



(51) International Patent Classification:

G10L 15/22 (2006.01) *G06F 1/16* (2006.01)
G10L 21/0208 (2013.01) *G10L 21/0216* (2013.01)
G10L 25/30 (2013.01)

(21) International Application Number:

PCT/US2016/062970

(22) International Filing Date:

20 November 2016 (20.11.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

14/978,015 22 December 2015 (22.12.2015) US

(71) Applicant: **INTEL CORPORATION** [US/US]; 2200 Mission College Boulevard, Santa Clara, California 95054-1549 (US).

(72) Inventors: **KAR, Swarnendu**; 1396 NE Carlabay Way, Apt. 135, Hillsboro, Oregon 97124 (US). **PHATAK, Gauri**; 18985 NW Northshore Ct., Portland, Oregon

97229 (US). **SHAH, Saurin**; 9241 NW Murdock Street, Portland, Oregon 97229 (US).

(74) Agent: **GUPTA, Rishi**; Patent Capital Group, 2816 Lago Vista Lane, Rockwall, Texas 75032 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,

[Continued on next page]

(54) Title: WEARER VOICE ACTIVITY DETECTION

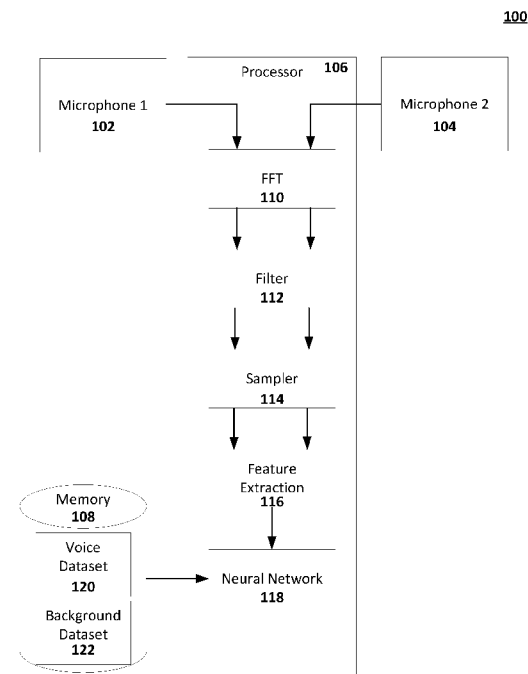


FIG. 1

[Continued on next page]



TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

(57) Abstract: Embodiments include a wearable device, such as a head-worn device. The wearable device includes a first microphone to receive a first sound signal from a wearer of the wearable device; a second microphone to receive a second sound signal from the wearer of the wearable device; and a processor to process the first sound signals and the second sound signals to determine that the first and second sound signals originate from the wearer of the wearable device.

WEARER VOICE ACTIVITY DETECTION

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of priority to U.S. Nonprovisional Patent Application No. 14/978,015 filed 22 December 2015 entitled, "WEARER VOICE ACTIVITY DETECTION", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] This disclosure pertains to wearer voice activity detection, and in particular, to wearer voice activity detection using bimodal microphones.

BACKGROUND

[0003] Single modal microphones for device-based speech recognition and dialog systems provide a way for a user to interact with a wearable device. Single modal microphones allow for a user to provide instructions to a wearable device to elicit responses, such as directions, information, responses to queries, confirmations of command executions, etc.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] FIG. 1 is a schematic block diagram of a wearable device that includes bimodal microphones.

[0005] FIG. 2 is a schematic block diagram of processing inputs from each of two microphones.

[0006] FIG. 3 is a schematic block diagram of using bimodal inputs to determine an origination of speech.

[0007] FIG. 4 is a process flow diagram for processing bimodal inputs.

DETAILED DESCRIPTION

[0008] This disclosure describes a wearable device, such as a head-worn device that includes two microphones. The use of two microphones can be used to clarify the originator of speech input to a dialog system. Though a single modality (either air or bone-conduction) by itself is not sufficiently reliable for achieving wearer voice activity detection (VAD), having two modalities simultaneously changes the way VAD is processed. The relative transfer function (between the two modes) has enough information to make a distinction between wearer speaking and ambient audio.

[0009] This disclosure describes wearable devices such as glasses/earphones with audio based command/control interface. Such a device uses voice commands for various purposes like navigation, browsing calendars, setting alarms and web search. One problem with these devices is that the source of audio cannot easily be ascertained as coming from the wearer of the device. For example, when Person-A wears a Google glass and Person-B says "ok google", the device still unlocks, which is clearly

not desirable. This disclosure addresses this problem by using a combination of hardware and software referred to herein as wearer voice activity detection (wearer VAD), which uses bimodal microphone signal to figure out whether the wearer is speaking or not. By bimodal, we mean a combination of an air microphone (the ordinary kind) and another microphone where the audio signal is transmitted through a non-air medium (example, bone conduction or in-ear mics).

[0010] In particular, this disclosure describes:

- 1) Capture and fusion of two different audio modalities on a wearable audio device; and
- 2) Software that harnesses the difference of modalities to compute a wearer voice activity indicator.

[0011] Potential uses of wearer VAD extend well beyond unlocking a device or triggering a conversation – it can help with traditionally difficult audio processing tasks like noise reduction, speech enhancement and speech separation.

[0012] In this IDF, a mode of a microphone can be defined as the medium in which sound is captured. Examples include bone conduction microphones, in-ear microphones, both of which use the vibration through a skull and ear-canal respectively as the medium of transfer of sound. Air microphones use air as the medium of transfer.

[0013] Audio captured simultaneously through two different modes jointly carries definitive evidence of wearer voice activity. The term “wearer” is used to mean any user wearing that particular device and no user-specific pre-registration is involved.

[0014] Sound produced by human beings is based on the shape of their vocal tracts and differs widely in frequency content as a result. Audio transmitted through non-air media (bone/in-ear) undergoes frequency distortion vis-à-vis the original signal.

[0015] By having two different modes of simultaneous capture, we can compute the intermodal relation between the signals, which happens to carry enough information to make a distinction. In other words, the absolute frequency content is not important in this case, it is the relative transfer function that provides the information for wearer VAD. When the source of sound is in the background, the two modalities of signal appear more identical in its frequency characteristics.

[0016] However, when the source of sound is the wearer, some frequencies in the bone conduction transmission are attenuated drastically, leading to a less identical appearance in some frequency bands. This distinction is what makes wearer VAD possible.

[0017] This disclosure makes use of multimodal microphones and can be used on any head mounted wearable where the device is in contact with the person’s nose, throat, skull or ear canal, e.g. Glasses, headbands, earphones, helmets, headphones.

[0018] This is to facilitate the use of either bone conduction or in-ear microphones. At least one of these types of mic is required in addition to an ordinary air microphone for the invention to work.

[0019] The algorithm for wearer VAD is based on making a binary decision on whether the source of sound is the wearer or something in the background. A binary classifier uses information contained in the relative transfer function between the two modalities of signal, e.g., bone-conduction and ordinary air microphones. The idea is that when the source of sound is in the background, the two modalities of signal appear more identical in its frequency characteristics. However, when the source of sound is the wearer, some frequencies in the bone-conduction transmission are attenuated drastically, leading to a less identical appearance in some frequency bands. This difference in relative transfer function is learnt by extracting several features in individual frequency sub-bands and performing a Neural network classification that is trained on data collected from a diverse population.

[0020] The Neural Network can be trained to differentiate between 2 situations – A) wearer is speaking (voice) and B) someone else is speaking (background). Audio can be collected from 16 different male and female subjects with different accents and voice kinds. For voice data, each subject was asked to utter a few phrases in both dialog and conversational mode. For background data, each subject was made to sit quiet wearing the device while an audio clip containing several phrases in dialog and conversational modes was played in the background at different intensity levels. In the testing, the training based on 16 subjects proved to be sufficiently robust and works for new and different test subjects.

[0021] FIG. 1 is a schematic block diagram of a wearable device 100 that includes bimodal microphones. Wearable device 100 may be a pair of smart glasses and earphones with audio based command/control interface, a helmet, earbuds. The wearable device 100 includes a first microphone 102 and a second microphone 104. The first microphone 102 and the second microphone 104 can be different modes. For example, the first microphone 102 can be an air microphone, while the second microphone 104 can be a bone conduction microphone or in-ear microphone. In some embodiments a third microphone can be used to encompass three types of modes. The first microphone 102 can receive a sound signal at the same time as the second microphone 104. The first microphone 102 can provide a first sound signal representative of the received sound signal to a processor 106. The second microphone 104 can provide a second sound signal representative of the received sound signal to the processor 106. The processor 106 can process the first and second sound signals to determine an originator of the received sound signal.

[0022] The processor can use a fast Fourier transform (FFT) 110 to filter the first and second sound signals in the frequency domain. Other filtering can be performed to reduce noise, etc. using filtering logic 112. The first and second sound signals can be sampled using a sampler 114. Features

can be extracted using a feature extraction module 116. The extracted features are fused and provided to a neural network 118.

[0023] The wearable device 100 also includes a memory 108. Memory 108 includes a voice dataset 120 and a background dataset 122. The voice dataset 120 and the background dataset 122 are preprogrammed into the wearable device 100 and are the result of training the neural network.

[0024] FIG. 2 is a schematic block diagram 200 of processing inputs from each of two microphones. FIG. 2 describes the feature extraction part of the algorithm. Mic 1 202 and Mic 2 204 receive a sound signal at substantially the same time. Audio from both the modalities are filtered by a fast Fourier transfer (FFT) 206 and FFT 208, respectively; each audio signal can undergo additional filtering 210 and 212. Each audio signal is subsampled based on 5 uniformly spaced bands between 400Hz and 3200Hz. For mic 1 202, the audio signal can be subsampled into five 16 millisecond frames X1, X2, X3, X4, and X5 to facilitate audio analysis. For mic 2 204, the audio signal can be subsampled into five 16 millisecond frames Y1, Y2, Y3, Y4, and Y5 to facilitate audio analysis. Frames from each subband-pair are fused to extract certain features such as energy ratio, mutual entropy, and correlation. For example, the data from corresponding subband frames from each microphone are combined to calculate each feature for those frames. Features from all 5 subband-pairs are concatenated to form feature vectors FE1, FE2, FE3, FE4, and FE5. For each pair of subbands, features are extracted from each audio signal, for a total of 15 features (e.g., 3 features per subband). FE1, therefore, is a concatenation of features from X1 and features from Y1, which forms the feature vector $F1 = \{\text{gain}(X1, Y1), \text{entropy}(X1, Y1), \text{correlation}(X1, Y1)\}$. F1, thus, has three values, each value representing a feature of the audio signal for each mode at a subband. The feature vectors FE1–FE5 form the input to the neural network shown in FIG. 3.

[0025] FIG. 3 is a schematic block diagram 300 of using extracted features from bimodal microphone signals to determine an origination of audio signals. FIG. 3 describes the classification algorithm and structural features for determining whether the audio signal originates from the wearer. The feature vectors FE1–FE5 are fed to a trained neural network 304 that is pre-trained with a diverse voice and background dataset (e.g., voice dataset 120 and background dataset 122 stored in memory 108 in FIG. 1). The neural network 304 is a system of non-linear functions, which are applied to the feature vectors FE1–FE5. The neural network 304 operates on each of the feature vector values. The neural network 304 outputs a voice probability Q[1] 306 that represents the probability that the audio signal originates from the wearer of the wearable device. Another output 1-Q[1] 308 can represent the probability that the audio signal originates as background noise.

[0026] The neural network 304 can include weighted rules that the neural network 304 applies to the feature vectors FE1–FE5. The neural network 304 is trained using example data prior to being worn by the wearer. The training can include using speech patterns from a variety of speakers wearing

a wearable device that includes two modes of microphones. As an example, when the audio signal originates from the wearer of the wearable device, there is a high correlation between the audio signals from each microphone at high frequencies. When the audio signal originates from a non-wearer, there is a high correlation between audio signals from each microphone in all frequencies. The neural network 304 applies the set of weighted rules to the feature vectors to provide the probability output.

[0027] In some implementations, a soft max function is applied to the output of the neural network 304 to accentuate the winning decision and convert scores to class-probabilities that add to 1 (e.g., $Q[1]$ 306 and $1-Q[1]$ 308). Since each frame is 16 milliseconds long, 60 probability values for each class per second can be generated. $Q[1]$ 306 can be considered the wearer's voice probability. The probability of voice $Q[1]$ 306 is further smoothed over time to get rid of spurious decisions (310). A threshold 312 is subsequently applied to make a final decision 314. For example, if the probability $Q[1]$ is greater than a predetermined threshold (e.g., a threshold value determined during training), the wearable device can determine that the audio signal is the voice of the wearer. The wearable device can then provide the audio signal to a dialog engine or other features of the wearable device.

[0028] In some cases, the wearable device can determine based on the probability threshold 312 that the audio signal does not originate from the wearer, in which case the wearable device can discard the audio signal or request clarification from the wearer as to whether the wearer is attempting to communicate with the wearable device.

[0029] FIG. 4 is a process flow diagram 400 for processing bimodal inputs. A first audio signal is received at a first microphone (402a). A second audio signal is received at a second microphone (402b). The first and second audio signals are from the same source at the same time. The first and second audio signals can be received from each microphone substantially simultaneously. The first audio signal can be filtered using FFT (404a). The second audio signal can be filtered using FFT (404b). The first audio signal can be sampled (406a). The second audio signal can be sampled (406b). The sampling can be performed at 8kHz, for example. The first audio signal can be subsampled based on 5 uniformly spaced bands between 400Hz and 3200Hz (408a). The second audio signal can be subsampled based on 5 uniformly spaced bands between 400Hz and 3200Hz (408b).

[0030] Features can be extracted from each frame of the subsampled first audio signal (410a). Features can also be extracted from each frame of the subsampled first audio signal (410b). Feature vectors can be created using the extracted features (412). The feature vectors can be processed using a neural network (414). The neural network can output a probability of whether the audio signal originated from the based on the feature vectors. The probability can be compared to a threshold level to conclude whether the audio signal originates from the wearer of the wearable device or from background signals/noise (416).

[0031] Example 1 is a wearable device that includes a first microphone to receive a first sound signal from a wearer of the wearable device; a second microphone to receive a second sound signal from the wearer of the wearable device; a processor to process the first sound signals and the second sound signals to determine that the first and second sound signals originate from the wearer of the wearable device.

[0032] Example 2 may include the subject matter of example 1, wherein the first microphone comprises an air microphone.

[0033] Example 3 may include the subject matter of any of examples 1 or 2, wherein the second microphone comprises one of a bone conduction microphone or an in-ear microphone.

[0034] Example 4 may include the subject matter of any of examples 1 or 2 or 3, wherein the processor is configured to process the first sound signal and the second sound signal by sampling the first sound signal and the second sound signal; extracting one or more features from the first sound signal and from the second sound signal; and determining that the first and second sound signals originate from the wearer by comparing extracted features from the first sound signal and from the second sound signal.

[0035] Example 5 may include the subject matter of any of claims 1 or 4, further comprising a neural network to process extracted features from each of the first sound signal and from the second sound signal.

[0036] Example 6 may include the subject matter of example 5, wherein the neural network is trained with a voice dataset and a background dataset, and wherein the neural network is configured to determine based on the voice dataset, the background dataset, and the extracted features from the first and second sound signals that the first and second sound signals original from the wearer of the wearable device.

[0037] Example 7 may include the subject matter of any of examples 1 or 4, further comprising a Fast Fourier Transform module to filter the first and second sound signals.

[0038] Example 8 may include the subject matter of any of examples 1 or 4, wherein the processor is configured to process the first sound signal and the second sound signal by splitting the first sound signal into a first set of subparts, the first set of subparts comprising a frame representing a portion in time of the first sound signal; splitting the second sound signal into a second set of subparts, the second set of subparts comprising a frame representing a portion in time of the second sound signal; combining a frame from the first set of subparts with a corresponding frame from the second set of subparts; and extracting a feature of the first and second sound signals based on combining the frame from the first set of subparts with the corresponding frame from the second set of subparts.

[0039] Example 9 may include the subject matter of example 1, wherein the wearable device comprises a head-worn device.

[0040] Example 10 is a method comprising receiving a first sound signal at a wearable device from a first microphone; receiving a second sound signal at the wearable device from a second microphone; determining that the first and second sound signals originate from the wearer of the wearable device.

[0041] Example 11 may include the subject matter of example 10, wherein the first microphone comprises an air microphone.

[0042] Example 12 may include the subject matter of any of examples 10 or 11, wherein the second microphone comprises one of a bone conduction microphone or an in-ear microphone.

[0043] Example 13 may include the subject matter of example 10, wherein processing the first sound signal and the second sound signal comprises sampling the first sound signal and the second sound signal; extracting one or more features from the first sound signal and from the second sound signal; and determining that the first and second sound signals originate from the wearer by comparing extracted features from the first sound signal and from the second sound signal.

[0044] Example 14 may include the subject matter of any of examples 10 or 13, further comprising processing the extracted features from each of the first sound signal and from the second sound signal using a neural network.

[0045] Example 15 may include the subject matter of example 14, wherein the neural network is trained with a voice dataset and a background dataset, and wherein the neural network is configured to determine based on the voice dataset, the background dataset, and the extracted features from the first and second sound signals that the first and second sound signals originate from the wearer of the wearable device.

[0046] Example 16 may include the subject matter of any of examples 10 or 13, further comprising filtering the first and second sound signals using a Fast Fourier Transform module.

[0047] Example 17 may include the subject matter of any of examples 10 or 13, wherein processing the first sound signal and the second sound signal comprises splitting the first sound signal into a first set of subparts, the first set of subparts comprising a frame representing a portion in time of the first sound signal; splitting the second sound signal into a second set of subparts, the second set of subparts comprising a frame representing a portion in time of the second sound signal; combining a frame from the first set of subparts with a corresponding frame from the second set of subparts; and extracting a feature of the first and second sound signals based on combining the frame from the first set of subparts with the corresponding frame from the second set of subparts.

[0048] Example 18 is a computer program product tangibly embodied on a non-transient computer readable medium, the computer program product comprising instructions operable when executed to receive from a first microphone a first sound signal from a wearer of the wearable device; receive from a second microphone a second sound signal from the wearer of the wearable device; and

process the first sound signals and the second sound signals to determine that the first and second sound signals originate from the wearer of the wearable device.

[0049] Example 19 may include the subject matter of example 18, wherein the first microphone comprises an air microphone.

[0050] Example 20 may include the subject matter of any of examples 18 or 19, wherein the second microphone comprises one of a bone conduction microphone or an in-ear microphone.

[0051] Example 21 may include the subject matter of example 18, wherein the processor is configured to process the first sound signal and the second sound signal by sampling the first sound signal and the second sound signal; extracting one or more features from the first sound signal and from the second sound signal; and determining that the first and second sound signals originate from the wearer by comparing extracted features from the first sound signal and from the second sound signal.

[0052] Example 22 may include the subject matter of any of examples 18 or 21, further comprising a neural network to process extracted features from each of the first sound signal and from the second sound signal.

[0053] Example 23 may include the subject matter of example 22, wherein the neural network is trained with a voice dataset and a background dataset, and wherein the neural network is configured to determine based on the voice dataset, the background dataset, and the extracted features from the first and second sound signals that the first and second sound signals originate from the wearer of the wearable device.

[0054] Example 24 may include the subject matter of any of examples 18 or 21, further comprising a Fast Fourier Transform module to filter the first and second sound signals.

[0055] Example 25 may include the subject matter of any of examples 18 or 21, wherein the processor is configured to process the first sound signal and the second sound signal by splitting the first sound signal into a first set of subparts, the first set of subparts comprising a frame representing a portion in time of the first sound signal; splitting the second sound signal into a second set of subparts, the second set of subparts comprising a frame representing a portion in time of the second sound signal; combining a frame from the first set of subparts with a corresponding frame from the second set of subparts; and extracting a feature of the first and second sound signals based on combining the frame from the first set of subparts with the corresponding frame from the second set of subparts.

[0056] Advantages of the present disclosure are readily apparent to those of skill in the art. Among the various advantages of the present disclosure include the following:

[0057] Aspects of the present disclosure can provide an enhanced user experience when using an interactive wearable device, such as a head-worn device, by using multiple microphones of different types of modes.

[0058] While this specification contains many specific implementation details, these should not be construed as limitations on the scope of any disclosures or of what may be claimed, but rather as descriptions of features specific to particular embodiments of particular disclosures. Certain features that are described in this specification in the context of separate embodiments can also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment can also be implemented in multiple embodiments separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claimed combination may be directed to a subcombination or variation of a subcombination.

[0059] Similarly, while operations are depicted in the drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system components in the embodiments described above should not be understood as requiring such separation in all embodiments, and it should be understood that the described program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

[0060] Thus, particular embodiments of the subject matter have been described. Other embodiments are within the scope of the following claims. In some cases, the actions recited in the claims can be performed in a different order and still achieve desirable results. In addition, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results.

CLAIMS:

1. A wearable device comprising:
a first microphone to receive a first sound signal from a wearer of the wearable device;
a second microphone to receive a second sound signal from the wearer of the wearable device;
a processor to process the first sound signals and the second sound signals to determine that the first and second sound signals originate from the wearer of the wearable device.
2. The wearable device of claim 1, wherein the first microphone comprises an air microphone.
3. The wearable device of any of claims 1 or 2, wherein the second microphone comprises one of a bone conduction microphone or an in-ear microphone.
4. The wearable device of claim 1, wherein the processor is configured to process the first sound signal and the second sound signal by:
sampling the first sound signal and the second sound signal;
extracting one or more features from the first sound signal and from the second sound signal;
and
determining that the first and second sound signals originate from the wearer by comparing extracted features from the first sound signal and from the second sound signal.
5. The wearable device of any of claims 1 or 4, further comprising a neural network to process extracted features from each of the first sound signal and from the second sound signal.
6. The wearable device of claim 5, wherein the neural network is trained with a voice dataset and a background dataset, and wherein the neural network is configured to determine based on the voice dataset, the background dataset, and the extracted features from the first and second sound signals that the first and second sound signals originate from the wearer of the wearable device.
7. The wearable device of any of claims 1 or 4, further comprising a Fast Fourier Transform module to filter the first and second sound signals.
8. The wearable device of any of claims 1 or 4, wherein the processor is configured to process the first sound signal and the second sound signal by:

splitting the first sound signal into a first set of subparts, the first set of subparts comprising a frame representing a portion in time of the first sound signal;

splitting the second sound signal into a second set of subparts, the second set of subparts comprising a frame representing a portion in time of the second sound signal;

combining a frame from the first set of subparts with a corresponding frame from the second set of subparts; and

extracting a feature of the first and second sound signals based on combining the frame from the first set of subparts with the corresponding frame from the second set of subparts.

9. The wearable device of claim 1, wherein the wearable device comprises a head-worn device.

10. A method comprising:

receiving a first sound signal at a wearable device from a first microphone;

receiving a second sound signal at the wearable device from a second microphone;

determining that the first and second sound signals originate from the wearer of the wearable device.

11. The method of claim 10, wherein the first microphone comprises an air microphone.

12. The method of any of claims 10 or 11, wherein the second microphone comprises one of a bone conduction microphone or an in-ear microphone.

13. The method of claim 10, wherein processing the first sound signal and the second sound signal comprises

sampling the first sound signal and the second sound signal;

extracting one or more features from the first sound signal and from the second sound signal;

and

determining that the first and second sound signals originate from the wearer by comparing extracted features from the first sound signal and from the second sound signal.

14. The method of any of claims 10 or 13, further comprising processing the extracted features from each of the first sound signal and from the second sound signal using a neural network.

15. The method of claim 14, wherein the neural network is trained with a voice dataset and a background dataset, and wherein the neural network is configured to determine based on the voice

dataset, the background dataset, and the extracted features from the first and second sound signals that the first and second sound signals original from the wearer of the wearable device.

16. The method of any of claims 10 or 13, further comprising filtering the first and second sound signals using a Fast Fourier Transform module.

17. The method of any of claims 10 or 13, wherein processing the first sound signal and the second sound signal comprises:

splitting the first sound signal into a first set of subparts, the first set of subparts comprising a frame representing a portion in time of the first sound signal;

splitting the second sound signal into a second set of subparts, the second set of subparts comprising a frame representing a portion in time of the second sound signal;

combining a frame from the first set of subparts with a corresponding frame from the second set of subparts; and

extracting a feature of the first and second sound signals based on combining the frame from the first set of subparts with the corresponding frame from the second set of subparts.

18. A computer program product tangibly embodied on a non-transient computer readable medium, the computer program product comprising instructions operable when executed to:

receive from a first microphone a first sound signal from a wearer of the wearable device;

receive from a second microphone a second sound signal from the wearer of the wearable device;

process the first sound signals and the second sound signals to determine that the first and second sound signals originate from the wearer of the wearable device.

19. The computer program product of claim 18, wherein the first microphone comprises an air microphone.

20. The computer program product of claim 18 or 19, wherein the second microphone comprises one of a bone conduction microphone or an in-ear microphone.

21. The computer program product of claim 18, further comprising instructions operable when executed to:

sample the first sound signal and the second sound signal;

extract one or more features from the first sound signal and from the second sound signal; and

determine that the first and second sound signals originate from the wearer by comparing extracted features from the first sound signal and from the second sound signal.

22. The computer program product of claim 18 or 21, further comprising a neural network to process extracted features from each of the first sound signal and from the second sound signal.

23. The computer program product of claim 22, wherein the neural network is trained with a voice dataset and a background dataset, and wherein the neural network is configured to determine based on the voice dataset, the background dataset, and the extracted features from the first and second sound signals that the first and second sound signals originate from the wearer of the wearable device.

24. The computer program product of claim 18 or 21, further comprising instructions operable when executed to filter the first and second sound signals based on a Fast Fourier Transform.

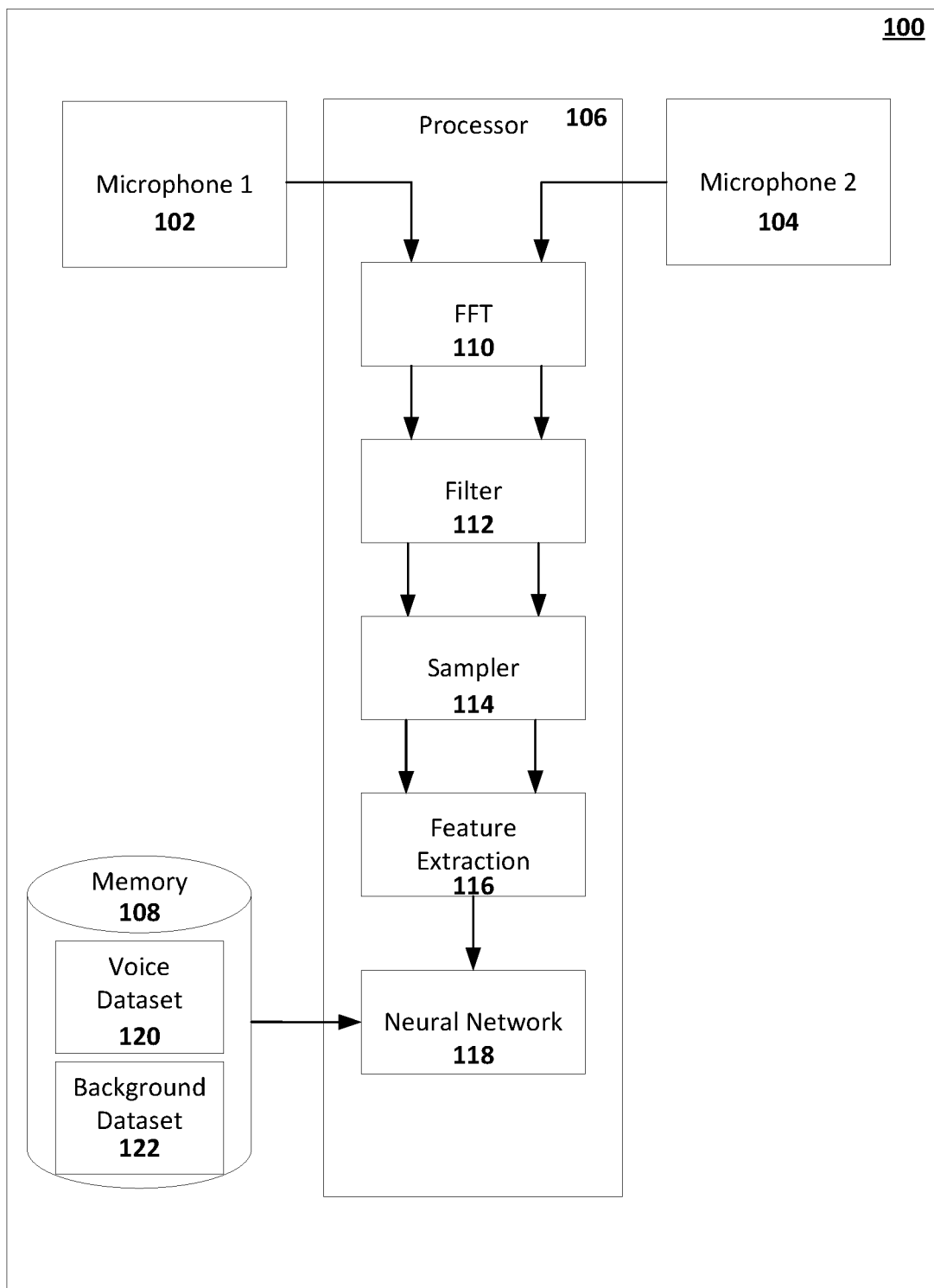
25. The computer program product of claim 18 or 21, further comprising instructions operable when executed to:

splitting the first sound signal into a first set of subparts, the first set of subparts comprising a frame representing a portion in time of the first sound signal;

splitting the second sound signal into a second set of subparts, the second set of subparts comprising a frame representing a portion in time of the second sound signal;

combining a frame from the first set of subparts with a corresponding frame from the second set of subparts; and

extracting a feature of the first and second sound signals based on combining the frame from the first set of subparts with the corresponding frame from the second set of subparts.

**FIG. 1**

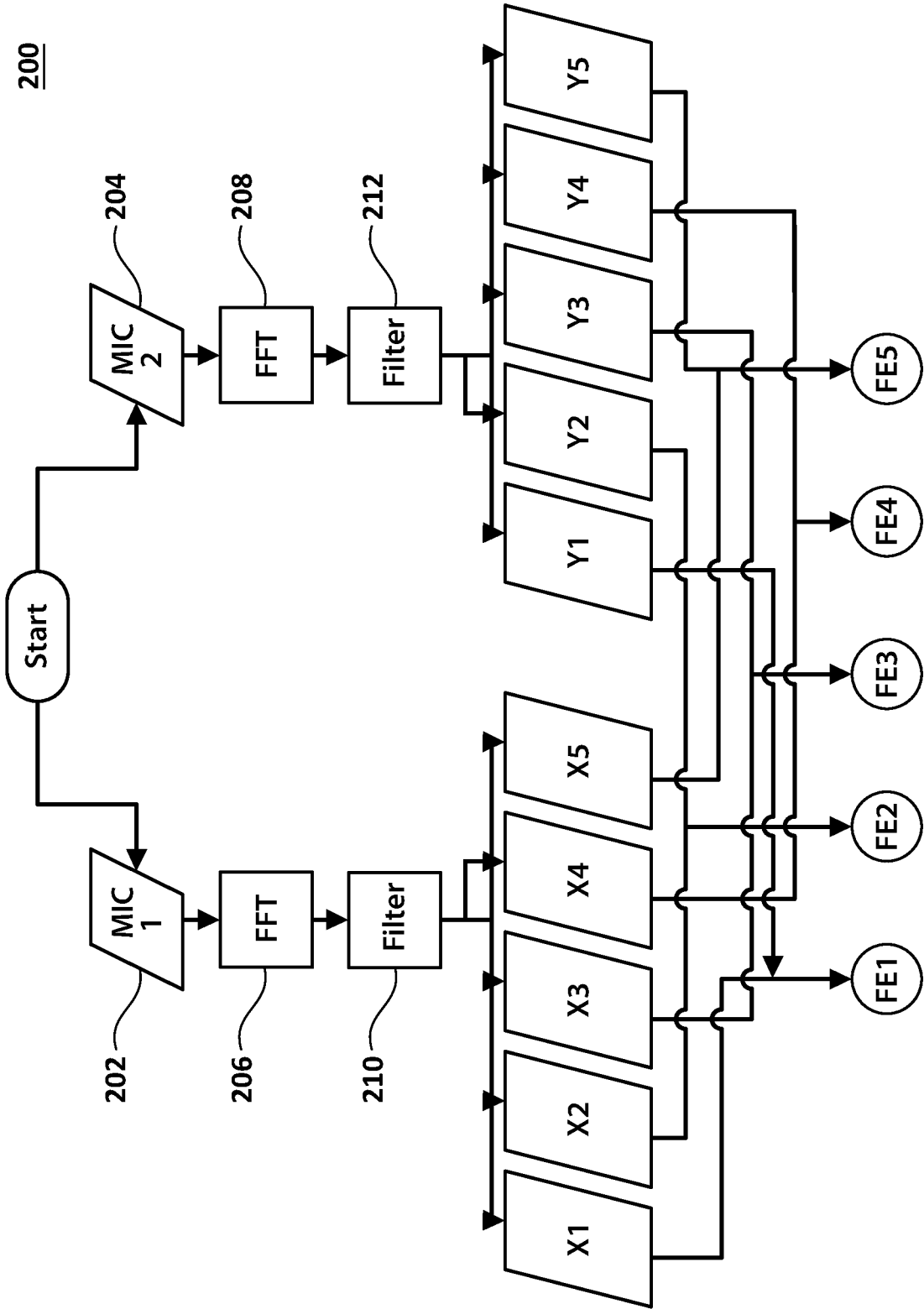


FIG. 2

300

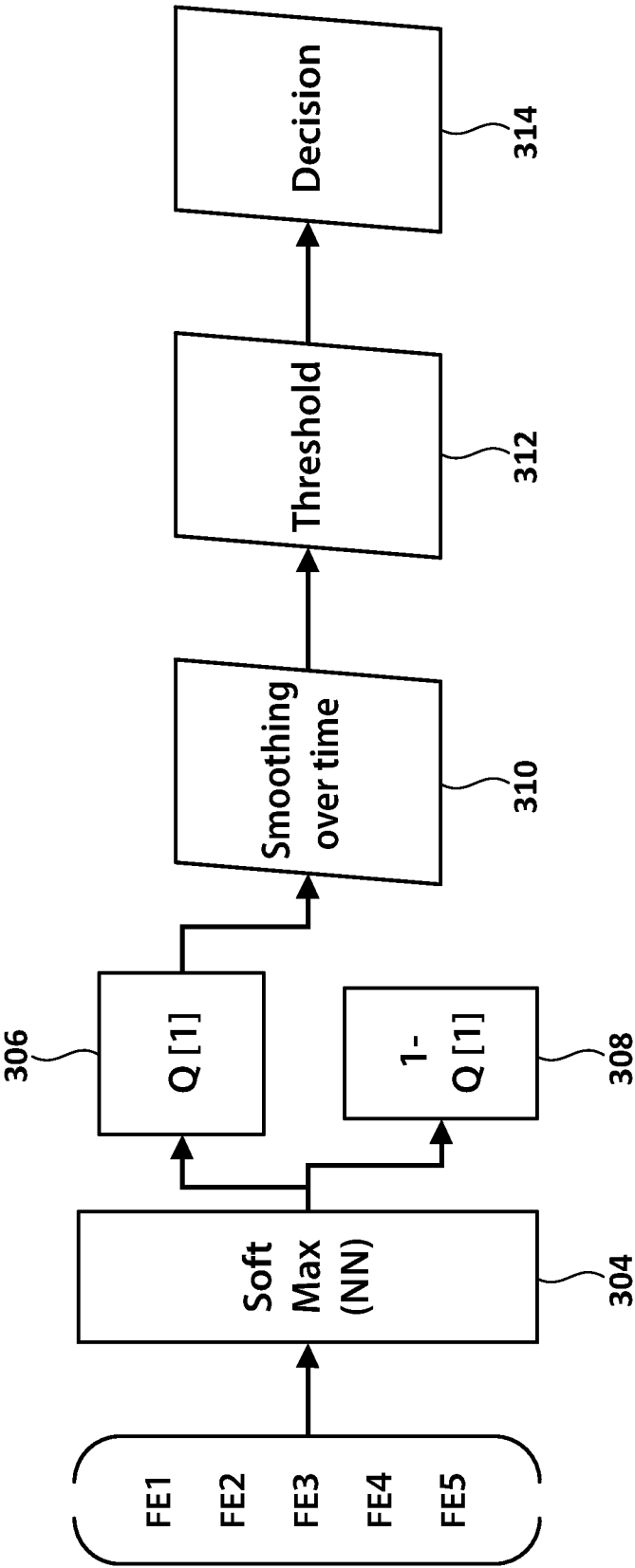
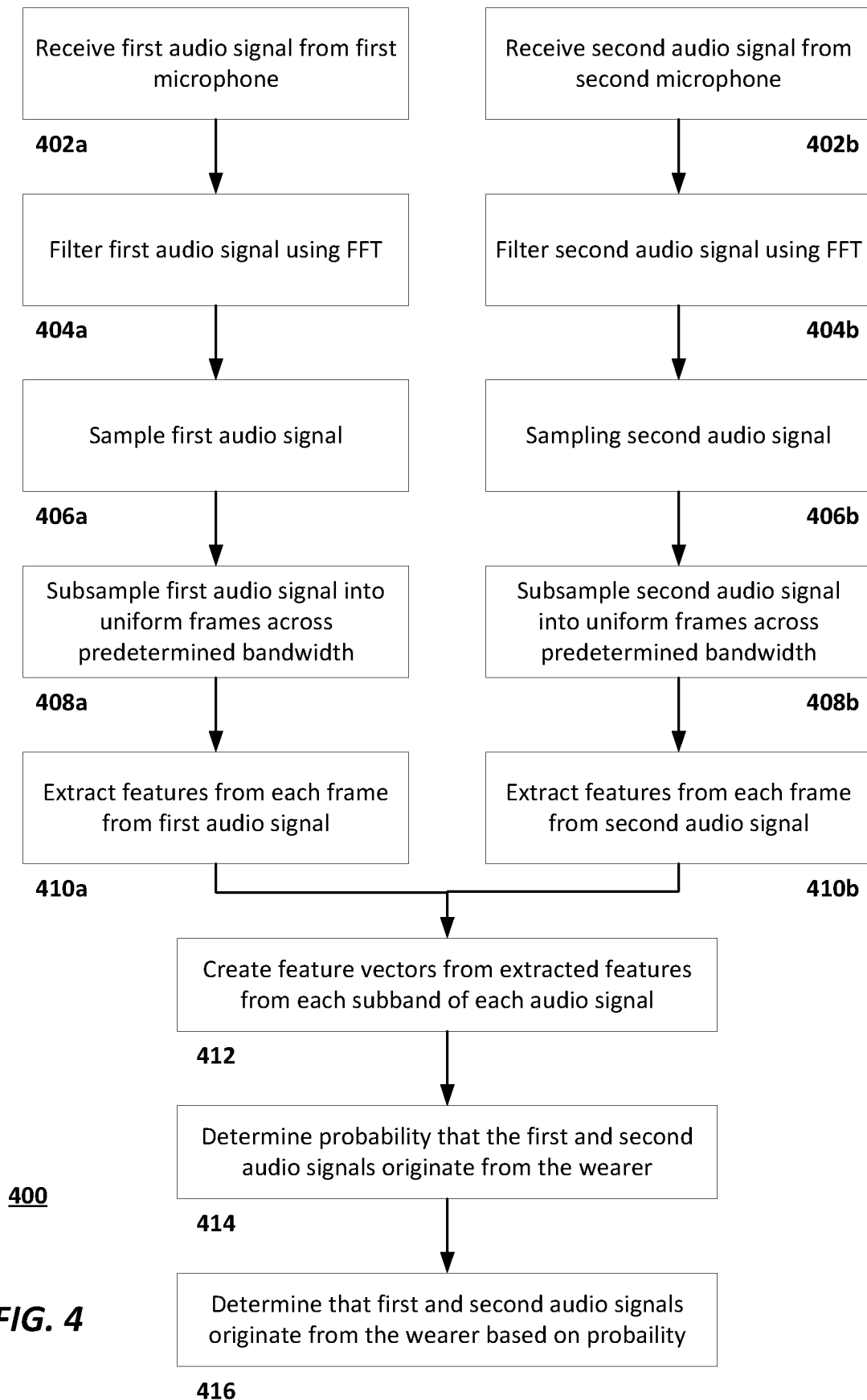


FIG. 3



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/062970**A. CLASSIFICATION OF SUBJECT MATTER****G10L 15/22(2006.01)i, G10L 21/0208(2013.01)i, G10L 25/30(2013.01)i, G06F 1/16(2006.01)i, G10L 21/0216(2013.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15/22; H04R 25/00; G10L 15/00; H04R 3/00; G10L 15/14; G06F 21/62; G10L 21/0208; G10L 25/30; G06F 1/16; G10L 21/0216

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: wearable device, bimodal microphone, sound signal, originate, voice activity detection, and similar terms.**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014-0010397 A1 (STARKEY LABORATORIES, INC.) 09 January 2014 See paragraphs [0004], [0007], and [0021]-[0026]; claim 1; and figures 1A-3.	1-3, 9-12, 18-20
Y		4-8, 13-17, 21-25
Y	US 7,254,538 B1 (HYNEK HERMAN SKY et al.) 07 August 2007 See column 3, lines 47-56; column 4, lines 7-10; and figure 2.	4-6, 13-15, 21-23
Y	US 8,625,819 B2 (STEVEN WAYNE GOLDSTEIN et al.) 07 January 2014 See column 7, line 62 - column 8, line 4; column 8, lines 49-58; and column 11, lines 51-55.	7-8, 16-17, 24-25
A	US 2012-0230526 A1 (TAO ZHANG) 13 September 2012 See paragraphs [0019]-[0020]; claim 1; and figures 1-2.	1-25
A	US 2013-0263284 A1 (GERALD R. BLACK et al.) 03 October 2013 See claims 1-3 and figures 3A-8.	1-25



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

10 March 2017 (10.03.2017)

Date of mailing of the international search report

10 March 2017 (10.03.2017)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

NHO, Ji Myong

Telephone No. +82-42-481-8528



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2016/062970

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014-0010397 A1	09/01/2014	EP 2242289 A1 EP 2242289 B1 US 2010-260364 A1 US 2015-0043765 A1 US 2016-029131 A1 US 8477973 B2 US 9094766 B2 US 9219964 B2	20/10/2010 28/12/2016 14/10/2010 12/02/2015 28/01/2016 02/07/2013 28/07/2015 22/12/2015
US 7254538 B1	07/08/2007	None	
US 8625819 B2	07/01/2014	US 2009-010456 A1 US 2014-126748 A1 US 2016-088391 A1 US 9204214 B2	08/01/2009 08/05/2014 24/03/2016 01/12/2015
US 2012-0230526 A1	13/09/2012	AT 540538 T CA 2639572 A1 DK 2040486 T3 EP 2040486 A2 EP 2040486 A3 EP 2040486 B1 US 2009-074201 A1 US 8031881 B2 US 9210518 B2	15/01/2012 18/03/2009 10/04/2012 25/03/2009 20/10/2010 04/01/2012 19/03/2009 04/10/2011 08/12/2015
US 2013-0263284 A1	03/10/2013	CA 2647194 A1 CA 2647194 C US 2010-0075631 A1 US 2010-0311390 A9 US 2016-086402 A1 US 9213861 B2 WO 2008-008101 A2 WO 2008-008101 A3	17/01/2008 16/08/2016 25/03/2010 09/12/2010 24/03/2016 15/12/2015 17/01/2008 17/04/2008