预测系数 (Perception Linear Predictive Coefficients，PLP)，具体提取过程为：首先对上述步骤一预处理过后的语音做短时快速傅里叶变换，然后使用 24 阶子带的梅尔三角滤波器组进行临界频率分析，接着对得到的语音能量谱进行等响度曲线预加重用以模拟人耳的听觉感知系统对频率的敏感度，即将能量谱用如下式中的等响度曲线进行加权。

[0060] $$E(\omega) = \frac{(\omega^2 + 56.8\times10^6)\omega^4}{(\omega^2 + 6.3\times10^6)^2 \times (\omega^2 + 0.38\times10^9)(\omega^6 + 9.58\times10^{26})}$$

[0061] 然后使用指数为 1/15 的幂函数对能量进行压缩，接下来做离散傅里叶反变换，用 dubin 算法计算 M 阶全极点模型，最后对得到的模型系数进一步转换为倒谱系数得到 PLP 特征参数。根据端点检测结果对采集到的信号中为语音信号之外的信号，如噪声或静音帧对应的特征参数直接丢弃。

[0062] 三、声纹模型训练

[0063] 本发明中的用户声纹模型训练采用的是总变化因子建模技术，该技术的目标在于

Petitioner Exhibit 1004, Page 671 of 1052

对语音信号中含有的说话人身份信息和周围环境噪声、麦克风传输等信道信息变化空间进行同时建模。给定一段用户的语音，经过上述步骤一和二的处理后，其与说话人及信道相关的高斯混合模型 GMM 的均值超向量 S 表示如下：

[0064] $S = m + Tv + \varepsilon$

[0065] 其中 m 为通用背景模型的均值超向量，T 矩阵为总变化空间，ε 表示残差，v 矢量是与当前目标说话人（用户）及信道相关的总变化因子，它是一个低维矢量，代表着当前说话人的身份信息，即用户的声纹模型。ε 和 v 均服从高斯分布：

[0066] $\varepsilon \sim N(0, I)$

[0067] $v \sim N(0, \Sigma)$

[0068] 其中 Σ 为对角协方差矩阵。对于本发明而言，为了得到 v，需要事先采用与电视机语音唤醒场景近似的语音数据集合，利用期望最大化算法 (EM) 训练得到通用背景模型，得出总变化空间和残差向量。对于任意电视用户的语音，就可利用已训练好的 T，ε 和 m 以及由用户语音构建的高斯混合模型 S，求出每个用户的总变化因子，即声纹模型。

[0069] 四、声纹匹配

[0070] 本发明使用的基于总变化因子的声纹识别系统中，采用两个总变化因子 v_1 和 v_2 之间的余弦距离来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

[0071] $$d(\mathbf{v}_1, \mathbf{v}_2) = \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \cdot \|\mathbf{v}_2\|}$$

[0072] 如果 d 值大于预先设定的门限值则认为这两段语音来自同一个说话人，声纹匹配成功，否则匹配失败。有研究结果显示，通过概率线性判别分析规整后的总变化因子的鉴别能力更强，可进一步提升声纹识别的性能，但由于本发明主要适用于家庭的电视唤醒，家庭成员数目并不多，一般为 1 到 7 之间，因此上述基于总变化因子的声纹识别算法已足以保证本发明在声纹识别方面的准确率。

[0073] 下面介绍本发明中涉及到的语音识别单元所采用的语音识别算法

[0074] 由于本发明主要用于电视唤醒和频道切换，用户使用时涉及到的唤醒词和用于表示电视频道信息的词语数量都非常少，因此语音识别单元中采用的语音识别算法只需一套简单的基于 GMM-HMM（基于高斯混合模型的隐马尔可夫模型）识别系统即可，且训练和识别过程均采用离线方法，这样能够大大提高语音识别的速度。构建一套 GMM-HMM 语音识别系统主要包括声学特征参数的提取，声学模型和语言模型训练，语音解码过程。具体实现步骤如下：

[0075] 一、声学特征参数的提取

[0076] 由于感知线性预测系数中同时含有说话人的身份信息和语义信息，因此上述介绍的声纹识别算法中的 PLP 参数提取方法同样用于语音识别，但 PLP 参数的维数为 52 维，同时为了增强声学特征的区分能力，需要将 52 维的 PLP 参数进行异方差线性判别分析 (Heteroscedastic Linear Discriminant Analysis, HLDA) 后再用于声学模型建模和解码。

[0077] 二、声学模型训练

[0078] 声学模型的训练采用最大似然估计准则，需要用到期望最大化算法 EM 和

Petitioner Exhibit 1004, Page 672 of 1052

Baum-Welch 算法，前者用来处理隐马尔可夫模型中由状态序列隐变量带来的不完全数据下的训练问题，而后者用于从训练数据中高效的累积统计量，从而有效抽取模型参数更新所需要的信息。对于本发明的应用，在进行声学模型训练时，隐马尔可夫模型的状态数取 5，每个状态使用一个含 8 个混合高斯成分的 GMM 模型表示，第 j 状态的高斯混合密度表达式如下：

[0079] $$b_j(x)=\sum_{m=1}^{M}c_{jm}p_{jm}(x)$$

[0080] 其中 $$\sum_{m=1}^{M}c_{jm}=1,\quad c_{jm}\geq 0$$

[0081] $$p_{jm}(x)=\frac{1}{(2\pi)^{D/2}\left|\Sigma_{jm}\right|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu_{jm})^{T}\Sigma_{jm}^{-1}(x-\mu_{jm})\right\}$$

[0082] M = 8 表示高斯混合数目，x 为任意语音帧的特征参数，D 为特征的维数，c_{jm}，μ_{jm} 和 Σ_{jm} 为第 j 状态的 GMM 模型中第 m 个高斯分量的权重，均值向量和协方差矩阵。

[0083] 三、语言模型训练

[0084] 目前主流语音识别所用到的语言模型绝大部分是 n- 元 (n-gram) 统计语言模型，是将词串的生成过程表示为一个词间跳转的 n-1 阶马尔可夫过程。训练和评价 n-gram 语言模型的指标为下式的模型复杂度 (Perplexity)：

[0085] $$H=-\lim_{K\to\infty}\frac{1}{K}\log_2\left(P(w_1,\cdots,w_K)\right)\approx-\frac{1}{K}\sum_{k=1}^{K}\log_2\left(P(w_k\mid w_{k-1},w_{k-2},\cdots,w_{k-N+1})\right)$$

[0086] 其中 $w_1,\cdots,w_K$ 为含有 K 个词的词序列，N 为语言模型的元数，本发明中取值为 2。从理论上来看，语言模型对词串的期望复杂度应尽可能小，这表示语言模型在对给定的历史词串进行当前词预测时，能拥有更高的确信度。实际上语言模型的训练准则就是使其对训练集合中句子的复杂度最小化。

[0087] 四、语音解码

[0088] 在给定声学模型和语言模型的情况下，解码器的好坏与语音识别的速度直接相关，从而直接决定了用户的体验。虽然本发明用来唤醒电视所用的词语有限，由它们构成的解码网络并不大，但本发明仍然推荐采用基于加权有限状态转换 (WFST) 的 Viterbi 解码技术，它可以将声学、语言模型和发音字典进行有效的整合，对输入的语音信号 PLP 特征参数在词网络中进行搜索和匹配，从而找到统计意义上最匹配的词串作为最终的识别结果。

[0089] 上述语音识别单元中的声学模型和语言模型均可采用目前主流的开源工具 HTK 或者 Kaldi 提前训练好并存储起来，在进行语音唤醒时，仅需调用存储好的声学和语言模型，离线对采集到的用户语音进行 PLP 特征参数提取和解码即可。实验证明，本发明方法具有训练时间短，声纹和语音识别速度都很快，且识别率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。另外，本发明在电视关机时，电视整体处于掉电状态，即电视机主控制器、语音识别单元、声纹识别单元、信息存储单元均断电，只有语音采集单元实时采集用户的语音信号消耗电能，而

Petitioner Exhibit 1004, Page 673 of 1052

该模块耗电非常少，大大节省了电能消耗。

[0090] 综上所述，本发明提供的基于声纹识别和语音识别进行个性化电视语音唤醒的实现方法，能够在电视处于关机状态时，用户直接说类似于“hello 电视”这样的唤醒词，电视机就可以实现直接开机并立即显示自己最喜爱的电视频道。本发明的实现方法还可应用于电视机的用户个性化自动语音换台，同时可被移植到通用的 DSP 高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

Petitioner Exhibit 1004, Page 674 of 1052

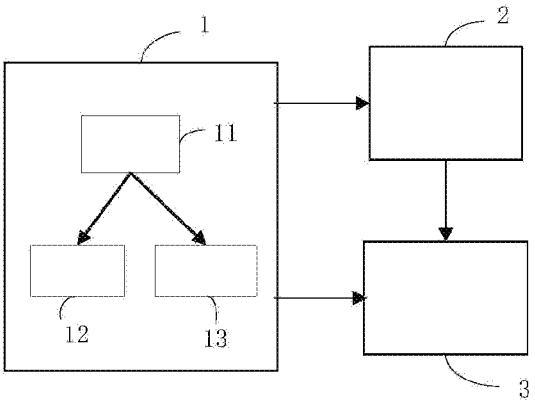


图 1

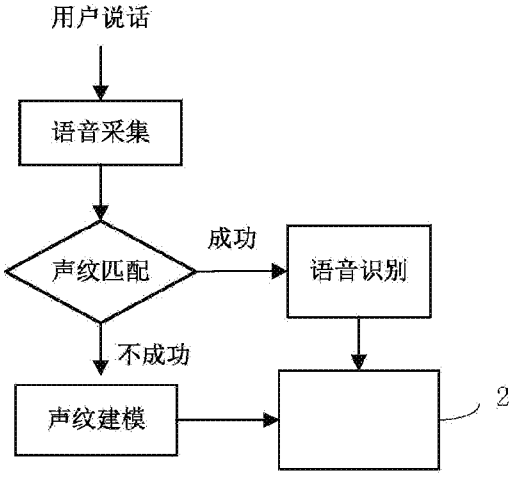


图 2

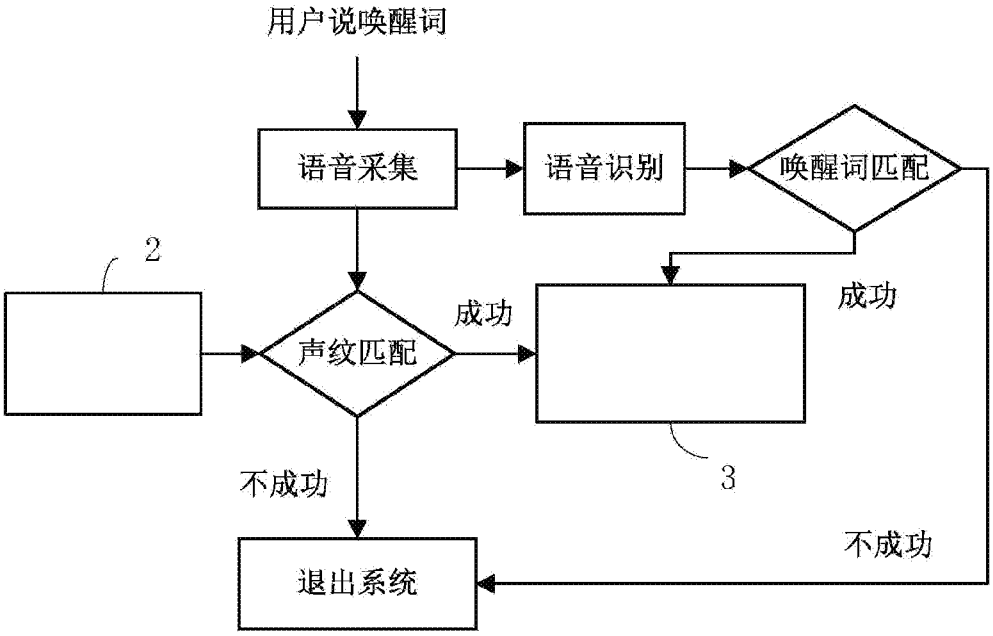


图 3

Petitioner Exhibit 1004, Page 675 of 1052

Publication number: CN105575395 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20160511
Application number: CN201410541444
Application date: 20141014
Priority: CN201410541444201410 14CN20141541444;
Assignee[std]: ZTE CORP ;
Inventor[std]: WANG JIE ;
International class[8]: G10L17/22; G10L15/22;
Family members: CN105575395 A

Title: VOICE WAKE-UP METHOD AND APPARATUS, TERMINAL, AND PROCESSING METHOD THEREOF

Title: 语音唤醒方法及装置、终端及其处理方法

Abstract:

The invention discloses a voice wake-up method and apparatus, a terminal, and a processing method thereof. The voice wake-up method comprises: a wake-up word voice signal sent out by a terminal user is collected and a wake-up word identification result and a vocal print feature of the wake-up word voice signal are obtained and stored; when a terminal is in a sleep state, a voice signal inputted externally is detected; a voice identification result and a vocal print feature of the voice signal are obtained; the voice identification result and the stored wake-up word identification result as well as a matching result of the vocal print feature and the stored vocal print feature are obtained; and according to the matching result, whether the terminal needs to be woken up is determined; and if so, the terminal is woken up. With the method, technical problems of low user wake-up rate, high wrong wake-up rate, and poor user experience according to the existing voice wake-up technology are solved.

Abstract:

本发明公开了一种语音唤醒方法及装置、终端及其处理方法。本发明的语音唤醒方法包括：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；在所述终端处于休眠状态下，检测外界输入的语音信号；获取所述语音信号的语音识别结果和声纹特征；获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端；本发明的语音唤醒方法能够

The present invention discloses a method and apparatus for voice, terminal and processing method to wake up. Voice wake-up method of the present invention includes: a wake-word speech signal acquisition end-user issue, get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and stored; the terminal is in a dormant state, the detection and external input speech signal; acquiring the speech recognition result voiceprint characteristics of the speech signal; obtaining the wake-up word speech

解决现有语音唤醒技术存在的用户唤醒率低、误唤醒率高以及用户体验差的技术问题。

recognition result of the recognition result and storage, as well as matches the voiceprint and voiceprint stored; according to the match results to determine whether the need to wake up the terminal, and if so, to wake the terminal; the voice of the present invention is a wake-up method can solve the prior art existing user voice wake wake low error rate and wake poor user experience technical problems.

Claims:

1.一种语音唤醒方法，其特征在于，包括如下步骤：

采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词 识别结果和声纹特征并存储；

在所述终端处于休眠状态下，检测外界输入的语音信号；

获取所述语音信号的语音识别结果和声纹特征；

获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存 储的声纹特征的匹配结果；

根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端。

2.如权利要求1所述的方法，其特征在于，在检测外界输入的语音信 号之后，获取所述语音信号的语音识别结果和声纹特征之前，还包括：

判断所述语音信号是否满足预设信号要求，若是，获取所述语音信号的语 音识别结果和声纹特征。

3.如权利要求2所述的方法，其特征在于，所述判断所述语音信号是 否满足预设信号要求的步骤包括：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行 匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

4.如权利要求3所述的方法，其特征在于，所述将所述用户语音数据 和所述背景噪声数据分别与预置的背景噪声模型进行匹配的步骤包括：

.

1 A speech wake method comprising the steps of: acquisition wake-word speech signals sent by the end-user, access to the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

In the terminal is in a dormant state, detection and external input speech signal;

Being the speech signal a speech recognition result voiceprint feature;

Get the wake-up word speech recognition result of the recognition result and storage, as well as matches the voiceprint and voiceprint characteristic features stored;

According to the matching results to determine whether the need to wake up the terminal, and if so, to wake the terminal.

Before

2. The method of claim 1 or claim 2, characterized in that, after the detection of the external input speech signal, the speech recognition result to obtain the speech signal and a voiceprint characteristic, further They include:

Analyzing the voice signal meets a preset signal requirements, if, in the speech signal to obtain speech recognition results and a voiceprint characteristic.

.

3 The method as claimed in claim 2, characterized in that the determination whether the speech signal is a signal meets a preset requirement comprises: of the speech signal is processed to obtain user data and voice background noise data;

将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分 值；

将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分 值；

所述根据匹配结果判断所述语音信号是否满足预设信号要求的步骤包括：

采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到 所述语音信号的匹配分值；

判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设 信号要求，若否，则判定不满足预设信号要求。

5.如权利要求1-4任一项所述的方法，其特征在于，所述根据所述匹 配结果判断是否需要唤醒所述终端的步骤包括：

当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成 功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时， 则判定需要唤醒所述终端。

6.如权利要求2-4任一项所述的方法，其特征在于，所述判断所述语 音信号是否满足预设信号要求的步骤包括：

多核音频处理器中的第一核单元唤醒后，所述第一核单元判断所述语音信 号是否满足预设信号要求；

所述获取所述语音信号的语音识别结果和声纹特征的步骤包括：

当所述第一核单元判断为是时，所述第一核单元唤醒所述多核音频处理器 中的第二核单元；

所述第二核单元获取所述语音信号的语音识别结果和声纹特征；

所述获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征 与存储的声纹特征的匹配结果的步骤包括：

所述第二核单元获取所述语音识别结果与存储的唤醒词识别结果、以及所 述声纹特征与存储的声纹特征的匹配结果；

所述根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终 端的步骤包括：

the user data and voice data of the background noise and background noise models are preset to match; determine whether the speech signal according to the matching result satisfies preset signal requirements.

4. The method as claimed in claim 3, characterized in that said voice data to said user data and said background noise and background noise models are preset to match steps include: the user voice data and background noise model of the pre-match points to get the first matching value;

The data of the background noise and background noise model of the pre-match match points to obtain a second value;

According to the matching results to determine whether the speech signal meets a preset signal required steps include:

With the default value of the first matching algorithm score and the second match scores were calculated match score of the speech signal; match score is determined whether the speech signal is greater than a preset threshold, if it is determined to meet the requirements of a preset signal, and if not, it is determined that the signal does not meet the preset requirements.

Methods

5 as claimed in claim any one of claims 1-4, characterized in that the said matching results according to the procedure judges whether the wake-up terminal comprising:

When the speech recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the voiceprint of the speech signal stored speech wakeup word voiceprint match is found, determination need to wake up the terminal.

.

6 The method as claimed in one any of claim 2-4, wherein said determining step includes the required preset signal of said speech signal meets:

After the first multi-core audio processor core unit wakes up, the first nuclear unit judges that the

所述第二核单元根据所述匹配结果判断是否需要唤醒所述终端，若是，则 所述第二核单元唤醒所述终端。

7.如权利要求6所述的方法，其特征在于，还包括：

当所述第一核单元判断为否时，所述第一核单元进入休眠状态；

和/或

当所述第二核单元判断为否时，所述第二核单元进入休眠状态。

8.如权利要求6所述的方法，其特征在于，所述第二核单元唤醒所述 终端的步骤包括：

所述第二核单元发送中断事件唤醒所述终端中处于休眠状态的主处理器。

9.一种终端处理方法，其特征在于，包括如下步骤：

利用权利要求1-5任一项所述的语音唤醒方法唤醒所述终端；

所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设 操作。

10.如权利要求9所述的方法，其特征在于，所述唤醒所述终端的步骤 包括：

根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

所述语音信号的语音识别结果执行预设的操作的步骤包括：

根据所述唤醒中断事件的类型执行对应的预设操作。

11.一种语音唤醒装置，其特征在于，包括：采样模块、检测模块、第 一获取模块、第二获取模块、第一判断模块和唤醒模块；

所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词 语音信号的唤醒词识别结果和声纹特征并存储；

所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

speech signal meets the preset signal requirements;

Get the voiceprint features speech recognition result of the speech signal comprises:

When the first nuclear unit is judged to be the first nuclear unit of the multi-core audio processor in the wake of a second nuclear unit;

The second nuclear unit acquired by the speech recognition result of the speech signal voiceprint feature;

Get the wake of the speech recognition result of the word recognition result is stored, and the matching results of the step of the voiceprint stored voiceprint features include:

The second nuclear unit acquires wake word recognition result of the speech recognition result and storage, as well as matches the voiceprint characteristic with the stored voiceprint said;

According to the result of the match is determined whether the terminal needs to wake up, if the step of the terminal include wake-up:

The second nuclear unit is required to wake the terminal according to the matching result of the determination, if so, the second nuclear unit to wake the terminal.

.

7 The method as claimed in claim 6, characterized in that, further comprising:

When the first core unit judgment is NO, the first nuclear unit goes to sleep; and / or

When the second nuclear unit judgment is NO, the second nuclear unit into hibernation.

.

8 as claimed in claim 6, characterized in that said second core said terminal means comprises wake-up steps:

The second nuclear unit sends an interrupt event to wake the dormant terminal in the main processor.

.

所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、 以及所述声纹特征与存储的声纹特征的匹配结果；

所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

12.如权利要求11所述的装置，其特征在于，还包括：第二判断模块；

所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第 一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信 号的语音识别结果和声纹特征。

13.如权利要求12所述的装置，其特征在于，所述第二判断模块用于：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行 匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

14.如权利要求12或13所述的装置，其特征在于，所述第一获取模块、 第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音频处理器的第 二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是 否满足预设信号要求，若是，唤醒所述第二核单元；

所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的 语音识别结果和声纹特征。

15.如权利要求11-13任一项所述的装置，其特征在于，所述第一判断 模块用于当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹 配成功，且所述语音信号的声纹特征与所述存储的语音

9 A terminal processing method, characterized by comprising the steps of:

Voice wake use claims 1-5 method according to any one of the wake-up terminal;

After the wake-up terminal, perform the corresponding operation according to preset voice recognition result of the speech signal.

.

10 The method as claimed in claim 9, characterized in that said terminal comprises the step of wake-up: generated based on the recognition result of the speech signal preset type of wake-up interrupt events to wake the terminal;

The default action steps speech recognition result of the speech signal include: type of interrupt event to perform the corresponding operation according to the preset wake.

.

11 A speech wake apparatus, characterized in that, comprising: a sampling module, a detection module, a first acquisition module, a second acquisition module, and a first judgment module wakeup module;

The sampling module is used to wake-word speech signal acquisition end-user issue, get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

The detection module for the terminal is in a dormant state, the detection of the external input speech signal;

The first acquisition module for acquiring speech recognition result of the speech signal voiceprint feature;

The second acquisition module for acquiring the speech recognition result wakeup word recognition result and storage, as well as matches the voiceprint and voiceprint stored;

The first module is used for judging according to the matching results to determine whether to wake the terminal; in the wake of the first judging module for judging module is YES, the wake-up terminal.

唤醒词声纹特征匹配成 功时，则判定需要唤醒所述终端。

16.一种终端，其特征在于，包括：处理模块和如权利要求11-13任一 项所述的语音唤醒装置；

所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音 信号的语音识别结果执行对应的预设操作。

17.如权利要求16所述终端，其特征在于，所述语音唤醒装置中唤醒模 块用于根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

.

12 apparatus as claimed in claim 11, characterized by further comprising: a second judging module;

Before After

in the second determining module for external detection module input speech signal, the first acquisition module acquiring the speech recognition result of the speech signal and a voiceprint characteristic, judging whether the speech signal satisfies preset signal requirements; the first module for obtaining said second determination means determines YES, obtaining a speech recognition result of the speech signal and a voiceprint characteristic.

.

13 The device of claim 12, characterized in that said second judging module for: of the speech signal is processed to obtain user data and voice background noise data; the user data and voice data of the background noise and background noise models are preset to match; determine whether the speech signal according to the matching result satisfies preset signal requirements.

14. Apparatus as claimed in claim 12 or 13, characterized in that said first acquisition module, a second acquisition module, the judging module and the first module wakeup integrated multi-core audio processor of the second nuclear unit; the second judgment module integrated in the multi-core audio processor's first core units;

The second determination module is used after the first wake-up core unit, it determines whether the speech signal is a signal complies with the preset requirement, and if so, waking the second core means;

The first acquisition module for, after said second core means wake, the speech signal to obtain a speech recognition result and a voiceprint characteristic.

15. Apparatus according to any one of claim 11-13, wherein said first determining module is used for the speech recognition result when the speech signal and wake word stored speech recognition result matches, and the voiceprint of the speech signal stored speech wakeup word voiceprint

match is successful, it is determined that the terminal needs to wake up.

.

16 A terminal, characterized by comprising: a speech according to any one of the processing modules 11-13 and a wake-up apparatus as claimed in claim;

After the processing module for the voice wake wake the terminal apparatus, performs the corresponding operation according to the preset speech recognition result of the speech signal.

17. The terminal as claimed in claim 16, wherein said speech wake-up means for generating a wake-up module wakeup preset type of the identification result of the speech signal interrupt event wake the terminal;

The processing module is used to perform the type of interrupt events corresponding operation according to the preset wake.

Description:

语音唤醒方法及装置、终端及其处理方法

技术领域

本发明涉及语音识别技术领域，尤其涉及一种语音唤醒方法及装置、终端 及其处理方法。

背景技术

语音在手机终端中已不单单存在于单独的程序应用中，而开始慢慢的作为 一种独立于手指触摸输入的新增输入方式，不需触碰手机，遥控和更智能化友好的服务于手机用户。

而语音唤醒技术为用户解放双手完全的语音指令遥控手机提供了前提，唤 醒主流技术是技术软件唤醒，但软件运行在手机系统启动为前提，为保证能随 时随地接收到用户的语音指令，语音引擎需要一直后台运行和监听，手机系统 将无法进入休眠待机的节电状态，此类语音唤醒技术的应用对电量消耗异常强 烈。目前语音技术在市场技术应用中在不单单考虑到功能友好性，还要关注手 机电量消耗，同时要求对用户使用中的不被语音过多打扰，所以目前各厂商对 于语音识别系统的唤醒方案多数采用休眠状态下硬件按键唤醒或者手机系统运 行时软件按钮点击唤醒。

Voice wake-up method and apparatus, terminal and approach

FIELD

The present invention relates to the field of speech recognition technology, particularly to a method and apparatus for voice wake, terminal and processing methods.

BACKGROUND voice in mobile terminals has not only exist in a separate program applications, and slowly began to add as an independent input finger touch input, without touching the phone, remote control and more intelligent and friendly service to mobile phone users.

The Voice wake-up technology to provide users with complete hands-free voice commands remote control handset provides premise wake wake mainstream technology software, but the software running on premise phone system startup, is guaranteed to receive the user's voice commands anytime, anywhere voice engine needs to run in the background and has been listening, mobile phone system will not enter standby power-saving sleep state, these wake-up voice application

随着技术人员的不断改进，目前已经出现语音低功耗唤醒技术，在手机休 眠状态下应用低功耗的芯片实现直接用语音指令唤醒手机操作系统，无需用户手动进行硬件和软件操作。

但对于现有技术的语音低功耗唤醒功能中，只能使用固定的唤醒词，通过 语音模型生成此预置固定唤醒词的声学模型。具体地现有语音低功耗唤醒功能的实现方法，包括如下：

录制一些大批量的"选定唤醒词"声音样本进行处理和人群声学特征培训 和提取。

将统计估算得出的"固定的唤醒词"声音特征预置在手机系统中。

根据用户个体输入的声音波形与统计生成的声音特征进行比对。

因低功耗性能音频芯片中运算处理能力有限，群体样本统计生成的声学模 型在单独用户个体使用时会有以下问题：

1、用户个体的发声方式差异性远远低于或超出采样统计模型特征提取的置 信区间，造成用户唤醒率低。如果样本量太小或者或人群特征太单一，例如， 全是普通话，全是成年人，会造成有些特定人群的声音例如儿童，老人的声音 对唤醒失效。

2、在噪音环境下复杂噪音环境生成了采样统计模型中类似的似然曲线，造 成误唤醒率高。不该唤醒时乱唤醒，造成用户困扰。

3、由于只能通过厂家事先录制的"固定的唤醒词"来唤醒终端，用户不能

够自定义唤醒词，不能满足用户个性化的需求，降低了用户体验。

因此，现有技术虽然实现了一定功耗控制下的语音唤醒，但是存在用户唤 醒率低、误唤醒率高以及用户体验差的技术等问题。

发明内容

本发明要解决的主要技术问题是，提供一种语音唤醒方法及装置、终端及 其处理方法,能够解决现有语音唤醒技术存在的用户唤醒率低、误唤醒率高以及 用户体验差的技术问题。

为解决上述技术问题，本发明提供一种语音唤醒方法，包括如下步骤：

technology for power consumption unusually strong. Speech technology currently in the market not only in technology application taking into account the functional-friendliness, but also concerned about the mobile phone power consumption, and to ask for too much is not the voice of the user to bother, so now most of the manufacturers wake up programs for speech recognition system use hibernation wakeup hardware keys or cell phone system is running software button clicks wake.

With the continuous improvement of technical personnel, has now appeared speech Low-power wake-up technology, the application of low-power sleep state in the mobile phone chip for direct use voice commands to wake the phone operating system, without requiring the user to manually operate the hardware and software.

However, the prior art voice low-power wake-up function, you can only use a fixed wake-word speech model generated by this preset fixed wake-acoustic model of the word. Specifically, existing voice low-power wake-up implementation functions, including the following:

Record number of large-volume "Selected wake word" crowd sound sample processing and acoustic features training and extraction.

The statistics estimated from a "fixed wake-up word" sound characteristics preset in the phone system.

The sound waveform individual user input and generate sound statistical characteristics for comparison. due to the low power performance audio chip arithmetic processing capacity is limited, the sample population statistics generated acoustic model will be in a separate individual users use the following questions:

1, the user's individual sound pattern difference is far below or above sampling statistical models feature extraction confidence interval, causing the user to wake low. If the sample size is too small or too simple or population characteristics, for example, all Mandarin, all adults, will result in sound Some specific groups, such as the voices of children, the elderly fail to wake up.

2, in noisy environments generate complex noise environment sampling statistical model similar likelihood curve, causing waking up rate. When

采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词 识别结果和声纹特征并存储；

在所述终端处于休眠状态下，检测外界输入的语音信号；

获取所述语音信号的语音识别结果和声纹特征；

获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存 储的声纹特征的匹配结果；

根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端。

进一步地，所述方法在检测外界输入的语音信号之后，获取所述语音信号 的语音识别结果和声纹特征之前，还包括：判断所述语音信号是否满足预设信号要求，若是，获取所述语音信号的语音识别结果和声纹特征。

进一步地，所述判断所述语音信号是否满足预设信号要求的步骤包括：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行 匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

进一步地，所述将所述用户语音数据和所述背景噪声数据分别与预置的背 景噪声模型进行匹配的步骤包括：

将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分 值；

将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分 值；

所述根据匹配结果判断所述语音信号是否满足预设信号要求的步骤包括：

采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到 所述语音信号的匹配分值；

判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设 信号要求，若否，则判定不满足预设信号要求。

should wake wake chaos, causing the user problems.

3, since the only pre-recorded "fixed wake-up word" by manufacturers to wake up the terminal, the user can not

Custom enough wake-up word, can not meet the needs of individual users, reducing the user experience.

Thus, although the prior art to achieve a certain power wake-up voice control, but there are a wake-up users is low, error rate, and wake up poor user experience technical problems.

Summary

The main technical problem to be solved by the present invention is to provide a method and apparatus for voice, terminal and approach wake, can solve the existing user voice technology exists wake wake low error rate and wake poor user experience technical problems.

To solve the above technical problem, the present invention provides a method for wake-up speech, comprising the steps of: acquisition wake-word speech signals sent by the end-user, access to the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

In the terminal is in a dormant state, detection and external input speech signal;

Being the speech signal a speech recognition result voiceprint feature;

Get the wake-up word speech recognition result of the recognition result and storage, as well as matches the voiceprint and voiceprint characteristic features stored;

According to the matching results to determine whether the need to wake up the terminal, and if so, to wake the terminal.

Before

Further, after detecting the speech signal the method of input, the speech recognition result to obtain a voiceprint characteristic of the speech signal, further comprising: determining whether the voice signal meets the requirements of a preset

进一步地，所述根据所述匹配结果判断是否需要唤醒所述终端的步骤包括：

当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成 功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时， 则判定需要唤醒所述终端。

进一步地，所述判断所述语音信号是否满足预设信号要求的步骤包括：

多核音频处理器中的第一核单元唤醒后，所述第一核单元判断所述语音信号是 否满足预设信号要求；

所述获取所述语音信号的语音识别结果和声纹特征的步骤包括：

当所述第一核单元判断为是时，所述第一核单元唤醒所述多核音频处理器 中的第二核单元；

所述第二核单元获取所述语音信号的语音识别结果和声纹特征；

所述获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征 与存储的声纹特征的匹配结果的步骤包括：

所述第二核单元获取所述语音识别结果与存储的唤醒词识别结果、以及所 述声纹特征与存储的声纹特征的匹配结果；

所述根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终 端的步骤包括：

所述第二核单元根据所述匹配结果判断是否需要唤醒所述终端，若是，则 所述第二核单元唤醒所述终端。

进一步地，所述方法还包括：

当所述第一核单元判断为否时，所述第一核单元进入休眠状态；

和/或

当所述第二核单元判断为否时，所述第二核单元进入休眠状态。

进一步地，所述第二核单元唤醒所述终端的步骤包括：

所述第二核单元发送中断事件唤醒所述终端中处于休眠状态的主处理器。

signal, and if so, Get the speech signal a speech recognition result voiceprint features.

Furthermore, the judgment of the speech signal meets the requirements of preset signal comprises: of the speech signal is processed to obtain user data and voice background noise data; the user data and voice data of the background noise and background noise models are preset to match; determine whether the speech signal according to the matching result satisfies preset signal requirements.

Step

Further, the voice data and the user data of the background noise, respectively preset to match the background noise model include: the user voice data and background noise model of the pre-match points to get the first matching value;

The data of the background noise and background noise model of the pre-match match points to obtain a second value;

According to the matching results to determine whether the speech signal meets a preset signal required steps include:

With the default value of the first matching algorithm score and the second match scores were calculated match score of the speech signal; match score is determined whether the speech signal is greater than a preset threshold, if it is determined to meet the requirements of a preset signal, and if not, it is determined that the signal does not meet the preset requirements.

Furthermore, according to the results to determine whether the steps need to wake up the terminal includes the matching:

When the speech recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the voiceprint of the speech signal stored speech wakeup word voiceprint match is found, determination need to wake up the terminal.

Furthermore, the judgment of the speech signal meets the requirements of preset signal comprises:

After the first multi-core audio processor core unit wakes up, the first nuclear unit determines

同样为了解决上述的技术问题，本发明还提供了一种终端处理方法，包括 如下步骤：

利用上述前五项中任一项所述的语音唤醒方法唤醒所述终端；

所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设 操作。

进一步地，所述唤醒所述终端的步骤包括：

根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终 端；

所述语音信号的语音识别结果执行预设的操作的步骤包括：

根据所述唤醒中断事件的类型执行对应的预设操作。

同样为了解决上述的技术问题，本发明还提供了一种语音唤醒装置，包括： 采样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模 块；

所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词 语音信号的唤醒词识别结果和声纹特征并存储；

所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、 以及所述声纹特征与存储的声纹特征的匹配结果；

所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

进一步地，所述装置还包括：第二判断模块；

所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第 一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信 号的语音识别结果和声纹特

whether the speech signal is preset to meet the signaling requirements;

Get the voiceprint features speech recognition result of the speech signal comprises:

When the first nuclear unit is judged to be the first nuclear unit of the multi-core audio processor in the wake of a second nuclear unit;

The second nuclear unit acquired by the speech recognition result of the speech signal voiceprint feature;

Get the wake of the speech recognition result of the word recognition result is stored, and the matching results of the step of the voiceprint stored voiceprint features include:

The second nuclear unit acquires wake word recognition result of the speech recognition result and storage, as well as matches the voiceprint characteristic with the stored voiceprint said;

According to the result of the match is determined whether the terminal needs to wake up, if the step of the terminal include wake-up:

The second nuclear unit is required to wake the terminal according to the matching result of the determination, if so, the second nuclear unit to wake the terminal.

Further, the method further comprises:

When the first core unit judgment is NO, the first nuclear unit goes to sleep; and / or

When the second nuclear unit judgment is NO, the second nuclear unit into hibernation.

Further, the second nuclear unit to wake the terminal comprises:

The second nuclear unit sends an interrupt event to wake the dormant terminal in the main processor.

The same technique to solve the above problems, the present invention also provides a method for processing a terminal, comprising the steps of: using voice above the top five in any one of the wake-up wake up the terminal;

After the wake-up terminal, perform the corresponding operation according to preset voice

征。

进一步地，所述第二判断模块用于：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行 匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

进一步地，所述第一获取模块、第二获取模块、所述第一判断模块和所述 唤醒模块集成在多核音频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是 否满足预设信号要求，若是，唤醒所述第二核单元；

所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的 语音识别结果和声纹特征。

进一步地，所述第一判断模块用于当所述语音信号的语音识别结果与所述 存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

同样为了解决上述的技术问题，本发明还提供了一种终端，包括处理模块 和上述所述的语音唤醒装置；

所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音 信号的语音识别结果执行对应的预设操作。

进一步地，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结 果产生预设类型的唤醒中断事件唤醒所述终端；

所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

本发明的有益效果是：

本发明提供了一种语音唤醒方法及装置、终端及其处理方法，能够提高用 户唤醒率、降低误唤醒率和提升用户体验。具体地，本发明的语音唤醒方法包括：采集终端用户发出的唤醒词语音信号，获取所

recognition result of the speech signal.

Further, the wake of the terminal comprises: generate preset type of wake-up interrupt events to wake the terminal according to the recognition result of the speech signal;

The default action steps speech recognition result of the speech signal include: type of interrupt event to perform the corresponding operation according to the preset wake.

Similarly to solve the above technical problem, the present invention also provides a wake-speech apparatus comprising: a sampling module, a detection module, a first acquisition module, a second acquisition module, and a first judgment module wakeup module;

The sampling module is used to wake-word speech signal acquisition end-user issue, get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

The detection module for the terminal is in a dormant state, the detection of the external input speech signal;

The first acquisition module for acquiring speech recognition result of the speech signal voiceprint feature;

The second acquisition module for acquiring the speech recognition result wakeup word recognition result and storage, as well as matches the voiceprint and voiceprint stored;

The first module is used for judging according to the matching results to determine whether to wake the terminal; in the wake of the first judging module for judging module is YES, the wake-up terminal.

Furthermore, the apparatus further comprises: a second judging module;

Before After

in the second determining module for external detection module input speech signal, the first acquisition module acquiring the speech recognition result of the speech signal and a voiceprint characteristic, judging whether the speech signal satisfies preset signal requirements;

述唤醒词语音信号的唤醒词 识别结果和声纹特征并存储；在所述终端处于休眠状态下，检测外界输入的语 音信号；获取所述语音信号的语音识别结果和声纹特征；获取所述语音识别结 果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果； 根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端；本发 明的语音唤醒方法，以终端用户自己声音作为声音样本，获取终端用户个体的 声音特征(即识别结果和声纹特征)作为声音样本特征；在进行语音唤醒时， 由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同(同一用户的发声 方式相同)，避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤 醒失败，提供了唤醒率；另外，本发明的方法加入了语音识别结果和声纹特征 的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端， 降低了误唤醒率；本发明的方法还可以供用户自定义唤醒词，满足用户个性化 的需求，提升了用户的体验。

另外，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施 例方法只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信 息安全。

附图说明

图1为本发明实施例一提供的一种语音唤醒方法的流程示意图；

图2为本发明实施例一提供的一种获取用户自定义语音唤醒词样本特征的 流程示意图；

图3为本发明实施例一提供的一种判断语音信号是否满足预设信号要求的 流程示意图；

图4为本发明实施例一提供的一种语音唤醒系统的结构示意图；

图5为本发明实施例一提供的一种语音唤醒系统进行语音唤醒的流程示意 图；

图6为本发明实施例二提供的一种终端处理方法的流程示意图；

图7为本发明实施例三提供的第一种语音唤醒装置的结构示意图；

图8为本发明实施例三提供的第二种语音唤醒装置的结构示意图；

图9为本发明实施例三提供的第三种语音唤醒装置的结构示意图；

the first module for obtaining said second determination means determines YES, obtaining a speech recognition result of the speech signal and a voiceprint characteristic.

Further, the second judging module for: of the speech signal is processed to obtain user data and voice background noise data; the user data and voice data of the background noise and background noise models are preset to match; determine whether the speech signal according to the matching result satisfies preset signal requirements.

Furthermore, the first acquisition module, a second acquisition module, the judging module and the first module is integrated in the wake of the second core multicore audio processor unit; a second judgment module integrated the first multi-core audio processor core unit;

The second determination module is used after the first wake-up core unit, it determines whether the speech signal is a signal complies with the preset requirement, and if so, waking the second core means;

The first acquisition module for, after said second core means wake, the speech signal to obtain a speech recognition result and a voiceprint characteristic.

Further, the first determining module is used when the speech recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the voiceprint of the speech signal with the stored when the wake-up word voice voiceprint match is successful, it is determined that the terminal needs to wake up.

The same technique to solve the above problems, the present invention also provides a terminal, comprising a processing module and said wake-up means of said speech;

After the processing module for the voice wake wake the terminal apparatus, performs the corresponding operation according to the preset speech recognition result of the speech signal.

Further, the voice wake-up means for generating a wake-up module preset type of wake-up interrupt wake-up event of the terminal based on the identification result of the speech signal;

图10为本发明实施例三提供的一种终端的结构示意图。

具体实施方式

下面通过具体实施方式结合附图对本发明作进一步详细说明。

实施例一：

考虑到现有固定唤醒词语音唤醒方法存在的用户唤醒率低、误唤醒率高以 及用户体验差的技术问题；本实施例提供了一种可以提高终端用户唤醒率、降低误唤醒率、提升用户体验的语音唤醒方法，如图1所示，本实施例的语音唤 醒方法包括如下步骤：

步骤101：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号 的唤醒词识别结果和声纹特征并存储。

本实施例方法可以采用语音采集装置，例如麦克风来采集终端用户发出的 语音信号。本实施例方法中获取唤醒词识别结果和声纹特征的时序不限，可以先获取识别结果，后获取声纹特征，也可以同时获取。

本实施例中唤醒词识别结果是对唤醒词语音信号进行语音识别得到的唤醒 词文字内容或者命令。本实施例中声纹特征是语音信号中可以代表用户发声方式的参数，例如声音频率、响度等等。

本步骤为进行语音唤醒之前的准备步骤，发生在用户自定义语音唤醒词的 过程中，本步骤可以由终端的AP侧执行，即终端的主芯片，例如手机主CPU中 ARM芯片来完成。

为了提高语音唤醒的精确性，本实施例方法在采集样本特征的阶段，可以 多次录制用户发出的唤醒词语音信号，然后根据多次录制的唤醒词语音信号得到最优的唤醒词识别结果和声纹特征。如图2所示，为终端获取用户自定义语 音唤醒词样本特征的过程，由终端的AP侧完成:

在终端启动用户自定义唤醒词设置之后，第一次录制用户所说的唤醒词语 音信号，然后生成唤醒词语音信号的唤醒词识别结果1和声纹特征1，将识别结果1和声纹特征1存储至自定义唤醒词存储区中；

终端第二次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的 唤醒词识别结果2和声纹特征2，当唤醒词识别结果2与唤醒词识别结果1相同时，利用声纹特征2优化声纹特征1，当唤醒词识别

The processing module is used to perform the type of interrupt events corresponding operation according to the preset wake.

Advantageous Effects

The present invention is:

The present invention provides a method and apparatus for voice, terminal and approach wake, wake-up can improve the user rate and reduce the error rate and enhance the user experience wake. In particular, the present invention is the voice wakeup method comprising: signal acquisition wake-word speech issued by the end user to get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and stored; the terminal is in a sleep state, detection external input speech signal; obtaining a speech recognition result voiceprint characteristics of the speech signal; obtaining the wake-up word speech recognition result of the recognition result and storage, as well as matches the voiceprint features and characteristics of the voiceprint stored; according to the matching results to determine whether the need to wake up the terminal, and if so, to wake the terminal; the voice of the present invention, methods for waking up to their voice as a sound sample end-user, end-users get the sound characteristics of individuals (ie, the recognition result harmony pattern feature) as an audio sample characteristics; wake during voice because sound sample features voiceprint and voiceprint wakeup extraction (same sound the same user mode), to avoid the user's individual sound pattern difference in sound Sample user's sound pattern caused wake-up failed to provide a wake-up rate; in addition, the method of the present invention is added to the speech recognition result voiceprint matching characteristics, can effectively avoid the noisy environment caused by waking up, so that the noise environment can not wake up the terminal, reduces the error rate to wake; the method of the present invention can also provide custom wake-up word, to meet the needs of individual users, to enhance the user experience.

In addition, because different users different sound pattern is different, that voiceprint, the application of the present embodiment method only to wake up my end user terminal, compared with the prior art, but also to improve the information security of the terminal.

结果2与唤醒词识别结果1 不相同时，重新录制；进入第三次录制或者从第一次开始重新录制；

终端第三次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的 唤醒词识别结果3和声纹特征3，当唤醒词识别结果3与唤醒词识别结果1相同时，利用声纹特征3优化上述已经优化过的声纹特征1；

依次类推，直到第n次录制完成。图2中n取4，本实施例方法中n的取值 可以根据实际需求进行设定。

步骤102：在所述终端处于休眠状态下，检测外界输入的语音信号。

本步骤中可以由终端的麦克风来检测外界的语音信号，在终端休眠状态时， 麦克风可以保持工作状态。

本步骤中终端处于休眠状态下从硬件的角度可以包括：终端的主处理器处 于休眠状态，从软件的角度可以包括：终端的操作系统处于休眠或锁定状态。

步骤103：获取所述语音信号的语音识别结果和声纹特征。

本步骤中获取语音识别结果和声纹特征没有时序限制，可以先获取语音识 别结果，后获取声纹特征，也可以先获取声纹特征，后获取语音识别结果，还可以同时获取语音识别结果和声纹特征。

本步骤中语音信号的语音识别结果是对语音信息进行语音识别得到的文字 内容或者命令。本实施例中声纹特征是语音信号中可以代表用户发声方式的参数，例如声音频率、响度等等。

步骤103-106为语音唤醒过程中主要步骤，其可以在一个低功耗音频芯片 执行，例如低功耗的DSP芯片。在进行语音唤醒之前，事先将上述获取的唤醒词识别结果和唤醒词语音信号的声纹特征下载到低功耗音频芯片的存储区，为 语音唤醒做准备。

步骤104：获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹 特征与存储的声纹特征的匹配结果。

本步骤中获取语音识别结果与存储的唤醒词识别结果的匹配结果方式包 括：将语音识别结果与存储的唤醒词识别结果进行匹配，获取所述声纹特征与 存储的声纹特征的匹配结果的方式包括：将获取的声纹特征与存储的声纹特征 进行匹配；本步骤中获取所述语音结果与存储的唤醒词识别结果匹配结果，与 获取声纹特征与存储的声纹特征的匹配结果过程

Brief Description of the

Figure 1 embodiment of the present invention to a speech provided a wake-up schematic flow diagram of a method embodiment;

2 the present invention method for obtaining user-defined voice wake-word sample characteristics provide a flow diagram illustrating the embodiment;

3 embodiment of the present invention method for judging whether the speech signal example of a preset signal provides a schematic flow diagram meet the requirements;

Figure 4 is a schematic of one configuration of the invention to provide a voice Example wake implementation of the system;

Figure 5 embodiment of the present invention to a speech provided a wake-up system for voice wake schematic flow diagram embodiment;

6 a schematic diagram of the process according to a second embodiment of a terminal to treatment;

Figure 7 is a schematic view of a first configuration of the invention to provide a wake-up voice Example Three apparatus embodiment;

8 a schematic view of a second structure of the present invention to provide a wake-up voice Example Three apparatus embodiment;

FIG. 9 is a schematic view of the structure of the invention to provide a third voice Example Three wakeup device embodiment;

10 the present invention structure diagram of an embodiment of a terminal provided in the third embodiment.

DETAILED DESCRIPTION

The following figures present invention will be described in further detail by specific embodiments in combination.

Example one:

Taking into account the existing fixed-word speech wake wake wake-user rate method exists, waking up rate and poor user experience technical problems; the present embodiment provides a wake-up can improve end-user rate, reduce the

时序不受限制，可以先执行前 者，后执行后者，或者同时执行前者和后者。

本实施例中的匹配结果包括：匹配成功与否，或者匹配分值；其中根据匹 配分值是否在预设范围内来判定是否匹配成功。

步骤105：根据所述匹配结果判断是否需要唤醒所述终端，若是，执行步骤 106，若否，则执行步骤107。

具体地，当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结 果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端；否则，判断不需要唤醒所述终端。

本步骤需要判断语音识别结果是否匹配成功和声纹特征是否匹配，这两个 判断过程的时序可以任意，具体包括以下情况：

1、先判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果 是否匹配成功，若是，则再判断所述语音信号的声纹特征与所述存储的语音唤 醒词声纹特征是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情 况下均判定不需要唤醒终端。

2、先判断所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否 匹配成功，若是，则再判断所述语音信号的语音识别结果与所述存储的语音唤 醒词识别结果是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情 况下均判定不需要唤醒终端。

3、同时判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结 果是否匹配成功、以及所述语音信号的声纹特征与所述存储的语音唤醒词声纹 特征是否匹配成功，若均成功，则判定需要唤醒所述终端；若有一个匹配不成 功则判定不需要唤醒终端。

本步骤中匹配成功可以包括：匹配分值在预设范围内，或者相似度在预设 范围之内。

步骤106：唤醒所述终端。

本步骤可以通过发送中断事件唤醒终端；当语音唤醒步骤在低功耗音频芯 片上执行时，低功耗音频芯片判断需要唤醒终端后，会上传中断事件给终端的主处理器唤醒处于休眠状态下的主处理器。

步骤107：不进行任何处理。

error rate to wake enhance the user experience of voice wake-up method, shown in Figure 1, the speech wake method of this embodiment includes the following steps:

Step 101: end-user acquisition wake-word speech signals emitted by obtaining the wake wakeup word recognition results voiceprint characteristic word speech signals and stored.

This embodiment of the method may be employed voice collecting device, such as a microphone to collect end-user voice signal sent. The method of this embodiment, the word recognition result obtained wake voiceprint characteristic timing is not limited, you can get the recognition result, after obtaining voiceprint may be acquired at the same time.

In this embodiment wake word recognition result is a wake-word speech recognition voice signal obtained wake-word text or commands. This embodiment voiceprint example is the speech signal may represent the user sound mode parameters, such as frequency of sound, loudness and so on.

In this step, a voice wake preparatory steps before occurred in a user-defined voice wake-up word in the process, this step may be performed by AP-side terminal, that is the main chip terminals, such as mobile phone main CPU ARM chip carry out.

In order to improve the accuracy of speech wake, the present embodiment method in the acquisition phase of sample characteristics, you can record multiple wake-word speech signals issued by the user, and then get the best wake-up word for word the speech signal wake-up times recorded results voiceprint recognition feature. As shown in Figure 2 for the terminal to obtain user-defined voice wake-word sample characteristics of the process, completed by the AP-side terminal:

In the terminal starts after a user-defined set wake-up word, the first recorded user of said wake-word speech signal, and then generate a wake wake-up signal word word speech recognition result of a voiceprint features 1, 1 and the recognition result voiceprint 1 storage to a custom wake-word memory area;

The second terminal of said user recorded wake-word speech signal, and then generate a wake wake-up word word speech recognition result

本实施例方法中当判断不需要唤醒终端时，可以忽略当前输入的语音信号， 执行唤醒步骤的芯片或器件可以进入休眠状态，以节省终端功耗。

本实施例提供的语音唤醒方法，以终端用户自己声音作为声音样本，获取 终端用户个体的声音特征(即识别结果和声纹特征)作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同(同 一用户的发声方式相同)，避免了用户个体的发声方式差异于声音样本用户的发 声方式导致唤醒失败，提供了唤醒率；另外，本实施例方法加入了语音识别结 果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境 无法唤醒终端，降低了误唤醒率；本发明的方法还可以供用户自定义唤醒词， 满足用户个性化的需求，提升了用户的体验。

进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本 实施例方法只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

考虑到噪声环境对用户输入的语音信号的影响，进一步提高用户唤醒率降 低误唤醒率，本实施例方法在上述步骤102之后，步骤103之前还可以包括：

判断所述语音信号是否满足预设信号要求，若是，则执行步骤103。

具体地，可以判断输入的语音信号的信噪比是否在预设范围内，若是，则 满足语音信号要求，或者语音信号的其他质量参数是否在预设范围内，若是，则满足语音信号要求，或者将输入的语音信号的声纹特征用用较简单的波形匹 配，若成功，则满足信号要求。该步骤可以在低功耗音频芯片上完成。

本实施例方法在判断语音信号不满足预设信号要求时，可以输入的语音信 号丢弃；或者当该判断步骤是在低功耗音频芯片上置信时，该芯片进入休眠状态，以节省电能。

优先地，如图3所示，上述判断所述语音信号是否满足预设信号要求的过 程可以包括：

步骤301：对所述语音信号进行处理得到用户语音数据和背景噪声数据。

由于外界输入的语音信号包括用户自己发出的语音信号和当前用户环境产 生的噪声信号，所以在进行信号判断时，可以将语音信号分离成用户语音数据和背景噪声数据。

signal 2 voiceprint characteristic 2, when the wake-up word recognition results 2 and 1 wake-up word recognition results are the same, optimize the use of voiceprint 2 voiceprint 1, when the wake-up word recognition results 2 and 1 wake word recognition results are not the same, re-recorded; recorded or re-enter the third record from the first start;

The third terminal of said user recorded wake-word speech signal, and then generate a wake wake-up word recognition results 3 voiceprint characteristic word speech signal 3, when the wake-up word recognition results 3 and 1 wake-up word recognition results are the same, optimize the use of voiceprint 3 above have been optimized voiceprint 1;

And so on, until the n-th recording is completed. Figure 2, n is 4, for example, the method of the present embodiment n values can be set according to actual demand.

Step 102: the terminal is in a dormant state, the voice signal detection and external input.

This step can be detected by the microphone terminal outside of the speech signal, when the terminal sleep state, the microphone can remain active.

In this step, the terminal is in hibernation from a hardware point of view may include: a main processor of the terminal is in a dormant state, from a software point of view may include: terminal operating system is in sleep or locked.

Step 103: obtaining a speech recognition result of the speech signal voiceprint features.

This step is obtaining a speech recognition result voiceprint feature no timing constraints, you can get the speech recognition result, after obtaining voiceprint, or you can get voiceprint, after obtaining the speech recognition result, you can also get voice results voiceprint recognition feature.

Speech recognition result

In this step, the voice signal is voice information obtained by speech recognition or text commands. This embodiment voiceprint example is the speech signal may represent a user utterance mode

步骤302：将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声 模型进行匹配。

本步骤中用户语音数据和背景噪声数据匹配时序不受限制，可以先后，也 可以同时。

本实施例方法可以预置一些较典型的生活场景下的稳态或非稳态噪声声学 统计模型，例如，汽车中，酒吧里，餐馆，家庭，办公室，地铁上，会议中等。

具体地，本步骤可以包括如下步骤：

将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分 值；

将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分 值。

步骤303：根据匹配结果判断所述语音信号是否满足预设信号要求。

本步骤中的匹配结果可以包括上述两个匹配过程得到的匹配分值，此时具 体地判断过程可以包括：

采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到 所述语音信号的匹配分值；

判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设 信号要求，若否，则判定不满足预设信号要求，若否，则不满足预设信号求。

例如对两个匹配分值进行加权处理得到一个匹配分值，然后判断该匹配分 值是否超出置信阀值，若是，则判定满足预设信号要求。

进一步考虑语音唤醒时终端的功耗问题，本实施例可以由一个音频处理器 来完成语音唤醒的主要过程，事先将声音样本特征(唤醒词识别结果和唤醒词语音信号的声纹特征)下载到该音频处理器的存储区；在终端处于休眠状态时 该音频处理器处于半sleep待机状态，由mic事件控制其工作唤醒；当mic检 测到声音输入时唤醒多核音频处理器，执行语音唤醒；这样可以使得终端的主 处理在休眠状态时，通过低功耗的音频处理器响应和处理外界输入的声音信息， 节省终端的功耗，提高终端的待机时间。

优先地，上述音频处理器可以采取分步式核心处理，Core1用于处理接收到 mic接收到外部声音信号的唤醒事件声音初步分析，即判断语音信号是否满足预 设要求，Core2用户后续的语音识别以及鉴权认证，即判断是否唤醒。当mic检 测到语音输入

parameters, such as the sound frequency, loudness and so on.

The main steps 103-106 to process step voice wake-up, which may be in a low-power audio chip performs, such as low-power DSP chip. Before voice wake, wake up in advance of the above word recognition results and the wake-up word speech signals voiceprint get downloaded to the low-power audio chip memory area, in preparation for the wake-up voice.

Step 104: Get wakeup word recognition result of the speech recognition result and storage, as well as matches the voiceprint and voiceprint stored.

In this step, get wake-up word recognition result of the speech recognition result matches were stored methods include: speech recognition result stored wake-up word recognition results were matched to obtain the voiceprint and voiceprint stored in way matching results include: the voiceprint features will be acquired and stored voiceprint match; this step get wake-up word recognition result matches the result with a stored voice, and access the stored voiceprint and voiceprint results matching process timing is not limited, you can execute the former, the latter after the execution, or both the execution of the former and the latter.

This embodiment results matching comprising: a successful match or not, or match scores; according to which the matching score is within a predetermined range to determine whether the match succeeds.

Step 105: according to the matching results to determine whether the need to wake up the terminal, and if so, step 106, and if not, proceed to step 107.

In particular, when the voice recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the characteristics of the voiceprint stored speech voice signal wake-up word voiceprint matching success when it is determined that the terminal needs to wake up; otherwise, the determination is no wakeup terminal.

This step needs to determine whether the results match the success of the speech recognition feature voiceprint matches, the timing of these two

时唤醒Core1，Core1执行判断语音信号是否满足预设要求的步骤； 当Core1判断为是时，唤醒Core2，并将语音信号传输给Core2(优先地，Core1 将语音信号存储在Core1、2共享的存储区中，由Core2提取)；Core2唤醒后执 行上述步骤103-106，完成唤醒鉴别，若通过，则由Core2向终端主处理器(例 如终端ARM芯片)传送硬件中断事件，终端主处理器接收到中断事件后被唤醒 由sleep模式切换到Normal模式正常运行。

应当理解的是：本实施例还可将语音识别结果匹配或者声纹匹配的过程放 在Core1中执行，只有当匹配成功时才唤醒Core2，然后Core2唤醒之后执行声 纹匹配或者语音结果匹配。

本实施例通过多核音频处理器的分步式核心处理设计，可以实现在节省终 端功耗或电能的效果，提升了用户体验。

下面通过如图4所示的语音唤醒系统来具体介绍本实施例方法；如图4所 示，该系统包括：DSP芯片(多核DSP芯片，可以包括至少两个核)、终端主CPU 中的ARM芯片和麦克风；其中DSP芯片包括两个核单元以及一个唤醒存储单元 (例如Flash)，第一核单元Core1和第二核单元Core2；ARM芯片包括唤醒词处 理单元和唤醒词存储单元；图4所示的系统进行语音唤醒的具体过程包括：

步骤501：ARM芯片中唤醒词处理单元录制用户所说的唤醒词语音信号，并 生成唤醒词语音识别结果和声纹特征，将唤醒词语音识别结果和声纹特征存储 在唤醒词存储单元中。

具体地可以生成方式可以参考上述图2的描述。

步骤502：ARM芯片将唤醒词语音识别结果、唤醒词语音信号的声纹特征和 预置的背景噪声模型下载到DSP芯片，并存储在唤醒存储单元中。

步骤503：在ARM芯片处于休眠状态下，麦克风动态监测外界的语音输入， 当检测到语音信号时发送唤醒事件唤醒Core1。

步骤504：Core1对检测到的语音信号进行处理得到用户语音数据和背景语 音数据。

步骤505：Core1将所述用户语音数据与所述预置的背景噪声模型进行匹配 得到第一匹配分值，将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

步骤506：Core1采用预设分值算法对所述第一匹配分值和所述第二匹配分 值进行计算得到所述语音信

processes can be any judgment, including the following:

1, first determine the speech recognition result and the stored speech of the speech signal wake-up word recognition result if the match is successful, and if so, then determine the voiceprint of the speech signal with the stored wakeup word speech voiceprint match is successful, if it is judged need to wake up the terminal; judgment is NO circumstances determination no wakeup terminal.

2, first determine the voiceprint of the speech signal with the stored speech wake term voiceprint match is successful, and if so, then judge the speech recognition result with the storage of the speech signal of voice wake word recognition result if the match is successful, if it is judged need to wake up the terminal; judgment is NO circumstances determination no wakeup terminal.

3, while the determination of the speech recognition result of the speech signal stored speech recognition result if the wake-up word matching is successful, and the voiceprint of the speech signal and the wake-up word stored speech characteristics whether the voiceprint successful match, if all is successful, it is determined that the terminal needs to wake up; if it is determined that a match is unsuccessful no wakeup terminal.

In this step, a match may include: match scores within a preset range, or similarity within a preset range.

Step 106: wake up the terminal.

In this step, you can wake up the terminal by sending an interrupt event; when the voice wake-up steps to be performed on a low-power audio chip, low-power audio chip Analyzing need to wake up after the terminal, the terminal will upload interrupt events to wake up the main processor main processor sleep state.

Step 107: no processing.

This embodiment of the method when the judge is not required to wake the terminal, you can ignore the current input speech signal, the wake-up procedure can chip or device into hibernation to save power terminal.

号的匹配分值。

步骤507：Core1判断所述语音信号的匹配分值是否大于预设阈值，若是， 则执行步骤508，若否，则执行步骤512。

步骤508：Core1将语音信号传输给Core2同时唤醒Core2。

步骤509：Core2唤醒后，对所述语音信号进行语音识别获得语音识别结果， 并与存储的唤醒词识别结果进行匹配，若匹配成功，则执行步骤510，若失败， 则执行步骤513。

步骤510：Core2提取所述语音信号的声纹特征，并与存储的声纹特征进行 匹配，若成功，执行步骤511，若否，执行步骤513。

步骤511：Core2发送唤醒中断事件给ARM芯片唤醒ARM芯片。

步骤512：Core1重新休眠，DSP芯片休眠。

步骤513：Core2重新休眠，语音唤醒流程结束。

DSP芯片中Core1和Core2的语音算处理能力不同功耗有差异，当DSP芯片 只有Core1工作时因为其运算能力和主频和处理能力有限，其运转时是DSP处 在相对低功耗工作状态(Normal)。如果DSP芯片Core2工作时其运算和处理能 力较高，功耗相对较高，Core2被唤醒时DSP芯片处于Work态。但唤醒DSP芯 片全速运行work状态的功耗也是远小于终端主CPU芯片Normal态的运行功耗。

本实施例可以将DSP芯片的运行状态分为低耗能Core1运行状态和相对高 耗能运算的Core2运行态。从上图5可以看出，即使在复杂强噪音环境下，只 要没有清晰人声特征输入，Core2将不被唤醒，会使DSP芯片处于稳定低功耗状 态。即使Core2被唤醒如果人声输入不是自定义的唤醒指令，或者唤醒指令不 是录制唤醒特征的用户输入，手机主CPU芯片也不会被唤醒。通过硬件低功耗DSP芯片的阶梯层次的唤醒机制，个体化唯一化的鉴权机制保证了手机语音唤醒 的功耗控制，和误醒率的降低。

应用本实施例方法，带有声纹特征的自定义唤醒词可以在降低了智能机中 语音检测功耗控制优化难题的基础上，更好的适应特殊人群或者个体的个性化需求，在便捷操作的同时，让用户的语音控制更贴近于人们生活实际使用场景。

实施例二：

During voice wake, since the sound;

Example embodiments provide a voice to wake this method to end-users own voice as a sound sample, end-users get the sound characteristics of individuals (ie, the recognition result voiceprint characteristic) as an audio sample characteristics Sample features voiceprint and voiceprint wakeup extraction (same same user sound pattern), to avoid the user individual sound pattern difference in the sound sample user sound pattern caused wake-up failed to provide a wake-up rate; in addition, Example embodiments of the present joined the speech recognition result voiceprint matching characteristics, can effectively avoid the noisy environment caused by waking up, so that the noise environment can not wake up the terminal, reducing the error rate wake; the method of the present invention may also supply custom wake-term, to meet the needs of individual users, to enhance the user experience.

Further, since different users different sound pattern is different, that voiceprint, the application of the present embodiment method only to wake up my end user terminal, compared with the prior art, but also to improve the information security of the terminal.

Taking into account the environmental impact of noise on the voice signal input by the user, to further enhance the user reduces the error rate to wake wakeup rate, the method of this embodiment, after step 102 and before step 103 may also include:

Analyzing the voice signal meets the requirements of a preset signal, and if so, step 103 is performed.

In particular, based on the input speech signal to noise ratio is within a preset range, and if so, to meet the requirements of the speech signal, or other quality parameters of the speech signal is within a preset range, and if so, to meet the voice signal requirements, or to enter the voiceprint of the speech signal with the use of a relatively simple waveform matching, if successful, will meet the signaling requirements. This step can be done on a low-power audio chip.

Example of the present embodiment in determining the speech signal does not meet the requirements of the preset signal, the speech signal can be input to discard; or when the judging step is confidence in the low-power audio chip, the

为进一步提升用户体验，在终端唤醒之后，还行可以进行多样性的快速和 自动响应，例如打开应用程序等，本实施例在上述语音唤醒方法的基础上，提供了一种终端处理方法，如图6所示，包括如下步骤：

步骤601：利用实施例一所述的语音唤醒方法唤醒所述终端。

具体唤醒过程参考实施例一的具体描述。

步骤602：所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对 应的预设操作。

上述步骤602可以由终端中的主处理器完成。

本实施例方法可以预先设置自动定义唤醒词对应的响应操作，不同的唤醒 词可以对应不同的响应操作；在用户通过语音唤醒词唤醒终端之后，终端根据 用户输入语音信号的识别结果执行对应的响应操作，例如自动打电话或者自动 打开聊天软件或者开启驾驶模式等。这里的唤醒词是文字内容，识别结果也是 用户所说的文字内容。

优先地，当通过唤醒中断事件来唤醒终端情况下，可以设置不同的唤醒词 对应不同的唤醒中断事件，如果一个用户预先录制多个唤醒词，不同唤醒词对应不同的唤醒中断事件，终端对不同的唤醒中断事件，执行不同的预设操作或 者架构不同的应用场景。这样达成即使在手机黑屏休眠时用户也可以通过不同 的用户唤醒指令，进行多样性的快速和自动响应。例如用户自定义唤醒词为"打 电话"时，在终端处于休眠状态下，在用户说出"打电话"时终端发出对应的 唤醒中断事件来唤醒自身，并且终端在唤醒之后会根据预设的电话号码拨打电 话。

因此，上述步骤601语音唤醒方法中唤醒所述终端的过程包括：根据所述 语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

上述步骤602具体包括：根据所述唤醒中断事件的类型执行对应的预设操 作。

下面列举应用本实施例方法的具体场景：

场景一：如果用户录制"急救SOS，救命，妈妈"此类和用户情急时呼救口 语化相关的指令做为唤醒词时，给用户一个默认的紧急唤醒的设置方法。允许用户设置的应急号码，即使在强噪音数据下，此指令的敏感度分值在用户无法 手动呼叫的时候第一时间协助用户拨打出急救电话，或者发送出急救信息给接 收者，同时后台开启录音或摄像功能存放在紧急文件目录中，并在wifi环境下 上传网络云盘(此目

chip enters sleep mode to conserve electricity. preferentially, shown in Figure 3, the above-described voice signal meets a preset determination required signal process may include:

Step 301: the voice of the voice signals are processed to obtain user data and background noise data.

Since the external input voice signal includes a speech signal emitted and the current user's own user environment generated noise signal, during signal determining can be a voice signal into voice data and user data background noise.

Step 302: the voice data and the user data of the background noise and the background noise are preset to match the model.

In this step, the user voice data and background noise unrestricted data matching the timing may have to be at the same time.

Example of the present embodiment can be preset number of stationary and non-stationary noise acoustic statistical model than the typical life scenarios, such as a car, a bar, a restaurant, home, office, subway, meeting moderate.

In particular, this step may comprise the steps of: the user voice data and background noise model of the pre-match points to get the first matching value;

The data of the background noise and background noise model matching the preset value to obtain a second division match.

Step 303: according to the matching results to determine whether the speech signal meets a preset signal requirements.

This step may include the above two matching results obtained by the matching process to match the score, this time to determine specific process may include:

With the default value of the first matching algorithm score and the second match scores were calculated match score of the speech signal; match scores to judge the speech signal is greater than a predetermined threshold, if it is determined to meet the requirements of the preset signal, if not, it is determined that the signal does not meet

录只能通过密码，用户手动删除，不被系统格式化删除)， 紧急文件生成时会以短信息的形式通知用户设置的紧急联系人。

场景二：当监听到用户说"OK，LetGo！"，手机根据外界噪音环境是否类 似汽车载内环境，根据陀螺仪和加速度传感器检测手机的动能位移增量变化，根据光感器检测是否手机屏幕置于非遮挡状态，为用户直接开启中兴驾驶助手。 用户拿起手机下车时，根据噪音环境和传感器动量检测的变化自动关闭驾驶助 手。

场景三：幼儿英文发音练习：通过自定义唤醒词提前录制小朋友正学习中 的英文语句，其成功识别后对应设置儿童喜欢的特色应用或登陆模式。一个很喜欢玩手机游戏的小朋友，家长通过自定义唤醒词关联手机儿童模式，为了引 导小朋友在游戏的同时巩固练习英文发音，可以将正在学习中的小朋友发音和 语句预置为唤醒词，只有小朋友记忆正确并发音标准后才可以进入儿童模式进 行游戏。不经意间规范小朋友的行为，提高学习热情。

其他场景：也可以通过手机在线服务或者内置不同的歇后语，对联，或时 尚俏皮话供用户选择录制，当用户说了对应的时尚俏皮话的内容时，允许用户。

本实施例方法提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒 率，实现趣味个性化的语音应用场景实现。

实施例三：

如图7所示，本实施例提供了一种语音唤醒装置，其特征在于，包括：采 样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块； 所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音 信号的唤醒词识别结果和声纹特征并存储；

所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、 以及所述声纹特征与存储的声纹特征的匹配结果；

所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

the preset requirements, and if not, do not seek to meet preset signal.

For example two match scores were weighted to obtain a matching score, the match points and then determine whether the value exceeded the threshold of confidence, if it is judged to meet the requirements of the preset signal.

Consider further voice terminal power consumption when the wake-up, the present embodiment may be composed of a main audio processor to complete voice wake-up process, prior sound sample characteristics (wake-up word recognition results and wake voiceprint word speech signals) to download the audio processor storage area; when the terminal is in a dormant state in a semi-sleep of the audio processor standby mode, the mic event controlled wake of its work; when detecting wake mic audio multicore processors sound input to the implementation of voice wake-up; this can make the main terminal processing in the sleep state, the low-power audio processor in response to external input and processing of sound information, save power terminals, improve end-of standby time. priority, said audio processor core can be taken step-wise process, Core1 mic for processing a received signal received by the external sound a wake-up event sounds preliminary analysis, which determines the voice signal meets the preset requirement, Core2 users to follow speech recognition and authentication authentication, that is, whether to wake up. Wake When mic input Core1 speech is detected, the preset step required Core1 execute judgment speech signal meets; when Core1 determined to be, wake Core2, and voice signals to the Core2 (preferentially, Core1 voice signal stored in Core1,2 shared storage area extracted by the Core2); Core2 wake after performing the above steps 103-106, to wake up to identify, if adopted by the Core2 processor to the main terminal (for example terminal ARM chips) and transfer the hardware interrupt event, terminal host processor receives an interrupt wake-up event after the switch from sleep mode to normal mode normal operation.

It should be understood: the speech recognition result can also be matched to the present embodiment or voiceprint matching process on Core1 performed only when the match is successful only wake Core2, and then Core2 wake

所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

本实施例提供的语音唤醒装置，以终端用户自己声音作为声音样本，获取 终端用户个体的声音特征(即识别结果和声纹特征)作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同(同 一用户的发声方式相同)，避免了用户个体的发声方式差异于声音样本用户的发 声方式导致唤醒失败，提供了唤醒率；另外，本实施例装置中加入了语音识别 结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环 境无法唤醒终端，降低了误唤醒率；本实施例的装置还可以供用户自定义唤醒 词，满足用户个性化的需求，提升了用户的体验。

进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本 实施例装置可实现只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

优先地，如图8所示，在上述装置的基础上，本实施例语音唤醒装置还包 括：第二判断模块；

所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第 一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信 号的语音识别结果和声纹特征。

优先地，所述第二判断模块用于：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行 匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

优先地，如图9所示，在上述装置的基础上，本实施例唤醒装置中所述第 一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音 频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第 一核单元中；

所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是 否满足预设信号要求，若是，唤醒所述第二核单元；

up after a voiceprint match result of the execution or voice match.

In this embodiment, a multi-core audio processor core sub-step process designed to achieve savings in power terminals or electrical effects, enhance the user experience.

The following voice wake-up system shown in FIG. 4 will be specifically described embodiment the method of the present embodiment; FIG. 4, the system comprising: DSP chip (multicore DSP chip may include at least two nuclei), the main CPU terminal the ARM chip and microphone; wherein the DSP chip includes two nuclear units, and a wake-up storage unit (such as Flash), the first and second nuclear unit Core1 nuclear unit Core2; ARM chip includes a processing unit and a wake-up wake-up word word storage unit; specific process system shown in Figure 4 comprises a voice wakeup:

Step 501: ARM chip in the wake of word processing unit recording the user said wake-word speech signal, and generates a wake-word speech recognition result voiceprint feature will wake word speech recognition result is stored in the voiceprint feature wake term storage unit. can specifically be described with reference to the above-described generation method of FIG.

Step 502: ARM chip will wake word speech recognition result, wake-word speech signals voiceprint and a preset background noise model downloaded to the DSP chip, and stored in the wake storage unit.

Step 503: In the ARM chip is in sleep mode, the dynamic monitoring external microphone voice input, sends wake-up event when the wake-up Core1 voice signal is detected.

Step 504: Core1 of detected voice signals are processed to obtain user speech data and background speech data.

Step 505: Core1 voice data to the user and the background noise model to obtain a first preset matching match score, the background noise model and the background noise data to obtain a first pre-matching two match score;

Step 506: Core1 using preset algorithm to score the first and second match score of the match

所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的 语音识别结果和声纹特征。

优先地，本实施例中所述第一判断模块用于当所述语音信号的语音识别结 果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

本实施例装置提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒 率。

如图10所示，本实施例还提供了一种终端，包括处理模块和如上任一所 述的语音唤醒装置；

所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音 信号的语音识别结果执行对应的预设操作。

优先地，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果 产生预设类型的唤醒中断事件唤醒所述终端；

所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

优先地，本实施例中处理模块集成在ARM芯片；所述语音唤醒装置集成在 低功耗音频芯片上。

本实施例提供的终端提高用户个体的语音唤醒率，降低噪音和他人声音的 误唤醒率，实现趣味个性化的语音应用场景。

以上内容是结合具体的实施方式对本发明所作的进一步详细说明，不能认 定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干简单推演或替换， 都应当视为属于本发明的保护范围。

points value, calculated match score of the speech signal.

Step 507: Core1 determination of the speech signal match score is greater than a predetermined threshold, If YES, step 508, and if not, step 512 is performed.

Step 508: Core1 voice signal to wake up at the same time Core2 Core2.

Step 509: After Core2 wake of the speech signal for speech recognition speech recognition result obtained, Wake word recognition results and stored to match, if the match is successful, step 510, if failed, step 513.

Step 510: Core2 extracted voiceprint of the speech signal, with stored voiceprint match, if successful, step 511, and if not, step 513.

Step 511: Core2 send a wakeup interrupt events to wake ARM ARM chip chip.

Step 512: Core1 back to sleep, DSP chip to sleep.

Step 513: Core2 back to sleep, wake voice flow ends.

There are differences DSP chip Core1 Core2 and voice processing capabilities of different calculation power only when the DSP chip Core1 work because of the limited computing power and clock speed and processing power when it is operating in a relatively low-power DSP job status (Normal). If the DSP chip Core2 work when the computational processing power and high power consumption is relatively high, when the Core2 wakes DSP chip in the Work state. But wake DSP chip running at full speed work state power is much smaller than the main CPU chip operating power terminal Normal state.

This embodiment of the operating state of the DSP chip can be divided into low-power operating state and Core1 relatively high energy-consuming operation Core2 run state. As can be seen from Figure 5, even in a complex environment of strong noise, not only to clear voice input feature, Core2 will not wake up, make DSP chip in a stable low-power state. Even if the person is awakened Core2 voice input is not custom instructions to wake, or wake-up instruction is not recorded wake-up feature user input, mobile phone main CPU chip will not be awakened. Low-power DSP

chip hardware ladder level wake-up mechanism, the only of individual authentication mechanisms to ensure the cell phone voice wake of power control, and reduce the error rate in the wake.

Example

embodiment of the present application, the custom with voiceprint wake term basis can reduce the power of intelligent speech detection machine control optimization problems, the better adapt to the individual needs of special groups or individuals, in at the same time easy to operate, allowing users to voice control closer to people's real life usage scenarios.

Second Embodiment:

To further enhance the user experience, after the wake-up terminal, the line can also be fast and automatic response to diversity, such as opening applications, etc., on the basis of the present embodiment, the above-described voice wake-up method, there is provided a terminal processing The method, as shown in Figure 6, comprising the steps of:

Step 601: Use a voice wake-up method described in Example wake the terminal.

Detailed description of a process of Reference Example

particular wake.

Step 602: After waking up in the terminal, according to a speech recognition result of the speech signal should perform the preset action. above step 602 can be done at the end of the main processor.

Example of the present embodiment may be preset to automatically defines the wake-word corresponding to the response action, different wake-up word may correspond to different response actions; after the user wakeup word wake terminals via voice, the terminal performs according to the identification result of user input speech signal corresponding response operations, such as automatic call or chat software to automatically open or turn driving mode. Here is the text of the wake-up word, but also the recognition result of said user text. priority, when the wake-up through an interrupt event to wake the terminal cases, you can set different wake-up

word corresponds to a different wake-up interrupt events, if a plurality of user pre-recorded wake-up word, different word corresponding to different wake wakeup interrupt events, terminal to different wake-up interrupt events, perform different preset operation or architecture different scenarios. Such reached even when the phone user can also awakened dormant black instruction by different users, for quick and automatic response diversity. For example, after a user-defined wake-up word is "call" in the dormant terminal, the terminal issues when the user says "Call" corresponds to the wake-up interrupt events to wake up to itself, and in the wake of the terminal according to a preset call the phone number.

Therefore, the above method in step 601 voice wake wake of the terminal process comprising: generating a preset type of wake-up interrupt events to wake the terminal according to the recognition result of the speech signal;

602 steps above include: wake-up according to the type of interrupt event executes the corresponding preset operation.

The following are applications of the present embodiment method specific scenarios:

Scene One: If the user record "emergency SOS, help, Mom" when such a desperate cry for help and user instructions related to oral language of the word as a wake-up, giving the user a default emergency wake-setting method. Allows the user to set the emergency number, even with strong noise data, the sensitivity value of this instruction the first time to assist users to make an emergency call when the user can not manually call, or send an emergency message to the recipient at the same time open the background recording or camera functions stored in the file directory, emergency, and in wifi network environment uploaded cloud disk (this directory can only be deleted by password, user manual, are not deleted formatted system), emergency file is generated in the form of short messages Notify the user to set emergency contact.

Scene Two: When listening to the user says "OK, LetGo!", A mobile phone according to the external environment is similar to the noise inside the car contained environment, according to the gyroscope and acceleration sensor to detect the phone's kinetic displacement incremental change, according to the light sensor to detect Are mobile

phone screen into a non-blocking state for the user to directly open ZTE driving assistant. When the user picks up the phone off, according to changes in the environment and noise detection sensor automatically turns off the momentum driving assistants.

Scene Three: Children practice English pronunciation: wake-word recording by customizing the children are learning English in the statement in advance, after the success of its children like identifying the corresponding set of characteristic application or landing mode. A very like to play mobile games of children, parents, custom wake-up word association Kid Mode phone, in order to guide the children in the game, while the consolidation exercise English pronunciation can be children are learning the pronunciation and statements preset wake-up word, only children memory and correct pronunciation standards before it can enter the game mode for children. Inadvertently regulate children's behavior, increase enthusiasm for learning.

Other scenarios: You can also phone or online service built different twisters, couplets, or when the recording is still one-liners for users to choose, when the user says the fashion pun content corresponding time, allowing the user.

This embodiment of a method to improve the individual user's voice wake-up rates, lower error rates wake-noise sound and others, to achieve interesting personalized voice Scenario.

Third Embodiment:

7, the present embodiment provides a wake-speech apparatus, characterized by comprising: a sampling module, a detection module, a first acquisition module, a second acquisition module, and a first judgment module wakeup module; the sampling module is used to wake-word speech signal acquisition end-user issue, get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

The detection module for the terminal is in a dormant state, the detection of the external input speech signal;

The first acquisition module for acquiring speech recognition result of the speech signal voiceprint feature;

The second acquisition module for acquiring the speech recognition result wakeup word recognition result and storage, as well as matches the voiceprint and voiceprint stored;

The first module is used for judging according to the matching results to determine whether to wake the terminal; in the wake of the first judging module for judging module is YES, the wake-up terminal.

During voice wake, since the sound;

Example provide voice wake present embodiment means for end-users own voice as a sound sample, end-users get the sound characteristics of individuals (ie, the recognition result voiceprint characteristic) as an audio sample characteristics Sample features voiceprint and voiceprint wakeup extraction (same same user sound pattern), to avoid the user individual sound pattern difference in the sound sample user sound pattern caused wake-up failed to provide a wake-up rate; in addition, this embodiment of the device is added to the speech recognition result voiceprint matching characteristics, can effectively avoid the noisy environment caused by waking up, so that the noise environment can not wake up the terminal, reducing the error rate wake; apparatus of the present embodiment can also provide the user custom wake term, to meet the needs of individual users, to enhance the user experience.

Further, since different users different sound pattern is different, that voiceprint, the application of the present embodiment of the device can be realized only to wake up my end user terminal, compared with the prior art, but also to improve the information security of the terminal.

Second judging module; : preferentially, 8, on the basis of the above means, in this case voice wake embodiment further includes means Before After

in the second determining module for external detection module input speech signal, the first acquisition module acquiring the speech recognition result of the speech signal and a voiceprint characteristic, judging whether the speech signal satisfies preset signal requirements; the first module for obtaining said second determination means determines YES, obtaining a speech recognition result of the speech signal and a voiceprint characteristic. priority, the second judging module for: of the speech signal is

processed to obtain user data and voice background noise data; the user data and voice data of the background noise and background noise models are preset to match; determine whether the speech signal according to the matching result satisfies preset signal requirements. preferentially, shown in Figure 9, in the above apparatus, the apparatus of the present embodiment in the wake of the first acquisition module, a second acquisition module, judging module, and the first wake-up module is integrated in the multi-core audio processor second nuclear unit; the second judgment module integrated in the multi-core audio processor's first core units;

The second determination module is used after the first wake-up core unit, it determines whether the speech signal is a signal complies with the preset requirement, and if so, waking the second core means;

The first acquisition module for, after said second core means wake, the speech signal to obtain a speech recognition result and a voiceprint characteristic. preferentially, in this case the judgment of the first voice recognition module for the speech signal when the result of the stored speech the voiceprint of the speech signal wake-up and implementation of word recognition result matches, and the stored voice wake-up word voiceprint match is successful, it is determined that the terminal needs to wake up.

Example apparatus of the present embodiment to improve the individual user's voice wake-up rates, lower error rate noise and wake others sound.

10, the present embodiment also provides a terminal includes a processing module, and took office as a wake-up means of said speech;

After the processing module for the voice wake wake the terminal apparatus, performs the corresponding operation according to the preset speech recognition result of the speech signal. priority, the voice wake-up means for generating a wake-up module preset type of wake-up interrupt events to wake the terminal according to the recognition result of the speech signal;

The processing module is used to perform the type of interrupt events corresponding operation according to the preset wake.

Example processing module

priority, the present embodiment is integrated in the ARM chip; the voice wake-up device is integrated in the low-power audio chip.

Terminal

This embodiment provides improved user individual voice wake-up rates, lower error rates wake-noise sound and others, to achieve interesting personalized voice scenarios.

The above is described in further detail with reference to specific embodiments of the present invention is made, the final version does not recognize the specific embodiment of the invention is only limited to these instructions. Those of ordinary skill in the art of the present invention, without departing from the inventive concept premise, you can also make a number of simple deduction or replacement, should be regarded as belonging to the scope of the present invention.

Cited by: CN105810196 A ; CN105898065 A ; CN106131292 A ; CN106297777 A ; CN106448663 A ; CN106502649 A ; WO18086033 A1 ;

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105575395 A
(43)申请公布日 2016.05.11

(21)申请号 201410541444.6

(22)申请日 2014.10.14

(71)申请人 中兴通讯股份有限公司
地址 518057 广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦

(72)发明人 王婕

(74)专利代理机构 深圳鼎合诚知识产权代理有限公司 44281
代理人 薛祥辉 李发兵

(51)Int.Cl.
G10L 17/22(2013.01)
G10L 15/22(2006.01)

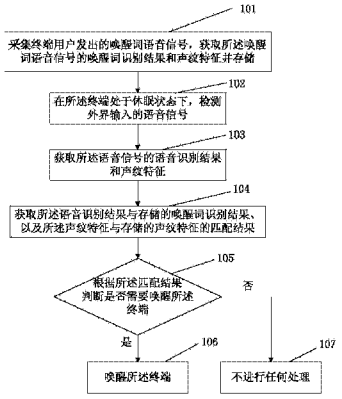
权利要求书3页 说明书12页 附图6页

(54)发明名称

语音唤醒方法及装置、终端及其处理方法

(57)摘要

本发明公开了一种语音唤醒方法及装置、终端及其处理方法。本发明的语音唤醒方法包括：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；在所述终端处于休眠状态下，检测外界输入的语音信号；获取所述语音信号的语音识别结果和声纹特征；获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端；本发明的语音唤醒方法能够解决现有语音唤醒技术存在的用户唤醒率低、误唤醒率高以及用户体验差的技术问题。



CN 105575395 A

1. 一种语音唤醒方法，其特征在于，包括如下步骤：

采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

在所述终端处于休眠状态下，检测外界输入的语音信号；

获取所述语音信号的语音识别结果和声纹特征；

获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端。

2. 如权利要求 1 所述的方法，其特征在于，在检测外界输入的语音信号之后，获取所述语音信号的语音识别结果和声纹特征之前，还包括：

判断所述语音信号是否满足预设信号要求，若是，获取所述语音信号的语音识别结果和声纹特征。

3. 如权利要求 2 所述的方法，其特征在于，所述判断所述语音信号是否满足预设信号要求的步骤包括：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

4. 如权利要求 3 所述的方法，其特征在于，所述将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配的步骤包括：

将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值；

将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

所述根据匹配结果判断所述语音信号是否满足预设信号要求的步骤包括：

采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值；

判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设信号要求，若否，则判定不满足预设信号要求。

5. 如权利要求 1-4 任一项所述的方法，其特征在于，所述根据所述匹配结果判断是否需要唤醒所述终端的步骤包括：

当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

6. 如权利要求 2-4 任一项所述的方法，其特征在于，所述判断所述语音信号是否满足预设信号要求的步骤包括：

多核音频处理器中的第一核单元唤醒后，所述第一核单元判断所述语音信号是否满足预设信号要求；

所述获取所述语音信号的语音识别结果和声纹特征的步骤包括：

当所述第一核单元判断为是时，所述第一核单元唤醒所述多核音频处理器中的第二核单元；

所述第二核单元获取所述语音信号的语音识别结果和声纹特征；

Petitioner Exhibit 1004, Page 707 of 1052

所述获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果的步骤包括：

所述第二核单元获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

所述根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端的步骤包括：

所述第二核单元根据所述匹配结果判断是否需要唤醒所述终端，若是，则所述第二核单元唤醒所述终端。

7. 如权利要求 6 所述的方法，其特征在于，还包括：

当所述第一核单元判断为否时，所述第一核单元进入休眠状态；

和 / 或

当所述第二核单元判断为否时，所述第二核单元进入休眠状态。

8. 如权利要求 6 所述的方法，其特征在于，所述第二核单元唤醒所述终端的步骤包括：

所述第二核单元发送中断事件唤醒所述终端中处于休眠状态的主处理器。

9. 一种终端处理方法，其特征在于，包括如下步骤：

利用权利要求 1-5 任一项所述的语音唤醒方法唤醒所述终端；

所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设操作。

10. 如权利要求 9 所述的方法，其特征在于，所述唤醒所述终端的步骤包括：

根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

所述语音信号的语音识别结果执行预设的操作的步骤包括：

根据所述唤醒中断事件的类型执行对应的预设操作。

11. 一种语音唤醒装置，其特征在于，包括：采样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块；

所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

12. 如权利要求 11 所述的装置，其特征在于，还包括：第二判断模块；

所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信号的语音识别结果和声纹特征。

13. 如权利要求 12 所述的装置，其特征在于，所述第二判断模块用于：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

Petitioner Exhibit 1004, Page 708 of 1052

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

14. 如权利要求 12 或 13 所述的装置，其特征在于，所述第一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是否满足预设信号要求，若是，唤醒所述第二核单元；

所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的语音识别结果和声纹特征。

15. 如权利要求 11-13 任一项所述的装置，其特征在于，所述第一判断模块用于当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

16. 一种终端，其特征在于，包括：处理模块和如权利要求 11-13 任一项所述的语音唤醒装置；

所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音信号的语音识别结果执行对应的预设操作。

17. 如权利要求 16 所述终端，其特征在于，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

Petitioner Exhibit 1004, Page 709 of 1052

语音唤醒方法及装置、终端及其处理方法

技术领域

[0001] 本发明涉及语音识别技术领域，尤其涉及一种语音唤醒方法及装置、终端及其处理方法。

背景技术

[0002] 语音在手机终端中已不单单存在于单独的程序应用中，而开始慢慢的作为一种独立于手指触摸输入的新增输入方式，不需触碰手机，遥控和更智能化友好的服务于手机用户。

[0003] 而语音唤醒技术为用户解放双手完全的语音指令遥控手机提供了前提，唤醒主流技术是技术软件唤醒，但软件运行在手机系统启动为前提，为保证能随时随地接收到用户的语音指令，语音引擎需要一直后台运行和监听，手机系统将无法进入休眠待机的节电状态，此类语音唤醒技术的应用对电量消耗异常强烈。目前语音技术在市场技术应用中在不单单考虑到功能友好性，还要关注手机电量消耗，同时要求对用户使用中的不被语音过多打扰，所以目前各厂商对于语音识别系统的唤醒方案多数采用休眠状态下硬件按键唤醒或者手机系统运行时软件按钮点击唤醒。

[0004] 随着技术人员的不断改进，目前已经出现语音低功耗唤醒技术，在手机休眠状态下应用低功耗的芯片实现直接用语音指令唤醒手机操作系统，无需用户手动进行硬件和软件操作。

[0005] 但对于现有技术的语音低功耗唤醒功能中，只能使用固定的唤醒词，通过语音模型生成此预置固定唤醒词的声学模型。具体地现有语音低功耗唤醒功能的实现方法，包括如下：

[0006] 录制一些大批量的“选定唤醒词”声音样本进行处理和人群声学特征培训和提取。

[0007] 将统计估算得出的“固定的唤醒词”声音特征预置在手机系统中。

[0008] 根据用户个体输入的声音波形与统计生成的声音特征进行比对。

[0009] 因低功耗性能音频芯片中运算处理能力有限，群体样本统计生成的声学模型在单独用户个体使用时会有以下问题：

[0010] 1、用户个体的发声方式差异性远远低于或超出采样统计模型特征提取的置信区间，造成用户唤醒率低。如果样本量太小或者或人群特征太单一，例如，全是普通话，全是成年人，会造成有些特定人群的声音例如儿童，老人的声音对唤醒失效。

[0011] 2、在噪音环境下复杂噪音环境生成了采样统计模型中类似的似然曲线，造成误唤醒率高。不该唤醒时乱唤醒，造成用户困扰。

[0012] 3、由于只能通过厂家事先录制的“固定的唤醒词”来唤醒终端，用户不能

[0013] 够自定义唤醒词，不能满足用户个性化的需求，降低了用户体验。

[0014] 因此，现有技术虽然实现了一定功耗控制下的语音唤醒，但是存在用户唤醒率低、误唤醒率高以及用户体验差的技术等问题。

Petitioner Exhibit 1004, Page 710 of 1052

发明内容

[0015] 本发明要解决的主要技术问题是，提供一种语音唤醒方法及装置、终端及其处理方法，能够解决现有语音唤醒技术存在的用户唤醒率低、误唤醒率高以及用户体验差的技术问题。

[0016] 为解决上述技术问题，本发明提供一种语音唤醒方法，包括如下步骤：

[0017] 采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

[0018] 在所述终端处于休眠状态下，检测外界输入的语音信号；

[0019] 获取所述语音信号的语音识别结果和声纹特征；

[0020] 获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0021] 根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端。

[0022] 进一步地，所述方法在检测外界输入的语音信号之后，获取所述语音信号的语音识别结果和声纹特征之前，还包括：判断所述语音信号是否满足预设信号要求，若是，获取所述语音信号的语音识别结果和声纹特征。

[0023] 进一步地，所述判断所述语音信号是否满足预设信号要求的步骤包括：

[0024] 对所述语音信号进行处理得到用户语音数据和背景噪声数据；

[0025] 将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

[0026] 根据匹配结果判断所述语音信号是否满足预设信号要求。

[0027] 进一步地，所述将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配的步骤包括：

[0028] 将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值；

[0029] 将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

[0030] 所述根据匹配结果判断所述语音信号是否满足预设信号要求的步骤包括：

[0031] 采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值；

[0032] 判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设信号要求，若否，则判定不满足预设信号要求。

[0033] 进一步地，所述根据所述匹配结果判断是否需要唤醒所述终端的步骤包括：

[0034] 当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

[0035] 进一步地，所述判断所述语音信号是否满足预设信号要求的步骤包括：

[0036] 多核音频处理器中的第一核单元唤醒后，所述第一核单元判断所述语音信号是否满足预设信号要求；

[0037] 所述获取所述语音信号的语音识别结果和声纹特征的步骤包括：

[0038] 当所述第一核单元判断为是时，所述第一核单元唤醒所述多核音频处理器中的第二核单元；

Petitioner Exhibit 1004, Page 711 of 1052

[0039] 所述第二核单元获取所述语音信号的语音识别结果和声纹特征；

[0040] 所述获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果的步骤包括：

[0041] 所述第二核单元获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0042] 所述根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端的步骤包括：

[0043] 所述第二核单元根据所述匹配结果判断是否需要唤醒所述终端，若是，则所述第二核单元唤醒所述终端。

[0044] 进一步地，所述方法还包括：

[0045] 当所述第一核单元判断为否时，所述第一核单元进入休眠状态；

[0046] 和／或

[0047] 当所述第二核单元判断为否时，所述第二核单元进入休眠状态。

[0048] 进一步地，所述第二核单元唤醒所述终端的步骤包括：

[0049] 所述第二核单元发送中断事件唤醒所述终端中处于休眠状态的主处理器。

[0050] 同样为了解决上述的技术问题，本发明还提供了一种终端处理方法，包括如下步骤：

[0051] 利用上述前五项中任一项所述的语音唤醒方法唤醒所述终端；

[0052] 所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0053] 进一步地，所述唤醒所述终端的步骤包括：

[0054] 根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0055] 所述语音信号的语音识别结果执行预设的操作的步骤包括：

[0056] 根据所述唤醒中断事件的类型执行对应的预设操作。

[0057] 同样为了解决上述的技术问题，本发明还提供了一种语音唤醒装置，包括：采样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块；

[0058] 所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

[0059] 所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

[0060] 所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

[0061] 所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0062] 所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

[0063] 所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

[0064] 进一步地，所述装置还包括：第二判断模块；

[0065] 所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

[0066] 所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信号的语音识别结果和声纹特征。

Petitioner Exhibit 1004, Page 712 of 1052

[0067] 进一步地，所述第二判断模块用于：

[0068] 对所述语音信号进行处理得到用户语音数据和背景噪声数据；

[0069] 将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

[0070] 根据匹配结果判断所述语音信号是否满足预设信号要求。

[0071] 进一步地，所述第一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

[0072] 所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是否满足预设信号要求，若是，唤醒所述第二核单元；

[0073] 所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的语音识别结果和声纹特征。

[0074] 进一步地，所述第一判断模块用于当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

[0075] 同样为了解决上述的技术问题，本发明还提供了一种终端，包括处理模块和上述所述的语音唤醒装置；

[0076] 所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0077] 进一步地，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0078] 所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

[0079] 本发明的有益效果是：

[0080] 本发明提供了一种语音唤醒方法及装置、终端及其处理方法，能够提高用户唤醒率、降低误唤醒率和提升用户体验。具体地，本发明的语音唤醒方法包括：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；在所述终端处于休眠状态下，检测外界输入的语音信号；获取所述语音信号的语音识别结果和声纹特征；获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端；本发明的语音唤醒方法，以终端用户自己声音作为声音样本，获取终端用户个体的声音特征（即识别结果和声纹特征）作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同（同一用户的发声方式相同），避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤醒失败，提供了唤醒率；另外，本发明的方法加入了语音识别结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端，降低了误唤醒率；本发明的方法还可以供用户自定义唤醒词，满足用户个性化的需求，提升了用户的体验。

[0081] 另外，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施例方法只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

Petitioner Exhibit 1004, Page 713 of 1052

附图说明

[0082] 图 1 为本发明实施例一提供的一种语音唤醒方法的流程示意图；

[0083] 图 2 为本发明实施例一提供的一种获取用户自定义语音唤醒词样本特征的流程示意图；

[0084] 图 3 为本发明实施例一提供的一种判断语音信号是否满足预设信号要求的流程示意图；

[0085] 图 4 为本发明实施例一提供的一种语音唤醒系统的结构示意图；

[0086] 图 5 为本发明实施例一提供的一种语音唤醒系统进行语音唤醒的流程示意图；

[0087] 图 6 为本发明实施例二提供的一种终端处理方法的流程示意图；

[0088] 图 7 为本发明实施例三提供的第一种语音唤醒装置的结构示意图；

[0089] 图 8 为本发明实施例三提供的第二种语音唤醒装置的结构示意图；

[0090] 图 9 为本发明实施例三提供的第三种语音唤醒装置的结构示意图；

[0091] 图 10 为本发明实施例三提供的一种终端的结构示意图。

具体实施方式

[0092] 下面通过具体实施方式结合附图对本发明作进一步详细说明。

[0093] 实施例一：

[0094] 考虑到现有固定唤醒词语音唤醒方法存在的用户唤醒率低、误唤醒率高以及用户体验差的技术问题；本实施例提供了一种可以提高终端用户唤醒率、降低误唤醒率、提升用户体验的语音唤醒方法，如图 1 所示，本实施例的语音唤醒方法包括如下步骤：

[0095] 步骤 101：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储。

[0096] 本实施例方法可以采用语音采集装置，例如麦克风来采集终端用户发出的语音信号。本实施例方法中获取唤醒词识别结果和声纹特征的时序不限，可以先获取识别结果，后获取声纹特征，也可以同时获取。

[0097] 本实施例中唤醒词识别结果是对唤醒词语音信号进行语音识别得到的唤醒词文字内容或者命令。本实施例中声纹特征是语音信号中可以代表用户发声方式的参数，例如声音频率、响度等等。

[0098] 本步骤为进行语音唤醒之前的准备步骤，发生在用户自定义语音唤醒词的过程中，本步骤可以由终端的 AP 侧执行，即终端的主芯片，例如手机主 CPU 中 ARM 芯片来完成。

[0099] 为了提高语音唤醒的精确性，本实施例方法在采集样本特征的阶段，可以多次录制用户发出的唤醒词语音信号，然后根据多次录制的唤醒词语音信号得到最优的唤醒词识别结果和声纹特征。如图 2 所示，为终端获取用户自定义语音唤醒词样本特征的过程，由终端的 AP 侧完成：

[0100] 在终端启动用户自定义唤醒词设置之后，第一次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的唤醒词识别结果 1 和声纹特征 1，将识别结果 1 和声纹特征 1 存储至自定义唤醒词存储区中；

[0101] 终端第二次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的唤醒词识别结果 2 和声纹特征 2，当唤醒词识别结果 2 与唤醒词识别结果 1 相同时，利用声纹特征

Petitioner Exhibit 1004, Page 714 of 1052

2 优化声纹特征 1，当唤醒词识别结果 2 与唤醒词识别结果 1 不相同时，重新录制；进入第三次录制或者从第一次开始重新录制；

[0102] 终端第三次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的唤醒词识别结果 3 和声纹特征 3，当唤醒词识别结果 3 与唤醒词识别结果 1 相同时，利用声纹特征 3 优化上述已经优化过的声纹特征 1；

[0103] 依次类推，直到第 n 次录制完成。图 2 中 n 取 4，本实施例方法中 n 的取值可以根据实际需求进行设定。

[0104] 步骤 102：在所述终端处于休眠状态下，检测外界输入的语音信号。

[0105] 本步骤中可以由终端的麦克风来检测外界的语音信号，在终端休眠状态时，麦克风可以保持工作状态。

[0106] 本步骤中终端处于休眠状态下从硬件的角度可以包括：终端的主处理器处于休眠状态，从软件的角度可以包括：终端的操作系统处于休眠或锁定状态。

[0107] 步骤 103：获取所述语音信号的语音识别结果和声纹特征。

[0108] 本步骤中获取语音识别结果和声纹特征没有时序限制，可以先获取语音识别结果，后获取声纹特征，也可以先获取声纹特征，后获取语音识别结果，还可以同时获取语音识别结果和声纹特征。

[0109] 本步骤中语音信号的语音识别结果是对语音信息进行语音识别得到的文字内容或者命令。本实施例中声纹特征是语音信号中可以代表用户发声方式的参数，例如声音频率、响度等等。

[0110] 步骤 103-106 为语音唤醒过程中主要步骤，其可以在一个低功耗音频芯片执行，例如低功耗的 DSP 芯片。在进行语音唤醒之前，事先将上述获取的唤醒词识别结果和唤醒词语音信号的声纹特征下载到低功耗音频芯片的存储区，为语音唤醒做准备。

[0111] 步骤 104：获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果。

[0112] 本步骤中获取语音识别结果与存储的唤醒词识别结果的匹配结果方式包括：将语音识别结果与存储的唤醒词识别结果进行匹配，获取所述声纹特征与存储的声纹特征的匹配结果的方式包括：将获取的声纹特征与存储的声纹特征进行匹配；本步骤中获取所述语音结果与存储的唤醒词识别结果匹配结果，与获取声纹特征与存储的声纹特征的匹配结果过程时序不受限制，可以先执行前者，后执行后者，或者同时执行前者和后者。

[0113] 本实施例中的匹配结果包括：匹配成功与否，或者匹配分值；其中根据匹配分值是否在预设范围内来判定是否匹配成功。

[0114] 步骤 105：根据所述匹配结果判断是否需要唤醒所述终端，若是，执行步骤 106，若否，则执行步骤 107。

[0115] 具体地，当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端；否则，判断不需要唤醒所述终端。

[0116] 本步骤需要判断语音识别结果是否匹配成功和声纹特征是否匹配，这两个判断过程的时序可以任意，具体包括以下情况：

[0117] 1、先判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果是否

Petitioner Exhibit 1004, Page 715 of 1052

匹配成功，若是，则再判断所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情况下均判定不需要唤醒终端。

[0118] 2、先判断所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否匹配成功，若是，则再判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情况下均判定不需要唤醒终端。

[0119] 3、同时判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果是否匹配成功、以及所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否匹配成功，若均成功，则判定需要唤醒所述终端；若有一个匹配不成功则判定不需要唤醒终端。

[0120] 本步骤中匹配成功可以包括：匹配分值在预设范围内，或者相似度在预设范围之内。

[0121] 步骤 106：唤醒所述终端。

[0122] 本步骤可以通过发送中断事件唤醒终端；当语音唤醒步骤在低功耗音频芯片上执行时，低功耗音频芯片判断需要唤醒终端后，会上传中断事件给终端的主处理器唤醒处于休眠状态下的主处理器。

[0123] 步骤 107：不进行任何处理。

[0124] 本实施例方法中当判断不需要唤醒终端时，可以忽略当前输入的语音信号，执行唤醒步骤的芯片或器件可以进入休眠状态，以节省终端功耗。

[0125] 本实施例提供的语音唤醒方法，以终端用户自己声音作为声音样本，获取终端用户个体的声音特征（即识别结果和声纹特征）作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同（同一用户的发声方式相同），避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤醒失败，提供了唤醒率；另外，本实施例方法加入了语音识别结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端，降低了误唤醒率；本发明的方法还可以供用户自定义唤醒词，满足用户个性化的需求，提升了用户的体验。

[0126] 进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施例方法只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

[0127] 考虑到噪声环境对用户输入的语音信号的影响，进一步提高用户唤醒率降低误唤醒率，本实施例方法在上述步骤 102 之后，步骤 103 之前还可以包括：

[0128] 判断所述语音信号是否满足预设信号要求，若是，则执行步骤 103。

[0129] 具体地，可以判断输入的语音信号的信噪比是否在预设范围内，若是，则满足语音信号要求，或者语音信号的其他质量参数是否在预设范围内，若是，则满足语音信号要求，或者将输入的语音信号的声纹特征用用较简单的波形匹配，若成功，则满足信号要求。该步骤可以在低功耗音频芯片上完成。

[0130] 本实施例方法在判断语音信号不满足预设信号要求时，可以输入的语音信号丢弃；或者当该判断步骤是在低功耗音频芯片上置信时，该芯片进入休眠状态，以节省电能。

[0131] 优先地，如图 3 所示，上述判断所述语音信号是否满足预设信号要求的过程可以包括：

[0132] 步骤 301：对所述语音信号进行处理得到用户语音数据和背景噪声数据。

[0133] 由于外界输入的语音信号包括用户自己发出的语音信号和当前用户环境产生的

Petitioner Exhibit 1004, Page 716 of 1052

噪声信号，所以在进行信号判断时，可以将语音信号分离成用户语音数据和背景噪声数据。

[0134] 步骤 302：将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配。

[0135] 本步骤中用户语音数据和背景噪声数据匹配时序不受限制，可以先后，也可以同时。

[0136] 本实施例方法可以预置一些较典型的生活场景下的稳态或非稳态噪声声学统计模型，例如，汽车中，酒吧里，餐馆，家庭，办公室，地铁上，会议中等。

[0137] 具体地，本步骤可以包括如下步骤：

[0138] 将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值；

[0139] 将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值。

[0140] 步骤 303：根据匹配结果判断所述语音信号是否满足预设信号要求。

[0141] 本步骤中的匹配结果可以包括上述两个匹配过程得到的匹配分值，此时具体地判断过程可以包括：

[0142] 采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值；

[0143] 判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设信号要求，若否，则判定不满足预设信号要求，若否，则不满足预设信号求。

[0144] 例如对两个匹配分值进行加权处理得到一个匹配分值，然后判断该匹配分值是否超出置信阀值，若是，则判定满足预设信号要求。

[0145] 进一步考虑语音唤醒时终端的功耗问题，本实施例可以由一个音频处理器来完成语音唤醒的主要过程，事先将声音样本特征（唤醒词识别结果和唤醒词语音信号的声纹特征）下载到该音频处理器的存储区；在终端处于休眠状态时该音频处理器处于半 sleep 待机状态，由 mic 事件控制其工作唤醒；当 mic 检测到声音输入时唤醒多核音频处理器，执行语音唤醒；这样可以使得终端的主处理在休眠状态时，通过低功耗的音频处理器响应和处理外界输入的声音信息，节省终端的功耗，提高终端的待机时间。

[0146] 优先地，上述音频处理器可以采取分步式核心处理，Core1 用于处理接收到 mic 接收到外部声音信号的唤醒事件声音初步分析，即判断语音信号是否满足预设要求，Core2 用户后续的语音识别以及鉴权认证，即判断是否唤醒。当 mic 检测到语音输入时唤醒 Core1，Core1 执行判断语音信号是否满足预设要求的步骤；当 Core1 判断为是时，唤醒 Core2，并将语音信号传输给 Core2（优先地，Core1 将语音信号存储在 Core1、2 共享的存储区中，由 Core2 提取）；Core2 唤醒后执行上述步骤 103-106，完成唤醒鉴别，若通过，则由 Core2 向终端主处理器（例如终端 ARM 芯片）传送硬件中断事件，终端主处理器接收到中断事件后被唤醒由 sleep 模式切换到 Normal 模式正常运行。

[0147] 应当理解的是：本实施例还可将语音识别结果匹配或者声纹匹配的过程放在 Core1 中执行，只有当匹配成功时才唤醒 Core2，然后 Core2 唤醒之后执行声纹匹配或者语音结果匹配。

[0148] 本实施例通过多核音频处理器的分步式核心处理设计，可以实现在节省终端功耗或电能的效果，提升了用户体验。

[0149] 下面通过如图 4 所示的语音唤醒系统来具体介绍本实施例方法；如图 4 所示，该系

Petitioner Exhibit 1004, Page 717 of 1052

统包括：DSP 芯片（多核 DSP 芯片，可以包括至少两个核）、终端主 CPU 中的 ARM 芯片和麦克风；其中 DSP 芯片包括两个核单元以及一个唤醒存储单元（例如 Flash），第一核单元 Core1 和第二核单元 Core2；ARM 芯片包括唤醒词处理单元和唤醒词存储单元；图 4 所示的系统进行语音唤醒的具体过程包括：

[0150] 步骤 501：ARM 芯片中唤醒词处理单元录制用户所说的唤醒词语音信号，并生成唤醒词语音识别结果和声纹特征，将唤醒词语音识别结果和声纹特征存储在唤醒词存储单元中。

[0151] 具体地可以生成方式可以参考上述图 2 的描述。

[0152] 步骤 502：ARM 芯片将唤醒词语音识别结果、唤醒词语音信号的声纹特征和预置的背景噪声模型下载到 DSP 芯片，并存储在唤醒存储单元中。

[0153] 步骤 503：在 ARM 芯片处于休眠状态下，麦克风动态监测外界的语音输入，当检测到语音信号时发送唤醒事件唤醒 Core1。

[0154] 步骤 504：Core1 对检测到的语音信号进行处理得到用户语音数据和背景语音数据。

[0155] 步骤 505：Core1 将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值，将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

[0156] 步骤 506：Core1 采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值。

[0157] 步骤 507：Core1 判断所述语音信号的匹配分值是否大于预设阈值，若是，则执行步骤 508，若否，则执行步骤 512。

[0158] 步骤 508：Core1 将语音信号传输给 Core2 同时唤醒 Core2。

[0159] 步骤 509：Core2 唤醒后，对所述语音信号进行语音识别获得语音识别结果，并与存储的唤醒词识别结果进行匹配，若匹配成功，则执行步骤 510，若失败，则执行步骤 513。

[0160] 步骤 510：Core2 提取所述语音信号的声纹特征，并与存储的声纹特征进行匹配，若成功，执行步骤 511，若否，执行步骤 513。

[0161] 步骤 511：Core2 发送唤醒中断事件给 ARM 芯片唤醒 ARM 芯片。

[0162] 步骤 512：Core1 重新休眠，DSP 芯片休眠。

[0163] 步骤 513：Core2 重新休眠，语音唤醒流程结束。

[0164] DSP 芯片中 Core1 和 Core2 的语音算处理能力不同功耗有差异，当 DSP 芯片只有 Core1 工作时因为其运算能力和主频和处理能力有限，其运转时是 DSP 处在相对低功耗工作状态 (Normal)。如果 DSP 芯片 Core2 工作时其运算和处理能力较高，功耗相对较高，Core2 被唤醒时 DSP 芯片处于 Work 态。但唤醒 DSP 芯片全速运行 work 状态的功耗也是远小于终端主 CPU 芯片 Normal 态的运行功耗。

[0165] 本实施例可以将 DSP 芯片的运行状态分为低耗能 Core1 运行状态和相对高耗能运算的 Core2 运行态。从上图 5 可以看出，即使在复杂强噪音环境下，只要没有清晰人声特征输入，Core2 将不被唤醒，会使 DSP 芯片处于稳定低功耗状态。即使 Core2 被唤醒如果人声输入不是自定义的唤醒指令，或者唤醒指令不是录制唤醒特征的用户输入，手机主 CPU 芯片也不会被唤醒。通过硬件低功耗 DSP 芯片的阶梯层次的唤醒机制，个体化唯一化的鉴权

Petitioner Exhibit 1004, Page 718 of 1052

机制保证了手机语音唤醒的功耗控制，和误醒率的降低。

[0166] 应用本实施例方法，带有声纹特征的自定义唤醒词可以在降低了智能机中语音检测功耗控制优化难题的基础上，更好的适应特殊人群或者个体的个性化需求，在便捷操作的同时，让用户的语音控制更贴近于人们生活实际使用场景。

[0167] 实施例二：

[0168] 为进一步提升用户体验，在终端唤醒之后，还行可以进行多样性的快速和自动响应，例如打开应用程序等，本实施例在上述语音唤醒方法的基础上，提供了一种终端处理方法，如图 6 所示，包括如下步骤：

[0169] 步骤 601：利用实施例一所述的语音唤醒方法唤醒所述终端。

[0170] 具体唤醒过程参考实施例一的具体描述。

[0171] 步骤 602：所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0172] 上述步骤 602 可以由终端中的主处理器完成。

[0173] 本实施例方法可以预先设置自动定义唤醒词对应的响应操作，不同的唤醒词可以对应不同的响应操作；在用户通过语音唤醒词唤醒终端之后，终端根据用户输入语音信号的识别结果执行对应的响应操作，例如自动打电话或者自动打开聊天软件或者开启驾驶模式等。这里的唤醒词是文字内容，识别结果也是用户所说的文字内容。

[0174] 优先地，当通过唤醒中断事件来唤醒终端情况下，可以设置不同的唤醒词对应不同的唤醒中断事件，如果一个用户预先录制多个唤醒词，不同唤醒词对应不同的唤醒中断事件，终端对不同的唤醒中断事件，执行不同的预设操作或者架构不同的应用场景。这样达成即使在手机黑屏休眠时用户也可以通过不同的用户唤醒指令，进行多样性的快速和自动响应。例如用户自定义唤醒词为“打电话”时，在终端处于休眠状态下，在用户说出“打电话”时终端发出对应的唤醒中断事件来唤醒自身，并且终端在唤醒之后会根据预设的电话号码拨打电话。

[0175] 因此，上述步骤 601 语音唤醒方法中唤醒所述终端的过程包括：根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0176] 上述步骤 602 具体包括：根据所述唤醒中断事件的类型执行对应的预设操作。

[0177] 下面列举应用本实施例方法的具体场景：

[0178] 场景一：如果用户录制“急救 SOS，救命，妈妈”此类和用户情急时呼救口语化相关的指令做为唤醒词时，给用户一个默认的紧急唤醒的设置方法。允许用户设置的应急号码，即使在强噪音数据下，此指令的敏感度分值在用户无法手动呼叫的时候第一时间协助用户拨打出急救电话，或者发送出急救信息给接收者，同时后台开启录音或摄像功能存放在紧急文件目录中，并在 wifi 环境下上传网络云盘（此目录只能通过密码，用户手动删除，不被系统格式化删除），紧急文件生成时会以短信息的形式通知用户设置的紧急联系人。

[0179] 场景二：当监听到用户说“OK，Let Go！”，手机根据外界噪音环境是否类似汽车载内环境，根据陀螺仪和加速度传感器检测手机的动能位移增量变化，根据光感器检测是否手机屏幕置于非遮挡状态，为用户直接开启中兴驾驶助手。用户拿起手机下车时，根据噪音环境和传感器动量检测的变化自动关闭驾驶助手。

[0180] 场景三：幼儿英文发音练习：通过自定义唤醒词提前录制小朋友正学习中的英文

Petitioner Exhibit 1004, Page 719 of 1052

语句，其成功识别后对应设置儿童喜欢的特色应用或登陆模式。一个很喜欢玩手机游戏的小朋友，家长通过自定义唤醒词关联手机儿童模式，为了引导小朋友在游戏的同时巩固练习英文发音，可以将正在学习中的小朋友发音和语句预置为唤醒词，只有小朋友记忆正确并发音标准后才可以进入儿童模式进行游戏。不经意间规范小朋友的行为，提高学习热情。

[0181] 其他场景：也可以通过手机在线服务或者内置不同的歇后语，对联，或时尚俏皮话供用户选择录制，当用户说了对应的时尚俏皮话的内容时，允许用户。

[0182] 本实施例方法提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒率，实现趣味个性化的语音应用场景实现。

[0183] 实施例三：

[0184] 如图 7 所示，本实施例提供了一种语音唤醒装置，其特征在于，包括：采样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块；所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

[0185] 所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

[0186] 所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

[0187] 所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0188] 所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

[0189] 所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

[0190] 本实施例提供的语音唤醒装置，以终端用户自己声音作为声音样本，获取终端用户个体的声音特征（即识别结果和声纹特征）作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同（同一用户的发声方式相同），避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤醒失败，提供了唤醒率；另外，本实施例装置中加入了语音识别结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端，降低了误唤醒率；本实施例的装置还可以供用户自定义唤醒词，满足用户个性化的需求，提升了用户的体验。

[0191] 进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施例装置可实现只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

[0192] 优先地，如图 8 所示，在上述装置的基础上，本实施例语音唤醒装置还包括：第二判断模块；

[0193] 所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

[0194] 所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信号的语音识别结果和声纹特征。

[0195] 优先地，所述第二判断模块用于：

[0196] 对所述语音信号进行处理得到用户语音数据和背景噪声数据；

[0197] 将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

Petitioner Exhibit 1004, Page 720 of 1052

[0198] 根据匹配结果判断所述语音信号是否满足预设信号要求。

[0199] 优先地，如图 9 所示，在上述装置的基础上，本实施例唤醒装置中所述第一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

[0200] 所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是否满足预设信号要求，若是，唤醒所述第二核单元；

[0201] 所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的语音识别结果和声纹特征。

[0202] 优先地，本实施例中所述第一判断模块用于当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

[0203] 本实施例装置提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒率。

[0204] 如图 10 所示，本实施例还提供了一种终端，包括处理模块和如上任一所述的语音唤醒装置；

[0205] 所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0206] 优先地，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0207] 所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

[0208] 优先地，本实施例中处理模块集成在 ARM 芯片；所述语音唤醒装置集成在低功耗音频芯片上。

[0209] 本实施例提供的终端提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒率，实现趣味个性化的语音应用场景。

[0210] 以上内容是结合具体的实施方式对本发明所作的进一步详细说明，不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干简单推演或替换，都应当视为属于本发明的保护范围。

Petitioner Exhibit 1004, Page 721 of 1052

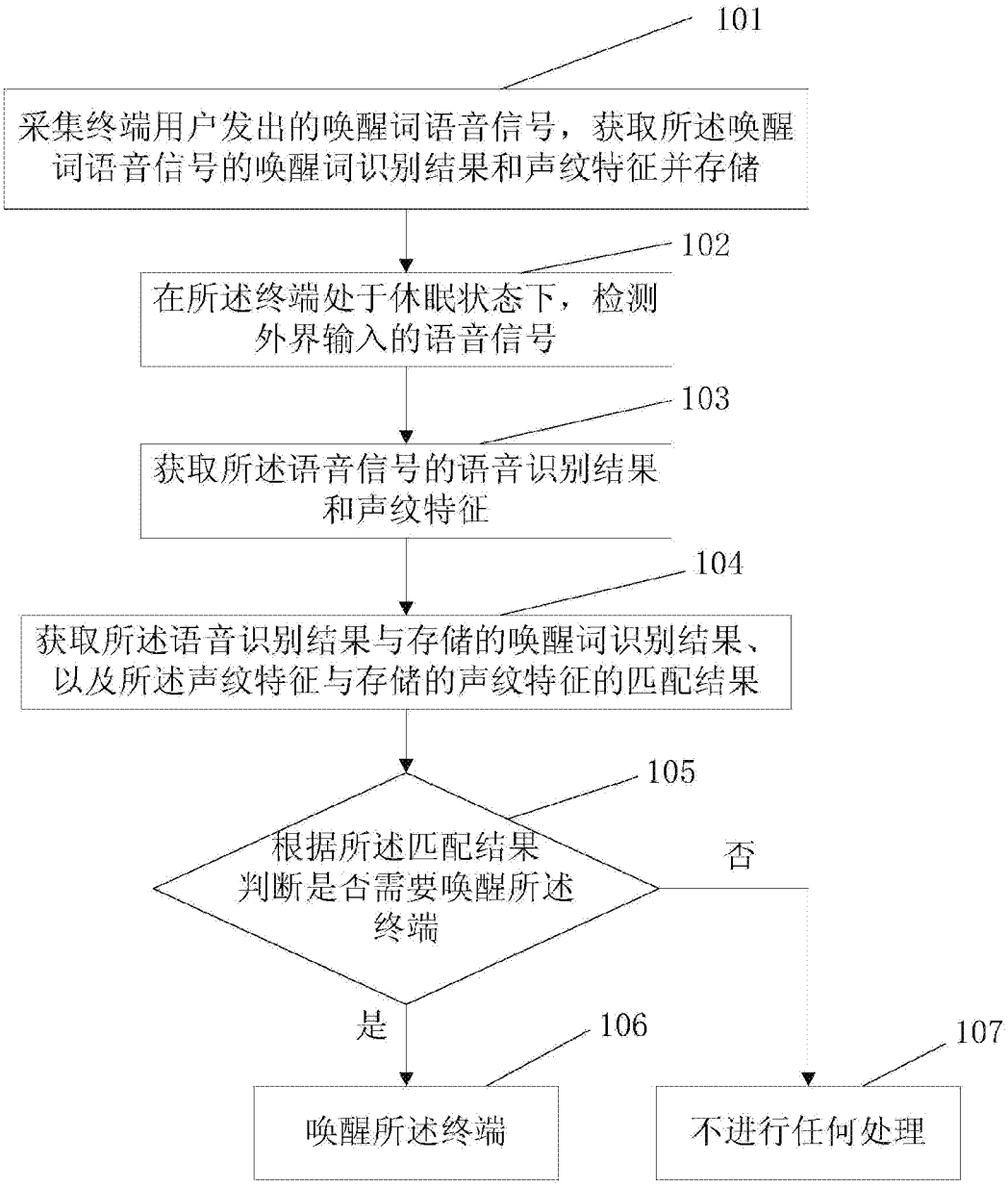


图 1

Petitioner Exhibit 1004, Page 722 of 1052

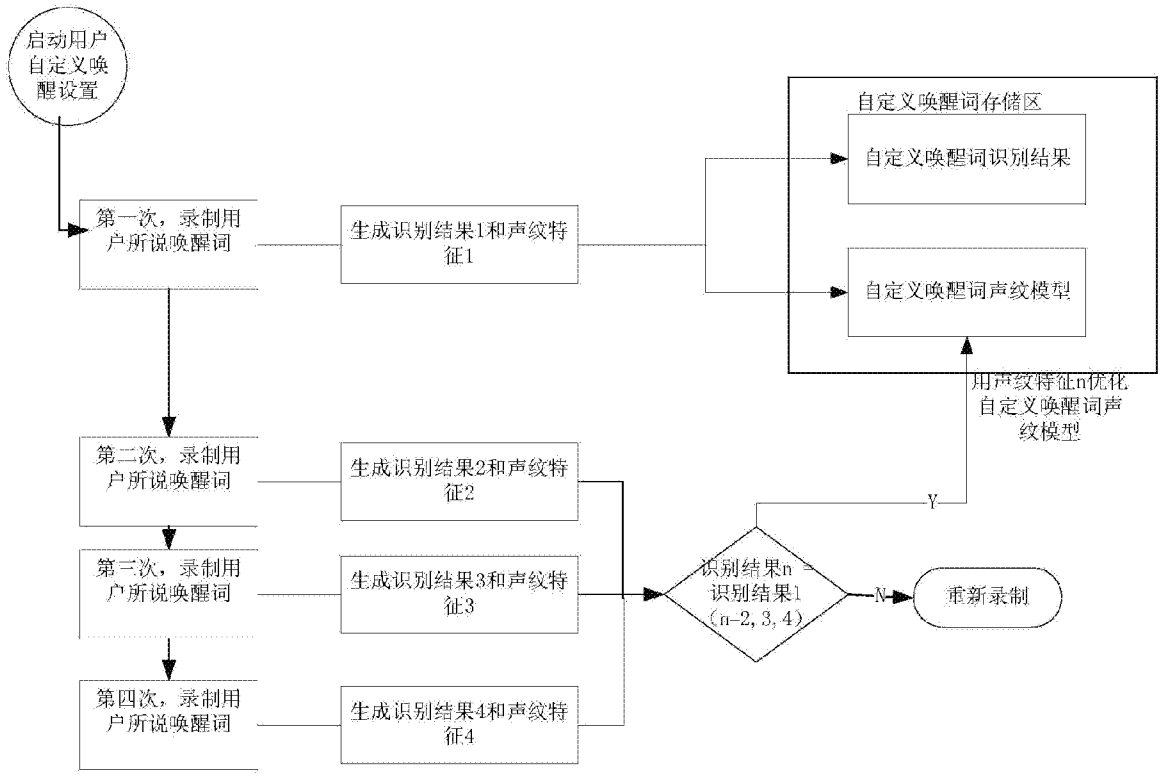


图 2

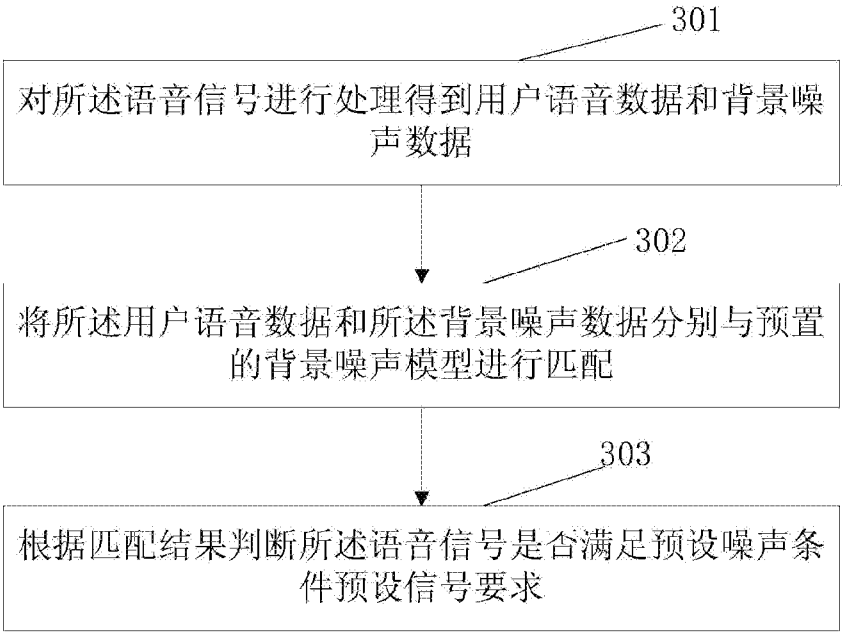


图 3

Petitioner Exhibit 1004, Page 723 of 1052

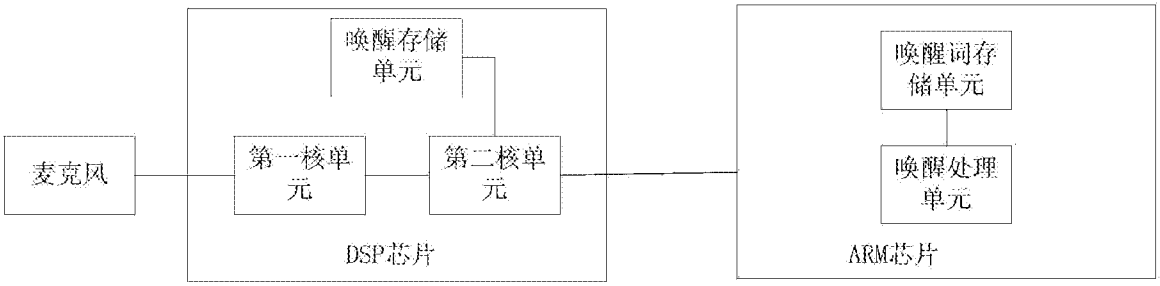


图 4

Petitioner Exhibit 1004, Page 724 of 1052

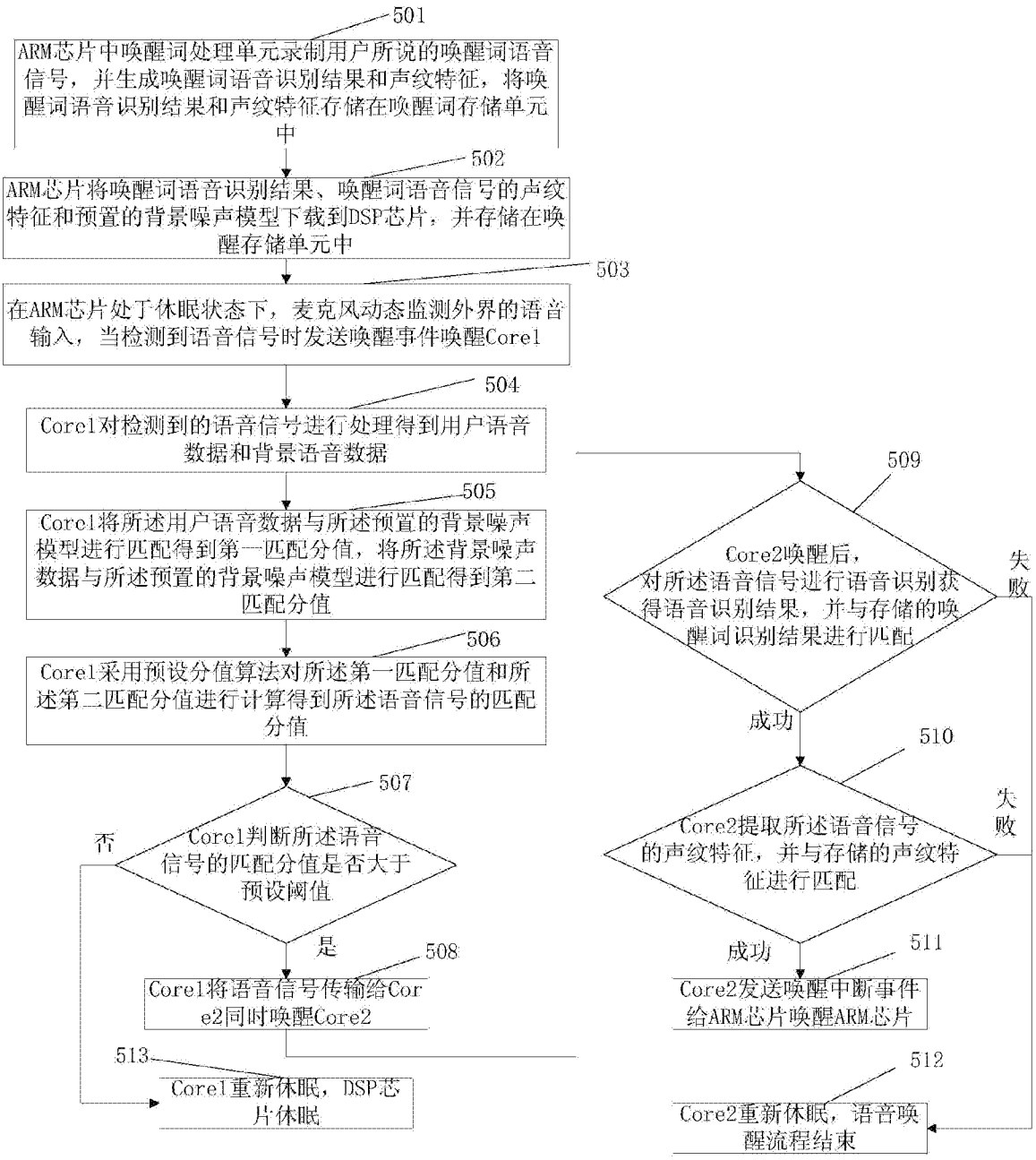


图 5

Petitioner Exhibit 1004, Page 725 of 1052

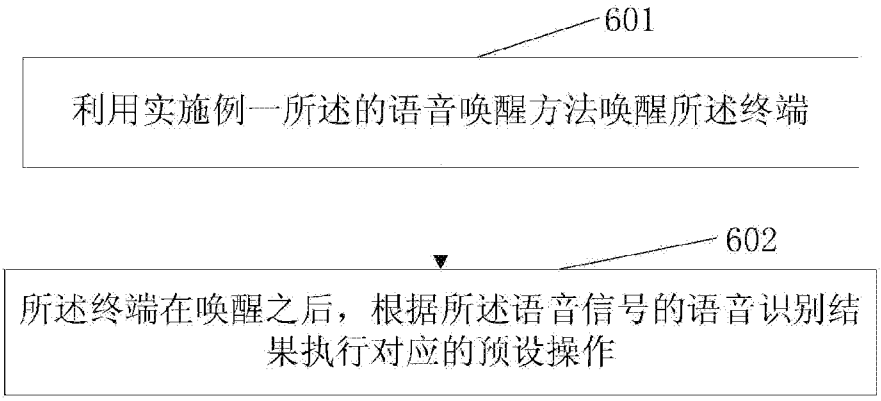


图 6

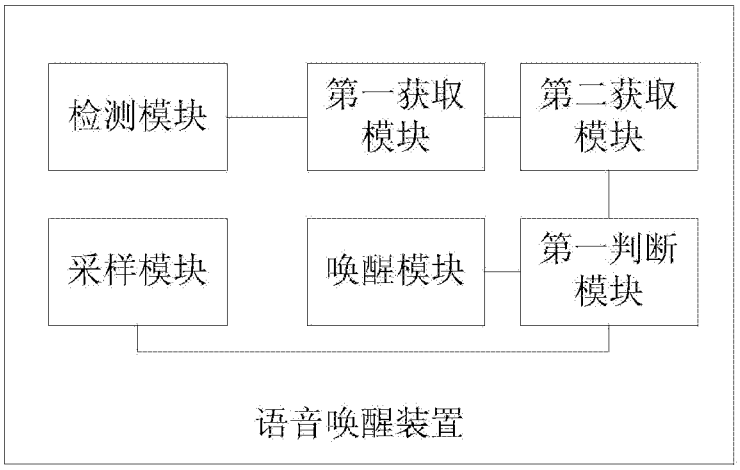


图 7

Petitioner Exhibit 1004, Page 726 of 1052

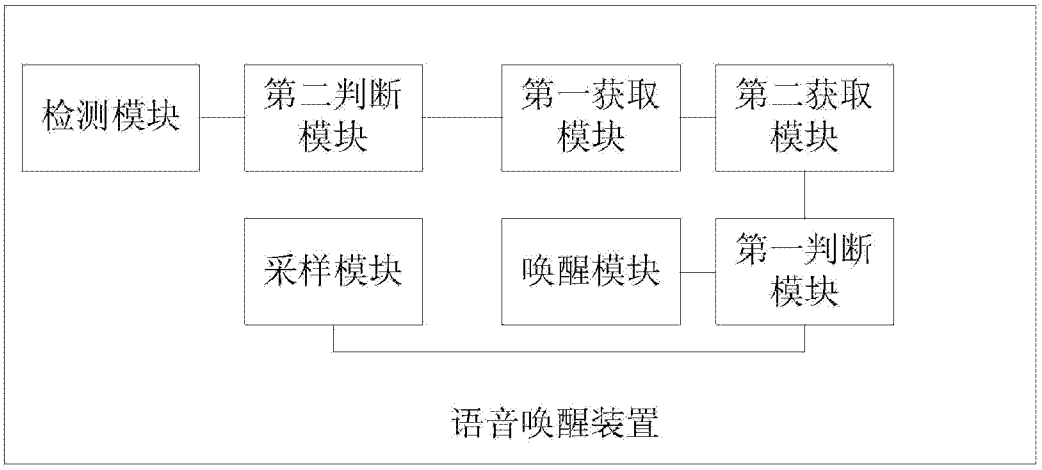


图 8

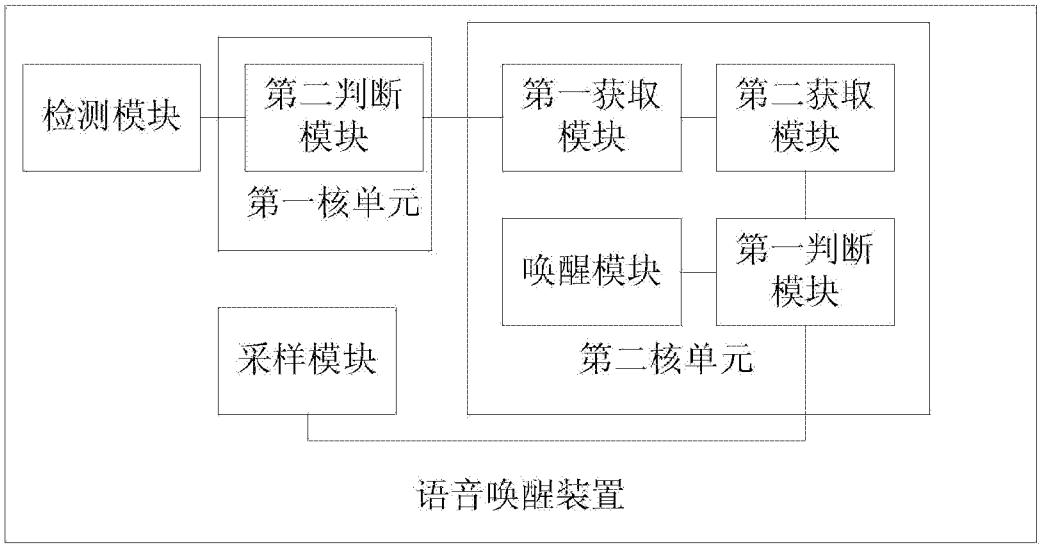


图 9

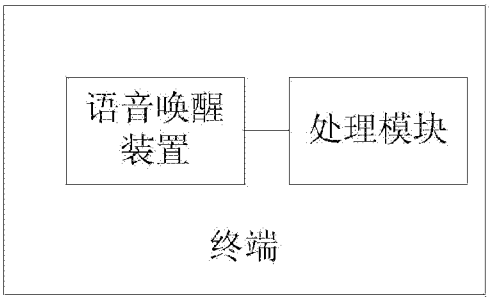


图 10

Petitioner Exhibit 1004, Page 727 of 1052

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 08/16/2019

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
08/16/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

IP.Inbox@nuance.com
N0484_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@wolfgreenfield.com

Office Action Summary	Application No. 15/238,238	Applicant(s) Ganong et al.	
	Examiner HUY C HO	Art Unit 2644	AIA (FITF) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 30 April 2019.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-21 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1-21 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 16 August 2016 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date 05/22/2019.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Continued Examination Under 37 CFR 1.114

2. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on **04/30/2019** has been entered.

Response to Arguments

3. Applicant's arguments filed **04/30/2019** have been fully considered but they are not persuasive for the following reasons.

With respect to the arguments that no combination of the cited references, i.e., Freeman and Mozer, would have taught or suggested "performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input" and "determining, while in the low power

mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase," and "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command", and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server".

The examiner respectfully disagrees because:

Freeman teaches "performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input" (*Freeman, Figure 2, voice-to-command analyzer 216, Figure 3, Col. 5 Lines 13-67, Col. 6 Lines 1-10; Figure 4, Boxes 402, 404, 406,408; Figure 5, Boxes 504, 506, 508, 510: a portable electronic device in a low power mode performing audio pickup to recognize a command from the audio/voice input*), and "determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase," (*Freeman, Figure 4, Boxes 402, 404, 406, 408: receiving*

audio input, determining whether the audio input pertains to at least one of the commands in a command set).

Mozer teaches "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]: a speech recognizer device receives acoustic input signal to process the acoustic input signal at stage 1 (lower power mode) listening for trigger words or phrases such as travel, purchases, entertainment, research, food, or electronics, stage 2 recognizer uses more accurate speech recognition algorithm to perform the speech recognition; Stage 2 recognizer (over a network device or a server) may use large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning);* and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server" (*Mozer, pp [36]: after stage 2*

recognizer perform the large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning, stage 1 or stage 2 recognizers may decide to save information from previous discussions to better classify, solve problems and assist).

Therefore, Freeman, in combination of Mozer, teach and suggest the argued features/limitations, i.e., "performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input" and "determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase," and "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command", and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server".

As such, the argued features/limitations were written such that they read upon Freeman and Mozer as discussed above.

Claim Rejections - 35 USC § 103

4. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

5. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

7. Claims 1-12, 15-19 and 21 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S.

Patent No. 8,977,255) and further in view of Mozer (U.S. Patent Publication No. 20103/0080167).

Regarding claim 1, (Currently Amended) Freeman teaches a device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to perform at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512);*

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input (*Freeman, Figure 2, voice-to-command analyzer 216, Figure 3, Col. 5 Lines 13-67, Col. 6 Lines 1-10; Figure 4, Boxes 402, 404, 406,408; Figure 5, Boxes 504, 506, 508, 510: a portable electronic device in a low power mode performing audio pickup to recognize a command from the audio/voice input*);

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase (*Freeman, Figure 4,*

Boxes 402, 404, 406, 408: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set);

in response to determining, that the at least one first word or phrase matches a trigger word or phrase (*Freeman, Figure 4, Boxes 402, 404, 406, 408: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set);*

Freeman does not teach "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server".

Mozer teaches "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer,*

Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]: a speech recognizer device receives acoustic input signal to process the acoustic input signal at stage 1 (lower power mode) listening for trigger words or phrases such as travel, purchases, entertainment, research, food, or electronics, stage 2 recognizer uses more accurate speech recognition algorithm to perform the speech recognition; Stage 2 recognizer (over a network device or a server) may use large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning); and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server" (*Mozer, pp [36]: after stage 2 recognizer perform the large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning, stage 1 or stage 2 recognizers may decide to save information from previous discussions to better classify, solve problems and assist).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 15, (Currently Amended) Freeman teaches at least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode (*Freeman, the Abstract, Figure 1, Col. 1 Lines 60-65*), the method comprising:

performing, at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*;

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the

acoustic input (*Freeman, Figure 2, voice-to-command analyzer 216, Figure 3, Col. 5 Lines 13-67, Col. 6 Lines 1-10; Figure 4, Boxes 402, 404, 406,408; Figure 5, Boxes 504, 506, 508, 510: a portable electronic device in a low power mode performing audio pickup to recognize a command from the audio/voice input*);

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase (*Freeman, Figure 4, Boxes 402, 404, 406, 408: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set).*

Freeman does not teach "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" and "respond to at least one voice command in the acoustic input recognized by the at least one server".

Mozer teaches "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3;*

Figures 4, 5, 6 and 8, pp [55]-[56]: a speech recognizer device receives acoustic input signal to process the acoustic input signal at stage 1 (lower power mode) listening for trigger words or phrases such as travel, purchases, entertainment, research, food, or electronics, stage 2 recognizer uses more accurate speech recognition algorithm to perform the speech recognition; Stage 2 recognizer (over a network device or a server) may use large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning); and respond to at least one voice command in the acoustic input recognized by the at least one server" (*Mozer, pp [36]: after stage 2 recognizer perform the large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning, stage 1 or stage 2 recognizers may decide to save information from previous discussions to better classify, solve problems and assist).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to

modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 21, (Currently Amended) Freeman teaches a device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*:

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input (*Freeman, Figure 2, voice-to-command analyzer 216, Figure 3, Col. 5 Lines 13-67, Col. 6 Lines 1-10; Figure 4, Boxes 402, 404, 406,408; Figure 5, Boxes 504, 506, 508, 510: a portable electronic device in a low power mode performing audio pickup to recognize a command from the audio/voice input*);

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase (*Freeman, Figure 4, Boxes 402, 404, 406, 408: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set).*

Freeman does not teach "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server".

Mozer teaches "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]: a speech recognizer device receives acoustic input signal to process the acoustic input*

signal at stage 1 (lower power mode) listening for trigger words or phrases such as travel, purchases, entertainment, research, food, or electronics, stage 2 recognizer uses more accurate speech recognition algorithm to perform the speech recognition; Stage 2 recognizer (over a network device or a server) may use large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning); and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server" (*Mozer, pp [36]: after stage 2 recognizer perform the large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning, stage 1 or stage 2 recognizers may decide to save information from previous discussions to better classify, solve problems and assist).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to

modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 2, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 3, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 4, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor (*Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 5, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech (*Freeman, Col. 4 Lines 45-62, Col. 5 Lines 13-33*).

Regarding claim 6, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input determining whether the at least one first word or phrase matches the trigger word or phrase (*Mozer, Paragraph(s) [33]; Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]*).

Regarding claim 7, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the second processor is configured to perform automatic speech recognition on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command (*Mozer, Paragraph(s) [36]*).

Regarding claim 8, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to determining whether the at least one first word or phrase matches the trigger word or phrase (*Mozer, Paragraph(s) [33]; Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]*).

Regarding claim 9, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Mozer, Paragraph(s) [26]-[28], [46]-[47]*).

Regarding claim 10, (Previously Presented) Freeman, as modified by Mozer, teaches the mobile device of claim 3, wherein the device is further configured to transition to an active mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 5, boxes 512).*

Regarding claim 11, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the low power mode, and wherein the trigger word or phrase, when detected, is sufficient to engage the device (*Mozer, Paragraph(s) [65]-[66]: specific words such as "vacation", "travel", etc.; Mozer, Paragraph(s) [33]; Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]).*

Regarding claim 12, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input (*Freeman, Figure 5, box 506).*

Regarding claim 16, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Mozer, Figure 1B, Paragraph(s) [31]-[33], [36]*).

Regarding claim 17, (Currently Amended) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to determining whether the at least one first word or phrase matches the trigger word or phrase (*Mozer, Paragraph(s) [33]; Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]*).

Regarding claim 18, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Freeman, Figure 5, boxes 504, 506, 508*).

Regarding claim 19, (Previously Presented) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing at least one task specified by the voice command with the device remaining in the low power mode (*Freeman, Figure 5, boxes 506, 508, 510*).

8. Claims 13, 14 and 20 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255), further in views of Mozer (U.S. Patent Publication No. 20103/0080167) and LeBeau et al. (U.S. Patent Publication No. 2012/0034904).

Regarding claim 13, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the mobile device, an occurrence of an event on the device, recent activity on the mobile device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual

cue selected from a group consisting of: time of day, motion of the mobile device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 14, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of

the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 20, (Previously Presented) Freeman, as modified by Mozer, does not teach the at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient

sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUY C HO whose telephone number is (571)270-1108. The examiner can normally be reached on M-F 8AM-5PM.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, KATHY WANG-HURST can be reached on (571)270-5371. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications

may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HUY C HO/
Primary Examiner, Art Unit 2644

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

CPC - Searched*		
Symbol	Date	Examiner
G10L15/20; G10L15/22; G10L15/24; G10L15/25; G10L2015/223; G10L2015/226; H04M2250/74; H04W52/0251; H04W52/0254.	4/20/2017	HH

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner
379	88.01	4/20/2017	HH

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
EAST Database Search	4/20/2017	HH
Inventor Search	4/20/2017	HH
Search update	1/10/2018	HH
Search update	8/20/2018	HH
Search update	02/24/2019	HH
Search update	08/13/2019	HH

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner
379	88.02	4/20/2017	HH
704	246-255	4/20/2017	HH

/HUY C HO/ Primary Examiner.Art Unit 2644	

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/03/18 21:04
S2	104	(("GANONG") near3 ("William")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S3	16	(("VAN") near3 ("MULBREGT") near3 ("Paul")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S4	38	(("SEJNOHA") near3 ("Vladimir")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S5	33	(("WILSON") near3 ("Glen")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:20
S6	220	(("NUANCE") near3 ("COMMUNICATIONS") near3 ("INC")).AANM.	USPAT	OR	OFF	2017/04/20 10:20
S7	43	("8977255" | "9112984" | "9361885" | "20070129949" | "20120329389" | "7574361" | "20100124896" | "20030040339" | "20100289994" | "20140274211" | "20140278435" | "8548176" | "20100185448" | "20110054899" | "20130080167" | "20130080171" | "20130289994" | "20130339028" | "20140012573" | "20140012586" | "5428707" | "5774859" | "5802305" | "5983186" | "6070140" | "6092043" | "6397186" | "6408396" | "6411926" | "6449496" | "6594630" | "6756700" | "6941265" | "6965786" | "7114090" | "7567827" | "7720683" | "7774204" | "8056070" | "8181046" | "8190420" | "8285545" | "8666751").PN.	US-PGPUB; USPAT	OR	OFF	2017/04/20 11:21
S8	5	"20120034904"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:10
S9	9	"9361885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:32
S10	33784	(379/88.01-88.02|G10L15/20|G10L15/22|G10L15/24|G10L15/25|G10L2015/223|G10L2015/226|H04M2250/74|H04W52/0251|H04W52/0254|).ccls,cpc.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:11
S11	1	("15".src. and "238238".ap.) or "15238238".RLAN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/04/20 16:12
S12	25130	S10 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT;	OR	ON	2017/04/20 16:13

			USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB			
S13	4617	S12 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S14	2692	S13 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S15	563	S14 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S16	362	S15 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S17	17	S16 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:33
S18	9564	(704/246-255).ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S19	8758	S18 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S20	1733	S19 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S21	1338	S20 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS;	OR	ON	2017/04/20 16:36

			EPO; JPO; DERWENT; IBM_TDB			
S22	212	S21 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S23	104	S22 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S24	3	S23 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S25	3	"9646610"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:41
S26	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/09/20 13:42
S27	8	"20100121636"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:42
S28	5	"20120035924"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:43
S29	6	"20120127072"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:45
S30	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/03/09 10:35
S31	12	"20140012573"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	ON	2018/03/09 10:53

			DERWENT; IBM_TDB			
S32	10	"20100009719"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2018/03/09 10:55
S33	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/05/30 09:37
S34	78	"9940936"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:18
S35	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2019/07/26 18:19
S36	3	"20140249817"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:20
S37	2	"20140365225"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:24
S38	2	"20160189706"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:26
S39	1050	UTSM	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/08/11 10:17

8/ 11/ 2019 12:00:26 PM
C:\ Users\ hho3\ Documents\ EAST\ Workspaces\ 15238238.wsp

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong, III	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		9,940,936	B2	Sejnoha et al.	04-10-2018
		2014-0249817	A1	Hart et al.	09-04-2014
		2014-0365225	A1	Haiut	12-11-2014
		2016-0189706	A1	Zopf et al.	06-30-2016

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		CN	101650943	A	Chinese Acad Inst Acoustics	02-17-2010	Y
		CN	103021409	A	Anhui Ustc Iflytek Co	04-03-2013	Y
		CN	103632668	A	Beijing Baidu Netcom Sci and Tec.	03-12-2014	Y
		CN	104575504	A	Univ Shanghai	04-29-2015	Y
		CN	105575395	A	Zte corp	05-11-2016	Y
		CN	106098059	A	Univ Shanghai Jiaotong	11-09-2016	Y
		EP	2 899 955	A1	Huawei Device	07-29-2015	
		WO	2014/066192	A1	Microsoft Corp.	05-01-2014	

OTHER ART — NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		International Search Report and Written Opinion for International Application No. PCT/CN2016/105343 mailed September 21, 2017 (N0484.71177WO00).	
		International Search Report and Written Opinion for International Application No. PCT/US2016/037495 mailed December 05, 2016 (N0484.71178WO00).	
		International Preliminary Report on Patentability for International Application No. PCT/US2016/037495 mailed December 27, 2018 (N0484.71178WO00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:
/HUY C HO/	08/13/2019

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7243457.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	04/20/2017	01/10/2018	08/20/2018	02/24/2019	08/13/2019				
	1	✓	✓	✓	✓	✓				
	2	✓	✓	✓	✓	✓				
	3	✓	✓	✓	✓	✓				
	4	✓	✓	✓	✓	✓				
	5	✓	✓	✓	✓	✓				
	6	✓	✓	✓	✓	✓				
	7	✓	✓	✓	✓	✓				
	8	✓	✓	✓	✓	✓				
	9	✓	✓	✓	✓	✓				
	10	✓	✓	✓	✓	✓				
	11	✓	✓	✓	✓	✓				
	12	✓	✓	✓	✓	✓				
	13	✓	✓	✓	✓	✓				
	14	✓	✓	✓	✓	✓				
	15	✓	✓	✓	✓	✓				
	16	✓	✓	✓	✓	✓				
	17	✓	✓	✓	✓	✓				
	18	✓	✓	✓	✓	✓				
	19	✓	✓	✓	✓	✓				
	20	✓	✓	✓	✓	✓				
	21				✓	✓				

Bibliographic Data

Application No: 15/238,238

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Huy C. Ho/ HH

Examiner's Signature Initials

Title: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
08/16/2016	455	2644	N0484.71090US04
RULE			

APPLICANTS

Nuance Communications, Inc., Burlington, MA,

INVENTORS

William F. Ganong III, Brookline, MA, UNITED STATES

Paul Adrian Van Mulbregt Wayland, MA, UNITED STATES

Vladimir Sejnoha Lexington, MA, UNITED STATES

Glen Edward Wilson Maynard, MA, UNITED STATES

CONTINUING DATA

This application is a CON of 13795933 03/12/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

08/25/2016

STATE OR COUNTRY

UNITED STATES

ADDRESS

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

FILING FEE RECEIVED

$1,600

Electronic Acknowledgement Receipt

EFS ID:	37132535
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Charles Nicholas Ruggiero/Joanne Romaniello
Filer Authorized By:	Charles Nicholas Ruggiero
Attorney Docket Number:	N0484.71090US04
Receipt Date:	11-SEP-2019
Filing Date:	16-AUG-2016
Time Stamp:	14:23:21
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		20190911_0013359USX1_4705A_Transmittal_POA_Statement.pdf	293310 b156e3fa1ef027b012af09d5a81584c5dda61b80	yes	4

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Power of Attorney	1	1
	Power of Attorney	2	2
	Assignee showing of ownership per 37 CFR 3.73	3	4

Warnings:

The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing

Information:

2	Application Data Sheet	20190911_0013359USX1_4705A_ADS_aia0014_Corrected_ADS.pdf	134527 7cc3f4870b63a75e305af36e9dcae4e38385b63e	no	9

Warnings:

Information:

This is not an USPTO supplied ADS fillable form

Total Files Size (in bytes):	427837

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0404.71098US04~~0013359USX1/4705A
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2:

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	William	F.	GANONG	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Brookline	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	118 Sumner Road		
Address 2			
City	Brookline	State/Province	MA
Postal Code	02146	Country	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
Mr.	Paul	Adrian	Van MULBREGT	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Wayland	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	30 Bow Road		
Address 2			
City	Wayland	State/Province	MA
Postal Code	01778	Country	US

Inventor 3 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
Mr.	Vladimir		SEJNOHA	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0484.71090US04~~0013359USX1/4705A
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

City	Lexington	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	10 Flintlock Road		
Address 2			
City	Lexington	State/Province	MA
Postal Code	02420	Country	US

Inventor 4 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Glen	Edward	WILSON	

Residence Information (Select One) (•) US Residency ( ) Non US Residency ( ) Active US Military Service

City	Maynard	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	13 Sherman Sreet		
Address 2			
City	Maynard	State/Province	MS
Postal Code	01754	Country	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	~~91997~~ 157176		
Email Address	DOCKETING@OGRP.COM	Add Email	Remove Email

Application Information:

Title of the Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND		
Attorney Docket Number	~~N0484.71090US04~~0013359USX1/4705A	Small Entity Status Claimed	☐
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	10	Suggested Figure for Publication (if any)	

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0484.71090US04~~0013359USX1/4705A
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	◉ Customer Number	○ US Patent Practitioner	○ Limited Recognition (37 CFR 11.9)
Customer Number	~~91337~~ 157176		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status			Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Continuation of	13/795933	2013-03-12

Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0484.71090US04~~ 0013359USX1/4705A
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

Remove

Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)

Additional Foreign Priority Data may be generated within this form by selecting the **Add** button.

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0484.71090US04~~0013359USX1/4705A
	Application Number	15/238,298
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0484.71090US04~~ 0013359USX1/4705A
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

◉ Assignee	○ Legal Representative under 35 U.S.C. 117	○ Joint Inventor
○ Person to whom the inventor is obligated to assign.	○ Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor:	

If the Applicant is an Organization check here. ☒

Organization Name	NUANCE COMMUNICATIONS, INC.

Mailing Address Information For Applicant:

Address 1	One Wayside Road		
Address 2			
City	Burlington	**State/Province**	MA
Country	US	Postal Code	01803
Phone Number		Fax Number	
Email Address			

Additional Applicant Data may be generated within this form by selecting the Add button.

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0484.71090US04~~ 0013359USX1/4705A
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

If the Assignee or Non-Applicant Assignee is an Organization check here. ☒

Organization Name	NUANCE COMMUNICATIONS, INC.

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1	One Wayside Road		
Address 2			
City	Burlington	**State/Province**	MA
Country i	US	Postal Code	01803
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button.

Signature:

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Charles N.J. Ruggiero/			Date (YYYY-MM-DD)	2019-09-11
First Name	Charles N.J.	Last Name	Ruggiero	Registration Number	28468

Additional Signature may be generated within this form by selecting the Add button.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~N0484.71090US04~~ 0013359USX1/4705A
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1 The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3 A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent CooperationTreaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/96 (08-12)
Approved for use through 01/31/2013. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Nuance Communications, Inc.

Application No./Patent No.: 13/795,933 Filed/Issue Date: 2013-03-12

Titled: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Nuance Communications, Inc., a corporation

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. ☑ The assignee of the entire right, title, and interest.

2. ☐ An assignee of less than the entire right, title, and interest (check applicable box):

   ☐ The extent (by percentage) of its ownership interest is _____%. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   ☐ There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. ☐ The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. ☐ The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. ☐ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

B. ☑ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Inventors To: Nuance Communications, Inc.

   The document was recorded in the United States Patent and Trademark Office at

   Reel 040342, Frame 0893, or for which a copy thereof is attached.

2. From: _____ To: _____

   The document was recorded in the United States Patent and Trademark Office at

   Reel _____, Frame _____, or for which a copy thereof is attached.

This collection of information is required by 37 CFR 3.73(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/AIA/96 (08-12)
Approved for use through 01/31/2013. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☑ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Charles N.J. Ruggiero/	2019-09-11
Signature	Date
Charles N.J. Ruggiero	28,468
Printed or Typed Name	Title or Registration Number

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82B (07-13)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

☑ I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above: 000157176

OR

☐ I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:

☑ The address associated with the above-mentioned Customer Number

OR

☐ The address associated with Customer Number:

OR

☐ Firm or Individual Name	----------				
Address	----------				
City	----------	State	--	Zip	-----
Country	--				
Telephone	----------	Email	----------		

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

NUANCE COMMUNICATIONS, INC.

☐ Inventor or Joint Inventor (title not required below)

☐ Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)

☑ Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)

☐ Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature		Date (Optional)	April 24, 2019
Name	Leanne Fitzgerald		
Title	VP, Associate General Counsel		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

☑ Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82A (07-13)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5, unless the application number and filing date are identified in the Power of Attorney by Applicant form. If neither form PTO/AIA/82A nor form PTO/AIA82B identifies the application to which the Power of Attorney is directed, the Power of Attorney will not be recognized in the application.

Application Number	15/238,238
Filing Date	2016-08-16
First Named Inventor	GANONG, William F.
Title	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Art Unit	2644
Examiner Name	Ho, Huy C.
Attorney Docket Number	0013359USX1/4705A

SIGNATURE of Applicant or Patent Practitioner

Signature	/Charles N.J. Ruggiero/	Date (Optional)	2019-09-11
Name	Charles N.J. Ruggiero	Registration Number	28,846
Title (if Applicant is a juristic entity)	Attorney for Applicant		
Applicant Name (if Applicant is a juristic entity)	Nuance Communications, Inc.		

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. If more than one applicant, use multiple forms.

☑ *Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	0013359USX1/4705A

CONFIRMATION NO. 5864

157176
Ohlandt Greeley Ruggiero & Perle, L.L.P.-VoiceBox
One Landmark Square, 10th Floor
Stamford, CT 06901

POA ACCEPTANCE LETTER

OC000000111122379

Date Mailed: 09/13/2019

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 09/11/2019.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/rmohamed/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04

CONFIRMATION NO. 5864

POWER OF ATTORNEY NOTICE

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

OC000000111122332

Date Mailed: 09/13/2019

NOTICE REGARDING CHANGE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 09/11/2019.

- The Power of Attorney to you in this application has been revoked by the applicant. Future correspondence will be mailed to the new address of record(37 CFR 1.33).

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/rmohamed/

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82B (07-13)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

[X] I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above: 91337

OR

[ ] I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:

[X] The address associated with the above-mentioned Customer Number

OR

[ ] The address associated with Customer Number:

OR

[ ] Firm or Individual Name					
Address					
City		State		Zip	
Country					
Telephone		Email			

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

Nuance Communications, Inc.

[ ] Inventor or Joint Inventor (title not required below)

[ ] Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)

[X] Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)

[ ] Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature		Date (Optional)	
Name	Scott Ouellette		
Title	Senior Director -- Intellectual Property		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

[ ] Total of forms are submitted.

7648792.1

Electronic Acknowledgement Receipt

EFS ID:	37621826
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	157176
Filer:	Andrew J. Tibbetts/Theressa Bresnahan
Filer Authorized By:	Andrew J. Tibbetts
Attorney Docket Number:	0013359USX1/4705A
Receipt Date:	31-OCT-2019
Filing Date:	16-AUG-2016
Time Stamp:	15:17:46
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		N048471090US04-POA-AJT.pdf	87804 405f54407d302ebd3b75f8776822e3d5f7b72e1e	yes	2

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Power of Attorney	1	1
	Power of Attorney	2	2

Warnings:

The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing

Information:

Total Files Size (in bytes):	87804

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82A (07-13)
Approved for use through 03/31/2021. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5, unless the application number and filing date are identified in the Power of Attorney by Applicant form. If neither form PTO/AIA/82A nor form PTO/AIA82B identifies the application to which the Power of Attorney is directed, the Power of Attorney will not be recognized in the application.

Application Number	15/238,238-Confirmation No. 5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Title	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

SIGNATURE of Applicant or Patent Practitioner

Signature	/Andrew J. Tibbetts/	Date (Optional)	October 31, 2019
Name	Andrew J. Tibbetts	Registration Number	65,139
Title (if Applicant is a juristic entity)	Attorney		
Applicant Name (if Applicant is a juristic entity)	Nuance Communications, Inc.		

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. If more than one applicant, use multiple forms.

☐ *Total of _____ forms are submitted.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: October 31, 2019 Electronic Signature for Theresa A. Bresnahan: /Theresa A. Bresnahan/

7654508.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	0013359USX1/4705A

CONFIRMATION NO. 5864

POWER OF ATTORNEY NOTICE

157176
Ohlandt Greeley Ruggiero & Perle, L.L.P.-Cerence
One Landmark Square, 10th Floor
Stamford, CT 06901

OC000000112379846

Date Mailed: 11/05/2019

NOTICE REGARDING CHANGE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 10/31/2019.

- The Power of Attorney to you in this application has been revoked by the applicant. Future correspondence will be mailed to the new address of record(37 CFR 1.33).

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/ltaba/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	0013359USX1/4705A

CONFIRMATION NO. 5864

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

POA ACCEPTANCE LETTER

OC000000112379868

Date Mailed: 11/05/2019

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 10/31/2019.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/ltaba/

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 15/238,238	Filing Date 08/16/2016	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $80 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $420 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	12/16/2019	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 21	Minus	** 21	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 3	Minus	*** 3	= 0		x $460 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/ANTHONY TYSON/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Andrew J. Tibbetts/Eileen MacKenzie
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 1 month with $0 paid	1251	1	200	200
Miscellaneous:				
			Total in USD ($)	**200**

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: December 16, 2019 Electronic Signature for: /Eileen M. MacKenzie/

RESPONSE TO NON-FINAL OFFICE ACTION UNDER 37 C.F.R. § 1.111

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Office Action dated August 16, 2019, please amend the above-identified U.S. patent application as follows:

Amendments to the Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 8 of this paper.

PTO/AIA/22 (03-13)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) N0484.71090US04
Application Number 15/238,238-Conf. #5864	Filed August 16, 2016
For METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	
Art Unit 2644	Examiner H. C. Ho

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	Micro Entity Fee	
[X] One month (37 CFR 1.17(a)(1))	$200	$100	$50	$ 200.00
[ ] Two months (37 CFR 1.17(a)(2))	$600	$300	$150	$
[ ] Three months (37 CFR 1.17(a)(3))	$1,400	$700	$350	$
[ ] Four months (37 CFR 1.17(a)(4))	$2,200	$1,100	$550	$
[ ] Five months (37 CFR 1.17(a)(5))	$3,000	$1,500	$750	$

[ ] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

[ ] A check in the amount of the fee is enclosed.

[X] Payment by credit card. Form PTO-2038 is attached.

[ ] The Director has already been authorized to charge fees in this application to a Deposit Account.

[X] The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 23/2825 .

[X] Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

[ ] applicant.

[X] attorney or agent of record. Registration number 65,139 .

[ ] attorney or agent acting under 37 CFR 1.34. Registration number _____ .

/Andrew J. Tibbetts/ — Signature
December 16, 2019 — Date

Andrew J. Tibbetts — Typed or printed name
617.646.8000 — Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

[ ] * Total of 1 forms are submitted.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: December 16, 2019 Electronic Signature for: /Eileen M. MacKenzie/

7788912.1

AMENDMENTS TO THE CLAIMS

Applicant submits below a complete listing of the current claims, including marked-up claims with insertions indicated by underlining and deletions indicated by strikeouts and/or double bracketing. This listing of claims replaces all prior versions, and listings, of claims in the application:

Listing of the Claims

1. (Currently amended) A device configured to operate in a lower power mode and a higher power mode, the device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in [[a low]] the lower power mode; and

at least one processor configured to perform at least one processing stage on the acoustic input, performing the at least one processing stage including:

determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, wherein the determining comprises:

performing automatic speech recognition (ASR), while in the [[low]] lower power mode, to recognize at least one word or phrase in the acoustic input;

determining, while in the [[low]] lower power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches ~~a trigger~~ the specific word or phrase;

in response to determining that the ~~at least one first word or phrase matches the trigger~~ acoustic input includes the specific word or phrase, requesting, via a network while in the [[low]] lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command,

wherein the device is configured to respond to voice commands recognized by the at least one server.

Petitioner Exhibit 1004, Page 791 of 1052

2. (Previously presented) The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.

3. (Previously presented) The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

4. (Previously presented) The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

5. (Previously presented) The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

6. (Currently amended) The device of claim 4, wherein the first processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the ~~at least one first word or phrase matches the trigger word or phrase~~ acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.

7. (Currently amended) The device of claim 4, wherein the second processor is configured to perform the ASR ~~automatic speech recognition~~ on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

Petitioner Exhibit 1004, Page 792 of 1052

8. (Currently amended) The device of claim 7, wherein the second processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the ~~at least one first word or phrase matches the trigger word or phrase~~ acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.

9. (Previously presented) The device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. (Currently amended) The device of claim 3, wherein the device is further configured to transition to ~~an active~~ the higher power mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command.

11. (Currently amended) The device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the [[low]] lower power mode, and wherein the ~~trigger~~ specific word or phrase, when detected, is sufficient to engage the ~~device~~ speech interface of the device.

12. (Previously presented) The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. (Previously presented) The device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

Petitioner Exhibit 1004, Page 793 of 1052

14. (Previously presented) The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

15. (Currently amended) At least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:

performing at least one processing stage on the acoustic input, the at least one processing stage including:

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input;

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase;

in response to determining that the at least one first word or phrase matches the trigger word or phrase:

requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes ~~at least~~ one or more voice ~~command~~ commands; and

responding to at least one voice ~~commands~~ command recognized by the at least one server.

16. (Original) The at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

Petitioner Exhibit 1004, Page 794 of 1052

17. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the at least one first word or phrase matches the trigger word or phrase.

18. (Original) The at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. (Currently amended) The at least one non-transitory computer readable medium of claim 15, ~~further comprising performing~~ wherein responding to the at least one voice command comprises initiating performance of at least one task specified by the voice command with the device remaining in the low power mode.

20. (Previously presented) The at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

21. (Previously presented) A device comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and

at least one processor configured to:

perform automatic speech recognition (ASR), while in [[a]] the low power mode, to recognize at least one word or phrase in the acoustic input;

determine, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase;

in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a

Petitioner Exhibit 1004, Page 795 of 1052

portion of the acoustic input to determine whether the acoustic input includes ~~at least~~ one or more voice ~~command~~ commands; and

in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, ~~respond to~~ perform at least one task to carry out the at least one voice command in the acoustic input recognized by the at least one server.

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Examiner Name	H. C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)200.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☒ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☐ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee** ☐ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**	**Multiple Dependent Claims**	
				Fee ($)	**Fee Paid ($)**
20 - or HP =	_____	x _____	= _____	_____	_____

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - or HP =	_____	x _____	= _____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____	= _____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1251 Extension for response within first month	200.00

SUBMITTED BY

Signature	/Andrew J. Tibbetts/	Registration No. (Attorney/Agent)	65,139	Telephone	617.646.8000
Name (Print/Type)	Andrew J. Tibbetts			Date	December 16, 2019

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: December 16, 2019 Electronic Signature for: /Eileen M. MacKenzie/

7788910.1

Electronic Acknowledgement Receipt

EFS ID:	38045764
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Andrew J. Tibbetts/Eileen MacKenzie
Filer Authorized By:	Andrew J. Tibbetts
Attorney Docket Number:	N0484.71090US04
Receipt Date:	16-DEC-2019
Filing Date:	16-AUG-2016
Time Stamp:	17:48:32
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$200
RAM confirmation Number	E2019BFH48492978
Deposit Account	232825
Authorized User	Eileen MacKenzie

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-AJT.pdf	26068 25670a5637e46c38fe690165a62b42f8653d0ed6	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-AJT.pdf	32619 04ef1e0c2796661eddc3aa4734e16d2fdf0c3ec0	no	1
Warnings:					
Information:					
3	Extension of Time	N048471090US04-EXT-AJT.pdf	24505 765732e0f817fb977ecd0bb5a08388a166b03897	no	1
Warnings:					
Information:					
4		N048471090US04-AMN-AJT.pdf	76125 7103d78aea1805c0bd015ed96d664161491dc178	yes	15

Multipart Description/PDF files in .zip description

Document Description	Start	End
Applicant Arguments/Remarks Made in an Amendment	8	15
Claims	2	7
Amendment/Req. Reconsideration-After Non-Final Reject	1	1

Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	31189 272b4f046368f1c3fb284d069b1ceda0166c26a9	no	2

Warnings:	
Information:	
Total Files Size (in bytes):	190506

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

REMARKS

In response to the Office Action mailed August 16, 2019, the Applicant respectfully requests reconsideration. Each of the rejections set forth in the Office Action have been carefully considered and are addressed below. To further prosecution of the application, the claims are amended herein. The claims as presented are believed to be in condition for allowance.

I. Summary of Telephone Interview with the Examiner

The Assignee's representatives appreciate the courtesies extended by Examiner Ho in granting and conducting a telephone interview on December 12, 2019. The Assignee was represented at the interview by Andrew Tibbetts (Reg. No. 65,139). The substance of the telephone interview is summarized in the discussion below. In brief, the Examiner suggested he appreciated the distinction between the references and the technology described by the application, but said he needed more time to review the claim language. In view of the Examiner's comments in the interview, some amendments are made to the claims herein. If the Examiner has any questions following review of these amendments and the remarks below, he is invited to contact the undersigned for a further telephone interview.

II. Claim Rejections under 35 U.S.C. §103

Independent claims 1, 15, and 21 were rejected under pre-AIA 35 U.S.C. §103(a) as allegedly being unpatentable over U.S. Patent No. 8,977,255 ("Freeman") in view of U.S. Patent Application Publication No. 2013/0080167 ("Mozer"). The Applicant respectfully disagrees that any of the claims would have been obvious over any combination of the cited references. Even if one of ordinary skill would have been led to combine the cited references in *some* manner, no such combination would have taught or suggested all limitations of any of the pending claims.

A. Discussion of Some Embodiments From The Specification

Some embodiments described in the present application relate to a technique for a device to determine whether an acoustic input includes a voice command, while the device remains in a low power mode (Published Specification, ¶0035). The device determines whether the acoustic input

Petitioner Exhibit 1004, Page 801 of 1052

includes a trigger word or phrase (e.g., "Hello Dragon") by (1) performing automatic speech recognition (ASR) on the acoustic input to recognize spoken words or phrases in the acoustic input; and (2) determining whether any of the recognized words or phrases match the trigger word or phrase (¶0087). For example, the device may determine whether any of the phrases recognized in the acoustic input match the trigger phrase "Hello Dragon" (¶0087).

In response to determining that a trigger word phrase was spoken in the acoustic input, the device may (while remaining in the low power mode) invoke more comprehensive ASR to determine whether a voice command is present in the acoustic input (¶0090). To invoke more comprehensive ASR, the device may transmit (while in the low power mode) the acoustic input to a server may perform more computationally expensive ASR to determine whether the acoustic input includes a voice command (¶0091). When the server determines that the acoustic input includes a voice command, the server may notify the device, such that the device responds to the voice command recognized by the server (¶0070).

The foregoing discussion is provided to assist the Examiner in appreciating some applications for various inventive aspects of the present disclosure. However, this discussion may not apply to each of the independent claims, and the language of the independent claims may differ in material respects from the discussion provided above. Thus, Applicant respectfully requests that careful consideration be given to the language of each of the independent claims and that each be addressed on its own merits, without relying on the discussion provided above. In this respect, Applicant does not rely on the discussion provided above to distinguish any of the claims over the cited art. Rather, Applicant relies only upon the language of the claims themselves and the arguments related specifically thereto.

B. Discussion of Mozer

Mozer describes a "background speech recognizer" (Mozer, ¶0020) that passively listens to a conversation between two or more speakers and, after inferring a topic of the conversation, displays information that might be relevant to the speakers (¶0066). Mozer thus describes that a device "may classify keywords that are recognized into pre-defined classifications" that are "topics that describe different areas of interest, such as travel, purchases, entertainment, research, food, or

Petitioner Exhibit 1004, Page 802 of 1052

electronics" (Mozer, ¶0033). In Mozer's system, once "a certain number of keywords for a specific classification are detected, then it may be determined that a topic associated with the classification is being discussed" and "***when one classification has enough key words classified with it, then stage 1 recognizer 102 may trigger stage 2 recognizer 104***" (¶0033, emphasis added). As an example, the device selects a classification of "travel" based on the words "vacation" and "flight" in an acoustic input and determines "***that the travel classification should be selected and the stage 2 recognizer 104 should be activated***" (¶0066, emphasis supplied). This "stage 2 recognizer" is a server-based recognizer, which analyzes the conversation based on the identified topic and determines relevant information to display (¶0067).

In particular, Mozer describes that the server "receives the trigger to activate and also may receive information that a conversation is occurring about the classification of 'travel'" (¶0067). "At this point, stage 2 recognizer 104 may take over listening to the conversation" (¶0067). The "stage 2 recognizer 104 may be able to decipher whole sentences and may hear a sentence 'Maybe we should do an activity in Ireland'" and "the classification of 'travel' may be used to determine content for the response" (¶0067). "At this point, a response may be determined that pictures of Ireland should be output with a coupon for a soccer game in Ireland" and "pictures of Ireland may be output to an interface, such as the tablet computer screen" (¶0067).

C. Independent Claim 1 Patentably Distinguishes Over Freeman and Mozer.

Independent claim 1 has been amended herein and recites "determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input."

In connection with rejecting previously-pending claim 1, the Office Action cited to Freeman's FIG. 4 and elements 402-408 that the Office Action characterized as relating to "determining whether the audio input pertains to at least one of the commands in a command set" (Office Action, pages 7-8). Such commands relate to performance of tasks at the request of a user (e.g., Freeman, 4:33-39). Freeman explains that in its system, it can receive and process such voice commands directly, "without any preparatory user action with respect to the portable electronic device" (Freeman, Abstract, 1:46-48, 3:56-58).

Petitioner Exhibit 1004, Page 803 of 1052

Claim 1 recites a determination of "whether the acoustic input includes at least one voice command" but, in contrast to Freeman, prior to that recites "determining … whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input." Freeman does not teach or suggest evaluating whether acoustic input includes a "voice command" **and**, separate from the voice command(s), such a "specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input." In fact, Freeman does not teach or suggest determining whether acoustic input includes any such a "specific word or phrase." In Freeman's system, the voice commands are input directly "without any preparatory user action."

Like Freeman, the passages of Mozer cited by the Office Action also describe that its system "does not need to be triggered (e.g., by a button press or trigger phrase) to begin speech recognition," but instead conducts recognition and classification on all speech content (Mozer, ¶¶0032-0033). As Mozer describes functionality that "does not need to be triggered" and thus does not determine whether a trigger phrase is present, the cited passages of Mozer do not teach or suggest "determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input," as recited by claim 1.

As neither Freeman nor Mozer would have taught or suggested such a determination of whether acoustic input includes such a "specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input," no combination of the references would have taught or suggested the quoted language of claim 1.

In addition, for at least the same reason, no combination of the references would have taught or suggest *responding* to "determining that the acoustic input includes the specific word or phrase" by "requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command," as recited by claim 1.

For at least these reasons, claim 1 patentably distinguishes over any combination of Freeman and Mozer. Claims 2-14 depend from claim 1 and patentably distinguish for at least the same reasons. Withdrawal of the rejections of claim 1-14 under §103 is respectfully requested.

Petitioner Exhibit 1004, Page 804 of 1052

D. Independent Claim 15 Patentably Distinguishes Over Freeman and Mozer.

For reasons that should be appreciated from the foregoing discussion, no combination of Freeman and Mozer would have taught or suggested "determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase" and "in response to determining that the at least one first word or phrase matches the trigger word or phrase" "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands."

For at least these reasons, claim 15 patentably distinguishes over any combination of Freeman and Mozer. Claims 16-20 depend from claim 15 and patentably distinguish for at least the same reasons. Withdrawal of the rejections of claim 15-20 under §103 is respectfully requested.

E. Independent Claim 21 Patentably Distinguishes Over Freeman and Mozer.

For at least the two reasons discussed below, no combination of Freeman and Mozer would have taught or suggested all limitations of claim 21.

i. Freeman and Mozer do not describe "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition."

The Office Action concedes that Freeman would not teach or suggest "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition" as recited by claim 21 (Office Action, p. 14).

The Office Action instead alleges that figures 4 and 6, ¶¶0031-0039, ¶¶0041-0044, and ¶¶0055-0056 of Mozer would have disclosed the "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition" recited by claim 1 (Office Action, p. 14). As discussed during the interview and highlighted in the discussion of the reference above, the cited passage explicitly describes that the device ("stage 1 recognizer") activates the server ("stage 2 recognizer") when the device determines that a conversation relates to a particular topic/classification (e.g., travel, food, electronics) (¶0033,

Petitioner Exhibit 1004, Page 805 of 1052

¶0066). In terms of activating the server, the device does not take any action when any particular word or phrase is recognized. For example, Mozer describes that the device may recognize the words "flight" and "vacation," but Mozer's device does not take any action based only on those words. Mozer's device does not (1) match those words to a trigger word or phrase, or (2) activate the server in response to detecting that the acoustic input includes those words (¶0066). In fact, Mozer's device does not take ***any*** particular action in response to recognition of any particular word/phrase, or in response to determining that any particular word/phrase "matches" a "trigger word or phrase." Rather, Mozer's device activates the server only when a classification of a conversation is determined (e.g., "travel") (¶0066). Mozer's server uses the classification identified by the device for determining content to provide to the device, and thus relies on the classification to begin processing. In Mozer, the device does not activate the server "***in response to determining that [] at least one word or phrase matches [a] trigger word or phrase***" (emphasis supplied), as recited in claim 21.

Accordingly, even if the references would have been combined in the manner alleged by the Office Action (which Applicant does not concede), the references would not have taught or suggested the above-quoted language of amended claim 21.

ii. Freeman and Mozer fail to describe "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server."

The server of Mozer does not determine whether an acoustic input includes any voice command. As can be appreciated from the discussion of the reference above, Mozer's server analyzes sentences of a conversation and outputs content (e.g., pictures and coupons) based on the sentences and a classification received from the device (¶0067). The server does not determine whether an acoustic input includes a voice command. As the server does not determine whether acoustic input includes a voice command, the device of Mozer does not receive any indication from the server that the acoustic input includes a voice command. It follows that Mozer does not teach or suggest "***in response to receiving an indication from the at least one server that the acoustic input***

Petitioner Exhibit 1004, Page 806 of 1052

includes at least one voice command, respond to the at least one voice command" (emphasis added) as recited in claim 21.

Freeman describes voice activation of a single device without interaction with a separate server. Thus Freeman does not describe that a device responds to a voice command "in response to receiving an indication from the at least one server," as recited in claim 21.

Accordingly, even if the references would have been combined in the manner alleged by the Office Action (which Applicant does not concede), the references would not have taught or suggested the above-quoted language of claim 21.

iii. Conclusion

For at least these reasons, claim 21 patentably distinguishes over any combination of Mozer and Freeman. Accordingly, withdrawal of the rejection under §103 of claim 21 is respectfully requested.

III. General Comments on Dependent Claims

As each of the dependent claims depends from a base claim that is believed to be in condition for allowance, Applicant believes that arguing the further distinguishing features of all of the dependent claims is unnecessary at this time. However, Applicant does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does Applicant concur that the basis for the rejection of any of the dependent claims is proper. Therefore, Applicant reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

Petitioner Exhibit 1004, Page 807 of 1052

CONCLUSION

In view of the foregoing, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after this response, the application is not in condition for allowance, the Examiner is requested to call the undersigned at the telephone number indicated below to discuss any outstanding issues.

Applicant believes no fee is due with this Amendment. However, if a fee is due with this Amendment, please charge our Deposit Account No. 23/2825, under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Dated: December 16, 2019

Respectfully submitted,

Electronic signature: /Andrew J. Tibbetts/
Andrew J. Tibbetts
Registration No.: 65,139
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [x] Fee Attached
- [x] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [x] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [ ] Other Enclosure(s) (please Identify below):

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Andrew J. Tibbetts/		
Printed name	Andrew J. Tibbetts		
Date	December 16, 2019	Reg. No.	65,139

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: December 16, 2019 Electronic Signature for: /Eileen M. MacKenzie/

7656568.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 12/18/2019

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
12/18/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

IP.Inbox@nuance.com
N0484_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@wolfgreenfield.com

Applicant-Initiated Interview Summary	**Application No.** 15/238,238	**Applicant(s)** Ganong et al.	
	Examiner HUY C HO	**Art Unit** 2644	**AIA (FITF) Status** No

All participants (applicant, applicants representative, PTO personnel):

(1) HUY C. HO. (3) _____.

(2) ANDREW J. TIBBETTS (Reg. No. 65,139). (4) _____.

Date of Interview: 12 December 2019.

Type: ☑ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☑ No.
If Yes, brief description: _____.

Issues Discussed ☐101 ☐112 ☐102 ☑103 ☐Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: 21.

Identification of prior art discussed: Mozer and Freeman.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Discussed the argued features/limitations , i.e., in response to determining the first word/phrase matches the trigger word/phrase, request at least one server perform recognition on a portion of the acoustic input ... ; and in response to receive an indication from the server ...., respond to at least one voice command ..." over the prior art references above; However, no agreement was reached at this time..

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HUY C HO/ Primary Examiner, Art Unit 2644	

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record
A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) 1.133 Interviews
Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.
All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiners responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicants correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:
- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:
1) A brief description of the nature of any exhibit shown or any demonstration conducted,-
2) an identification of the claims discussed,
3) an identification of the specific prior art discussed,
4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
5) a brief identification of the general thrust of the principal arguments presented to the examiner,
(The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
6) a general indication of any other pertinent matters discussed, and
7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicants record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiners version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, Interview Record OK on the paper recording the substance of the interview along with the date and the examiners initials.

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 2	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		6,006,175	A	Holzrichter	12-21-1999
		6,859,776	B1	Cohen et al.	02-22-2005
		9,087,520	B1	Salvador	07-21-2015
		9,558,749	B1	Secker-Walker et al.	01-31-2017
		9,747,899	B2	Pogue et al.	08-29-2017
		10,332,525	B2	Secker-Walker et al.	06-25-2019
		2012-0281885	A1	Syrdal et al.	11-08-2012
		2014-0039888	A1	Taubman et al.	02-06-2014
		2014-0163978	A1	Basye et al.	06-12-2014
		2015-0006176	A1	Pogue et al.	01-01-2015
		2015-0053779	A1	Adamek et al.	02-26-2015
		2016-0078869	A1	Syrdal et al.	03-17-2016
		2019-0073999	A1	Premont et al.	03-07-2019
		2019-0287526	A1	Ren et al.	09-19-2019
		2019-0311715	A1	Pfeffinger et al.	10-10-2019
		2019-0355365	A1	Kim et al.	11-21-2019

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		International Preliminary Report on Patentability for International Application No. PCT/CN2016/105343 mailed May 23, 2019 (N0484.71177WO00).	

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7977632.1

FORM PTO-1449/A and B (modified PTO/SB/08) **INFORMATION DISCLOSURE STATEMENT BY APPLICANT**				APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
				FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
				FIRST NAMED INVENTOR: William F. Ganong	
Sheet	2	of	2	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7977632.1

PATENT COOPERATION TREATY

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING
TRANSMITTAL OF COPY OF INTERNATIONAL
PRELIMINARY REPORT ON PATENTABILITY
(CHAPTER I OF THE PATENT COOPERATION
TREATY)
(PCT Rule 44bis.1(c))

To:

CCPIT PATENT AND TRADEMARK LAW OFFICE
8th Floor, Vantone New World Plaza
2 Fuchengmenwai Street, Xicheng District
Beijing 100037
CHINE

Date of mailing *(day/month/year)* 23 May 2019 (23.05.2019)	
Applicant's or agent's file reference IEE160077PCT	**IMPORTANT NOTICE**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/CN2016/105343	10 November 2016 (10.11.2016)	

Applicant
NUANCE COMMUNICATIONS, INC.

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Xiaofan Tang

e-mail: pct.team2@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference IEE160077PCT	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/CN2016/105343	International filing date *(day/month/year)* 10 November 2016 (10.11.2016)	Priority date *(day/month/year)*
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant NUANCE COMMUNICATIONS, INC.		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☐	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 14 May 2019 (14.05.2019)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Xiaofan Tang e-mail: pct.team2@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
100037
China 8th Floor, Vantone New World Plaza, 2 Fuchengmenwai Street, Xicheng District, Beijing

CCPIT PATENT AND TRADEMARK LAW OFFICE

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	**04 August 2017**
Applicant's or agent's file reference **IEE160077PCT**	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/CN2016/105343	**10 November 2016**	

International Patent Classification (IPC) or both national classification and IPC
G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G06F 1/32(2006.01)i

Applicant
NUANCE SOFTWARE TECHNOLOGY (BEIJING) CO., LTD.

1. This opinion contains indications relating to the following items:
   - [x] Box No. I Basis of the opinion
   - [ ] Box No. II Priority
   - [ ] Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - [ ] Box No. IV Lack of unity of invention
   - [x] Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
   - [ ] Box No. VI Certain documents cited
   - [ ] Box No. VII Certain defects in the international application
   - [ ] Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

   If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

   If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

   For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/	Date of completion of this opinion	Authorized officer
STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA **China 6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088**	**31 July 2017**	**KANG,Dandan**
Facsimile No. **(86—10) 62019451**		Telephone No. **(86-10)61648170**

Form PCT/ISA/237 (Cover sheet) (July 2011)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. [ ] This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing filed or furnished:
   a. (means)
      - [ ] on paper
      - [ ] in electronic form
   b. (time)
      - [ ] in the international application as filed
      - [ ] together with the international application in electronic form
      - [ ] subsequently to this Authority for the purposes of search

4. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Petitioner Exhibit 1004, Page 818 of 1052

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-20	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-20	NO
Industrial applicability (IA)	Claims	1-20	YES
	Claims	None	NO

2. Citations and explanations :

[1] D1: US 2016189706 A1

[2] D2: CN 101650943 A

[3] D1 discloses a wake up phrase detection method and an electronic device(see description paragraphs [0030]-[0032], [0049]-[0070],[0086]-[0092], figures 2-5,8): electronic device may comprise a smart phone; FIG.4 depicts a training system 400, a user is prompted to utter their custom wake-up phrase(WUP), the utterances are captured as user training sequences 402(corresponding to "receiving acoustic input of a user speaking a wake-up word"); user training sequences 402 is provided to VAD 404, then to feature extraction component 406 to derive the feature vectors 408(corresponding to "acoustic features derived from the acoustic input"); the extracted features 408 are decoded by concatenated phoneme ASR engine 410 using phoneme model DB 414 which is implemented in a memory or storage medium and store phoneme models 206, and produce a phoneme transcription of the WUP, a Hidden Markov Model(HMM) is trained offline to generate a phoneme model 206(corresponding to "acoustic model"); based on the results produced by ASR 410, phoneme model concatenation component 416 determines the best phonemes models and concatenates them to produce an initial WUP model consist of the concatenated phoneme models(corresponding to "a sequence of speech units corresponding to the wake-up word spoken by the user" and "a first sequence of speech units" in claim 9 or "a second sequence of speech units" in claim 16); the WUP model used for WUP detector is stored in WUP model DB 418 which is implemented in a memory or storage medium ; FIG.5 depicts a user-specific wake up phrase detection system 500, a low-power processor is included and operates in a stand-by mode of the electronic device, receiving a wake-up phrase 514 as an input by a user(corresponding to "receiving acoustic input from the user speaking while the user device is operating in a low-power mode"), WUP 514 is provided to VAD 504 and then to feature extractor and WUP detector(WUPD) component 506 to detect the presence of the WUP in the captured audio by comparing the feature extracted to the WUP model stored in a non-volatile RAM; the presence of the WUP triggers the device to wake up, or come out of a stand-by mode or a sleep mode into a normal mode of operation(corresponding to "exiting the low-power mode if it is determined that the user spoke the wake-up word").

[4] D2 discloses a non-native speech recognition system and method (see description page 5 line 3 – page 6 line 9 , figures 1-3): a native language(corresponding to "the source language") model is provided for the native state decoding module, the non-native language(corresponding to "the target language") speech is decoded by the native state decoding module using the native language model, thus obtaining a native state decode sequence "b aa zh " (see figure 3).

[5] **Novelty and Inventive step**

[6] Claim 1 differs from D1 in: (1) the acoustic input of a user is spoken in a target language, and the acoustic model is trained in a source language different than the target language.

[7] Claim 5 differs from D1 in the above (1), and (2) at least one microphone configured to obtain acoustic information from the environment of the user device.

[8] Claim 9 differs from D1 in the above (1), (2) and (3) applying acoustic features derived from the acoustic input to the acoustic model to obtain a second sequence of speech units; determining if the user spoke the wake-up word at least in part by comparing the first sequence of speech units to the second sequence of speech units.

Petitioner Exhibit 1004, Page 819 of 1052

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

[9] Claim 16 differs from D1 in the above (1) and (3).

[10] Therefore, claims 1-20 are novel (PCT Article 33(2)).

[11] The feature (1) has already been disclosed by D2. The person skilled in the art would easily acquire the technical feature of D2 to decode the acoustic input.

[12] The feature (2) of obtaining acoustic information by microphone is general solution in the art.

[13] As to the feature (3), it is common practice for the person skilled in the art to determine the wake-up word by comparing the first sequence of speech units to the second sequence of speech units which is obtained by applying acoustic features derived from the acoustic input to the acoustic model instead of the WUP model.

[14] Therefore, claims 1,5,9,16 do not involve an inventive step (PCT Article 33 (3)).

[15] Dependent claims 2-4,6-8,10-15,17-20 do not involve an inventive step (PCT Article 33 (3)).

[16] As to claims 2,6,10,17,4,8,13,18, it is common practice to adapt the acoustic model using the acoustic input and the sequence, and associate at least one task with the wake-up word and initiate performance of the task based on the wake-up detection.

[17] As to claims 3,7,11,12,14,15,19,20, those additional features have been disclosed by D1.

[18] **Industrial applicability**

[19] Claims 1-20 can find industrial applicability in the technical field of wake-up detecting (PCT Article 33 (4)).

Petitioner Exhibit 1004, Page 820 of 1052

DOCKET NO.: N0484.71090US04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: _____March 6, 2020_____ Signature: _____/Miranda W. Stopa/_____ (Miranda W. Stopa)

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed more than three months after the filing date of this application and after the mailing date of a first Office Action, but before the mailing date of any of a Final Action under 37 C.F.R. § 1.113, a Notice of Allowance under 37 C.F.R. § 1.311, or an action that otherwise closes prosecution in this application.

Please charge our Credit Card in the amount of $240.00 covering the fee set forth in 37 C.F.R. § 1.17(p).

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
PCT/CN2016/105343	**05-23-2019**	**International Preliminary Report on Patentability**	**N0484.71177WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Andrew J. Tibbetts/

Andrew J. Tibbetts
Registration No.: 65,139
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: N0484.71090US04
Date: March 6, 2020

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Andrew J. Tibbetts/Miranda Stopa
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	1806	1	240	240
		Total in USD ($)		**240**

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):
  - Form PTO - 1449
  - Copy of Cited Reference

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Andrew J. Tibbetts/		
Printed name	Andrew J. Tibbetts		
Date	March 6, 2020	Reg. No.	65,139

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: _____March 6, 2020_____ Signature: _____/Miranda W. Stopa/_____ (Miranda W. Stopa)

7977631.1

Electronic Acknowledgement Receipt

EFS ID:	38788302
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Andrew J. Tibbetts/Miranda Stopa
Filer Authorized By:	Andrew J. Tibbetts
Attorney Docket Number:	N0484.71090US04
Receipt Date:	06-MAR-2020
Filing Date:	16-AUG-2016
Time Stamp:	10:42:35
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$240
RAM confirmation Number	E202036A42532423
Deposit Account	232825
Authorized User	Miranda Stopa

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-AJT.pdf	26192 c8edaa13325f0d99997f01cb386e6e91c272450b	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US04-FEE-AJT.pdf	32694 8a35b36e3dc389cb235c3e66a4e4721fa3db6b6e	no	1
Warnings:					
Information:					
3	Transmittal Letter	N048471090US04-IDSBODY-AJT.pdf	31360 354473ea9e2b7aa4dffeaba6a23c272e45c9c21e	no	3
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	N048471090US04-IDSSB08-AJT.pdf	25834 4dca2dc24f7df952a4f96a96a56c7913d3389b7d	no	2
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
5	Other Reference-Patent/App/Search documents	N048471177WO00_IPRP_Mailed_05-23-2019.pdf	229934 272c61976195a4f379de3b487dd23070ac8ab8f8	no	6
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	30877 260e290a0d81bbecae476a2a0c70b4871d3583b0	no	2
Warnings:					

Information:	
Total Files Size (in bytes):	376891

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong
Examiner Name	H. C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)240.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☒ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☐ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee** ☐ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	**Extra Sheets**	**Number of each additional 50 or fraction thereof**	**Fee ($)**	**Fee Paid ($)**
_____ - 100 =	_____ /50 =	_____ (round **up** to a whole number) x	_____ =	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1806 Submission of an Information Disclosure Statement	240.00

SUBMITTED BY

Signature	/Andrew J. Tibbetts/	Registration No. (Attorney/Agent)	65,139	Telephone	617.646.8000
Name (Print/Type)	Andrew J. Tibbetts			Date	March 6, 2020

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: March 6, 2020 Signature: /Miranda W. Stopa/ (Miranda W. Stopa)

7977629.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 03/13/2020

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

NOTIFICATION DATE	DELIVERY MODE
03/13/2020	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

IP.Inbox@nuance.com
N0484_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@wolfgreenfield.com

Office Action Summary	Application No. 15/238,238	Applicant(s) Ganong et al.	
	Examiner HUY C HO	Art Unit 2644	AIA (FITF) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 16 December 2019.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☑ This action is **FINAL.** 2b) ☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-21 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1-21 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 16 August 2016 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Response to Arguments

2. Applicant's arguments filed **12/16/2019** have been fully considered but they are moot and not persuasive for the following reasons.

a/ moot because claims 1, 6, 8, 10, 11, 15, 19 and 21 have been amended;

b/ not persuasive because:

A. With respect to the arguments that the independent claim 1 (Currently Amended) patentably distinguishes over Freeman and Mozer.

The examiner respectfully disagrees with the arguments that the combination of Freeman and Mozer has/have not taught or suggested the argued amended limitations, i.e., "determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input", because Freeman teaches a portable electronic device (Freeman, Figure 1) is operating in a low power mode (Figure 5, box 504) to monitor audio inputs from a surrounding

environment (Freeman, Figure 4, boxes 402, 404) in order to determine if the audio inputs include commands (Freeman, Figure 4, boxes 406, 408). The portable electronic device has a data store 220 (Figure 2, box 220) to store commands and macros (macros are considered groups or sets of commands) and when the portable electronic device uses a voice-to-command analyzer 213 to monitor and to determine if there is/are command(s) included in the audio input(s) by comparing the commands to the stored commands and macros from the data store 220, the portable electronic device executes further or subsequent steps (Freeman, Figure 4, boxes 408, 410, Col. 6 Lines 10-47; Figure 5, boxes 506, 508, 510 and 512, Col. 6 Lines 48-67, Col. 7 Lines 1-32).

Therefore, Freeman teaches and suggests the argued limitations, i.e., “determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input”.

Mozer teaches a device 100 comprises a speech recognizer 101 (Mozer, Figure 1A), wherein the speech recognizer 101 includes multiple stages recognizers, e.g., stage 1 and stage 2 (Mozer, Figure 1B) configured to listen for acoustic input signals. Stage 1 recognizer is a speech recognition that is at a low power mode and always-on for recognizing keywords that are picked up from the acoustic input signals in the surrounding and

providing subsequent steps (Mozer, Figures 4, 6, pp [32]-[33], [66]).

Therefore, Mozer teaches and suggests the argued limitations, i.e., "determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input".

As such, Freeman, in combination of Mozer, teach and suggest the argued features/limitations above, i.e., "determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input".

B. With respect to arguments to independent claim 15:

With respect to the arguments that no combination of Freeman and Mozer would have taught or suggested "determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase" and "in response to determining that the at least one first word or phrase matches the trigger word or phrase" "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands."

The examiner respectfully disagrees because Freeman teaches: “determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase” (Freeman, Figure 5, boxes 504, 506, 508, 510, Col. 6 Lines 50-67, Col. 7 Lines 1-10; Figure 2, voice to command analyzer 216 receives audio inputs to compare if matches with stored commands 220; Figure 6, boxes 602-608). Thus Freeman discloses and teaches the limitation, i.e., determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase.

Mozer teaches a device comprises a speech recognizer with multiple stage recognizers, e.g., stage 1 (low power stage) and stage 2 (higher power stage), that receives acoustic input signals where the stage 1 recognizer at the low power level makes a determination of the acoustic input signals comprising any keywords compared to the stored keywords from the data store 206 (Mozer, Figure 2, stage 1, pp [34]-[35]). When stage 1 recognizer identifies a keyword match (Figure 4), it activates stage 2 recognizer to use a more accurate speech recognition algorithm to recognize and analyze the acoustic input signals, wherein the stage 2 recognizer can be a server (Figure 1B, pp [31]). Therefore, Mozer teaches and discloses the argued

limitations, i.e., "determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase" and "in response to determining that the at least one first word or phrase matches the trigger word or phrase" "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands."

As such, the combination of Freeman and Mozer teach and suggest the argued limitations above.

B. With respect to arguments to independent claim 21:

i. Freeman and Mozer do not describe "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition."

The examiner respectfully disagrees because Mozer teaches a device comprises a speech recognizer with multiple stage recognizers, e.g., stage 1 (low power stage) and stage 2 (higher power stage), that receives acoustic input signals where the stage 1 recognizer at the low power level makes a determination of the acoustic input signals comprising any keywords compared to the stored keywords from the data store 206 (Mozer, Figure 2, stage 1, pp [34]-[35]). When stage 1 recognizer identifies a

keyword match (Figure 4), it activates stage 2 recognizer to use a more accurate speech recognition algorithm to recognize and analyze the acoustic input signals, wherein the stage 2 recognizer can be a server (Figure 1B, pp [31]). Therefore, Mozer teaches and discloses the argued limitations, i.e., "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition."

ii. Freeman and Mozer fail to describe "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server."

The examiner respectfully disagrees because Mozer teaches after stage 1 recognizer recognizes a keyword and triggers stage 2 recognizer (also acts as a server) for recognizing further specific information then stage 2 recognizer outputs a response based on the ranked specific information (Mozer, Figure 1B, pp [56]: stage 2 can be a server; Figure 4, pp [54]; Figures 5, 6, pp [56]-[59]). Thus Mozer teaches and suggests "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, respond to at least one voice command in the acoustic input recognized by the at least one server".

As such, the argued features/limitations were written such that they read upon Freeman and Mozer as discussed above.

Claim Rejections - 35 USC § 103

3. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

4. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

6. Claims 1-12, 15-19 and 21 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255) and further in view of Mozer (U.S. Patent Publication No. 20103/0080167).

Regarding claim 1, (Currently Amended) Freeman teaches a device configured to operate in a lower power mode and a higher power mode, the device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in a lower power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to perform at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512);*

determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, wherein the determining comprises *(Freeman, Figure 4, boxes 408, 410, Col. 6 Lines 10-47; Figure 5, boxes 506, 508, 510 and 512, Col. 6 Lines 48-67, Col. 7 Lines 1-32).*

performing automatic speech recognition (ASR), while in the lower power mode, to recognize at least one word or phrase in the acoustic input (*Freeman, Figure 2, voice-to-command analyzer 216, Figure 3, Col. 5 Lines 13-67, Col. 6 Lines 1-10; Figure 4, Boxes 402, 404, 406,408; Figure 5, Boxes 504, 506, 508, 510: a portable electronic device in a low power mode performing audio pickup to recognize a command from the audio/voice input*);

determining, while in the lower power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches the specific word or phrase (*Freeman, Figure 4, Boxes 402, 404, 406, 408: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set);*

in response to determining, that the acoustic input includes the specific word or phrase (*Freeman, Figure 4, Boxes 402, 404, 406, 408, Col. 6 Lines 10-47: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set; Figure 5, boxes 506, 508, 510 and 512, Col. 6 Lines 48-67, Col. 7 Lines 1-32).*

Freeman does not teach “requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to

determine whether the acoustic input includes at least one voice command, wherein the device is configured to respond to voice commands recognized by the at least one server".

Mozer teaches "requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command" (*Mozer, Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]: a speech recognizer device receives acoustic input signal to process the acoustic input signal at stage 1 (lower power mode) listening for trigger words or phrases such as travel, purchases, entertainment, research, food, or electronics, stage 2 recognizer uses more accurate speech recognition algorithm to perform the speech recognition; Stage 2 recognizer (over a network device or a server) may use large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning)*; and "the device is configured to respond to voice commands recognized by the at least one server" (*Mozer, pp [36]: after stage 2 recognizer perform the large-vocabulary continuous speech recognition*

(LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning, stage 1 or stage 2 recognizers may decide to save information from previous discussions to better classify, solve problems and assist).

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 15, (Currently Amended) Freeman teaches at least one non-transitory computer readable medium storing instructions that, when executed on at least one processor, perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode (*Freeman, the Abstract, Figure 1, Col. 1 Lines 60-65*), the method comprising:

performing, at least one processing stage on the acoustic input (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512);*

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input (*Freeman, Figure 2, voice-to-command analyzer 216, Figure 3, Col. 5 Lines 13-67, Col. 6 Lines 1-10; Figure 4, Boxes 402, 404, 406,408; Figure 5, Boxes 504, 506, 508, 510: a portable electronic device in a low power mode performing audio pickup to recognize a command from the audio/voice input*);

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase (*Freeman, Figure 4, Boxes 402, 404, 406, 408: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set).*

Freeman does not teach "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice command" and "responding to at least one voice command recognized by the at least one server".

Mozer teaches "requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice command" (*Mozer, Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]: a speech recognizer device receives acoustic input signal to process the acoustic input signal at stage 1 (lower power mode) listening for trigger words or phrases such as travel, purchases, entertainment, research, food, or electronics, stage 2 recognizer uses more accurate speech recognition algorithm to perform the speech recognition; Stage 2 recognizer (over a network device or a server) may use large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning)*; and "responding to at least one voice command recognized by the at least one server" (*Mozer, pp [36]: after stage 2 recognizer perform the large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning, stage 1 or stage 2 recognizers may decide to*

save information from previous discussions to better classify, solve problems and assist).

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 21, (Currently Amended) Freeman teaches a device (*Freeman, the Abstract, Figure 1*), comprising:

at least one input configured to receive acoustic input from an environment of the device while the device is operating in the low power mode (*Freeman, Figure 3, voice input, Figure 5, box 504)*; and

at least one processor configured to (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512)*:

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the

acoustic input (*Freeman, Figure 2, voice-to-command analyzer 216, Figure 3, Col. 5 Lines 13-67, Col. 6 Lines 1-10; Figure 4, Boxes 402, 404, 406,408; Figure 5, Boxes 504, 506, 508, 510: a portable electronic device in a low power mode performing audio pickup to recognize a command from the audio/voice input*);

determining, while in the low power mode, whether at least one first word or phrase of the recognized at least one word or phrase matches a trigger word or phrase (*Freeman, Figure 4, Boxes 402, 404, 406, 408: receiving audio input, determining whether the audio input pertains to at least one of the commands in a command set).*

Freeman does not teach "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one or more voice commands" and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, perform at least one task to carry out the at least one voice command in the acoustic input recognized by the at least one server".

Mozer teaches "in response to determining that the at least one first word or phrase matches the trigger word or phrase, request that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one or more voice commands" (*Mozer, Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]: a speech recognizer device receives acoustic input signal to process the acoustic input signal at stage 1 (lower power mode) listening for trigger words or phrases such as travel, purchases, entertainment, research, food, or electronics, stage 2 recognizer uses more accurate speech recognition algorithm to perform the speech recognition; Stage 2 recognizer (over a network device or a server) may use large-vocabulary continuous speech recognition (LVCSR) techniques to describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning)*; and "in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, perform at least one task to carry out the at least one voice command in the acoustic input recognized by the at least one server" (*Mozer, pp [36]: after stage 2 recognizer perform the large-vocabulary continuous speech recognition (LVCSR) techniques to*

describe the language of a specific topic (language model) and converts the acoustic input signal into a probable word trellis that is then accurately parsed using a statistical parser to extract meaning, stage 1 or stage 2 recognizers may decide to save information from previous discussions to better classify, solve problems and assist).

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman by incorporating teachings of Mozer, system and method for speech recognition wherein the speech queries and commands are recognized from the acoustic input signals by multi-staged process and one of the multi-staged process could be done via a different remote device such as a server, thus providing improvement accuracy for the speech recognizing process of acoustic input signals.

Regarding claim 2, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]).*

Regarding claim 3, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Freeman, Figure 4, box 408; Figure 5, boxes 504, 508, 510, 512; and Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 4, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor (*Mozer, Figure 1B, Paragraph(s) [31]-[32], [36]*).

Regarding claim 5, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech (*Freeman, Col. 4 Lines 45-62, Col. 5 Lines 13-33*).

Regarding claim 6, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the first processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input determining whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase (*Mozer, Paragraph(s) [33]; Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]*).

Regarding claim 7, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 4, wherein the second processor is configured to perform the ASR on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command (*Mozer, Paragraph(s) [36]*).

Regarding claim 8, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the second processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determining whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase (*Mozer, Paragraph(s) [33];*

Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]).

Regarding claim 9, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Mozer, Paragraph(s) [26]-[28], [46]-[47])*.

Regarding claim 10, (Currently Amended) Freeman, as modified by Mozer, teaches the mobile device of claim 3, wherein the device is further configured to transition to the higher power mode to respond to the voice command when either the at least one first processing stage or the at least one second processing stage determines that the acoustic input includes a voice command (*Freeman, Figure 5, boxes 512)*.

Regarding claim 11, (Currently Amended) Freeman, as modified by Mozer, teaches the device of claim 1, wherein at least one task specified by the voice command is performed with the device remaining in the lower power mode, and wherein the specific word or phrase, when detected, is sufficient to engage the speech interface of the device (*Mozer, Paragraph(s) [65]-*

[66]: specific words such as "vacation", "travel", etc.; Mozer, Paragraph(s) [33]; Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]).

Regarding claim 12, (Previously Presented) Freeman, as modified by Mozer, teaches the device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input (*Freeman, Figure 5, box 506).*

Regarding claim 16, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 16, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage (*Mozer, Figure 1B, Paragraph(s) [31]-[33], [36]).*

Regarding claim 17, (Currently Amended) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising performing limited vocabulary automatic speech recognition on the acoustic input to determining whether the at least one first word or phrase matches the trigger word or phrase (*Mozer,*

Paragraph(s) [33]; Figure 1B, pp [31]-[39]; Figure 2, pp [41]-[44]; Figure 3; Figures 4, 5, 6 and 8, pp [55]-[56]).

Regarding claim 18, (Original) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, further comprising providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed (*Freeman, Figure 5, boxes 504, 506, 508).*

Regarding claim 19, (Currently Amended) Freeman, as modified by Mozer, teaches the at least one non-transitory computer readable medium of claim 15, wherein responding to the at least one voice command comprising at least one task specified by the voice command with the device remaining in the low power mode (*Freeman, Figure 5, boxes 506, 508, 510).*

8. Claims 13, 14 and 20 are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Freeman et al. (U.S. Patent No. 8,977,255), further in views of Mozer (U.S. Patent Publication No. 20103/0080167) and LeBeau et al. (U.S. Patent Publication No. 2012/0034904).

Regarding claim 13, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the mobile device, an occurrence of an event on the device, recent activity on the mobile device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the mobile device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43])*.

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient

sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 14, (Previously Presented) Freeman, as modified by Mozer, does not teach the device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches "performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to

modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Regarding claim 20, (Previously Presented) Freeman, as modified by Mozer, does not teach the at least one non-transitory computer readable medium of claim 15, further comprising evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

LeBeau teaches “performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on

the device, recent activity on the device, and state of a battery powering the device." (*LeBeau, Figure 1B, 140, Paragraph(s) [41]-[43]).*

Therefore, it would have been obvious to a person of ordinary skill in the art before the invention was made to modify Freeman and Mozer, by incorporating teachings of LeBeau, system and method for monitoring ambient sounds for voice input of mobile device with further advantage of consideration of the context associated with the mobile device such as current location of the device, date and time, schedule, power source, etc., thus providing an improvement in recognizing of ambient sounds for voice input around the device with automatic and intelligent determination the device context associated with the device rather waiting for explicit direction from the user of the device for beginning monitoring for the voice input from the ambient sounds.

Conclusion

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL.** See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUY C HO whose telephone number is (571)270-1108. The examiner can normally be reached on M-F 8AM-5PM.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner’s supervisor, KATHY WANG-HURST can be reached on (571)270-5371. The fax phone number for the

organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HUY C HO/
Primary Examiner, Art Unit 2644

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	04/20/2017	01/10/2018	08/20/2018	02/24/2019	08/13/2019	02/18/2020	03/04/2020		
	1	✓	✓	✓	✓	✓	✓	✓		
	2	✓	✓	✓	✓	✓	✓	✓		
	3	✓	✓	✓	✓	✓	✓	✓		
	4	✓	✓	✓	✓	✓	✓	✓		
	5	✓	✓	✓	✓	✓	✓	✓		
	6	✓	✓	✓	✓	✓	✓	✓		
	7	✓	✓	✓	✓	✓	✓	✓		
	8	✓	✓	✓	✓	✓	✓	✓		
	9	✓	✓	✓	✓	✓	✓	✓		
	10	✓	✓	✓	✓	✓	✓	✓		
	11	✓	✓	✓	✓	✓	✓	✓		
	12	✓	✓	✓	✓	✓	✓	✓		
	13	✓	✓	✓	✓	✓	✓	✓		
	14	✓	✓	✓	✓	✓	✓	✓		
	15	✓	✓	✓	✓	✓	✓	✓		
	16	✓	✓	✓	✓	✓	✓	✓		
	17	✓	✓	✓	✓	✓	✓	✓		
	18	✓	✓	✓	✓	✓	✓	✓		
	19	✓	✓	✓	✓	✓	✓	✓		
	20	✓	✓	✓	✓	✓	✓	✓		
	21				✓	✓	✓	✓		

Bibliographic Data

Application No: 15/238,238

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Huy C. Ho/ HH

Examiner's Signature Initials

Title: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
08/16/2016	455	2644	N0484.71090US04
RULE			

APPLICANTS

Nuance Communications, Inc., Burlington, MA,

INVENTORS

William F. Ganong III, Brookline, MA, UNITED STATES

Paul Adrian Van Mulbregt Wayland, MA, UNITED STATES

Vladimir Sejnoha Lexington, MA, UNITED STATES

Glen Edward Wilson Maynard, MA, UNITED STATES

CONTINUING DATA

This application is a CON of 13795933 03/12/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

08/25/2016

STATE OR COUNTRY

UNITED STATES

ADDRESS

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

FILING FEE RECEIVED

$1,600

DOCKET NO.: N0484.71090US04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: _____March 6, 2020_____ Signature: _____/Miranda W. Stopa/_____ (Miranda W. Stopa)

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed more than three months after the filing date of this application and after the mailing date of a first Office Action, but before the mailing date of any of a Final Action under 37 C.F.R. § 1.113, a Notice of Allowance under 37 C.F.R. § 1.311, or an action that otherwise closes prosecution in this application.

Please charge our Credit Card in the amount of $240.00 covering the fee set forth in 37 C.F.R. § 1.17(p).

7977633.1

/HUY C HO/ 03/09/2020

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
PCT/CN2016/105343	**05-23-2019**	**International Preliminary Report on Patentability**	**N0484.71177WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

7977633.1

/HUY C HO/ 03/09/2020

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Andrew J. Tibbetts/
Andrew J. Tibbetts
Registration No.: 65,139
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: N0484.71090US04
Date: March 6, 2020

7977633.1

/HUY C HO/ 03/09/2020

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	11	(9,747,899 | 10,332,525 | 2012/0281885 | 2014/0039888 | 2014/0163978 | 2015/0006176 | 2015/0053779 | 2016/00788692019/0073999 2019/0287526 2019/0311715 2019/0355365).PN.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/06 14:59
L2	13	(9558749 | 9747899 | 10332525 | 2012/0281885 | 2014/0039888 | 2014/0163978 | 2015/0006176 | 2015/0053779 | 2016/0078869 | 2019/0073999 | 2019/0287526 | 2019/0311715 | 2019/0355365).PN.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/06 15:00
L3	3	"20120281885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/06 15:05
L4	121	(9558749 | 9747899 | 10332525 | 20120281885 | 20140039888 | 20140163978 | 20150006176 | 20150053779 | 20160078869 | 20190073999 | 20190287526 | 20190311715 | 2190355365)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/06 15:06
S1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/03/18 21:04
S2	104	(("GANONG") near3 ("William")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S3	16	(("VAN") near3 ("MULBREGT") near3 ("Paul")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S4	38	(("SEJNOHA") near3 ("Vladimir")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S5	33	(("WILSON") near3 ("Glen")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:20
S6	220	(("NUANCE") near3 ("COMMUNICATIONS") near3 ("INC")).AANM.	USPAT	OR	OFF	2017/04/20 10:20
S7	43	("8977255" | "9112984" | "9361885" | "20070129949" | "20120329389" |	US-PGPUB; USPAT	OR	OFF	2017/04/20 11:21

		"7574361" | "20100124896" | "20030040339" | "20100289994" | "20140274211" | "20140278435" | "8548176" | "20100185448" | "20110054899" | "20130080167" | "20130080171" | "20130289994" | "20130339028" | "20140012573" | "20140012586" | "5428707" | "5774859" | "5802305" | "5983186" | "6070140" | "6092043" | "6397186" | "6408396" | "6411926" | "6449496" | "6594630" | "6756700" | "6941265" | "6965786" | "7114090" | "7567827" | "7720683" | "7774204" | "8056070" | "8181046" | "8190420" | "8285545" | "8666751").PN.				
S8	5	"20120034904"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:10
S9	9	"9361885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:32
S10	33784	(379/88.01-88.02 |G10L15/20 |G10L15/22 |G10L15/24 |G10L15/25 |G10L2015/223 |G10L2015/226 |H04M2250/74 |H04W52/0251 |H04W52/0254 |).ccls,cpc.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:11
S11	1	("15".src. and "238238".ap.) or "15238238".RLAN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/04/20 16:12
S12	25130	S10 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:13
S13	4617	S12 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S14	2692	S13 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31

S15	563	S14 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S16	362	S15 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S17	17	S16 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:33
S18	9564	(704/246-255).ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S19	8758	S18 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S20	1733	S19 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S21	1338	S20 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S22	212	S21 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S23	104	S22 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	ON	2017/04/20 16:36

			DERWENT; IBM_TDB			
S24	3	S23 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S25	3	"9646610"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:41
S26	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/09/20 13:42
S27	8	"20100121636"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:42
S28	5	"20120035924"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:43
S29	6	"20120127072"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:45
S30	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/03/09 10:35
S31	12	"20140012573"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2018/03/09 10:53
S32	10	"20100009719"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2018/03/09 10:55
S33	2	("20160358605").PN.	US-PGPUB;	OR	ON	2018/05/30

			USPAT; USOCR; DERWENT			09:37
S34	78	"9940936"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:18
S35	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2019/07/26 18:19
S36	3	"20140249817"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:20
S37	2	"20140365225"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:24
S38	2	"20160189706"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:26
S39	1050	UTSM	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/08/11 10:17
S40	169	"6006175"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/06 12:27
S41	72	"6859776"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/06 12:36
S42	11	"9087520"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	ON	2020/03/06 12:42

			DERWENT; IBM_TDB			
S43	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/03/06 12:52
S44	23	"9558749"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/06 12:53

3/6/2020 4:16:43 PM
C:\Users\hho3\Documents\EAST_OLD2\Workspaces\15238238.wsp

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 2	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		6,006,175	A	Holzrichter	12-21-1999
		6,859,776	B1	Cohen et al.	02-22-2005
		9,087,520	B1	Salvador	07-21-2015
		9,558,749	B1	Secker-Walker et al.	01-31-2017
		9,747,899	B2	Pogue et al.	08-29-2017
		10,332,525	B2	Secker-Walker et al.	06-25-2019
		2012-0281885	A1	Syrdal et al.	11-08-2012
		2014-0039888	A1	Taubman et al.	02-06-2014
		2014-0163978	A1	Basye et al.	06-12-2014
		2015-0006176	A1	Pogue et al.	01-01-2015
		2015-0053779	A1	Adamek et al.	02-26-2015
		2016-0078869	A1	Syrdal et al.	03-17-2016
		2019-0073999	A1	Premont et al.	03-07-2019
		2019-0287526	A1	Ren et al.	09-19-2019
		2019-0311715	A1	Pfeffinger et al.	10-10-2019
		2019-0355365	A1	Kim et al.	11-21-2019

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		International Preliminary Report on Patentability for International Application No. PCT/CN2016/105343 mailed May 23, 2019 (N0484.71177WO00).	

EXAMINER:	DATE CONSIDERED:
/HUY C HO/	03/06/2020

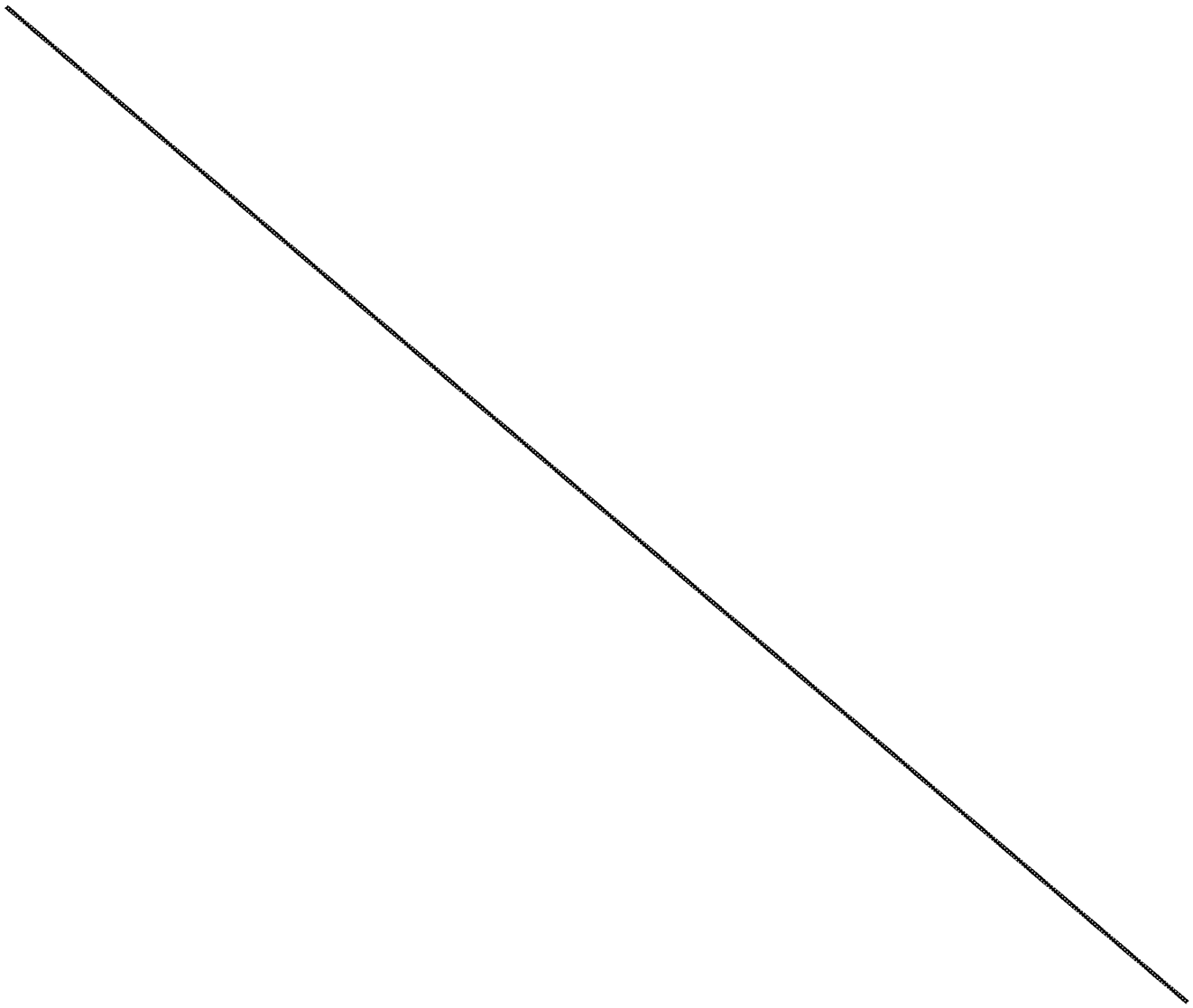
EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7977632.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
				FIRST NAMED INVENTOR: William F. Ganong	
Sheet	2	of	2	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]



EXAMINER:	DATE CONSIDERED:
/HUY C HO/	03/06/2020

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7977632.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

CPC - Searched*		
Symbol	Date	Examiner
G10L15/20; G10L15/22; G10L15/24; G10L15/25; G10L2015/223; G10L2015/226; H04M2250/74; H04W52/0251; H04W52/0254.	4/20/2017	HH

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner
379	88.01	4/20/2017	HH

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
EAST Database Search	4/20/2017	HH
Inventor Search	4/20/2017	HH
Search update	1/10/2018	HH
Search update	8/20/2018	HH
Search update	02/24/2019	HH
Search update	08/13/2019	HH
Search update	02/18/2020	HH
Search update	03/04/2020	HH

/HUY C HO/ Primary Examiner.Art Unit 2644	

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner
379	88.02	4/20/2017	HH
704	246-255	4/20/2017	HH

/HUY C HO/ Primary Examiner.Art Unit 2644	

PATENT COOPERATION TREATY

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING
TRANSMITTAL OF COPY OF INTERNATIONAL
PRELIMINARY REPORT ON PATENTABILITY
(CHAPTER I OF THE PATENT COOPERATION
TREATY)
(PCT Rule 44bis.1(c))

To:

JOHANNES, Marc, S.
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)* 23 August 2018 (23.08.2018)	
Applicant's or agent's file reference N0484.71176	**IMPORTANT NOTICE**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/017317	10 February 2016 (10.02.2016)	

Applicant
NUANCE COMMUNICATIONS, INC.

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Agnès Wittmann-Regis

e-mail: pct.team6@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference N0484.71176	**FOR FURTHER ACTION** See item 4 below	
International application No. PCT/US2016/017317	International filing date *(day/month/year)* 10 February 2016 (10.02.2016)	Priority date *(day/month/year)*
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant NUANCE COMMUNICATIONS, INC.		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 6 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☐	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 14 August 2018 (14.08.2018)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Agnès Wittmann-Regis e-mail: pct.team6@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/017317	10.02.2016	

International Patent Classification (IPC) or both national classification and IPC
INV. G10L15/32 ADD. G10L21/0216

Applicant
NUANCE COMMUNICATIONS, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	see form PCT/ISA/210	Zimmermann, Elko Telephone No. +49 89 2399-0

Form PCT/ISA/237 (Cover Sheet) (January 2015)

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	2-17, 19, 20
	No:	Claims	1, 18
Inventive step (IS)	Yes:	Claims	6, 11, 13, 15
	No:	Claims	1-5, 7-10, 12, 14, 16-20
Industrial applicability (IA)	Yes:	Claims	1-20
	No:	Claims	

2. Citations and explanations

see separate sheet

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1 The subject-matter of claim 1 is not new (Article 33(2) PCT), because EP 1 511 010 A1 discloses a system for detecting a designated wake-up word ("das Wort 'ich'", column 6, line 13), the system comprising:

a plurality of microphones ("Mikrofonanordnung 12" of figure 1, column 5, line 3) to detect acoustic information from a physical space having a plurality of acoustic zones ("Raumrichtungen", see figure 2 and paragraph [0025]);

at least one processor configured to:

receive a first acoustic signal representing the acoustic information received by the plurality of microphones ("Mikrofonanordnung 12");

process ("Optimierungseinheit 10" of figure 1, column 7, line 20) the first acoustic signal to identify content of the first acoustic signal originating from each of the plurality of acoustic zones ("Raumrichtungen");

provide a plurality of second acoustic signals ("vorverarbeitete Mikrofonsignale", column 6, line 33), each of the plurality of second acoustic signals substantially corresponding to the content identified as originating from a respective one of the plurality of acoustic zones ("aus unterschiedlichen Raumrichtungen", column 6, line 35); and

performing automatic speech recognition ("Spracherkenner 11" of figure 1, column 6, line 24) on each of the plurality of second acoustic signals to determine whether the designated wake-up word ("das Wort 'ich'") was spoken (see column 6, lines 13-23).

2 The objection raised against claim 1 is accordingly raised against corresponding claim 18.

3 The features of claims 1 and 2 could serve as basis for a new claim that meets the requirements of Articles 33(2) and (3) PCT.

In order to clarify the difference with respect to the prior art cited above the "wake-up word" claimed in claim 1 should be defined more precisely in the new claim, for example in view of the last three lines of page 1 and the first two lines of page 2, which teach that unless and until the system detects the wake-up word, the system assumes that the acoustic input received from the environment is not directed or intended for the system and will not process the acoustic input further, and once the wake-up word has been detected, the system begins recognizing subsequent speech spoken by the user.

Further, the additional steps of claim 2 should be defined more precisely in the new claim, for example in view of the last sentence of paragraph [0021], which teaches that once a first acoustic zone has been selected, the system performs speech recognition on sounds produced from the selected acoustic zone whilst excluding sounds produced from other acoustic zones.

4 Claim 18 could be amended accordingly.

PTO/AIA/22 (03-13)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)	Docket Number (Optional) N0484.71090US04
Application Number 15/238,238-Conf. #5864	Filed August 16, 2016
For METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	
Art Unit 2644	Examiner H. C. Ho

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a reply in the above-identified application.

The requested extension and fee are as follows (check time period desired and enter the appropriate fee below):

	Fee	Small Entity Fee	Micro Entity Fee	
[x] One month (37 CFR 1.17(a)(1))	$200	$100	$50	$ 200.00
[ ] Two months (37 CFR 1.17(a)(2))	$600	$300	$150	$
[ ] Three months (37 CFR 1.17(a)(3))	$1,400	$700	$350	$
[ ] Four months (37 CFR 1.17(a)(4))	$2,200	$1,100	$550	$
[ ] Five months (37 CFR 1.17(a)(5))	$3,000	$1,500	$750	$

[ ] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

[ ] A check in the amount of the fee is enclosed.

[x] Payment by credit card. ~~Form PTO-2038 is attached~~.

[ ] The Director has already been authorized to charge fees in this application to a Deposit Account.

[x] The Director is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 23/2825 .

[x] Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

[ ] applicant.

[ ] attorney or agent of record. Registration number _____.

[x] attorney or agent acting under 37 CFR 1.34. Registration number 72,100 .

/David S. Gesner/ — Signature
July 13, 2020 — Date

David S. Gesner — Typed or printed name
617.646.8000 — Telephone Number

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. Submit multiple forms if more than one signature is required, see below*.

[ ] * Total of 1 forms are submitted.

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 13, 2020 Electronic Signature for : /Trish McDonald/

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
Johannes, Marc S.
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
ETATS-UNIS D'AMERIQUE

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

Applicant's or agent's file reference N0484.71176	Date of mailing *(day/month/year)* 12 May 2016 (12-05-2016)
International application No. PCT/US2016/017317	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
Applicant NUANCE COMMUNICATIONS, INC.	International filing date *(day/month/year)* 10 February 2016 (10-02-2016)

1. [X] The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the International Application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the International Search Report.

   How? Directly to the International Bureau of WIPO, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Fascimile No.: (41-22) 338.82.70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 - 9.011.

2. [ ] The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. [ ] **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

   [ ] the protest together with the decision thereon has been transmitted to the International Bureau together with any applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

   [ ] no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **19 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the International Searching Authority European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040 Fax: (+31-70) 340-3016	Authorized officer WAGNER, Judit Tel: +49 (0)89 2399-8023

RECEIVED
Wolf, Greenfield & Sacks, P.C.
MAY 19 2016

Docketed ✓ Already Docketed
Not Required
Initials 1st K 2nd eg

Form PCT/ISA/220 (July 2014)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
N0484.71176		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US2016/017317	10 February 2016 (10-02-2016)	
Applicant		
NUANCE COMMUNICATIONS, INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**
   a. With regard to the **language,** the international search was carried out on the basis of:
      [X] the international application in the language in which it was filed
      [ ] a translation of the international application into _____, which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b))
   b. [ ] This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).
   c. [ ] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. [ ] **Certain claims were found unsearchable** (See Box No. II)

3. [ ] **Unity of invention is lacking** (see Box No III)

4. With regard to the **title,**
   [X] the text is approved as submitted by the applicant
   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**
   [X] the text is approved as submitted by the applicant
   [ ] the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority

6. With regard to the **drawings,**
   a. the figure of the **drawings** to be published with the abstract is Figure No. 1
      [X] as suggested by the applicant
      [ ] as selected by this Authority, because the applicant failed to suggest a figure
      [ ] as selected by this Authority, because this figure better characterizes the invention
   b. [ ] none of the figures is to be published with the abstract

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/017317

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/32
ADD. G10L21/0216

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 511 010 A1 (DAIMLER CHRYSLER AG [DE]) 2 March 2005 (2005-03-02)	1-5, 7-10,12, 14,16-20
A	paragraph [0018] - paragraph [0026]; figures 1,2	6,11,13, 15
A	US 2015/106085 A1 (LINDAHL ARAM M [US]) 16 April 2015 (2015-04-16) abstract; figure 3 paragraph [0002] - paragraph [0010]	1-20
A	US 2012/310646 A1 (HU JWU-SHENG [TW] ET AL) 6 December 2012 (2012-12-06) abstract; figure 1 paragraph [0008]	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 May 2016	12/05/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Zimmermann, Elko

3

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/017317

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1511010 A1	02-03-2005	DE 10339973 A1	17-03-2005
		EP 1511010 A1	02-03-2005
		JP 2005079098 A	24-03-2005
		US 2005049864 A1	03-03-2005
US 2015106085 A1	16-04-2015	NONE	
US 2012310646 A1	06-12-2012	TW 201250670 A	16-12-2012
		US 2012310646 A1	06-12-2012

Information on Search Strategy - Pilot phase (see OJ 2015, A86)
The type of information contained in this sheet may change during the pilot for improving the usefulness of this new service.

Application Number
PCT/US2016/017317

TITLE: TECHNIQUES FOR SPATIALLY SELECTIVE WAKE-UP WORD RECOGNITION AND RELATED SYSTEMS AND METHODS

APPLICANT: NUANCE COMMUNICATIONS, INC.

IPC CLASSIFICATION: G10L15/32, G10L21/0216

EXAMINER: Zimmermann, Elko

CONSULTED DATABASES: WPI, DOSYS

CLASSIFICATION SYMBOLS DEFINING EXTENT OF THE SEARCH:

IPC:

CPC: G10L15/32, G10L2021/02166

FI/F-TERMS:

KEYWORDS OR OTHER ELEMENTS FEATURING THE INVENTION:
Once a wake-up word has been detected in one of a plurality of acoustic zones, sounds from this acoustic zone are speech recognized.

EPO FORM P04A42

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/017317	10.02.2016	

International Patent Classification (IPC) or both national classification and IPC
INV. G10L15/32 ADD. G10L21/0216

Applicant
NUANCE COMMUNICATIONS, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	see form PCT/ISA/210	Zimmermann, Elko Telephone No. +49 89 2399-0

Form PCT/ISA/237 (Cover Sheet) (January 2015)

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	2-17, 19, 20
	No:	Claims	1, 18
Inventive step (IS)	Yes:	Claims	6, 11, 13, 15
	No:	Claims	1-5, 7-10, 12, 14, 16-20
Industrial applicability (IA)	Yes:	Claims	1-20
	No:	Claims	

2. Citations and explanations

see separate sheet

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1 The subject-matter of claim 1 is not new (Article 33(2) PCT), because EP 1 511 010 A1 discloses a system for detecting a designated wake-up word ("das Wort 'ich'", column 6, line 13), the system comprising:

a plurality of microphones ("Mikrofonanordnung 12" of figure 1, column 5, line 3) to detect acoustic information from a physical space having a plurality of acoustic zones ("Raumrichtungen", see figure 2 and paragraph [0025]);

at least one processor configured to:

receive a first acoustic signal representing the acoustic information received by the plurality of microphones ("Mikrofonanordnung 12");

process ("Optimierungseinheit 10" of figure 1, column 7, line 20) the first acoustic signal to identify content of the first acoustic signal originating from each of the plurality of acoustic zones ("Raumrichtungen");

provide a plurality of second acoustic signals ("vorverarbeitete Mikrofonsignale", column 6, line 33), each of the plurality of second acoustic signals substantially corresponding to the content identified as originating from a respective one of the plurality of acoustic zones ("aus unterschiedlichen Raumrichtungen", column 6, line 35); and

performing automatic speech recognition ("Spracherkenner 11" of figure 1, column 6, line 24) on each of the plurality of second acoustic signals to determine whether the designated wake-up word ("das Wort 'ich'") was spoken (see column 6, lines 13-23).

2 The objection raised against claim 1 is accordingly raised against corresponding claim 18.

3 The features of claims 1 and 2 could serve as basis for a new claim that meets the requirements of Articles 33(2) and (3) PCT.

In order to clarify the difference with respect to the prior art cited above the "wake-up word" claimed in claim 1 should be defined more precisely in the new claim, for example in view of the last three lines of page 1 and the first two lines of page 2, which teach that unless and until the system detects the wake-up word, the system assumes that the acoustic input received from the environment is not directed or intended for the system and will not process the acoustic input further, and once the wake-up word has been detected, the system begins recognizing subsequent speech spoken by the user.

Further, the additional steps of claim 2 should be defined more precisely in the new claim, for example in view of the last sentence of paragraph [0021], which teaches that once a first acoustic zone has been selected, the system performs speech recognition on sounds produced from the selected acoustic zone whilst excluding sounds produced from other acoustic zones.

4 Claim 18 could be amended accordingly.

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
- [ ] Fee Attached
- [x] Amendment/Reply
- [ ] After Final
- [ ] Affidavits/declaration(s)
- [x] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
- [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
- [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

Request for Continued Examination

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/David S. Gesner/		
Printed name	David S. Gesner		
Date	July 13, 2020	Reg. No.	72,100

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 13, 2020 Electronic Signature for : /Trish McDonald/

8040462.1

FEE TRANSMITTAL

Complete if known	
Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong
Examiner Name	H. C. Ho
Art Unit	2644
Practitioner Docket No.	N0484.71090US04

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)2,100.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☒ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☒ Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☐ Charge fee(s) indicated below

☐ Charge fee(s) indicated below, **except for the filing fee**

☒ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17, **except for any excess claims fees or multiple dependent claim fee**

☐ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17

☒ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	300	150*	75	660	330	165	760	380	190	
Design	200	100	50	160	80	40	600	300	150	
Plant	200	100	50	420	210	105	620	310	155	
Reissue	300	150	75	660	330	165	2,200	1,100	550	
Provisional	280	140	70	0	0	0	0	0	0	

* The $150 small entity status filing fee for a utility application is further reduced to $75 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	100	50	25
Each independent claim over 3 (including Reissues)	460	230	115
Multiple dependent claims	820	410	205

Total Claims		**Extra Claims**		**Fee ($)**		**Fee Paid ($)**
_____	- 21 or HP =	_____	x	_____	=	_____

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims		**Extra Claims**		**Fee ($)**		**Fee Paid ($)**
_____	- 3 or HP =	_____	x	_____	=	_____

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		**Extra Sheets**		**Number of each additional 50 or fraction thereof**		**Fee ($)**		**Fee Paid ($)**
_____	- 100 =	_____	/50 =	_____	(round **up** to a whole number) x	_____	=	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): 1820 Request for continued examination (RCE) - 2nd and ...	1,900.00
1251 Extension for response within first month	200.00

SUBMITTED BY

Signature	/David S. Gesner/	Registration No. (Attorney/Agent)	72,100	Telephone	617.646.8000
Name (Print/Type)	David S. Gesner			Date	July 13, 2020

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 13, 2020 Electronic Signature for : /Trish McDonald/

8195774.1

Electronic Acknowledgement Receipt

EFS ID:	39986979
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	David Scott Gesner/Patricia McDonald
Filer Authorized By:	David Scott Gesner
Attorney Docket Number:	N0484.71090US04
Receipt Date:	13-JUL-2020
Filing Date:	16-AUG-2016
Time Stamp:	19:09:32
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$2100
RAM confirmation Number	E20207CJ09543286
Deposit Account	232825
Authorized User	Patricia McDonald

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.17 (Patent application and reexamination processing fees)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-DSG.pdf	26331 7eabb6422a1d89e960648b57faff22282cfcc07f	no	1
Warnings:					
Information:					
2	Extension of Time	N048471090US04-EXT-DSG.pdf	24257 464ae583e87c710229e2264842ffd7f039512822	no	1
Warnings:					
Information:					
3	Fee Worksheet (SB06)	N048471090US04-FEE-DSG.pdf	32746 ae8a0fce22f48a2cee6933ca13157a2d239ae8fe	no	1
Warnings:					
Information:					
4	Request for Continued Examination (RCE)	N048471090US04-RCE-DSG.pdf	28023 4c87f99f612ac0e343b20493038b9374179c6552	no	1
Warnings:					
This is not a USPTO supplied RCE SB30 form.					
Information:					
5		N048471090US04-AMN-DSG.pdf	71504 cd0f7119bec1cb4cd23a1913dfc11f031a45a26e	yes	16

Multipart Description/PDF files in .zip description

Document Description	Start	End
Amendment Submitted/Entered with Filing of CPA/RCE	1	1
Claims	2	7

	Applicant Arguments/Remarks Made in an Amendment		8	16	
Warnings:					
Information:					
6	Transmittal Letter	N048471090US04-IDSBODY-DSG.pdf	34646 e3c2337a698069da66c7e3e2ba308564d26b6bc8	no	4
Warnings:					
Information:					
7	Information Disclosure Statement (IDS) Form (SB08)	N048471090US04-IDS1449-DSG.pdf	22735 3e0674174cd68173dd32786884f71abac28c5488	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
8	Other Reference-Patent/App/Search documents	N048471176WO00_IPRP_mailed_08-23-2018.pdf	217029 3803a030c7771a33c779daa4f02127800016e689	no	7
Warnings:					
Information:					
9	Other Reference-Patent/App/Search documents	N048471176WO00_ISRWO_mailed_05-12-2016.PDF	1028806 9acf358d2fae3caef656fafef816e729ecd72e64	no	10
Warnings:					
Information:					
10	Foreign Reference	EP1511010A1.pdf	729001 ead4b4ef6aa4830d9af0195d4694bc54392799b9	no	14
Warnings:					
Information:					
11	Fee Worksheet (SB06)	fee-info.pdf	32760 5dc01d595cbedecfe9f8fba650de006cc4133eea	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	2247838		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Espacenet

Bibliographic data: EP1511010 (A1) — 2005-03-02

Control of a microphone array using feedback of a speech recognition system, and speech recognizion using said array

Inventor(s): KALTENMEIER ALFRED DR ING [DE]; LINHARD KLAUS DR ING [DE] ± (KALTENMEIER, ALFRED DR.ING, ; LINHARD, KLAUS DR.ING)

Applicant(s): DAIMLER CHRYSLER AG [DE] ± (DAIMLERCHRYSLER AG)

Classification: - international: ***F21V29/02; G10L15/20;*** *F21W101/10; F21Y101/02; G10L21/0216;* (IPC1-7): G10L15/20
- cooperative: **G10L15/20**; G10L2021/02166

Application number: EP20040013378 20040607 Global Dossier

Priority number (s): DE2003139973 20030829

Also published as: EP1511010 (B1) DE10339973 (A1) JP2005079098 (A) US2005049864 (A1) US7505901 (B2)

Abstract of EP1511010 (A1)

An optimizing unit (OU) (10) controls a microphone system (MS) (12) regarding its receiving parameters. The MS picks up and processes voice signals. After processing the signals via the OU, the MS receiving parameters are controlled so that recognition performance is optimized for a voice identifier (11) subordinate to the OU. An independent claim is also included for a device for controlling a voice recognition system.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 511 010 A1

(12) EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:
02.03.2005 Patentblatt 2005/09

(51) Int Cl.[7]: **G10L 15/20**

(21) Anmeldenummer: **04013378.7**

(22) Anmeldetag: **07.06.2004**

(84) Benannte Vertragsstaaten:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PL PT RO SE SI SK TR
Benannte Erstreckungsstaaten:
AL HR LT LV MK

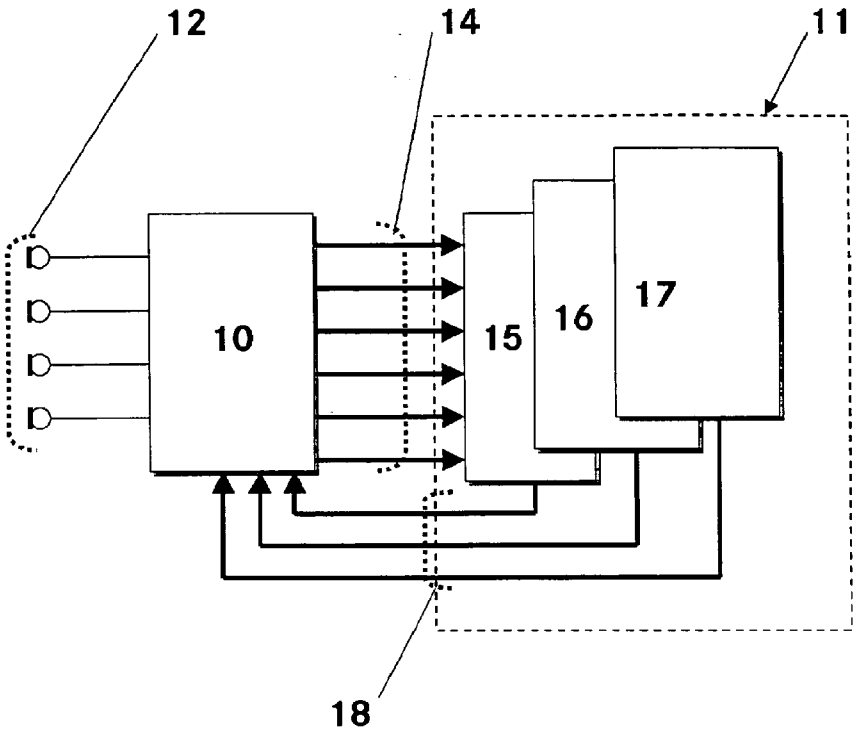
(30) Priorität: **29.08.2003 DE 10339973**

(71) Anmelder: **DaimlerChrysler AG**
70567 Stuttgart (DE)

(72) Erfinder:
- **Kaltenmeier, Alfred Dr.Ing**
  89075 Ulm (DE)
- **Linhard, Klaus Dr.Ing**
  89601 Schelklingen (DE)

(54) Intelligentes akustisches Mikrofon-Frontend mit Spracherkenner-Feedback

(57) Hierbei wird zur Regelung des Spracherkennungssystems eine richtungsselektive Mikrofonanordnung (12) durch ein Optimierungseinheit (10) in ihrer Empfangscharakteristik gesteuert. Die hiermit empfangenen Sprachsignale werden sodann zumindest zeitweise in einem Spracherkenner (11) parallel verarbeitet. Auf Grundlage der von dem Spracherkenner (11) gelieferten Ergebnisse wird sodann über die Optimierungseinheit (10) die Empfangscharakteristik der Mikrofonanordnung (12) so gesteuert, dass die Erkennungsleistung des nachgeordneten Spracherkenners optimiert wird.



Figur 1

EP 1 511 010 A1

Printed by Jouve, 75001 PARIS (FR)

Beschreibung

[0001] Die Erfindung betrifft ein intelligentes akustisches Mikrofon-Frontend und ein Verfahren zum Betrieb eines solchen nach dem Oberbegriff der Patentansprüche 1 und 6.

[0002] In modernen Kraftfahrzeugen werden zunehmend Systeme installiert, welche über Spracheingabe bedient werden können. Hierbei sollen solche Systeme durch unterschiedliche Fahrzeuginsassen aus unterschiedlichen Positionen bedienbar sein. Auch ist es Ziel, im Rahmen moderner Insassenkommunikationsanlagen die einzelnen Plätzen im Fahrzeuginnenraum mit diesen spezifisch zugeordneten Mikrofonen und Lautsprechern auszustatten. Auch hierbei sollte sichergestellt werden, dass der einzelnen Insasse unabhängig von der physikalischen Gestaltung des Insassenkommunikationssystems zu jeder Zeit und in jeder Position komfortabel kommunizieren kann.

[0003] Die japanische Schrift JP 2002-091469 zeigt ein Spracherkennungssystem, welches zur Aufnahme von Sprachsignalen über ein durch eine Einheit zur Strahlformung (Beamforming Unit) schaltbares, richtungsselektives Mikrofonarray verfügt. Die Einheit zur Strahlformung umfasst hierbei einen Richtungserkenner, welcher aus den aufgenommenen Sprachsignalen die auf die Sprechrichtung bezogene Winkelablagen erkennt. Hierzu werden aus unterschiedlichen Winkelablagen hinsichtlich der Empfangscharakteristik des Mikrofonarrays Signale empfangen und einer Schwellwertuntersuchung zugeführt. Im Rahmen dieser Schwellwertuntersuchung liefert der Richtungserkenner als Erkennungsergebnis die wahrscheinlichste Richtung, aus welcher das Sprachsignal herrührt. Dieses Ergebnis wird sodann von der Einheit zur Strahlformung zur Steuerung des Mikrofonarrays herangezogen, so dass das Empfangsdiagramm des Mikrofonarrays möglichst auf die Richtung hin ausgerichtet wird, aus welcher das wahrscheinlich beste Sprachsignal herrührt.

[0004] Aus der Offenlegungsschrift EP 1 081 682 A2 ist ein Spracherkennungssystem für eine Insassenkommunikationsanlage bekannt, welche eine Vielzahl unterschiedlich positionierter Mikrofonsysteme umfasst. Die Signale der einzelnen Mikrofonsysteme werden parallel erfasst und hinsichtlich der Parameter Signal-Rausch-Verhältnis und des im Signal enthaltenen mittleren Sprachpegels untersucht. Diejenigen Signale, deren Parameter bestimmte Schwellwerte überschreiten, werden einem Spracherkenner zugeführt, welcher sequentiell die einzelnen Signale bearbeitet, wobei mit dem stärksten Signal begonnen wird. Aus den aus den Einzelsignalen resultierenden positiven Erkennungsergebnissen wird der Sprachinhalt der Signale ermittelt. Durch die parallele Untersuchung der an unterschiedlichen Stellen aufgenommenen Mikrofonsignale ist ein Sprecher nicht mehr daran gebunden, die Sprachinformation in eine bestimmte Richtung hin zu sprechen; auch wird es möglich, dass er während des Sprechvorgangs die Sprechrichtung ändert, ohne dass Information verloren geht.

[0005] Eine weitere Spracherkennungseinrichtung zur parallelen Verarbeitung mehrerer paralleler Mikrofonkanäle wird in der deutschen Offenlegungsschrift DE 100 30 105 A1 beschrieben. Hierbei werden die einzelnen Mikrofonkanäle unabhängig voneinander einer Merkmalsextraktion unterzogen und erst im Rahmen einer gemeinsamen Klassifikation zusammengeführt. Auf diese Weise ist in der der Klassifikation zugeführten Information noch die den einzelnen Kanälen individuell zuordenbare Information enthalten. Auch wird es möglich die Vorverarbeitung in den einzelnen Kanälen individuell zu gestalten und die daraus resultierenden Auswirkungen auf die Signale unabhängig zu beobachten.

[0006] Aufgabe der Erfindung ist es ein Spracherkennungssystem zu schaffen, welches zum einen eine verbesserte Spracherfassung aufweist und zum anderen eine optimierte Steuerung der Spracherfassung ermöglicht.

[0007] Die Aufgabe wird durch ein Verfahren und eine zur Durchführung des Verfahrens geeignete Vorrichtung mit den Merkmalen der Patentansprüche 1 und 6 gelöst. Vorteilhafte Ausgestaltungen und Weiterbildungen der Erfindung sind in den Unteransprüchen beschrieben.

[0008] Zur Regelung des Spracherkennungssystems wird eine Mikrofonanordnung (12) durch eine Optimierungseinheit (10) in ihrer Empfangscharakteristik modifiziert, wobei mittels der Mikrofonanordnung (12) Sprachsignale empfangen werden. Diese Sprachsignale werden sodann in einem Spracherkenner (11) zumindest zeitweise parallel verarbeitet. Auf Grundlage der von dem Spracherkenner (11) gelieferten Ergebnisse wird sodann über die Optimierungseinheit (10) die Empfangscharakteristik der Mikrofonanordnung (12) so gesteuert, dass die Erkennungsleistung des der Optimierungseinheit (10) nachgeordneten Spracherkenners (11) optimiert wird. In erfinderischer Weise werden hierbei dem Spracherkenner die empfangenen Sprachsignale parallel oder zumindest quasi-parallel zugeführt über die Sprachkanäle (14), so dass dieser unterschiedliche Sprachsignale zumindest Zeitweise parallel verarbeiten kann. Die Erkennungsergebnisse werden sodann hinsichtlich ihrer Qualität miteinander verglichen, wobei auf Grundlage dieses Vergleichs diejenigen Sprachsignale auswählt und weiterverarbeitet werden, welche das Potential für eine bestmögliche Erkennungsleistung haben. Desweiteren erhält auf Grundlage der Erkennungsergebnisse die Optimierungseinheit (10) durch den Spracherkenner die notwendigen Regelsignale (18), um die Empfangscharakteristik der Mikrofonanordnung (12) mittels einer Optimierungseinheit (10) in Bezug auf diejenigen Sprachkanäle (14) zu optimieren, welche die Sprachsignale führen, welche das Potential für eine bestmögliche Erkennungsleistung haben.

Petitioner Exhibit 1004, Page 902 of 1052

[0009] Im Sinne dieser Erfindung ist selbstverständlich eine quasi-parallele Signalverarbeitung oder Signalweiterleitung mit einer parallelen Signalverarbeitung oder Signalweiterleitung gleichzusetzen. Bei einer quasi-parallelen Signalverarbeitung oder Signalweiterleitung erfolgt die Verarbeitung oder Weiterleitung zwar teilweise sequentiell (auf Grund der Prozessordimensionierung oder mangelnder Kapazität der Übertragungskanäle (14)), die Verarbeitungsergebnisse werden aber so behandelt, als ob der Verarbeitungs- oder Weiterleitungsvorgang tatsächlich parallel bzw. gleichzeitig verlaufen wäre.

[0010] In besonders vorteilhafte Weise erlaubt die Erfindung eine intelligente Optimierung der Empfangscharakteristik des Mikrofonarrays (12), da die Regelung auf den Erkennungsergebnissen eines Spracherkenners (11) basiert und nicht nur auf Grundlage rein abstrakter, physikalischer Signalparameter (Stärke des Signalpegels oder Verhältnisse der Signale an den Ausgängen der Einzelmikrofone des Mikrofonarrays).

[0011] Als Mikrofonanordnung eignet sich in besonderer Weise eine Mikrofonarray, also eine Zusammenschaltung einer Anordnung von mehreren Einzelmikrofonen. Hierbei können die Einzelmikrofone zu einer oder mehreren Gruppen zusammengeführt werden, wodurch sich unterschiedliche Richtungsselektivitäten in der Empfangscharakteristik der einzelnen Mikrofonarrays erzielen läst. Am Ausgang der Mikrofonanordnung (12) können sodann parallel oder quasi-parallel Sprachsignale aus -unterschiedlichen Raumrichtungen abgegriffen und einer Weiterverarbeitung in einem Spracherkenner (11) zugeführt werden.

[0012] Es ist denkbar die Empfangscharakteristik des Mikrofonarrays (12) auf unterschiedliche Weise zu modifizieren und in Bezug auf unterschiedliche Empfangsrichtungen hin zu optimieren:

- Einerseits ist es hierbei denkbar die Richtungsselektivität (Richtwirkung) der Empfangscharakteristik des Mikrofonarrays (12) aktiv zu steuern.

- Andererseits kann die Empfangscharakteristik auch dahingehend optimiert werden, dass mehreren der Einzelmikrofone des Mikrofonarrays Einheiten zur Geräuschreduktion nachgeschaltet werden.

[0013] Diese beiden vorteilhaften Optimierungsverfahren und Ausgestaltungsmöglichkeiten können einzeln oder auch in Kombination realisiert werden. So wird in besonders vorteilhafter Weise das Spracherkennungssystem so ausgestaltet, dass mehreren der Einzelmikrofone des Mikrofonarrays (12) Einheiten zur Geräuschreduktion zugeordnet werden, und die Optimierungseinheit (10) ausgehend von den Regelsignalen (18) des Spracherkenners (11) sowohl die Richtungsselektivität als auch die Geräuschreduktion steuert und modifiziert.

[0014] So aber gleichwohl auch denkbar, das die Mikrofonanordnung (12) als Mikrofonarray so auszugestalten, dass dessen Empfangscharakteristik eine oder mehrere spezifische Richtungsselektivitäten aufweist, welche unveränderlich, fest vorgegeben sind, sodass die Optimierungseinheit (10) nur mehrere der den Einzelmikrofonen nachgeschalteten Einheiten zur Geräuschreduktion ansteuert und so die Empfangscharakteristik des Mikrofonarrays (12) modifiziert. Auf diese Weise lassen sich Schaltungsaufwand und Kosten in Bezug auf das Mikrofonarray (12) minimieren. In gleicher Weise ist es selbstverständlich auch denkbar auf die mehreren der Einzelmikrofone nachgeschalteten Einheiten zur Geräuschreduktion zu verzichten.

[0015] Eine andere vorteilhafte Ausgestaltung der Mikrofonanordnung (12), welche sich insbesondere dann eignet wenn aus physikalischen oder designtechnischen Gründen nur ein geringer Einbauplatz vorhanden ist, besteht darin, dass nur ein einzelnen Mikrofon benutzt wird. Hierbei wird dessen Ausgangssignal auf mehrere parallele oder quasi-parallele Sprachkanäle aufgetrennt, welchen für diese Sprachkanäle spezifische Signalaufbereitungseinheiten (insbesondere Einheiten zur Geräuschreduktion) zugeordnet sind. Auf die Funktionsparameter dieser Signalaufbereitungseinheiten wird durch die Optimierungseinheit (10) sodann im Rahmen der Spracherkennung einfluss genommen, so dass diejenige sprachkanal-spezifische Signalaufbereitung optimiert wird, welche die besten Erkennungsergebnisse im Spracherkenner (11) liefert.

[0016] Die Erfindung lässt sich somit in besonders vorteilhafter Weise mit den unterschiedlichsten Arten von Mikrofonanordnungen koppeln, hierdurch in ihrer Funktionalität beeinflusst zu werden. Hierdurch lässt sich die Erfindung auch auf die unterschiedlichsten Einsatzorte (Kraftfahrzeug (->geringes Platzangebot), Konferenzraum (-> großzügiges Platzangebot)) und Einsatzbedingungen (eine Sprecherposition mit unterschiedlichen Sprechern oder aber mehrere gleichzeitige Sprecherpositionen mit starken Umgebungsgeräuschen) anpassen.

[0017] Nachfolgend wird anhand von Ausführungsbeispielen und mit Hilfe von Figuren die Erfindung im Detail erläutert. Dabei zeigen:

Fig. 1 zeigt den Aufbau einer vorteilhaften Ausführung des mehrstufigen, rückgekoppelten Sprachenerkennungssystems.

Fig. 2 zeigt die Variabilität. der durch die Einheit zur Strahlformung gesteuerten Empfangscharakterisiken des Mikrofonarrays auf.

Fig. 3 zeigt beispielhaft eine mögliche statistische Auswertung der Signale des Mikrofonarrays zur Erkennung und Lokalisation mehrerer Sprecher.

[0018] Mittels der Figuren werden insbesondere Aus-

Petitioner Exhibit 1004, Page 903 of 1052

führungsbeispiele beschrieben, bei welchen die Richtungsselektivität (Richtwirkung) der Empfangscharakteristik der Mikrofonanordnung (12) durch die Optimierungseinheit (10) beeinflusst wird. Es ist für den Fachmann jedoch naheliegend diese Ausführungsbeispiele auch auf Anwendungsfälle zu übertragen, bei welchem die Empfangscharakteristik der Mikrofonanordnung (12) zusätzlich oder nur mittels der Steuerung von Einheiten zur Geräuschreduktion, welche mehreren der Einzelmikrofonen nachgeschalteten sind, modifiziert wird. Entsprechendes gilt für den Fall, bei welchem die Mikrofonanordnung (12) nur ein einziges Mikrofon gebildet wird, dessen Ausgangssignal auf mehrere parallele oder quasi-parallele Kanäle mit kanal-spezifischer Signalaufbereitung (insbesondere Geräuschreduktion) aufgespaltet wird.

[0019] In besonders vorteilhafter Weise lässt sich das erfindungsgemäße Spracherkennungssystem ausgestalten, indem der Spracherkenner (11) dergestalt ausgebildet wird, dass die Spracherkennung im Rahmen eines mehrstufigen Prozesses erfolgt. Dabei erfolgt durch eine oder mehrere dieser Verarbeitungsstufen zum einen eine Auswahl von Sprachsignalen, welche das Potential für eine bestmögliche Erkennungsleistung haben und einer Weiterverarbeitung zugeführt werden sollen. Zum anderen stellen diese einzelnen Verarbeitungsstufen einzeln oder gemeinsam der Optimierungseinheit (10) die notwendigen Regelsignale (18) zur Verfügung, um die Empfangscharakteristik des Mikrofonarrays (12) zu optimieren. Bei dem dargestellten Fall, dass die Mikrofonanordnung (12) aus einem richtungsselektiven Mikrofonarray besteht, wird durch die Optimierungseinheit (10) insbesondere die Richtungsselektivität der Empfangscharakteristik der Mikrofonanordnung (12) in Bezug auf diejenigen Richtungen hin zu optimieren, aus welchen diejenigen Sprachsignale empfangen werden, welche das Potential für eine bestmögliche Erkennungsleistung haben.

[0020] Wie in Figur 1 aufgezeigt, wird in gewinnbringender Weise der Spracherkenner aus zumindest drei aufeinander folgenden Verarbeitungseinheiten gebildet, welche zumindest

- einen Klassifikator (15), zur Erkennung von Lauteinheiten und/oder Lautuntereinheiten von Lauteinheiten,
- eine Einheit zur Worterkennung (16), zur Erkennung gesprochener Worte,
- und ein Einheit zur Erkennung und Interpretation sinnvoller Sätze (17) umfasst.

[0021] Hierbei sind die einzelnen Verarbeitungseinheiten (15, 16, 17) des Spracherkenners (11) so miteinander verbunden, dass mehrere der zuvor parallel verarbeiteten Signale parallel oder quasi-parallel an die nachfolgende Verarbeitungsstufe wertergereicht werden können. So wird nicht nur das als ‚bestes' Signal identifizierte Sprachsignal weiterverarbeitet, sondern es ist möglich auch weitere ‚gute' Sprachsignale weiter zu berücksichtigen. Eine derartige Ausgestaltung ist insbesondere deshalb besonders vorteilhaft, da es sich in der Praxis manchmal zeigt, dass bestimmte Signale in einer der Verarbeitungsstufen zwar exzellente Ergebnisse liefern, in der darauffolgenden Verarbeitungsstufe aber nicht zu gebrauchen sind; hier liefern dann oft diejenigen Sprachsignale, welche in der vorausgehenden Verarbeitungsstufe mit nicht so gutem Ergebnis abgeschnitten haben im gesamten ein besseres Ergebnis. Wird beispielsweise das gesprochene Wort "ich" durch störende Umfeldgeräusche zu einem scheinbaren "mich", so würde dieser Fehler von der Einheit zur Sprachinterpretation (17) meist noch richtig auf "ich" umgedeutet, wenn dieser durch den Worterkenner parallel sowohl die Worte "mich" als auch "ich" zugeführt würden. Würde der Worterkenner (16) aber nur das Wort, welches er mit der höchsten Wahrscheinlichkeit erkennt weiter geleitet, so hätte die Einheit zur Sprachinterpretation (17) keine Möglichkeit zu Erkennen, dass in diesem Zusammenhang das Wort "ich" richtig ist.

[0022] Dem Spracherkenner (11) werden durch die Optimierungseinheit (10) mehrere Mikrofonsignale über Zuleitungen (15) parallel oder quasi parallel zugeführt. Hierbei kann es sich um Signale unterschiedlichster Art handeln. In einfachster Form entsprechen die parallel geführten Signale den Einzelsignalen der einzelnen die Mikrofonanordnung (10) bildenden Einzelmikrofone. Sehr wohl ist es aber auch denkbar, über die einzelnen parallel geführten Signalkanäle bereits durch die Optimierungseinheit (10) vorverarbeitete Mikrofonsignale zu führen. So ist es möglich, dass die Signalkanäle Signale aus unterschiedlichen Raumrichtungen (21) führen, welche durch adaptives Beamforming und zusammenführen der Signale mehrerer Einzelmikrofone (12) generiert wurden. Sehr wohl kann es sich aber auch um die Sprachsignale aus einer einzelnen Raumrichtung ($21_{(1-k)}$) handeln, welche mit unterschiedlich gestalteten Empfangscharakteristiken des Mikrofonarrays (beispielsweise: breite/schmale Empfangskeule, unterschiedlichste Empfangsparameter und -empfindlichkeiten oder unterschiedlichen Methoden bzw. Parameter bei der Geräuschreduktion)empfangen wurden. Mittels der Rückkopplung der Steuersignale vom Spracherkenner zum Optimierungseinheit (10) kann die Empfangscharakteristik des Mikrofonarrays also auch hinsichtlich der Anzahl und Art von Charakteristiken und auf Grundlage unterschiedlichster optimierter Geräuschreduktions-Parameter gesteuert werden.

[0023] Die Optimierungseinheit (10) erhält ihre Steuersignale über eine oder mehrere Steuerleitungen (18) vom Spracherkenner (11) beziehungsweise von einer seiner Untereinheiten (15, 16, 17). Der Klassifikator (15) kann beispielsweise bewirken, dass der Empfang von denjenigen Raumrichtungen unterdrückt wird, welche nur akustische Signale liefern, welche gar keine Sprach-

Petitioner Exhibit 1004, Page 904 of 1052

signale sind. Durch die Steuersignale der Einheit zur Worterkennung (16) kann die Empfangscharakteristik weiter optimiert werden, um ein geeignetes Empfangssignal zu generieren, aus welchem Wörter optimal extrahiert werden können. In Bezug auf den oben angeführten Beispielsfall, bei welchem das Wort "ich" durch Störsignale, zumindest bei bestimmten Empfangscharakterisiken als Wort "mich" vom Mikrofonarray empfangen wurde, kann durch gezielte Steuerung durch die Einheit zur Sprachinterpretation die Empfangscharakteristik des Mikrofonarrays ergänzend dergestalt optimiert werden, dass die in diesem Falle verfälschenden Störgeräusche bestmöglich weiter unterdrückt werden.

[0024] Dieses besonders vorteilhafte Konzept der mehrstufigen, parallelisierten Steuerung der Empfangscharakteristik der Mikrofonanordnung (12) erlaubt eine bestmögliche Elimination von Störeinflüssen und resultiert in einer optimalen Spracherkennung.

[0025] Beispielhaft zeigt Figur 2 die Variabilität der durch die Optimierungseinheit (10) gesteuerten Empfangscharakterisiken des Mikrofonarrays aus M Einzelmikrofonen (12) auf. Hierbei weist die Optimierungseinheit (10) N parallele Ausgänge (15) zur Übermittlung der Mikrofonsignale an den Spracherkenner (11) auf. Wie zuvor bereits ausgeführt muss die Anzahl N der Ausgänge (15) der Optimierungseinheit (10) nicht mit der Anzahl M der Einzelmikrofone (12) an seinem Eingang übereinstimmen, da die parallelen Ausgangssignale im allgemeinen nicht einfache Kopien der Ausgangssignale der Einzelmikrofone des Mikrofonarrays entsprechen, sondern meist aus teilweise komplexen Fusionen dieser Signale herrühren. Im der unteren Hälfte der Figur 2 sind beispielhaft sechs mögliche Empfangscharakteristika des Mikrofonarrays aufgezeigt. So ist es beispielsweise denkbar, Empfangscharakteristiken zu bilden, welche selektiv bezüglich unterschiedlicher Raumrichtungen sind (A, B, C) oder welche unterschiedliche Richtungsselektivität aufweisen (D, E) oder welche mit einer Nullstelle (zum Beispiel in Richtung einer dominanten Störquelle, wie einem geöffneten Fenster)versehen sind (F).

[0026] Bei dem in Figur 2 aufgezeigten Beispiel finden sich ein Sprecher (22) in etwa zwischen den virtuellen Raumsektoren 2 und 3, so dass mit den im unteren Teil der Figur 2 mit B, D, E und F bezeichneten, durch die Optimierungseinheit(10) gebildeten Empfangscharakterisiken die beste Erkennungsleitung zu erwarten ist. Aus dieser Figur 2 wird jedoch zugleich auch klar, dass es im Vorfeld einer eigentlichen Spracherkennung schwierig ist sich auf eine dieser vier Empfangscharakteristiken B, D, E und F festzulegen, so dass sicher besser ist, alle aus den vier Empfangscharakteristiken resultierenden Signale parallel im Spracherkenner (11) zu verarbeiten und erst im Nachhinein eine Auswahl oder gar Optimierung der Empfangscharakteristik vorzunehmen.

[0027] Die Adaption der Empfangscharakteristik (Beamforming)der Mikrofonanordnung (12) kann mittels bekannter Verfahren beispielsweise mittels der Methode nach Griffith-Jim oder dem Frost-Beamforming erfolgen.

[0028] Es ist zudem oft üblich einer Einheit zur Adaption der Empfangscharakteristik eines Mikrofons (Mikrofon-Beamformer) ein einkanaliges Geräuschreduktionsverfahren, wie zum Beispiel, das bekannte Verfahren der Spektralen Subtraktion (im Spektralbereich des Mikrofonsignals wird ein Geräusch-Schätzwert subtrahiert), nachzuschalten. Einfache Geräuschreduktionsverfahren benutzen beispielsweise nur die Spektrale Subtraktion. Im vorliegenden Fall kann das einkanalige Verfahren zur Nachverarbeitung der einzelnen Ausgänge der Einheit zur Adaption der Empfangscharakteristik verwendet werden. Die Spektrale Subtraktion bringt ihrerseits Parameter mit, wie unter anderem ein Faktor mit der Geräusch-Schätzwerte vor der Subtraktion gewichtet wird, oder eine untere Schwelle im Spektrum die nach der Subtraktion nicht unerschritten werden darf (der sogenannte Spectral Floor). Der Geräusch-Gewichtungsfaktor und die untere Schwelle des Spektrums sind Parameter. Aus jedem der Ausgänge des Mikrofon-Beamformers können nun wiederum mehrere Ausgänge erzeugt werden, wenn das Ausgangssignal des Beamformers mehrere Parameter-Varianten einer Spektralen Subtraktion durchläuft.

[0029] In besonders vorteilhafter Weise lässt sich das erfindungsgemäße Spracherkennungssystem dahingehend modifizieren, dass es in die Lage versetzt wird, in der Mikrofonanordnung (12) empfangenen Signalen mehrere Sprecher (30, 31) zu erkennen. Werden mehrere Sprecher (30, 31) erkannt, so können sodann die über die Optimierungseinheit (10) unterschiedlich wählbaren Empfangscharakteristiken auf die Sprechrichtung der unterschiedlichen Sprecher hin optimiert werden. Figur 3 dient der Erläuterung eines besonders vorteilhaften Verfahrens zur Erkennung mehrer Sprecher (30, 31) in dem von der Mikrofonanordnung (12) aufgenommenen Signal. Hierbei wird über einen bestimmten Zeitraum eine Statistik darüber geführt, wie oft jeder einzelne der N-Ausgänge der Optimierungseinheit (10) durch den Spracherkenner (11) als derjenige der N Sprachsignalkanäle (15) identifiziert wird, welcher das ‚beste' Signal liefert. Aus der Kenntnis der den einzelnen Sprachsignalkanälen (15) zuzuordnenden Empfangscharakteristik der Optimierungseinheit (10) kann sodann die Anzahl der Sprecher (30, 31) geschätzt werden. Bei dem in Figur 3 aufgezeigten Beispiel befindet sich einer der Sprecher (30) in Raumsektor 3 und einer Sprecher (31) in Raumsektor 1. Das System verfügt über N Sprachsignalkanäle (15) denen jeweils eine auf einen von N nebeneinanderliegenden Raumsektoren (21) spezifizierte Empfangscharakteristiken zugeordnet ist. Wird nun über einen gewissen Zeitraum, während dessen die Sprecher 30 und 31 reden, eine Statistik über die Qualität der Sprachsignale in den einzelnen Kanälen geführt, so ergibt sich in etwa die in Figur 3 aufgezeigte Statistik. In besonders vorteilhafte Weise

Petitioner Exhibit 1004, Page 905 of 1052

wurde die hier dargestellte statistische Auswertung anhand eines Histogramms erstellt, wobei andere statistische Auswerteverfahren gleichsam ebenfalls zur Anwendung kommen können. Bei einer statistischen Auswertung auf Basis eines Histogramms werden im Hinblick auf das in Figur 3 dargestellte Beispiel entlang der Ordinate des Diagramms unterschiedliche räumlich nebeneinander liegende Empfangscharakteristiken (Raumrichtungen) aufgelistet. Dabei wird abgebildet, wie häufig ein aus deiner dieser bestimmten Raumrichtungen (21) empfangenes Signal zu einem brauchbaren Erkennungsergebnis beigetragen hat. Auf diese Weise kann sodann auf unterschiedliche Sprecher (30, 31) geschlossen werden, wenn zwischen zwei im Histogramm in ihrer Häufigkeit deutlich hervortretenden Raumrichtungen wenigstens eine Raumrichtung gegeben ist, welche eine hierzu signifikant geringe Häufigkeit aufweist. Selbstverständlich kann durch den Fachmann dieses Verfahren vergleichbar auf eine Optimierungseinheit mit unterschiedlichsten Richtungsselektivitäten in der Empfangscharakteristik adaptiert werden.

[0030] Eine weitere gewinnbringende Verbesserung des erfindungsgemäßen Spracherkennungssystems lässt sich dadurch erzielen, dass für den Fall, dass auf das Vorhandensein mehrere Sprecher geschlossen wird, dem Spracherkennungssystem wenigstens ein weiterer Spracherkenner zugeschaltet wird. Die Zuschaltung erfolgt idealer Weise direkt hinter der Optimierungseinheit (10), so dass gewissermaßen eine Parallelschaltung der Spracherkenner entsteht. Auf diese Weise lässt sich die Verarbeitungsleistung weiter steigern und optimieren, insbesondere indem jedem einzelnen der Spracherkenner spezifisch diejenigen Sprachsignale zugeleitet werden, welche aus denjenigen Raumrichtungen stammen, welche den einzelnen der mehreren Sprecher zugeordnet sind.

Patentansprüche

1. Verfahren zur Regelung eines Spracherkennungssystems,
   bei welchem eine Mikrofonanordnung durch ein Optimierungseinheit (10) in ihrer Empfangscharakteristik gesteuert wird,
   wobei mittels der Mikrofonanordnung Sprachsignale empfangen und verarbeitet werden,
   und bei dem in Folge der Verarbeitung der Sprachsignale über die Optimierungseinheit (10) die Empfangscharakteristik der Mikrofonanordnung (12) so gesteuert wird, dass die Erkennungsleistung des der Optimierungseinheit(10) nachgeordneten Spracherkenners (11) optimiert wird,
   dadurch gekennzeichnet,
   dass dem Spracherkenner (11) empfangenen Sprachsignale über mehrere parallele oder quasiparallele Sprachkanäle (15) zugeführt werden,
   dass der Spracherkenner (11) diese Sprachkanäle zumindest zeitweise parallel verarbeitet und die Erkennungsergebnisse hinsichtlich ihrer Qualität miteinander vergleicht,
   und **dass** auf Grundlage dieses Vergleichs
   zum einen die Sprachsignale derjenigen Sprachkanäle (14) auswählt und weiterverarbeitet werden, welche das Potential für eine bestmögliche Erkennungsleistung haben,
   und zum anderen die Optimierungseinheit (10) durch den Spracherkenner die notwendigen Regelsignale (18) erhält, um die Empfangscharakteristik des Mikrofonarrays (12) in bezug auf diejenigen Sprachkanäle (14) zu optimieren, über welche diejenigen Sprachsignale empfangen werden, welche das Potential für eine bestmögliche Erkennungsleistung haben.

2. Verfahren zur Regelung eines Spracherkennungssystems nach Anspruch 1,
   dadurch gekennzeichnet,
   dass die Spracherkennung im Spracherkenner (11) im Rahmen eines mehrstufigen Prozesses erfolgt, welcher zumindest nacheinander folgend die Verarbeitungsstufen

   - Klassifikation (15), zur Erkennung von Lauteinheiten und/oder Lautuntereinheiten von Lauteinheiten,
   - Worterkennung (16), zur Erkennung gesprochener Worte,
   - und Interpretation, zur Erkennung und Interpretation sinnvoller Sätze (17) umfasst,

   und **dass** durch eine oder mehrere dieser Verarbeitungsstufen
   zum einen eine Auswahl von Sprachsignalen getroffen wird, welche das Potential für eine bestmögliche Erkennungsleistung haben, und welche einer Weiterverarbeitung zugeführt werden sollen,
   und zum anderen der Optimierungseinheit (10) notwendige Regelsignale (18) zur Verfügung gestellt werden, um die Empfangscharakteristik des Mikrofonarrays (12) in Bezug auf diejenigen Sprachkanäle (14) zu optimieren, aus welchen diejenigen Sprachsignale empfangen werden, welche das Potential für eine bestmögliche Erkennungsleistung haben.

3. Verfahren zur Regelung eines Spracherkennungssystems nach einem der vorhergehenden Ansprüche,
   dadurch gekennzeichnet,
   dass eine Statistik darüber geführt wird, für welche Empfangscharakteristiken des Mikrofonarrays optimale Erkennungsleistungen erzielt wurden,
   und das ausgehend von dieser Statistik entschie-

Petitioner Exhibit 1004, Page 906 of 1052

den wird, ob die von dem Spracherkennungssystem verarbeiteten Sprachsignale von unterschiedlichen Sprechern stammen.

4. Verfahren zur Regelung eines Spracherkennungssystems nach Anspruch 3,
dadurch gekennzeichnet,
dass die Statistik in Art eines Histogramms geführt wird, welches entlang seiner Ordinate unterschiedliche räumlich nebeneinander liegende Raumrichtungen auflistet, und abbildet, wie häufig ein aus einer dieser bestimmten Raumrichtungen empfangenes Signal zu einem brauchbaren Erkennungsergebnis beigetragen hat,
wobei dann auf unterschiedliche Sprecher entschieden wird, wenn zwischen zwei im Histogramm in ihrer Häufigkeit deutlich hervortretenden Raumrichtungen wenigstens eine Raumrichtung gegeben ist, welche eine hierzu signifikant geringe Häufigkeit aufweist.

5. Verfahren zur Regelung eines Spracherkennungssystem nach einem der Ansprüche 1 oder 4,
dadurch gekennzeichnet,
dass für den Fall, dass auf das Vorhandensein mehrerer Sprecher (30, 31) geschlossen wird, dem Spracherkennungssystem wenigstens ein weiterer Spracherkenner zugeschaltet wird,
welcher im wesentlichen die selben Funktionalitäten wie der bereits im Spracherkennungssystem enthaltene Spracherkenner aufweist,
und welchem von der Optimierungseinheit (10) diejenigen Sprachsignale zugeleitet werden, welche aus den Raumrichtungen stammen, welche wenigstens einem der mehreren Sprecher zugeordnet sind.

6. Vorrichtung zur Regelung eines Spracherkennungssystems,
bestehend aus einer Mikrofonanordnung mit einer Optimierungseinheit (10) zur Modifikation der Empfangscharakteristik der Mikrofonanordnung (12),
welche über wenigstens einen Spracherkenner (11) verfügt, welchem die Sprachsignale zugeführt werden und welcher diese verarbeitet,
wobei der Spracherkenner (11) mit der Optimierungseinheit (10) zur Modifikation Empfangscharakteristik der Mikrofonanordnung (12) in Verbindung steht, so dass über diese Verbindung die Empfangscharakteristik dergestalt geregelt werden kann, dass die Erkennungsleistung des der Optimierungseinheit (10) nachgeordneten Spracherkenners (11) optimiert wird,
dadurch gekennzeichnet,
dass dem Spracherkenner (11) die Sprachsignale über mehrere parallele Signalleitungen (14) zugeführt werden,
dass dem Spracherkenner (11) mit der Optimierungseinheit (10) dergestalt in Verbindung steht, dass ihm die aus Sprachsignale über Sprachkanäle (14) parallel oder quasi-parallel zugeführt werden,
dass der Spracherkenner (11) so ausgestaltet ist, dass er diese unterschiedlichen Sprachsignale parallel oder quasi-parallel verarbeiten und die Erkennungsergebnisse hinsichtlich ihrer Qualität miteinander vergleichen kann,
und **dass** auf Grundlage dieses Vergleichs
zum einen die Sprachsignale derjenigen Sprachkanäle auswählt und weiterverarbeitet werden, welche das Potential für eine bestmögliche Erkennungsleistung haben,
und zum anderen die Optimierungseinheit (10) durch den Spracherkenner die notwendigen Regelsignale (18) erhält, um die Empfangscharakteristik des Mikrofonarrays in Bezug auf diejenigen Sprachkanäle (14) hin zu optimieren, welche die Sprachsignale leiten, welche das Potential für eine bestmögliche Erkennungsleistung haben.

7. Vorrichtung zur Regelung eines Spracherkennungssystems nach Anspruch 6,
dadurch gekennzeichnet,
dass der Spracherkenner mehrstufig ausgestaltet ist, und zumindest enthält:

- einen Klassifikator (15), zur Erkennung von Lauteinheiten und/oder Lautuntereinheiten von Lauteinheiten,
- eine Einheit zur Erkennung gesprochener Worte in den empfangen Sprachsignalen (16), zur Erkennung gesprochener Worte,
- und eine Einheit zur Interpretation des empfangenen Sprachsignals (17), zur Erkennung und Interpretation sinnvoller Sätze umfasst,

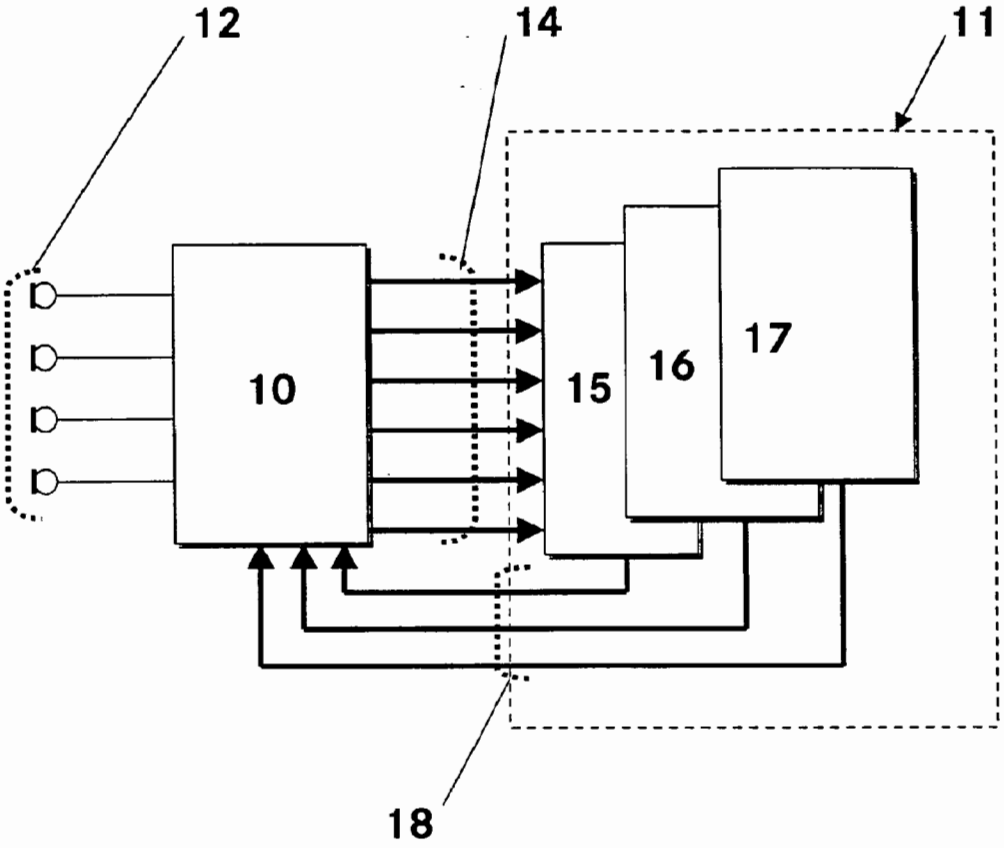
und **dass** durch eine oder mehrere dieser Verarbeitungsstufen
zum einen eine Auswahl von Sprachsignalen getroffen wird, welche das Potential für eine bestmögliche Erkennungsleistung haben, und welche einer Weiterverarbeitung zugeführt werden sollen,
und zum anderen der Optimierungseinheit (10) notwendige Regelsignale (18) zur Verfügung gestellt werden, um die Empfangscharakteristik des Mikrofonarrays (12) in Bezug auf diejenigen Sprachkanäle (14) hin zu optimieren, welche diejenigen Sprachsignale führen, welche das Potential für eine bestmögliche Erkennungsleistung haben.

8. Vorrichtung zur Regelung eines Spracherkennungssystem nach einem der Ansprüche 6 bis 7,
dadurch gekennzeichnet,
dass die Vorrichtung so ausgestaltet ist, dass für

Petitioner Exhibit 1004, Page 907 of 1052

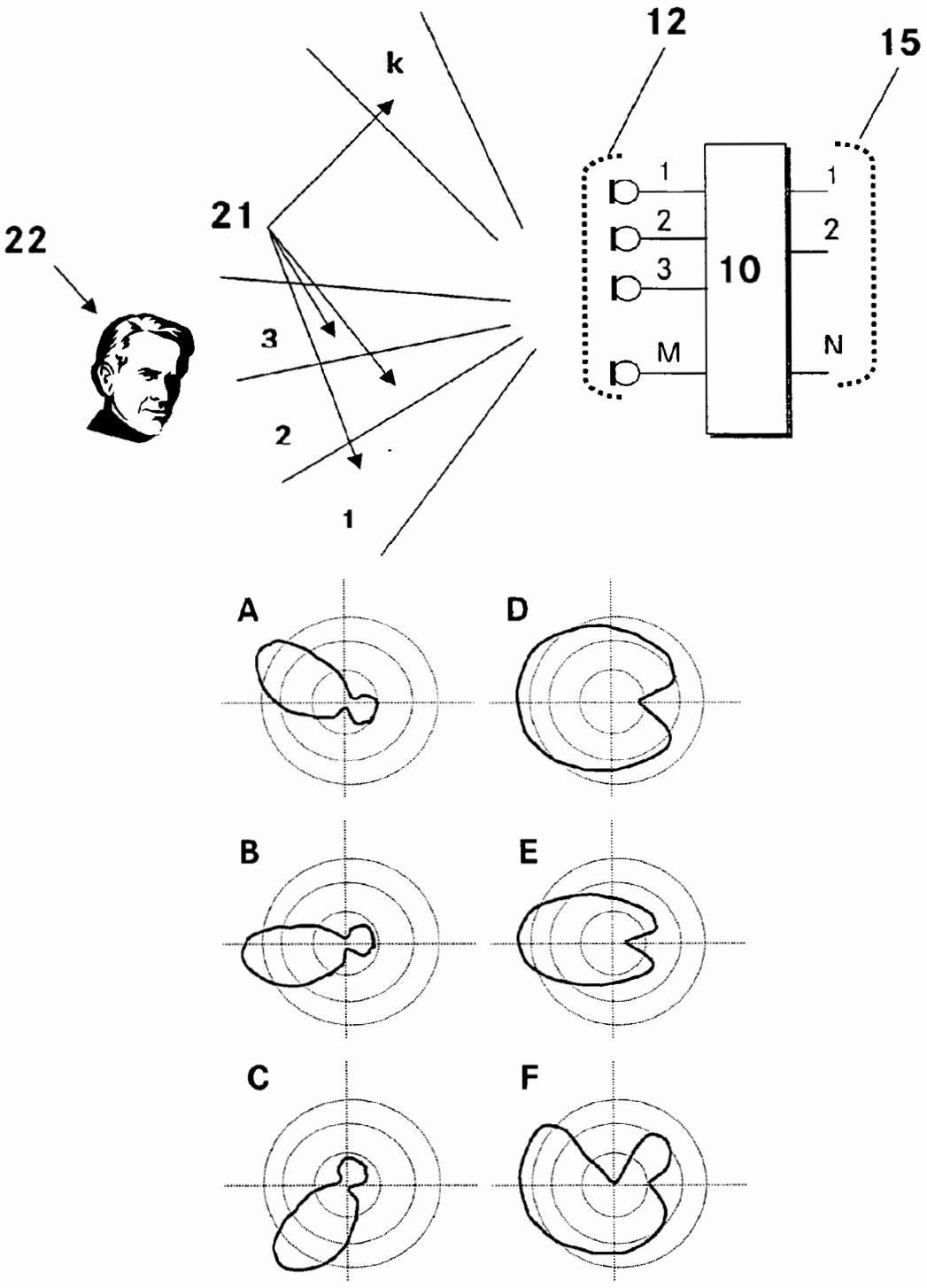
den Fall dass die empfangenen Sprachsignale mehreren Sprechern zuzuordnen sind wenigstens ein weiterer Spracherkenner zugeordnet werden kann, welcher im wesentlichen die selben Funktionalitäten wie der bereits im Spracherkennungssystem enthaltene Spracherkenner (11) aufweist,
und welcher mit der Optimierungseinheit (10) dergestalt in Verbindung steht, dass ihm diejenigen Sprachsignale zugeleitet werden, welche aus den Raumrichtungen stammen, welche wenigstens einem der mehreren Sprecher zugeordnet sind.

Petitioner Exhibit 1004, Page 908 of 1052



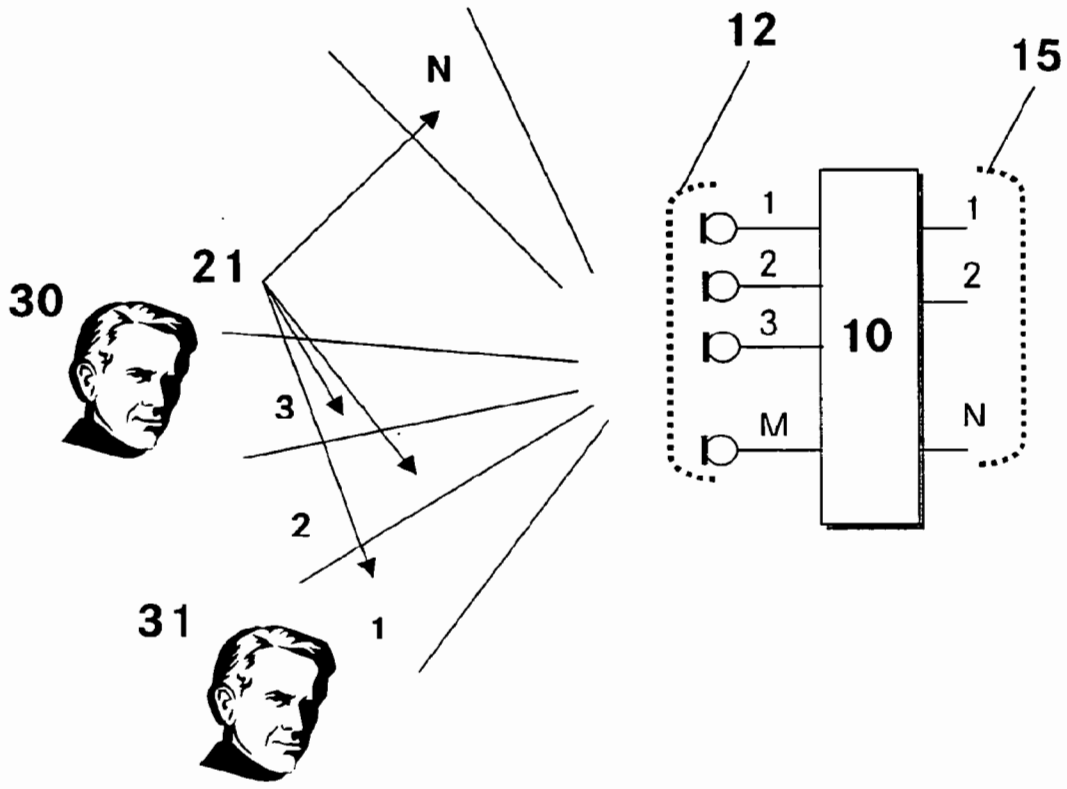
Figur 1

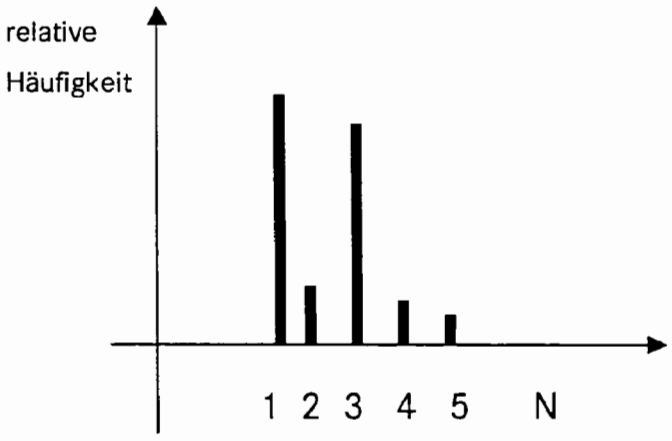
Petitioner Exhibit 1004, Page 909 of 1052



Figur 2

Petitioner Exhibit 1004, Page 910 of 1052





Figur 3

Petitioner Exhibit 1004, Page 911 of 1052

Europäisches Patentamt

EUROPÄISCHER RECHERCHENBERICHT

Nummer der Anmeldung
EP 04 01 3378

EINSCHLÄGIGE DOKUMENTE

Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Betrifft Anspruch	KLASSIFIKATION DER ANMELDUNG (Int.Cl.7)
A	SHIMIZU Y ET AL: "Speech recognition based on space diversity using distributed multi-microphone" IEEE ICASSP 2000, Bd. 3, 5. Juni 2000 (2000-06-05), Seiten 1747-1750, XP010507697 * Absatz [05.1] *	1,6	G10L15/20
A	SELTZER M L ET AL: "Speech recognizer-based microphone array processing for robust hands-free speech recognition" 2002 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS (CAT. NO.02CH37334) IEEE PISCATAWAY, NJ, USA, Bd. 1, 13. Mai 2002 (2002-05-13), - 17. Mai 2002 (2002-05-17) Seiten 897-900, XP002310104 ORLANDO FL ISBN: 0-7803-7402-9 * Abbildung 1b * * Zusammenfassung *	1,6	RECHERCHIERTE SACHGEBIETE (Int.Cl.7) G10L
D,A	EP 1 081 682 A (PIONEER CORP) 7. März 2001 (2001-03-07) * Absätze [0069] - [0071], [0080] * * Abbildung 10 *	1,6	
A	US 5 684 924 A (STANLEY BARBARA ANN ET AL) 4. November 1997 (1997-11-04) * Abbildungen 2,3 *	1,6	
A	US 6 122 613 A (BAKER JAMES K) 19. September 2000 (2000-09-19) * Zusammenfassung *		

Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt

Recherchenort	Abschlußdatum der Recherche	Prüfer
München	10. Dezember 2004	Krembel, L

KATEGORIE DER GENANNTEN DOKUMENTE

X : von besonderer Bedeutung allein betrachtet
Y : von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie
A : technologischer Hintergrund
O : nichtschriftliche Offenbarung
P : Zwischenliteratur

T : der Erfindung zugrunde liegende Theorien oder Grundsätze
E : älteres Patentdokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist
D : in der Anmeldung angeführtes Dokument
L : aus anderen Gründen angeführtes Dokument

& : Mitglied der gleichen Patentfamilie, übereinstimmendes Dokument

EPO FORM 1503 03.82 (P04C03)

Petitioner Exhibit 1004, Page 912 of 1052

ANHANG ZUM EUROPÄISCHEN RECHERCHENBERICHT ÜBER DIE EUROPÄISCHE PATENTANMELDUNG NR. EP 04 01 3378

In diesem Anhang sind die Mitglieder der Patentfamilien der im obengenannten europäischen Recherchenbericht angeführten Patentdokumente angegeben.
Die Angaben über die Familienmitglieder entsprechen dem Stand der Datei des Europäischen Patentamts am
Diese Angaben dienen nur zur Unterrichtung und erfolgen ohne Gewähr.

10-12-2004

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 1081682 A	07-03-2001	JP 2001075594 A	23-03-2001
		EP 1081682 A2	07-03-2001
US 5684924 A	04-11-1997	KEINE	
US 6122613 A	19-09-2000	DE 69814589 D1	18-06-2003
		DE 69814589 T2	25-03-2004
		EP 0954848 A1	10-11-1999
		WO 9834217 A1	06-08-1998

EPO FORM P0461

Für nähere Einzelheiten zu diesem Anhang : siehe Amtsblatt des Europäischen Patentamts, Nr.12/82

Petitioner Exhibit 1004, Page 913 of 1052

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		2002-0193989	A1	Geilhufe et al.	12-19-2002
		2008-0118080	A1	Gratke et al.	05-22-2008
		2009-0055178	A1	Coon	02-26-2009
		2012-0197637	A1	Gratke et al.	08-02-2012
		2012-0310646	A1	Hu et al.	12-06-2012
		2015-0106085	A1	Lindahl	04-16-2015
		2016-0039356	A1	Talwar et al.	02-11-2016
		2016-0314782	A1	Klimanis	10-27-2016
		2017-0116983	A1	Furukawa et al.	04-27-2017

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		EP	1511010	A1	Daimler Chrysler Ag	03-02-2005	Y-Abstract Only

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		International Preliminary Report on Patentability for International Application No. PCT/US2016/017317 mailed August 23, 2018 (N0484.71176WO00).	
		International Search Report and Written Opinion for International Application No. PCT/US2016/017317 mailed May 12, 2016 (N0484.71176WO00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8207852.1

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	David Scott Gesner/Patricia McDonald
Attorney Docket Number:	N0484.71090US04

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 1 month with $0 paid	1251	1	200	200
Miscellaneous:				
RCE- 2ND AND SUBSEQUENT REQUEST	1820	1	1900	1900
		Total in USD ($)		**2100**

REMARKS

In response to the Office Action dated March 13, 2020, reconsideration is respectfully requested. To further the prosecution of this application, amendments have been made in the claims, and each of the rejections set forth in the Office Action has been carefully considered and is addressed below. The claims as presented are believed to be in condition for allowance.

Claims 1-21 were previously pending in this application. Claims 1, 11, 15, 17, 18, 20 and 21 are amended. Claims 10 and 16 are cancelled. No claims are added. As a result, claims 1-9, 11-15 and 17-21 are pending for examination, with claims 1, 15 and 21 being independent. No new matter has been added.

Claim Rejections under 35 U.S.C. §103

Each of independent claims 1, 15, and 21 stands rejected under 35 U.S.C. §103(a) as purportedly being obvious over U.S. Patent No. 8,977,255 to Freeman et al. (hereinafter "Freeman") in view of U.S. Patent Application Publication No. 2013/0080167 to Mozer (hereinafter "Mozer"). The rejection of each of independent claims 1, 15 and 21 on this basis is respectfully traversed.

A. Brief Overview Of Some Embodiments

Some embodiments described in the Specification relate to techniques whereby a device determines, while in a low power mode, whether an acoustic input includes a voice command (Published Specification, ¶[0035]). In some techniques, a two-stage process is used by a device in a low power mode to determine whether an acoustic input includes a voice command (¶¶[0087]-[0091]). First, the device performs computationally-inexpensive automatic speech recognition (ASR) on the acoustic input to see whether any recognized words or phrases match a trigger word or phrase (¶[0087]). For example, the device may determine whether any recognized words or phrases match the trigger phrase "Hello Dragon" (¶[0087]).

After it is determined that the acoustic input includes a trigger word phrase, the device, while still in the low power mode, invokes more comprehensive ASR to determine whether the

Petitioner Exhibit 1004, Page 917 of 1052

acoustic input includes a voice command (¶[0090]). This may be performed in any of various ways. In some embodiments, the device, while still in the low power mode, may transmit the acoustic input to a server so that the server can perform more computationally-expensive ASR to determine whether the acoustic input includes a voice command (¶[0091]). If the server determines that a voice command is included in the acoustic input, the server may notify the device, so that the device may respond to the voice command recognized by the server (¶0070).

The foregoing overview is provided to assist the Examiner in appreciating some aspects of certain embodiments. However, this overview may not apply to each of the independent claims, and the language of each independent claim may differ in material respects from the foregoing overview. The Examiner is thus respectfully requested to give careful consideration to the language of each independent claim, and to address each on its own merits, without relying on the overview above. In this respect, the Assignee does not rely on the overview to distinguish any of the claims over the cited art, but rather relies only upon the language of the claims themselves and the arguments presented in the sections that follow.

B. Rejection Of Independent Claim 1

Amended claim 1 recites a device configured to operate in a lower power mode and a higher power mode. The device comprises: at least one processor configured to perform at least one processing stage on an acoustic input received from an environment of the device while the device is operating in the lower power mode. Performing the at least one processing stage comprises: determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, wherein the determining comprises: performing automatic speech recognition (ASR), while in the lower power mode, to recognize at least one word or phrase in the acoustic input; determining, while in the lower power mode, whether the recognized at least one word or phrase matches the specific word or phrase; in response to determining that the acoustic input includes the specific word or phrase, requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine

Petitioner Exhibit 1004, Page 918 of 1052

whether the acoustic input includes at least one voice command; in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode; and responding to the at least one voice command, in either the lower power mode or the higher power mode, depending on the selecting.

Amended claim 1 patentably distinguishes over any combination of the asserted references for at least two reasons. First, Freeman and Mozer do not disclose or suggest, either individually in combination, "determining in [a] lower power mode whether [ ] acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of [a] device is to be engaged for provision of subsequent input," and then separately "determin[ing] whether the [same] acoustic input includes at least one voice command," as claim 1 recites. Second, Freeman and Mozer do not disclose or suggest, either individually or in combination, a device which, "in response to receiving an indication [ ] that [an] acoustic input includes at least one voice command, select[s] whether to remain in [a] lower power mode or to transition to a higher power mode" and "respond[s] to the at least one voice command in either the lower power mode or the higher power mode, depending on the selecting," as claim 1 recites. These distinctions are discussed in turn below.

1. The Asserted References Do Not Disclose Or Suggest Determining That Acoustic Input Includes A Word Or Phrase Indicating That A Voice Interface Is To Be Engaged For Provision Of Subsequent Input, And Then Separately Determining Whether The Same Acoustic Input Includes At Least One Voice Command

In the "Response To Arguments" section, the Office Action contends that Freeman and Mozer both disclose a process whereby a device first determines in a lower power mode whether acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, and then separately determines whether the same acoustic input includes at least one voice command. This contention is unsupported, by either of the references.

Freeman does not disclose or suggest determining that an acoustic input includes a specific word or phrase, ***and then separately*** determining whether the same acoustic input includes at least

Petitioner Exhibit 1004, Page 919 of 1052

one voice command. In fact, Freeman explicitly states that a user need not supply a specific word or phrase to indicate that a voice interface is to be engaged before subsequently supplying a voice command. In this respect, Freeman touts that voice commands supplied by a user are processed directly "without any preparatory user action with respect to the portable electronic device" (Abstract). In portions of Freeman which are cited by the Office Action, Freeman discloses that a user's acoustic input is analyzed to determine whether it includes a voice command, and if so, the command is automatically and directly processed (see, e.g., steps 404, 408, 410 of Fig. 4, steps 508, 510, 514, 516 of FIG. 5; col. 6, line 15 – col. 7, line 10). Freeman says nothing about first determining whether an acoustic input includes a specific word or phrase that indicates that a voice interface is to be engaged for provision of subsequent input, ***and then separately*** determining whether the same acoustic input includes at least one voice command, as claim 1 recites.

Mozer also does not satisfy these limitations of claim 1. Like Freeman, Mozer expressly touts that the disclosed system "does not need to be triggered (e.g., by [ ] a trigger phrase) to begin speech recognition" (¶¶[0032]-[0033]). In Mozer's system, a stage 1 recognizer passively listens to a conversation to determine whether it includes keywords (¶¶[0032]-[0033]). The keywords have each been assigned to a classification (¶¶[0032]-[0033]). If enough keywords in a particular classification are detected, the stage 1 recognizer initiates a stage 2 recognizer, and provides the identified classification to it (¶¶[0033], [0067]). The stage 2 recognizer then commences listening to the conversation, deciphers words which have been spoken after it was initiated, and retrieves information relating to the conversation (¶[0067]). For example, if the stage 2 recognizer determines that the part of the conversation which took place after it commenced listening relates to potential activities in Ireland, then it may determine that the participants are discussing a trip to Ireland, and retrieve pictures of Ireland, or a coupon for a soccer game (¶0067).

Because Mozer's stage 2 recognizer only begins listening to a conversation after being initiated by the stage 1 recognizer, the stage 1 recognizer and stage 2 recognizer separately process ***different acoustic inputs***. Thus, even if one assumed for argument's sake that a keyword detected by the stage 1 recognizer could reasonably be interpreted as "a specific word or phrase indicating that a voice interface is to be engaged for provision of subsequent input," and that the conversation deciphered by the stage 2 recognizer could reasonably be interpreted as a "voice command" (which

Petitioner Exhibit 1004, Page 920 of 1052

the Assignee does not concede), Mozer does not disclose or suggest determining whether the "specific word or phrase" and "at least one voice command" are included ***in the same acoustic input***, as claim 1 recites, since in Mozer's system the "specific word or phrase" and the conversation reside in different acoustic inputs.

As a result, Freeman and Mozer do not satisfy, either individually or in combination, the claim 1 limitations directed to determining whether an acoustic input includes a specific word or phrase that indicates that a voice interface is to be engaged for provision of subsequent input, and then separately determining whether the same acoustic input includes at least one voice command. Claim 1 patentably distinguishes over any combination of the asserted references for at least this reason.

2. The Asserted References Do Not Disclose Or Suggest A Device Which, In Response To Receiving An Indication That An Acoustic Input Includes At Least One Voice Command, Selects Whether To Remain In A Lower Power Mode Or Transition To A Higher Power Mode, And Responds To The At Least One Voice Command In Either The Lower Power Mode Or The Higher Power Mode, Depending On The Selecting

Claim 1 also patentably distinguishes over any combination of the asserted references because Freeman and Mozer do not disclose or suggest, either individually or in combination, a device which, "in response to receiving an indication [ ] that [an] acoustic input includes at least one voice command, select[s] whether to remain in [a] lower power mode or to transition to a higher power mode, and respond[s] to the at least one voice command in either the lower power mode or the higher power mode, depending on the selecting," as claim 1 recites.

Freeman says nothing whatsoever about a device selecting whether to remain in a lower power mode or transition to a higher power mode in response to receiving an indication that an acoustic input includes at least one voice command. In fact, in a portion of Freeman which is cited by the Office Action, Freeman explicitly discloses automatically transitioning to a higher power mode upon determining that an acoustic input includes a voice command (FIG. 5, steps 510, 512).

Mozer does not remedy this deficiency of Freeman, as Mozer says nothing whatsoever about a device which, "in response to receiving an indication [ ] that [an] acoustic input includes at least

Petitioner Exhibit 1004, Page 921 of 1052

one voice command, select[s] whether to remain in [a] lower power mode or to transition to a higher power mode, and respond[s] to the at least one voice command in either the lower power mode or the higher power mode, depending on the selecting," as claim 1 recites.

3. Conclusion Regarding Claim 1

For each of the foregoing reasons, claim 1 patentably distinguishes over any combination of the asserted references. Withdrawal of the rejection of claim 1 as purportedly being obvious over Freeman in view of Mozer is therefore respectfully requested.

Claims 2-9 and 11-14 depend from claim 1, and are allowable for at least the same reasons.

B. Rejection Of Independent Claim 15

Claim 15 recites at least one non-transitory computer readable medium storing instructions that, when executed by at least one processor, cause the at least one processor to perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode. The method comprises: performing at least one processing stage on the acoustic input, the at least one processing stage comprising: performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input; determining, while in the low power mode, whether the recognized at least one word or phrase matches a trigger word or phrase; in response to determining that the at least one first word or phrase matches the trigger word or phrase, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands; in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode, and responding to the at least one voice command recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.

Petitioner Exhibit 1004, Page 922 of 1052

It should be appreciated from the discussion above regarding claim 1 that Freeman and Mozer do not satisfy, either individually or in combination, the claim 15 limitations directed to determining whether a recognized at least one word or phrase matches a trigger word or phrase, and then separately determining whether the same acoustic input includes one or more voice commands. It should also be appreciated from the discussion above regarding claim 1 that Freeman and Mozer do not satisfy, either individually or in combination, the claim 15 limitations directed to selecting, in response to receiving an indication that an acoustic input includes at least one voice command, whether to remain in a lower power mode or transition to a higher power mode, and responding to the at least one recognized voice command, in either the lower power mode or the higher power mode, depending on the selecting.

Accordingly, claim 15 patentably distinguishes over any combination of Freeman and Mozer. Withdrawal of the rejection of claim 15 as purportedly being obvious over Freeman in view of Mozer is respectfully requested.

Claims 17-20 depend from claim 15, and are allowable for at least the same reasons.

C. Rejection Of Independent Claim 21

Claim 21 recites a device comprising at least one processor, configured to process acoustic input from an environment of the device while the device is operating in a low power mode. The processing comprises: performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input; determining, while in the low power mode, whether the recognized at least one word or phrase matches a trigger word or phrase; in response to determining that the at least one first word or phrase matches the trigger word or phrase, requesting that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands; and in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode, and performing at least one task to carry out the at least one voice command in

Petitioner Exhibit 1004, Page 923 of 1052

the acoustic input recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.

It should be appreciated from the discussion above regarding claim 1 that Freeman and Mozer do not satisfy, either individually or in combination, the claim 21 limitations directed to determining whether a recognized at least one word or phrase matches a trigger word or phrase, and then separately determining whether the same acoustic input includes one or more voice commands. It should also be appreciated from the discussion above that Freeman and Mozer do not satisfy, either individually or in combination, the claim 21 limitations directed to selecting, in response to receiving an indication that an acoustic input includes at least one voice command, whether to remain in a lower power mode or transition to a higher power mode, and performing at least one task to carry out the at least one voice command in either the lower power mode or the higher power mode, depending on the selecting.

Accordingly, withdrawal of the rejection of claim 21 as purportedly being obvious over Freeman in view of Mozer is therefore respectfully requested.

General Comments on Dependent Claims

As each dependent claim depends from a base claim that is believed to be in condition for allowance, the Assignee believes that arguing the further distinguishing features of all of the dependent claims is unnecessary at this time. However, the Assignee does not necessarily concur with the interpretation of any of the dependent claims set forth in the Office Action, and does not necessarily concur that the basis for rejecting any dependent claim is proper. Therefore, the Assignee reserves the right to address in the future the patentability of any dependent claim not specifically addressed herein.

Petitioner Exhibit 1004, Page 924 of 1052

CONCLUSION

In view of the foregoing, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after this response, the application is not in condition for allowance, the Examiner is requested to call the undersigned at the telephone number indicated below to discuss any outstanding issues.

Dated: July 13, 2020

Respectfully submitted,

Electronic signature: /David S. Gesner/
David S. Gesner
Registration No.: 72,100
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Petitioner Exhibit 1004, Page 925 of 1052

Request for Continued Examination (RCE) Transmittal		
Address to: Mail Stop RCE Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450	Application Number	15/238,238-Conf. #5864
	Filing Date	August 16, 2016
	First Named Inventor	William F. Ganong
	Art Unit	2644
	Examiner Name	H. C. Ho
	Attorney Docket Number	N0484.71090US04

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.

Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8, 1995, to any international application that does not comply with the requirements of 35 U.S.C 371, or to any design application. See Instruction Sheet for RCEs (not to be submitted to the USPTO on page 2.)

1. Submission required under 37 CFR 1.114 Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the order in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s) entered, applicant must request non-entry of such amendment(s).
   - a. [ ] Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.
     - i. [ ] Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
     - ii. [ ] Other _____
   - b. [x] Enclosed
     - i. [x] Amendment/Reply
     - ii. [ ] Affidavit(s)/ Declaration(s)
     - iii. [ ] Information Disclosure Statement (IDS)
     - iv. [x] Other Extension of Time
2. Miscellaneous
   - a. [ ] Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of _____ months. (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)
   - b. [ ] Other _____
3. Fees The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.
   - a. [ ] The Director is hereby authorized to charge the following fees, any underpayment of fees, or credit any overpayments, to Deposit Account No. 23/2825 .
     - i. [x] RCE fee required under 37 CFR 1.17(e)
     - ii. [x] Extension of time fee (37 CFR 1.136 and 1.17)
     - iii. [ ] Other _____
   - b. [ ] Check in the amount of $ _____ enclosed
   - c. [x] Payment by credit card (~~Form PTO-2038 enclosed~~)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED			
Signature	/David S. Gesner/	Date	July 13, 2020
Name (Print/Type)	David S. Gesner	Registration No.	72,100

~~***CERTIFICATE OF MAILING OR TRANSMISSION***~~			
~~I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop RCE, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date shown below.~~			
~~Signature~~			
~~Name (Print/Type)~~		~~Date~~	

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 13, 2020 Electronic Signature for : /Trish McDonald/

8195775.1

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 13, 2020 Electronic Signature for : /Trish McDonald/

RESPONSE FILED WITH REQUEST FOR CONTINUED EXAMINATION

Mail Stop RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

In response to the Office Action dated March 13, 2020, please amend the above-identified U.S. patent application as follows:

Claim Amendments are reflected in the claim listing which begins on page 2 of this paper.

Remarks/Arguments begin on page 8 of this paper.

DOCKET NO.: N0484.71090US04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: July 13, 2020 Electronic Signature for : /Trish McDonald/

MAIL STOP RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed before the mailing of a first Office Action after the filing of a Request for Continued Examination under 37 C.F.R. § 1.114.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending application that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/074,953	08-02-2018	Prémont et al.	N0484.71176US00

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR .

The above-identified co-pending application may include subject matter that is technically-related to subject matter of the present application and may contain claims that recite elements similar to those recited in claims of the present application. In handling prosecution of the above-identified application, the Examiner for this other application presumably will search for relevant prior art and, in some circumstances, may take positions about the prior art and/or the interpretation of the claims that may be material to the examination of the present application.

Accordingly, the Examiner in the present application is requested to review the file history of the above-identified co-pending application to see whether there is any information that the Examiner believes may be relevant to the prosecution of the present application. It is assumed that the Examiner has access to the file history of the above-identified application. However, if the Examiner would like copies of any or all of the information included in the file history of the above-identified application, the Examiner is asked to contact the undersigned representative.

The undersigned would like to bring to the Examiner's attention the enclosed search reports or other communications from corresponding or related International or Foreign National Applications:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
PCT/US2016/017317	**08-23-2018**	**International Preliminary Report on Patentability**	**N0484.71176WO00**
PCT/US2016/017317	**05-12-2016**	**International Search Report and Written Opinion**	**N0484.71176WO00**

Petitioner Exhibit 1004, Page 929 of 1052

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

Petitioner Exhibit 1004, Page 930 of 1052

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US04 from which the undersigned is authorized to draw.

Respectfully submitted,

By:_/David S. Gesner/_____
David S. Gesner
Registration No.: 72,100
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: N0484.71090US04
Date: July 13, 2020

CLAIMS

1. (Currently amended) A device configured to operate in a lower power mode and a higher power mode, the device comprising:

~~at least one input configured to receive acoustic input from an environment of the device while the device is operating in the lower power mode; and~~

at least one processor configured to perform at least one processing stage on [[the]] an acoustic input received from an environment of the device while the device is operating in the lower power mode, performing the at least one processing stage ~~including~~ comprising:

determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, wherein the determining comprises:

performing automatic speech recognition (ASR), while in the lower power mode, to recognize at least one word or phrase in the acoustic input;

determining, while in the lower power mode, whether ~~at least one first word or phrase of~~ the recognized at least one word or phrase matches the specific word or phrase;

in response to determining that the acoustic input includes the specific word or phrase, requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command;

in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode; and

responding to the at least one voice command, in either the lower power mode or the higher power mode, depending on the selecting. ~~, wherein the device is configured to respond to voice commands recognized by the at least one server.~~

2. (Previously presented) The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.

3. (Previously presented) The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

4. (Previously presented) The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

5. (Previously presented) The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

6. (Previously presented) The device of claim 4, wherein the first processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.

7. (Previously presented) The device of claim 4, wherein the second processor is configured to perform the ASR on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

Petitioner Exhibit 1004, Page 933 of 1052

8. (Previously presented) The device of claim 7, wherein the second processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.

9. (Previously presented) The device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.

10. (Cancelled)

11. (Currently amended) The device of claim 1, wherein at least one task specified by the voice command is performed with the device ~~remaining~~ selecting to remain in the lower power mode. ~~, and wherein the specific word or phrase, when detected, is sufficient to engage the speech interface of the device.~~

12. (Previously presented) The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

13. (Previously presented) The device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

14. (Previously presented) The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

15. (Currently amended) At least one non-transitory computer readable medium storing instructions that, when executed [[on]] by at least one processor, cause the at least one processor to perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:

performing at least one processing stage on the acoustic input, the at least one processing stage ~~including~~ comprising:

performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input;

determining, while in the low power mode, whether ~~at least one first word or phrase of~~ the recognized at least one word or phrase matches a trigger word or phrase;

in response to determining that the recognized at least one ~~first~~ word or phrase matches the trigger word or phrase, [[:]] requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands;

in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode, and

responding to the at least one voice command recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.

Petitioner Exhibit 1004, Page 935 of 1052

16. (Cancelled)

17. (Currently amended) The at least one non-transitory computer readable medium of claim 15, wherein the method ~~further comprising~~ comprises performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the at least one first word or phrase matches the trigger word or phrase.

18. (Currently amended) The at least one non-transitory computer readable medium of claim 15, wherein the method ~~further comprising~~ comprises providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

19. (Previously presented) The at least one non-transitory computer readable medium of claim 15, wherein responding to the at least one voice command comprises initiating performance of at least one task specified by the voice command with the device remaining in the low power mode.

20. (Currently amended) The at least one non-transitory computer readable medium of claim 15, wherein the method further ~~comprising~~ comprises evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

21. (Currently amended) A device comprising: ~~at least one input configured to receive acoustic input from an environment of the device while the device is operating in a low power mode; and~~

at least one processor, configured to process acoustic input from an environment of the device while the device is operating in a low power mode, the processing comprising:

Petitioner Exhibit 1004, Page 936 of 1052

~~perform~~ performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input;

~~determine~~ determining, while in the low power mode, whether ~~at least one first word or phrase of~~ the recognized at least one word or phrase matches a trigger word or phrase;

in response to determining that the recognized at least one ~~first~~ word or phrase matches the trigger word or phrase, ~~request~~ requesting that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands; [[and]]

in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode; and

performing ~~perform~~ at least one task to carry out the at least one voice command in the acoustic input recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.

Petitioner Exhibit 1004, Page 937 of 1052

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 15/238,238	Filing Date 08/16/2016	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $80 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $420 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	07/13/2020	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 19	Minus	** 20	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 3	Minus	*** 3	= 0		x $460 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/MARQUITA D JONES/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

91337 7590 09/02/2020

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
HO, HUY C

ART UNIT	PAPER NUMBER
2644	

DATE MAILED: 09/02/2020

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

TITLE OF INVENTION: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1000	$0.00	$0.00	$1000	12/02/2020

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Maintenance fees are due in utility patents issuing on applications filed on or after Dec. 12, 1980. It is patentee's responsibility to ensure timely payment of maintenance fees when due. More information is available at www.uspto.gov/PatentMaintenanceFees.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

91337 7590 09/02/2020

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

TITLE OF INVENTION: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1000	$0.00	$0.00	$1000	12/02/2020

EXAMINER	ART UNIT	CLASS-SUBCLASS
HO, HUY C	2644	455-556100

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

❑ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. Fees submitted: ❑ Issue Fee ❑ Publication Fee (if required) ❑ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

❑ Electronic Payment via EFS-Web ❑ Enclosed check ❑ Non-electronic payment by credit card (Attach form PTO-2038)

❑ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. _____

5. **Change in Entity Status** (from status indicated above)

❑ Applicant certifying micro entity status. See 37 CFR 1.29

❑ Applicant asserting small entity status. See 37 CFR 1.27

❑ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

Petitioner Exhibit 1004, Page 940 of 1052

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

91337 7590 09/02/2020

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	

ART UNIT	PAPER NUMBER
2644	

DATE MAILED: 09/02/2020

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 15/238,238	Applicant(s) Ganong et al.	
	Examiner HUY C HO	Art Unit 2644	AIA (FITF) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to 07/13/2020.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 1-9,11-15 and 17-21 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information , please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☑ Interview Summary (PTO-413), Paper No./Mail Date. 20200825.
5. ☐ Examiner's Amendment/Comment
6. ☑ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/HUY C HO/
Primary Examiner, Art Unit 2644

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Continued Examination Under 37 CFR 1.114

2. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on **07/13/2020** has been entered.

Response to Arguments

3. Applicant's arguments filed **07/13/2020** have been fully considered and are persuasive. The rejection of claims 1-9, 11-15 and 17-21 has been withdrawn.

Allowable Subject Matter

4. Claims 1-9, 11-15 and 17-21 are allowed.

5. The following is an examiner’s statement of reasons for allowance:

The arguments made in the Applicant Arguments/Remarks on **07/13/2020** have been fully considered and are persuasive. Further search has been updated. However, no reference is found alone or in any combination with the prior art of record to provide the teachings, suggestions or disclosures of the unique claimed limitations and features recited from the independent claims 1, 15 and 21 (and thus the dependent claims 2-9, 11-14 and 17-20) as filed on **07/13/2020.**

Therefore, claims 1-9, 11-15 and 17-21 are deemed allowed.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUY C HO whose telephone number is (571)270-1108. The examiner can normally be reached on M-F 8AM-5PM.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is

encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, KATHY WANG-HURST can be reached on (571)270-5371. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/HUY C HO/
Primary Examiner, Art Unit 2644

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

CPC - Searched*		
Symbol	Date	Examiner
G10L15/20; G10L15/22; G10L15/24; G10L15/25; G10L2015/223; G10L2015/226; H04M2250/74; H04W52/0251; H04W52/0254.	4/20/2017	HH
G10L15/20|G10L15/22|G10L15/24|G10L15/25|G10L2015/223| G10L2015/226|H04M2250/74|H04W52/0251|H04W52/0254|H04W52 /0254|H04W52/0261|H04W52/0251|H04M1/72519G10L15/02; H4W40/005;G1L15/30; G10L15/22; G10L2015/226; H04W88/02; Y02D70/00; Y02D30/70; Y02D70/164; Y02D70/144; H04M2250/74; G10L2015/223; Y02D70/142	08/27/2020	HH

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner
379	88.01	4/20/2017	HH

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

/HUY C HO/ Primary Examiner.Art Unit 2644	

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	Art Unit
	HUY C HO	2644

Search Notes		
Search Notes	**Date**	**Examiner**
EAST Database Search	4/20/2017	HH
Inventor Search	4/20/2017	HH
Search update	1/10/2018	HH
Search update	8/20/2018	HH
Search update	02/24/2019	HH
Search update	08/13/2019	HH
Search update	02/18/2020	HH
Search update	03/04/2020	HH
Search update	08/27/2020	HH

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**
379	88.02	4/20/2017	HH
704	246-255	4/20/2017	HH

/HUY C HO/ Primary Examiner.Art Unit 2644	

PATENT COOPERATION TREATY

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING TRANSMITTAL OF COPY OF INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY (CHAPTER I OF THE PATENT COOPERATION TREATY)
(PCT Rule 44bis.1(c))

To:

CCPIT PATENT AND TRADEMARK LAW OFFICE
8th Floor, Vantone New World Plaza
2 Fuchengmenwai Street, Xicheng District
Beijing 100037
CHINE

Date of mailing *(day/month/year)* 23 May 2019 (23.05.2019)	
Applicant's or agent's file reference IEE160077PCT	**IMPORTANT NOTICE**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/CN2016/105343	10 November 2016 (10.11.2016)	

Applicant
NUANCE COMMUNICATIONS, INC.

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty)

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Xiaofan Tang e-mail: pct.team2@wipo.int

Form PCT/IB/326 (January 2004) /HUY C HO/ 03/09/2020

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference IEE160077PCT	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/CN2016/105343	International filing date *(day/month/year)* 10 November 2016 (10.11.2016)	Priority date *(day/month/year)*
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant NUANCE COMMUNICATIONS, INC.		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☐	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 14 May 2019 (14.05.2019)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Xiaofan Tang e-mail: pct.team2@wipo.int

Form PCT/IB/373 (January 2004)

03/09/2020

/HUY C HO/

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
100037
China 8th Floor, Vantone New World Plaza, 2
Fuchengmenwai Street, Xicheng District, Beijing

CCPIT PATENT AND TRADEMARK LAW OFFICE

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	**04 August 2017**
Applicant's or agent's file reference **IEE160077PCT**	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/CN2016/105343	**10 November 2016**	

International Patent Classification (IPC) or both national classification and IPC
G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G06F 1/32(2006.01)i

Applicant
NUANCE SOFTWARE TECHNOLOGY (BEIJING) CO., LTD.

1. This opinion contains indications relating to the following items:

- [x] Box No. I Basis of the opinion
- [ ] Box No. II Priority
- [ ] Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- [ ] Box No. IV Lack of unity of invention
- [x] Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- [ ] Box No. VI Certain documents cited
- [ ] Box No. VII Certain defects in the international application
- [ ] Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/	Date of completion of this opinion	Authorized officer
STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA **China 6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088**	**31 July 2017**	**KANG,Dandan**
Facsimile No. **(86—10) 62019451**		Telephone No. **(86-10)61648170**

Form PCT/ISA/237 (Cover sheet) (July 2011)

/HUY C HO/ 03/09/2020

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. [ ] This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing filed or furnished:
   a. (means)
      - [ ] on paper
      - [ ] in electronic form
   b. (time)
      - [ ] in the international application as filed
      - [ ] together with the international application in electronic form
      - [ ] subsequently to this Authority for the purposes of search

4. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Form PCT/ISA/237 (Box No. I) (July 2011)

03/09/2020

/HUY C HO/

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-20	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-20	NO
Industrial applicability (IA)	Claims	1-20	YES
	Claims	None	NO

2. Citations and explanations :

[1] D1: US 2016189706 A1

[2] D2: CN 101650943 A

[3] D1 discloses a wake up phrase detection method and an electronic device(see description paragraphs [0030]-[0032], [0049]-[0070],[0086]-[0092], figures 2-5,8): electronic device may comprise a smart phone; FIG.4 depicts a training system 400, a user is prompted to utter their custom wake-up phrase(WUP), the utterances are captured as user training sequences 402(corresponding to "receiving acoustic input of a user speaking a wake-up word"); user training sequences 402 is provided to VAD 404, then to feature extraction component 406 to derive the feature vectors 408(corresponding to "acoustic features derived from the acoustic input"); the extracted features 408 are decoded by concatenated phoneme ASR engine 410 using phoneme model DB 414 which is implemented in a memory or storage medium and store phoneme models 206, and produce a phoneme transcription of the WUP, a Hidden Markov Model(HMM) is trained offline to generate a phoneme model 206(corresponding to "acoustic model"); based on the results produced by ASR 410, phoneme model concatenation component 416 determines the best phonemes models and concatenates them to produce an initial WUP model consist of the concatenated phoneme models(corresponding to "a sequence of speech units corresponding to the wake-up word spoken by the user" and "a first sequence of speech units" in claim 9 or "a second sequence of speech units" in claim 16); the WUP model used for WUP detector is stored in WUP model DB 418 which is implemented in a memory or storage medium ; FIG.5 depicts a user-specific wake up phrase detection system 500, a low-power processor is included and operates in a stand-by mode of the electronic device, receiving a wake-up phrase 514 as an input by a user(corresponding to "receiving acoustic input from the user speaking while the user device is operating in a low-power mode"), WUP 514 is provided to VAD 504 and then to feature extractor and WUP detector(WUPD) component 506 to detect the presence of the WUP in the captured audio by comparing the feature extracted to the WUP model stored in a non-volatile RAM; the presence of the WUP triggers the device to wake up, or come out of a stand-by mode or a sleep mode into a normal mode of operation(corresponding to "exiting the low-power mode if it is determined that the user spoke the wake-up word").

[4] D2 discloses a non-native speech recognition system and method (see description page 5 line 3 – page 6 line 9 , figures 1-3): a native language(corresponding to "the source language") model is provided for the native state decoding module, the non-native language(corresponding to "the target language") speech is decoded by the native state decoding module using the native language model, thus obtaining a native state decode sequence "b aa zh " (see figure 3).

[5] **Novelty and Inventive step**

[6] Claim 1 differs from D1 in: (1) the acoustic input of a user is spoken in a target language, and the acoustic model is trained in a source language different than the target language.

[7] Claim 5 differs from D1 in the above (1), and (2) at least one microphone configured to obtain acoustic information from the environment of the user device.

[8] Claim 9 differs from D1 in the above (1), (2) and (3) applying acoustic features derived from the acoustic input to the acoustic model to obtain a second sequence of speech units; determining if the user spoke the wake-up word at least in part by comparing the first sequence of speech units to the second sequence of speech units.

Form PCT/ISA/237 (Box No. V) (July 2011)

03/09/2020

/HUY C HO/

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

[9] Claim 16 differs from D1 in the above (1) and (3).

[10] Therefore, claims 1-20 are novel (PCT Article 33(2)).

[11] The feature (1) has already been disclosed by D2. The person skilled in the art would easily acquire the technical feature of D2 to decode the acoustic input.

[12] The feature (2) of obtaining acoustic information by microphone is general solution in the art.

[13] As to the feature (3), it is common practice for the person skilled in the art to determine the wake-up word by comparing the first sequence of speech units to the second sequence of speech units which is obtained by applying acoustic features derived from the acoustic input to the acoustic model instead of the WUP model.

[14] Therefore, claims 1,5,9,16 do not involve an inventive step (PCT Article 33 (3)).

[15] Dependent claims 2-4,6-8,10-15,17-20 do not involve an inventive step (PCT Article 33 (3)).

[16] As to claims 2,6,10,17,4,8,13,18, it is common practice to adapt the acoustic model using the acoustic input and the sequence, and associate at least one task with the wake-up word and initiate performance of the task based on the wake-up detection.

[17] As to claims 3,7,11,12,14,15,19,20, those additional features have been disclosed by D1.

[18] **Industrial applicability**

[19] Claims 1-20 can find industrial applicability in the technical field of wake-up detecting (PCT Article 33 (4)).

/HUY C HO/ 03/09/2020

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/238,238	ATTY. DOCKET NO.: N0484.71090US04
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: August 16, 2016	CONFIRMATION NO.: 5864
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		2002-0193989	A1	Geilhufe et al.	12-19-2002
		2008-0118080	A1	Gratke et al.	05-22-2008
		2009-0055178	A1	Coon	02-26-2009
		2012-0197637	A1	Gratke et al.	08-02-2012
		2012-0310646	A1	Hu et al.	12-06-2012
		2015-0106085	A1	Lindahl	04-16-2015
		2016-0039356	A1	Talwar et al.	02-11-2016
		2016-0314782	A1	Klimanis	10-27-2016
		2017-0116983	A1	Furukawa et al.	04-27-2017

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		EP	1511010	A1	Daimler Chrysler Ag	03-02-2005	Y-Abstract Only

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		International Preliminary Report on Patentability for International Application No. PCT/US2016/017317 mailed August 23, 2018 (N0484.71176WO00).	
		International Search Report and Written Opinion for International Application No. PCT/US2016/017317 mailed May 12, 2016 (N0484.71176WO00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:
/HUY C HO/	08/21/2020

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

8207852.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H/

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/08/27 09:05
S1	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/03/18 21:04
S2	104	(("GANONG") near3 ("William")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S3	16	(("VAN") near3 ("MULBREGT") near3 ("Paul")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S4	38	(("SEJNOHA") near3 ("Vladimir")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:19
S5	33	(("WILSON") near3 ("Glen")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2017/04/20 10:20
S6	220	(("NUANCE") near3 ("COMMUNICATIONS") near3 ("INC")).AANM.	USPAT	OR	OFF	2017/04/20 10:20
S7	43	("8977255" | "9112984" | "9361885" | "20070129949" | "20120329389" | "7574361" | "20100124896" | "20030040339" | "20100289994" | "20140274211" | "20140278435" | "8548176" | "20100185448" | "20110054899" | "20130080167" | "20130080171" | "20130289994" | "20130339028" | "20140012573" | "20140012586" | "5428707" | "5774859" | "5802305" | "5983186" | "6070140" | "6092043" | "6397186" | "6408396" | "6411926" | "6449496" | "6594630" | "6756700" | "6941265" | "6965786" | "7114090" | "7567827" | "7720683" | "7774204" | "8056070" | "8181046" | "8190420" | "8285545" | "8666751").PN.	US-PGPUB; USPAT	OR	OFF	2017/04/20 11:21

S8	5	"20120034904"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:10
S9	9	"9361885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/04/20 12:32
S10	33,784	(379/88.01-88.02 |G10L15/20 |G10L15/22 |G10L15/24 |G10L15/25 |G10L2015/223 |G10L2015/226 |H04M2250/74 |H04W52/0251 |H04W52/0254 |).ccls,cpc.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:11
S11	1	("15".src. and "238238".ap.) or "15238238".RLAN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/04/20 16:12
S12	25,130	S10 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:13
S13	4,617	S12 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S14	2,692	S13 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S15	563	S14 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO;	OR	ON	2017/04/20 16:31

			JPO; DERWENT; IBM_TDB			
S16	362	S15 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:31
S17	17	S16 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:33
S18	9,564	(704/246-255).ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S19	8,758	S18 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S20	1,733	S19 and ((voice audio speech) near4 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S21	1,338	S20 and ((voice speech) near4 (recogniz$5))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S22	212	S21 and (trigger)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	ON	2017/04/20 16:36

			DERWENT; IBM_TDB			
S23	104	S22 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S24	3	S23 and ((low saving idle) near7 (power battery energy) near7 (mode mood state status))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2017/04/20 16:36
S25	3	"9646610"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:41
S26	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2017/09/20 13:42
S27	8	"20100121636"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:42
S28	5	"20120035924"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:43
S29	6	"20120127072"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2017/09/20 13:45

S30	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/03/09 10:35
S31	12	"20140012573"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2018/03/09 10:53
S32	10	"20100009719"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2018/03/09 10:55
S33	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2018/05/30 09:37
S34	78	"9940936"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:18
S35	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2019/07/26 18:19
S36	3	"20140249817"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:20
S37	2	"20140365225"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/07/26 18:24
S38	2	"20160189706"	US-PGPUB; USPAT;	OR	ON	2019/07/26 18:26

			USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB			
S39	1,050	UTSM	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2019/08/11 10:17
S40	169	"6006175"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/06 12:27
S41	72	"6859776"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/06 12:36
S42	11	"9087520"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/06 12:42
S43	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/03/06 12:52
S44	23	"9558749"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/06 12:53
S45	11	(9,747,899 | 10,332,525 | 2012/0281885 | 2014/0039888 | 2014/0163978 | 2015/0006176 | 2015/0053779 | 2016/00788692019/0073999	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	OFF	2020/03/06 14:59

		2019/0287526 2019/0311715 2019/0355365).PN.	DERWENT; IBM_TDB			
S46	13	(9558749 | 9747899 | 10332525 | 2012/0281885 | 2014/0039888 | 2014/0163978 | 2015/0006176 | 2015/0053779 | 2016/0078869 | 2019/0073999 | 2019/0287526 | 2019/0311715 | 2019/0355365).PN.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/06 15:00
S47	3	"20120281885"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/06 15:05
S48	121	(9558749 | 9747899 | 10332525 | 20120281885 | 20140039888 | 20140163978 | 20150006176 | 20150053779 | 20160078869 | 20190073999 | 20190287526 | 20190311715 | 2190355365)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/06 15:06
S49	2	"20160189706"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/03/09 09:55
S50	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/03/09 10:22
S51	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/03/09 10:22
S52	2	"20020196228"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/09 10:36
S53	2	S52	US-PGPUB; USPAT; USOCR; FPRS; EPO;	OR	OFF	2020/03/09 10:36

			JPO; DERWENT; IBM_TDB			
S54	2	("20190042173").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/03/09 11:06
S55	104	"5508720"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/03/09 11:10
S56	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/08/21 20:43
S57	974	("20010047264" | "20020002461" | "20020010584" | "20020031262" | "20020032564" | "20020035474" | "20020042707" | "20020046025" | "20020059066" | "20020059068" | "20020067308" | "20020069063" | "20020077817" | "20020091511" | "20020099552" | "20020103641" | "20020116171" | "20020116185" | "20020116189" | "20020128827" | "20020133347" | "20020135565" | "20020138265" | "20020143551" | "20020154160" | "20020164000" | "20020184189" | "20020198714" | "20030078766" | "20030088414" | "20030097210" | "20030099335" | "20030120494" | "20030135740" | "20030167335" | "20030190074" | "20030212961" | "20030233230" | "20030234824" | "20040054535" | "20040054690" | "20040085162" | "20040106432" | "20040135701" | "20040145607" | "20040162741" | "20040176958" | "20040186714" | "20040193420" | "20040199387" | "20040220798" | "20040236778" | "20050002507" | "20050015772" | "20050027385" | "20050033582" | "20050045373" | "20050049880" | "20050055403" | "20050058438" | "20050071332" | "20050080625" | "20050086059" | "20050091118" | "20050100214" | "20050102614" |	US-PGPUB; USPAT; USOCR	OR	ON	2020/08/21 21:59

"20050108001" | "20050108074" |
"20050114124" | "20050114140" |
"20050119897" | "20050125235" |
"20050132301" | "20050143972" |
"20050152602" | "20050165607" |
"20050182616" | "20050182628" |
"20050182629" | "20050192801" |
"20050196733" | "20050203747" |
"20050203991" | "20050228665" |
"20050273626" | "20050288936" |
"20050289463" | "20060009973" |
"20060018492" | "20060061488" |
"20060095848" | "20060106592" |
"20060106594").PN. OR
("20060106595" | "20060111906" |
"20060117002" | "20060122834" |
"20060143007" | "20060156252" |
"20060193518" | "20060217967" |
"20060235700" | "20060252457" |
"20060293886" | "20070006098" |
"20070021956" | "20070027732" |
"20070033003" | "20070038436" |
"20070041361" | "20070050191" |
"20070050712" | "20070055514" |
"20070055525" | "20070055529" |
"20070058832" | "20070073540" |
"20070088556" | "20070094026" |
"20070100790" | "20070106674" |
"20070118377" | "20070118378" |
"20070124149" | "20070135949" |
"20070174188" | "20070185754" |
"20070185917" | "20070198269" |
"20070208569" | "20070208579" |
"20070208726" | "20070211071" |
"20070225980" | "20070276651" |
"20070276714" | "20070276810" |
"20070282595" | "20080012950" |
"20080015864" | "20080021708" |
"20080034032" | "20080052063" |
"20080052073" | "20080056579" |
"20080071544" | "20080077384" |
"20080079566" | "20080082332" |
"20080082338" | "20080082390" |
"20080082651" | "20080091406" |
"20080118143" | "20080120112" |
"20080120342" | "20080126100" |
"20080129520" | "20080131006" |
"20080140413" | "20080140416" |
"20080140652" | "20080140657" |
"20080189114" | "20080221880" |
"20080221889" | "20080221903" |
"20080228463" | "20080228490" |
"20080228496" | "20080240569" |

"20080247519" | "20080249770" |
"20080256613" | "20080270118" |
"20080281510" | "20080300878" |
"20080313335" | "20080319763" |
"20090006100" | "20090006343" |
"20090011709" | "20090018835" |
"20090028435" | "20090030800" |
"20090055179").PN. OR
("20090058823" | "20090070097" |
"20090076792" | "20090076796" |
"20090077165" | "20090094033" |
"20090100049" | "20090112572" |
"20090112677" | "20090123071" |
"20090125477" | "20090144049" |
"20090150156" | "20090157401" |
"20090164441" | "20090171664" |
"20090187577" | "20090191895" |
"20090204409" | "20090222488" |
"20090239552" | "20090253463" |
"20090271109" | "20090271175" |
"20090271178" | "20090287583" |
"20090290718" | "20090299745" |
"20090299849" | "20090300391" |
"20090306980" | "20090307162" |
"20100005081" | "20100019834" |
"20100023320" | "20100036660" |
"20100042400" | "20100049514" |
"20100057457" | "20100070899" |
"20100081456" | "20100088020" |
"20100088100" | "20100131273" |
"20100138215" | "20100138416" |
"20100145700" | "20100146442" |
"20100161554" | "20100204986" |
"20100217604" | "20100228540" |
"20100231474" | "20100235341" |
"20100257160" | "20100262599" |
"20100277579" | "20100278320" |
"20100278453" | "20100280983" |
"20100286985" | "20100299133" |
"20100299142" | "20100312547" |
"20100318576" | "20100324905" |
"20100332235" | "20100332280" |
"20100332348" | "20110047072" |
"20110054901" | "20110060807" |
"20110076994" | "20110082688" |
"20110112827" | "20110112921" |
"20110119049" | "20110125540" |
"20110130958" | "20110131036" |
"20110131045" | "20110143811" |
"20110144999" | "20110161076" |
"20110161309" | "20110175810" |
"20110184730" | "20110195758" |
"20110218855" | "20110224972" |

"20110231182" | "20110231188" |
"20110231474").PN. OR
("20110260861" | "20110264643" |
"20110279368" | "20110288861" |
"20110298585" | "20110306426" |
"20110314404" | "20120002820" |
"20120016678" | "20120020490" |
"20120022787" | "20120022857" |
"20120022860" | "20120022868" |
"20120022869" | "20120022870" |
"20120022872" | "20120022874" |
"20120022876" | "20120023088" |
"20120034904" | "20120035908" |
"20120035924" | "20120035931" |
"20120035932" | "20120042343" |
"20120084086" | "20120137367" |
"20120149394" | "20120173464" |
"20120197998" | "20120214517" |
"20120245719" | "20120245944" |
"20120265528" | "20120271635" |
"20120271676" | "20120309363" |
"20120310642" | "20120310649" |
"20120311583" | "20120311584" |
"20120311585" | "20120330661" |
"20130006638" | "20130110515" |
"20130110518" | "20130110519" |
"20130110520" | "20130111348" |
"20130111487" | "20130115927" |
"20130117022" | "20130185074" |
"20130185081" | "20130325443" |
"3704345" | "3828132" | "3979557" |
"4278838" | "4282405" | "4310721" |
"4348553" | "4653021" | "4680805" |
"4688195" | "4692941" | "4718094" |
"4724542" | "4726065" | "4727354" |
"4776016" | "4783807" | "4811243" |
"4819271" | "4827518" | "4827520" |
"4829576" | "4829583" | "4833712" |
"4839853" | "4852168" | "4862504" |
"4878230" | "4903305" | "4905163" |
"4914586" | "4914590" | "4944013" |
"4955047" | "4965763" | "4974191" |
"4975975" | "4977598" | "4992972" |
"5007098" | "5010574" | "5020112" |
"5021971" | "5022081" | "5027406" |
"5031217" | "5032989" | "5040218" |
"5047614" | "5057915").PN. OR
("5072452" | "5091945" | "5127053" |
"5127055" | "5128672" | "5133011" |
"5142584" | "5164900" | "5165007" |
"5179627" | "5179652" | "5194950" |
"5197005" | "5199077" | "5202952" |
"5208862" | "5216747" | "5220639" |

"5220657" | "5222146" | "5230036" |
"5235680" | "5267345" | "5268990" |
"5282265" | "5289562" | "5291286" |
"5293448" | "5293452" | "5297170" |
"5301109" | "5303406" | "5309359" |
"5317507" | "5317647" | "5325297" |
"5325298" | "5327498" | "5333236" |
"5333275" | "5345536" | "5349645" |
"5353377" | "5377301" | "5377303" |
"5384892" | "5384893" | "5386494" |
"5386556" | "5390279" | "5396625" |
"5400434" | "5404295" | "5412756" |
"5412804" | "5412806" | "5418951" |
"5424947" | "5434777" | "5444823" |
"5455888" | "5469529" | "5471611" |
"5475587" | "5479488" | "5491758" |
"5491772" | "5493677" | "5495604" |
"5500905" | "5502790" | "5502791" |
"5515475" | "5536902" | "5537618" |
"5544264" | "5555343" | "5574823" |
"5577164" | "5577241" | "5578808" |
"5579436" | "5581655" | "5584024" |
"5596676" | "5596994" | "5608624" |
"5613036" | "5617507" | "5619694" |
"5621859" | "5621903" | "5642464" |
"5642519" | "5644727" | "5649060" |
"5664055" | "5675819" | "5682539" |
"5687077" | "5696962" | "5701400" |
"5706442" | "5710886" | "5712957" |
"5715468" | "5721827" | "5727950" |
"5729694" | "5732216" | "5732390" |
"5734750" | "5734791" | "5737734" |
"5742705" | "5748974" | "5749081" |
"5757979" | "5759101" | "5777614" |
"5790978" | "5794050" | "5794182" |
"5794207" | "5794237").PN. OR
("5799276" | "5812697" | "5812698" |
"5822743" | "5825881" | "5826261" |
"5828999" | "5835893" | "5839106" |
"5841902" | "5845255" | "5857184" |
"5860063" | "5862223" | "5864806" |
"5864844" | "5867799" | "5873056" |
"5875437" | "5878394" | "5884323" |
"5895464" | "5895466" | "5899972" |
"5909666" | "5913193" | "5915236" |
"5915249" | "5920836" | "5920837" |
"5930408" | "5930769" | "5933822" |
"5936926" | "5940811" | "5941944" |
"5943670" | "5948040" | "5956699" |
"5960394" | "5960422" | "5963924" |
"5966126" | "5970474" | "5974146" |
"5982891" | "5983179" | "5987132" |
"5987140" | "5987404" | "5987440" |

"5991441" | "5999908" | "6016471" |
"6023684" | "6024288" | "6026345" |
"6026375" | "6026388" | "6026393" |
"6029132" | "6038533" | "6052656" |
"6055514" | "6055531" | "6064767" |
"6064959" | "6064960" | "6069648" |
"6070139" | "6070147" | "6073097" |
"6076051" | "6076088" | "6078914" |
"6081750" | "6081774" | "6088671" |
"6088731" | "6092043" | "6094649" |
"6101468" | "6105865" | "6108627" |
"6119101" | "6122616" | "6125356" |
"6144938" | "6161084" | "6173261" |
"6173279" | "6177905" | "6188999" |
"6195641" | "6205456" | "6208971" |
"6226403" | "6233559" | "6233578" |
"6240303" | "6246981" | "6249606" |
"6259826" | "6260011" | "6260013" |
"6260024" | "6266637" | "6275824" |
"6282507" | "6285786" | "6289124" |
"6308149" | "6311189" | "6317594" |
"6317707" | "6317831" | "6321092" |
"6324512" | "6334103" | "6353794" |
"6356854" | "6356905" | "6366883" |
"6366884" | "6397186").PN. OR
("6421672" | "6434522" | "6434524" |
"6442518" | "6446076" | "6449620" |
"6453281" | "6453292" | "6460029" |
"6466654" | "6477488" | "6487534" |
"6489951" | "6499013" | "6501937" |
"6505158" | "6505175" | "6505183" |
"6510417" | "6513063" | "6519565" |
"6523061" | "6523172" | "6526382" |
"6526395" | "6532444" | "6532446" |
"6546388" | "6553344" | "6556983" |
"6584464" | "6598022" | "6598039" |
"6601026" | "6601234" | "6604059" |
"6615172" | "6615175" | "6615220" |
"6622121" | "6622136" | "6625583" |
"6628808" | "6631346" | "6633846" |
"6643401" | "6647260" | "6650735" |
"6654740" | "6665639" | "6665640" |
"6665641" | "6680675" | "6684187" |
"6691064" | "6691090" | "6691111" |
"6691151" | "6697780" | "6697824" |
"6701294" | "6711585" | "6718324" |
"6721728" | "6735632" | "6742021" |
"6757362" | "6757718" | "6766320" |
"6772123" | "6778951" | "6778952" |
"6778962" | "6778970" | "6792082" |
"6807574" | "6810379" | "6813491" |
"6829603" | "6832194" | "6839464" |
"6839669" | "6842767" | "6847966" |

"6847979" | "6851115" | "6859931" |
"6865533" | "6895257" | "6895380" |
"6895558" | "6901399" | "6912498" |
"6912499" | "6924828" | "6928614" |
"6931384" | "6934684" | "6937975" |
"6937986" | "6957076" | "6960734" |
"6964023" | "6980949" | "6980955" |
"6985865" | "6988071" | "6996531" |
"6999927" | "7020685" | "7024363" |
"7027974" | "7031530" | "7036128" |
"7050977" | "7058569" | "7062428" |
"7069560" | "7072686" | "7084758" |
"7085723" | "7092887" | "7092928" |
"7093693" | "7107204").PN. OR
("7127046" | "7127403" | "7136710" |
"7137126" | "7139714" | "7139722" |
"7152070" | "7177798" | "7197460" |
"7200559" | "7203646" | "7216073" |
"7216080" | "7225125" | "7228278" |
"7233790" | "7233904" | "7266496" |
"7269556" | "7277854" | "7290039" |
"7292579" | "7299033" | "7302392" |
"7302686" | "7310600" | "7315818" |
"7324947" | "7349953" | "7362738" |
"7376556" | "7376645" | "7379874" |
"7386449" | "7389224" | "7392185" |
"7398209" | "7403938" | "7409337" |
"7415100" | "7418392" | "7426467" |
"7427024" | "7447635" | "7454351" |
"7460652" | "7467087" | "7475010" |
"7483832" | "7483894" | "7487089" |
"7496498" | "7496512" | "7502738" |
"7508373" | "7522927" | "7523108" |
"7526466" | "7528713" | "7529671" |
"7529676" | "7539656" | "7543232" |
"7546382" | "7548895" | "7552055" |
"7555431" | "7558730" | "7571106" |
"7577522" | "7580551" | "7580576" |
"7599918" | "7603381" | "7620549" |
"7624007" | "7634409" | "7636657" |
"7640160" | "7647225" | "7649454" |
"7657424" | "7664638" | "7672841" |
"7676026" | "7684985" | "7684990" |
"7693715" | "7693719" | "7693720" |
"7698131" | "7702500" | "7702508" |
"7707027" | "7707032" | "7707267" |
"7711565" | "7711672" | "7716056" |
"7720674" | "7720683" | "7721301" |
"7725307" | "7725318" | "7725320" |
"7725321" | "7729904" | "7729916" |
"7734461" | "7747616" | "7752152" |
"7756868" | "7774204" | "7783486" |
"7801729" | "7809570" | "7809610" |

		"7818176" | "7818291" | "7822608" | "7823123" | "7826945" | "7831426" | "7840400" | "7840447").PN. OR ("7853574" | "7853664" | "7873519" | "7873654" | "7881936" | "7885844" | "7890652" | "7912702" | "7917367" | "7917497" | "7920678" | "7925525" | "7930168" | "7930197" | "7949529" | "7949534" | "7974844" | "7974972" | "7983915" | "7983917" | "7983997" | "7986431" | "7987151" | "7996228" | "7999669" | "8000453" | "8005664" | "8005679" | "8015006" | "8024195" | "8032383" | "8036901" | "8041570" | "8041611" | "8050500" | "8055708" | "8065155" | "8065156" | "8069046" | "8069422" | "8073681" | "8078473" | "8082153" | "8095364" | "8099289" | "8107401" | "8112275" | "8112280" | "8117037" | "8122353" | "8131557" | "8138912" | "8140335" | "8165886" | "8166019" | "8188856" | "8190359" | "8195467" | "8204238" | "8205788" | "8219407" | "8285551" | "8285553" | "8290777" | "8290778" | "8290781" | "8296146" | "8296153" | "8296383" | "8301456" | "8311834" | "8370158" | "8371503" | "8374871" | "8447612" | "8498857" | "RE34562").PN. OR ("8977255").URPN.				
S58	2	"20020193989"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:09
S59	4	"20080118080"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:10
S60	36	"20090055178"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:11

S61	4	"20120197637"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:17
S62	3	"20120310646"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:18
S63	5	"20150106085"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:20
S64	5	"20160039356"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:22
S65	3	"20160314782"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:23
S66	13	"20170116983"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:24
S67	37	"1511010"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:25

S68	10	"20050049864"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:27
S69	5	"20050106085"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:34
S70	52	"2005106085"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:35
S71	4	"2012310646"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:36
S72	3	"20120310646"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/21 22:36
S73	183	("2013/0080167").URPN.	USPAT	OR	ON	2020/08/21 22:36
S74	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/08/24 11:14
S75	217	((voice audio speech) near4 (order$3 command$3)) same (server) same ((low$2 idle idling) near5 (power energy))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/24 11:46

S76	45	S75 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/24 12:11
S77	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/08/25 11:53
S78	1,042	((acoustic ambient audio) near5 (ambient input environment surrounding)) same ((voice speech) near4 (command order request$3)) same (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 13:40
S79	444	S78 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 13:42
S80	306	S79 and ((voice speech) near4 (command order request$3) near6 (detect$3 recogni$6))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 13:43
S81	87	S80 and (keyword)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 13:44
S82	742	("20010047264" | "20020032564" | "20020046025" | "20020055844" | "20020067308" | "20020069063" | "20020077817" | "20020103641" | "20020154160" | "20020163544" | "20020164000" | "20020173299" | "20020198714" | "20030013483" | "20030033151" | "20030120494" | "20030206199" | "20030212961" | "20030233237" | "20040030559" | "20040054535" | "20040054539" |	US-PGPUB; USPAT; USOCR	OR	ON	2020/08/25 14:35

"20040054690" | "20040135701" |
"20040162741" | "20040193420" |
"20040236778" | "20040260438" |
"20050015772" | "20050021336" |
"20050055403" | "20050071332" |
"20050080625" | "20050091118" |
"20050102614" | "20050108001" |
"20050108074" | "20050114124" |
"20050114140" | "20050119897" |
"20050125235" | "20050143972" |
"20050165607" | "20050182629" |
"20050196733" | "20050288936" |
"20060018492" | "20060035632" |
"20060061488" | "20060106592" |
"20060106594" | "20060106595" |
"20060111906" | "20060117002" |
"20060122834" | "20060143007" |
"20060148520" | "20060156252" |
"20070022178" | "20070027732" |
"20070050191" | "20070055529" |
"20070088556" | "20070100790" |
"20070106674" | "20070118377" |
"20070124149" | "20070135949" |
"20070174188" | "20070185754" |
"20070185917" | "20070208726" |
"20070276714" | "20070276810" |
"20070282595" | "20080015864" |
"20080021708" | "20080034032" |
"20080052063" | "20080071544" |
"20080079566" | "20080082390" |
"20080082651" | "20080104541" |
"20080120112" | "20080122796" |
"20080129520" | "20080140657" |
"20080189110" | "20080189360" |
"20080201306" | "20080221880" |
"20080221903").PN. OR
("20080228496" | "20080247519" |
"20080249770" | "20080300878" |
"20080300886" | "20080313335" |
"20080319763" | "20090006099" |
"20090006100" | "20090006343" |
"20090030800" | "20090055179" |
"20090058823" | "20090076796" |
"20090077165" | "20090094033" |
"20090100049" | "20090112572" |
"20090112677" | "20090150156" |
"20090157401" | "20090164441" |
"20090171664" | "20090191895" |
"20090239552" | "20090253463" |
"20090287583" | "20090290718" |
"20090299745" | "20090299849" |
"20090306980" | "20090307162" |
"20100005081" | "20100023320" |

"20100036660" | "20100042400" |
"20100081456" | "20100088020" |
"20100138215" | "20100145700" |
"20100146442" | "20100204986" |
"20100217604" | "20100228540" |
"20100235341" | "20100235793" |
"20100257160" | "20100262599" |
"20100277579" | "20100280983" |
"20100286985" | "20100299142" |
"20100312547" | "20100318576" |
"20100332235" | "20100332348" |
"20110047072" | "20110060807" |
"20110082688" | "20110112827" |
"20110112921" | "20110119049" |
"20110125540" | "20110130958" |
"20110131036" | "20110131045" |
"20110143811" | "20110144999" |
"20110161076" | "20110161309" |
"20110175810" | "20110184730" |
"20110195758" | "20110218855" |
"20110231182" | "20110231188" |
"20110260861" | "20110264643" |
"20110279368" | "20110306426" |
"20110314404" | "20120002820" |
"20120016678" | "20120020490" |
"20120022787" | "20120022857" |
"20120022860" | "20120022868" |
"20120022869" | "20120022870" |
"20120022874" | "20120022876" |
"20120023088").PN. OR
("20120034904" | "20120035908" |
"20120035924" | "20120035931" |
"20120035932" | "20120042343" |
"20120137367" | "20120149394" |
"20120173464" | "20120214517" |
"20120265528" | "20120271676" |
"20120311583" | "20130110518" |
"20130110520" | "20190189125" |
"3704345" | "3828132" | "3979557" |
"4278838" | "4282405" | "4310721" |
"4348553" | "4653021" | "4688195" |
"4692941" | "4718094" | "4724542" |
"4726065" | "4727354" | "4776016" |
"4783807" | "4811243" | "4819271" |
"4827520" | "4829576" | "4833712" |
"4839853" | "4852168" | "4862504" |
"4878230" | "4903305" | "4905163" |
"4914586" | "4914590" | "4944013" |
"4955047" | "4965763" | "4974191" |
"4977598" | "4992972" | "5010574" |
"5020112" | "5021971" | "5022081" |
"5027406" | "5031217" | "5032989" |
"5040218" | "5047617" | "5057915" |

"5072452" | "5091945" | "5127053" |
"5127055" | "5128672" | "5133011" |
"5142584" | "5164900" | "5165007" |
"5179652" | "5194950" | "5197005" |
"5199077" | "5202952" | "5208862" |
"5216747" | "5220639" | "5220657" |
"5222146" | "5230036" | "5235680" |
"5267345" | "5268990" | "5282265" |
"5291286" | "5293448" | "5293452" |
"5297170" | "5301109" | "5303406" |
"5309359" | "5317507" | "5317647" |
"5325297" | "5325298" | "5327498" |
"5333236" | "5333275" | "5345536" |
"5349645" | "5353377" | "5377301" |
"5377303" | "5384892" | "5384893" |
"5386494" | "5386556" | "5390279" |
"5396625" | "5400434" | "5404295" |
"5412756" | "5412804" | "5412806" |
"5418951" | "5424947" | "5434777" |
"5444823").PN. OR ("5455888" |
"5469529" | "5471611" | "5475587" |
"5479488" | "5491772" | "5493677" |
"5495604" | "5502790" | "5502791" |
"5515475" | "5536902" | "5537618" |
"5574823" | "5577164" | "5577241" |
"5578808" | "5579436" | "5581655" |
"5584024" | "5596676" | "5596994" |
"5608624" | "5613036" | "5617507" |
"5619694" | "5621859" | "5621903" |
"5642464" | "5642519" | "5644727" |
"5664055" | "5675819" | "5682539" |
"5687077" | "5696962" | "5701400" |
"5706442" | "5710886" | "5712957" |
"5715468" | "5721827" | "5727950" |
"5729694" | "5732390" | "5734791" |
"5737734" | "5748974" | "5749081" |
"5759101" | "5777614" | "5790978" |
"5794050" | "5794182" | "5794207" |
"5794237" | "5799276" | "5822743" |
"5825881" | "5826261" | "5828999" |
"5835893" | "5839106" | "5845255" |
"5857184" | "5860063" | "5862233" |
"5864806" | "5864844" | "5867799" |
"5873056" | "5875437" | "5884323" |
"5893063" | "5895464" | "5895466" |
"5899972" | "5913193" | "5915249" |
"5930769" | "5933822" | "5936926" |
"5940811" | "5941944" | "5943670" |
"5948040" | "5956699" | "5960422" |
"5963924" | "5966126" | "5970474" |
"5974146" | "5982891" | "5987132" |
"5987140" | "5987404" | "5987440" |
"6016471" | "6023684" | "6024288" |

"6026345" | "6026375" | "6026388" |
"6026393" | "6029132" | "6038533" |
"6052656" | "6055514" | "6055531" |
"6064960" | "6070139" | "6070147" |
"6076051" | "6076088" | "6078914" |
"6081750" | "6081774" | "6088671" |
"6088731" | "6094649" | "6105865" |
"6108627" | "6119101" | "6122616" |
"6125356").PN. OR ("6144938" |
"6161084" | "6173261" | "6173279" |
"6177905" | "6188999" | "6195641" |
"6205456" | "6208971" | "6233559" |
"6233578" | "6246981" | "6260024" |
"6266637" | "6275824" | "6282507" |
"6285786" | "6308149" | "6311189" |
"6317594" | "6317707" | "6317831" |
"6321092" | "6334103" | "6356854" |
"6356905" | "6366883" | "6366884" |
"6421672" | "6434524" | "6446076" |
"6449620" | "6453281" | "6453292" |
"6460029" | "6466654" | "6477488" |
"6487534" | "6499013" | "6501937" |
"6505158" | "6505175" | "6505183" |
"6510417" | "6513063" | "6523061" |
"6523172" | "6526382" | "6526395" |
"6532444" | "6532446" | "6546388" |
"6553344" | "6556983" | "6584464" |
"6598039" | "6601026" | "6601234" |
"6604059" | "6615172" | "6615175" |
"6615220" | "6625583" | "6631346" |
"6633846" | "6647260" | "6650735" |
"6654740" | "6665639" | "6665640" |
"6665641" | "6680675" | "6684187" |
"6691064" | "6691111" | "6691151" |
"6697780" | "6697824" | "6701294" |
"6711585" | "6718324" | "6721728" |
"6735632" | "6742021" | "6757362" |
"6757718" | "6766320" | "6778951" |
"6778952" | "6778962" | "6778970" |
"6792082" | "6807574" | "6810379" |
"6813491" | "6829603" | "6832194" |
"6842767" | "6847966" | "6847979" |
"6851115" | "6859931" | "6895380" |
"6895558" | "6901399" | "6912499" |
"6924828" | "6928614" | "6931384" |
"6937975" | "6937986" | "6944594" |
"6957076" | "6964023" | "6980949" |
"6980955" | "6985865" | "6988071" |
"6996531" | "6999927" | "7020685" |
"7027974" | "7036128" | "7050976" |
"7050977").PN. OR ("7058569" |
"7062428" | "7069560" | "7084758" |
"7092887" | "7092928" | "7093693" |

"7109970" | "7127046" | "7127403" |
"7136710" | "7137126" | "7139714" |
"7139722" | "7152070" | "7177798" |
"7197460" | "7200559" | "7203646" |
"7216073" | "7216080" | "7225125" |
"7233790" | "7233904" | "7266496" |
"7277854" | "7290039" | "7292579" |
"7299033" | "7302686" | "7310600" |
"7324947" | "7349953" | "7376556" |
"7376645" | "7379874" | "7386449" |
"7389224" | "7392185" | "7398209" |
"7403938" | "7409337" | "7415100" |
"7418392" | "7426467" | "7427024" |
"7447635" | "7454351" | "7467087" |
"7475010" | "7483894" | "7487089" |
"7496498" | "7496512" | "7502738" |
"7508373" | "7522927" | "7523108" |
"7526466" | "7528713" | "7529671" |
"7529676" | "7539656" | "7543232" |
"7546382" | "7548895" | "7552055" |
"7555431" | "7571106" | "7577522" |
"7599918" | "7603381" | "7620549" |
"7624007" | "7634409" | "7636657" |
"7640160" | "7647225" | "7649454" |
"7657424" | "7672841" | "7676026" |
"7684985" | "7693715" | "7693720" |
"7698131" | "7702500" | "7702508" |
"7707027" | "7707032" | "7707267" |
"7711565" | "7711672" | "7716056" |
"7720674" | "7720683" | "7725307" |
"7725318" | "7725320" | "7725321" |
"7729904" | "7729916" | "7734461" |
"7747616" | "7752152" | "7756868" |
"7774204" | "7783486" | "7801729" |
"7809570" | "7809610" | "7818176" |
"7818291" | "7822608" | "7823123" |
"7826945" | "7831426" | "7840400" |
"7840447" | "7853574" | "7873519" |
"7873654" | "7881936" | "7885844" |
"7890652").PN. OR ("7912702" |
"7917367" | "7917497" | "7920678" |
"7925525" | "7930168" | "7930197" |
"7949529" | "7974844" | "7974972" |
"7983915" | "7983917" | "7983997" |
"7986431" | "7987151" | "7996228" |
"7999669" | "8000453" | "8005679" |
"8015006" | "8024195" | "8036901" |
"8041570" | "8041611" | "8055708" |
"8065155" | "8065156" | "8069046" |
"8069422" | "8073681" | "8078473" |
"8082153" | "8095364" | "8099289" |
"8107401" | "8112275" | "8112280" |
"8117037" | "8138912" | "8140335" |

		"8165886" | "8166019" | "8188856" | "8190359" | "8195467" | "8204238" | "8205788" | "8219407" | "8285551" | "8285553" | "8290778" | "8290781" | "8296146" | "8296153" | "8301456" | "8311834" | "8370158" | "8371503" | "8374871" | "8447612" | "8694322" | "RE34562").PN. OR ("10540976").URPN.				
S83	2	"20140244273"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 14:37
S84	18	"10381001"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 14:40
S85	2	"20190355365"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:15
S86	2	"20190311715"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:18
S87	2	("20160358605").PN.	US-PGPUB; USPAT; USOCR; DERWENT	OR	ON	2020/08/25 23:26
S88	299,278	(379/88.01-88.02 |G10L15/20 |G10L15/22 |G10L15/24 |G10L15/25 |G10L2015/223 |G10L2015/226 |H04M2250/74 |H04W52/0251 |H04W52/0254 |H04W52/0254 |H04W52/0261 |H04W52/0251 |H04M1/72519 |G10L15/02 |H04W40/005 |G10L15/30 |G10L15/22 |G10L2015/226 |H04W88/02 |Y02D70/00 |Y02D30/70 |Y02D70/164	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:34

		|Y02D70/144 |H04M2250/74 |G10L2015/223 |Y02D70/142).ccls,cpc.				
S89	171,277	S88 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:37
S90	13,411	S89 and ((voice audio speech) near4 (request command order$3))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:38
S91	9,009	S90 and ((voice speech) near4 (recogniz$5 detect$3 input$4))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:38
S92	1,241	S91 and ((low saving idl$3) near5 (power battery energy))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:39
S93	1,402	S91 and ((low$2 saving idl$3) near5 (power battery energy))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2020/08/25 23:39
S94	867	S93 and (server)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 23:40
S95	551	S94 and ((acoustic ambient audio) same (ambient input environment surrounding) same (voice speech))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO;	OR	OFF	2020/08/25 23:41

			DERWENT; IBM_TDB			
S96	443	S94 and ((acoustic ambient audio) with (ambient input environment surrounding) with (voice speech))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 23:41
S97	351	S94 and ((acoustic ambient audio) near5 (ambient input environment surrounding) near5 (voice speech))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2020/08/25 23:42
S98	448	("4525793" | "5819220" | "5926789" | "5946050" | "5953392" | "5953700" | "5956683" | "5960399" | "5963618" | "6078886" | "6098043" | "6101473").PN. OR ("6408272").URPN.	US-PGPUB; USPAT; USOCR	OR	ON	2020/08/26 08:38

8/27/2020 10:49:05 AM
C:\Users\hho3\Documents\EAST_OLD2\Workspaces\15238238.wsp

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	**Art Unit**
	HUY C HO	2644

CPC						
Symbol					**Type**	**Version**
G10L	/	15	/	22	F	2013-01-01
H04W	/	52	/	0251	I	2013-01-01
H04W	/	52	/	0254	I	2013-01-01
H04M	/	1	/	72519	I	2013-01-01
G10L	/	15	/	02	I	2013-01-01
G10L	/	15	/	30	I	2013-01-01
H04W	/	40	/	005	I	2013-01-01
H04W	/	52	/	0261	I	2013-01-01
H04M	/	2250	/	74	A	2013-01-01
Y02D	/	30	/	70	A	2020-08-01
G10L	/	2015	/	223	A	2013-01-01
G10L	/	2015	/	226	A	2013-01-01
H04W	/	88	/	02	A	2013-01-01

CPC Combination Sets							
Symbol					**Type**	**Set**	**Ranking**	**Version**
	/		/					

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	19	
/HUY C HO/ Primary Examiner, Art Unit 2644	27 August 2020	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	9

U.S. Patent and Trademark Office — Part of Paper No.: 20200825

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	**Art Unit**
	HUY C HO	2644

INTERNATIONAL CLASSIFICATION					
CLAIMED					
H04M	/	1	/	64	

NON-CLAIMED					
	/		/		

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**
379	88.01

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					
379	88.01		88.02			

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	19	
/HUY C HO/ Primary Examiner, Art Unit 2644	27 August 2020	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	9

U.S. Patent and Trademark Office — Part of Paper No.: 20200825

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15/238,238	Ganong et al.
	Examiner	**Art Unit**
	HUY C HO	2644

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
1	1		10	17	19										
2	2	12	11	18	20										
3	3	13	12	19	21										
4	4	10	13												
5	5	11	14												
6	6	14	15												
7	7		16												
8	8	15	17												
9	9	16	18												

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	19	
/HUY C HO/ Primary Examiner, Art Unit 2644	27 August 2020	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	9

U.S. Patent and Trademark Office — Part of Paper No.: 20200825

Examiner-Initiated Interview Summary	Application No. 15/238,238	Applicant(s) Ganong et al.	
	Examiner HUY C HO	Art Unit 2644	AIA (FITF) Status No

All participants (applicant, applicant's representative, PTO personnel):

(1) HUY C. HO. (3) _____.

(2) DAVID S. GESNER (Reg. No. 72,100). (4) _____.

Date of Interview: 25 August 2020.

Type: ☑ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☐ No.
If Yes, brief description: _____.

Issues Discussed ☐ 101 ☐ 112 ☐ 102 ☐ 103 ☑ Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: 1.

Identification of prior art discussed: Mozer.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Discussed the supports for the new amended limitations and features from claims 1, 15 and 21. The claims are under condition for allowance..

Applicant recordation instructions: It is not necessary for applicant to provide a separate record of the substance of interview.

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☐ Attachment

/HUY C HO/ Primary Examiner, Art Unit 2644	

Bibliographic Data

Application No: 15/238,238

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Huy C. Ho/ (Examiner's Signature) | HH (Initials)

Title: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
08/16/2016	455	2644	N0484.71090US04
RULE			

APPLICANTS

Nuance Communications, Inc., Burlington, MA,

INVENTORS

William F. Ganong III, Brookline, MA, UNITED STATES

Paul Adrian Van Mulbregt, Wayland, MA, UNITED STATES

Vladimir Sejnoha, Lexington, MA, UNITED STATES

Glen Edward Wilson, Maynard, MA, UNITED STATES

CONTINUING DATA

This application is a CON of 13795933 03/12/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

08/25/2016

STATE OR COUNTRY

UNITED STATES

ADDRESS

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

FILING FEE RECEIVED

$1,600

Electronic Acknowledgement Receipt

EFS ID:	40922172
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	David Scott Gesner/Patricia McDonald
Filer Authorized By:	David Scott Gesner
Attorney Docket Number:	N0484.71090US04
Receipt Date:	22-OCT-2020
Filing Date:	16-AUG-2016
Time Stamp:	17:16:13
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US04-TRN-DSG.pdf	26195 5b15401b35ac5b2f3dc4593c9d881aa751459cb2	no	1

Warnings:

Information:					
2	Applicant summary of interview with examiner	N048471090US04-ISUM-DSG.pdf	28089 203494d8d5ec24bb615fd604d7640fc0e8679101	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	54284	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Docket No.: N0484.71090US04
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	15/238,238
Confirmation No.:	5864
Filed:	August 16, 2016
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: October 22, 2020 Electronic Signature for : /Trish McDonald/

SUMMARY OF TELEPHONE INTERVIEW WITH EXAMINER

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

The undersigned thanks Examiner Ho for the courtesies extended in the conducting the telephone interview conducted which occurred on Monday, October 19, 2020. The substance of the telephone interview is summarized herein.

The undersigned initiated the telephone interview after reviewing the returned Form PTO-1449/A and B (hereinafter "the form") which was appended to the Notice Of Allowance And Fee(s) Due of September 2, 2020, because the form raises questions about whether certain information cited by the Assignee has been considered by the Examiner.

In this respect, the form states, at the bottom, "ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C.H./", and includes two lines added by the Examiner. The topmost of these lines is placed in the section captioned "FOREIGN PATENT DOCUMENTS"

and the bottom-most of these lines is placed below the section captioned "OTHER ART – NON PATENT LITERATURE DOCUMENTS". The undersigned explained during the telephone interview that the presence of these lines makes it unclear whether (1) the foreign patent document listed on the form was not considered (given that the topmost line goes through that section of the form), (2) the non-patent literature documents listed on the form were not considered (given that they reside between the two lines placed on the form), or (3) all of the documents listed on the form have been considered. With respect to (3), it was explained that if all of the documents listed on the form have been considered, it is unclear why the Examiner placed any lines on the form at all.

The Examiner evidenced an appreciation for these points, and indicated that all of the information listed on the form has been considered. The telephone interview then concluded.

The Assignee believes that no fee is due with this correspondence. If a fee is due, please charge Deposit Account No. 23/2825, from which the undersigned is authorized to draw, under Docket No. N0484.71090US04.

Dated: October 22, 2020

Respectfully submitted,

By__/David S. Gesner/_____
David S. Gesner
Registration No.: 72,100
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

8639547.1

Doc Code: TRAN.LET
Document Description: Transmittal Letter

USED IN LIEU OF PTO/SB/21 (07-09)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/238,238-Conf. #5864
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	H. C. Ho
Attorney Docket Number	N0484.71090US04

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

Summary of Telephone Interview with Examiner

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/David S. Gesner/		
Printed name	David S. Gesner		
Date	October 22, 2020	Reg. No.	72,100

Certificate of Electronic Filing under 37 C.F.R. § 1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: October 22, 2020 Electronic Signature for : /Trish McDonald/

8639545.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04	5864

7590 12/18/2020

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER	
HO, HUY C	
ART UNIT	PAPER NUMBER
2644	
NOTIFICATION DATE	DELIVERY MODE
12/18/2020	ELECTRONIC

Notice of Abandonment

This application is abandoned in view of:

1. ☐ The applicant's failure to timely file a proper reply to the Office letter mailed on _____.
   (a) ☐ A reply was received on _____ (with a Certificate of Mailing or Transmission date _____), which is after the expiration of the period for reply (including a total extension of ____ month(s)) which expired on _____.
   (b) ☐ No reply has been received.
2. ☑ Applicant's failure to timely pay the required issue fee and publication fee, if applicable, within the statutory period of three months from the mailing date of the Notice of Allowance (PTOL-85).
   (a) ☐ The issue fee and publication fee, if applicable, was received on _____ (with a Certificate of Mailing or Transmission date _____), which is after the expiration of the statutory period for payment of the issue fee (and publication fee) set in the Notice of Allowance (PTOL-85).
   (b) ☐ The submitted fee of $_____ is insufficient. A balance of $_____ is due.
   The issue fee required by 37 CFR 1.18 is $_____.
   The publication fee, if required by 37 CFR 1.18(d), is $_____.
   (c) ☑ The issue fee and publication fee, if applicable, has not been received.
3. ☐ Applicant's failure to timely file corrected drawings as required by, and within the three-month period set in, the Notice of Allowability (PTO-37).
   (a) ☐ Proposed corrected drawings were received on _____ (with a Certificate of Mailing or Transmission dated _____), which is after the expiration of the period for reply.
   (b) ☐ No corrected drawing have been received.
4. ☐ Applicant's failure to timely file the inventor's oath or declaration no later than the date on which the issue fee was paid as required by the Notice Requiring Inventor's Oath or Declaration (PTO-2306).
   (a) ☐ An inventor's oath or declaration was received on _____ (with a Certificate of Mailing or Transmission date _____), which is after the date on which the issue fee was paid.
   (b) ☐ While an oath or declaration (or substitute statement) for one or more inventors was received, an oath or declaration (or substitute statement) for at least one other inventor has not been received.
   (c) ☐ No inventor's oath or declaration has been received.
5. ☐ Drawings received on _____ were disapproved by examiner. See examiner's response dated _____.
6. ☐ Corrected drawings were received on _____ , which is after the expiration of the one-month period for reply set in examiner's response dated _____.
7. ☐ No corrected drawings have been received in reply to one-month period set in examiner's response dated _____.
8. ☐ The reason(s) below:

Petitions to revive under 37 CFR 1.137(a) or request to withdraw the holding of abandonment under 37 CFR 1.181, should be promptly filed to minimize any negative effects on patent term.

A Gezahegn

(571)-272-4200 or 1(888)-786-0101
Patent Publication Branch
Office of Data Management

FORM PTO-ABN0 (Rev. 03/17)

Electronic Acknowledgement Receipt

EFS ID:	42549028
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Rachel Ann Smith/Kristine Grand
Filer Authorized By:	Rachel Ann Smith
Attorney Docket Number:	N0484.71090US04
Receipt Date:	26-APR-2021
Filing Date:	16-AUG-2016
Time Stamp:	15:28:56
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Power of Attorney	TRANSMITTAL_POA_CERN0103PUS1_120291USCON2.PDF	2085160 2ddf9b4f7c2a933cf2ecab8a4a117d2ede69d4bb	no	3

Warnings:

Information:	
Total Files Size (in bytes):	2085160

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B or equivalent) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5. If the Power of Attorney by Applicant form is not accompanied by this transmittal form or an equivalent, the Power of Attorney will not be recognized in the application.

Application Number	15/238,238
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong III
Title	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Art Unit	2644
Examiner Name	Huy C. Ho
Attorney Docket Number	CERN0103PUS1 (12-0291-US-CON-2)

SIGNATURE of Applicant or Patent Practitioner

Signature	/Rachel A. Smith/	Date	
Name	Rachel A. Smith	Telephone	248-358-4400
Registration Number	66,437		

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d)

* Total of 1 Power of Attorney form is attached

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82B (07-13)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

☑ I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above: 179598

OR

☐ I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:

☑ The address associated with the above-mentioned Customer Number

OR

☐ The address associated with Customer Number:

OR

☐ Firm or Individual Name					
Address					
City		State		Zip	
Country					
Telephone		Email			

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

CERENCE OPERATING COMPANY

☐ Inventor or Joint Inventor (title not required below)

☐ Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)

☑ Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)

☐ Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature	/Leanne Fitzgerald/	Date (Optional)	2021-04-21
Name	Leanne Fitzgerald		
Title	General Counsel and Secretary		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

☑ Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04

CONFIRMATION NO. 5864

IMPROPER CFR REQUEST

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

OC000000125235184

Date Mailed: 04/29/2021

RESPONSE TO REQUEST FOR CORRECTED FILING RECEIPT

Power of Attorney, Claims, Fees, System Limitations, and Miscellaneous

In response to your request for a corrected Filing Receipt, the Office is unable to comply with your request because:

- Any request to correct or update the name of the applicant must include an application data sheet (ADS) in compliance with 37 CFR 1.76 specifying the correct or updated name of the applicant in the applicant information section. Any request to change the applicant after an original applicant has been specified under 37 CFR 1.46(b) must include a new ADS in compliance with 37 CFR 1.76 specifying the applicant in the applicant information section and comply with 37 CFR 3.71 and 3.73. See 37 CFR 1.46(c).

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/mteklemichael/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04

CONFIRMATION NO. 5864

IMPROPER CPOA LETTER

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

OC000000125234881

Date Mailed: 04/29/2021

NOTICE REGARDING POWER OF ATTORNEY

This is in response to the power of attorney filed 04/26/2021. The power of attorney in this application is not accepted for the reason(s) listed below:

- The power of attorney has not been accepted because the party who is giving power has not been identified. Power of attorney may only be signed by the applicant for patent (37 CFR 1.42) or the patent owner. A party who is not the applicant must become the applicant in accordance with 37 CFR 1.46(c) and appoint any power of attorney in compliance with 37 CFR 3.71 and 3.73. For a reissue application, reexamination proceeding, or supplemental examination proceeding, a patent owner who was not the applicant under 37 CFR 1.46 must appoint any power of attorney in compliance with 37 CFR 3.71 and 3.73. See 37 CFR 1.32(b)(4).

/mteklemichael/

Office of Data Management, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

179598
Brooks Kushman / Cerence
1000 Town Center
22nd Floor
Southfield, MI 48075
UNITED STATES

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B or equivalent) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5. If the Power of Attorney by Applicant form is not accompanied by this transmittal form or an equivalent, the Power of Attorney will not be recognized in the application.

Application Number	15/238,238
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong III
Title	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Art Unit	2644
Examiner Name	Huy C. Ho
Attorney Docket Number	CERN0103PUS1 (12-0291-US-CON-2)

SIGNATURE of Applicant or Patent Practitioner

Signature	/Rachel A. Smith/	Date	
Name	Rachel A. Smith	Telephone	248-358-4400
Registration Number	66,437		

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d)

* Total of 1 Power of Attorney form is attached

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82B (07-13)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

☑ I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above: 179598

OR

☐ I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:

☑ The address associated with the above-mentioned Customer Number

OR

☐ The address associated with Customer Number:

OR

☐ Firm or Individual Name					
Address					
City		State		Zip	
Country					
Telephone		Email			

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

CERENCE OPERATING COMPANY

☐ Inventor or Joint Inventor (title not required below)

☐ Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)

☑ Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)

☐ Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature	/Leanne Fitzgerald/	Date (Optional)	2021-04-21
Name	Leanne Fitzgerald		
Title	General Counsel and Secretary		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

☑ Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: CERENCE OPERATING COMPANY

Application No./Patent No.: 15/238,238 Filed/Issue Date: 8/16/2016

Titled: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

CERENCE OPERATING COMPANY, a CORPORATION

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. ☑ The assignee of the entire right, title, and interest.

2. ☐ An assignee of less than the entire right, title, and interest (check applicable box):

   ☐ The extent (by percentage) of its ownership interest is _____%. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   ☐ There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. ☐ The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. ☐ The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. ☐ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

B. ☑ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Vladimir Sejnoha; Paul Adrian Van Mulbregt; Glen Edward Wilson; William F. Ganong III To: NUANCE COMMUNICATIONS, INC.

   The document was recorded in the United States Patent and Trademark Office at
   Reel 040342, Frame 0893, or for which a copy thereof is attached.

2. From: NUANCE COMMUNICATIONS, INC. To: CERENCE OPERATING COMPANY

   The document was recorded in the United States Patent and Trademark Office at
   Reel 055927, Frame 0620, or for which a copy thereof is attached.

This collection of information is required by 37 CFR 3.73(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☐ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Rachel A. Smith/	5/3/2021
Signature	Date
Rachel A. Smith	66,437
Printed or Typed Name	Title or Registration Number

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Electronic Acknowledgement Receipt

EFS ID:	42619372
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Rachel Ann Smith/Kristine Grand
Filer Authorized By:	Rachel Ann Smith
Attorney Docket Number:	N0484.71090US04
Receipt Date:	04-MAY-2021
Filing Date:	16-AUG-2016
Time Stamp:	16:59:54
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Power of Attorney	TRANSMITTAL_POA_CERN0103PUS1_120291USCON2.PDF	2085160 2ddf9b4f7c2a933cf2ecab8a4a117d2ede69d4bb	no	3

Warnings:

Information:					
2	Assignee showing of ownership per 37 CFR 3.73	373_STATEMENT_CERN0103PUS1_12091USCON2.PDF	127722 9c0fe044021dd37e31a5e4fb71cd35a75d3371a9	no	3
Warnings:					
Information:					
		Total Files Size (in bytes):	2212882		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04

CONFIRMATION NO. 5864

IMPROPER CPOA LETTER

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

OC00000125439476

Date Mailed: 05/10/2021

NOTICE REGARDING POWER OF ATTORNEY

This is in response to the power of attorney filed 05/04/2021. The power of attorney in this application is not accepted for the reason(s) listed below:

- The revocation is not signed by the applicant, the assignee of the entire interest, or one particular principal attorney having the authority to revoke.

/tlulu/

Office of Data Management, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

179598
Brooks Kushman / Cerence
1000 Town Center
22nd Floor
Southfield, MI 48075
UNITED STATES

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B or equivalent) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5. If the Power of Attorney by Applicant form is not accompanied by this transmittal form or an equivalent, the Power of Attorney will not be recognized in the application.

Application Number	15/238,238
Filing Date	August 16, 2016
First Named Inventor	William F. Ganong III
Title	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Art Unit	2644
Examiner Name	Huy C. Ho
Attorney Docket Number	CERN0103PUS1 (12-0291-US-CON-2)

SIGNATURE of Applicant or Patent Practitioner

Signature	/Rachel A. Smith/	Date	
Name	Rachel A. Smith	Telephone	248-358-4400
Registration Number	66,437		

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d)

* Total of 1 Power of Attorney form is attached

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82B (07-13)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

☑ I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above: 179598

OR

☐ I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:

☑ The address associated with the above-mentioned Customer Number

OR

☐ The address associated with Customer Number:

OR

☐ Firm or Individual Name					
Address					
City		State		Zip	
Country					
Telephone		Email			

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

CERENCE OPERATING COMPANY

☐ Inventor or Joint Inventor (title not required below)

☐ Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)

☑ Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)

☐ Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature	/Leanne Fitzgerald/	Date (Optional)	2021-04-21
Name	Leanne Fitzgerald		
Title	General Counsel and Secretary		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

☑ Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

MEIK PFEFFINGER

Group Art Unit: 2644

Examiner: Huy C. Ho

Serial No.: 15/238238

Filed: August 16, 2016

For: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Attorney Docket No.: CERN0103PUS1
12–0291–US–CON–2

REQUEST FOR CORRECTED APPLICATION DATA SHEET

Mail Stop
Commissioner for Patents
U.S. Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Commissioner:

Applicant respectfully submits herewith a Corrected Application Data sheet to update the Applicant in response to the Notice Regarding Power of Attorney dated May 10, 2021. Other details including client reference numbers have been updated as well.

Please charge any fees or credit any overpayments as a result of the filing of this paper to our Deposit Account No. 02-3978.

Respectfully submitted,

MEIK PFEFFINGER

By: /Rachel A. Smith/
Rachel A. Smith
Reg. No. 66,437
Attorney for Applicant

Date: May 25, 2021

BROOKS KUSHMAN P.C.
1000 Town Center, 22nd Floor
Southfield, MI 48075-1238
Phone: 248-358-4400
Fax: 248-358-3351

Petitioner Exhibit 1004, Page 1013 of 1052

UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #	RECEIPT DATE / TIME	ATTORNEY DOCKET #
15/238,238	**05/25/2021 03:42:39 PM ET**	**N0484.71090US04**

Title of Invention

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Application Information

APPLICATION TYPE	Utility - Nonprovisional Application under 35 USC 111(a)	PATENT #	-
CONFIRMATION #	5864	FILED BY	Heather Greenwell
PATENT CENTER #	60201390	FILING DATE	08/16/2016
CUSTOMER #	91337	FIRST NAMED INVENTOR	William F. Ganong III
CORRESPONDENCE ADDRESS	-	AUTHORIZED BY	Rachel Smith

Documents

TOTAL DOCUMENTS: 5

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
CERN0103PUS1_Transmittal_Letter.PDF	2	Transmittal Letter	95 KB
CERN0103PUS1_Corrected_ADS.pdf	9	Application Data Sheet	7493 KB
Warning: This is not a USPTO supplied ADS fillable form. Data in the form cannot be automatically loaded to other USPTO systems.			
373_STATEMENT_CERN0103PUS1_12091USCON2.PDF	3	Assignee showing of ownership per 37 CFR 3.73	111 KB
TRANSMITTAL_POA_CERN0103PUS1_120291USCON2.	3	Power of Attorney	2036 KB

PDF			
CERN0103PUS1_Chain_of_Title.pdf	2	Miscellaneous Incoming Letter	135 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
CERN0103PUS1_Transmittal_Letter.PDF	B2104EA0060BDE26F6EB494599CC5BCFF93724E68EF89EFD7FCF711E2DC784789AB53D6CE56117DC23F4FF7D7F789F5BCD6BFA900248818D7124390D6D21DC9D
CERN0103PUS1_Corrected_ADS.pdf	912DD2958E737923A5046F0F9F61B92B6B3E89F09E28EB02E8382F3156D142BD11D5393BD4E0BE263921AD51D0F25AD880EC99F825EE0D0B6CE9EEDA7F54E296
373_STATEMENT_CERN0103PUS1_12091USCON2.PDF	C5FFFF19C25F0985EF1189D70FBC125AC8DB127ABA15B34455CBCB8E73AC8D2E698BCADB1E64EDA64933811FD9BF2D4B7DC6F8E47FEB7960423DDD7DD6759CDD
TRANSMITTAL_POA_CERN0103PUS1_120291USCON2.PDF	53EA979826196EC131135A40EECCBE8F9DDB0C1E5F60B6B888F886EF2F64A0A3F025B3528466334A9BB70BA62D7E3A44111537B18BE24B310486ABAB997CB658
CERN0103PUS1_Chain_of_Title.pdf	C9797959AF5A69C34EF8E97FB2462CADA3AB9AE375278209FAD2B8B0CD7D3B8685672555B7C87015E5E263E0F0F5A84723B9CC54AFE4E9EA04D91DE9A45265C9

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage

submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES
PATENT AND TRADEMARK OFFICE

Assignment abstract of title for Application 15238238

Invention title/Inventor	Patent	Publication	Application	PCT	International registration
METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND William F. Ganong III, Paul Adrian Van Mulbregt, Vladimir Sejnoha, Glen Edward Wilson	None	20160358605 Dec 8, 2016	15238238 Aug 16, 2016		

Assignments (2 total)

Assignment 2

Reel/frame	Execution date	Date recorded	Properties	Pages
055927/0620	Apr 15, 2021	Apr 15, 2021	28	7

Conveyance
ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).

Assignors
NUANCE COMMUNICATIONS, INC.

Correspondent
VASILIA KELLY
15 WAYSIDE ROAD
CERENCE INC.
BURLINGTON, MA 01803

Assignee
CERENCE OPERATING COMPANY
15 WAYSIDE ROAD
BURLINGTON, MASSACHUSETTS 01803

Assignment 1

Reel/frame	Execution date	Date recorded	Properties	Pages
040342/0893	Mar 14, 2013	Nov 16, 2016	1	7

Conveyance
ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).

Assignors
SEJNOHA, VLADIMIR
VAN MULBREGT, PAUL ADRIAN
WILSON, GLEN EDWARD
GANONG, WILLIAM F., III

Correspondent
THOMAS S. CHLEBECK
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210

Assignee
NUANCE COMMUNICATIONS, INC.
ONE WAYSIDE ROAD
BURLINGTON, MASSACHUSETTS 01803

CORRECTED APPLICATION DATA SHEET

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2:

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	William	F.	GANONG	III

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Brookline	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	118 Sumner Road		
Address 2			
City	Brookline	State/Province	MA
Postal Code	02146	Country	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Paul	Adrian	Van MULBREGT	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Wayland	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	30 Bow Road		
Address 2			
City	Wayland	State/Province	MA
Postal Code	01778	Country	US

Inventor 3 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Vladimir		SEJNOHA	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

City	Lexington	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	10 Flintlock Road		
Address 2			
City	Lexington	State/Province	MA
Postal Code	02420	Country	US

Inventor 4 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Glen	Edward	WILSON	

Residence Information (Select One) (•) US Residency ( ) Non US Residency ( ) Active US Military Service

City	Maynard	State/Province	MA	Country of Residence	US

Mailing Address of Inventor:

Address 1	~~13 Sherman Sreet~~ 13 Sherman Street		
Address 2			
City	Maynard	State/Province	~~MS~~ MA
Postal Code	01754	Country	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	~~157176~~ 179598		
Email Address	~~DOCKETING@OGRP.COM~~	Add Email	Remove Email

Application Information:

Title of the Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND		
Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)	Small Entity Status Claimed	☐
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	10	Suggested Figure for Publication (if any)	

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	◉ Customer Number	○ US Patent Practitioner	○ Limited Recognition (37 CFR 11.9)
Customer Number	~~157176~~ 179598		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status			Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Continuation of	13/795933	2013-03-12

Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

Remove

Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)

Additional Foreign Priority Data may be generated within this form by selecting the **Add** button.

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

◉ Assignee	○ Legal Representative under 35 U.S.C. 117	○ Joint Inventor
○ Person to whom the inventor is obligated to assign.	○ Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor:

If the Applicant is an Organization check here. ☒

Organization Name	~~NUANCE COMMUNICATIONS, INC.~~ CERENCE OPERATING COMPANY

Mailing Address Information For Applicant:

Address 1	~~One Wayside Road~~ 15 Wayside Road		
Address 2			
City	Burlington	**State/Province**	MA
Country	US	Postal Code	01803
Phone Number		Fax Number	
Email Address			

Additional Applicant Data may be generated within this form by selecting the Add button.

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

If the Assignee or Non-Applicant Assignee is an Organization check here. ☒

Organization Name	~~NUANCE COMMUNICATIONS, INC.~~ CERENCE OPERATING COMPANY

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1	~~One Wayside Road~~ 15 Wayside Road		
Address 2			
City	Burlington	**State/Province**	MA
Country	US	Postal Code	01803
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button.

Signature:

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Rachel A. Smith/			Date (YYYY-MM-DD)	2021-05-25
First Name	Rachel	Last Name	Smith	Registration Number	66437

Additional Signature may be generated within this form by selecting the Add button.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	~~0013359USX1/4705A~~ CERN0103PUS1 (12-0291-US-CON-2)
	Application Number	15/238,238
Title of Invention	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND	

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1 The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3 A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent CooperationTreaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: CERENCE OPERATING COMPANY

Application No./Patent No.: 15/238,238 Filed/Issue Date: August 16, 2016

Titled: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

CERENCE OPERATING COMPANY, a CORPORATION

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. ☑ The assignee of the entire right, title, and interest.

2. ☐ An assignee of less than the entire right, title, and interest (check applicable box):

   ☐ The extent (by percentage) of its ownership interest is _____%. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   ☐ There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. ☐ The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. ☐ The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. ☐ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

B. ☑ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Vladimir Sejnoha; Paul Adrian Van Mulbregt; Glen Edward Wilson; William F. Ganong III To: NUANCE COMMUNICATIONS, INC.

   The document was recorded in the United States Patent and Trademark Office at
   Reel 040342, Frame 0893, or for which a copy thereof is attached.

2. From: NUANCE COMMUNICATIONS, INC. To: CERENCE OPERATING COMPANY

   The document was recorded in the United States Patent and Trademark Office at
   Reel 055927, Frame 0620, or for which a copy thereof is attached.

This collection of information is required by 37 CFR 3.73(b). The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☐ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Rachel A. Smith/	2021-05-25
Signature	Date
Rachel A. Smith	66,437
Printed or Typed Name	Title or Registration Number

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC PAYMENT RECEIPT

APPLICATION #	RECEIPT DATE / TIME	ATTORNEY DOCKET #
15/238,238	05/25/2021 03:42:39 PM ET	N0484.71090US04

Title of Invention

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Application Information

APPLICATION TYPE	Utility - Nonprovisional Application under 35 USC 111(a)	PATENT #	-
CONFIRMATION #	5864	FILED BY	Heather Greenwell
PATENT CENTER #	60201390	AUTHORIZED BY	Rachel Smith
CUSTOMER #	91337	FILING DATE	08/16/2016
CORRESPONDENCE ADDRESS	-	FIRST NAMED INVENTOR	William F. Ganong III

Payment Information

PAYMENT METHOD	PAYMENT TRANSACTION ID	PAYMENT AUTHORIZED BY
DA / 023978	E20215OF43540207	Heather Greenwell

PRE-AUTHORIZED ACCOUNT	PRE-AUTHORIZED CATEGORY
023978	37 CFR 1.16 (National application filing, search, and examination fees); 37 CFR 1.17 (Patent application and reexamination processing fees); 37 CFR 1.20 (Post issuance fees)

FEE CODE	DESCRIPTION	ITEM PRICE($)	QUANTITY	ITEM TOTAL($)
1830	PROCESSING FEE, EXCEPT IN PROVISIONAL APPLICATIONS	140.00	1	140.00
			TOTAL AMOUNT:	$140.00

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d)

and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	N0484.71090US04

CONFIRMATION NO. 5864

POWER OF ATTORNEY NOTICE

91337
Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

OC000000125891153

Date Mailed: 05/27/2021

NOTICE REGARDING CHANGE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 05/25/2021.

- The Power of Attorney to you in this application has been revoked by the assignee who has intervened as provided by 37 CFR 3.71. Future correspondence will be mailed to the new address of record(37 CFR 1.33).

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/sharris/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/238,238	08/16/2016	William F. Ganong III	CERN0103PUS1(12-0291-US-C

CONFIRMATION NO. 5864

179598
Brooks Kushman / Cerence
1000 Town Center
22nd Floor
Southfield, MI 48075

POA ACCEPTANCE LETTER

OC000000125891279

Date Mailed: 05/27/2021

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 05/25/2021.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/sharris/

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

91337 7590 09/02/2020

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

Rachel A. Smith	(Typed or printed name)
/Rachel A. Smith/	(Signature)
July 7, 2021	(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	~~N0484-71090US04~~ CERN0103PUS1 (12-0291-US-CON-2)	5864

TITLE OF INVENTION: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1000	$0.00	$0.00	$1000	12/02/2020

EXAMINER	ART UNIT	CLASS-SUBCLASS
HO, HUY C	2644	455-556100

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 Brooks Kushman, P.C.
2
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: CERENCE OPERATING COMPANY

(B) RESIDENCE: (CITY and STATE OR COUNTRY): BURLINGTON, MA

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. Fees submitted: ☒ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

☒ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☒ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. 02-3978

5. **Change in Entity Status** (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29

☐ Applicant asserting small entity status. See 37 CFR 1.27

☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /Rachel A. Smith/ Date July 7, 2021

Typed or printed name Rachel A. Smith Registration No. 66,437

PTOL-85 Part B (08-18) Approved for use through 01/31/2020 OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Petitioner Exhibit 1004, Page 1035 of 1052

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Rachel Ann Smith/Kristine Grand
Attorney Docket Number:	CERN0103PUS1(12-0291-US-C

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
UTILITY APPL ISSUE FEE	1501	1	1000	1000

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**1000**

Electronic Acknowledgement Receipt

EFS ID:	43187100
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	179598
Filer:	Rachel Ann Smith/Kristine Grand
Filer Authorized By:	Rachel Ann Smith
Attorney Docket Number:	CERN0103PUS1(12-0291-US-C
Receipt Date:	07-JUL-2021
Filing Date:	16-AUG-2016
Time Stamp:	13:35:05
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1000
RAM confirmation Number	E202177D35250229
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	ISSUE_FEE_TRANSMITTAL_CERN0103PUS1_120291USCON2.PDF	782152 5c7368ea6fc7df93f3d6dce2f113b9287c238f6f	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30440 33edbc63a09e5b1b1430e0a0123b32d5366c35f8	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	812592		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Decision Date : July 7, 2021

In re Application of :

William Ganong

Application No : 15238238

Filed : 16-Aug-2016

Attorney Docket No : CERN0103PUS1(12-0291-US-C

DECISION ON PETITION

UNDER CFR 1.137(a)

This is an electronic decision on the petition under 37 CFR 1.137(a), filed July 7, 2021 , to revive the above-identified application.

The petition is **GRANTED.**

The above-identified application became abandoned for failure to reply in a timely manner to the Notice of Allowance and Issue Fee(s) Due. The date of abandonment is the day after the expiration date of the period set for reply in the Notice.

The electronic petition satisfies the conditions for revival pursuant to the provisions of 37 CFR 1.137(a) in that (1) the reply in the form of payment of the Issue Fee and the Publication Fee (if necessary); (2) the petition fee as set forth in 37 CFR 1.17 (m); (3) the drawing correction and/or other deficiencies (if necessary); and (4) the required statement of unintentional delay have been received. Accordingly, the Issue Fee payment is accepted as having been unintentionally delayed.

Telephone inquiries concerning this decision should be directed to the Patent Electronic Business Center (EBC) at 866-217-9197.

This application file is being directed to the Office of Data Management.

Office of Petitions

Electronic Acknowledgement Receipt

EFS ID:	43185253
Application Number:	15238238
International Application Number:	
Confirmation Number:	5864
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	179598
Filer:	Rachel Ann Smith/Kristine Grand
Filer Authorized By:	Rachel Ann Smith
Attorney Docket Number:	CERN0103PUS1(12-0291-US-C
Receipt Date:	07-JUL-2021
Filing Date:	16-AUG-2016
Time Stamp:	13:29:44
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$2100
RAM confirmation Number	E202177D29380104
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	ISSUE_FEE_TRANSMITTAL_CERN0103PUS1_120291USCON2.PDF	782152 5c7368ea6fc7df93f3d6dce2f113b9287c238f6f	no	1
Warnings:					
Information:					
2	Petition automatically granted by EFS	petition-request.pdf	39976 e48a0145de731a4c7b8cc48890f818ca68e1e167	no	2
Warnings:					
Information:					
3	Fee Worksheet (SB06)	fee-info.pdf	30756 834faac5e6c607958dcd93fe3435010bc7598da5	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	852884		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Electronic Patent Application Fee Transmittal

Application Number:	15238238
Filing Date:	16-Aug-2016
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Rachel Ann Smith/Kristine Grand
Attorney Docket Number:	CERN0103PUS1(12-0291-US-C

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
PET. REVIVE ABANDON APP, DELAY PYMT-RESP	1453	1	2100	2100
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**2100**

Doc Code: PET.AUTO Document Description: Petition automatically granted by EFS-Web	PTO/SB/64 U.S. Patent and Trademark Office Department of Commerce

Electronic Petition Request	PETITION FOR REVIVAL OF AN APPLICATION FOR PATENT ABANDONED UNINTENTIONALLY UNDER 37 CFR 1.137(a)
Application Number	15238238
Filing Date	16-Aug-2016
First Named Inventor	William Ganong
Art Unit	2644
Examiner Name	HUY HO
Attorney Docket Number	CERN0103PUS1(12-0291-US-C
Title	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

The above-identified application became abandoned for failure to file a timely and proper reply to a notice or action by the United States Patent and Trademark Office. The date of abandonment is the day after the expiration date of the period set for reply in the office notice or action plus any extensions of time actually obtained.

APPLICANT HEREBY PETITIONS FOR REVIVAL OF THIS APPLICATION

NOTE: A grantable petition requires the following items:
(1) Petition fee;
(2) Reply and/or issue fee;
(3) Terminal disclaimer with disclaimer fee – required for all utility and plant applications filed before June 8, 1995; and for all design applications;
(4) Statement that the entire delay was unintentional.

Petition Fee

◯ Small Entity	
◯ Micro Entity	
◉ Regular Undiscounted	

Issue Fee and Publication Fee :

Issue Fee and Publication Fee are not due.

☒ Issue Fee Transmittal is attached

Drawing corrections and/ or other deficiencies.

◉ Drawing corrections and/ or other deficiencies are not required

○ I certify, in accordance with 37 CFR 1.4.(d)(4), that drawing corrections and/ or other deficiencies have previously been filed on

○ Drawing corrections and/ or other deficiencies are attached.

☒ **STATEMENT:** The entire delay in filing the required reply from the due date for the required reply until the filing of a grantable petition under 37 CFR 1.137(a) was unintentional.

Petitioner is reminded that a delay resulting from a deliberately chosen course of action on the part of the applicant does not become an "unintentional" delay within the meaning of 37CFR 1.137 because:

- the applicant does not consider the claims to be patentable over the references relied upon in an outstanding Office action;
- the applicant does not consider the allowed or patentable claims to be of sufficient breadth or scope to justify the financial expense of obtaining a patent;
- the applicant does not consider any patent to be of sufficient value to justify the financial expense of obtaining the patent; or
- the applicant remains interested in eventually obtaining a patent, but simply seeks to defer patent fees and patent prosecution expenses.

Petitioner is further reminded that an intentional delay resulting from a deliberate course of action chosen by the applicant is not affected by:

- the correctness of the applicant's (or applicant's representative's) decision to abandon the application or not to seek or persist in seeking revival of the application;
- the correctness or propriety of a rejection, or other objection, requirement, or decision by the Office; or
- the discovery of new information or evidence, or other change in circumstances subsequent to the abandonment or decision not to seek or persist in seeking revival.

NOTE: Where the petition under 37 CFR 1.137(a) is filed more than two years after the date the application became abandoned, the United States Patent and Trademark Office requires an additional explanation of the circumstances surrounding the delay that establishes the entire delay was unintentional. This requirement is in addition to the requirement to provide a statement that the entire delay was unintentional. See *Clarification of the Practice for Requiring Additional Information in Petitions Filed in Patent Applications and Patents Based on Unintentional Delay*, 85 FR 12222 (March 2, 2020). See MPEP 711.03(c)(II)(C)-(F) for additional guidance on the information required to establish that the entire delay was unintentional.

THIS PORTION MUST BE COMPLETED BY THE SIGNATORY OR SIGNATORIES

I certify, in accordance with 37 CFR 1.4(d)(4) that I am:

◉ An attorney or agent registered to practice before the Patent and Trademark Office who has been given power of attorney in this application.

○ An attorney or agent registered to practice before the Patent and Trademark Office, acting in a representative capacity.

○ A sole inventor

○ A joint inventor; I certify that I am authorized to sign this submission on behalf of all of the inventors as evidenced by the power of attorney in the application

○ A joint inventor; all of whom are signing this e-petition.

Signature	/Rachel A. Smith/
Name	Rachel A. Smith
Registration Number	66437

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

91337 7590 09/02/2020

Nuance Communications, Inc.
c/o Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

Rachel A. Smith	(Typed or printed name)
/Rachel A. Smith/	(Signature)
July 7, 2021	(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/16/2016	William F. Ganong III	~~N0484-71090US04~~ CERN0103PUS1 (12-0291-US-CON-2)	5864

TITLE OF INVENTION: METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1000	$0.00	$0.00	$1000	12/02/2020

EXAMINER	ART UNIT	CLASS-SUBCLASS
HO, HUY C	2644	455-556100

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 Brooks Kushman, P.C.
2
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: CERENCE OPERATING COMPANY

(B) RESIDENCE: (CITY and STATE OR COUNTRY): BURLINGTON, MA

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. Fees submitted: ☒ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

☒ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☒ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. 02-3978

5. **Change in Entity Status** (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29

☐ Applicant asserting small entity status. See 37 CFR 1.27

☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /Rachel A. Smith/ Date July 7, 2021

Typed or printed name Rachel A. Smith Registration No. 66,437

Petitioner Exhibit 1004, Page 1047 of 1052

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/238,238	08/16/2016	2644	1600	CERN0103PUS1(12-0291-US-C	20	2

CONFIRMATION NO. 5864

CORRECTED FILING RECEIPT

179598
Brooks Kushman / Cerence
1000 Town Center
22nd Floor
Southfield, MI 48075

OC000000127023030

Date Mailed: 07/16/2021

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

William F. Ganong III, Brookline, MA;
Paul Adrian Van Mulbregt, Wayland, MA;
Vladimir Sejnoha, Lexington, MA;
Glen Edward Wilson, Maynard, MA;

Applicant(s)

CERENCE OPERATING COMPANY, Burlington, MA;

Assignment For Published Patent Application

Nuance Communications, Inc., Burlington, MA

Power of Attorney: The patent practitioners associated with Customer Number 179598

Domestic Priority data as claimed by applicant

This application is a CON of 13/795,933 03/12/2013

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 08/25/2016

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 15/238,238**

Projected Publication Date: Not Applicable

Non-Publication Request: No

Early Publication Request: No

Title

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Preliminary Class

455

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/238,238	08/10/2021	11087750	CERN0103PUS1(12-0291-US-C	5864

179598 7590 07/21/2021

Brooks Kushman / Cerence
1000 Town Center
22nd Floor
Southfield, MI 48075

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 0 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Data Management (ODM) at (571)-272-4200.

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

William F. Ganong III, Brookline, MA;
CERENCE OPERATING COMPANY, Burlington, MA;
Paul Adrian Van Mulbregt, Wayland, MA;
Vladimir Sejnoha, Lexington, MA;
Glen Edward Wilson, Maynard, MA;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.