



(51) International Patent Classification:

G10L 17/10 (2013.01) *G07C 9/00* (2006.01)
G06F 21/32 (2013.01) *H04R 29/00* (2006.01)
H04L 9/32 (2006.01) *G10L 21/0216* (2013.01)
H04W 12/00 (2009.01) *G10L 25/78* (2013.01)
G10L 25/03 (2013.01)

(21) International Application Number:

PCT/GB2018/051765

(22) International Filing Date:

25 June 2018 (25.06.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/526,034 28 June 2017 (28.06.2017) US

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

(72) Inventors: **ALONSO, César**; Calle Galileo, 94, 4D, PC. 28015 Madrid (ES). **LESSO, John Paul**; 8 Galachlaw Shot, Edinburgh Lothian EH10 7JF (GB).

(74) Agent: **HASELTINE LAKE LLP**; Redcliff Quay, 120 Redcliff Street, 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: MAGNETIC DETECTION OF REPLAY ATTACK

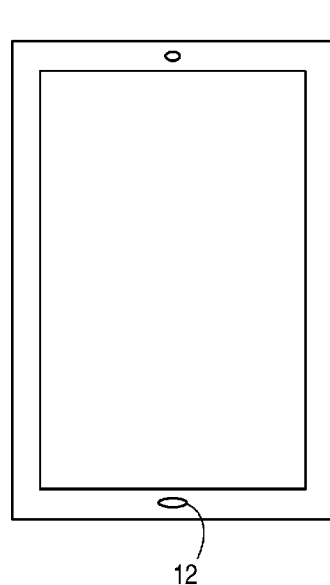


Figure 1

(57) Abstract: A method of detecting a replay attack on a voice biometrics system comprises: receiving an audio signal representing speech; detecting a magnetic field; determining if there is a correlation between the audio signal and the magnetic field; and if there is a correlation between the audio signal and the magnetic field, determining that the audio signal may result from a replay attack.



MAGNETIC DETECTION OF REPLAY ATTACK

Technical Field

- 5 Embodiments described herein relate to methods and devices for detecting a replay attack on a voice biometrics system.

Background

- 10 Voice biometrics systems are becoming widely used. In such a system, a user trains the system by providing samples of their speech during an enrolment phase. In subsequent use, the system is able to discriminate between the enrolled user and non-registered speakers. Voice biometrics systems can in principle be used to control access to a wide range of services and systems.

- 15 One way for a malicious party to attempt to defeat a voice biometrics system is to obtain a recording of the enrolled user's speech, and to play back the recording in an attempt to impersonate the enrolled user and to gain access to services that are intended to be restricted to the enrolled user.

- 20 This is referred to as a replay attack, or as a spoofing attack.

Summary

- 25 According to an aspect of the present invention, there is provided a method of detecting a replay attack on a voice biometrics system. The method comprises: receiving an audio signal representing speech; detecting a magnetic field; determining if there is a correlation between the audio signal and the magnetic field; and if there is a correlation between the audio signal and the magnetic field, determining that the audio
30 signal may result from a replay attack.

- Determining if there is a correlation between the audio signal and the magnet field may comprise: identifying first periods during which the audio signal contains speech; identifying second periods during which the magnetic field differs from a baseline; and
35 determining if the first and second periods are substantially the same.

The method may comprise determining that the first and second periods are substantially the same if more than 60% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 60% of the second periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech. The method may comprise determining that the first and second periods are substantially the same if or more than 80% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 80% of the second periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech.

Determining if there is a correlation between the audio signal and the magnetic field may comprise: sampling the detected magnetic field at a first sample rate; sampling the audio signal at a second sample rate; and determining if there is a correlation between the sampled audio signal and the sampled detected magnetic field.

The method may comprise: receiving a series of values of a signal representing a magnetic field strength; forming an average value of the magnetic field strength over a period of time; and subtracting the average value of the magnetic field strength from the series of values of the signal representing the magnetic field strength to form said detected magnetic field.

The method may comprise: obtaining a digital audio signal at a third sample rate, and undersampling said digital audio signal to form said audio signal at said second sample rate.

The second sample rate may be approximately equal to said sample rate.

The step of determining if there is a correlation between the sampled audio signal and the sampled detected magnetic field comprises performing a mathematical correlation operation on the sampled audio signal and the sampled detected magnetic field to obtain an output correlation function, and determining if a peak value of the output correlation function exceeds a predetermined threshold.

The method may further comprise: determining a direction of a source of said audio signal representing speech; determining a direction of a source of said magnetic field; and determining that the audio signal may result from a replay attack if the direction of the source of said audio signal representing speech corresponds to the direction of the source of said magnetic field.

According to an aspect of the present invention, there is provided a system for detecting a replay attack on a voice biometrics system, the system being configured for: receiving an audio signal representing speech; detecting a magnetic field; determining if there is a correlation between the audio signal and the magnetic field; and, if there is a correlation between the audio signal and the magnetic field, determining that the audio signal may result from a replay attack.

The system may be configured for determining if there is a correlation between the audio signal and the magnet field by: identifying first periods during which the audio signal contains speech; identifying second periods during which the magnetic field differs from a baseline; and determining if the first and second periods are substantially the same.

The system may be configured for determining that the first and second periods are substantially the same if more than 60% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 60% of the second periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech.

The system may be configured for determining that the first and second periods are substantially the same if or more than 80% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 80% of the second periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech.

The system may be configured for determining if there is a correlation between the audio signal and the magnet field by: sampling the detected magnetic field at a first sample rate; sampling the audio signal at a second sample rate; and determining if

there is a correlation between the sampled audio signal and the sampled detected magnetic field.

5 The system may be configured for: receiving a series of values of a signal representing a magnetic field strength; forming an average value of the magnetic field strength over a period of time; and subtracting the average value of the magnetic field strength from the series of values of the signal representing the magnetic field strength to form said detected magnetic field.

10 The system may be configured for: obtaining a digital audio signal at a third sample rate, and undersampling said digital audio signal to form said audio signal at said second sample rate.

15 The second sample rate may be approximately equal to said sample rate.

The system may be configured for determining if there is a correlation between the sampled audio signal and the sampled detected magnetic field by performing a mathematical correlation operation on the sampled audio signal and the sampled detected magnetic field to obtain an output correlation function, and determining if a
20 peak value of the output correlation function exceeds a predetermined threshold.

The system may be further configured for: determining a direction of a source of said audio signal representing speech; determining a direction of a source of said magnetic field; and determining that the audio signal may result from a replay attack if the
25 direction of the source of said audio signal representing speech corresponds to the direction of the source of said magnetic field.

According to an aspect of the present invention, there is provided a method of detecting a replay attack on a voice biometrics system. The method comprises:
30 receiving an audio signal representing speech; detecting a magnetic field; and if a strength of the magnetic field exceeds a threshold value, determining that the audio signal may result from a replay attack.

According to an aspect of the present invention, there is provided a system for
35 detecting a replay attack on a voice biometrics system, the system being configured for: receiving an audio signal representing speech; detecting a magnetic field; and if a

strength of the magnetic field exceeds a threshold value, determining that the audio signal may result from a replay attack.

According to an aspect of the present invention, there is provided a method of
5 detecting a replay attack on a voice biometrics system. The method comprises:
receiving an audio signal representing speech; determining a direction of a source of
said audio signal representing speech; detecting a magnetic field; determining a
direction of a source of said magnetic field; and determining that the audio signal may
result from a replay attack if the direction of the source of said audio signal
10 representing speech corresponds to the direction of the source of said magnetic field.

The method may comprise receiving the audio signal representing speech from
multiple microphones.

15 The method may comprise detecting components of the magnetic field in three
orthogonal directions.

According to an aspect of the present invention, there is provided a system for
detecting a replay attack on a voice biometrics system, the system being configured
20 for: receiving an audio signal representing speech; determining a direction of a source
of said audio signal representing speech; detecting a magnetic field; determining a
direction of a source of said magnetic field; and determining that the audio signal may
result from a replay attack if the direction of the source of said audio signal
representing speech corresponds to the direction of the source of said magnetic field.

25 The system may be configured for receiving the audio signal representing speech from
multiple microphones.

The system may be configured for detecting components of the magnetic field in three
30 orthogonal directions.

According to an aspect of the present invention, there is provided a device comprising
a system according to any of the above aspects. The device may comprise a mobile
telephone, an audio player, a video player, a mobile computing platform, a games
35 device, a remote controller device, a toy, a machine, or a home automation controller
or a domestic appliance.

According to an aspect of the present invention, there is provided a computer program product, comprising a computer-readable tangible medium, and instructions for performing a method according to any one of the previous aspects.

5

According to an aspect of the present invention, there is provided a non-transitory computer readable storage medium having computer-executable instructions stored thereon that, when executed by processor circuitry, cause the processor circuitry to perform a method according to any one of the previous aspects.

10

According to an aspect of the present invention, there is provided a device comprising said non-transitory computer readable storage medium. The device may comprise a mobile telephone, an audio player, a video player, a mobile computing platform, a games device, a remote controller device, a toy, a machine, or a home automation controller or a domestic appliance.

15

Brief Description of Drawings

20 For a better understanding of the present invention, and to show how it may be put into effect, reference will now be made to the accompanying drawings, in which:-

Figure 1 illustrates a smartphone.

25 Figure 2 is a schematic diagram, illustrating the form of the smartphone.

Figure 3 illustrates a first situation in which a replay attack is being performed;

Figure 4 illustrates a second situation in which a replay attack is being performed;

30

Figure 5 is a flow chart illustrating a method in accordance with the invention;

Figure 6 is a block diagram of a system for implementing one method;

35 Figure 7 illustrates a result of a method;

Figure 8 illustrates a result of a second method;

Figure 9 illustrates further situations in which a replay attack is being performed; and

5 Figure 10 illustrates a further system for implementing one method.

Detailed Description of Embodiments

10 The description below sets forth example embodiments according to this disclosure. Further example embodiments and implementations will be apparent to those having ordinary skill in the art. Further, those having ordinary skill in the art will recognize that various equivalent techniques may be applied in lieu of, or in conjunction with, the
15 embodiments discussed below, and all such equivalents should be deemed as being encompassed by the present disclosure.

Figure 1 illustrates a smartphone 10, having a microphone 12 for detecting ambient sounds. In normal use, the microphone is of course used for detecting the speech of a user who is holding the smartphone 10.

20

Figure 2 is a schematic diagram, illustrating the form of the smartphone 10.

Specifically, Figure 2 shows various interconnected components of the smartphone 10. It will be appreciated that the smartphone 10 will in practice contain many other
25 components, but the following description is sufficient for an understanding of the present invention.

Thus, Figure 2 shows the microphone 12 mentioned above. In certain embodiments, the smartphone 10 is provided with multiple microphones 12, 12a, 12b, etc.

30

Figure 2 also shows a memory 14, which may in practice be provided as a single component or as multiple components. The memory 14 is provided for storing data and program instructions.

Figure 2 also shows a processor 16, which again may in practice be provided as a single component or as multiple components. For example, one component of the processor 16 may be an applications processor of the smartphone 10.

- 5 Figure 2 also shows a transceiver 18, which is provided for allowing the smartphone 10 to communicate with external networks. For example, the transceiver 18 may include circuitry for establishing an internet connection either over a WiFi local area network or over a cellular network.
- 10 Figure 2 also shows audio processing circuitry 20, for performing operations on the audio signals detected by the microphone 12 as required. For example, the audio processing circuitry 20 may filter the audio signals or perform other signal processing operations.
- 15 Figure 2 also shows at least one sensor 22. In embodiments of the present invention, the sensor is a magnetic field sensor for detecting a magnetic field. For example, the sensor 22 may be a Hall effect sensor, that is able to provide separate measurements of the magnet field strength in three orthogonal directions.
- 20 In this embodiment, the smartphone 10 is provided with voice biometric functionality, and with control functionality. Thus, the smartphone 10 is able to perform various functions in response to spoken commands from an enrolled user. The biometric functionality is able to distinguish between spoken commands from the enrolled user, and the same commands when spoken by a different person. Thus, certain
- 25 embodiments of the invention relate to operation of a smartphone or another portable electronic device with some sort of voice operability, for example a tablet or laptop computer, a games console, a home control system, a home entertainment system, an in-vehicle entertainment system, a domestic appliance, or the like, in which the voice biometric functionality is performed in the device that is intended to carry out the
- 30 spoken command. Certain other embodiments relate to systems in which the voice biometric functionality is performed on a smartphone or other device, which then transmits the commands to a separate device if the voice biometric functionality is able to confirm that the speaker was the enrolled user.
- 35 In some embodiments, while voice biometric functionality is performed on the smartphone 10 or other device that is located close to the user, the spoken commands

are transmitted using the transceiver 18 to a remote speech recognition system, which determines the meaning of the spoken commands. For example, the speech recognition system may be located on one or more remote server in a cloud computing environment. Signals based on the meaning of the spoken commands are then
5 returned to the smartphone 10 or other local device.

One attempt to deceive a voice biometric system is to play a recording of an enrolled user's voice in a so-called replay or spoof attack.

10 **Figure 3** shows an example of a situation in which a replay attack is being performed. Thus, in Figure 3, the smartphone 10 is provided with voice biometric functionality. In this example, the smartphone 10 is in the possession, at least temporarily, of an attacker, who has another smartphone 30. The smartphone 30 has been used to record the voice of the enrolled user of the smartphone 10. The smartphone 30 is
15 brought close to the microphone inlet 12 of the smartphone 10, and the recording of the enrolled user's voice is played back. If the voice biometric system is unable to detect that the enrolled user's voice that it detects is a recording, the attacker will gain access to one or more services that are intended to be accessible only by the enrolled user.

20 It is known that smartphones, such as the smartphone 30, are typically provided with loudspeakers that are of relatively low quality due to size constraints. Thus, the recording of an enrolled user's voice played back through such a loudspeaker will not be a perfect match with the user's voice, and this fact can be used to identify replay attacks. For example, loudspeakers may have certain frequency characteristics, and if
25 these frequency characteristics can be detected in a speech signal that is received by the voice biometrics system, it may be considered that the speech signal has resulted from a replay attack.

Figure 4 shows a second example of a situation in which a replay attack is being
30 performed, in an attempt to overcome the method of detection described above. Thus, in Figure 4, the smartphone 10 is provided with voice biometric functionality. Again, in this example, the smartphone 10 is in the possession, at least temporarily, of an attacker, who has another smartphone 40. The smartphone 40 has been used to record the voice of the enrolled user of the smartphone 10.

35

In this example, the smartphone 40 is connected to a high quality loudspeaker 50. Then, the microphone inlet 12 of the smartphone 10 is positioned close to the loudspeaker 50, and the recording of the enrolled user's voice is played back through the loudspeaker 50. As before, if the voice biometric system is unable to detect that
5 the enrolled user's voice that it detects is a recording, the attacker will gain access to one or more services that are intended to be accessible only by the enrolled user.

In this example, the loudspeaker 50 may be of high enough quality that the recording of the enrolled user's voice played back through the loudspeaker will not be reliably
10 distinguishable from the user's voice, and so the audio features of the speech signal cannot be used to identify the replay attack.

However, it is appreciated that many loudspeakers, and particularly high quality loudspeakers, are electromagnetic loudspeakers in which an electrical audio signal is
15 applied to a voice coil, which is located between the poles of a permanent magnet, causing the coil to move rapidly backwards and forwards. This movement causes a diaphragm attached to the coil to move backwards and forwards, creating sound waves. It is recognised here that, if a device such as the smartphone 10 is positioned close to a loudspeaker while it is playing back sounds, there will be corresponding
20 changes in the magnetic field, which will be detectable by a magnetic field sensor 22.

Figure 5 is a flow chart, illustrating a method of detecting a replay attack on a voice biometrics system, and **Figure 6** is a block diagram illustrating functional blocks in the
25 voice biometrics system.

Specifically, in step 60 in the method of Figure 5, an audio signal is received on an input 80 of the system shown in Figure 6. For example, in a device as shown in Figure 2, the audio signal received on the input 80 may be the audio signal detected by the microphone 12, or may be the sum of the audio signals detected by the microphones if
30 there is more than one.

At the same time, in step 62 in the method of Figure 5, an input signal is received on an input 82 of the system shown in Figure 6. The input signal received on the input 82 is received from a magnetometer. For example, when the method is performed in a
35 device such as a smartphone or a tablet computer, the device will typically include a

three-axis magnetometer, which generates an output signal containing separate measurements of the magnetic field strength in three orthogonal directions.

5 In some embodiments, the input signal received from the magnetometer is passed to a first pre-processing block 84. For example, if the signal received from the magnetometer contains separate measurements of the magnetic field strength in three orthogonal directions, these can be combined to provide a single measurement of the magnetic field strength. The measurement of the magnetic field strength could be found as the square root of the sum of the squares of the three separate
10 measurements of the magnetic field strength in the three orthogonal directions.

Further, the aim of the system is to determine any magnetic field that is generated by a nearby object such as a loudspeaker. In order to obtain the most useful information about this, one possibility is to process the input signal received from the
15 magnetometer in order to remove the effects of the Earth's magnetic field. For example, this can be achieved by forming an average value of the magnetic field strength, for example over a period of at least several seconds, and possibly several minutes or hours, and subtracting this from each individual measurement to obtain an instantaneous measurement of the magnetic field generated by artificial sources.
20 When the measurements of the magnetic field strength in three orthogonal directions are considered separately, these values will depend heavily on the orientation of the device in the Earth's magnetic field. The orientation can be determined from the signals generated by accelerometers present in the device, and so the orientation can be taken into account when determining the artificial magnetic field that is generated by
25 a nearby object such as a loudspeaker.

Typically, in a smartphone, a magnetometer generates a digital signal, with a sampling rate in the region of 80-120Hz, which can be applied as the input signal on the input 82
30 of the system.

In some embodiments, the audio signal received on the input 80 is passed to a second pre-processing block 86. For example, if the audio signal is received in an analog form, the pre-processing block 86 may comprise an analog-to-digital converter for converting the signal to a digital form.
35

In some embodiments, the pre-processing block 86 may comprise a digital or analog filter for correcting for expected non-linearities in the frequency response of a loudspeaker whose presence is being detected. Thus, the relationship between the magnetic field and the frequency in a typical loudspeaker will have a notch shape, that is, at one particular frequency around the mechanical resonance of the speaker, the magnetic field will be particularly low. The pre-processing block 86 may then apply a filter having a similar characteristic to the received audio signal, in order to improve a degree of correlation between the audio signal and the magnetic field.

10 In this illustrated embodiment, the second pre-processing block 86 comprises a decimation block. If the analog-to-digital converter in the second pre-processing block 86 generates a digital audio signal at a sampling rate that exceeds the sampling rate of the magnetometer signal, then samples of the digital audio signal are discarded, so that the resulting sample rate is approximately equal to the sampling rate, R , of the magnetometer signal, for example the audio sampling rate should be in the range 15 $0.5R$ - $2R$, more preferably $0.8R$ - $1.2R$. For example, if the input signal received from the magnetometer has a sampling rate in the region of 80-120Hz, and the analog-to-digital converter in the second pre-processing block 86 has a sampling rate of 40kHz (as would be typical for an analog-to-digital converter that would typically be present in a device such as a smartphone, and could be used for an accurate digital representation of an analog audio signal), then only one in every 400 samples of the analog-to-digital converter would be retained, so that the resulting sample rate is $40\text{kHz}/400 = 100\text{Hz}$.

Alternatively, the magnetometer signal may be upsampled to the sampling rate of the audio signal, by interleaving zero-value samples between the samples of the magnetometer signal.

In step 64 in the method of Figure 5, it is determined whether there is a correlation between the audio signal and the magnetic field. Thus, in Figure 6 the outputs of the first and second pre-processing blocks 84, 86 are passed to a correlation block 88.

The correlation block 88 can operate in different ways.

Figure 7 illustrates a first method of determining whether there is a correlation between the audio signal and the magnetic field.

Thus, Figure 7(a) illustrates the form of the decimated audio signal generated by the second pre-processing block 86, while Figure 7(b) illustrates the form of the magnetometer output. In both Figure 7(a) and Figure 7(b), the horizontal axes represent time. More specifically, the units on the horizontal axes are samples of the
5 respective digital samples. In each case, the sample rate is ≈ 100 Hz, and so 1000 samples is ≈ 10 seconds. In both Figure 7(a) and Figure 7(b), the vertical axes represent the strengths of the respective signals, in arbitrary units. In the case of Figure 7(b), the effect of the Earth's magnetic field has been removed by forming an average value of the magnetic field strength, for example over a period of several
10 seconds. This average value is then taken as a baseline, and subtracted from each individual measurement. Figure 7(b) then shows these individual measurements as differences from that baseline, representing an instantaneous measurement of the magnetic field generated by artificial sources.

15 In this example, the sample rate of the decimated audio signal generated by the second pre-processing block 86 is exactly the same as the sample rate of the magnetometer output. Therefore, Figure 7, which shows equal numbers of samples of both signals, covers equal periods of time for both inputs.

20 It can be seen that the audio signal contains a noticeable input during periods from about samples 15-85, samples 115-260, samples 300-365, samples 395-545, etc. For example, it may be determined that the audio signal contains a relevant input when the magnitude of a sample value exceeds a threshold value, or when a sample value
25 magnitude, averaged over a relatively small number of samples, exceeds a threshold value. It can therefore be assumed that the user's speech is present during the periods of time during which these samples were taken.

It can also be seen that the magnetometer output also contains a noticeable input during the same periods. For example, it may be determined that the magnetometer
30 output contains a relevant input when the magnitude of a sample value exceeds a threshold value, or when a sample value magnitude, averaged over a relatively small number of samples, exceeds a threshold value. It can therefore be assumed the device was in the presence of a loudspeaker that was generating sounds during the periods of time during which these samples were taken.

If the device was in the presence of a loudspeaker that was generating sounds at the same time as the microphone was detecting speech, then this may suggest that the speech was being played by the loudspeaker, and hence that the device was the object of a replay attack.

5

The correlation block 88 may therefore identify first periods during which the audio signal contains speech, and may identify second periods during which there is a significant magnetic field.

10 In step 66 in the method of Figure 5, it is determined whether the audio signal may result from a replay attack. Thus, in Figure 6 the results of the determinations of the correlation block 88 are passed to a decision block 90, which determines if the first and second periods are substantially the same. If so, it is determined that the audio signal may result from a replay attack.

15

For example, the decision block 90 may determine that the first and second periods are substantially the same if more than 60% of the first periods during which the audio signal contains speech overlap with second periods during which there is a significant magnetic field, and/or more than 60% of the second periods during which there is a significant magnetic field overlap with first periods during which the audio signal contains speech, or more than 80% of the first periods during which the audio signal contains speech overlap with second periods during which there is a significant magnetic field, and/or more than 80% of the second periods during which there is a significant magnetic field overlap with first periods during which the audio signal contains speech.

25

The method illustrated in Figure 7 is particularly effective when the audio signal and the magnetometer output are not subject to large amounts of noise. **Figure 8** illustrates a second method of determining whether there is a correlation between the audio signal and the magnetic field.

30

This second method forms the mathematical cross-correlation between the decimated audio signal generated by the second pre-processing block 86 and the samples of the magnetometer output. That is, the sequence of samples in one of the signals is correlated with a delayed copy of the other signal, for a range of delay values. The degree of correlation will be a function of the delay, which can conveniently be

35

measured by the number of sample periods by which the delayed version has been delayed. Conventionally, autocorrelations are performed on the two signals, and the size of the correlation for any delay value is normalised against the sizes of the autocorrelations of the two signals at zero delay.

5

Thus, the correlation $R_{xy}[n]$ between the two signals $x[m]$ and $y[m]$, as a function of the number of samples, n , by which the second signal is delayed, is given by:

$$R_{xy}[n] = \sum_{m=-\infty}^{\infty} x^*[m]y[n+m]$$

10 and, after normalisation:

$$r_{xy}[n] = \frac{R_{xy}[n]}{\sqrt{R_{xx}[0]R_{yy}[0]}}$$

Figure 8 illustrates the result of obtaining the cross-correlation in one illustrative example. Specifically, the trace 100 shows the result obtained when the audio input is obtained from a live user's speech, while the trace 102 shows the result obtained when the audio input is obtained by a playback of the user's speech through a loudspeaker.

It can be seen that the trace 100 fluctuates, but there is no clear pattern. The trace 102 fluctuates in a similar way, but it is noticeable that, at one or two particular delay values 104, there is a very high degree of correlation. These particular delay values correspond to a delay of zero, that is, the two signals are correlated.

This can be assumed to be a consequence of the fact that the audio input is obtained by a playback of the user's speech through a loudspeaker, and hence that the loudspeaker is generating a magnetic field in synchronisation with the sounds that it is generating.

This method picks out the correlation, even when either or both of the audio signal and the magnetometer output contain significant amounts of noise.

In step 66 in the method of Figure 5, it is determined on the basis of the correlation whether the audio signal may result from a replay attack. Thus, in Figure 6 the results

of the determinations of the correlation block 88 are passed to a decision block 90, which determines if the correlation is such that it should be determined that the audio signal may result from a replay attack. For example, in some embodiments the decision block 90 determines that the audio signal may result from a replay attack if the peak value of the cross-correlation (or, in particular, the peak value of the cross-correlation, which may occur at a delay value corresponding to synchronisation of the audio input and the measured magnetic field) exceeds a threshold value. Figure 8 shows the result of performing the cross-correlation on one frame of data (for example comprising 1000 samples or about 10 seconds) of data. Where the input signal continues for longer than this, the decision block 90 may determine that the audio signal may result from a replay attack by considering multiple frames of data, for example by determining that the audio signal may have resulted from a replay attack if the peak value of the cross-correlation exceeds a threshold value in every frame, or if the peak value of the cross-correlation, averaged over several frames, exceeds a threshold value.

In addition to determining whether the variation in the magnetic field is correlated in time with the variation in the audio signal, it is also possible to determine whether the direction of a source of the magnetic field corresponds to a direction of a source of the audio signal.

Figure 9 illustrates how this can be done. Specifically, Figure 9 illustrates two further situations in which a replay attack is being performed.

Figure 9(a) shows a situation with a loudspeaker 120 placed on a surface, and a smartphone 122 being held vertically in front of the loudspeaker and facing it. The smartphone 122 is shown partially in section, so that internal components of the smartphone can be shown.

Specifically, Figure 9(a) shows three microphones 124, 126, and 128 located respectively near the centre of the top edge of the smartphone, near the bottom left corner of the smartphone (as the user looks at the front of the smartphone), and near the bottom right corner of the smartphone.

In addition, Figure 9(a) shows a three-axis magnetometer 130, which produces separate measurements of the magnetic field strengths in the x, y, and z directions as shown in Figure 9(a).

5 The loudspeaker 120 is an electromagnetic loudspeaker, in which sound is produced by the movement of a coil, with the coil being moved by a magnetic field. As shown in Figure 9(a), the magnetic field M is oriented out of the front face 132 of the loudspeaker 120. Provided that the smartphone 122 is positioned sufficiently close to the front of the loudspeaker 120, it can be assumed that the direction of the magnetic field sensed
10 by the magnetometer 130 will be generally in the z direction. Thus, if the measurements of the magnetic field strengths in the x, y, and z directions show that the magnetic field in the z direction is predominant, then it can be assumed that the smartphone 122 is positioned close to the front of the loudspeaker 120 in the orientation illustrated in Figure 9(a).

15

The signals received from the three microphones 124, 126, 128 can also be used to determine the direction of the source of the audio signal, by using known techniques. For example, in the situation shown in Figure 9(a), the audio signal generated by the loudspeaker 120 will be received by the three microphones 124, 126, 128 at essentially
20 the same time. This can be used to determine the approximate direction in which the source of the audio signal is located.

Thus, in this example, it can be determined that the direction of the source of the magnetic field generally corresponds to the direction of the source of the audio signal.
25 This can be used to provide further confirmation that the audio signal is the result of a replay attack.

Figure 9(b) shows an alternative situation with the loudspeaker 120 placed on a surface, and the smartphone 122 being placed face upwards on the same surface.

30

In this situation, provided that the smartphone 122 is positioned sufficiently close to the front of the loudspeaker 120, it can be assumed that the direction of the magnetic field sensed by the magnetometer 130 will be generally in the y direction. Thus, if the measurements of the magnetic field strengths in the x, y, and z directions show that the
35 magnetic field in the y direction is predominant, then it can be assumed that the

smartphone 122 is positioned close to the front of the loudspeaker 120 in the orientation illustrated in Figure 9(b).

Again, the signals received from the three microphones 124, 126, 128 can also be
5 used to determine the location of the source of the audio signal, by using known techniques. For example, in the situation shown in Figure 9(b), the audio signal generated by the loudspeaker 120 will be received by the two microphones 126, 128 at essentially the same time, and slightly before the audio signal is received by the microphone 124. This can be used to determine the approximate direction in which the
10 source of the audio signal is located.

Thus, again, it can be determined that the source of the magnetic field generally corresponds to the source of the audio signal. This can be used to provide further confirmation that the audio signal is the result of a replay attack.
15

In other embodiments, the method comprises receiving an audio signal representing speech, and detecting a magnetic field. In these embodiments, if a strength of the magnetic field exceeds a threshold value, it is determined that the audio signal may
20 result from a replay attack. These embodiments are particularly suitable when the replay attack is generated using a loudspeaker that contains a large magnet, and hence the presence of a large magnetic field is indicative of a replay attack. In such cases, the magnetic field strength may be several times, and possibly orders of magnitude, greater than the baseline magnetic field strength caused by the Earth's
25 magnetic field. Therefore, in these cases, it is not necessary to determine the baseline magnetic field strength and subtract it from the individual measurements.

In other embodiments, the method comprises receiving an audio signal representing speech, and detecting a magnetic field. It is possible to determine a direction of a
30 source of said audio signal representing speech, for example using beamforming techniques if the signal is detected using multiple microphones. It is also possible to determine a direction of a source of said magnetic field. If the direction of the source of said audio signal representing speech corresponds to the direction of the source of said magnetic field, it may be determined that the audio signal may result from a replay
35 attack.

Figure 10 illustrates a system for determining whether a direction of a source of a magnetic field generally corresponds to a direction of a source of an audio signal representing speech, whether this result is used as further confirmation that the audio signal is the result of a replay attack, in conjunction with other methods, or whether it is used as the sole indication that the audio signal is the result of a replay attack.

Figure 10 shows a processor 150. The processor 150 receives input signals from multiple microphones 12a, ..., 12n, after suitable conditioning in pre-processing blocks 152a, ..., 152n. Similarly, the processor 150 receives separate input signals from the three magnetometers 154, 156, 158 that measure the magnetic field strength in the three orthogonal directions, multiple again after suitable conditioning in pre-processing blocks 160, 162, 164.

The processor 150 may separately calculate the direction of the source of the audio signal, for example using standard beamforming techniques, and the direction of the source of the magnetic field, and may then check for correlation between them.

Alternatively, the processor 150 may be a Neural Network, pre-trained using samples of representative loudspeakers with various orientations relative to the target device.

There are therefore disclosed methods and systems that can be used for detecting situations that may indicate that a received audio signal is the result of a replay attack.

The skilled person will recognise that some aspects of the above-described apparatus and methods may be embodied as processor control code, for example on a non-volatile carrier medium such as a disk, CD- or DVD-ROM, programmed memory such as read only memory (Firmware), or on a data carrier such as an optical or electrical signal carrier. For many applications embodiments of the invention will be implemented on a DSP (Digital Signal Processor), ASIC (Application Specific Integrated Circuit) or FPGA (Field Programmable Gate Array). Thus the code may comprise conventional program code or microcode or, for example code for setting up or controlling an ASIC or FPGA. The code may also comprise code for dynamically configuring re-configurable apparatus such as re-programmable logic gate arrays. Similarly the code may comprise code for a hardware description language such as

Verilog TM or VHDL (Very high speed integrated circuit Hardware Description Language). As the skilled person will appreciate, the code may be distributed between a plurality of coupled components in communication with one another. Where appropriate, the embodiments may also be implemented using code running on a field-
5 (re)programmable analogue array or similar device in order to configure analogue hardware.

Note that as used herein the term module shall be used to refer to a functional unit or block which may be implemented at least partly by dedicated hardware components
10 such as custom defined circuitry and/or at least partly be implemented by one or more software processors or appropriate code running on a suitable general purpose processor or the like. A module may itself comprise other modules or functional units. A module may be provided by multiple components or sub-modules which need not be co-located and could be provided on different integrated circuits and/or running on
15 different processors.

Embodiments may be implemented in a host device, especially a portable and/or battery powered host device such as a mobile computing device for example a laptop or tablet computer, a games console, a remote control device, a home automation
20 controller or a domestic appliance including a domestic temperature or lighting control system, a toy, a machine such as a robot, an audio player, a video player, or a mobile telephone for example a smartphone.

It should be noted that the above-mentioned embodiments illustrate rather than limit
25 the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. The word “comprising” does not exclude the presence of elements or steps other than those listed in a claim, “a” or “an” does not exclude a plurality, and a single feature or other unit may fulfil the functions of several units recited in the claims. Any reference
30 numerals or labels in the claims shall not be construed so as to limit their scope.

CLAIMS

1. A method of detecting a replay attack on a voice biometrics system, the method comprising:
 - 5 receiving an audio signal representing speech;
 - detecting a magnetic field;
 - determining if there is a correlation between the audio signal and the magnetic field; and
 - if there is a correlation between the audio signal and the magnetic field,
 - 10 determining that the audio signal may result from a replay attack.
2. A method according to claim 1, wherein determining if there is a correlation between the audio signal and the magnet field comprises:
 - identifying first periods during which the audio signal contains speech;
 - 15 identifying second periods during which the magnetic field differs from a baseline;
 - and
 - determining if the first and second periods are substantially the same.
3. A method according to claim 2, comprising determining that the first and second
 - 20 periods are substantially the same if more than 60% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 60% of the second periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech.
4. A method according to claim 3, comprising determining that the first and second
 - periods are substantially the same if or more than 80% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 80% of the second
 - 30 periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech.
5. A method according to claim 1, wherein determining if there is a correlation between the audio signal and the magnet field comprises:
 - 35 sampling the detected magnetic field at a first sample rate;
 - sampling the audio signal at a second sample rate; and

determining if there is a correlation between the sampled audio signal and the sampled detected magnetic field.

6. A method according to claim 5, comprising:
- 5 receiving a series of values of a signal representing a magnetic field strength;
forming an average value of the magnetic field strength over a period of time; and
subtracting the average value of the magnetic field strength from the series of
values of the signal representing the magnetic field strength to form said detected
magnetic field.
- 10
7. A method according to claim 5 or 6, comprising:
- obtaining a digital audio signal at a third sample rate, and undersampling said
digital audio signal to form said audio signal at said second sample rate.
- 15
8. A method according to claim 5, 6 or 7, wherein said second sample rate is
approximately equal to said sample rate.
9. A method according to claim 5, 6, 7 or 8, wherein the step of determining if there
is a correlation between the sampled audio signal and the sampled detected magnetic
20 field comprises performing a mathematical correlation operation on the sampled audio
signal and the sampled detected magnetic field to obtain an output correlation function,
and determining if a peak value of the output correlation function exceeds a
predetermined threshold.
- 25
10. A method according to any preceding claim, further comprising:
- determining a direction of a source of said audio signal representing speech;
determining a direction of a source of said magnetic field; and
determining that the audio signal may result from a replay attack if the direction of
the source of said audio signal representing speech corresponds to the direction of the
30 source of said magnetic field.
11. A system for detecting a replay attack on a voice biometrics system, the system
being configured for:
- receiving an audio signal representing speech;
- 35 detecting a magnetic field;

determining if there is a correlation between the audio signal and the magnetic field; and

if there is a correlation between the audio signal and the magnetic field, determining that the audio signal may result from a replay attack.

5

12. A system according to claim 11, configured for determining if there is a correlation between the audio signal and the magnet field by:

identifying first periods during which the audio signal contains speech;

identifying second periods during which the magnetic field differs from a baseline;

10 and

determining if the first and second periods are substantially the same.

13. A system according to claim 12, configured for determining that the first and second periods are substantially the same if more than 60% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 60% of the second periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech.

14. A system according to claim 13, configured for determining that the first and second periods are substantially the same if or more than 80% of the first periods during which the audio signal contains speech overlap with second periods during which the magnetic field differs significantly from a baseline, and/or more than 80% of the second periods during which the magnetic field differs significantly from a baseline overlap with first periods during which the audio signal contains speech.

15. A system according to claim 11, configured for determining if there is a correlation between the audio signal and the magnet field by:

sampling the detected magnetic field at a first sample rate;

sampling the audio signal at a second sample rate; and

determining if there is a correlation between the sampled audio signal and the sampled detected magnetic field.

16. A system according to claim 15, configured for:

receiving a series of values of a signal representing a magnetic field strength;

forming an average value of the magnetic field strength over a period of time; and

subtracting the average value of the magnetic field strength from the series of values of the signal representing the magnetic field strength to form said detected magnetic field.

- 5 17. A system according to claim 15 or 16, configured for:
 obtaining a digital audio signal at a third sample rate, and undersampling said
 digital audio signal to form said audio signal at said second sample rate.
- 10 18. A system according to claim 15, 16 or 17, wherein said second sample rate is
 approximately equal to said sample rate.
- 15 19. A system according to claim 15, 16, 17 or 18, configured for determining if there
 is a correlation between the sampled audio signal and the sampled detected magnetic
 field by performing a mathematical correlation operation on the sampled audio signal
 and the sampled detected magnetic field to obtain an output correlation function, and
 determining if a peak value of the output correlation function exceeds a predetermined
 threshold.
- 20 20. A system according to any of claims 11 to 19, further configured for:
 determining a direction of a source of said audio signal representing speech;
 determining a direction of a source of said magnetic field; and
 determining that the audio signal may result from a replay attack if the direction of
 the source of said audio signal representing speech corresponds to the direction of the
 source of said magnetic field.
- 25 21. A method of detecting a replay attack on a voice biometrics system, the method
 comprising:
 receiving an audio signal representing speech;
 detecting a magnetic field; and
 if a strength of the magnetic field exceeds a threshold value, determining that the
 audio signal may result from a replay attack.
- 30 22. A system for detecting a replay attack on a voice biometrics system, the system
 being configured for:
 receiving an audio signal representing speech;
 detecting a magnetic field; and
- 35

if a strength of the magnetic field exceeds a threshold value, determining that the audio signal may result from a replay attack.

23. A method of detecting a replay attack on a voice biometrics system, the method
5 comprising:
 receiving an audio signal representing speech;
 determining a direction of a source of said audio signal representing speech;
 detecting a magnetic field;
 determining a direction of a source of said magnetic field; and
10 determining that the audio signal may result from a replay attack if the direction of
the source of said audio signal representing speech corresponds to the direction of the
source of said magnetic field.
24. A method as claimed in claim 23, comprising receiving the audio signal
15 representing speech from multiple microphones.
25. A method as claimed in claim 23 or 24, comprising detecting components of the
magnetic field in three orthogonal directions.
- 20 26. A system for detecting a replay attack on a voice biometrics system, the system
being configured for:
 receiving an audio signal representing speech;
 determining a direction of a source of said audio signal representing speech;
 detecting a magnetic field;
25 determining a direction of a source of said magnetic field; and
 determining that the audio signal may result from a replay attack if the direction of
the source of said audio signal representing speech corresponds to the direction of the
source of said magnetic field.
- 30 27. A system as claimed in claim 26, configured for receiving the audio signal
representing speech from multiple microphones.
28. A system as claimed in claim 26 or 27, configured for detecting components of
the magnetic field in three orthogonal directions.

29. A device comprising a system as claimed in any of claims 11 to 20, 22, or 26 to 28.

30. A device as claimed in claim 29, wherein the device comprises a mobile
5 telephone, an audio player, a video player, a mobile computing platform, a games device, a remote controller device, a toy, a machine, or a home automation controller or a domestic appliance.

31. A computer program product, comprising a computer-readable tangible medium,
10 and instructions for performing a method according to any one of claims 1 to 10, 21, or 23 to 25.

32. A non-transitory computer readable storage medium having computer-executable
instructions stored thereon that, when executed by processor circuitry, cause the
15 processor circuitry to perform a method according to any one of claims 1 to 10, 21, or 23 to 25.

33. A device comprising the non-transitory computer readable storage medium as
claimed in claim 32.

20

34. A device as claimed in claim 33, wherein the device comprises a mobile
telephone, an audio player, a video player, a mobile computing platform, a games
device, a remote controller device, a toy, a machine, or a home automation controller
or a domestic appliance.

25

1/11

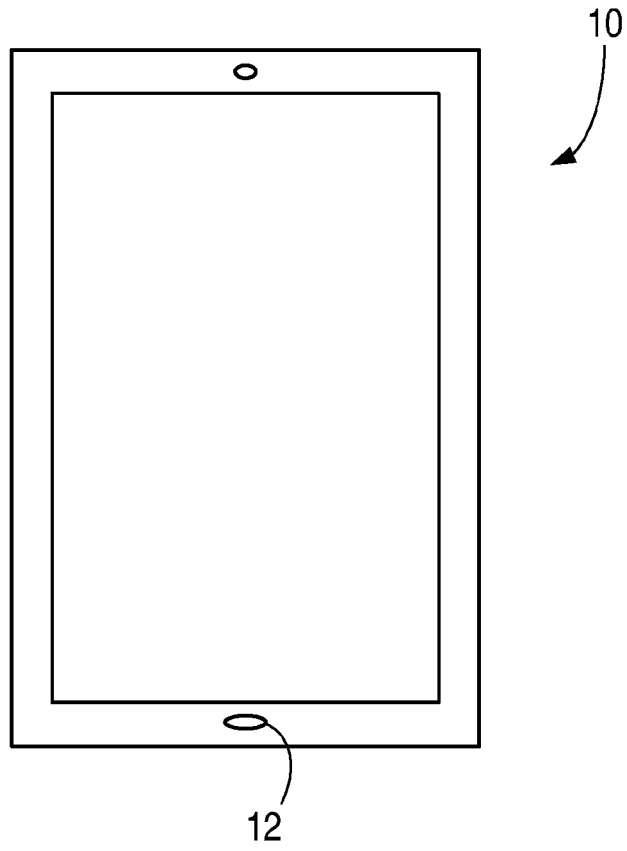


Figure 1

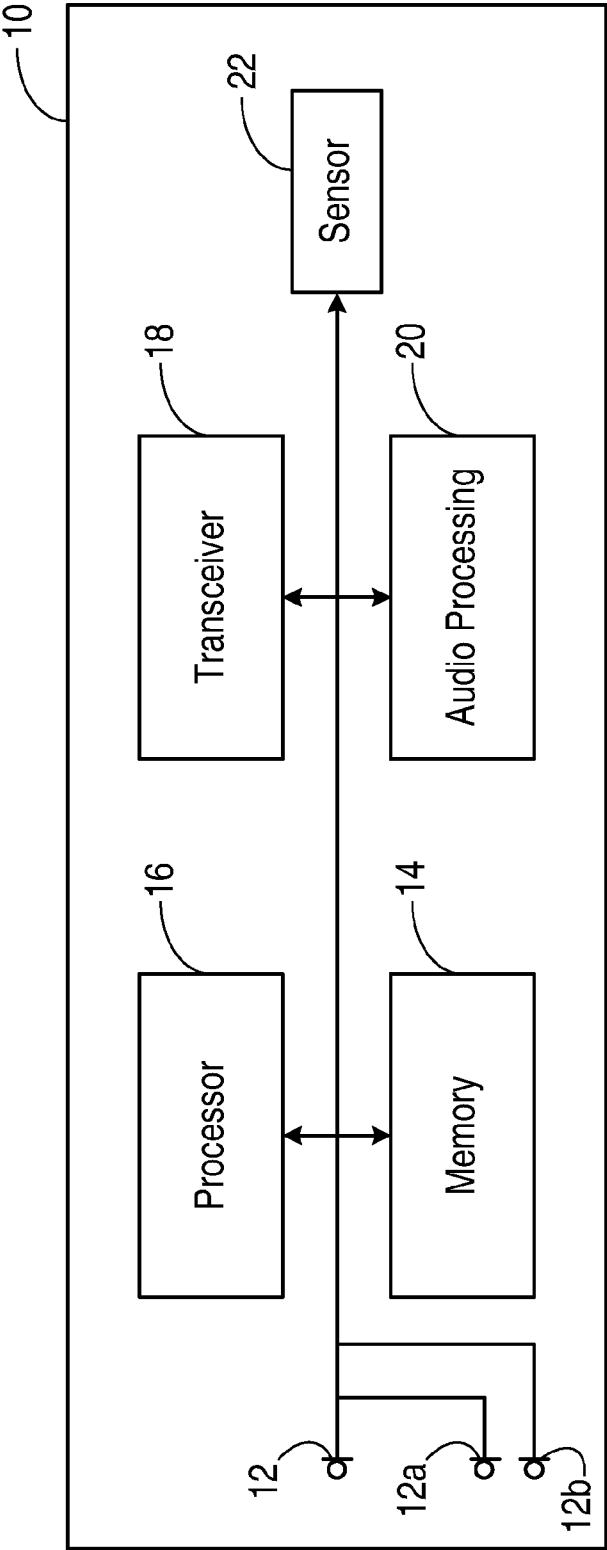


Figure 2

3/11

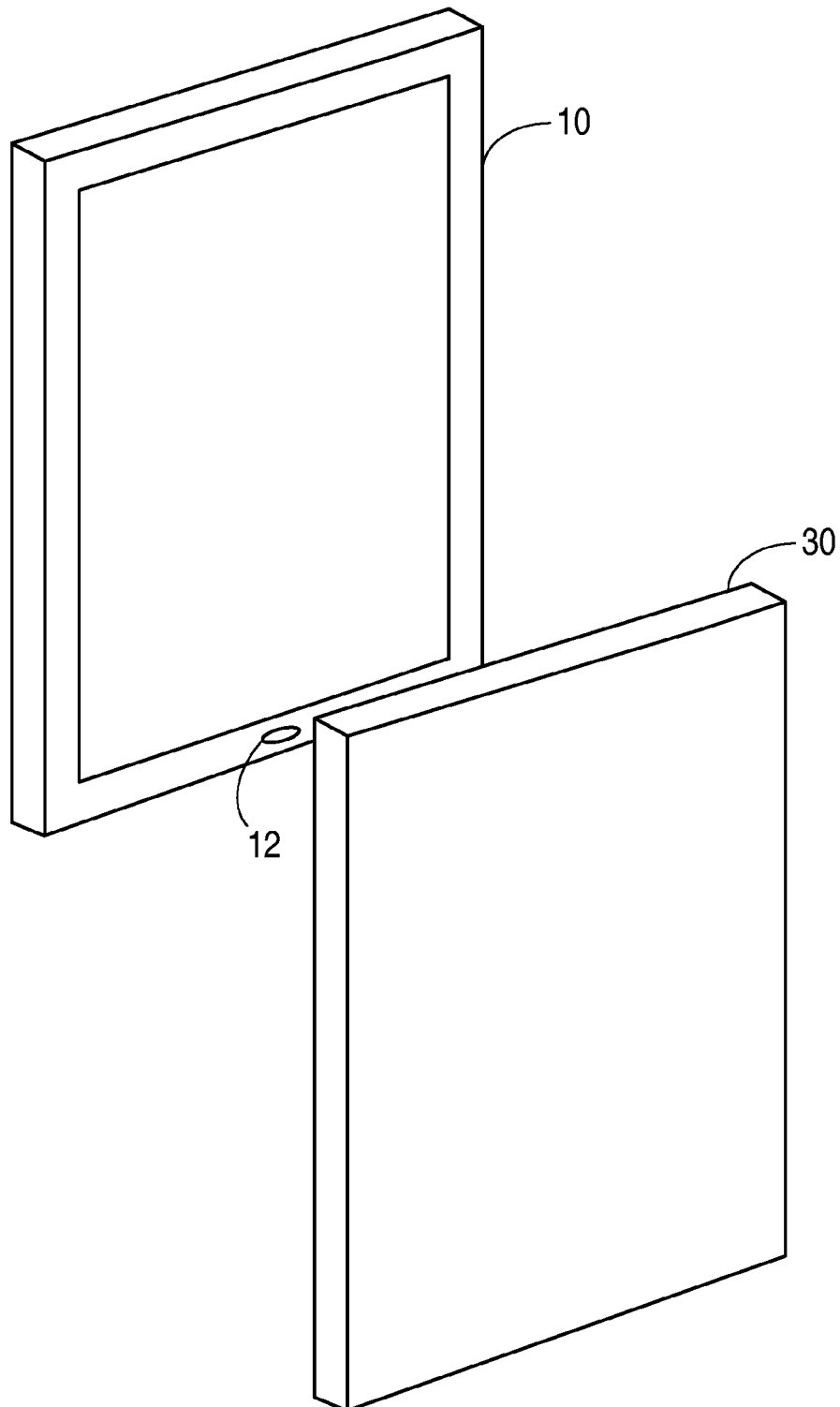


Figure 3

4/11

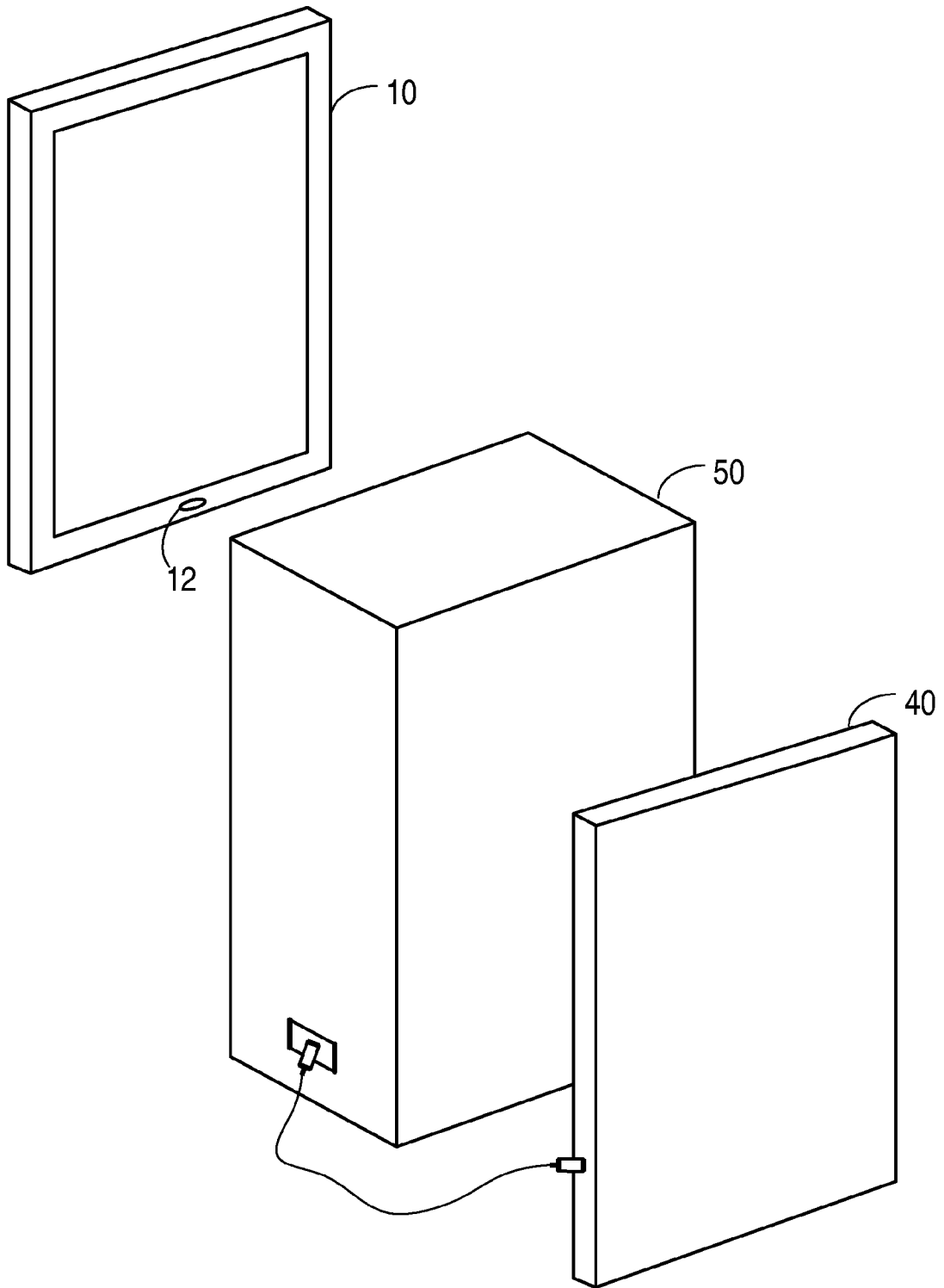


Figure 4

5/11

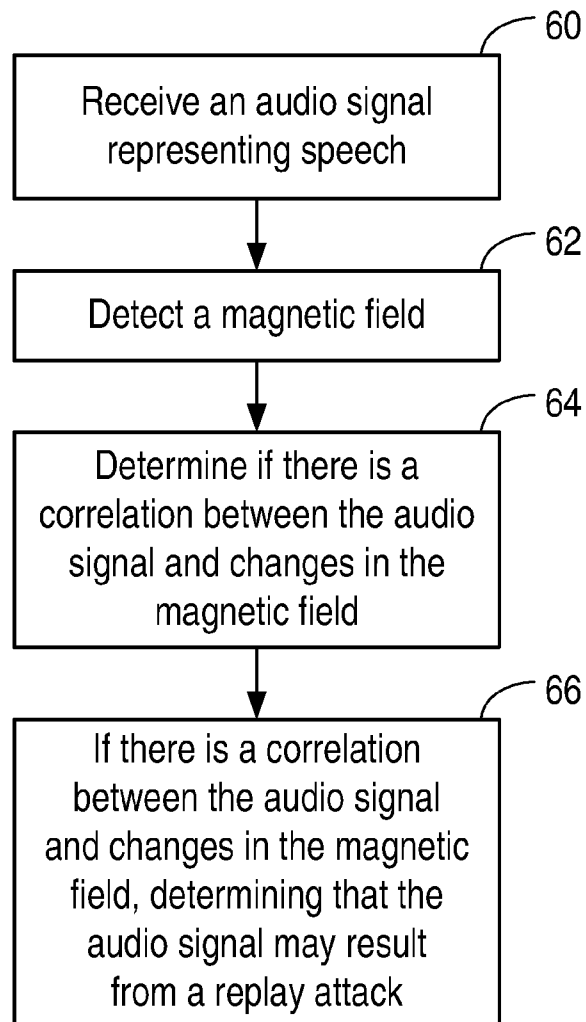


Figure 5

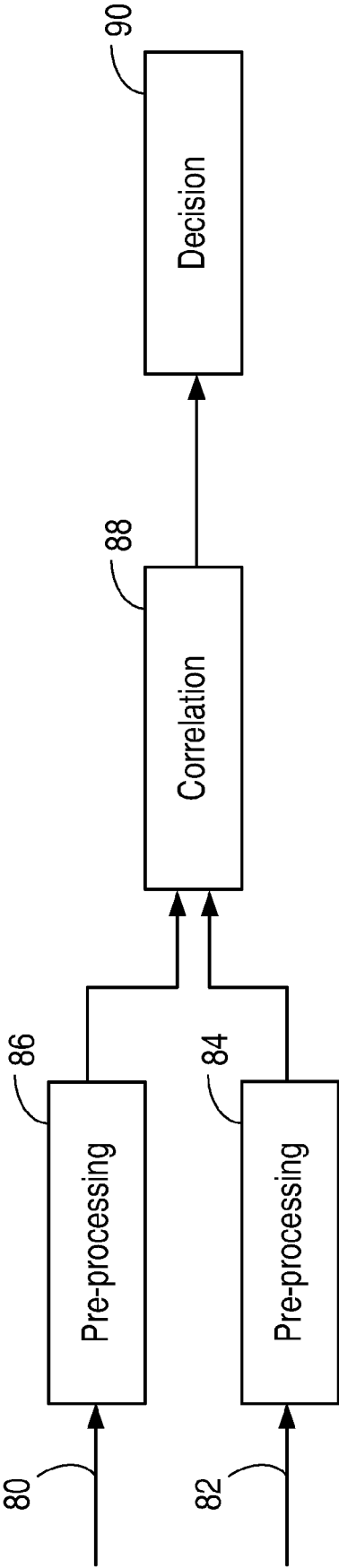


Figure 6

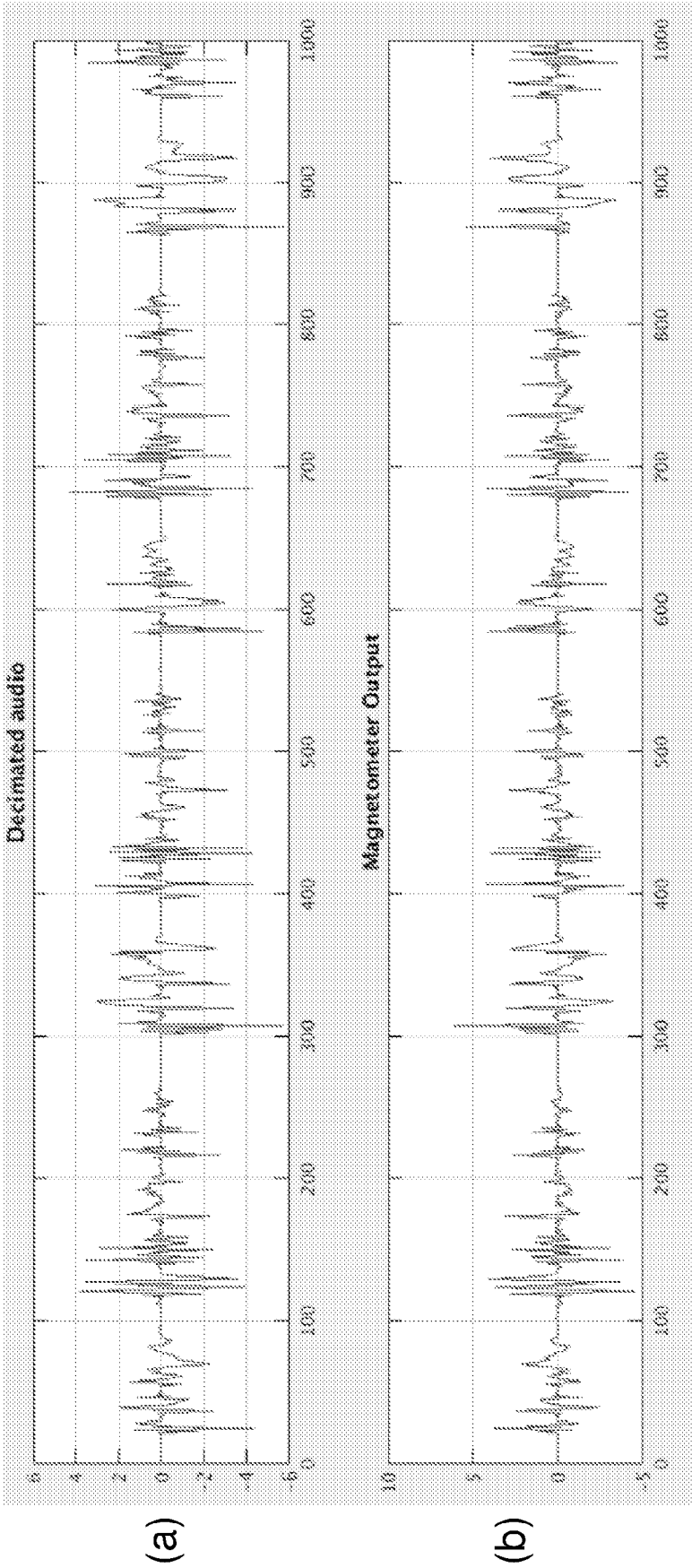


Figure 7

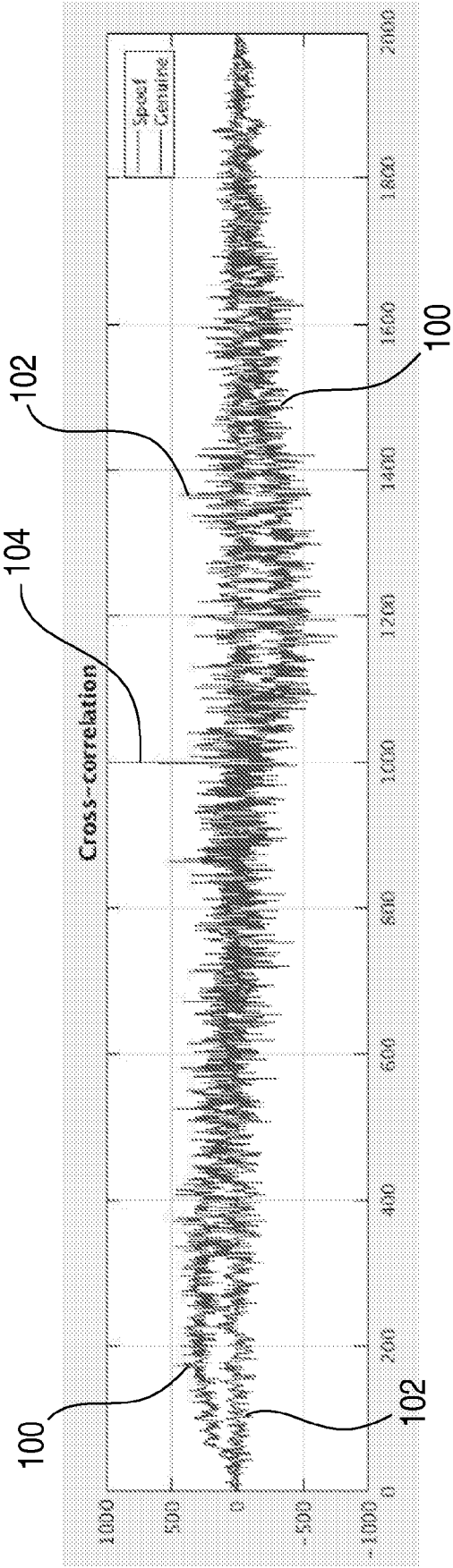


Figure 8

9/11

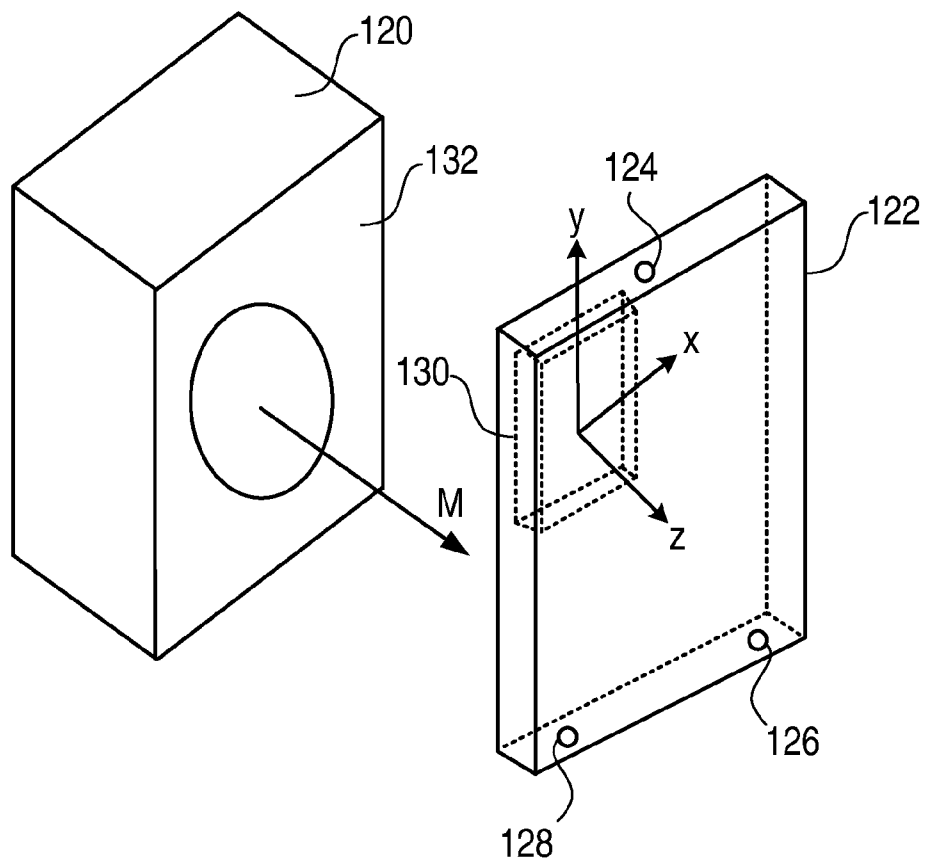


Figure 9(a)

10/11

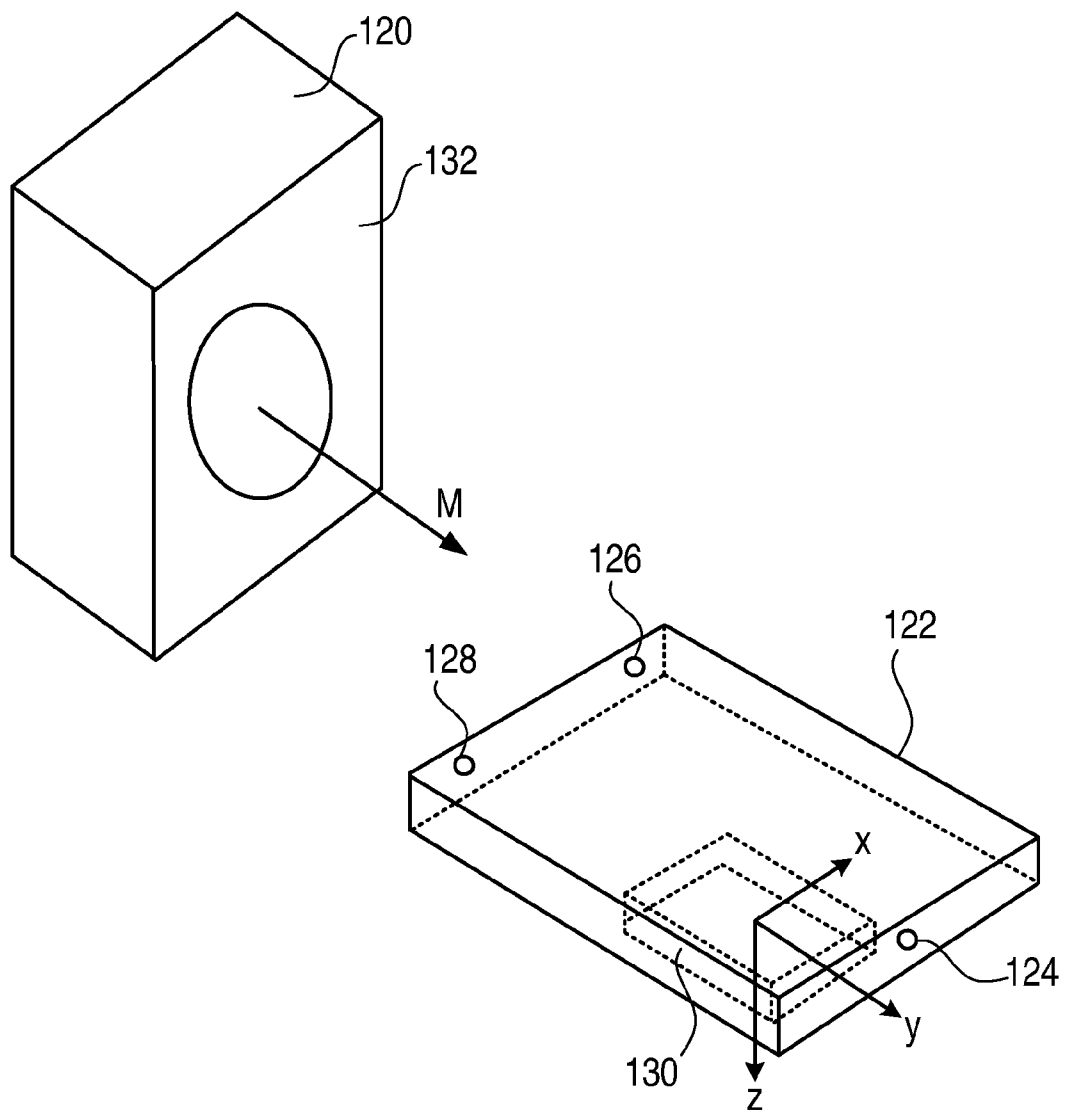


Figure 9(b)

11/11

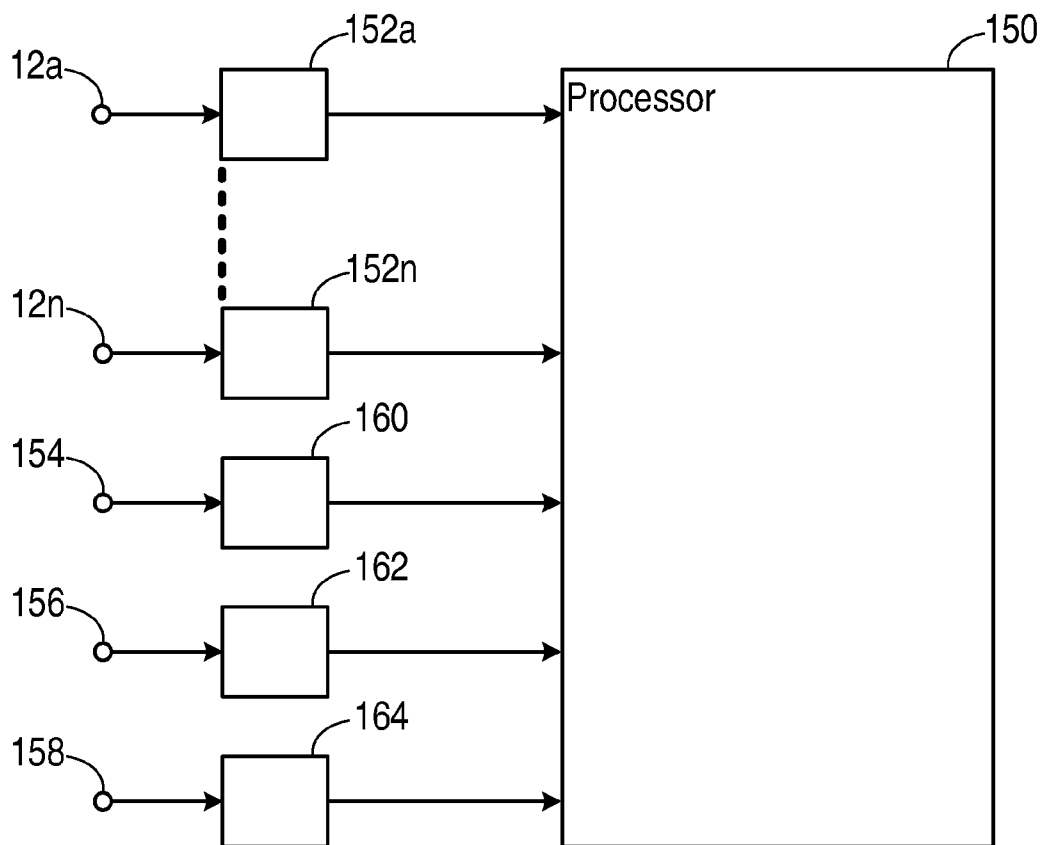


Figure 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2018/051765

A. CLASSIFICATION OF SUBJECT MATTER INV. G10L17/10 G06F21/32 H04L9/32 H04W12/00 G10L25/03 G07C9/00 H04R29/00 ADD. G10L21/0216 G10L25/78		
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 H04L H04W G07C H04R		
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	CHEN SI ET AL: "You Can Hear But You Cannot Steal: Defending Against Voice Impersonation Attacks on Smartphones", PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING SYSTEMS, IEEE COMPUTER SOCIETY, US, 5 June 2017 (2017-06-05), pages 183-195, XP033122979, ISSN: 1063-6927, DOI: 10.1109/ICDCS.2017.133 [retrieved on 2017-07-13]	1,5-8, 11, 15-18, 21,22, 31-34
A	figure 4 p. 183, left-hand column, section I. Introduction, first paragraph p. 185, section III. Problem Formulation, Section A. Adversary Model, Type 1: Voice Replay Attack section IV. The Proposed Solution p. 187, left-hand column, section A. -/--	2-4,9, 10, 12-14, 19,20, 23-30
☒ 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 31 July 2018		Date of mailing of the international search report 16/08/2018
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/GB2018/051765

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	System Architecture, second paragraph, line 2 from bottom p. 187, left-hand column, last paragraph, last line to p. 187, right-hand column, first paragraph, first line p. 187, left-hand column, second paragraph, line 4 p. 189, left-hand column, second paragraph, lines 1-5 from bottom ----- US 9 171 548 B2 (VELIUS GEORGE A [US]; WHELAN DAVID A [US]; BOEING CO [US]) 27 October 2015 (2015-10-27) figure 1 column 1, lines 9-15 column 3, lines 24-59 -----	1-34
A	CN 106 297 772 A (UNIV WUHAN) 4 January 2017 (2017-01-04) figure 3 -----	1-34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2018/051765

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 9171548	B2	27-10-2015	GB 2493849 A	20-02-2013
			JP 6106373 B2	29-03-2017
			JP 2013045104 A	04-03-2013
			US 2013043977 A1	21-02-2013

CN 106297772	A	04-01-2017	NONE	
