



(51) International Patent Classification:

G10L 17/10 (2013.01) G06K 9/00 (2006.01)

G06F 21/32 (2013.01) A61B 5/12 (2006.01)

G06F 21/40 (2013.01)

(21) International Application Number:

PCT/GB2019/052302

(22) International Filing Date:

15 August 2019 (15.08.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

16/118,950 31 August 2018 (31.08.2018) US

(71) Applicant: **CIRRUS LOGIC INTERNATIONAL SEMICONDUCTOR LIMITED** [GB/GB]; 7B Nightingale Way, Quatermile, Edinburgh Central Scotland EH3 9EG (GB).

(72) Inventor: **LESSO, John Paul**; 8 Galachlaw Shot, Edinburgh Lothian EH10 7JF (GB).

(74) Agent: **HALL, Christopher**; Haseltine Lake Kempner LLP, Redcliff Quay, 120 Redcliff Street, Bristol, Bristol BS1 6HU (GB).

(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, JO, JP, KE, KG, KH, 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, 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).

Published:

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

(54) Title: BIOMETRIC AUTHENTICATION

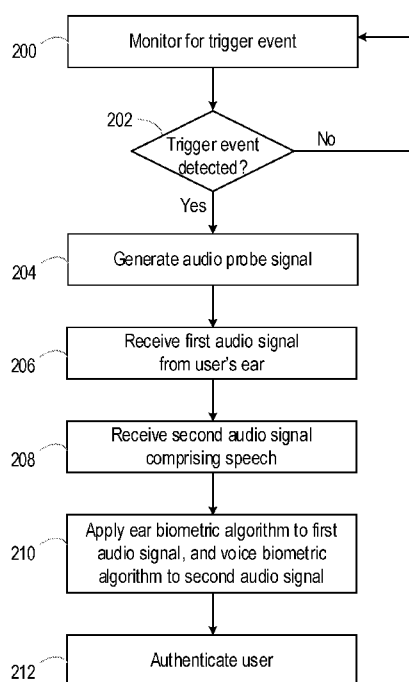


Figure 2

(57) Abstract: A method for authenticating a user of an electronic device is disclosed. The method comprises: responsive to detection of a trigger event indicative of a user interaction with the electronic device, generating an audio probe signal to play through an audio transducer of the electronic device; receiving a first audio signal comprising a response of the user's ear to the audio probe signal; receiving a second audio signal comprising speech of the user; and applying an ear biometric algorithm to the first audio signal and a voice biometric algorithm to the second audio signal to authenticate the user as an authorised user.

BIOMETRIC AUTHENTICATION

Technical field

Embodiments of the disclosure relate to biometric authentication, and particularly to
5 methods, apparatus and computer-readable mediums for authenticating a user based
on ear biometric data.

Background

It is known that the acoustic properties of a user's ear, whether the outer parts (known
10 as the pinna or auricle), the ear canal or both, differ substantially between individuals
and can therefore be used as a biometric to identify the user. One or more
loudspeakers or similar transducers positioned close to or within the ear generate a
stimulus, and one or more microphones similarly positioned close to or within the ear
detect the response of the ear to the stimulus. One or more features may be extracted
15 from the response signal, and used to characterize an individual.

For example, the ear canal is a resonant system, and therefore one feature which may
be extracted from the response signal is the resonant frequency of the ear canal. If the
measured resonant frequency (i.e. in the response signal) differs from a stored
20 resonant frequency for the user, a biometric algorithm coupled to receive and analyse
the response signal may return a negative result. Other features of the response signal
may be similarly extracted and used to characterize the individual. For example, the
features may comprise one or more mel frequency cepstrum coefficients. More
generally, the transfer function between the stimulus and the measured response
25 signal (or features of the transfer function) may be determined, and compared to a
stored transfer function (or stored features of the transfer function) which is
characteristic of the user.

Other acoustic properties of a user's ear include otoacoustic emissions, whether
30 spontaneous (i.e., generated by the ear without external stimulation) or evoked (i.e.,
generated by the ear in response to a stimulus signal such as one or more pure tones,
a tone burst, etc). Such otoacoustic emissions may also differ between individuals and
can therefore be used to discriminate between individuals or to identify a particular
individual.

One problem faced by biometric algorithms is the need to achieve acceptable performance in two respects. First, the algorithm should provide acceptable security so that unauthorised users are not falsely recognized as authorised users. The likelihood that the algorithm will accept an access attempt by an unauthorised user is known as the false acceptance rate (FAR), and should be kept low if the algorithm is to provide reasonable security. Second, the algorithm should work reliably, so that authorised users are not falsely rejected as unauthorised. The likelihood that the algorithm will reject an access attempt by an authorised user is known as the false rejection rate (FRR), and should also be kept low if the algorithm is not to prove frustrating for authorised users seeking authentication.

The problem is that these two performance requirements conflict with each other. A low FRR can be achieved by relaxing the requirements for a user to achieve authentication. However, this will also have the consequence of increasing the FAR. Conversely, a low FAR can be achieved by making the requirements for a user to achieve authentication stricter. However, this will have the consequence of increasing the FRR.

One way to decrease both FAR and FRR is to increase the efficacy of the biometric algorithm itself. However, designing the algorithm to achieve high performance is difficult. Further, the efficacy may depend on factors which are outside the designers' control. For example, the efficacy of the algorithm may depend on the quality of the biometric data (e.g., noise levels, etc), or the discriminatory nature of the biometric itself.

One known technique for improving the efficacy of the authentication is to combine (or "fuse") multiple authentication processes together, whether biometric or not. For example, a user may be required to input a security password or pin number in addition to providing a voice input (e.g., for voice biometric authentication). The authentication processes may be combined using score-level fusion, or result-level fusion. In the former case, separate scores are obtained for each authentication process indicating the likelihood that a user is an authorised user. Those scores are then combined and compared to a threshold to determine whether the user should be authenticated as the authorised user. In the latter case, separate results are obtained for each authentication process (e.g., through separate comparison of each score to a corresponding threshold) and then combined. For example, a user may be

authenticated only if the individual authentication results are all positive. By fusing multiple authentication processes in this way, the FAR and FRR values for the combined authentication process can be much lower than the FAR and FRR values associated with an individual authentication process.

5

However, it can be time consuming and inconvenient to provide multiple inputs as part of an authentication process. Users typically want authentication to be both reliable and secure, while taking as little time as possible. Preferably, the authentication process should even be invisible to the user, such that they are authenticated without even realising that an authentication process has taken place.

10

Summary

Embodiments of the present disclosure seek to address these and other problems.

15 For example, in one aspect the present disclosure provides a method for authenticating a user of an electronic device. The method comprises: responsive to detection of a trigger event indicative of a user interaction with the electronic device, generating an audio probe signal to play through an audio transducer of the electronic device; receiving a first audio signal comprising a response of the user's ear to the audio probe
20 signal; receiving a second audio signal comprising speech of the user; and applying an ear biometric algorithm to the first audio signal and a voice biometric algorithm to the second audio signal to authenticate the user as an authorised user.

In one embodiment, the trigger event comprises a predetermined keyword or
25 keyphrase spoken by the user. In such an embodiment, the method may further comprise additionally applying the voice biometric algorithm to the predetermined keyword or keyphrase (which may also be comprised within the second audio signal) to authenticate the user as an authorised user.

30 Another aspect of the disclosure provides an authentication device for authenticating a user of an electronic device. The authentication device comprises: an audio signal generation module for generating, responsive to detection of a trigger event indicative of a user interaction with the electronic device, an audio probe signal to play through an audio transducer of the electronic device; one or more inputs for receiving a first audio
35 signal comprising a response of the user's ear to the audio probe signal, and a second audio signal comprising speech of the user; and a biometric authentication module for

utilizing an ear biometric algorithm on the first audio signal and a voice biometric algorithm on the second audio signal to authenticate the user as an authorised user.

5 In one embodiment, the trigger event comprises a predetermined keyword or keyphrase spoken by the user. In such an embodiment, the biometric authentication module is further for utilizing the voice biometric algorithm on the predetermined keyword or keyphrase to authenticate the user as an authorised user.

10 A further aspect of the disclosure provides an electronic device comprising an authentication device as set out above. The electronic device may be portable and/or battery powered, such as a smartphone, an audio player, a mobile or cellular phone, or a handset.

Brief description of the drawings

15 For a better understanding of examples of the present disclosure, and to show more clearly how the examples may be carried into effect, reference will now be made, by way of example only, to the following drawings in which:

20 Figure 1a shows a personal audio device according to one configuration;

Figure 1b shows a personal audio device according to another configuration;

Figure 2 is a flowchart of a method according to embodiments of the disclosure;

25 Figure 3 is a flowchart of a method according to further embodiments of the disclosure; and

Figure 4 is a schematic diagram of a biometric authentication system according to embodiments of the disclosure.

30

Detailed description

As noted above, ear biometric data may be acquired by the generation of an acoustic stimulus (which may be audible or non-audible, e.g., ultrasonic), and the detection of an acoustic response of the ear to the acoustic stimulus. One or more features may be
35 extracted from the response signal, and used to characterize the individual. The acoustic response of the ear may relate to the acoustic transfer function of the ear,

and/or otoacoustic emissions from the ear. Alternatively or additionally, otoacoustic emissions may be spontaneous, and detected without any requirement for an acoustic stimulus.

- 5 Ear biometric data may be acquired using a personal audio device. As used herein, the term “personal audio device” is any electronic device which is suitable for, or configurable to, provide audio playback substantially to only a single user. Some examples of suitable personal audio devices are shown in Figures 1a and 1b.
- 10 **Figure 1a** shows the application of an electronic device arrangement 100 to a user’s ear, with the arrangement 100 comprising a personal audio device 150 (in the illustrated embodiment, an in-ear headphone or earphone, insert headphone, or ear bud) and a host device 160 (in the illustrated embodiment, a smartphone). In such an embodiment, the functionality discussed below may be provided solely in the personal
- 15 audio device 150, solely in the host device 160, or split between both devices (i.e. with each device performing different aspects). **Figure 1b** shows the application of a personal audio device 170 (in the illustrated embodiment, a smartphone) to a user’s ear. In this example, therefore, there is no separate host device and the functionality discussed below is carried out in the personal audio device itself.
- 20
- Figure 1a shows a schematic diagram of a user’s ear, comprising the (external) pinna or auricle 112a, and the (internal) ear canal 112b. An electronic device arrangement 100 comprises a personal audio device 150 and a host device 160.
- 25 As noted above, in the illustrated embodiment the personal audio device 150 comprises an in-ear headphone (or earphone), insert headphone, or ear bud. This headphone is configured to be partially or totally inserted within the ear canal 112b, and may provide a relatively tight seal between the ear canal 112b and the external environment (i.e. it may be acoustically closed or sealed). Alternative personal audio
- 30 devices include circum-aural headphones (e.g., those headphones comprising a shell which substantially surrounds and encloses the auricle), supra-aural headphones (e.g., those headphones which do not surround or enclose the user’s ear, but rather sit on the auricle 112a), intra-concha headphones or earphones (e.g., those headphones which sit inside the user’s concha cavity), and smartphones or mobile phones etc (see
- 35 Figure 1b).

The personal audio device 150 comprises one or more loudspeakers 152 and one or more microphones 154. The one or more loudspeakers 152 are positioned on an internal surface of the headphone, and arranged to generate acoustic signals (e.g., audible or inaudible signals, such as ultrasonic waves) towards the user's ear and particularly the ear canal 112b. The one or more microphones 154 (hereinafter, "ear microphones") are also positioned on the internal surface of the headphone, and arranged to detect acoustic signals within the internal volume defined by the headphone (in the illustrated embodiment, the ear canal 112b).

10 The headphone may be able to perform active noise cancellation, to reduce the amount of noise experienced by the user of the headphone. Active noise cancellation operates by detecting a noise (i.e. with a microphone), and generating a signal (i.e. with a loudspeaker) that has the same amplitude as the noise signal but is opposite in phase. The generated signal thus interferes destructively with the noise and so
15 lessens the noise experienced by the user. Active noise cancellation may operate on the basis of feedback signals, feedforward signals, or a combination of both. Feedforward active noise cancellation utilizes one or more microphones on an external surface of the headphone, operative to detect the environmental noise before it reaches the user's ear. The detected noise is processed quickly, and the cancellation
20 signal generated so as to match the incoming noise as it arrives at the user's ear. Feedback active noise cancellation utilizes one or more error microphones positioned on the internal surface of the headphone, operative to detect the combination of the noise and the audio playback signal generated by the one or more loudspeakers. This combination is used in a feedback loop, together with knowledge of the audio playback
25 signal, to adjust the cancelling signal generated by the loudspeaker and so reduce the noise. The ear microphone 154 shown in Figure 1a may therefore form part of an active noise cancellation system, for example, as an error microphone.

The personal audio device 150 may comprise a further microphone 156 positioned
30 such that, in use, the microphone 156 lies close to the user's mouth and therefore primarily detects the user's voice. The microphone 156 may be termed the voice microphone herein. In the illustrated embodiment, the voice microphone 156 is provided within the wired connection to the host device 160. In other embodiments, the voice microphone 156 may be provided on the end of an arm as part of a headset, for
35 example.

The personal audio device 150 and the host device 160 may be operatively coupled together, such that signals and/or data can be passed between them. For example, signals detected by the microphones 154, 156 may be passed from the personal audio device 150 to the host device 160; audio signals to be played back to the user may be passed from the host device 160 to the personal audio device. In some embodiments, biometric authentication results may be passed between the personal audio device 150 and the host device 160.

In the illustrated embodiment, the personal audio device 150 is coupled to the host device 160 via a wired connection. Those skilled in the art will appreciate of course that any operative connection between the two devices may be provided, such as a wireless connection (e.g., Bluetooth (RTM)).

Figure 1b shows an alternative personal audio device 170, which is a mobile or cellular phone or handset. The handset 170 comprises one or more loudspeakers 172 for audio playback to the user, one or more ear microphones 174 which are similarly positioned to detect acoustic signals in the vicinity of the user's ear, and one or more voice microphones 176 to detect the user's voice.

In use, the handset 170 is held close to the user's ear so as to provide audio playback (e.g. during a call). While a tight acoustic seal is not achieved between the handset 170 and the user's ear, the handset 170 is typically held close enough that an acoustic stimulus applied to the ear via the one or more loudspeakers 172 generates a response from the ear which can be detected by the one or more ear microphones 174. As with the other devices, the loudspeaker(s) 172 and ear microphone(s) 174 may form part of an active noise cancellation system.

Figure 2 is a flowchart of a method for authenticating a user as an authorised user of an electronic device. The method may utilize any of the personal audio devices discussed above with respect to Figures 1a and 1b, for example, and may be performed by an authentication device which is arranged exclusively within the personal audio device, exclusively within a host device coupled to that personal audio device, or which is distributed between the host device and the personal audio device.

In step 200, the authentication device monitors for an event which triggers the authentication process (hereinafter, "trigger event"). Various events are contemplated as trigger events according to embodiments of the disclosure.

- 5 In one embodiment, the trigger event is indicative of a user interaction with the authentication device. The trigger event may alternatively or additionally be indicative of a user's desire to interact with the personal audio device using speech. For example, the trigger event may be associated with activation of a virtual assistant, e.g., Siri, Cortana, Alexa (all RTM), etc.

10

- The trigger event may comprise actuation of a physical or virtual button on the personal audio device. Personal audio devices often have such buttons in order to send control signals to the personal audio device and/or a host device. Different control signals may be generated based on the type of button actuation (e.g., a short or long press) or on a sequence of multiple button actuations (e.g., a chain of two or three button presses in sequence). The button may physically move in response to the user interaction (e.g., a physical button), or comprise a capacitive or ultrasonic sensor (e.g., a virtual button). In another example, a personal audio device may detect finger taps or other user interactions with the device via a microphone (such as the ear microphones 154, 174 and/or the voice microphones 156, 176). Pulses of audio detected by the microphone may be determined as being indicative of a tap of the user's fingers on the microphone. In this way, the microphone itself functions as a button. Trigger events in which a button is actuated (or its functional equivalent) consume small amounts of power.

- 25 In another example, the user interaction may be a spoken interaction. For example, the user may speak a keyword, i.e. a predetermined word or phrase ("trigger phrase"), such as "Hey Siri", "OK Google", etc, which is detected by a trigger phrase detect module in the authentication device.

- 30 In another embodiment, the trigger event comprises an in-ear (or on-ear) detection event. For example, the trigger event may comprise a determination that the personal audio device (i.e. at least a loudspeaker and a microphone forming part of the personal audio device) is being placed in or on a user's ear. In another example, the trigger event may comprise a determination that the personal audio device is being actively worn by the user (e.g. an earbud or headset is positioned on or in a user's ear). Various mechanisms are suitable for such a purpose. For example, a microphone in
- 35

the personal audio device may detect an audio signal (e.g. a response signal to another audio signal transmitted by the personal audio device) and analyse the audio signal to determine whether it is characteristic of a small volume (i.e. the volume inside the ear canal and/or within a headphone shell). Alternatively or additionally, an optical
5 mechanism in the personal audio device may detect the presence of light and so conclude that the personal audio device is not applied to the user's head. Alternatively or additionally, an accelerometer in the personal audio device may detect movement which is inconsistent with the motion of a user's head. Alternatively or additionally, an accelerometer in the personal audio device may detect movement which is consistent
10 with part of the personal audio device (e.g. an earbud or headset) being placed on or in a user's ear.

In step 202, it is determined whether a trigger event has been detected. If no trigger event has been detected, the method returns to step 200 and the authentication device
15 continues to monitor for a trigger event. If a trigger event is detected, the authentication process is initiated and the method proceeds to step 204.

In step 204, the authentication device generates an audio probe signal to be played through an audio transducer of the personal audio device in the vicinity of the user's
20 ear. For example, the audio probe signal may be played back through the loudspeakers 152, 172.

The audio probe signal may be audible or inaudible. In embodiments where the audio probe signal is inaudible, the signal may be ultrasonic (e.g., greater than 20kHz or in
25 the range 20kHz to 50kHz). In embodiments where the audio probe signal is audible, and where the trigger event is associated with a user's desire to interact with the authentication device via speech, the audio probe signal may comprise a prompt to the user to speak. The prompt may be explicit (e.g., speech from a virtual assistant) or implicit (e.g., a notification tone associated with speech input).

30

In embodiments in which the audio probe signal is audible and the trigger event comprises an in-ear (or on-ear) detection event, then the audio probe signal may be played at the same time or during playback of an in-ear (or on-ear) audio signal which is emitted by the personal audio device on application of the personal audio device to a
35 user (e.g. when an earbud or headset is placed in or on an ear). The in-ear (or on-ear)

audio signal may be louder than the audio probe signal (e.g. to mask the sound of the probe signal from the user).

The audio probe signal may comprise any suitable signal which causes an acoustic response of the user's ear. For example, the audio probe signal may comprise a chirp of signals at one or more frequencies, white noise, a sound tone, music, speech, etc. In one embodiment, the audio probe signal may be selected or modulated to contain power at frequencies which stimulate frequencies in the acoustic response of the user's ear which are discriminative between different users. Frequencies which may be discriminative between different users include frequency bands between approximately 1 kHz and approximately 3 kHz; between approximately 5 kHz and approximately 7 kHz; and between approximately 10 kHz and approximately 12 kHz. For example, the speech output of a virtual assistant may be modified to contain power at frequencies which stimulate an acoustic response at those discriminative frequencies. Particular music samples may be chosen, or user-selected music adapted, to contain power at the appropriate frequencies.

In step 206, the authentication device receives a first audio signal comprising a response of the user's ear to play back of the audio probe signal. As noted above, the response may comprise the audio probe signal as modified by a transfer function associated with the user's ear. Alternatively or additionally, the response may comprise an otoacoustic emission from the user's ear. The first audio signal may be generated by a microphone in the vicinity of the user's ear, such as the ear microphones 154, 174.

In step 208, the authentication device further receives a second audio signal comprising speech of the user. The speech may comprise one or more commands or command phrases, e.g., instructions to be carried out by the personal audio device and/or the host device. The second audio signal may be generated by a voice microphone, such as the microphones 156, 176.

In embodiments where the audio probe signal comprises a prompt to the user to speak, the second audio signal may be received in response to that prompt.

In order to conserve power, the authentication device may monitor for the speech only in a limited time window. The time window may begin upon the audio probe signal

ending, for example, and continue for a predetermined period of time. Alternatively, the time window may overlap with transmission of the audio probe signal. For example, if the probe signal is inaudible, the user will not be aware of it and consequently may provide speech (e.g., command speech) before or during the audio probe signal. Even
5 if the probe signal is audible, a user may not wait for the audio probe signal before speaking. Consequently, the time window may be defined so as to overlap with transmission of the audio probe signal (e.g., starting before or during transmission of the audio probe signal) to increase the likelihood that all of the speech is acquired.

10 Alternatively or additionally, the time window may be defined based on speech detected in the first audio signal. In this regard, it is noted that speech detected in the first audio signal is highly likely to correspond to speech from the user rather than another person, owing to the location of the ear microphones close to, or inside the user's ear. That is, speech from the user detected by the ear microphone may have
15 been conducted primarily through the user's skull and/or jaw bone and have a relatively high amplitude, while speech from others is likely to be attenuated (e.g., owing to a relatively tight acoustic coupling around the ear canal 112b).

The time window may therefore be defined based on those times at which the first
20 audio signal is substantially correlated with the second audio signal. Alternatively or additionally, the time window may be defined based on those times at which speech input is detected in the first audio signal.

In step 210, the authentication device performs an ear biometric algorithm on the first
25 audio signal, and a voice biometric algorithm on the second audio signal, to authenticate the user as an authorised user.

Each biometric algorithm may comprise the extraction of one or more features from the audio signal, and the comparison of those extracted features to corresponding features
30 in a stored template for an authorised user. For example, a user will typically undergo a process known as enrolment, in which biometric data is obtained in a secure environment (e.g., once the user has already been authenticated via an alternative mechanism, or is deemed authenticated), and features are extracted from that biometric data and stored in a template for that user. The user may be required to
35 provide multiple biometric inputs to increase the reliability of the features extracted for the purposes of the template. For example, several samples of ear biometric data may

be acquired, or the user may be required to provide multiple spoken inputs as voice biometric data.

In step 210, the relevant features are extracted from the first and second audio signals, and compared to the features in those stored templates. Thus, one or more features are extracted from the first audio signal, and compared to corresponding features of the ear biometric template for the authorised user. Similarly, one or more features are extracted from the second audio signal, and compared to corresponding features of the voice biometric template for the authorised user.

One or more biometric scores may be generated, indicative of the likelihood that the user (i.e., the user providing the biometric input) is the authorised user. For example, the score may be based on the mathematical distance (e.g., the cosine vector similarity or any other suitable comparison algorithm) between the obtained features and the stored features. The one or more biometric scores may then be compared to one or more threshold values in order to decide whether the user should be authenticated as the authorised user or not.

The outputs of the voice and ear biometric algorithms may be combined or fused in order to authenticate the user. Various methods of biometric fusion are known in the art, and the present disclosure is not limited in that respect. For example, separate biometric scores may be generated for the ear biometric algorithm and the voice biometric algorithm, and these biometric scores compared to separate threshold values in order to generate separate biometric results. These results may then be fused to generate an overall authentication decision or result (result-level fusion). For example, the user may be authenticated as the authorised user only if both ear and voice biometric results are positive. Alternatively, the separate biometric scores may be combined to generate an overall score, with this overall score then being compared to a threshold value (score-level fusion).

In embodiments where the trigger event comprises a trigger phrase (i.e. a keyword or keyphrase) uttered by the user (e.g., a predetermined phrase or word, such as "Hey Siri", "OK Google", etc), a voice biometric algorithm may additionally be performed on features extracted from that trigger phrase, and the overall authentication decision or result based additionally on that voice biometric algorithm.

For example, the voice biometric algorithm may be performed on the predetermined phrase to obtain a first voice biometric score, and also on the command speech (e.g. from the second audio signal) to obtain a second voice biometric score. The first and second voice biometric scores may be fused with the ear biometric score to obtain an overall biometric score which can be compared to one or more thresholds to obtain an overall biometric result. Alternatively, as noted above, separate biometric results may be obtained for each biometric process (e.g., based on the first and second voice biometric scores and the ear biometric score) and the results fused to obtain an overall biometric result.

In step 212, the user is authenticated as the authorised user, or denied authentication, based on the ear and voice biometric algorithms performed in step 210.

Figure 3 is a flowchart of a method according to further embodiments of the disclosure.

Again, the method may be performed by an authentication device which is arranged exclusively within a personal audio device, exclusively within a host device coupled to that personal audio device, or which is distributed between the host device and the personal audio device.

In step 300, the authentication device monitors for an event which triggers the authentication process (hereinafter, "trigger event"). Various events are contemplated as trigger events according to embodiments of the disclosure.

In one embodiment, the trigger event is indicative of a physical interaction with the authentication device. The trigger event may alternatively or additionally be indicative of a user's desire to interact with the personal audio device using speech. For example, the trigger event may be associated with activation of a virtual assistant, e.g., Siri, Cortana, Alexa (all RTM), etc.

The trigger event may comprise actuation of a physical or virtual button on the personal audio device. Personal audio devices often have such buttons in order to send control signals to the personal audio device and/or a host device. Different control signals may be generated based on the type of button actuation (e.g., a short or long press) or on a sequence of multiple button actuations (e.g., a chain of two or three button presses in sequence). The button may physically move in response to the user interaction (e.g., a physical button), or comprise a capacitive or ultrasonic sensor (e.g., a virtual button).

In another example, a personal audio device may detect finger taps or other user interactions with the device via a microphone (such as the ear microphones 154, 174 and/or the voice microphones 156, 176). Pulses of audio detected by the microphone may be determined as being indicative of a tap of the user's fingers on the microphone.

- 5 In this way, the microphone itself functions as a button. Trigger events in which a button is actuated (or its functional equivalent) consume small amounts of power.

In step 302, it is determined whether a trigger event has been detected. If no trigger event has been detected, the method returns to step 300 and the authentication device
10 continues to monitor for a trigger event. If a trigger event is detected, the authentication process is initiated and the method proceeds to step 304.

In step 304, the authentication device generates an audio probe signal to be played through an audio transducer of the personal audio device in the vicinity of the user's
15 ear. For example, the audio probe signal may be played back through the loudspeakers 152, 172.

The audio probe signal is audible, and may comprise a prompt to the user to speak. The prompt may be explicit (e.g., speech from a virtual assistant) or implicit (e.g., a
20 notification tone associated with speech input).

The audio probe signal may comprise any suitable signal which causes an acoustic response of the user's ear. For example, the audio probe signal may comprise a chirp of signals at one or more frequencies, white noise, a sound tone, music, speech, etc.
25 In one embodiment, the audio probe signal may be selected or modulated to contain power at frequencies which stimulate frequencies in the acoustic response of the user's ear which are discriminative between different users. Frequencies which may be discriminative between different users include frequency bands between approximately 1 kHz and approximately 3 kHz; between approximately 5 kHz and approximately
30 7 kHz; and between approximately 10 kHz and approximately 12 kHz. For example, the speech output of a virtual assistant may be modified to contain power at frequencies which stimulate an acoustic response at those discriminative frequencies. Particular music samples may be chosen, or user-selected music adapted, to contain power at the appropriate frequencies.

In step 306, the authentication device receives a first audio signal comprising a response of the user's ear to play back of the audio probe signal. As noted above, the response may comprise the audio probe signal as modified by a transfer function associated with the user's ear. Alternatively or additionally, the response may
5 comprise an otoacoustic emission from the user's ear. The first audio signal may be generated by a microphone in the vicinity of the user's ear, such as the ear microphones 154, 174.

In step 308, the authentication device performs an ear biometric algorithm on the first
10 audio signal. Thus, one or more features are extracted from the first audio signal, and compared to corresponding features of an ear biometric template for an authorised user (or the templates for multiple authorised users if they exist). An ear biometric score is generated, indicative of the likelihood that the user (i.e., the user providing the ear biometric input) is an authorised user, based on the ear biometric algorithm. For
15 example, the score may be based on the mathematical distance (e.g., the cosine vector similarity or any other suitable comparison algorithm) between the obtained features and the stored features.

In step 310, and in parallel to steps 306 and 308 in the illustrated embodiment, the
20 authentication device further receives a second audio signal comprising speech of the user in response to the prompt contained in the audio probe signal. The speech may comprise one or more commands or command phrases, e.g., instructions to be carried out by the personal audio device and/or the host device. The second audio signal may be generated by a voice microphone, such as the microphones 156, 176.

25

In order to conserve power, the authentication device may monitor for the speech only in a limited time window. The time window may begin upon the audio probe signal ending, for example, and continue for a predetermined period of time.

30 Alternatively or additionally, the time window may be defined based on speech detected in the first audio signal. In this regard, it is noted that speech detected in the first audio signal is highly likely to correspond to speech from the user rather than another person, owing to the location of the ear microphones close to, or inside the user's ear. That is, speech from the user detected by the ear microphone may have
35 been conducted primarily through the user's skull and/or jaw bone and have a relatively high amplitude, while speech from others is likely to be attenuated. The time window

may therefore be defined based on those times at which the first audio signal is substantially correlated with the second audio signal. Alternatively or additionally, the time window may be defined based on those times at which speech input is detected in the first audio signal.

5

In step 312, the authentication device performs a voice biometric algorithm on the second audio signal. Thus, one or more features are extracted from the second audio signal, and compared to corresponding features of the voice biometric template for the authorised user (or multiple voice biometric templates for multiple authorised users if they exist).

10

In step 314, the outputs of the outputs of the voice and ear biometric algorithms are combined or fused in order to authenticate the user. In the illustrated embodiment, result-level fusion is utilized to fuse the results of the ear and voice biometric algorithms, i.e., the ear and voice biometric scores are compared to separate thresholds to generate corresponding results, and the results are combined. Alternatively, the separate biometric scores may be combined to generate an overall score, with this overall score then being compared to a threshold value (score-level fusion).

15

20

In step 316, it is determined whether the user was authenticated as a result of the combined voice and ear biometric algorithms. If the authentication is positive, the method proceeds to step 318, in which the command spoken by the user in the second audio signal is executed. If the authentication is negative, the method proceeds to step 320 in which the command spoken by the user is denied or not executed. In this regard, speech recognition may be performed on the second audio signal (e.g., the gated parts of the second audio signal) in order to determine the semantic meaning of the command speech. Speech recognition may be performed locally (e.g., in the host device or the personal audio device if they have sufficient processing capability), or remotely in a speech recognition service. In the latter case, the second audio signal may be transmitted to the speech recognition service to be processed.

25

30

Figure 4 shows a schematic diagram of an authentication system 400 according to embodiments of the disclosure. The authentication system 400 comprises an authentication device 402 or processing circuitry (which may be implemented on one or more integrated circuits), a memory 404, and one or more input and output devices

35

described below. As noted above, the authentication device 402 may be provided exclusively within a personal audio device, exclusively within a host device coupled to the personal audio device, or distributed between the personal audio device and the host device. The memory 404 may similarly be provided within the personal audio device, within the host device, or distributed between the personal audio device and the host device.

The system 400 comprises a loudspeaker 406, which may be a microspeaker. The loudspeaker 406 may be arranged within the personal audio device, such that it is located next to a user's ear during use. For example, the loudspeaker 406 may correspond to either of the loudspeakers 152, 172 described above with respect to Figures 1a and 1b.

The system 400 further comprises an ear microphone 408. The ear microphone 408 may be arranged within the personal audio device, such that it is located next to a user's ear during use. For example, the ear microphone 408 may correspond to either of the microphones 154, 174 described above with respect to Figures 1a and 1b.

The system 400 further comprises a voice microphone 410. The voice microphone 410 may be arranged within the personal audio device, such that it is located next to a user's mouth during use. For example, the voice microphone 410 may correspond to either of the microphones 156, 176 described above with respect to Figures 1a and 1b.

The device 402 comprises a control module 412 which is operative to receive or detect a trigger event, e.g., as described above with respect to steps 200 and/or 300. Responsive to detection of the trigger event, the control module 412 generates an audio probe signal to be played back to the user via the loudspeaker 406.

The audio probe signal may be audible or inaudible. In embodiments where the audio probe signal is inaudible, the signal may be ultrasonic (e.g., greater than 20kHz or in the range 20kHz to 50kHz). In embodiments where the audio probe signal is audible, and where the trigger event is associated with a user's desire to interact with the authentication device via speech, the audio probe signal may comprise a prompt to the user to speak. The prompt may be explicit (e.g., speech from a virtual assistant) or implicit (e.g., a notification tone associated with speech input).

The audio probe signal may comprise any suitable signal which causes an acoustic response of the user's ear. For example, the audio probe signal may comprise a chirp of signals at one or more frequencies, white noise, a sound tone, music, speech, etc. In one embodiment, the audio probe signal may be selected or modulated to contain
5 power at frequencies which stimulate frequencies in the acoustic response of the user's ear which are discriminative between different users. Frequencies which may be discriminative between different users include frequency bands between approximately 1 kHz and approximately 3 kHz; between approximately 5 kHz and approximately 7 kHz; and between approximately 10 kHz and approximately 12 kHz. For example,
10 the speech output of a virtual assistant may be modified to contain power at frequencies which stimulate an acoustic response at those discriminative frequencies. Particular music samples may be chosen, or user-selected music adapted, to contain power at the appropriate frequencies.

15 The ear microphone 408 detects the acoustic response of the user's ear, and generates a first audio signal. The first audio signal is detected by the microphone 408 in the time domain. However, the features extracted for the purposes of the biometric process may be in the frequency domain (in that it is the frequency response of the user's ear which is characteristic). The device 402 therefore comprises a Fourier
20 transform module 414, which converts the first audio signal to the frequency domain. For example, the Fourier transform module 414 may implement a fast Fourier transform (FFT). In some examples the biometric process may not be in the frequency domain, so the Fourier transform module may be omitted.

25 The output of the Fourier transform module 414 is provided to a biometric authentication module 416, which is configured to perform an ear biometric algorithm on the signal, e.g., as described above with respect to steps 210 and 308. Thus the biometric authentication module 416 extracts one or more features from the first audio signal and compares those features to a stored ear template 420 for an authorised user
30 (e.g., an "ear print"), for example as stored in the memory 404.

In some embodiments the biometric authentication module 416 may be designed to extract features with foreknowledge of the nature of the audio probe signal, for example knowing the spectrum of the audio probe signal, so that the response or transfer
35 function may be appropriately normalised. In other embodiments the Fourier transform module 414 may comprise a second input to monitor the audio probe signal (e.g.

playback music) and hence provide the biometric authentication module 416 with information about the audio probe signal or its spectrum so that the biometric authentication module may calculate the transfer function from the audio probe signal to the received acoustic waveform, from which it may derive the desired feature parameters.

The voice microphone 410 generates a second audio signal comprising speech of the user (e.g., in response to a prompt in the audio probe signal). The second audio signal is also provided to the biometric authentication module 416, which is configured to perform a voice biometric algorithm on the signal, e.g., as described above with respect to steps 210 and 312. Thus the biometric authentication module 416 extracts one or more features from the second audio signal and compares those features to a stored voice template 422 for an authorised user (e.g., a “voice print”), for example as stored in the memory 404. As noted above, the biometric authentication module 416 may additionally perform a voice biometric algorithm on a trigger phrase or word uttered by the user (e.g., and detected as a trigger event to commence authentication).

As noted above, in some embodiments the second audio signal is gated, and thus in the illustrated embodiment the authentication device 402 further comprises a gating module 418 which is operative to receive the second audio signal from the voice microphone 410, and to provide a gated signal to the biometric authentication module 416. The gating may be such that the biometric algorithm is performed only on the second audio signal in a limited time window. The time window may begin upon the audio probe signal ending, for example, and continue for a predetermined period of time. In such embodiments, the control module 412 may provide a gating signal to the gating module 418 indicating when the audio probe signal has ended.

Alternatively or additionally, the time window may be defined based on speech detected in the first audio signal. Thus, in the illustrated embodiment, the first audio signal is also provided to the gating module 418. The time window may therefore be defined based on those times at which the first audio signal is substantially correlated with the second audio signal. Alternatively or additionally, the time window may be defined based on those times at which speech input is detected in the first audio signal, e.g., using a voice activity detector (not illustrated).

The biometric authentication module 416 is operative to authenticate the user as an authorised user based on the ear biometric algorithm and the voice biometric algorithm. An overall biometric authentication result may be output from the biometric authentication module 416 to, for example, the control module 412 for further
5 processing (e.g., performing or denying a requested task).

Thus embodiments of the present disclosure provide a convenient mechanism for authenticating a user as an authorised user based on ear and voice biometrics.

10 Embodiments described above have focussed on an implementation in which ear biometrics are performed on signals detected in a single ear. It will be appreciated by those skilled in the art that the embodiments may straightforwardly be adapted to take into consideration biometric data obtained from both ears of a user. Thus, where the description above discloses acquiring data from an ear, data may similarly be acquired
15 from two ears. Biometric algorithms may similarly be performed on data from both ears, and this may be combined as described above, i.e. separate biometric authentication scores combined to form a combined score on which an overall decision is determined, or separate biometric authentication decisions which are then combined to determine an overall decision.

20

Embodiments may be implemented in an electronic, portable and/or battery powered host device such as a smartphone, an audio player, a mobile or cellular phone, or a handset. Embodiments may be implemented on one or more integrated circuits provided within such a host device. Embodiments may be implemented in a personal
25 audio device configurable to provide audio playback to a single person, such as a smartphone, a mobile or cellular phone, headset, headphones, earphones. Again, embodiments may be implemented on one or more integrated circuits provided within such a personal audio device. In yet further alternatives, embodiments may be implemented in a combination of a host device and a personal audio device. For
30 example, embodiments may be implemented in one or more integrated circuits provided within the personal audio device, and one or more integrated circuits provided within the host device.

It should be understood—especially by those having ordinary skill in the art with the
35 benefit of this disclosure—that that the various operations described herein, particularly in connection with the figures, may be implemented by other circuitry or other hardware

components. The order in which each operation of a given method is performed may be changed, and various elements of the systems illustrated herein may be added, reordered, combined, omitted, modified, etc. It is intended that this disclosure embrace all such modifications and changes and, accordingly, the above description should be
5 regarded in an illustrative rather than a restrictive sense.

Similarly, although this disclosure makes reference to specific embodiments, certain modifications and changes can be made to those embodiments without departing from the scope and coverage of this disclosure. Moreover, any benefits, advantages, or
10 solutions to problems that are described herein with regard to specific embodiments are not intended to be construed as a critical, required, or essential feature or element.

Further embodiments likewise, with the benefit of this disclosure, will be apparent to those having ordinary skill in the art, and such embodiments should be deemed as
15 being encompassed herein.

CLAIMS

- 1.A method for authenticating a user of an electronic device, the method comprising:
- 5 responsive to detection of a trigger event indicative of a user interaction with the electronic device, generating an audio probe signal to play through an audio transducer of the electronic device;
- receiving a first audio signal comprising a response of the user's ear to the audio probe signal;
- 10 receiving a second audio signal comprising speech of the user; and
- applying an ear biometric algorithm to the first audio signal and a voice biometric algorithm to the second audio signal to authenticate the user as an authorised user.
- 15 2.The method according to claim 1, wherein the audio probe signal comprises a prompt to the user to speak, and wherein the second audio signal is received in response to the prompt.
- 3.The method according to claim 1, wherein the ear biometric algorithm comprises
- 20 generation of an ear biometric score based on the comparison, and voice biometric algorithm comprises generation of a voice biometric score based on the comparison.
- 4.The method according to claim 3, wherein the ear biometric score and the voice
- 25 biometric score are fused to generate a combined biometric score, and wherein the combined biometric score is compared to one or more threshold values to generate a combined biometric result.
- 5.The method according to claim 3, wherein the ear biometric score is compared to
- 30 a first threshold value to generate an ear biometric result, wherein the voice biometric score is compared to a second threshold value to generate a voice biometric result, and wherein the ear biometric result is combined with voice biometric result to generate a combined biometric result.
- 35 6.An authentication device for authenticating a user of an electronic device, the authentication device comprising:

an audio signal generation module for generating, responsive to detection of a trigger event indicative of a user interaction with the electronic device, an audio probe signal to play through an audio transducer of the electronic device;

5 one or more inputs for receiving a first audio signal comprising a response of the user's ear to the audio probe signal, and a second audio signal comprising speech of the user; and

a biometric authentication module for utilizing an ear biometric algorithm on the first audio signal and a voice biometric algorithm on the second audio signal to authenticate the user as an authorised user.

10

7. The authentication device according to claim 6, wherein the audio probe signal comprises a prompt to the user to speak, and wherein the second audio signal is received in response to the prompt.

15

8. The authentication device according to claim 6, wherein the ear biometric algorithm comprises generation of an ear biometric score based on the comparison, and the voice biometric algorithm comprises generation of a voice biometric score based on the comparison.

20

9. The authentication device according to claim 8, wherein the ear biometric score and the voice biometric score are fused to generate a combined biometric score, and wherein the combined biometric score is compared to one or more threshold values to generate a combined biometric result.

25

10. The authentication device according to claim 8, wherein the ear biometric score is compared to a first threshold value to generate an ear biometric result, wherein the voice biometric score is compared to a second threshold value to generate a voice biometric result, and wherein the ear biometric result is combined with voice biometric result to generate a combined biometric result.

30

11. The authentication device according to claim 8, wherein the ear biometric score is generated based on a comparison between one or more features extracted from the first audio signal and corresponding features of an ear biometric template for the authorised user.

35

12. The authentication device according to claim 8, wherein the voice biometric score is generated based on a comparison between one or more features extracted from the second audio signal and corresponding features of a voice biometric template for the authorised user.
- 5 13. The authentication device according to claim 6, further comprising a gating module configured to monitor the second audio signal for the speech of the user in a time window.
- 10 14. The authentication device according to claim 13, wherein the time window is defined in dependence on a gating signal corresponding to detection of bone-conducted speech in the first audio signal.
- 15 15. The authentication device according to claim 6, wherein the audio probe signal is modulated to stimulate the response of the user's ear at one or more frequencies which are discriminative between different users.
- 20 16. The authentication device according to claim 6, wherein the response of the user's ear comprises one or more of a frequency response of the user's ear canal and an oto-acoustic emission of the user's ear.
17. The authentication device according to claim 6, wherein the trigger event comprises a spoken predetermined keyword or keyphrase by the user.
- 25 18. The authentication device according to claim 6, wherein the trigger event comprises a physical interaction by the user with the electronic device.
19. The authentication device according to claim 6, wherein the trigger event comprises an in-ear or an on-ear detection event.
- 30 20. The authentication device according to claim 6, wherein the speech of the user in response to the prompt comprises a command for execution by the electronic device.
- 35 21. An electronic device comprising the authentication device according to claim 6.

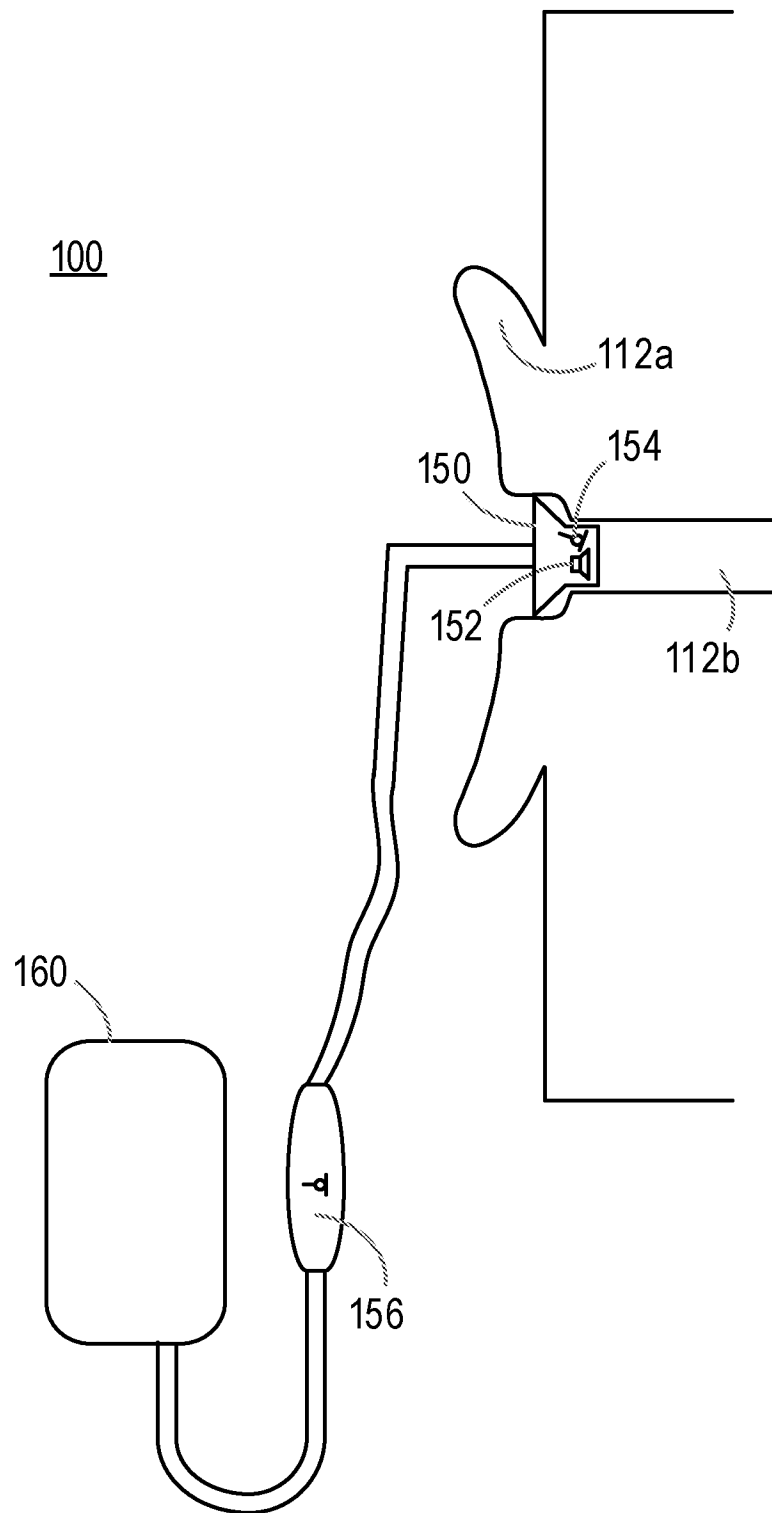


Figure 1a

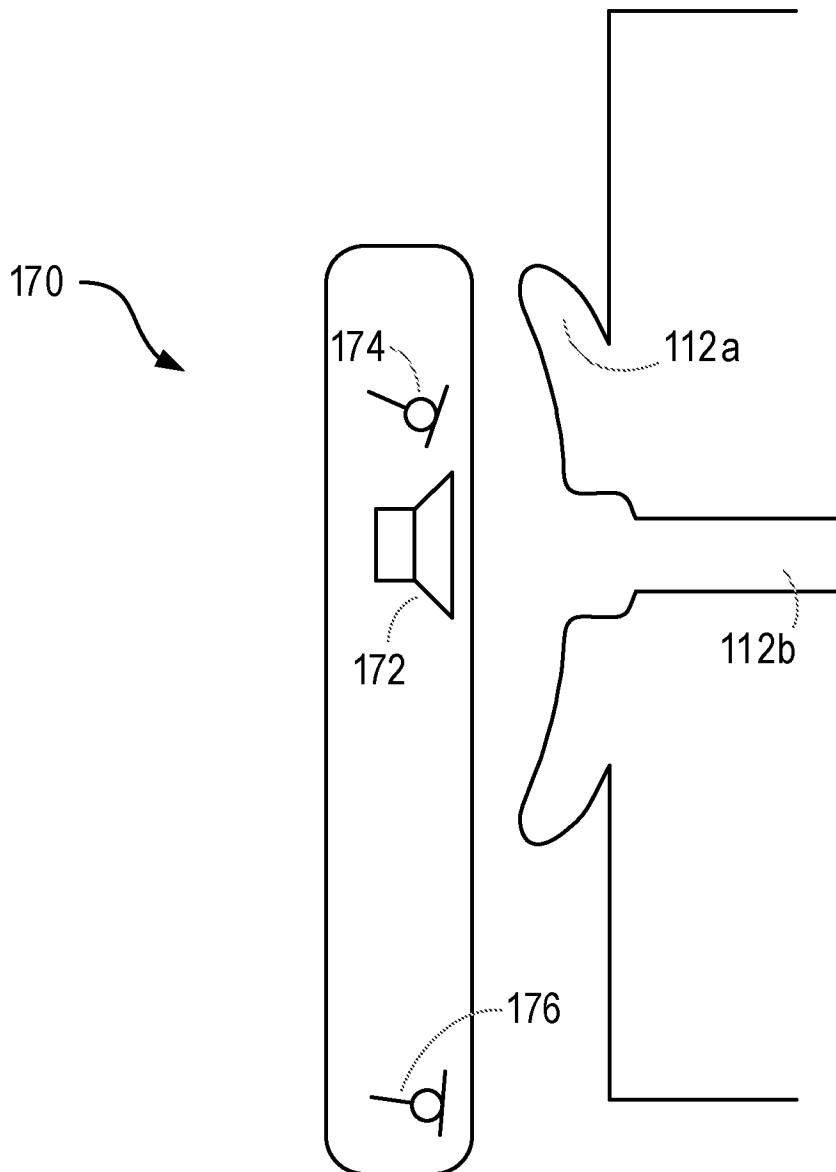


Figure 1b

3/5

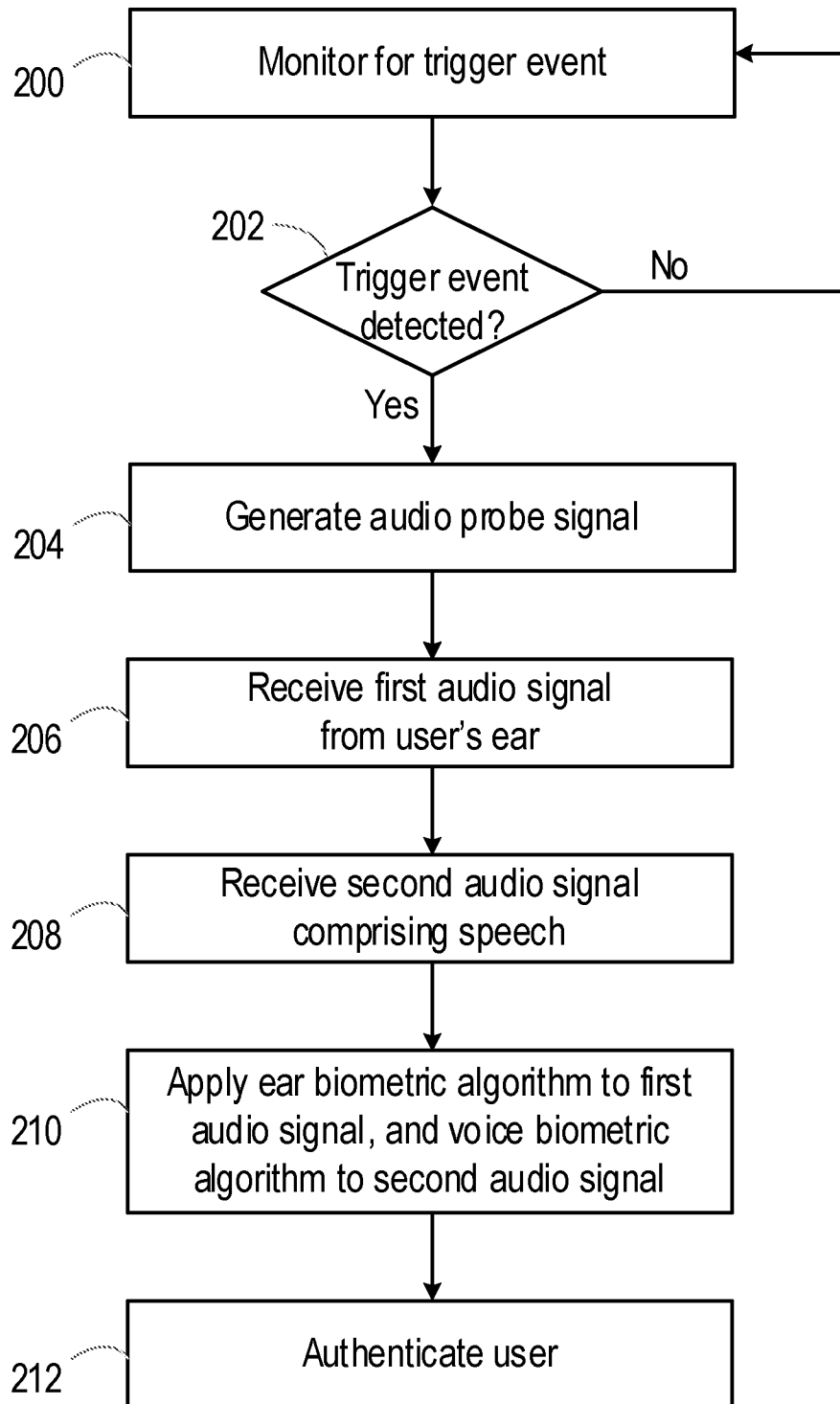


Figure 2

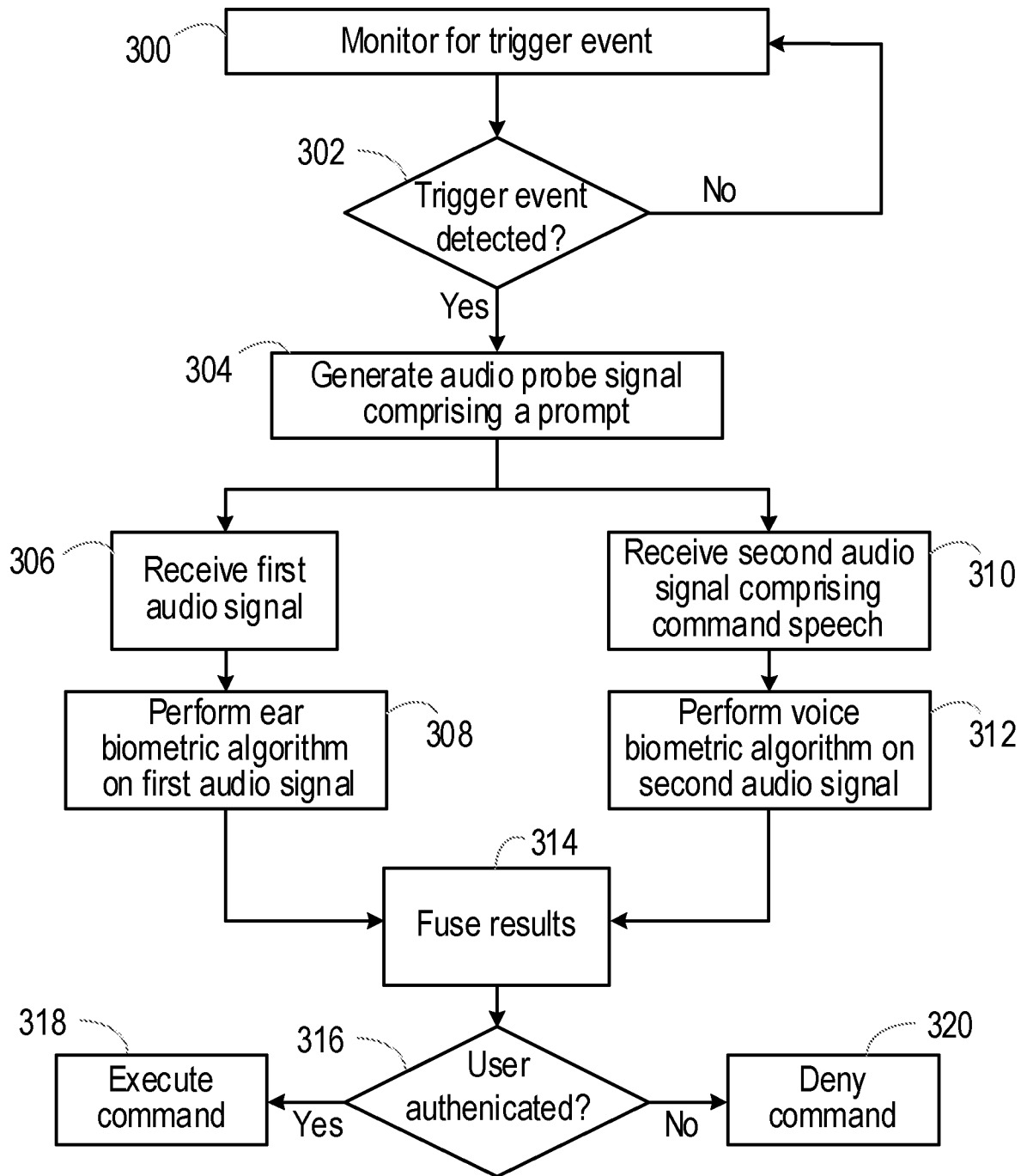


Figure 3

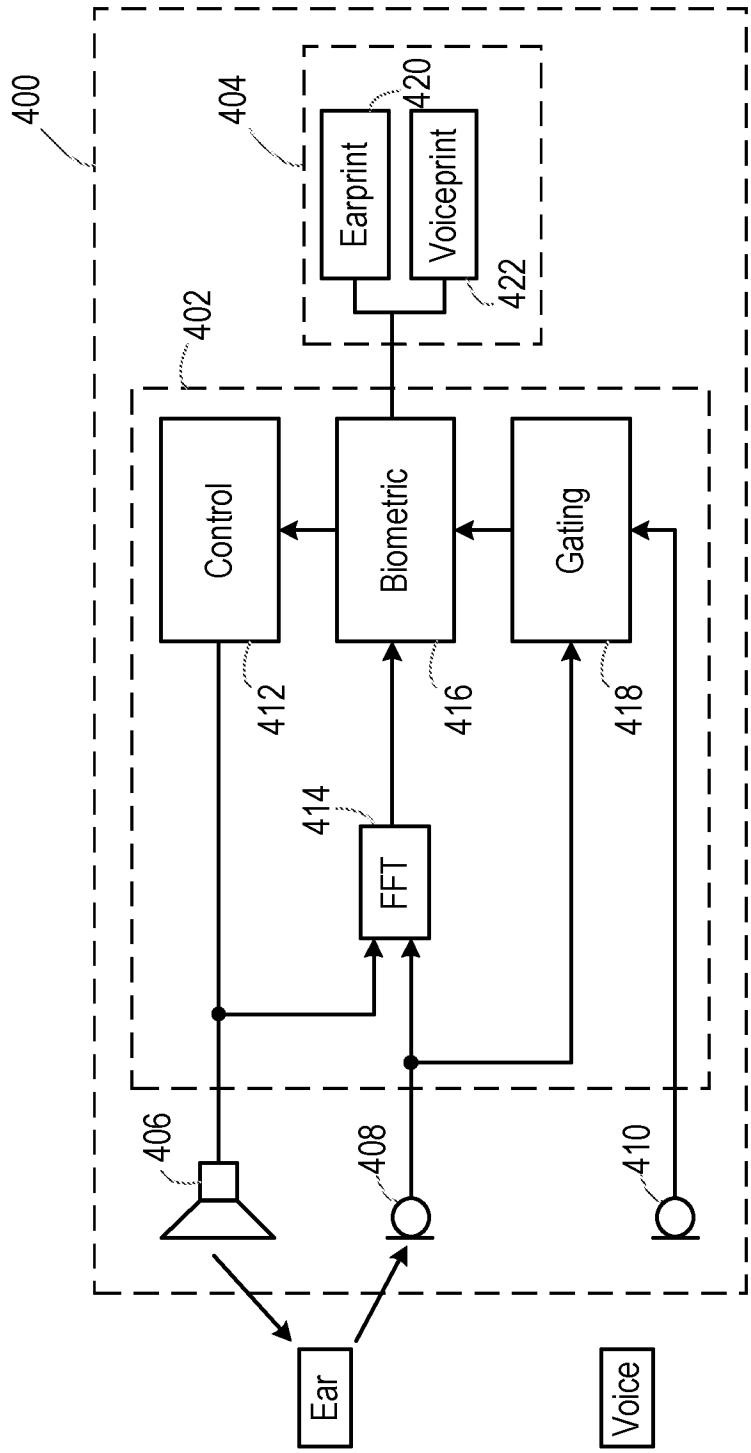


Figure 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2019/052302

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L17/10 G06F21/32 G06F21/40 G06K9/00 A61B5/12
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 G06F G06K A61B

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, COMPENDEX, INSPEC, 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 2012/110341 A1 (BEIGI HOMAYOON [US]) 3 May 2012 (2012-05-03) paragraphs [0003], [0005], [0012], [0056], [0111] - [0115], [0117] - [0118], [0122] - [0123], [0146] - [0147] -----	1-21
X	US 2018/096120 A1 (BOESEN PETER VINCENT [DE]) 5 April 2018 (2018-04-05) figures 1,2 paragraphs [0003], [0020], [0021], [0030] -----	1,6,21
X,P	WO 2019/008389 A1 (CIRRUS LOGIC INT SEMICONDUCTOR LTD [GB]) 10 January 2019 (2019-01-10) figures 4,5 ----- -/-	1,6,21



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

17 September 2019

Date of mailing of the international search report

02/10/2019

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Chétry, Nicolas

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2019/052302

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	WO 2019/008392 A1 (CIRRUS LOGIC INT SEMICONDUCTOR LTD [GB]) 10 January 2019 (2019-01-10) figures 3-5 -----	1,6,21
X,P	WO 2019/008387 A1 (CIRRUS LOGIC INT SEMICONDUCTOR LTD [GB]) 10 January 2019 (2019-01-10) figure 3a -----	1,6,21
X,P	WO 2019/145708 A1 (CIRRUS LOGIC INT SEMICONDUCTOR LTD [GB]) 1 August 2019 (2019-08-01) figure 18 -----	1,6,21

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2019/052302

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012110341	A1	03-05-2012	NONE
US 2018096120	A1	05-04-2018	NONE
WO 2019008389	A1	10-01-2019	US 2019012445 A1 10-01-2019 WO 2019008389 A1 10-01-2019
WO 2019008392	A1	10-01-2019	US 2019012447 A1 10-01-2019 WO 2019008392 A1 10-01-2019
WO 2019008387	A1	10-01-2019	US 2019012448 A1 10-01-2019 WO 2019008387 A1 10-01-2019
WO 2019145708	A1	01-08-2019	NONE