



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.
14/752,400	06/26/2015	Yue Liu	AM2-1887US	8487
29150 7590 11/02/2016 LEE & HAYES, PLLC 601 W. RIVERSIDE AVENUE SUITE 1400 SPOKANE, WA 99201			EXAMINER KIM, JONATHAN C	
			ART UNIT 2659	PAPER NUMBER
			NOTIFICATION DATE 11/02/2016	DELIVERY MODE 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):

lhpto@leehayes.com

Office Action Summary	Application No. 14/752,400	Applicant(s) LIU ET AL.	
	Examiner JONATHAN KIM	Art Unit 2659	AIA (First Inventor to File) Status Yes

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

Period for Reply

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

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

Status

1) ☒ Responsive to communication(s) filed on 8/23/2016.
☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on ____.

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

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

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

Disposition of Claims*

5) ☒ Claim(s) 1-23 is/are pending in the application.
5a) Of the above claim(s) ____ is/are withdrawn from consideration.

6) ☒ Claim(s) 1-5 is/are allowed.

7) ☒ Claim(s) 6-23 is/are rejected.

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

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

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

Application Papers

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

11) ☒ The drawing(s) filed on 6/26/2015 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

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

Certified copies:

a) ☐ All b) ☐ Some** c) ☐ None of the:

1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. ____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

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

Attachment(s)

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

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

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

4) ☐ Other: ____.

DETAILED ACTION

This Office Action is in response to the correspondence filed by the applicant on 8/23/2016.

Notice of Pre-AIA or AIA Status

The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA.

Response to Amendment

The Amendment filed on 8/23/2016 has been entered.

Claims 1-23 have been amended by Applicant.

Claims 1-23 remain pending in the application.

Response to Arguments

Applicant's arguments with respect to the rejection of claims 1-5 under 103 have been fully considered and **are persuasive**. The rejections have been withdrawn.

Applicant's arguments with respect to the rejection of claims 6 and 17 under 103 have been fully considered and are moot upon a further consideration and a new ground(s) of rejection made under AIA 35 U.S.C. 103 as being unpatentable over **LINDAHL (US 2015/0106085 A1)**, and further in view of **VISSER (US 2011/0264447 A1)**. Please see the rejections below for more details.

Applicant's arguments with respect to the rejection of claims 6-23 under 101 have been fully considered and are **not persuasive**.

All the determining steps are nothing more than an abstract concept. Because the Claims are so broadly stated and so much is left to the judgment that even two humans would not perform the limitations the same way. The steps may be done with a human without the aid of a

Art Unit: 2659

machine and they lack specificity to the degree that they give the human free hand in performing the steps *such that no two humans would do each step the same way*. Specifically, the lack of detail in the step of “determining that the first audio signal corresponds to the identified direction in which the non-human sound source is located”, which are central to the purpose of the method invented by the applicant, makes the Claims very broad such that the claim(s) as a whole, considering all claim elements both individually and in combination, do not amount to significantly more than an abstract idea. The Claims need meaningful limitations that go beyond generally linking the use of an abstract idea to a particular technological environment. As is, the steps of independent Claims and their respective dependent Claims can be performed by a human being. Therefore, the steps are all abstract and the Claim as a whole is abstract. Please see the rejections below for more details.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 6-23 are rejected under 35 U.S.C. 101 because the claimed invention is directed to a judicial exception (i.e., a law of nature, a natural phenomenon, or an abstract idea) without significantly more.

The Supreme Court has long held that “[l]aws of nature, natural phenomena, and abstract ideas are not patentable.” *Alice Corp. Pty. Ltd. v. CLS Bank Int’l*, 134 S. Ct. 2347, 2354 (2014) (quoting *Assoc. for Molecular Pathology v. Myriad Genetics, Inc.*, 133 S. Ct. 2107, 2116 (2013) (internal quotation marks omitted)). The “abstract ideas” category embodies the longstanding

Art Unit: 2659

rule that an idea, by itself, is not patentable. *Alice Corp.*, 134 S. Ct. at 2355 (quoting *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972)).

In *Alice*, the Supreme Court sets forth an analytical “framework for distinguishing patents that claim laws of nature, natural phenomena, and abstract ideas [or mental processes^[1]] from those that claim patent-eligible applications of those concepts.” *Id.* at 2355 (citing *Mayo Collaborative Servs. v. Prometheus Labs., Inc.*, 132 S. Ct. 1289, 1296–97 (2012)). The first step in the analysis is to “determine whether the claims at issue are directed to one of those patent-ineligible concepts.” *Id.* If the claims are directed to a patent-ineligible concept, the second step in the analysis is to consider the elements of the claims “individually and ‘as an ordered combination’” to determine whether there are additional elements that “‘transform the nature of the claim’ into a patent-eligible application.” *Id.* (quoting *Mayo*, 132 S. Ct. at 1298, 1297). In other words, the second step is to “search for an ‘inventive concept’—*i.e.*, an element or combination of elements that is ‘sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the [ineligible concept] itself’”. *Id.* (brackets in original) (quoting *Mayo*, 132 S. Ct. at 1294). The prohibition against patenting an abstract idea “‘cannot be circumvented by attempting to limit the use of the formula to a particular technological environment’ or adding ‘insignificant post-solution activity.’” *Bilski v. Kappos*, 561 U.S. 593, 610–11 (2010) (citation omitted).

In applying the framework set out in *Alice*, and as the first step of the analysis, examiner found Applicant’s claims 6 and 17 are directed to a patent-ineligible abstract concept of performing speech recognition based on identified direction of a noise source. All the steps of Applicant’s

^[1] See *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972) (“Phenomena of nature, though just discovered, *mental processes*, and abstract intellectual concepts are not patentable, as they are the basic tools of scientific and technological work.”). (Emphasis added).

Art Unit: 2659

claims 6 and 17 are an abstract concept that **could be performed in the human mind**, or by a human using a pen and paper. Because the Claims are so broadly stated and so much is left to the judgment that even two humans would not perform the limitations the same way. The above steps may be done with a human without the aid of a machine and they lack specificity to the degree that they give the human free hand in performing the steps *such that no two humans would do each step the same way*. Specifically, the lack of detail in the step of “*determining that the first audio signal corresponds to the identified direction in which the non-human sound source is located*”, which is central to the purpose of the method invented by the applicant, makes the Claims very broad such that the claim(s) as a whole, considering all claim elements both individually and in combination, do not amount to significantly more than an abstract idea. The Claims need meaningful limitations that go beyond generally linking the use of an abstract idea to a particular technological environment. As is, the steps of independent Claims and their respective dependent Claims can be performed by a human being. Therefore, the steps are all abstract and the Claim as a whole is abstract. *CyberSource Corp.* 654 F.3d at 1372. “[M]ental processes—or processes of human thinking—standing alone are not patentable even if they have practical application.” *In re Comiskey*, 554 F.3d 967, 979 (Fed. Cir. 2009); *see also Gottschalk v. Benson*, 409 U.S. 63, 67 (1972) (“Phenomena of nature . . . , *mental processes*, and abstract intellectual concepts are not patentable, as they are basic tools of scientific and technological work.” (emphasis added)).

For the second step of the *Alice* analysis, we find Applicant's claims 6 and 17 do not have any additional limitations that represent a patent-eligible application of the abstract concept. *Alice*, 134 S. Ct. at 2357. As recognized by the Federal Circuit in *Ultramercial Inc. v. Hulu, LLC*, 772 F.3d 709, 714 (Fed. Cir. 2014) (quoting *Bilski v. Kappos*, 561 U.S. at 594), the “machine-or-transformation” test can provide a “‘useful clue’ in the second step of the *Alice*

Art Unit: 2659

framework.” However, neither of Applicant’s process claims 6 and 17 is tied to any machine; nor any of the recited steps of Applicant’s claims 6 and 17 transforms “a particular article into a different state or thing.” *In re Bilski*, 545 F.3d 943, 954 (Fed. Cir. 2008), *aff’d*, *Bilski v. Kappos*, 561 U.S. 593 (2010). According to *In re Bilski*, the transformation (1) must involve an underlying article from one state to a different state or thing, and (2) must be central to the purpose of Appellants’ claimed process. *Id.* While an underlying article can be intangible, such as electrical signals, and transformation can include data transformation, data must represent a physical object or an article. *See Arrhythmia Research Tech., Inc. v. Corazonic Corp.*, 958 F.2d 1053, 1059 (Fed. Cir. 1992); *In re Abele*, 684 F.2d 902, 905–06 (CCPA 1982).

In this case, Examiner finds Applicant’s claims 6 and 17 are neither “tied to a particular machine or apparatus” nor do they “transform a particular article into a different state or thing” as required by *In re Bilski*. In particular, Examiner notes that none of recited steps in Applicant’s independent claims 6 and 17 refers to a specific machine by reciting structural limitations of any apparatus or to any specific operations that would cause a machine to be the mechanism to perform these steps. As such, Examiner maintains the previous 101 rejection.

Claim Rejections - 35 USC § 103

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

A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102 of this title, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 6-9, 11-12, and 16-21 are rejected under 35 U.S.C. 103 as being unpatentable over **LINDAHL (US 2015/0106085 A1)**, and in further view of **VISSER (US 2011/0264447 A1)**.

Art Unit: 2659

1. **Regarding Claim 6, LINDAHL discloses a processor-implemented method, comprising:**
receiving a first audio signal corresponding to a first direction, the first audio signal emphasizing first sound from the first direction (**LINDAHL Fig. 3 Units 105~107 – “microphone”; Par 6 – “e.g., the microphones may detect speech on different acoustic planes ... by being oriented in different directions to enable directional speech pickup.”**);
receiving a second audio signal corresponding to a second direction, the second audio signal emphasizing second sound from the second direction (**LINDAHL Fig. 3 Units 105~107 – “microphone”; Par 6 – “e.g., the microphones may detect speech on different acoustic planes ... by being oriented in different directions to enable directional speech pickup.”; Par 22 – “Each of these microphones 105-107 may be omnidirectional but picks up sound on a different acoustic plane as result of their varying locations and orientations ... one microphone 106 may be closest to the talker’s mouth and hence better able to pick up speech of the user ... while a second microphone 105 and a third microphone 107 may be used as reference microphone and an error microphone, respectively ...”**);
identifying the first direction as an identified direction in which a non-human sound source is located (**LINDAHL Par 58 – “The microphone corresponding to the highest identified confidence level ... is then identified to be an “unobstructed” or unoccluded microphone suitable to receive speech input from the user ...”; Par 64 – “This disparity between the two SR engines may be a result of the different acoustic planes on which their respective microphone A and microphone B pick up sound, and/or may be result of microphone B being partially occluded.”**);
analyzing the first audio signal to identify a representation of speech (**LINDAHL Fig. 4 – “Identified Frames from Microphone B”; Par 64 – “However, the second speech recognition engine recognizes the word in the fifth frame with a relatively mediocre confidence level of forty percent.”**);
determining that the first audio signal corresponds to the identified direction in which the non-human sound source is located (**LINDAHL Par 58 – “The microphone corresponding to the highest identified confidence level ... is then identified to be an “unobstructed” or unoccluded**

Art Unit: 2659

microphone suitable to receive speech input from the user ...”; In other words, the other directional microphones (e.g., Mic B or Mic C in Figs 3 and 4) are determined to be the non-human sound source, such as a noisy source.);

[analyzing the first audio signal using a first standard to detect an utterance of a trigger expression]

(LINDAHL Fig. 4 – “Identified Frames from Microphone B”; Par 64 – “However, the second speech recognition engine recognizes the word in the fifth frame with a relatively mediocre confidence level of forty percent.”);

analyzing the second audio signal using a second standard to detect the utterance of the trigger expression **(LINDAHL Fig. 4 – “Identified Frames from Microphone A”; Par 63 – “In this example, the speech recognition engine identifies the recognized word or words of the command with a relatively high confidence level of seventy-five percent.”);**

[wherein the first standard is more stringent than the second standard]; and

performing speech recognition on subsequent speech, based at least in part on the utterance of the trigger expression **(LINDAHL Par 58 – “The microphone corresponding to the highest identified confidence level ... is then identified to be an “unobstructed” or unoccluded microphone suitable to receive speech input from the user. Subsequently, the decision logic 330 may provide an identification of this preferred microphone to the long phrase recognition processor 220 ...”).**

LINDAHL is silent to the bracketed limitations: [wherein the first standard is more stringent than the second standard].

VISSER discloses a system/method for speech recognition/detection using multiple microphones,

[wherein the first standard is more stringent than the second standard] **(VISSER Par 165 – “The threshold value may be determined heuristically, and it may be desirable to use different threshold values depending on one or more factors such as signal-to-noise ratio (SNR), noise floor, etc. (e.g., to use a higher threshold value when the SNR is low) ...”; In other words, for the channel with a high level of noise (the first audio signal), the SNR is expected to be low. VISSER**

suggests to use a higher threshold for a noisy signal than a clean signal, e.g., a signal with a high SNR.).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of LINDAHL to include different thresholds based on the presence of a noise source, as taught by VISSER. One of ordinary skill would have been motivated to include different thresholds based on the presence of a noise source in order to improve speech detection (**VISSER Par 163-169**).

2. **Regarding Claim 7, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 6, further comprising:

producing a first score indicating a first likelihood that the trigger expression is represented by the first audio signal (**LINDAHL Fig. 4 – “Identified Frames from Microphone B”; Par 64 – “However, the second speech recognition engine recognizes the word in the fifth frame with a relatively mediocre confidence level of forty percent.”**);

comparing the first score to a first threshold (**LINDAHL Par 66 – “In one embodiment, the decision logic may evaluate the confidence levels received from the speech recognition engines, for example, comparing the confidence levels to a predetermined threshold”**);

producing a second score indicating a second likelihood that the trigger expression is represented by the second audio signal (**LINDAHL Fig. 4 – “Identified Frames from Microphone A”; Par 63 – “In this example, the speech recognition engine identifies the recognized word or words of the command with a relatively high confidence level of seventy-five percent.”**); and

comparing the second score to a second threshold (**LINDAHL Par 66 – “In one embodiment, the decision logic may evaluate the confidence levels received from the speech recognition engines, for example, comparing the confidence levels to a predetermined threshold”**);

wherein the first threshold is greater than the second threshold (**VISSER Par 165 – “The threshold value may be determined heuristically, and it may be desirable to use different threshold values depending on one or more factors such as signal-to-noise ratio (SNR), noise floor, etc. (e.g., to use a higher threshold value when the SNR is low) ...”**; In other words, for the channel with a high

Art Unit: 2659

level of noise (the first audio signal), the SNR is expected to be low. VISSER suggests to use a higher threshold for a noisy signal than a clean signal, e.g., a signal with a high SNR.).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of LINDAHL to include different thresholds based on the presence of a noise source, as taught by VISSER. One of ordinary skill would have been motivated to include different thresholds based on the presence of a noise source in order to reduce a false positive rate (**VISSER Par 163-169**).

3. **Regarding Claim 8, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 7, wherein producing the first score comprises:

creating an acoustic model of an utterance represented by the first audio signal (**LINDAHL Par 52 – “A speech recognition engine 320 may derive one or more words spoken by a user ... in the speech input of the user and the statistical representations of the phonemes from the acoustic model ... ”**); and

comparing the acoustic model to one or more reference acoustic models that correspond to the trigger expression (**LINDAHL Par 50 – “A command or other predetermined phrase can be detected by the decision logic 330, by comparing a word or phrase recognized by any one of the SR engines 320 to an expected or target word or phrase.”**).

4. **Regarding Claim 9, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 6, wherein:

analyzing the first audio signal using the first standard comprises performing expression detection in accordance with a first detection algorithm (**LINDAHL Fig. 3 – “Speech Recognition Engine 320A-C”; Fig. 4 – “Identified Frames from Microphone B”; Par 64 – “However, the second speech recognition engine recognizes the word in the fifth frame with a relatively mediocre confidence level of forty percent.”**); and

analyzing the second audio signal using the second standard comprises performing expression detection in accordance with a second detection algorithm (**LINDAHL Fig. 3 – “Speech Recognition Engine**

Art Unit: 2659

320A-C”; Fig. 4 – “Identified Frames from Microphone A”; Par 63 – “In this example, the speech recognition engine identifies the recognized word or words of the command with a relatively high confidence level of seventy-five percent.”).

5. **Regarding Claim 11, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 9, wherein the first algorithm comprises analyzing the first audio signal to detect the utterance of the trigger expression after a time period during which speech is not detected (**LINDAHL Fig. 4 – “Identified Frames from Microphone B”; Note that speech is not detected for the first two frames.**).

6. **Regarding Claim 12, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 6, wherein identifying the first direction as the identified direction further comprises determining that the first audio signal represents speech during at least a threshold amount of time during a past time period (**LINDAHL Par 68 – “... the command may be recognized from the content of a single frame (provided that the frame is defined to be long enough in time.)”; In other words, if the analysis frame is long enough (above a threshold amount of time), the command/speech/non-speech can be determined.**).

7. **Regarding Claim 16, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 6, wherein analyzing the first audio signal to identify the representation of speech comprises:
determining one or more features of the first audio signal (**LINDAHL Par 52 – “...on the phonemes in the speech input of the user and the statistical representations of the phonemes from the acoustic model ... ”; Par 56 – “... a speech recognition engine may include a voice activity detection (VAD) ... may determine if there is speech present in a received audio signal.”; VISSER Par 112 – “ ... configured to classify a segment as active (e.g., speech) or inactive (e.g., noise) based on one or**

Art Unit: 2659

more factors such as frame energy, signal-to-noise ration, periodicity, autocorrelation of speech and/or residual ..., zero crossing rate ...”); and

comparing the one or more features to one or more reference signals that are known to contain human speech (**Par 56 – “... a speech recognition engine may include a voice activity detection (VAD) ... may determine if there is speech present in a received audio signal.”; VISSER Par 112 – “ ... configured to classify a segment as active (e.g., speech) or inactive (e.g., noise) based on one or more factors such as frame energy, signal-to-noise ration, periodicity, autocorrelation of speech and/or residual ..., zero crossing rate ...”).**

8. **Regarding Claim 17, LINDAHL discloses a processor-implemented method, comprising:**
 receiving a first audio signal representing sound from a first area of an environment (**LINDAHL Fig. 3 Units 105~107 – “microphone”; Par 6 – “e.g., the microphones may detect speech on different acoustic planes ... by being oriented in different directions to enable directional speech pickup.”**);
 receiving a second audio signal representing sound from a second area of the environment (**LINDAHL Fig. 3 Units 105~107 – “microphone”; Par 6 – “e.g., the microphones may detect speech on different acoustic planes ... by being oriented in different directions to enable directional speech pickup.”; Par 22 – “Each of these microphones 105-107 may be omnidirectional but picks up sound on a different acoustic plane as result of their varying locations and orientations ... one microphone 106 may be closest to the talker’s mouth and hence better able to pick up speech of the user ... while a second microphone 105 and a third microphone 107 may be used as reference microphone and an error microphone, respectively ...”**);
 determining that the first audio signal represents sound generated by a non-human sound source (**LINDAHL Par 58 – “The microphone corresponding to the highest identified confidence level ... is then identified to be an “unobstructed” or unoccluded microphone suitable to receive speech input from the user ...”; Par 64 – “This disparity between the two SR engines may be a result of the different acoustic planes on which their respective microphone A and microphone B pick up sound, and/or may be result of microphone B being partially occluded.”**);

Art Unit: 2659

analyzing the first audio signal using a first standard to detect an utterance of a trigger expression (**LINDAHL Fig. 4 – “Identified Frames from Microphone B”; Par 64 – “However, the second speech recognition engine recognizes the word in the fifth frame with a relatively mediocre confidence level of forty percent.”**), [the first standard based at least in part on the non-human sound source represented in the first audio signal];

analyzing the second audio signal using a second standard to detect the utterance of the trigger expression (**LINDAHL Fig. 4 – “Identified Frames from Microphone A”; Par 63 – “In this example, the speech recognition engine identifies the recognized word or words of the command with a relatively high confidence level of seventy-five percent.”**); and

performing speech recognition on subsequent speech, based at least in part on the utterance of the trigger expression (**LINDAHL Par 58 – “The microphone corresponding to the highest identified confidence level ... is then identified to be an “unobstructed” or unoccluded microphone suitable to receive speech input from the user. Subsequently, the decision logic 330 may provide an identification of this preferred microphone to the long phrase recognition processor 220 ...”**).

LINDAHL is silent to the bracketed limitation: [the first standard based at least in part on the non-human sound source represented in the first audio signal].

VISSER discloses a system/method for speech recognition/detection using multiple microphones, [the first standard based at least in part on the non-human sound source represented in the first audio signal] (**VISSER Par 165 – “The threshold value may be determined heuristically, and it may be desirable to use different threshold values depending on one or more factors such as signal-to-noise ratio (SNR), noise floor, etc. (e.g., to use a higher threshold value when the SNR is low) ...”**).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of **LINDAHL** to include a standard based on a non-human source represented in an audio signal, as taught by **VISSER**.

One of ordinary skill would have been motivated to include a standard based on a non-human source represented in an audio signal in order to improve speech detection (**VISSER Par 163-169**).

9. **Regarding Claim 18, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 17, wherein the first standard is more stringent than the second standard (**VISSER Par 165 – “The threshold value may be determined heuristically, and it may be desirable to use different threshold values depending on one or more factors such as signal-to-noise ratio (SNR), noise floor, etc. (e.g., to use a higher threshold value when the SNR is low) ...”; In other words, for the channel with a high level of noise (the first audio signal), the SNR is expected to be low. VISSER suggests to use a higher threshold for a noisy signal than a clean signal, e.g., a signal with a high SNR.)**).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of LINDAHL to include different thresholds based on the presence of a noise source, as taught by VISSER. One of ordinary skill would have been motivated to include different thresholds based on the presence of a noise source in order to improve speech detection (**VISSER Par 163-169**).

10. **Regarding Claim 19, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 17, further comprising:
producing a first score indicating a first likelihood that the trigger expression is represented by the first audio signal (**LINDAHL Fig. 4 – “Identified Frames from Microphone B”; Par 64 – “However, the second speech recognition engine recognizes the word in the fifth frame with a relatively mediocre confidence level of forty percent.”**) and comparing the first score to a first threshold (**LINDAHL Par 66 – “In one embodiment, the decision logic may evaluate the confidence levels received from the speech recognition engines, for example, comparing the confidence levels to a predetermined threshold”**);
producing a second score indicating a second likelihood that the trigger expression is represented by the second audio signal (**LINDAHL Fig. 4 – “Identified Frames from Microphone A”; Par 63 – “In this example, the speech recognition engine identifies the recognized word or words of the command with a relatively high confidence level of seventy-five percent.”**); and

Art Unit: 2659

comparing the second score to a second threshold (**LINDAHL Par 66** – “In one embodiment, the decision logic may evaluate the confidence levels received from the speech recognition engines, for example, comparing the confidence levels to a predetermined threshold”);

wherein the first threshold is greater than the second threshold (**VISSER Par 165** – “The threshold value may be determined heuristically, and it may be desirable to use different threshold values depending on one or more factors such as signal-to-noise ratio (SNR), noise floor, etc. (e.g., to use a higher threshold value when the SNR is low) ...”; In other words, for the channel with a high level of noise (the first audio signal), the SNR is expected to be low. VISSER suggests to use a higher threshold for a noisy signal than a clean signal, e.g., a signal with a high SNR.).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of LINDAHL to include different thresholds based on the presence of a noise source, as taught by VISSER. One of ordinary skill would have been motivated to include different thresholds based on the presence of a noise source in order to reduce a false positive rate (**VISSER Par 163-169**).

11. **Regarding Claim 20, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 17, further comprising:

Determining that the trigger expression is represented in the first audio signal using a first detection algorithm (**LINDAHL Fig. 3** – “Speech Recognition Engine 320A-C”; **Fig. 4** – “Identified Frames from Microphone B”; **Par 64** – “However, the second speech recognition engine recognizes the word in the fifth frame with a relatively mediocre confidence level of forty percent.”); and

Determining that the trigger expression is represented in the second audio signal using a second detection algorithm (**LINDAHL Fig. 3** – “Speech Recognition Engine 320A-C”; **Fig. 4** – “Identified Frames from Microphone A”; **Par 63** – “In this example, the speech recognition engine identifies the recognized word or words of the command with a relatively high confidence level of seventy-five percent.”),

Art Unit: 2659

wherein the first detection algorithm is different than the second detection algorithm (**LINDAHL Par 51 – “The speech recognition engines 320A-C may ... consistent acoustic models may be stored for individual speech recognition 320A-C in individual files or other data structures.”**).

12. **Regarding Claim 21, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 17, wherein determining that the first audio signal represents sound generated by the non-human sound source comprises determining that the first audio signal represents speech during at least a threshold amount of time during a past time period (**LINDAHL Par 68 – “... the command may be recognized from the content of a single frame (provided that the frame is defined to be long enough in time.)”; In other words, if the analysis frame is long enough (above a threshold amount of time), the command/speech/non-speech can be determined.**).

Claims 10 is rejected under 35 U.S.C. 103 as being unpatentable over **LINDAHL (US 2015/0106085 A1)** in view of **VISSER (US 2011/0264447 A1)**, and in further view of **KIM (US 2013/0332168 A1)**.

13. **Regarding Claim 10, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 9, wherein the first algorithm comprises analyzing the first audio signal to detect at least two utterances of the trigger expression [within a particular time period]. (**LINDAHL Par 9 – “... for example, to at most ten (10) words, and/or the wake command is limited, for example, to at most five (5) words ...”**).

LINDAHL in view of VISSER is silent to the bracketed limitations: [within a particular time period]

KIM discloses a system/method for keyword detection comprising: analyzing the first audio signal to detect at least two utterances of the trigger expression [within a particular time period] (**KIM Par 24 – “In one embodiment, the voice recognition functionality may terminate after a particular time period between spoken utterance (e.g., a two second silence, three second silence, etc.”**).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of **LINDAHL in view of VISSER** to include a time period between utterances, as taught by **KIM**.

Art Unit: 2659

One of ordinary skill would have been motivated to include a time period between utterances in order to accurately recognize continuous speech signals (**KIM Par 24**).

Claims 13 and 22 are rejected under 35 U.S.C. 103 as being unpatentable over **LINDAHL (US 2015/0106085 A1)** in view of **VISSER (US 2011/0264447 A1)**, and in further view of **SINGHAL (US 2005/0096898 A1)**.

14. **Regarding Claim 13, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 6.

LINDAHL in view of VISSER is silent to wherein identifying the first direction as the identified direction further comprises determining that the first audio signal has audio frequencies outside of a range of spoken speech frequencies.

SINGHAL discloses a method/system for speech signal processing, wherein identifying the first direction as the identified direction further comprises determining that the first audio signal has audio frequencies outside of a range of spoken speech frequencies (**SINGHAL Par 51-53 – “Human speech, on the other hand, ordinary ranges from about 300Hz to about 4,000 Hz ... The greater the energy in the higher part of the spectrum in comparison to the lower part of the spectrum, the greater the likelihood that the audio communication is associated with music ...”**).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of **LINDAHL in view of VISSER** to determining audio frequencies outside of spoken speech frequencies, as taught by **SINGHAL**.

One of ordinary skill would have been motivated to include determining audio frequencies outside of spoken speech frequencies in order to accurately identify speech/non-speech signals (**SINGHAL Par 53**).

15. **Regarding Claim 22, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 17.

Art Unit: 2659

LINDAHL in view of VISSER is silent to wherein determining that the first audio signal represents sound generated by the non-human source comprises determining that the first audio signal has audio frequencies outside of a range of human speech frequencies.

SINGHAL discloses a method/system for speech signal processing, wherein determining that the first audio signal represents sound generated by the non-human source comprises determining that the first audio signal has audio frequencies outside of a range of human speech frequencies (**SINGHAL Par 51-53 – “Human speech, on the other hand, ordinary ranges from about 300Hz to about 4,000 Hz ... The greater the energy in the higher part of the spectrum in comparison to the lower part of the spectrum, the greater the likelihood that the audio communication is associated with music ...”**).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of **LINDAHL in view of VISSER** to determining audio frequencies outside of spoken speech frequencies, as taught by **SINGHAL**.

One of ordinary skill would have been motivated to include determining audio frequencies outside of spoken speech frequencies in order to accurately identify speech/non-speech signals (**SINGHAL Par 53**).

Claims 14 and 15 are rejected under 35 U.S.C. 103 as being unpatentable over **LINDAHL (US 2015/0106085 A1)** in view of **VISSER (US 2011/0264447 A1)**, and in further view of **JARVINEN (US 2016/0125867 A1)**.

16. **Regarding Claim 14, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 6.

LINDAHL in view of VISSER is silent to wherein identifying the first direction as the identified direction further comprises receiving an indication from a user that the nonhuman sound source is located in the first direction.

JARVINEN disclose a method/apparatus for detecting and analyzing audio signals with an array of microphones, wherein identifying the first direction as the identified direction further comprises receiving an indication from a user that the nonhuman sound source is located in the first direction (**JARVINEN Par 217 – “... the user can give indication where he would like a masking sound to be positioned (or**

Art Unit: 2659

where the most annoying noise source is located). The indication could be given by touching at desired direction on a user interface ...”).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of LINDAHL in view of VISSER to include an indication from a user, as taught by JARVINEN.

One of ordinary skill would have been motivated to include an indication from a user in order to accurately locate the source of noise (**JARVINEN Par 217**).

17. **Regarding Claim 15, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 6.

LINDAHL in view of VISSER is silent to wherein identifying the first direction as the identified direction further comprises receiving an indication from a user that the nonhuman sound source is located in a direction corresponding to the position of the user.

JARVINEN disclose a method/apparatus for detecting and analyzing audio signals with an array of microphones, wherein identifying the first direction as the identified direction further comprises receiving an indication from a user that the nonhuman sound source is located in a direction corresponding to the position of the user (**JARVINEN Par 217 – “... the user can give indication where he would like a masking sound to be positioned (or where the most annoying noise source is located). The indication could be given by touching at desired direction on a user interface, where the user is positioned in the centre ...”).**

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of LINDAHL in view of VISSER to include an indication from a user, as taught by JARVINEN.

One of ordinary skill would have been motivated to include an indication from a user in order to accurately locate the source of noise (**JARVINEN Par 217**).

Claim 23 is rejected under 35 U.S.C. 103 as being unpatentable over **LINDAHL (US 2015/0106085 A1)** in view of **VISSER (US 2011/0264447 A1)**, and in further view of **IVANOV (US 2014/0278393 A1)**.

Art Unit: 2659

18. **Regarding Claim 23, LINDAHL in view of VISSER** discloses the processor-implemented method of claim 17.

LINDAHL in view of VISSER is silent to accessing a database including a plurality of audio content; identifying audio content of the plurality of audio content corresponding to the sound produced by the non-human sound source; and canceling the identified content from the first audio signal.

IVANOV discloses a method/system for a voice recognition with multiple microphones comprising: accessing a database including a plurality of audio content (**IVANOV Par 41 – “The system memory 107 may store the database of noise suppression algorithms 109 and nay other useful information such as an associated memory table that can be used to compare observed SNR and noise types to select a suitable noise suppression algorithm”**);

identifying audio content of the plurality of audio content corresponding to the sound produced by the non-human sound source (**IVANOV Par 41 – “In operation block 223, the noise type classifier 170 receives the buffered voice signal 111, and may also receive signal-to-noise ratio information from SNR estimator 160 via control signal 129. The noise type classifier 170 assigns a noise type ...”**);

and

canceling the identified content from the first audio signal (**IVANOV Par 41 – “In operation block 225, the noise suppressor algorithm selector 180proceeds to select an appropriate noise suppressor algorithm for the observed conditions (i.e. observed SNR and noise type.)”**).

It would have been obvious to one of ordinary skill in the art at the time of filing to modify the method of LINDAHL in view of VISSER to include identifying audio content for noise suppression, as taught by IVANOV.

One of ordinary skill would have been motivated to include identifying audio content for noise suppression in order to appropriately reduce noise (**IVANOV Par 41**).

Allowable Subject Matter

Claims 1-5 are allowed.

Conclusion

THIS ACTION IS MADE FINAL. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to JONATHAN KIM whose telephone number is (571)272-3327. The examiner can normally be reached on Monday-Friday 7:30AM to 5:00PM EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Pierre-Louis Desir can be reached on (571)272-7799. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2659

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

/JONATHAN KIM/
Examiner, Art Unit 2659

/PIERRE-LOUIS DESIR/
Supervisory Patent Examiner, Art Unit 2659

Notice of References Cited	Application/Control No. 14/752,400	Applicant(s)/Patent Under Reexamination LIU ET AL.	
	Examiner JONATHAN KIM	Art Unit 2659	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-2015/0106085 A1	04-2015	Lindahl; Aram M.	G10L15/32	704/231
*	B	US-2011/0264447 A1	10-2011	Visser; Erik	G10L25/78	704/208
*	C	US-2013/0332168 A1	12-2013	Kim; Byoungju	G10L15/22	704/251
*	D	US-2005/0096898 A1	05-2005	Singhal, Manoj	G10H1/125	704/205
*	E	US-2016/0125867 A1	05-2016	Jarvinen; Kari Juhani	H04S7/30	381/73.1
*	F	US-2014/0278393 A1	09-2014	Ivanov; Plamen A.	G10L15/20	704/233
	G	US-				
	H	US-				
	I	US-				
	J	US-				
	K	US-				
	L	US-				
	M	US-				

FOREIGN PATENT DOCUMENTS

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

NON-PATENT DOCUMENTS

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

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

U.S. Patent and Trademark Office
PTO-892 (Rev. 01-2001)

Notice of References Cited

Part of Paper No. 20161021