Electronic Acknowledgement Receipt

EFS ID:	15190582
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Customer Number:	91337
Filer:	Marc S. Johannes/Jennifer Leveille
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US01
Receipt Date:	12-MAR-2013
Filing Date:	
Time Stamp:	14:58:51
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 of New Application	N048471090US01-TRN-MSJ.pdf	33098 21499fb569c72f9030ec584ca3bb32cfa2d6812	no	1

Warnings:	
Information:	

2	Application Data Sheet	N048471090US01-ADS-MSJ.PDF	138272 84132aad2ff6e98ab1c386a3c7159c324c37b1d9	no	4
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
3		N048471090US01-UTLAPL-MSJ.pdf	311464 6eccb57e225ad55f74ee4708ae8f7c2f83f36b34	yes	68

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Specification	1	63
	Claims	64	67
	Abstract	68	68

Warnings:					
Information:					
4	Drawings-only black and white line drawings	N048471090US01-DRW-MSJ.PDF	2194415 0b6dfc356a3a50b0b52dd3f648bd4eab6a8418d4	no	10
Warnings:					
Information:					
		Total Files Size (in bytes):	2677249		

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.

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

BACKGROUND

[0001] 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.).

[0002] 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.

[0003] 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

Petitioner Exhibit 1004, Page 3 of 1518

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.

[0004] 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

[0005] 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.

[0006] 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.

[0007] 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.

[0008] 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

Petitioner Exhibit 1004, Page 4 of 1518

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.

[0009] 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.

[0010] 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

[0011] 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,

Petitioner Exhibit 1004, Page 5 of 1518

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.

[0012] 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.

[0013] 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.

Petitioner Exhibit 1004, Page 6 of 1518

BRIEF DESCRIPTION OF DRAWINGS

[0014] Various aspects and embodiments of the application will be described with reference to the following figures.

[0015] FIG. 1 illustrates an exemplary mobile device on which voice response techniques may be implemented;

[0016] FIG. 1B illustrates exemplary system components of an illustrative mobile device on which voice response techniques may be implemented;

[0017] 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;

[0018] 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;

[0019] 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;

[0020] 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;

[0021] 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;

[0022] 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;

[0023] 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 7 of 1518

[0024] 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

[0025] FIG. 10 illustrates an exemplary computer system on which techniques described herein may be implemented.

DETAILED DESCRIPTION

[0026] 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).

[0027] 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.

[0028] 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

Petitioner Exhibit 1004, Page 8 of 1518

power down different components when entering a low power mode. However, mobile 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.

[0029] 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.

[0030] 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).

[0031] 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

Petitioner Exhibit 1004, Page 9 of 1518

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).

[0032] 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.

[0033] 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.

[0034] 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

Petitioner Exhibit 1004, Page 10 of 1518

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.

[0035] 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.

[0036] 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.

Petitioner Exhibit 1004, Page 11 of 1518

[0037] 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 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.

[0038] 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.

[0039] 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.

[0040] Following below are more detailed descriptions of various concepts related to, and embodiments of, methods and apparatus for implementing techniques for

Petitioner Exhibit 1004, Page 12 of 1518

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.

[0041] FIG. 1A illustrates 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.

[0042] 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.

[0043] 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

Petitioner Exhibit 1004, Page 13 of 1518

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.

[0044] 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.

[0045] 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.

[0046] 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.

Petitioner Exhibit 1004, Page 14 of 1518

[0047] 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 activity in the environment and obtain the resulting acoustic input for further processing to assess whether the acoustic input includes a voice command.

[0048] 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.

[0049] 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

Petitioner Exhibit 1004, Page 15 of 1518

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.

[0050] 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 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.

[0051] 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.

[0052] 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.

Petitioner Exhibit 1004, Page 16 of 1518

[0053] 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 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).

[0054] 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.

[0055] 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

Petitioner Exhibit 1004, Page 17 of 1518

spoken. Some non-limiting examples of speech processing stages that may be utilized are discussed in further detail below.

[0056] 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 selected to generally optimize false positive and false negative rates while attempting to keep power consumption to a minimum.

[0057] 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.

[0058] 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,

Petitioner Exhibit 1004, Page 18 of 1518

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).

[0059] 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.

[0060] 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

Petitioner Exhibit 1004, Page 19 of 1518

overhead to engage the components that provide the contextual cues and process this information).

[0061] 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 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.

[0062] 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.

Petitioner Exhibit 1004, Page 20 of 1518

[0063] 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.

[0064] 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 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.

[0065] 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.

[0066] 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

Petitioner Exhibit 1004, Page 21 of 1518

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.

[0067] 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.

[0068] 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 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).

[0069] 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,

Petitioner Exhibit 1004, Page 22 of 1518

initiation of the processes to perform the actions needed to respond to the voice command may be invoked.

[0070] 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.

[0071] 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 determining its content. Some exemplary VAD and speech processing stages are discussed in further detail below.

[0072] 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.

Petitioner Exhibit 1004, Page 23 of 1518

[0073] 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.

[0074] 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 the primary processor in a low power mode until and/or unless the primary processor is needed.

[0075] 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.

Petitioner Exhibit 1004, Page 24 of 1518

[0076] 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.

[0077] 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. 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.

[0078] 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

Petitioner Exhibit 1004, Page 25 of 1518

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.

[0079] 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.

[0080] 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 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.

[0081] 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

Petitioner Exhibit 1004, Page 26 of 1518

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.

[0082] 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.

[0083] 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 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.

[0084] 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

Petitioner Exhibit 1004, Page 27 of 1518

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.

[0085] 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.

[0086] 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).

[0087] 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

Petitioner Exhibit 1004, Page 28 of 1518

facilitate relatively fast, relatively low power ASR to obtain information about whether acoustic input includes a voice command.

[0088] 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.

[0089] 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 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.

[0090] According to some embodiments, generally unrestricted ASR (e.g., ASR performed without restricting or limiting the available vocabulary of the ASR component

Petitioner Exhibit 1004, Page 29 of 1518

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.

[0091] 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.

[0092] Such classification may be a binary decision, or the classifier (e.g., clusters) may include more numerous classifications. According to some embodiments, 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

Petitioner Exhibit 1004, Page 30 of 1518

classifier processing stages are not limited to any one, particular or combination of techniques.

[0093] 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.

[0094] 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.

[0095] 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.

[0096] 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 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).

[0097] 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

Petitioner Exhibit 1004, Page 31 of 1518

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.

[0098] 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.

[0099] 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 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.

Petitioner Exhibit 1004, Page 32 of 1518

[00100] 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.

[00101] 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).

[00102] 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.

[00103] 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 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

Petitioner Exhibit 1004, Page 33 of 1518

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.

[00104] 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.

[00105] 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).

[00106] 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 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

Petitioner Exhibit 1004, Page 34 of 1518

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.

[00107] 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.

[00108] 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.

[00109] 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 35 of 1518

[00110] 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.

[00111] 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.

[00112] 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 36 of 1518

[00113] 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.

[00114] 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.

[00115] 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

Petitioner Exhibit 1004, Page 37 of 1518

situations where the confidence is very high that the acoustic input includes a voice command (e.g., this may be achieved by modifying thresholds in one or more processing stages when a low battery condition is detected).

[00116] 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.

[00117] 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).

[00118] 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.

Petitioner Exhibit 1004, Page 38 of 1518

[00119] Speaker recognition techniques may be used as part of the process of determining whether acoustic input includes an actionable voice command. According 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.

[00120] 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.

[00121] 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.

Petitioner Exhibit 1004, Page 39 of 1518

[00122] 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 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.

[00123] FIG. 1B 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. 1B is provided to demonstrate some concepts related to generally low power "always-on" functionality, in accordance with some embodiments.

[00124] 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. 1B is provided to describe exemplary techniques for exploiting the processing capabilities of mobile devices that have multiple processors. As such, the configuration in FIG. 1B 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.

[00125] 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

Petitioner Exhibit 1004, Page 40 of 1518

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 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.

[00126] 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.

[00127] 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.

[00128] 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

Petitioner Exhibit 1004, Page 41 of 1518

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

[00129] 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).

[00130] 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.

[00131] 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

Petitioner Exhibit 1004, Page 42 of 1518

(e.g., when the secondary processor cannot itself reach a conclusion with enough certainty).

[00132] 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 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.

[00133] 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.

[00134] 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).

[00135] 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

Petitioner Exhibit 1004, Page 43 of 1518

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

[00136] 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.

[00137] 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.

Petitioner Exhibit 1004, Page 44 of 1518

[00138] 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.

[00139] 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.

[00140] 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.

[00141] 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

Petitioner Exhibit 1004, Page 45 of 1518

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.

[00142] 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 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.

[00143] 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.

[00144] 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

Petitioner Exhibit 1004, Page 46 of 1518

voice content. For example, a voting scheme, which may be weighted or not, may be used based on results from multiple VAD processing stages.

[00145] 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 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.

[00146] 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.

[00147] 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

Petitioner Exhibit 1004, Page 47 of 1518

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.

[00148] 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 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.

[00149] 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

Petitioner Exhibit 1004, Page 48 of 1518

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.

[00150] 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 (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).

[00151] 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.

Petitioner Exhibit 1004, Page 49 of 1518

[00152] 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.

[00153] 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 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.

[00154] 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

Petitioner Exhibit 1004, Page 50 of 1518

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.

[00155] 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 requiring the user to first speak an explicit voice trigger because speaking the command that the user would like carried out.

[00156] 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.

[00157] 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.

Petitioner Exhibit 1004, Page 51 of 1518

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.

[00158] 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 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.

[00159] 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.

[00160] 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.

Petitioner Exhibit 1004, Page 52 of 1518

[00161] 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.

[00162] 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 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.

[00163] 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

Petitioner Exhibit 1004, Page 53 of 1518

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.

[00164] 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 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.

[00165] 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

Petitioner Exhibit 1004, Page 54 of 1518

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.

[00166] 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 subsequent speech input to act upon or is presently processing any remainder of speech contained in acoustic input 805.

[00167] 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.

Petitioner Exhibit 1004, Page 55 of 1518

[00168] 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.

[00169] 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.

[00170] 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.

Petitioner Exhibit 1004, Page 56 of 1518

[00171] 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.

[00172] 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 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.

[00173] 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

Petitioner Exhibit 1004, Page 57 of 1518

input can be provided as an audio input to ASR component 930 (e.g., transmitted over a network).

[00174] 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 implementation. Any of the ASR processing stages described herein may be performed by ASR component 930.

[00175] 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

Petitioner Exhibit 1004, Page 58 of 1518

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).

[00176] 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.

[00177] 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

Petitioner Exhibit 1004, Page 59 of 1518

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.

[00178] 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 and/or identify one or more tasks or actions specified in the acoustic input and/or any parameters associated with identified tasks.

[00179] 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

Petitioner Exhibit 1004, Page 60 of 1518

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).

[00180] 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).

[00181] 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-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.

[00182] 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.

Petitioner Exhibit 1004, Page 61 of 1518

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.

[00183] 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.

[00184] 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 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.

[00185] 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

Petitioner Exhibit 1004, Page 62 of 1518

functionality to provide data, and may include control apparatus to perform I/O functionality.

[00186] 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).

[00187] 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 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.

[00188] 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.

Petitioner Exhibit 1004, Page 63 of 1518

[00189] 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.

[00190] 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.

[00191] All definitions, as defined and used herein, should be understood to control over dictionary definitions, and/or ordinary meanings of the defined terms.

[00192] 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. 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.

Petitioner Exhibit 1004, Page 64 of 1518

[00193] 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.

[00194] 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).

[00195] 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.

[00196] 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 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 65 of 1518

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
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
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::	Glenn
Family Name::	Wilson
City of Residence::	Burlington
State or Province of Residence::	MA
Country of Residence::	US
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

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::	N0484.71090US01
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets::	10
Small Entity?::	No
Petition included?::	No
Secrecy Order in Parent Appl.?::	No
Authorization to Permit Access::	No

Representative Information

Representative Customer Number:: 91337

Domestic Priority Information

Foreign Priority Information

Applicant Information

Non-Applicant Assignee Information

Signature:

A signature of the applicant or representative is required in accordance with 37 CFR 1.33 and 10.18. Please see 37 CFR 1.4(d) for the form of the signature.			
Signature	/Marc S. Johannes/	Date	March 12, 2013
Name (Print/Type)	Marc S. Johannes	Registration No. (Attorney/Agent)	64,978

PTO/AIA/15 (07-12)
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.	N0484.71090US01
First 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.

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

1. [ ] **Fee Transmittal Form.** (PTO/SB/17 or equivalent)
2. [ ] **Applicant claims small entity status.** See 37 CFR 1.27.
3. [X] **Specification.** [*Total Pages* 68]
   Both the claims and abstract must start on a new page
   (For information on the preferred arrangement, see MPEP § 608.01(a))
4. [X] **Drawing(s).** *(35 U.S.C. 113)* [*Total Sheets* 10]
5. **Inventor's Oath or Declaration.** [*Total Sheets*]
   (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. [ ] A copy from a prior application (37 CFR 1.63(d))
6. [X] **Application Data Sheet.** **See Note below.*
   See 37 CFR 1.76 (PTO/AIA/14 or equivalent)
7. [ ] **CD-ROM or CD-R.**
   In duplicate, large table or Computer Program *(Appendix)*
   [ ] Landscape Table on CD
8. **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

ACCOMPANYING APPLICATION PARTS

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

18. CORRESPONDENCE ADDRESS

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

Name					
Address					
City		State		Zip Code	
Country		Telephone		Email	

Signature	/Marc S. Johannes/	Date	March 12, 2013
Name (Print/Type)	Marc S. Johannes	Registration No. (Attorney/Agent)	64,978

3030103.1

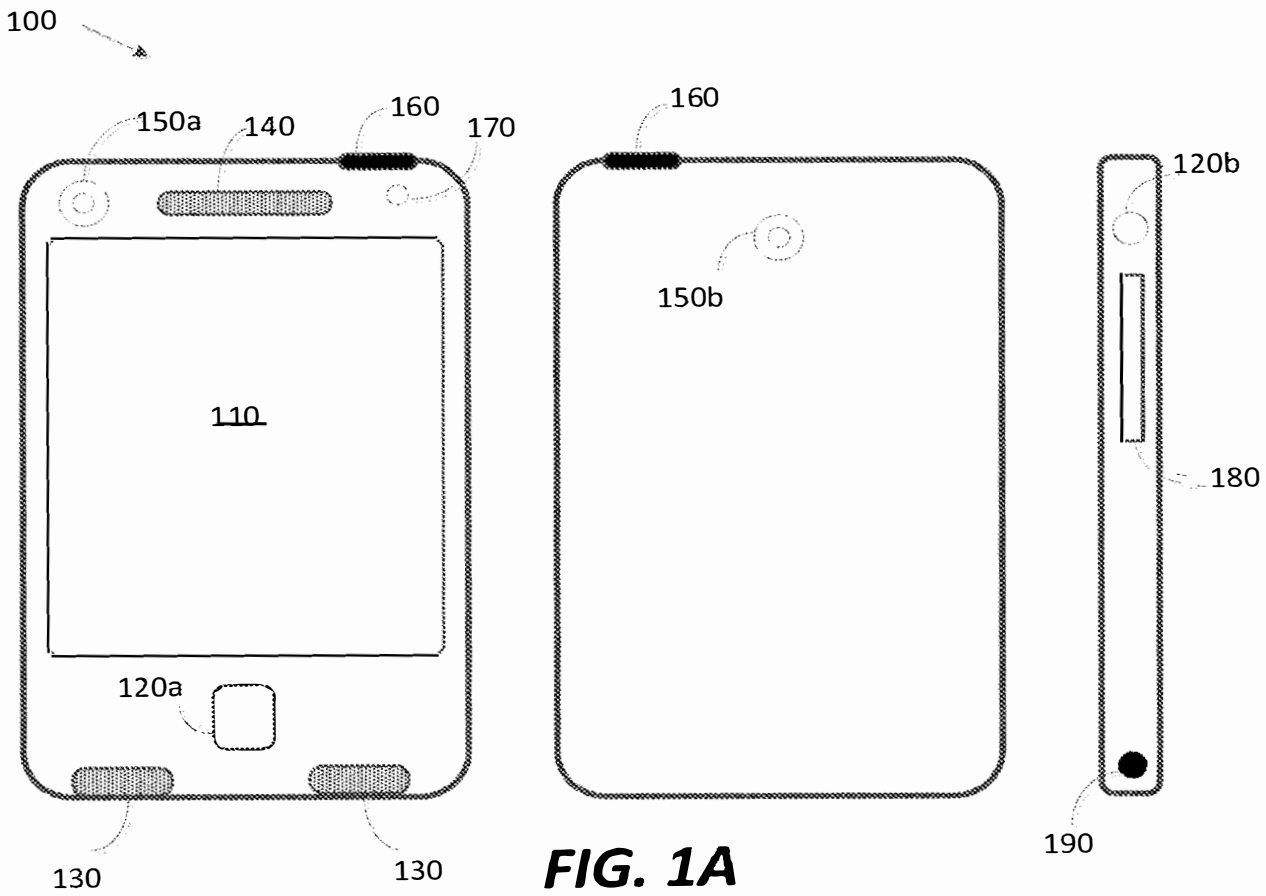

100
150a
140
160
170
110
120a
130
130
160
150b
120b
180
190

FIG. 1A

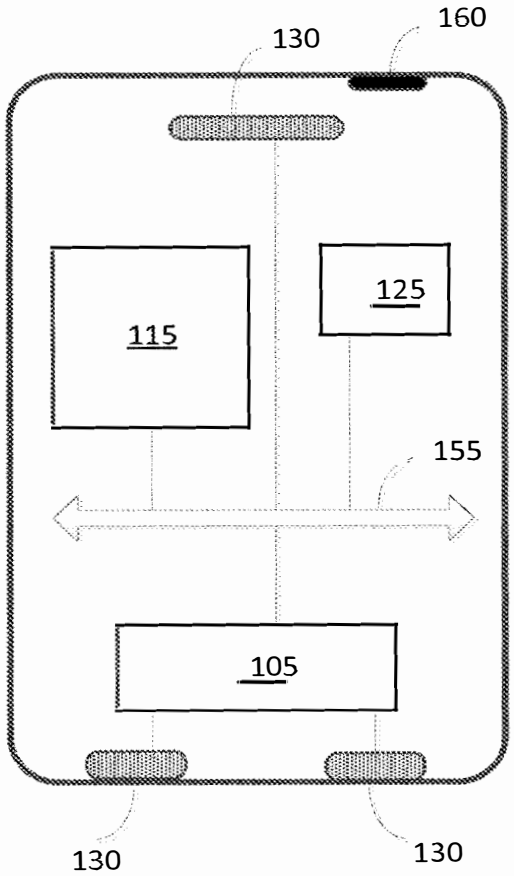

130
160
115
125
155
105
130
130

FIG. 1B

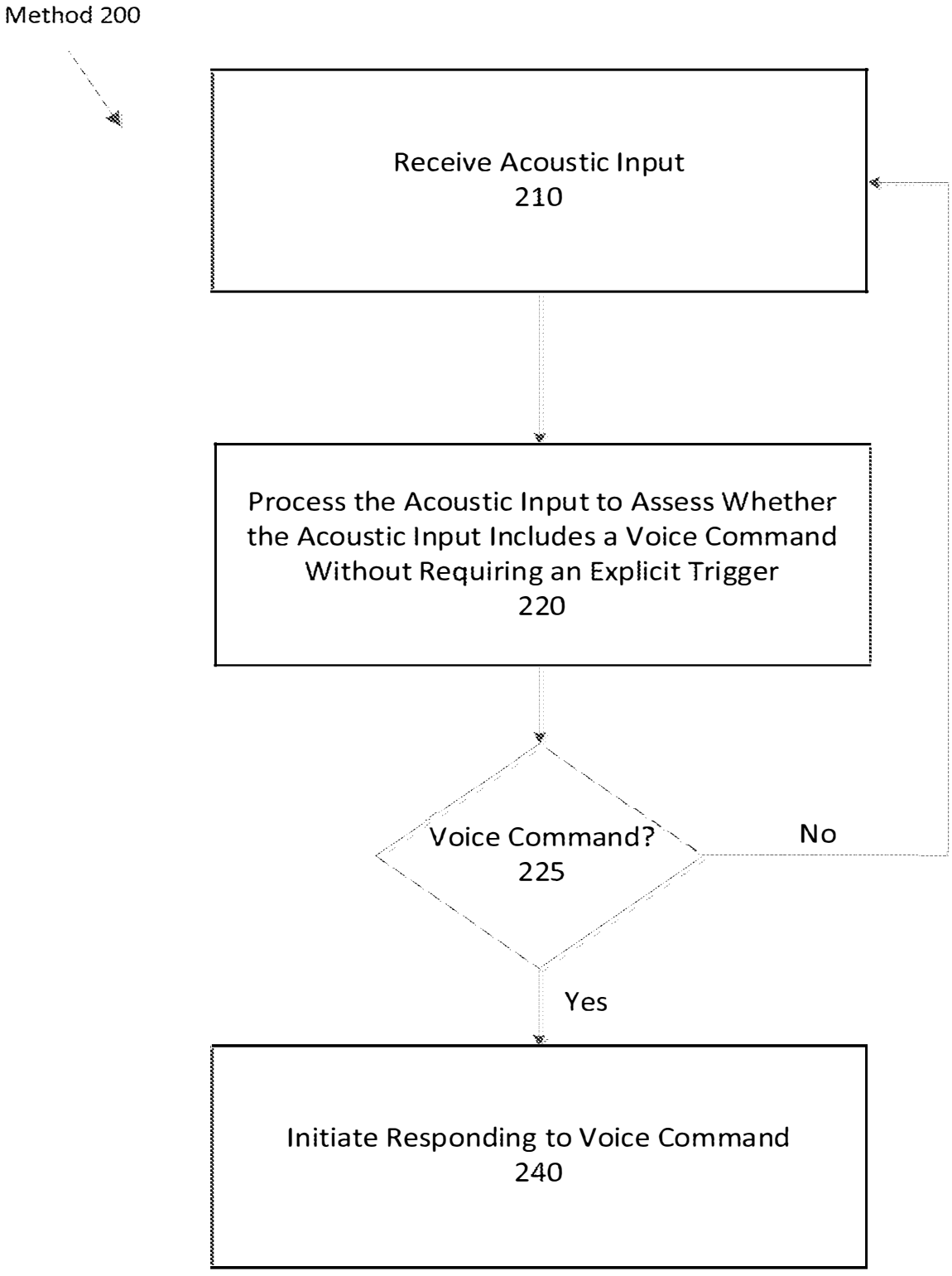

Method 200
Receive Acoustic Input
210
Process the Acoustic Input to Assess Whether the Acoustic Input Includes a Voice Command Without Requiring an Explicit Trigger
220
Voice Command?
225
No
Yes
Initiate Responding to Voice Command
240

FIG. 2

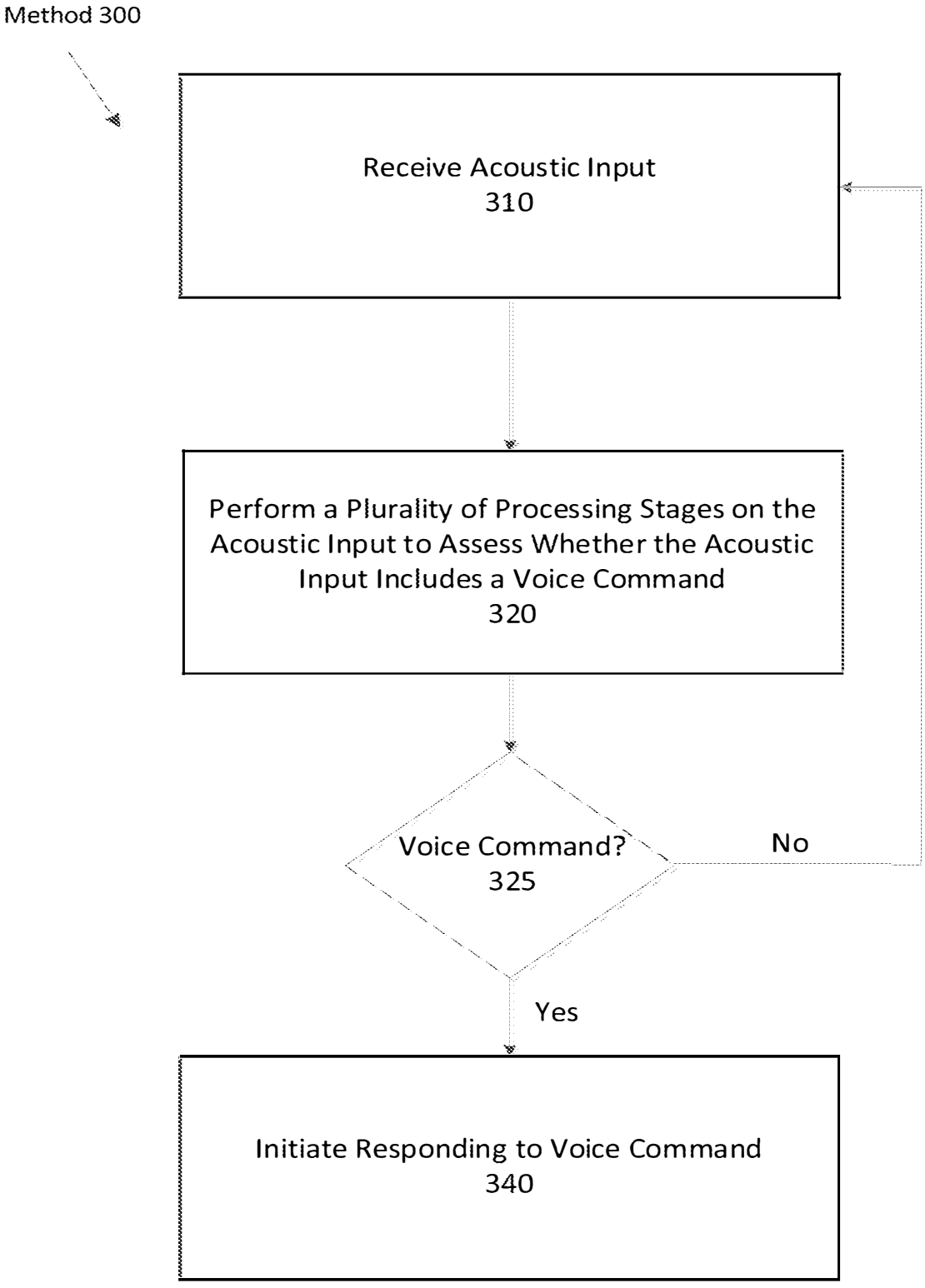

Method 300
Receive Acoustic Input
310
Perform a Plurality of Processing Stages on the Acoustic Input to Assess Whether the Acoustic Input Includes a Voice Command
320
Voice Command?
325
No
Yes
Initiate Responding to Voice Command
340

FIG. 3

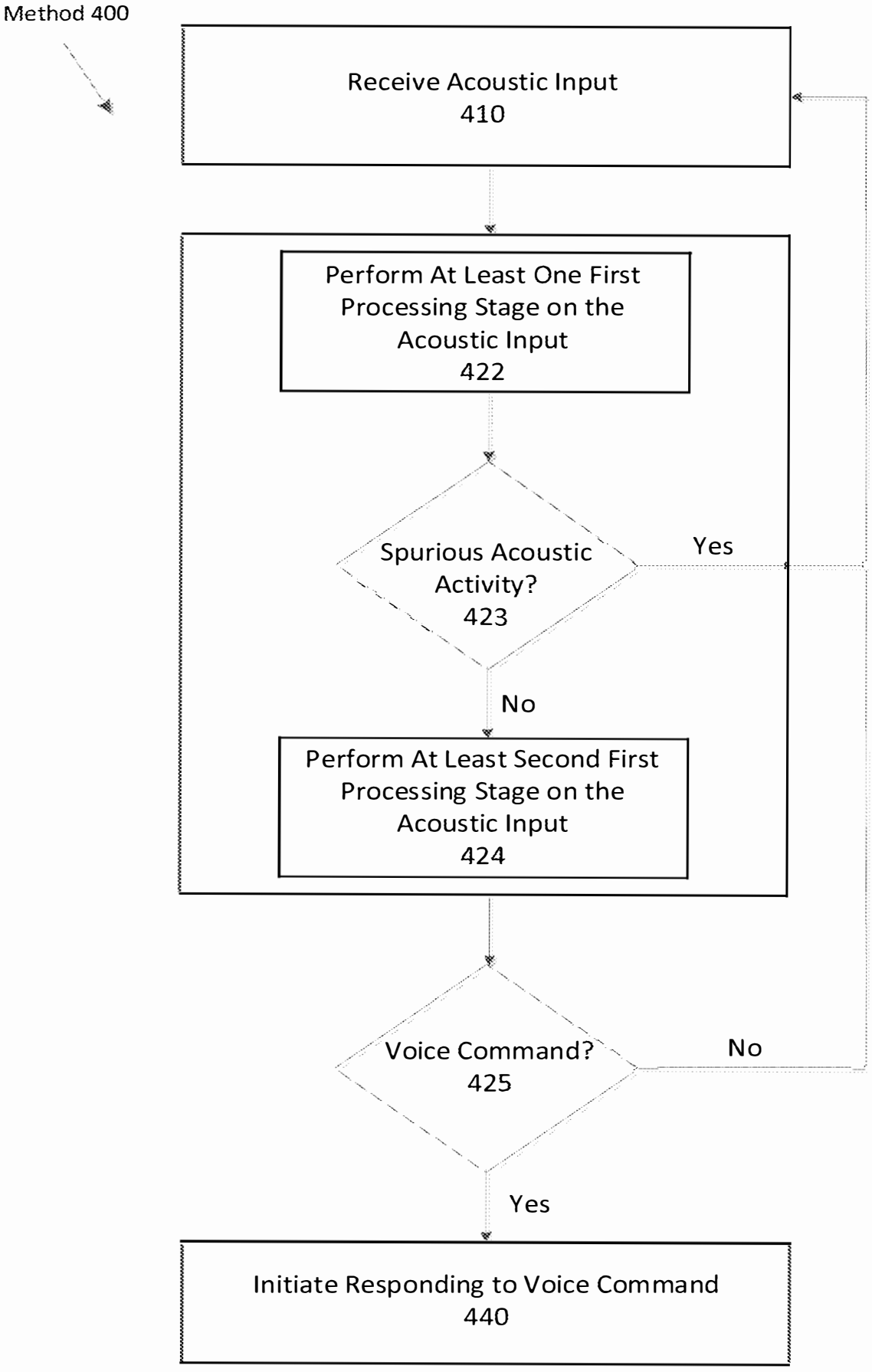

Method 400
Receive Acoustic Input
410
Perform At Least One First Processing Stage on the Acoustic Input
422
Spurious Acoustic Activity?
423
Yes
No
Perform At Least Second First Processing Stage on the Acoustic Input
424
Voice Command?
425
No
Yes
Initiate Responding to Voice Command
440

FIG. 4

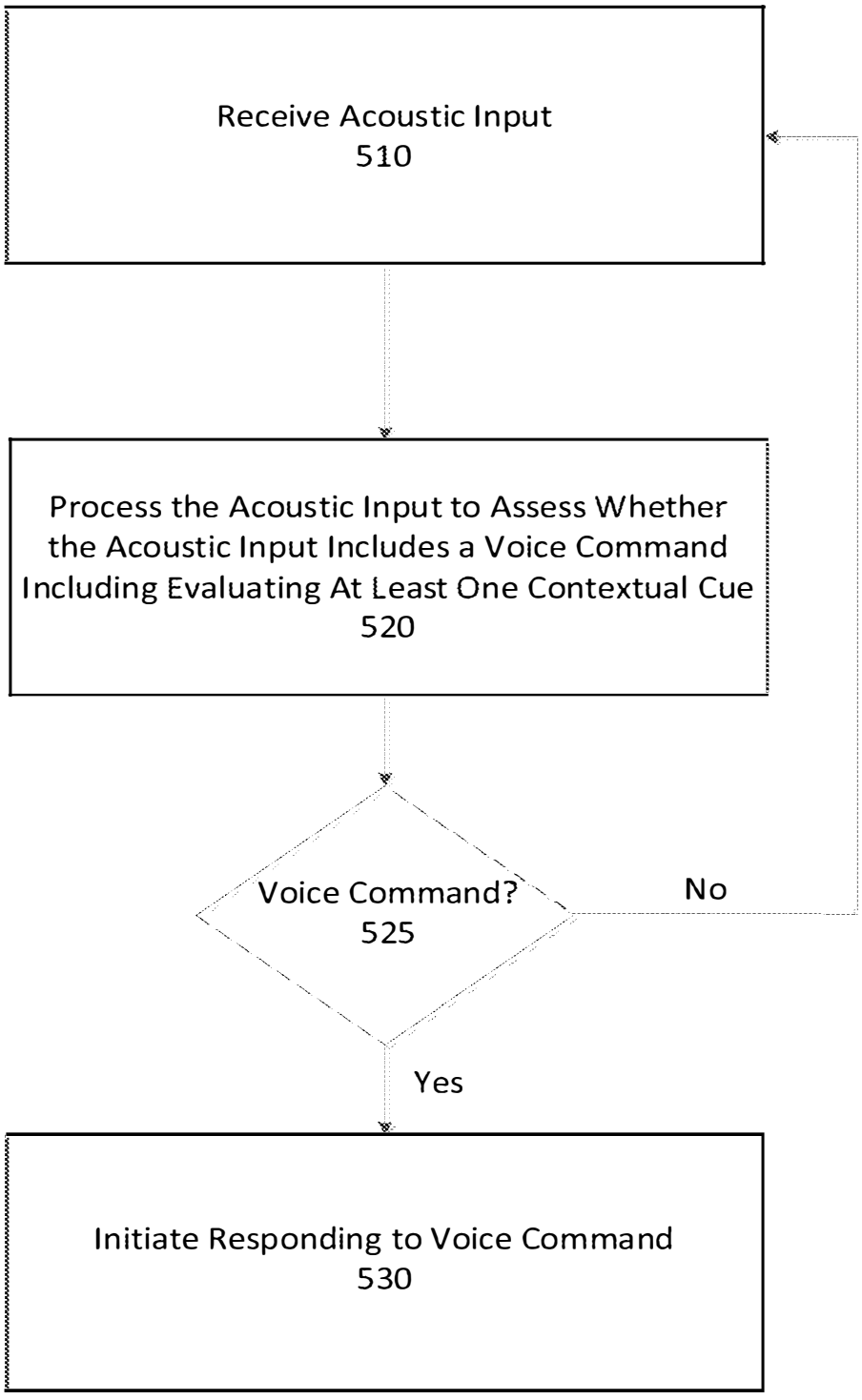

Receive Acoustic Input
510
Process the Acoustic Input to Assess Whether the Acoustic Input Includes a Voice Command Including Evaluating At Least One Contextual Cue
520
Voice Command?
525
No
Yes
Initiate Responding to Voice Command
530

FIG. 5

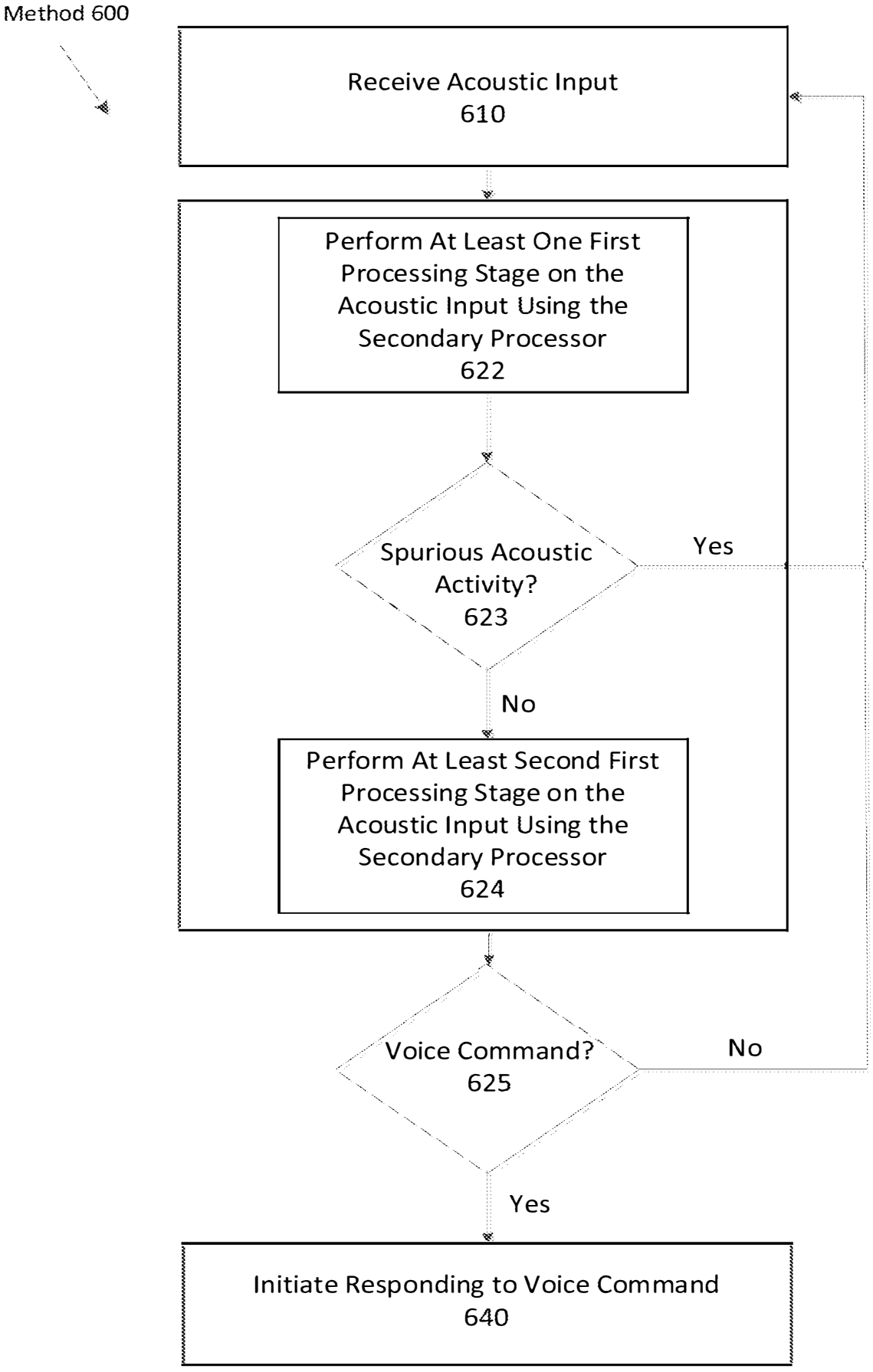

Method 600
Receive Acoustic Input
610
Perform At Least One First Processing Stage on the Acoustic Input Using the Secondary Processor
622
Spurious Acoustic Activity?
623
Yes
No
Perform At Least Second First Processing Stage on the Acoustic Input Using the Secondary Processor
624
Voice Command?
625
No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

FIG. 6

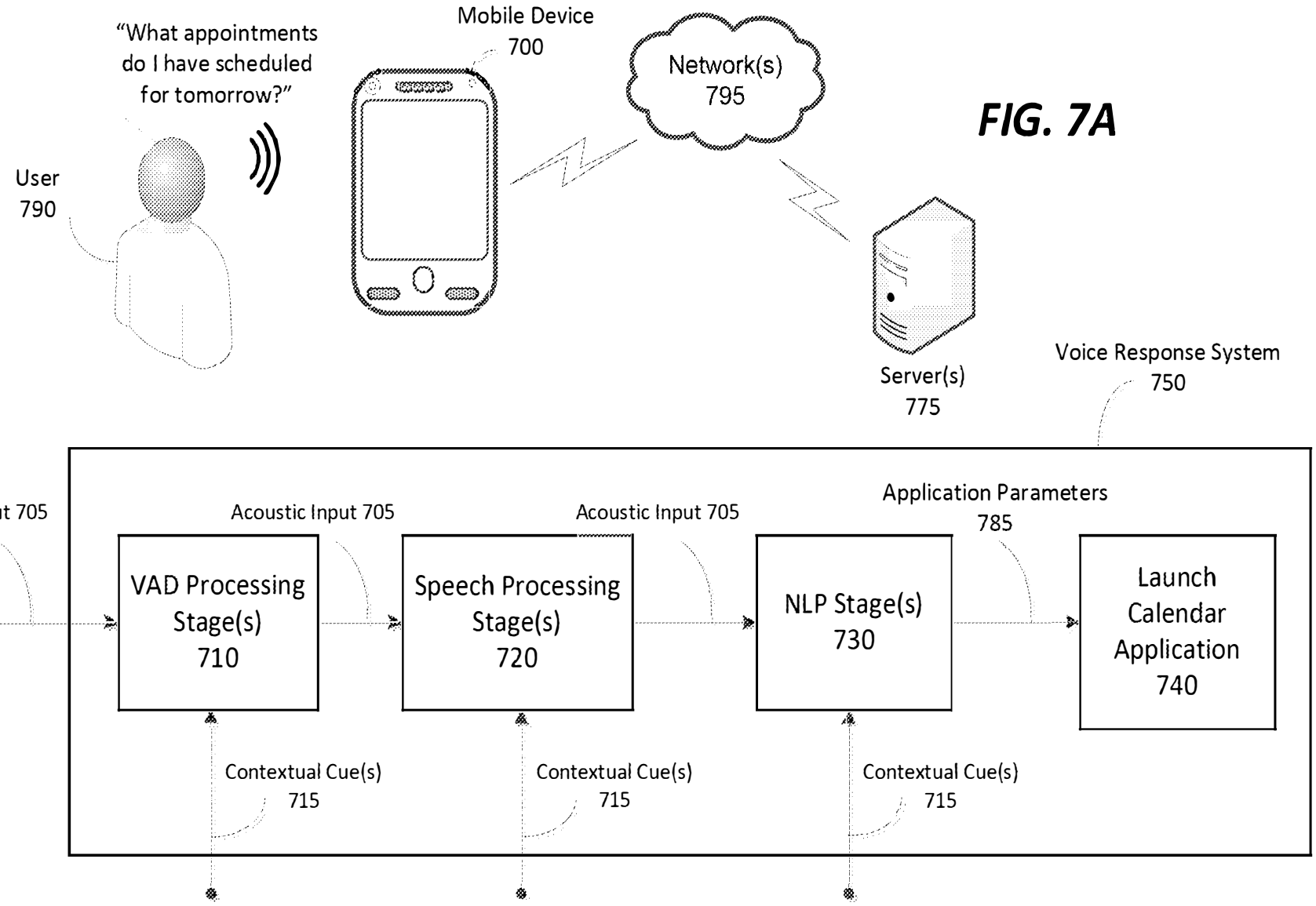

"What appointments do I have scheduled for tomorrow?"
User 790
Mobile Device 700
Network(s) 795
Server(s) 775
Voice Response System 750
Acoustic Input 705
Acoustic Input 705
Acoustic Input 705
Application Parameters 785
VAD Processing Stage(s) 710
Speech Processing Stage(s) 720
NLP Stage(s) 730
Launch Calendar Application 740
Contextual Cue(s) 715
Contextual Cue(s) 715
Contextual Cue(s) 715

FIG. 7A

FIG. 7B

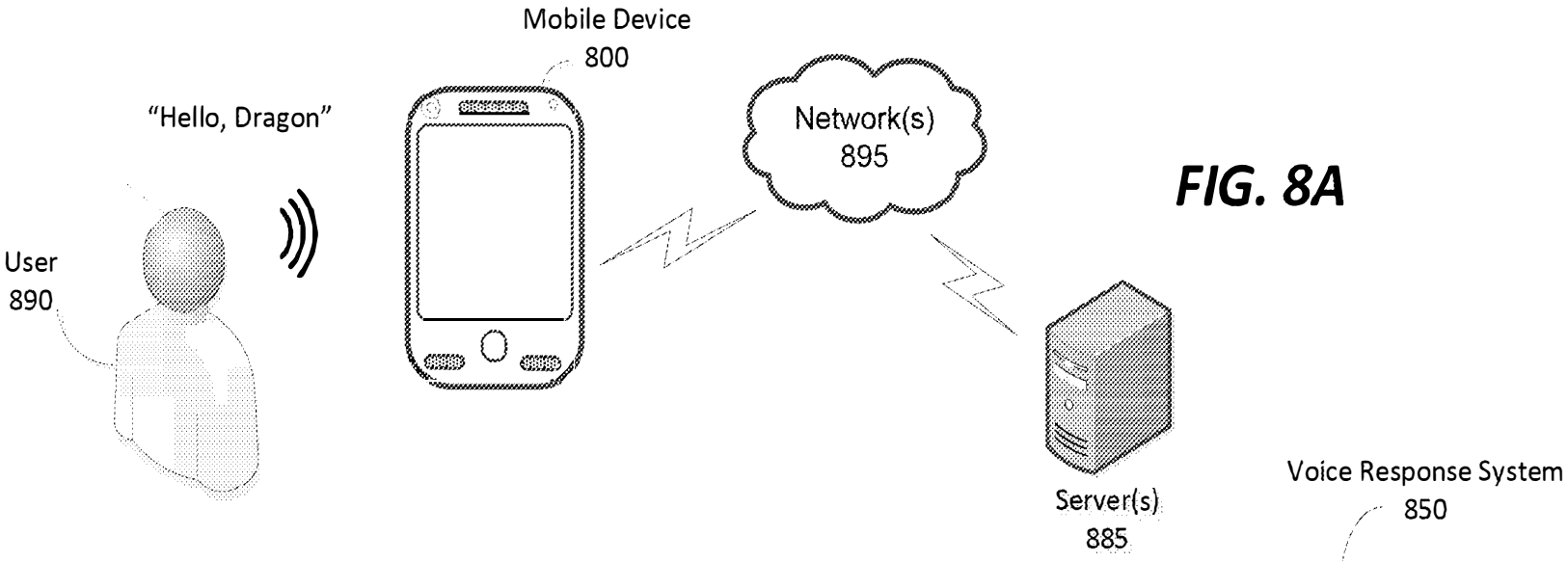

Mobile Device
800
"Hello, Dragon"
Network(s)
895
User
890
Server(s)
885

FIG. 8A

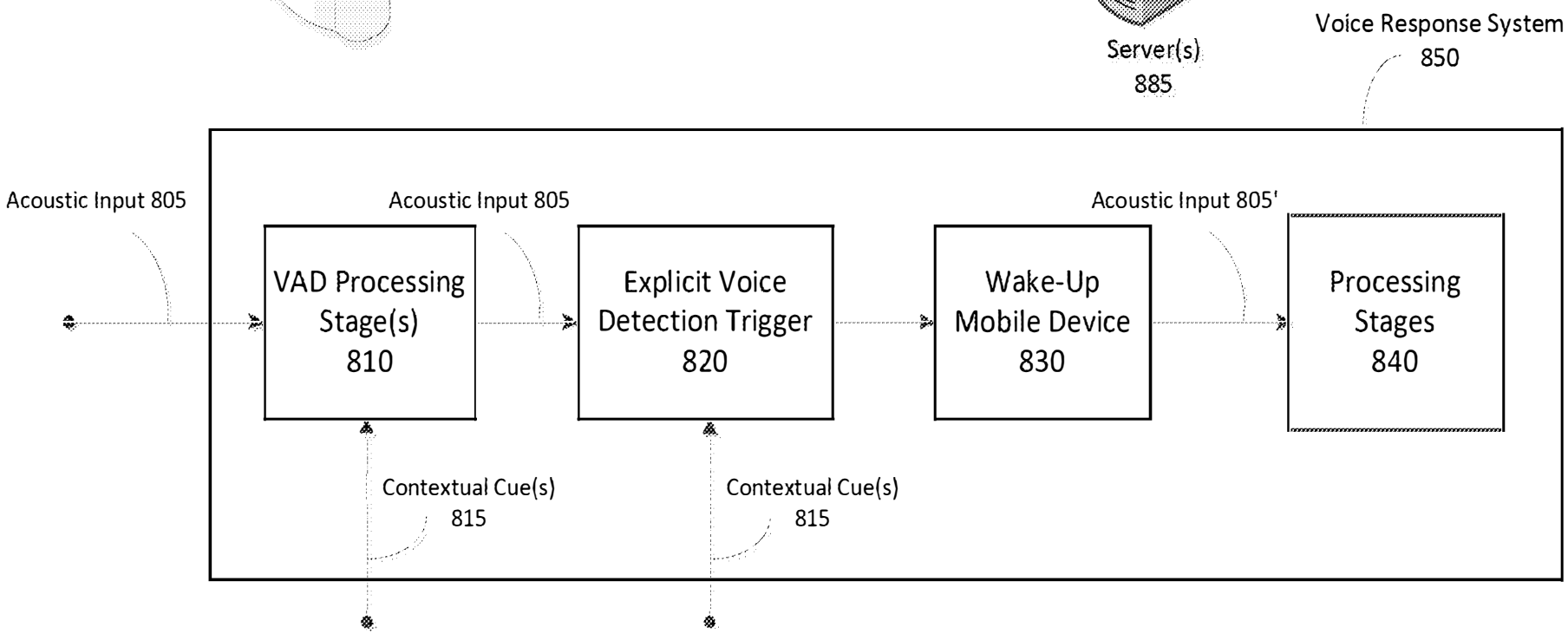

Voice Response System
850
Acoustic Input 805
Acoustic Input 805
Acoustic Input 805'
VAD Processing Stage(s) 810
Explicit Voice Detection Trigger 820
Wake-Up Mobile Device 830
Processing Stages 840
Contextual Cue(s) 815
Contextual Cue(s) 815

FIG. 8B

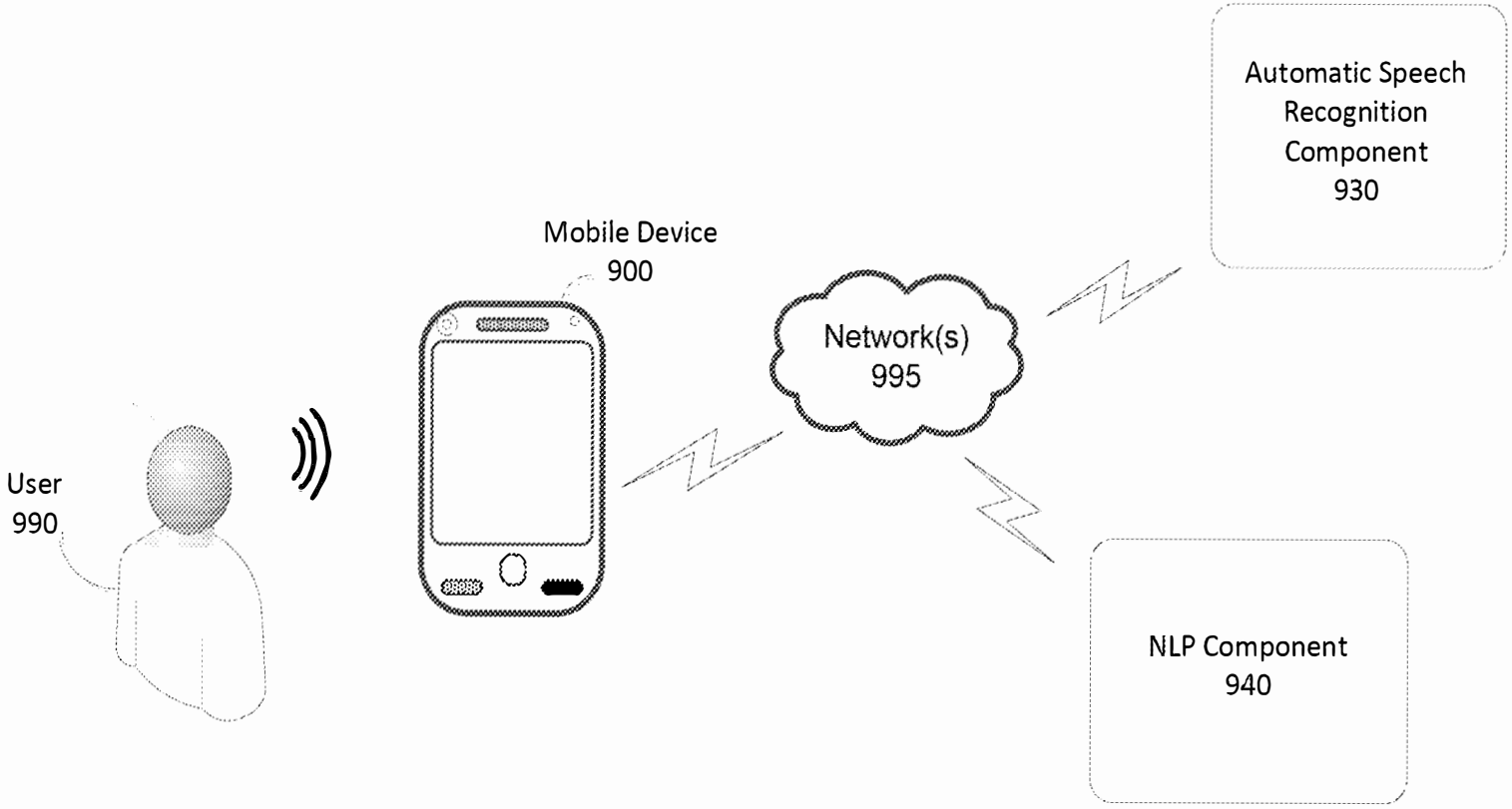

Automatic Speech Recognition Component 930
Mobile Device 900
Network(s) 995
User 990
NLP Component 940

FIG. 9

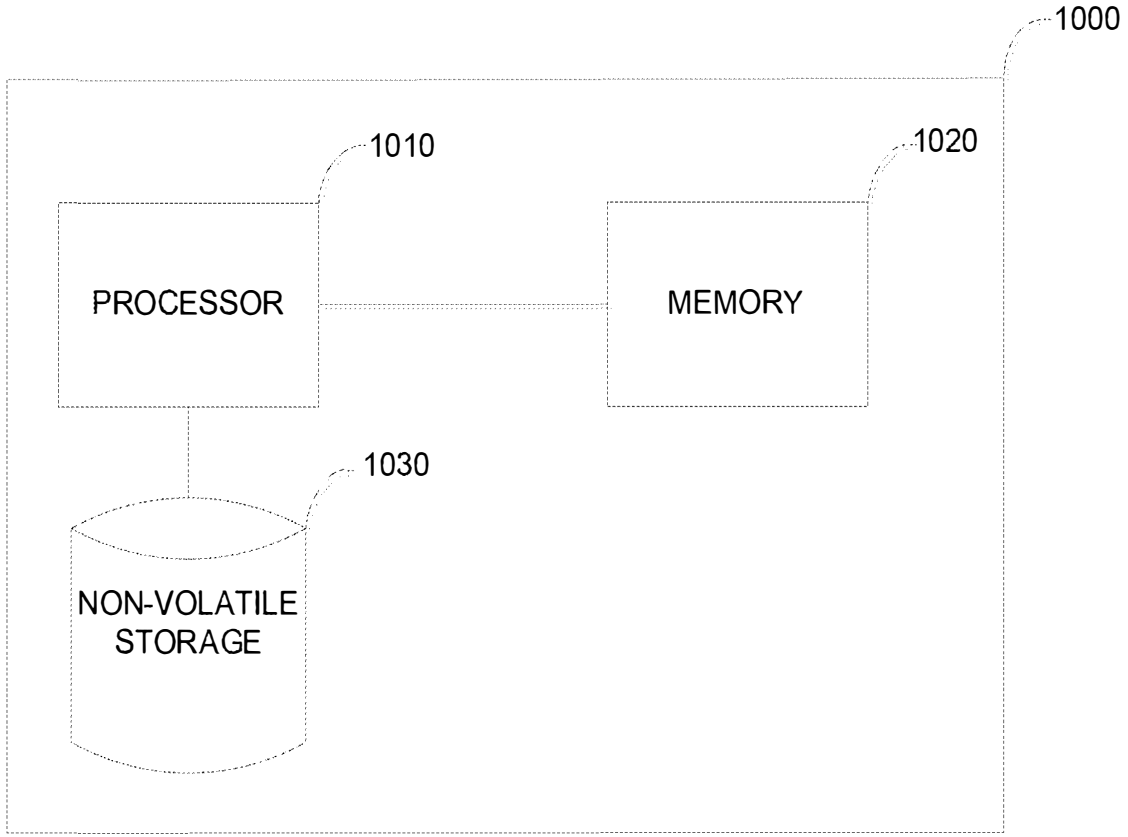


FIG. 10

CLAIMS

1. 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.

2. The method of claim 1, further comprising transitioning the mobile device from the low power mode to an active mode when it is detected that the acoustic input includes a voice command.

3. The method of claim 1, wherein the at least one contextual cue includes at least one of a time the acoustic input was received, motion of the mobile device, location of the mobile device, occurrence of an event on the mobile device, recent activity on the mobile device, a stored sample of the acoustic environment of the mobile device, and a power source of the mobile device.

4. The method of claim 3, wherein the at least one contextual cue includes the time the acoustic input was received, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on the time the acoustic input was received.

5. The method of claim 3, wherein the at least one contextual cue includes recent activity on the mobile device, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on at least one of how recently the

Petitioner Exhibit 1004, Page 81 of 1518

user activated the mobile device, how recently the user issued a voice command and how recently the user performed at least one action on the mobile device.

6. The method of claim 3, wherein the at least one contextual cue includes an occurrence of an event on the mobile device, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on whether the acoustic input is received proximate in time with the occurrence of the event.

7. The method of claim 3, wherein the at least one contextual cue includes motion of the mobile device received as an input from an accelerometer of the mobile device, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on an assessment of whether the user moved the mobile device proximate in time with receiving the acoustic input based, at least in part, on the input from the accelerometer.

8. The method of claim 3, wherein the at least one contextual cue includes the location of the mobile device, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on determining whether the mobile device is located in at least one of a vehicle, a home or a place of work of the user of the mobile device.

9. The method of claim 3, wherein the at least one contextual cue includes the power source of the mobile device, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on at least one of determining whether the mobile device is being powered using the battery or connected to wall power, and a state of the battery if the mobile device is being powered using the battery.

10. The method of claim 3, wherein the at least one contextual cue includes the sample of the acoustic environment obtained by the mobile device by storing acoustic input previously received by the mobile device and determined not to include a voice command, the method further comprising subtracting the sample of the acoustic environment from the acoustic input.

Petitioner Exhibit 1004, Page 82 of 1518

11. The method of claim 1, wherein a sequence of performing the plurality of processing stages is determined, at least in part, by the at least one contextual cue.

12. The method of claim 1, wherein the mobile device includes a primary processor and a secondary processor, and wherein at least one first processing stage of the plurality of processing stages is performed by the secondary processor prior to engaging the primary processor to process the acoustic input.

13. The method of claim 12, wherein at least one second processing stage of the plurality of processing stages is performed by the primary processor only after the at least one first processing stage is performed and neither concludes that the acoustic input includes a voice command nor concludes that the acoustic input is the result of spurious acoustic activity.

14. The method of claim 12, wherein performing the at least one first processing stage includes performing at least one voice activity detection processing stage including performing at least one of 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 phone loop analysis to evaluate whether the acoustic input includes speech.

15. The method of claim 12, 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.

16. The method of claim 12, wherein performing the at least one second processing stage comprises performing at least one speech processing stage including at least one of automatic speech recognition on the acoustic input and at least one natural language processing stage to understand the meaning of the voice command.

Petitioner Exhibit 1004, Page 83 of 1518

17. The method of claim 1, wherein the mobile device is capable of connecting to at least one server via a network, the method further comprising transmitting at least a portion of the acoustic information to the at least one server via the network for processing by the at least one server.

18. The method of claim 1, further comprising performing one or more default actions when a specific task could not be ascertained from the acoustic input.

19. 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.

20. 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.

Petitioner Exhibit 1004, Page 84 of 1518

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 85 of 1518

SCORE Placeholder Sheet for IFW Content

Application Number: 13795933 Document Date: 03/12/2013

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.

- Drawings – Other than Black and White Line Drawings

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.

To access the documents in the SCORE database, refer to instructions developed by SIRA.

At the time of document entry (noted above):

- Examiners may access SCORE content via the eDAN interface.
- Other USPTO employees can bookmark the current SCORE URL (http://es/ScoreAccessWeb/).
- External customers may access SCORE content via the Public and Private PAIR interfaces.

Form Revision Date: February 8, 2006

Electronic Acknowledgement Receipt

EFS ID:	15366173
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Customer Number:	91337
Filer:	Marc S. Johannes/Jennifer Leveille
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US01
Receipt Date:	27-MAR-2013
Filing Date:	
Time Stamp:	14:09:51
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$1740
RAM confirmation Number	459
Deposit Account	232825
Authorized User	JOHANNES, MARC

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

Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.19 (Document supply fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.20 (Post Issuance fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US01-TRN-MSJ.pdf	25437 9a94ea4987fecb234861cbe36304466c702ab453	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US01-FEE-MSJ.pdf	31888 a3763b86a341558c1abac7baeaaf4ce845974593	no	1
Warnings:					
Information:					
3	Applicant Response to Pre-Exam Formalities Notice	N048471090US01-NTMP-MSJ.pdf	23778 5de35f41a5646cc7f0aa6dcbd102bebbef989f82	no	2
Warnings:					
Information:					
4	Oath or Declaration filed	N048471090US01-OATH-MSJ.PDF	313946 c9c4b24c3086f84f24d02ecbc0de24bfdb513ce9	no	2
Warnings:					
Information:					
5	Application Data Sheet	N048471090US01-CADS-MSJ.pdf	21075 b2a72502b082b29587dae18af63b8494664c9a09	no	5
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
6	Fee Worksheet (SB06)	fee-info.pdf	36851 eb646fdc87f1560ae2e323448b92f25957e4f3f0	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	452975		

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:	13795933
Filing Date:	
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Filer:	Marc S. Johannes/Jennifer Leveille
Attorney Docket Number:	N0484.71090US01

Filed as Large Entity

Utility under 35 USC 111(a) Filing Fees

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:				
Late Filing Fee for Oath or Declaration	1051	1	140	140
Petition:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**1740**

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 13/795933 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 2014

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 92 of 1518

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 13/795933 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 93 of 1518

Application No. (if known): 13/795,933 Attorney Docket No.: N0484.71090US01

Certificate of Electronic Filing Under 37 C.F.R. § 1.8

I hereby certify that this correspondence is being transmitted via the Office electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4):

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

on March 27, 2013 .
Date

/Jennifer Leveille/
Signature

Jennifer Leveille
Typed or printed name of person signing Certificate

N/A
Registration Number, if applicable

617.646.8364
Telephone Number

Note: Each paper must have its own certificate of mailing, or this certificate must identify each submitted paper.

Corrected Application Data Sheet (5 pages)

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 person are required to respond to a collection of information unless it displays a valid OMB control number

FEE TRANSMITTAL

Complete if known	
Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Examiner Name	Not Yet Assigned
Art Unit	N/A
Practitioner Docket No.	N0484.71090US01

☐ 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 ($) 1,740.00

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 ($)**	**Fee Paid ($)**
20 - 20 or HP =	_____	x _____	= _____

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

Indep. Claims	**Extra Claims**	**Fee ($)**	**Fee Paid ($)**
3 - 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): 1051 Surcharge - Late filing fee, search fee, ...	140.00

SUBMITTED BY

Signature	/Marc S. Johannes/	Registration No. (Attorney/Agent)	64,978	Telephone	617.646.8000
Name (Print/Type)	Marc S. Johannes			Date	March 27, 2013

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: March 27, 2013 Electronic Signature for Jennifer Leveille: /Jennifer Leveille/

3050812.1

Corrected 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
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
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::	~~Glenn~~Glen
Middle Name::	Edward
Family Name::	Wilson
City of Residence::	~~Burlington~~Maynard
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	~~One Wayside Road~~13 Sherman Street

City of mailing address::	~~Burlington~~Maynard
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	~~01803~~01754

Correspondence Information

Correspondence Customer Number::	91337

Application Information

Application Number::	13/795,933
Filing Date::	03/12/13
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::	N0484.71090US01
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets::	10
Small Entity?::	No
Petition included?::	No

Secrecy Order in Parent Appl.?:: No

Authorization to Permit Access:: No

Representative Information

Representative Customer Number:: 91337

Domestic Priority Information

Foreign Priority Information

Applicant 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

Non-Applicant Assignee Information

Signature:

A signature of the applicant or representative is required in accordance with 37 CFR 1.33 and 10.18. Please see 37 CFR 1.4(d) for the form of the signature.			
Signature	/Marc S. Johannes/	Date	March 27, 2013
Name (Print/Type)	Marc S. Johannes	Registration No. (Attorney/Agent)	64,978

Electronic Acknowledgement Receipt

EFS ID:	15366173
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong, III
Customer Number:	91337
Filer:	Marc S. Johannes/Jennifer Leveille
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US01
Receipt Date:	27-MAR-2013
Filing Date:	
Time Stamp:	14:09:51
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$1740
RAM confirmation Number	459
Deposit Account	232825
Authorized User	JOHANNES, MARC

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

Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.19 (Document supply fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.20 (Post Issuance fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US01-TRN-MSJ.pdf	25437 9a94ea4987fecb234861cbe36304466c702ab453	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	N048471090US01-FEE-MSJ.pdf	31888 a3763b86a341558c1abac7baeaaf4ce845974593	no	1
Warnings:					
Information:					
3	Applicant Response to Pre-Exam Formalities Notice	N048471090US01-NTMP-MSJ.pdf	23778 5de35f41a5646cc7f0aa6dcbd102bebbef989f82	no	2
Warnings:					
Information:					
4	Oath or Declaration filed	N048471090US01-OATH-MSJ.PDF	313946 c9c4b24c3086f84f24d02ecbc0de24bfdb513ce9	no	2
Warnings:					
Information:					
5	Application Data Sheet	N048471090US01-CADS-MSJ.pdf	21075 b2a72502b082b29587dae18af63b8494664c9a09	no	5
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
6	Fee Worksheet (SB06)	fee-info.pdf	36851 eb646fdc87f1560ae2e323448b92f25957e4f3f0	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	452975		

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/SB/21 (07-09)
Approved for use through 07/31/2012. 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.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	11

Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	N0484.71090US01

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
- [x] 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):

Submission of Declaration and Filing Fees (2 pages)
Declaration (2 pages)
Corrected Application Data Sheet (5 pages)

Remarks

Notice to File Missing Parts not yet received.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Marc S. Johannes/		
Printed name	Marc S. Johannes		
Date	March 27, 2013	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 electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: March 27, 2013 Electronic Signature for Jennifer Leveille: /Jennifer Leveille/

3050814.1

Docket No.: N0484.71090US01
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	William F. Ganong, III et al.
Application No.:	13/795,933
Confirmation No.:	2321
Filed:	March 12, 2013
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Not Yet Assigned
Art Unit:	N/A

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: March 27, 2013 Electronic Signature for Jennifer Leveille: /Jennifer Leveille/

SUBMISSION OF DECLARATION AND FILING FEES

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

Dear Madam:

Applicant respectfully submits a Declaration and a Corrected Application Data Sheet.

Please charge our Credit Card in the amount of $1,740.00 covering the fees set forth in 37 C.F.R. § 1.16(f), 1.16(a), 1.16(k), and 1.16(o). 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 (or with any paper hereafter filed in this application by this firm) to our Deposit Account No. 23/2825, under Docket No. N0484.71090US01.

3050816.1

Dated: March 27, 2013

Respectfully submitted,

Electronic signature: /Marc S. Johannes/
Marc S. Johannes
Registration No.: 64,978
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

PTO/SB/21 (07-09)
Approved for use through 07/31/2012. 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.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Art Unit	N/A
Examiner Name	Not Yet Assigned
Total Number of Pages in This Submission: 4	Attorney Docket Number: N0484.71090US01

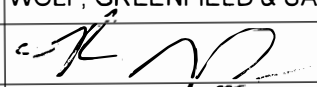
ENCLOSURES (*Check all that apply*)

- [ ] 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
- [x] 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):

Statement Under 37 CFR 3.73(c)

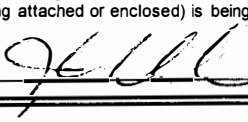
Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature			
Printed name	Richard F. Giunta		
Date	April 4, 2013	Reg. No.	36,149

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: April 4, 2013 Signature:  (Jeanne W. Chub)

3059459.1

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: March 12, 2013

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. [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, III To: Nuance Communications, Inc.
      The document was recorded in the United States Patent and Trademark Office at
      Reel 030096, Frame 0850, or for which a copy thereof is attached.

   2. From: Paul A. Van Mulbregt To: Nuance Communications, Inc.
      The document was recorded in the United States Patent and Trademark Office at
      Reel 030096, Frame 0850, or for which a copy thereof is attached.

3059460.1

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 030096 , Frame 0850 , or for which a copy thereof is attached.

4. From: Glen Wilson To: Nuance Communications, Inc.
The document was recorded in the United States Patent and Trademark Office at
Reel 030096 , Frame 0850 , 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.

Signature	April 4, 2013 Date
Richard F. Giunta Printed or Typed Name	Attorney for Assignee Title or Registration Number

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: April 4, 2013 Signature: _____ (Jeanne W. Chub)

Electronic Acknowledgement Receipt

EFS ID:	15430084
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Richard F. Giunta/Jeanne Chub
Filer Authorized By:	Richard F. Giunta
Attorney Docket Number:	N0484.71090US01
Receipt Date:	04-APR-2013
Filing Date:	12-MAR-2013
Time Stamp:	11:09:22
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	N0484-71090US01-TRN-RFG.pdf	44272 6b▪26832e7515cb3▪a6▪3eb81bf0376bc2▪dd6b1	no	1

Warnings:	
Information:	

2	Power of Attorney	N0484-71090US01-POA-RFG.pdf	52761 a67f0339a143a0b12d0920e1108e4e19711696c3	no	1
Warnings:					
Information:					
3	Assignee showing of ownership per 37 CFR 3.73.	N0484-71090US01-373c-RFG.pdf	96686 a61d05647c24b949f1e230acf60c3ff95cf6953b	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	193719		

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/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

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 13/795,933

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))	3 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

AMENDMENT A

	(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($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
Independent (37CFR 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	

AMENDMENT B

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)	OR	RATE($)	ADDITIONAL FEE($)
Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
Independent (37CFR 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	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
13/795,933	03/12/2013	William F. Ganong III	N0484.71090US01

CONFIRMATION NO. 2321

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

IMPROPER CFR REQUEST

OC000000060224242

Date Mailed: 04/09/2013

RESPONSE TO REQUEST FOR CORRECTED FILING RECEIPT

Continuity, Priority Claims, Petitions, and Non-Publication Requests

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 change the inventorship once the inventorship has been set must be by way of a request under 37 CFR 1.48(a) for a nonprovisional application or 37 CFR 1.48(d) for a provisional application.

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).
- The request to correct inventorship under 37 CFR 1.48(a) filed is deficient because the fee set forth in 37 CFR 1.17(i) has not been submitted.

/hteffera/

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

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
13/795,933	03/12/2013	William F. Ganong III	N0484.71090US01

CONFIRMATION NO. 2321

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

FORMALITIES LETTER

OC000000060216858

Date Mailed: 04/09/2013

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Filing Date Granted

An application number and filing date have been accorded to this application. The application is informal since it does not comply with the regulations for the reason(s) indicated below. Applicant is given TWO MONTHS from the date of this Notice within which to correct the informalities indicated below. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

The required item(s) identified below must be timely submitted to avoid abandonment:

- A substitute specification in compliance with 37 CFR 1.52, 1.121(b)(3), and 1.125, is required. The substitute specification must be submitted with markings and be accompanied by a clean version (without markings) as set forth in 37 CFR 1.125(c) and a statement that the substitute specification contains no new matter (see 37 CFR 1.125(b)). The specification, claims, and/or abstract page(s) submitted is not acceptable and cannot be scanned or properly stored because:
  - The application contains drawings, but the specification does not contain a brief description of the several views of the drawings as required by 37 CFR 1.74 and 37 CFR 1.77(b)(7).
- Replacement drawings in compliance with 37 CFR 1.84 and 37 CFR 1.121(d) are required. The drawings submitted are not acceptable because:
  - More than one figure is present and each figure is not labeled "Fig." with a consecutive Arabic numeral (1, 2, etc.) or an Arabic numeral and capital letter in the English alphabet (A, B, etc.)(see 37 CFR 1.84(u)(1)). See Figure(s) 1. A brief description of the several views of the drawings (see 37 CFR 1.74) should be added or amended to correspond to the corrected numbering of the figures. See also 37 CFR 1.77(b)(7).

Applicant is cautioned that correction of the above items may cause the specification and drawings page count to exceed 100 pages. If the specification and drawings exceed 100 pages, applicant will need to submit the required application size fee.

Replies must be received in the USPTO within the set time period or must include a proper Certificate of Mailing or Transmission under 37 CFR 1.8 with a mailing or transmission date within the set time period. For more information and a suggested format, see Form PTO/SB/92 and MPEP 512.

Replies should be mailed to:

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Registered users of EFS-Web may alternatively submit their reply to this notice via EFS-Web.
https://sportal.uspto.gov/authenticate/AuthenticateUserLocalEPF.html

For more information about EFS-Web please call the USPTO Electronic Business Center at **1-866-217-9197** or visit our website at http://www.uspto.gov/ebc.

If you are not using EFS-Web to submit your reply, you must include a copy of this notice.

/hteffera/

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

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	GRPART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
13/795,933	03/12/2013	2653	1740	N0484.71090US01	20	3

CONFIRMATION NO. 2321

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

FILING RECEIPT

OC000000060216857

Date Mailed: 04/09/2013

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;
Glenn Wilson, Burlington, MA;

Applicant(s)

William F. Ganong III, Brookline, MA;
Paul Adrian Van Mulbregt, Wayland, MA;
Vladimir Sejnoha, Lexington, MA;
Glenn Wilson, Burlington, MA;

Assignment For Published Patent Application

Nuance Communications, Inc., Burlington, MA

Power of Attorney: None

Domestic Applications for which benefit is claimed - None.
A proper domestic benefit claim must be provided in an Application Data Sheet in order to constitute a claim for domestic benefit. See 37 CFR 1.76 and 1.78.

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.

If Required, Foreign Filing License Granted: 04/02/2013

Petitioner Exhibit 1004, Page 116 of 1518

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 13/795,933**

Projected Publication Date: To Be Determined - pending completion of Corrected Papers

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:

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 (1-866-999-4158).

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.

Electronic Acknowledgement Receipt

EFS ID:	15958963
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
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/Jennifer Leveille
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US01
Receipt Date:	05-JUN-2013
Filing Date:	12-MAR-2013
Time Stamp:	16:28:00
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	N048471090US01-TRN-MSJ.pdf	25364 3156d74948cbfb74f18a0f670d2d1801845c7a3c	no	1

Warnings:	
Information:	

2	Applicant Response to Pre-Exam Formalities Notice	N048471090US01-RCAP-MSJ.pdf	23840 bff3247781d9b596577fc055c99695172876 03a7	no	2
Warnings:					
Information:					
3	Miscellaneous Incoming Letter	N048471090US01-STMT-MSJ.pdf	24112 99ea6587b59c146d57bd10053a73adf884f 5bbb0	no	2
Warnings:					
Information:					
4	Drawings-only black and white line drawings	N048471090US01-RDRW-MSJ.PDF	64530 4b61a2e14c7647f113972a58738aa86f942a a8b0	no	1
Warnings:					
Information:					
5	Specification	N048471090US01-MUSPEC-MSJ.pdf	197834 fd64cac40ca646b09e103f3d833a6d3c5fce 3391	no	66
Warnings:					
Information:					
6	Specification	N048471090US01-SUBSPEC-MSJ.pdf	200524 39a8523f9a84dd7cd8e5b4e5559ee126a3e 1e2c5	no	66
Warnings:					
Information:					
		Total Files Size (in bytes):	536204		

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.: N0484.71090US01
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	13/795,933
Confirmation No.:	2321
Filed:	March 12, 2013
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Not Yet Assigned
Art Unit:	2653

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: June 5, 2013 Electronic Signature for Jennifer Leveille: /Jennifer Leveille/

RESPONSE TO NOTICE TO CORRECTED APPLICATION PAPERS

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

Dear Madam:

In response to the Notice to File Corrected Application Papers – Filing Date Granted mailed April 9, 2013 (Notice), a Substitute Specification believed to be in compliance with 37 CFR §§1.52, 1.121(b)(3) and 1.125 and one (1) sheet of Replacement Drawings (Figs. 1A, 1B, 1C, 1D) are herein submitted. Also enclosed is a Marked-Up Specification and a Statement under 37 CFR 1.125(b). In response to the indication in the Notice that more than one view is labeled as "FIG. 1A," one sheet of Replacement Drawings are provided with the multiple views labeled in accordance with 37 C.F.R. 1.84. Amendments in the Substitute Specification were made to comport with the Replacing Drawing.

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. N0484.71090US01.

Dated: June 5, 2013

Respectfully submitted,

Electronic signature: /Marc S. Johannes/
Marc S. Johannes
Registration No.: 64,978
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Docket No.: N0484.71090US01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	13/795,933
Confirmation No.:	2321
Filed:	March 12, 2013
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Not Yet Assigned
Art Unit:	2653

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: June 5, 2013 Electronic Signature for Jennifer Leveille: /Jennifer Leveille/

STATEMENT UNDER 37 CFR §1.125

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

Dear Madam:

In response to the Notice to File Corrected Application Papers mailed April 9, 2013, a substitute specification is enclosed and entry in the above-captioned application is requested.

The enclosed substitute specification, believed to be in compliance with 37 CFR §§1.52, 1.121(b)(3) and 1.125, contains no new matter. A marked-up copy showing the changes as required under 37 CFR §1.125(c) is also enclosed. Please replace the specification with the enclosed substitute specification.

No fee is believed due with this response. However, if a fee is due, please charge our Deposit Account No. 23/2825 under Docket No. N0484.71090US01 from which the undersigned is authorized to draw.

Dated: June 5, 2013

Respectfully submitted,

Electronic signature: /Marc S. Johannes/
Marc S. Johannes
Registration No.: 64,978
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

PTO/SB/21 (07-09)
Approved for use through 07/31/2012. 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.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Art Unit	2653
Examiner Name	Not Yet Assigned
Attorney Docket Number	N0484.71090US01

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)
- [x] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [x] 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):

Substitute Specification - Marked up version
Substitute Specification - Clean version
Statement Under 37 CFR §1.125

Remarks

(1) Sheet of Replacement Drawings (Figs. 1A, 1B, 1C, 1D)

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Marc S. Johannes/		
Printed name	Marc S. Johannes		
Date	June 5, 2013	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 electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: June 5, 2013 Electronic Signature for Jennifer Leveille: /Jennifer Leveille/

3129181.1

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

BACKGROUND

[0001] 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.).

[0002] 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.

[0003] 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

Petitioner Exhibit 1004, Page 126 of 1518

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.

[0004] 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

[0005] 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.

[0006] 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.

[0007] 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.

[0008] 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.

[0009] 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.

[0010] 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

[0011] 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.

[0012] 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.

[0013] 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

[0014] Various aspects and embodiments of the application will be described with reference to the following figures.

[0015] FIGS. 1A, 1B and 1C illustrate ~~FIG. 1 illustrates~~ an exemplary mobile device on which voice response techniques may be implemented;

[0016] FIG. 1D~~1B~~ illustrates exemplary system components of an illustrative mobile device on which voice response techniques may be implemented;

[0017] 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;

[0018] 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;

[0019] 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;

[0020] 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;

[0021] 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;

[0022] 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;

[0023] 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;

[0024] 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

[0025] FIG. 10 illustrates an exemplary computer system on which techniques described herein may be implemented.

DETAILED DESCRIPTION

[0026] 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).

[0027] 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.

[0028] 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 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.

[0029] 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.

[0030] 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).

[0031] 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).

[0032] 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.

[0033] 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.

[0034] 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.

[0035] 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.

[0036] 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.

[0037] 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 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.

[0038] 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.

[0039] 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.

[0040] 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.

[0041] FIGS. 1A-1C illustrate ~~FIG. 1A illustrates~~ 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.

[0042] 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.

[0043] 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.

[0044] 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.

[0045] 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.

[0046] 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.

[0047] 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 activity in the environment and obtain the resulting acoustic input for further processing to assess whether the acoustic input includes a voice command.

[0048] 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.

[0049] 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.

[0050] 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 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.

[0051] 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.

[0052] 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.

[0053] 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 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).

[0054] 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.

[0055] 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.

[0056] 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 selected to generally optimize false positive and false negative rates while attempting to keep power consumption to a minimum.

[0057] 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.

[0058] 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).

[0059] 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.

[0060] 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).

[0061] 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 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.

[0062] 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.

[0063] 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.

[0064] 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

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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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 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).

[0069] 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.

[0070] 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.

[0071] 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 determining its content. Some exemplary VAD and speech processing stages are discussed in further detail below.

[0072] 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.

[0073] 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.

[0074] 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 the primary processor in a low power mode until and/or unless the primary processor is needed.

[0075] 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.

[0076] 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.

[0077] 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. 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.

[0078] 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.

[0079] 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.

[0080] 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 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.

[0081] 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.

[0082] 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.

[0083] 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 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.

[0084] 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.

[0085] 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.

[0086] 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).

[0087] 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.

[0088] 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.

[0089] 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 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.

[0090] 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.

[0091] 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.

[0092] Such classification may be a binary decision, or the classifier (e.g., clusters) may include more numerous classifications. According to some embodiments, 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.

[0093] 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.

[0094] 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.

[0095] 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.

[0096] 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 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).

[0097] 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.

[0098] 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.

[0099] 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 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.

[00100] 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.

[00101] 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).

[00102] 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.

[00103] 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 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.

[00104] 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.

[00105] 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).

[00106] 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 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.

[00107] 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.

[00108] 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.

[00109] 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.).

[00110] 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.

[00111] 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.

[00112] 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."

[00113] 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.

[00114] 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.

[00115] 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 command (e.g., this may be achieved by modifying thresholds in one or more processing stages when a low battery condition is detected).

[00116] 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.

[00117] 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).

[00118] 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.

[00119] Speaker recognition techniques may be used as part of the process of determining whether acoustic input includes an actionable voice command. According 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.

[00120] 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.

[00121] 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.

[00122] 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 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.

[00123] FIG. 1D~~1B~~ 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~~1B~~ is provided to demonstrate some concepts related to generally low power "always-on" functionality, in accordance with some embodiments.

[00124] 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~~1B~~ 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~~1B~~ 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.

[00125] 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 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.

[00126] 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.

[00127] 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.

[00128] 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 acoustic input from analog to digital, perform one or more operations on the acoustic input, etc.

[00129] 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).

[00130] 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.

[00131] 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).

[00132] 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 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.

[00133] 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.

[00134] 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).

[00135] 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, the acoustic input can be processed using network resources without further exhausting the battery power of the mobile device.

[00136] 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.

[00137] 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.

[00138] 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.

[00139] 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.

[00140] 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.

[00141] 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.

[00142] 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 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.

[00143] 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.

[00144] 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.

[00145] 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 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.

[00146] 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.

[00147] 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.

[00148] 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 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.

[00149] 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.

[00150] 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 (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).

[00151] 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.

[00152] 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.

[00153] 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 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.

[00154] 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.

[00155] 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

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

[00156] 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.

[00157] 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.

[00158] 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 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.

[00159] 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.

[00160] 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.

[00161] 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.

[00162] 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 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.

[00163] 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.

[00164] 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 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.

[00165] 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.

[00166] 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 subsequent speech input to act upon or is presently processing any remainder of speech contained in acoustic input 805.

[00167] 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.

[00168] 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.

[00169] 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.

[00170] 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.

[00171] 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.

[00172] 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 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.

[00173] 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).

[00174] 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 implementation. Any of the ASR processing stages described herein may be performed by ASR component 930.

[00175] 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).

[00176] 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.

[00177] 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.

[00178] 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 and/or identify one or more tasks or actions specified in the acoustic input and/or any parameters associated with identified tasks.

[00179] 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).

[00180] 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).

[00181] 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-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.

[00182] 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.

[00183] 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.

[00184] 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 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.

[00185] 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.

[00186] 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).

[00187] 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 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.

[00188] 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.

[00189] 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.

[00190] 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.

[00191] All definitions, as defined and used herein, should be understood to control over dictionary definitions, and/or ordinary meanings of the defined terms.

[00192] 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. 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.

[00193] 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.

[00194] 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).

[00195] 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.

[00196] 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 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:

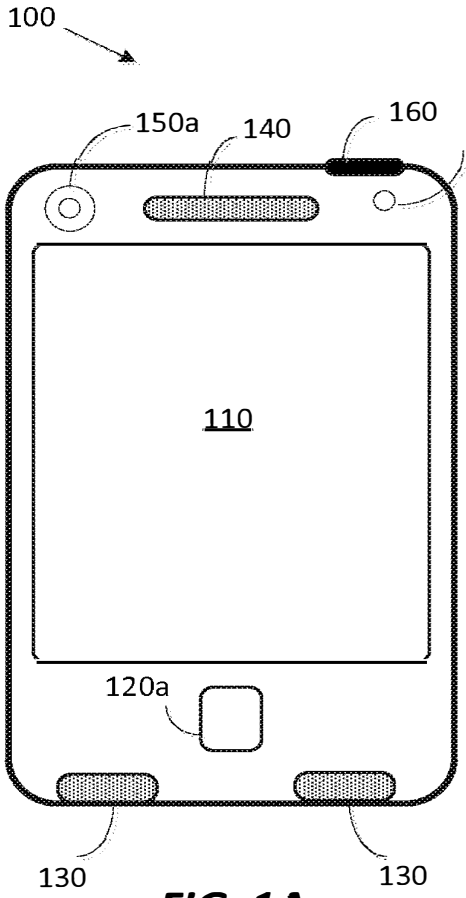


FIG. 1A

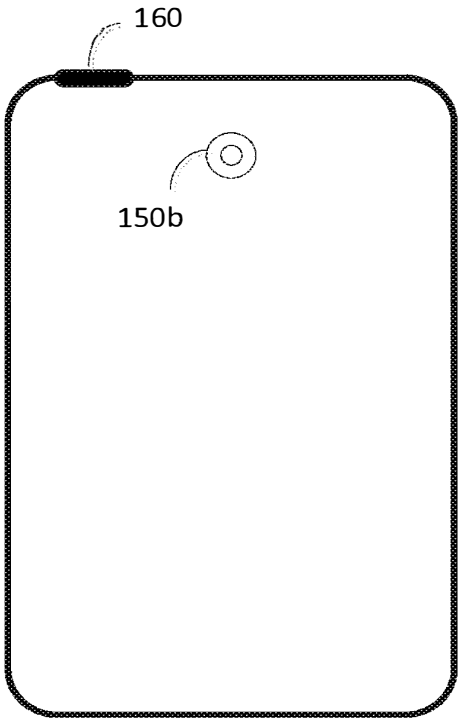


FIG. 1B

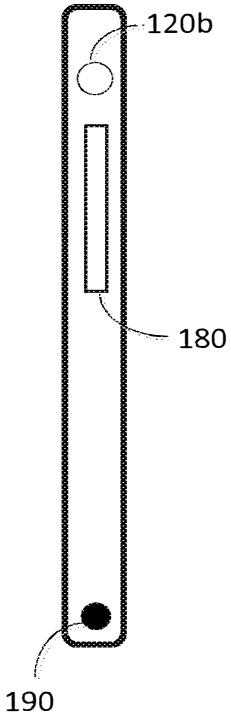


FIG. 1C

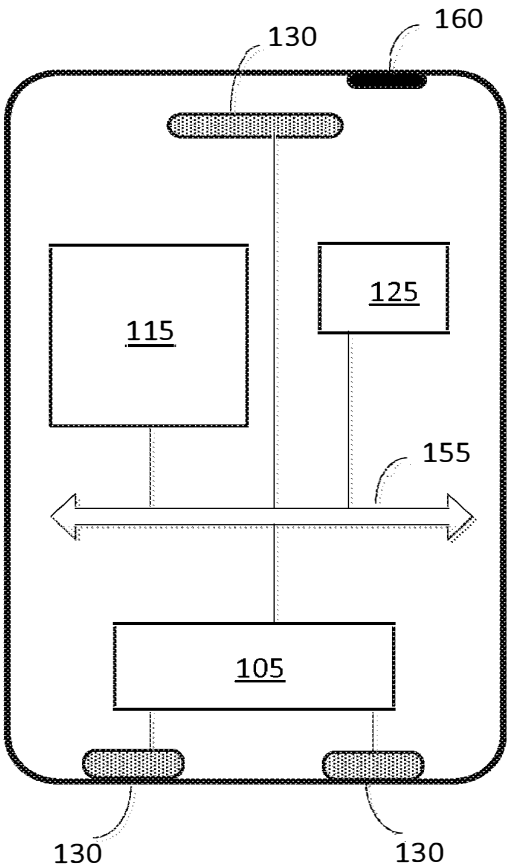


FIG. 1D

METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

BACKGROUND

[0001] 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.).

[0002] 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.

[0003] 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.

[0004] 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

[0005] 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.

[0006] 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.

[0007] 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.

[0008] Some embodiments include a method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is

Petitioner Exhibit 1004, Page 194 of 1518

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.

[0009] 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.

[0010] 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

[0011] 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,

Petitioner Exhibit 1004, Page 195 of 1518

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.

[0012] 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.

[0013] 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

Petitioner Exhibit 1004, Page 196 of 1518

input includes a voice command.

BRIEF DESCRIPTION OF DRAWINGS

[0014] Various aspects and embodiments of the application will be described with reference to the following figures.

[0015] FIGS. 1A, 1B and 1C illustrate an exemplary mobile device on which voice response techniques may be implemented;

[0016] FIG. 1D illustrates exemplary system components of an illustrative mobile device on which voice response techniques may be implemented;

[0017] 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;

[0018] 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;

[0019] 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;

[0020] 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;

[0021] 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;

[0022] 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;

[0023] 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

Petitioner Exhibit 1004, Page 197 of 1518

includes a voice command when the mobile device is operating in a low power mode, in accordance with some embodiments;

[0024] 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

[0025] FIG. 10 illustrates an exemplary computer system on which techniques described herein may be implemented.

DETAILED DESCRIPTION

[0026] 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).

[0027] 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.

[0028] 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

Petitioner Exhibit 1004, Page 198 of 1518

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 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.

[0029] 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.

[0030] 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).

[0031] 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

Petitioner Exhibit 1004, Page 199 of 1518

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).

[0032] 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.

[0033] 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.

Petitioner Exhibit 1004, Page 200 of 1518

[0034] 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.

[0035] 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.

[0036] 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

Petitioner Exhibit 1004, Page 201 of 1518

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.

[0037] 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 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.

[0038] 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.

[0039] 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

Petitioner Exhibit 1004, Page 202 of 1518

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.

[0040] 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.

[0041] 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.

[0042] 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

Petitioner Exhibit 1004, Page 203 of 1518

(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.

[0043] 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.

[0044] 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.

[0045] 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

Petitioner Exhibit 1004, Page 204 of 1518

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.

[0046] 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.

[0047] 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 activity in the environment and obtain the resulting acoustic input for further processing to assess whether the acoustic input includes a voice command.

[0048] 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.

[0049] 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

Petitioner Exhibit 1004, Page 205 of 1518

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.

[0050] 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 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.

[0051] 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,

Petitioner Exhibit 1004, Page 206 of 1518

according to some embodiments, the user is free to simply begin speaking to the mobile device to engage the voice response system.

[0052] 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.

[0053] 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 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).

[0054] 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

Petitioner Exhibit 1004, Page 207 of 1518

evaluate the acoustic input to determine the content of speech contained therein to assess whether the user has spoken a voice command.

[0055] 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.

[0056] 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 selected to generally optimize false positive and false negative rates while attempting to keep power consumption to a minimum.

[0057] 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

Petitioner Exhibit 1004, Page 208 of 1518

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.

[0058] 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).

[0059] 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.

Petitioner Exhibit 1004, Page 209 of 1518

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.

[0060] 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).

[0061] 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 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

Petitioner Exhibit 1004, Page 210 of 1518

determine whether the acoustic input includes a voice command are discussed in further detail below.

[0062] 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.

[0063] 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.

[0064] 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 211 of 1518

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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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

Petitioner Exhibit 1004, Page 212 of 1518

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 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).

[0069] 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.

[0070] 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.

[0071] 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

Petitioner Exhibit 1004, Page 213 of 1518

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

[0072] 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.

[0073] 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.

[0074] 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

Petitioner Exhibit 1004, Page 214 of 1518

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

[0075] 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.

[0076] 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.

Petitioner Exhibit 1004, Page 215 of 1518

[0077] 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. 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.

[0078] 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.

[0079] 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

Petitioner Exhibit 1004, Page 216 of 1518

spoken. Other processing stages beside the illustrative processing stages described herein may be used, as the aspects are not limited in this respect.

[0080] 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 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.

[0081] 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.

[0082] 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

Petitioner Exhibit 1004, Page 217 of 1518

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.

[0083] 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 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.

[0084] 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.

Petitioner Exhibit 1004, Page 218 of 1518

[0085] 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.

[0086] 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).

[0087] 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.

[0088] 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

Petitioner Exhibit 1004, Page 219 of 1518

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.

[0089] 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 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.

[0090] 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

Petitioner Exhibit 1004, Page 220 of 1518

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.

[0091] 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.

[0092] Such classification may be a binary decision, or the classifier (e.g., clusters) may include more numerous classifications. According to some embodiments, 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.

[0093] 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

Petitioner Exhibit 1004, Page 221 of 1518

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.

[0094] 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.

[0095] 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.

[0096] 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 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).

[0097] 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

Petitioner Exhibit 1004, Page 222 of 1518

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.

[0098] 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.

[0099] 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 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.

[00100] 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.

Petitioner Exhibit 1004, Page 223 of 1518

[00101] 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).

[00102] 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.

[00103] 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 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.

Petitioner Exhibit 1004, Page 224 of 1518

[00104] 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.

[00105] 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).

[00106] 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 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

Petitioner Exhibit 1004, Page 225 of 1518

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.

[00107] 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.

[00108] 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.

[00109] 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 226 of 1518

[00110] 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.

[00111] 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.

[00112] 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

Petitioner Exhibit 1004, Page 227 of 1518

"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."

[00113] 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.

[00114] 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.

[00115] 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

Petitioner Exhibit 1004, Page 228 of 1518

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

[00116] 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.

[00117] 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).

[00118] 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

Petitioner Exhibit 1004, Page 229 of 1518

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.

[00119] Speaker recognition techniques may be used as part of the process of determining whether acoustic input includes an actionable voice command. According 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.

[00120] 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.

[00121] 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

Petitioner Exhibit 1004, Page 230 of 1518

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.

[00122] 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 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.

[00123] 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.

[00124] 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

Petitioner Exhibit 1004, Page 231 of 1518

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.

[00125] 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 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.

[00126] 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.

[00127] 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

Petitioner Exhibit 1004, Page 232 of 1518

processor that render the secondary processor advantageous from either a power standpoint, an efficiency standpoint, or both.

[00128] 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 acoustic input from analog to digital, perform one or more operations on the acoustic input, etc.

[00129] 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).

[00130] 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.

Petitioner Exhibit 1004, Page 233 of 1518

[00131] 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).

[00132] 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 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.

[00133] 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.

Petitioner Exhibit 1004, Page 234 of 1518

[00134] 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).

[00135] 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, the acoustic input can be processed using network resources without further exhausting the battery power of the mobile device.

[00136] 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

Petitioner Exhibit 1004, Page 235 of 1518

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.

[00137] 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.

[00138] 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.

[00139] 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.

[00140] 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,

Petitioner Exhibit 1004, Page 236 of 1518

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.

[00141] 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.

[00142] 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 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.

[00143] 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

Petitioner Exhibit 1004, Page 237 of 1518

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.

[00144] 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.

[00145] 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 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.

Petitioner Exhibit 1004, Page 238 of 1518

[00146] 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.

[00147] 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.

[00148] 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 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

Petitioner Exhibit 1004, Page 239 of 1518

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.

[00149] 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.

[00150] 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 (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

Petitioner Exhibit 1004, Page 240 of 1518

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).

[00151] 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.

[00152] 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.

[00153] 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 explicit trigger phrase is not required may operate either by evaluating, but no relying on,

Petitioner Exhibit 1004, Page 241 of 1518

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.

[00154] 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.

[00155] 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 242 of 1518

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

[00156] 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.

[00157] 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.

[00158] 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

Petitioner Exhibit 1004, Page 243 of 1518

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 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.

[00159] 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.

[00160] 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.

[00161] 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.

Petitioner Exhibit 1004, Page 244 of 1518

[00162] 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 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.

[00163] 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.

[00164] Mobile device 800 may include a single processor, or may include multiple processors. When mobile device 800 includes a single processor (or includes

Petitioner Exhibit 1004, Page 245 of 1518

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 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.

[00165] 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

Petitioner Exhibit 1004, Page 246 of 1518

circumstances in which only a single processor is utilized and when multiple processors are utilized.

[00166] 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 subsequent speech input to act upon or is presently processing any remainder of speech contained in acoustic input 805.

[00167] 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.

[00168] 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

Petitioner Exhibit 1004, Page 247 of 1518

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.

[00169] 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.

[00170] 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.

[00171] 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

Petitioner Exhibit 1004, Page 248 of 1518

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.

[00172] 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 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.

[00173] 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).

[00174] 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

Petitioner Exhibit 1004, Page 249 of 1518

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

[00175] 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).

Petitioner Exhibit 1004, Page 250 of 1518

[00176] 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.

[00177] 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,

Petitioner Exhibit 1004, Page 251 of 1518

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.

[00178] 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 and/or identify one or more tasks or actions specified in the acoustic input and/or any parameters associated with identified tasks.

[00179] 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

Petitioner Exhibit 1004, Page 252 of 1518

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).

[00180] 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).

[00181] 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-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.

[00182] 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

Petitioner Exhibit 1004, Page 253 of 1518

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.

[00183] 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.

[00184] 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 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.

[00185] 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

Petitioner Exhibit 1004, Page 254 of 1518

functionality to provide data, and may include control apparatus to perform I/O functionality.

[00186] 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).

[00187] 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 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.

[00188] 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

Petitioner Exhibit 1004, Page 255 of 1518

functionality of the program modules may be combined or distributed as desired in various embodiments.

[00189] 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.

[00190] 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.

[00191] All definitions, as defined and used herein, should be understood to control over dictionary definitions, and/or ordinary meanings of the defined terms.

[00192] 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. 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

Petitioner Exhibit 1004, Page 256 of 1518

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.

[00193] 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.

[00194] 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).

[00195] 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.

[00196] 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 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.

Petitioner Exhibit 1004, Page 257 of 1518

What is claimed is:

Petitioner Exhibit 1004, Page 258 of 1518

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	GRPART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
13/795,933	03/12/2013	2653	1740	N0484.71090US01	20	3

CONFIRMATION NO. 2321

UPDATED FILING RECEIPT

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



Date Mailed: 06/12/2013

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;
Glenn Wilson, Burlington, MA;

Applicant(s)

William F. Ganong III, Brookline, MA;
Paul Adrian Van Mulbregt, Wayland, MA;
Vladimir Sejnoha, Lexington, MA;
Glenn Wilson, Burlington, MA;

Assignment For Published Patent Application

Nuance Communications, Inc., Burlington, MA

Power of Attorney: None

Domestic Applications for which benefit is claimed - None.
A proper domestic benefit claim must be provided in an Application Data Sheet in order to constitute a claim for domestic benefit. See 37 CFR 1.76 and 1.78.

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.

If Required, Foreign Filing License Granted: 04/02/2013

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 13/795,933**

Projected Publication Date: 09/18/2014

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:

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.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 13/795,933

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))	3 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 (37CFR 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	RATE($)	ADDITIONAL FEE($)		RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37CFR 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.

Application No. (if known): 13/795,933 Attorney Docket No.: N0484.71090US01

Certificate of Electronic Filing Under 37 C.F.R. § 1.8

I hereby certify that this correspondence is being transmitted via the Office electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4):

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

on April 7, 2014 .
Date

/Stefanie M. Grenier/
Signature

Stefanie M. Grenier
Typed or printed name of person signing Certificate

Registration Number, if applicable

617.646.8331
Telephone Number

Note: Each paper must have its own certificate of mailing, or this certificate must identify each submitted paper.

Corrected Application Data Sheet (6 pages)

Docket No.: N0484.71090US01
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	13/795,933
Confirmation No.:	2321
Filed:	March 12, 2013
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Not Yet Assigned
Art Unit:	N/A

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: April 7, 2014 Electronic Signature for Stefanie M. Grenier: /Stefanie M. Grenier/

REQUEST TO CORRECT INVENTORSHIP UNDER 37 C.F.R. § 1.48(a)

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

Dear Madam:

Applicant hereby requests under 37 CFR 1.48(a) to correct the name and residential information of an inventor in the above-identified application. The application was filed listing inventor Glenn Wilson, whose legal name should appear as "Glen Edward Wilson". A corrected Application Data Sheet showing the correct name and residential information of inventor Glen Edward Wilson is being filed herewith. In addition, the corrected Application Data Sheet also includes correction of the Applicant in the Application information section. The Applicant should appear as "Nuance Communications, Inc."

Applicant requests that an updated Filing Receipt be issued indicating this change in inventorship and Applicant. Applicant additionally requests that all pertinent U.S. Patent and Trademark Office records relating to the subject application be changed to reflect this correction

The required processing fee of $140.00 as set forth in 37 C.F.R. § 1.17(i) is being paid using the credit card payment option of the USPTO's Electronic Filing System. If additional fees are due, please charge them to our Deposit Account No. 23/2825, under Docket No. N0484.71090US01, from which the undersigned is authorized to draw.

Dated: April 7, 2014

Respectfully submitted,

Electronic signature: /Marc S. Johannes/
Marc S. Johannes
Registration No.: 64,978
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

Corrected 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::	~~Glenn~~Glen
Middle Name::	Edward
Family Name::	Wilson
City of Residence::	~~Burlington~~Maynard
State or Province of Residence::	MA

Corrected 13795933 03/12/2013

3499024.1

Country of Residence::	US
Street of mailing address::	~~One Wayside Road~~13 Sherman Street
City of mailing address::	~~Burlington~~Maynard
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	~~01803~~01754

Correspondence Information

Correspondence Customer Number::	91337

Application Information

Application Number::	13/795,933
Filing Date::	03/12/13
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::	N0484.71090US01
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 Order pursuant to 37 CFR 5.2::	No
Authorization to Permit Access to the Instant Application by the Participating Offices::	No

Representative Information

Representative Customer Number::	91337

Domestic Priority Information

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~~

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

Signature:

A signature of the applicant or representative is required in accordance with 37 CFR 1.33 and 10.18. Please see 37 CFR 1.4(d) for the form of the signature.			
Signature	/Marc S. Johannes/	Date	April 7, 2014
Name (Print/Type)	Marc S. Johannes	Registration No. (Attorney/Agent)	64,978

PTO/SB/21 (07-09)
Approved for use through 07/31/2012. 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.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	14

Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Art Unit	N/A
Examiner Name	Not Yet Assigned
Attorney Docket Number	N0484.71090US01

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
- [ ] 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
- [ ] 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 to Correct Inventorship Under 37 C.F.R. § 1.48(a)
- Corrected Application Data Sheet
-Statement Under 37 CFR 3.73(c)

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Marc S. Johannes/		
Printed name	Marc S. Johannes		
Date	April 7, 2014	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 electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: April 7, 2014 Electronic Signature for Stefanie M. Grenier: /Stefanie M. Grenier/

3509533.1

Electronic Acknowledgement Receipt

EFS ID:	18691592
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
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/Stefanie Grenier
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US01
Receipt Date:	07-APR-2014
Filing Date:	12-MAR-2013
Time Stamp:	15:19:41
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$140
RAM confirmation Number	1883
Deposit Account	232825
Authorized User	JOHANNES, MARC

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

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	N048471090US01-TRN-MSJ.pdf	25449 9da3762759e809990def1b4899c1a42afbba5fab	no	1
Warnings:					
Information:					
2	Request under Rule 48 correcting inventorship	N048471090US01-RTCI-MSJ.pdf	26134 ac3d34dde44a5711736f48bd0a23580c53ded4b6	no	2
Warnings:					
Information:					
3	Application Data Sheet	N048471090US01-CADS-MSJ.pdf	24409 71f2d1c40da60a8210d90ad5cd7a6eafc53c7315	no	7
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
4	Power of Attorney	N048471090US01-POA-MSJ.pdf	140020 1ec1dcc8518c9d90f6964d737390af565981b2ef	no	1
Warnings:					
Information:					
5	Assignee showing of ownership per 37 CFR 3.73.	N048471090US01-373-MSJ.pdf	264892 9ad3349dda005287158b99958b43be2300c81275	no	2
Warnings:					
Information:					
6	Fee Worksheet (SB06)	N048471090US01-FEE-MSJ.pdf	29415 73500e67a8a251254f4d69b32207e6f6c302f4fe	no	1
Warnings:					
Information:					
7	Fee Worksheet (SB06)	fee-info.pdf	30237 a44a2169cc7d2fa184e0f894c2e0e8d27dea1726	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	540556		

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/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: March 12, 2013

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, III To: Nuance Communications, Inc.
      The document was recorded in the United States Patent and Trademark Office at
      Reel 030096, Frame 0850, or for which a copy thereof is attached.

   2. From: Paul A. Van Mulbregt To: Nuance Communications, Inc.
      The document was recorded in the United States Patent and Trademark Office at
      Reel 030096, Frame 0850, or for which a copy thereof is attached.

3059460.1

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 030096 , Frame 0850 , or for which a copy thereof is attached.

4. From: Glen Wilson To: Nuance Communications, Inc.
The document was recorded in the United States Patent and Trademark Office at
Reel 030096 , Frame 0850 , 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.

Signature	April 4, 2013 Date
Richard F. Giunta Printed or Typed Name	Attorney for Assignee Title or Registration Number

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: April 4, 2013 Signature: _____ (Jeanne W. Chub)

PTO/SB/17i (03-13)
Approved for use through 03/31/2013. 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.

PROCESSING FEE Under 37 CFR 1.17(i) TRANSMITTAL

(Fees are subject to annual revision)

Send completed form to: Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Art Unit	N/A
Examiner Name	Not Yet Assigned
Practitioner Docket No.	N0484.71090US01

Enclosed is a paper filed under 37 CFR 1.48 **that requires a processing fee (37 CFR 1.17(i)(1) or (2)).**
Payment of $ 140.00 **is enclosed.**

This form should be included with the above-mentioned paper and faxed, or mailed to the Office using the appropriate Mail Stop, if applicable. *For transmittal of petition fees under 37 CFR 1.17(f), (g) or (h), see form PTO/SB/17p.*

[ ] 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.

Payment of Fees

[X] The Commissioner is hereby authorized to charge the following fees to Deposit Account No. 23/2825 :

[X] processing fee under 37 CFR 1.17(i)(1) [ ] processing fee under 37 CFR 1.17(i)(2) [X] any deficiency of fees and credit of any overpayments

[ ] Check in the amount of $ _____ is enclosed. [X] Payment via EFS-Web.

[X] Payment by credit card (~~Form PTO-2038 or equivalent is enclosed~~). Do not provide credit card information on this form.

Processing Fees under 37 CFR 1.17(i)(1):

Undiscounted Fee $140 (Fee Code 1053) Small Entity Fee $70 (Fee Code 2053) Micro Entity Fee $35 (Fee Code 3053)

For papers filed under:
§ 1.28(c)(3) – for processing a nonitemized fee deficiency based on an error in small entity status.
§ 1.41 – for supplying the name or names of the inventor or inventors after the filing date without an oath or declaration as prescribed by §1.63, except in provisional applications.
§ 1.48 – for correcting inventorship, except in provisional applications.
§ 1.52(d) – for processing a nonprovisional application filed with a specification in a language other than English.
§ 1.53(b)(3) – to convert a provisional application filed under § 1.53(c) into a nonprovisional application under §1.53(b).
§ 1.55 – for entry of late priority papers.
§ 1.71(g)(2) for processing a belated amendment under § 1.71(g).
§ 1.99(e) – for processing a belated submission under § 1.99.
§ 1.102(e) – for requesting prioritized examination of an application.
§ 1.103(b) – for requesting limited suspension of action, continued prosecution application for a design patent (§ 1.53(d)).
§ 1.103(c) – for requesting limited suspension of action, request for continued examination (§ 1.114).
§ 1.103(d) – for requesting deferred examination of an application.
§ 1.291(c)(5) - for processing a second or subsequent protest by the same real party in interest.
§ 1.497(d) – for filing an oath or declaration pursuant to 35 U.S.C. 371(c)(4) naming an inventive entity different from the inventive entity set forth in the international stage
§ 3.81 – for a patent to issue to assignee, assignment submitted after payment of the issue fee.

Processing Fees under 37 CFR 1.17(i)(2):

Fee $130 (small and micro entity discounts are not available for the 37 CFR 1.17(i)(2) processing fee)

For papers filed under:
§ 1.217 – for processing a redacted copy of a paper submitted in the file of an application in which a redacted copy was submitted for the patent application publication. **FEE CODE 1808**
§ 1.221 – for requesting voluntary publication or republication of an application. **FEE CODE 1803**

SUBMITTED BY

Signature	/Marc S. Johannes/	Registration No. (Attorney/Agent)	64,978	Telephone	617.646.8000
Name (Print/Type)	Marc S. Johannes			Date	April 7, 2014

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: April 7, 2014 Electronic Signature for Stefanie M. Grenier: /Stefanie M. Grenier/

3508150.1

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

Electronic Patent Application Fee Transmittal

Application Number:	13795933
Filing Date:	12-Mar-2013
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Filer:	Marc S. Johannes/Stefanie Grenier
Attorney Docket Number:	N0484.71090US01

Filed as Large Entity

Utility under 35 USC 111(a) Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
PROCESSING FEE, EXCEPT PROV. APPLS.	1830	1	140	140
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
			Total in USD ($)	**140**

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	GRPART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
13/795,933	03/12/2013	2644	1740	N0484.71090US01	20	3

CONFIRMATION NO. 2321

REPLACEMENT FILING RECEIPT

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

OC000000067662888

Date Mailed: 04/10/2014

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 Wilson, Maynard, MA;

Applicant(s)

Nuance Communications, Inc., Burlington, MA

Assignment For Published Patent Application

Nuance Communications, Inc., Burlington, MA

Power of Attorney: The patent practitioners associated with Customer Number 91337

Domestic Applications for which benefit is claimed - None.
A proper domestic benefit claim must be provided in an Application Data Sheet in order to constitute a claim for domestic benefit. See 37 CFR 1.76 and 1.78.

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.

If Required, Foreign Filing License Granted: 04/02/2013
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 13/795,933**
Projected Publication Date: 09/18/2014
Non-Publication Request: No

Petitioner Exhibit 1004, Page 282 of 1518

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:

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
13/795,933	03/12/2013	William F. Ganong III	N0484.71090US01

CONFIRMATION NO. 2321

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

POA ACCEPTANCE LETTER

OC000000067662944

Date Mailed: 04/10/2014

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 04/07/2014.

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.

/dgela/

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

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 13/795,933	ATTY. DOCKET NO.: N0484.71090US01
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: March 12, 2013	CONFIRMATION NO.: 2321
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: Matthew W. Genack

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
		7,720,683		Vermeulen et al.	05-18-2010
		7,567,827		Kim	07-28-2009
		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
		8,666,751		Murthi et al.	03-04-2014
		2010-0185448		Meisel	07-22-2010
		2011-0054899		Phillips et al.	03-03-2011
		2013-0080167		Mozer	03-28-2013
		2013-0080171		Mozer et al.	03-28-2013
		2013-0289994		Newman et al.	10-31-2013
		2013-0339028		Rosner et al.	12-19-2013
		2014-0012586		Rubin et al.	01-09-2014
		2014-0012573		Hung et al.	01-09-2014

[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.

Electronic Acknowledgement Receipt

EFS ID:	19041887
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
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/Sara Douglas
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US01
Receipt Date:	15-MAY-2014
Filing Date:	12-MAR-2013
Time Stamp:	15:54:11
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		3564088_1-N048471090US01-IDS-MSJ.PDF	33489 4f78aa[redacted]b3fa5718651705e1[redacted]fa8a[redacted]a2a8 3b00	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:			

2	Miscellaneous Incoming Letter	N048471090US01-TRN-MSJ.PDF	26252 5546564dd9271cb9e2e030397658774019 23ffdf	no	1
Warnings:					
Information:					

Total Files Size (in bytes):	59741

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.: N0484.71090US01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	13/795,933
Confirmation No.:	2321
Filed:	March 12, 2013
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	Matthew W. Genack
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 15th day of May, 2014.

__/Sara J. L. Douglas/_____
Sara J. L. Douglas

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

Madam:

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

Petitioner Exhibit 1004, Page 289 of 1518

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.
*13/795829	03-12-2013	Sejnoha et al.	N0484.71090US00
*13/795,960	03-12-2013	Ganong et al.	N0484.71090US02

*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 co-pending applications identified above may include subject matter which 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 the other applications will assumedly search for relevant prior art and, in some circumstances, 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 undersigned respectfully requests that the Examiner review the file histories of the above-identified 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, and the undersigned does not plan to provide copies of items from these files. However, if the Examiner would like copies of any or all of the information included in the file histories of the above-identified applications, or any other information related to the present application, the Examiner is respectfully requested to contact the undersigned.

Petitioner Exhibit 1004, Page 290 of 1518

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 291 of 1518

An early and favorable action is hereby requested.

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. N0484.71090US01 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Marc S. Johannes/

Marc S. Johannes, Reg. No. 64,978
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: N0484.71090US01
Date: May 15, 2014
xNDDx

PTO/SB/21 (07-09)
Approved for use through 07/31/2012. 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.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	Matthew W. Genack
Attorney Docket Number	N0484.71090US01

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	/Marc S. Johannes/		
Printed name	Marc S. Johannes		
Date	May 15, 2014	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 electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: _May 15, 2014_____ Signature: _/Sara J. L. Douglas/_____ (Sara J. L. Douglas)

3563168.1

PTO/SB/21 (07-09)
Approved for use through 07/31/2012. 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.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	13/795,933-Conf. #2321
Filing Date	March 12, 2013
First Named Inventor	William F. Ganong, III
Art Unit	2644
Examiner Name	Huy C. Ho
Attorney Docket Number	N0484.71090US01

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
  Copy of Cited Reference

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Marc S. Johannes/		
Printed name	Marc S. Johannes		
Date	July 17, 2014	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 electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: July 17, 2014 Signature: /Sara J. L. Douglas/ (Sara J. L. Douglas)

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 13/795,933	ATTY. DOCKET NO.: N0484.71090US01
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: March 12, 2013	CONFIRMATION NO.: 2321
	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		

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.

3634134.1

DOCKET NO.: N0484.71090US01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong
Application No.:	13/795,933
Confirmation No.:	2321
Filed:	March 12, 2013
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 17th day of July, 2014.

__/Sara J. L. Douglas/_____
Sara J. L. Douglas

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

Madam:

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: Compliance with 37 C.F.R. § 1.704(d)

Each item of information contained in this Information Disclosure Statement

(i) was first cited in any communication from a patent office in a counterpart foreign or international application or from the United States Patent and Trademark Office, and this

Petitioner Exhibit 1004, Page 296 of 1518

communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement; or

(ii) is a communication that was issued by a patent office in a counterpart foreign or international application or by the United States Patent and Trademark Office, and this communication was not received by any individual designated in 37 C.F.R. § 1.56(c) more than thirty (30) days prior to the filing of the Information Disclosure Statement.

PART III: 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/US2014/024270	**June 16, 2014**	**International Search Report and Written Opinion**	**N0484.71090WO00**

PART IV: 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;

Petitioner Exhibit 1004, Page 297 of 1518

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 filed, asserted to be filed or which should have been filed herewith to our Deposit Account No. 23/2825 under Docket No. N0484.71090US01 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Marc S. Johannes/
Marc S. Johannes, Reg. No. 64,978
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: N0484.71090US01
Date: July 17, 2014
xNDDx

3634134.1

PATENT COOPERATION TREATY

IFD

From the INTERNATIONAL SEARCHING AUTHORITY

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)

To:
Johannes, Marc S.
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)*	16 June 2014 (16-06-2014)
Applicant's or agent's file reference N0484.71090	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US2014/024270	International filing date *(day/month/year)* 12 March 2014 (12-03-2014)
Applicant NUANCE COMMUNICATIONS, INC.	

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.

Where? 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 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. 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. Following the expiration of 30 months from the priority date, these comments will also be made available to the public.

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 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.

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.
JUN 23 2014
Docketed ✓ Already Docketed
Not Required
Initials 1st 2nd

Form PCT/ISA/220 (July 2010)

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.
N0484.71090		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US2014/024270	12 March 2014 (12-03-2014)	12 March 2013 (12-03-2013)

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. 3

[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/US2014/024270

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/22 G10L15/28 G10L25/78 G10L15/08
ADD.

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	US 6 070 140 A (TRAN BAO Q [US]) 30 May 2000 (2000-05-30) figures 1-3,6 column 2, lines 39-52 column 4, line 33 - column 5, line 10 column 5, line 42 - column 6, line 8 column 18, lines 17-28 column 18, line 67 - column 19, line 5 -----	1-42

☐ 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
5 June 2014	16/06/2014

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	Tilp, Jan

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2014/024270

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6070140 A	30-05-2000	NONE	

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/US2014/024270	12.03.2014	12.03.2013

International Patent Classification (IPC) or both national classification and IPC
INV. G10L15/22 G10L15/28 G10L25/78 G10L15/08

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	Tilp, Jan Telephone No. +49 89 2399-4753

Form PCT/ISA/237 (Cover Sheet) (July 2009)

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 43bis.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:

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	7-9, 15, 19, 20, 28-30, 36, 39
	No:	Claims	1-6, 10-14, 16-18, 21-27, 31-35, 37, 38, 40-42
Inventive step (IS)	Yes:	Claims	
	No:	Claims	1-42
Industrial applicability (IA)	Yes:	Claims	1-42
	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 Reference is made to the following document:

D1 US 6 070 140 A (TRAN BAO Q [US]) 30 May 2000 (2000-05-30)

2 The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of **claims 1, 2-6, 10-14, 16-18, 21-24, 26, 27, 31-33 and 40-42** is not new in the sense of **Article 33(2) PCT**.

2.1 Document D1 discloses (the references in parentheses applying to D1) - using the present wording of independent **claim 1**:

"A method of monitoring an acoustic environment of a mobile device (column 2, lines 39-44; figures 1-3, 6; column 4, line 33 - column 5, line 9; column 5, lines 42-55), the method comprising:

receiving, by the mobile device, acoustic input from the environment of the mobile device (column 4, lines 66, 67);

detecting whether the acoustic input includes a voice command from a user without requiring receipt of an explicit trigger from the user (column 4, lines 44, 52, 53; column 5, lines 4, 5; column 5, lines 42-55); and

initiating responding to the detected voice command (column 4, lines 55-58; column 5, lines 53-55)."

Consequently, claim 1 fails to fulfill the requirements of the PCT with regard to novelty (Article 33(2) PCT).

2.2 D1 also discloses (the references in parentheses applying to D1) - using the present wording of independent **claim 23**:

"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 (column 2, lines 39-44; figures 1-3, 6; column 4, line 33 - column 5, line 10; column 5, lines 42-55), the method comprising:

receiving acoustic input from the environment of the mobile device while the mobile device is operating in the low power mode (column 4, lines 49-54, 66 and 67);

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 (column 4, lines 44, 52, 53; column 5, lines 4, 5; column 2, lines 39-44; column 5, lines 45-55); and

using at least one contextual cue to assist in detecting whether the acoustic input includes a voice command (column 5, lines 7-10)."

Hence, claim 23 does not fulfill the requirements of Article 33(2) PCT, either.

2.3 The objections raised above in sections 2.1 and 2.2 accordingly apply, mutatis mutandis, to independent **claims 21, 22, 41 and 42**, the respective subject-matter of which thus also fails to fulfill the requirements of the PCT with regard to novelty (Article 33(2) PCT).

2.4 The subject-matter of dependent **claims 2-6, 10-14 and 16-18** is also not new (Article 33(2) PCT), since D1 already discloses (the references in parentheses applying to D1)

i) that - using the wording of **claim 2** - "receiving acoustic input comprises receiving acoustic input while the mobile device is operating in a low power mode" (column 2, lines 41-44; column 4, lines 49-54),

ii) that - using the wording of **claim 3** - "detecting whether the acoustic input includes a voice command from the user comprises performing a plurality of processing stages on the acoustic input to evaluate whether the acoustic input includes the voice command or corresponds to spurious acoustic activity" (column 4, lines 44, 52, 53, 66, 67; column 5, lines 4, 5 and 45-55; column 2, lines 39-44),

iii) that - using the wording of **claim 4** - "performing at least one of the plurality of processing stages provides an indication of confidence that the acoustic input includes a voice command, the method further comprising selecting a subsequent processing stage to next process the acoustic input based, at least in part, on the indication of confidence from the at least one of the plurality of processing stages" (column 5, lines 53-55, wherein said indication of confidence is generally implicit to any kind of speech recognition),

iv) - using the wording of **claim 5** - "discontinuing further processing of the acoustic input if a given processing stage of the plurality of processing stages determines that the acoustic input resulted from spurious acoustic activity; initiating responding to the voice command if the given processing stage determines that the acoustic input includes a voice command; and processing

the acoustic input in a subsequent processing stage selected from the plurality of processing stages if the given processing stage cannot conclude whether the acoustic input includes a voice command" (column 5, lines 45-50, 53-55 and 58-62),

v) that - using the wording of **claim 6** - "performing the plurality of processing stages includes performing at least one voice activity detection processing stage" (column 4, line 66 - column 5, line 5),

vi) that - using the wording of **claim 10** - "performing at least one of the plurality of processing stages comprises performing at least one biometric analysis on the acoustic input to evaluate whether the acoustic input was received from the user of the mobile device" (column 18, lines 17-19, 27, 28),

vii) that - using the wording of **claim 11** - "performing the plurality of processing stages comprises performing at least one speech processing stage" (column 5, lines 48-54),

viii) that - using the wording of **claim 12** - "performing the at least one speech processing stage comprises performing limited vocabulary automatic speech recognition on the acoustic input" (column 5, lines 48-54),

ix) that - using the wording of **claim 13** - "performing the at least one speech processing stage comprises performing classification of the acoustic input including applying one or more statistical classifiers to evaluate whether the recognized speech is characteristic of a voice command" (column 5, lines 48-54, wherein a neural network-based command recognizer represents a statistical classifier),

x) that - using the wording of **claim 14** - "performing the at least one speech processing stage includes performing at least one natural language processing stage" (column 18, lines 24-27),

xi) that - using the wording of **claim 16** - "the mobile device includes a primary processor and a secondary processor, and wherein at least one first processing stage of the plurality of processing stages is performed by the secondary processor prior to engaging the primary processor to process the acoustic input" (figure 6, processor elements 24 and 36; column 5, lines 45-50 and 55-57),

xii) that - using the wording of **claim 17** - "at least one second processing stage of the plurality of processing stages is performed by the primary processor only after the at least one first processing stage neither concludes

that the acoustic input includes a voice command nor concludes that the acoustic input is the result of spurious acoustic activity" (figure 6, processor elements 24 and 36; column 5, lines 45-50 and 55-57), and

xiii) that - using the wording of **claim 18** - "performing at least one of the plurality of processing stages 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" (column 5, lines 7-10).

2.5 In light of the objections raised above in sections 2.4 ii), vii), viii) and x)-xii), dependent **claims 25, 34, 35, 37 and 38** also fail to fulfill the requirements of the PCT with regard to novelty (Article 33(2) PCT).

2.6 Moreover, the subject-matter of dependent **claims 24, 26, 27 and 31-33** is also not new (Article 33(2) PCT), since D1 already discloses (the references in parentheses applying to D1)

i) - using the wording of **claim 24** - "transitioning the mobile device from the low power mode to an active mode when it is detected that the acoustic input includes a voice command" (column 2, lines 49-51),

ii) that - using the wording of **claim 26** - "the at least one contextual cue includes the time the acoustic input was received, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on the time the acoustic input was received" (column 18, line 67 - column 19, line 5),

iii) that - using the wording of **claim 27** - "the at least one contextual cue includes recent activity on the mobile device, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on at least one of how recently the user activated the mobile device, how recently the user issued a voice command and how recently the user performed at least one action on the mobile device" (column 5, lines 53-57 and column 5, line 63 - column 6, line 8, wherein training represents a recent activity influencing subsequent detection),

iv) that - using the wording of **claim 31** - "the at least one contextual cue includes the power source of the mobile device, the method further comprising influencing evaluation of at least one of the plurality of processing stages based on at least one of determining whether the mobile device is being

powered using the battery or connected to wall power, and a state of the battery if the mobile device is being powered using the battery" (column 5, lines 7-10),

v) that - using the wording of **claim 32** - "the at least one contextual cue includes the sample of the acoustic environment obtained by the mobile device by storing acoustic input previously received by the mobile device and determined not to include a voice command, the method further comprising subtracting the sample of the acoustic environment from the acoustic input" (column 2, lines 47-52),

vi) that - using the wording of **claim 33** - "a sequence of performing the plurality of processing stages is determined, at least in part, by the at least one contextual cue" (column 5, lines 45-62, wherein depending on the contextual cue whether or not the "CPU 14" has previously put itself to sleep, the sequence of performing voice (command) detection by wake-up logic 24 and neural network 36 is determined), and

vii) - using the wording of **claim 40** - "performing one or more default actions when a specific task could not be ascertained from the acoustic input" (generally implicit to any method using speech recognition).

3 Each of the combinations of features claimed in dependent **claims 7-9** (**claim 36**, accordingly)**, 15, 19** (**claim 39** accordingly)**, 20 and 28-30** comes within the scope of the customary practice followed by persons skilled in the art, especially as the advantages thus achieved can readily be foreseen. Consequently, the respective subject-matter of said claims lacks an inventive step in the sense of **Article 33(3) PCT**.

Electronic Acknowledgement Receipt

EFS ID:	19608149
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
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/Sara Douglas
Filer Authorized By:	Marc S. Johannes
Attorney Docket Number:	N0484.71090US01
Receipt Date:	17-JUL-2014
Filing Date:	12-MAR-2013
Time Stamp:	14:32:33
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	N048471090US01-TRN-MSJ.pdf	26445 082dd0968a2cb5fe789d75296e03b965f2e4c808	no	1

Warnings:	
Information:	

2		N048471090US01-IDS-MSJ.pdf	31671 d48bb244fab8409d40813a424620850bae4b8f1b	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:

3	Other Reference-Patent/App/Search documents	N048471090WO00_ISR_mailed_2014-06-16.pdf	580645 4ff077cbcf3defe0d8774c8b7cab94be0f1ac988	no	11

Warnings:

Information:

Total Files Size (in bytes):	638761

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
13/795,933	03/12/2013	William F. Ganong III	N0484.71090US01

CONFIRMATION NO. 2321

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

PUBLICATION NOTICE

OC000000070837204

Title:METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND

Publication No.US-2014-0274203-A1
Publication Date:09/18/2014

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

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 13/795,933	ATTY. DOCKET NO.: N0484.71090US01
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: March 12, 2013	CONFIRMATION NO.: 2321
	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

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 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: /Huy Ho/ (09/11/2014)	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.

3634134.1 ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.H

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 13/795,933	ATTY. DOCKET NO.: N0484.71090US01
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: March 12, 2013	CONFIRMATION NO.: 2321
	FIRST NAMED INVENTOR: William F. Ganong	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: Matthew W. Genack

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
		7,720,683		Vermeulen et al.	05-18-2010
		7,567,827		Kim	07-28-2009
		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
		8,666,751		Murthi et al.	03-04-2014
		2010-0185448		Meisel	07-22-2010
		2011-0054899		Phillips et al.	03-03-2011
		2013-0080167		Mozer	03-28-2013
		2013-0080171		Mozer et al.	03-28-2013
		2013-0289994		Newman et al.	10-31-2013
		2013-0339028		Rosner et al.	12-19-2013
		2014-0012586		Rubin et al.	01-09-2014
		2014-0012573		Hung et al.	01-09-2014

[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 Ho/ (09/11/2014)	

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.

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

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	13795933	GANONG ET AL.
	Examiner	Art Unit
	HUY C HO	2644

CPC- SEARCHED		
Symbol	Date	Examiner

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner
455	563	9/25/2014	HH

SEARCH NOTES		
Search Notes	Date	Examiner
EAST Database Search	9/25/2014	HH
Inventor Search	9/25/2014	HH

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner
455	79	9/25/2014	HH
379	88.01, 88.02	9/25/2014	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
L1	2	"20130289994"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 23:54
L2	2	2013/0080171	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 23:54
L3	55	((William) near2 (Ganong)).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2014/09/29 23:55
L4	5	((Paul) near2 ("Van Mulbregt")).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2014/09/29 23:55
L5	31	((Vladimir) near2 (Sejnoha)).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2014/09/29 23:55
L6	25	((Glen) near2 (Wilson)).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2014/09/29 23:55
S1	84	("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; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/11 15:08
S2	1	"US 20140274203"	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2014/09/25 13:39
S3	6	"8094838"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/25 13:54
S4	6	"8094838"	US-PGPUB;	OR	OFF	2014/09/25

			USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB			13:54
S5	6	"8094838"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/25 14:38
S6	36	(jacoby near2 keith).in.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/25 14:39
S7	3	"US 20080172221"	US-PGPUB; USPAT; USOCR; DERWENT	OR	OFF	2014/09/25 14:39
S8	3	"20140012573"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 17:24
S9	5	"20030040339"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 19:50
S10	5	"20030040339"	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 19:50
S11	903	455/563.ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 20:36
S12	239	S11 and (voice near3 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 20:37
S13	162	S12 and (GPS location)	US-PGPUB; USPAT;	OR	OFF	2014/09/29 20:38

			USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB			
S14	18	S13 and (acceleration accelerometer)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 20:39
S15	14	S14 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 20:39
S16	4	S15 and (vehicle vehicular automobile car)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 20:40
S17	10166	379/88.01-88.2.ccls.	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:23
S18	1003	S17 and (voice near3 command)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:23
S19	367	S18 and (voice near3 recogniz$5)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:24
S20	7	S19 and (acceleration accelerometer)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:24
S21	253	S19 and (GPS location)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:24

S22	6	S15 and (vehicle vehicular automobile car home office)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:25
S23	211	S21 and (vehicle vehicular automobile car home office)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:25
S24	211	S23 and (@ad<"20130312" @rlad<"20130312")	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:25
S25	13	S24 and (low$2 near3 (power energy))	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2014/09/29 21:26

9/ 29/ 2014 11:56:07 PM
C:\ Users\ hho3\ Documents\ EAST\ Workspaces\ 13795933.wsp

Notice of References Cited	Application/Control No. 13/795,933	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	Classification
*	A	US-2014/0012573	01-2014	Hung et al.	704/233
*	B	US-2013/0289994	10-2013	Newman et al.	704/254
*	C	US-2013/0080171	03-2013	Mozer et al.	704/251
*	D	US-8,620,389	12-2013	Schrager, Martin	455/569.1
*	E	US-2010/0009719	01-2010	Oh et al.	455/563
	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	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.

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.
13/795,933	03/12/2013	William F. Ganong III	N0484.71090US01	2321

91337 7590 10/06/2014

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
10/06/2014	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