(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 913 489**

(51) Int. CI.:

G10L 17/16 (2013.01)
G10L 17/24 (2013.01)
G07C 9/37 (2010.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **31.10.2017 E 17199379 (3)**

(97) Fecha y número de publicación de la concesión europea: **13.04.2022 EP 3319085**

(54) Título: **Método y sistema de autenticación por biometría vocal de un usuario**

(30) Prioridad:

07.11.2016 FR 1660734

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
02.06.2022

(73) Titular/es:

PW GROUP (100.0%)
20 rue Euler
75008 Paris, FR

(72) Inventor/es:

LIBERT, GRÉGORY;
PETROVSKI CHOLLET, DIJANA y
KHEMIRI, HOUSSEMEDDINE

(74) Agente/Representante:

PONTI & PARTNERS, S.L.P.

DESCRIPCIÓN

Método y sistema de autenticación por biometría vocal de un usuario

[0001] La presente invención se refiere a un procedimiento y a un sistema de autenticación por biometría vocal de un usuario.

[0002] El reconocimiento de un hablante por un método de biometría vocal es un proceso que comienza a ser utilizado en diferentes aplicaciones.

[0003] El documento QI LI ET AL: «Recent advancements in automatic speaker authentification», IEEE ROBOTICS & AUTOMATION MAGAZINE, vol. 6, no. 1, 1 de marzo de 1999 describe un procedimiento de autenticación de usuarios. El documento US 2003/200087 A1 describe otro procedimiento de autentificación de usuarios.

[0004] En general, este tipo de procedimientos puede encontrar aplicaciones en los sistemas de control de acceso, por ejemplo, a locales o a otros destinos como el acceso a servicios bancarios, administrativos, etc.

[0005] Por lo tanto, estos procedimientos y sistemas de autentificación deben ser lo más fiables posible para contrarrestar problemas como imposturas o ataques de manera general.

[0006] En efecto, se sabe que estas imposturas o estos ataques pueden ser de naturalezas diferentes, como por ejemplo, la repetición, la transformación de la voz o incluso la síntesis de ésta.

[0007] La repetición es una forma de usurpación de identidad, mediante la cual un impostor ataca un sistema de verificación del hablante, reproduciendo una secuencia de palabras del hablante objetivo que previamente había grabado.

[0008] Se pueden distinguir dos tipos de repeticiones, a saber, las repeticiones no técnicas o repeticiones de micrófono o las repeticiones técnicas, también conocidas como repeticiones de transmisión o tratamiento.

[0009] La repetición no técnica o repetición de micrófono es una repetición que no requiere conocimientos técnicos especiales.

[0010] Su punto de realización se encuentra en el micrófono del sistema.

[0011] Este ataque consiste en reproducir delante del micrófono un archivo de audio del hablante objetivo que ha sido pregrabado anteriormente, con un dispositivo de tipo smartphone o grabador, ordenador, etc.

[0012] La repetición técnica o repetición de transmisión o tratamiento requiere, por su parte, competencias técnicas particulares.

[0013] Su punto de realización se sitúa en el nivel de la transmisión o del tratamiento de la señal audio.

[0014] En este tipo de ataques, se supone que el impostor ha podido tener acceso al canal de transmisión o de tratamiento en general, archivos de audio o de palabra, por ejemplo, por piratería, para inyectar directamente en el sistema el archivo de audio pregrabado del hablante objetivo.

[0015] La diferencia entre estos dos tipos de repetición es que en la repetición no técnica, la respuesta de impulso de los altavoces del dispositivo de repetición, así como el lugar donde se realizó el ataque, se suma a la señal de audio pregrabada por el impostor.

[0016] El objetivo de la invención es proponer mejoras a este tipo de procedimientos y sistemas de autentificación, para mejorar aún más su fiabilidad y su resistencia a los ataques.

[0017] Para este fin, la invención tiene por objeto un procedimiento de autentificación por biometría vocal de un usuario tal como se define en la reivindicación 1.

[0018] Según otro aspecto, la invención también tiene por objeto un sistema de autentificación por biometría vocal de un usuario tal como se define en la reivindicación 5.

[0019] La invención se entenderá mejor con la lectura de la descripción que se ofrece a continuación, proporcionada únicamente a modo de ejemplo y hecha en referencia a los dibujos anexos, en los que:

- la figura 1 es un organigrama que ilustra una parte de un sistema de autentificación que ilustra las zonas de repetición

no técnicas y técnicas, y
- la figura 2 es un organigrama de un ejemplo de realización de un procedimiento de visualización según la invención.

[0020] Estas figuras, y en particular la figura 1, ilustran parte de un sistema de autentificación mediante biometría vocal de un usuario.

[0021] En esta figura 1, el usuario del sistema de autentificación se designa por la referencia general 1.

[0022] Este usuario dispone de un sistema de micrófono, designado por la referencia general 2, conectado a medios de extracción de características vocales, designados por la referencia general 3.

[0023] El resto de la cadena de tratamiento no se ilustra, en la medida en que esta figura 1 se da únicamente para definir lo que, en la presente solicitud, constituye una repetición no técnica y una repetición técnica y donde pueden realizarse los ataques correspondientes.

[0024] De hecho y como se ha descrito anteriormente, una repetición no técnica o repetición de micrófono es una repetición que no requiere conocimientos técnicos especiales y su punto de realización se encuentra en el sistema de micrófono 2.

[0025] Esta repetición consiste en un ataque durante el cual se reproduce delante del micrófono un archivo de audio del hablante objetivo, que fue grabado anteriormente, con un dispositivo de tipo smartphone, tableta, etc.

[0026] La repetición no técnica es por lo tanto una repetición que se sitúa al nivel de la zona designada por la referencia general 4 en esta figura 1.

[0027] La repetición técnica o repetición de transmisión o tratamiento es, por su parte, una repetición que requiere competencias técnicas y se lleva a cabo a nivel de la transmisión o tratamiento de la señal, es decir, a partir de la zona designada por la referencia general 5 en esta figura 1.

[0028] En este ataque, se supone que el impostor pudo tener acceso al canal de transmisión o tratamiento de archivos de audio, por ejemplo, por piratería, e inyecta directamente el archivo de audio pregrabado del hablante objetivo en la cadena de transmisión o tratamiento.

[0029] Como se ha indicado anteriormente también, la invención se propone mejorar los procedimientos y sistemas de este tipo, para mejorar su resistencia a este tipo de ataques.

[0030] En la figura 2 se ilustra un procedimiento de autentificación por biometría vocal de un usuario, según la invención.

[0031] Este procedimiento incluye una fase previa de referenciación de un usuario autorizado.

[0032] Esta fase está designada por la referencia general 10 en esta figura.

[0033] Durante esta fase, un usuario designado por la referencia general 11 en esta figura pronuncia al menos una vez una frase de referencia.

[0034] Esto se realiza, por ejemplo, a través de un sistema de micrófono, designado por la referencia general 12.

[0035] Esta frase de referencia pronunciada durante esta fase previa de referenciación del usuario, se transforma a continuación en una serie de símbolos de referencia, mediante una transformación estadística común a todos los usuarios a los que se hace referencia en el sistema.

[0036] Esta transformación estadística puede ser, por ejemplo, una transformación cuyo aprendizaje se hace de manera no supervisada.

[0037] A modo de ejemplo, esta transformación estadística utiliza modelos de Markov ocultos.

[0038] Esta operación de transformación se denomina también Método MMC y está designada por la referencia general 13 en esta figura 2.

[0039] Esta transformación permite obtener una serie de caracteres designados, por ejemplo, por la referencia general 14 en esta figura.

[0040] Todos los usuarios a los que se hace referencia pasan entonces por esta fase previa de referenciación,

para constituir una base de datos de usuarios autorizados en el sistema.

[0041] El procedimiento según la invención comprende por otra parte una fase de prueba de autentificación.

[0042] Esta fase de prueba de autentificación se designa con la referencia general 15 en esta figura 2.

[0043] Durante esta fase de prueba de autentificación, un usuario candidato, designado por la referencia general 16 en esta figura, pronuncia al menos una vez, la frase de referencia.

[0044] Esto se realiza, por ejemplo, a través de medios de micrófono, designados por la referencia general 17 en esta figura.

[0045] Esta frase pronunciada durante esta fase de prueba de autentificación 15, también se transforma de la misma manera que la frase de referencia pronunciada durante la fase previa de referenciación 10, utilizando la misma transformación, en una serie de símbolos que es una serie de símbolos candidata.

[0046] En esta figura 2, la transformación está designada por la referencia general 18 y también implementa, por ejemplo, los modelos de Markov ocultos.

[0047] La serie de símbolos candidata obtenida después de la transformación se designa por la referencia general 19.

[0048] La secuencia de símbolos candidata 19, obtenida tras la transformación de la frase pronunciada por el usuario candidato durante esta fase de autentificación, se compara a continuación con la secuencia de símbolos de referencia 14.

[0049] Esta comparación se designa, por ejemplo, con la referencia general 20 en esta figura 2.

[0050] Se obtiene entonces un resultado de comparación entre las series, designado por la referencia general 21 en esta figura 2.

[0051] Este resultado de comparación 21 se compara a continuación con al menos un umbral predeterminado, para decidir si el usuario candidato que ha pronunciado la frase durante la fase de prueba 15 es efectivamente un usuario autorizado y por lo tanto autenticar este último.

[0052] Esta comparación del resultado de comparación con al menos un umbral predeterminado se designa por la referencia general 22 en esta figura 2 y la decisión se obtiene en 23.

[0053] El resultado 21 de la comparación realizada en 20 es un grado de similitud o distancia, calculado entre las dos series de símbolos.

[0054] Esta distancia es la distancia de Levenshtein. De manera general, el resultado de comparación 21 es entonces comparado con umbrales predeterminados para detectar repeticiones tal como se han descrito anteriormente.

[0055] En efecto, el resultado de comparación 21 se compara con umbrales predeterminados, como por ejemplo, dos, en el procedimiento según la invención, para detectar repeticiones por reproducción de un registro del usuario autorizado, captado sin su conocimiento, cuando el usuario pronuncia la frase de referencia o durante la transmisión o el tratamiento de la misma, en el resto del sistema.

[0056] Éstos son los ataques de repetición no técnica y técnica descritos anteriormente.

[0057] Como se ha indicado anteriormente, durante la fase previa de referenciación 10 y la fase de prueba de autentificación 15, el usuario puede pronunciar al menos una vez una frase.

[0058] En particular, se puede hacer que el usuario repita la frase de referencia al menos dos veces seguidas.

[0059] Asimismo, esto permite, por ejemplo, comparar las secuencias de símbolos sucesivos correspondientes y comparar el resultado de esta comparación con al menos un umbral predeterminado para detectar problemas de entorno sonoro, especialmente cuando estos son ruidosos.

[0060] Por supuesto pueden contemplarse otros modos de realización de este procedimiento y de este sistema.

[0061] La verificación del hablante consiste entonces en determinar si un hablante es efectivamente el que pretende ser.

[0062] El procedimiento y el sistema según la invención disponen como entrada de una muestra de palabra y de una identidad proclamada de un usuario con una referencia.

[0063] Se calcula una medida de similitud o distancia entre la muestra y la referencia del hablante correspondiente a la identidad programada.

[0064] Durante esta medición, el sistema acepta o rechaza al hablante.

[0065] En la verificación del hablante dependiente del texto, el texto pronunciado por el hablante para ser reconocido del sistema es el mismo que el que pronunció para crear su referencia.

[0066] Por lo tanto, el desafío de un sistema de verificación del hablante dependiente del texto es modelar tanto las características del hablante como el contenido léxico de la frase pronunciada.

[0067] Por lo tanto, un sistema de verificación del hablante que dependa del texto debería ser capaz de rechazar a un hablante que haya pronunciado una frase diferente a la de su referencia.

[0068] En el procedimiento y el sistema según la invención, se aplican métodos de segmentación de datos de audio con modelos estadísticos aprendidos de una manera no supervisada para la verificación del hablante dependiente del texto.

[0069] Con estos métodos, los datos de audio se convierten en una cadena de símbolos. Por lo tanto, los datos de audio de referencia y de prueba se pueden comparar y se puede medir un grado de similitud o distancia entre ellos.

[0070] Para medir la distancia o similitud entre dos archivos de audio convertidos en secuencias de símbolos, se utiliza preferentemente la distancia de Levenshtein.

[0071] Al establecer un umbral, se puede aceptar o rechazar al hablante y detectar que la frase pronunciada es efectivamente la frase de referencia.

[0072] No hace falta decir que, por supuesto, se pueden considerar otros modos de realización.

REIVINDICACIONES

1. Procedimiento de autentificación por biometría vocal de un usuario, que comprende una fase previa de referenciación (10) de un usuario autorizado, en la que dicho usuario pronuncia al menos una vez una frase de referencia, **caracterizada porque** dicha frase se transforma en una serie de símbolos de referencia (14) mediante una transformación estadística (13) común a todos los usuarios a los que se hace referencia, y una fase de prueba de autentificación (15), que comprende una primera etapa en la que un usuario candidato pronuncia al menos una vez la frase de referencia y esta frase pronunciada se transforma de la misma manera que la frase de referencia en la fase previa, utilizando la misma transformación (18), en una serie de símbolos candidata (19), y una segunda etapa durante la cual la serie de símbolos candidata (19) se compara (en 20) con la serie de símbolos de referencia (14) para determinar un resultado de comparación (21) y este resultado (21) se compara (en 22) con al menos un umbral predeterminado, para decidir (en 23) si el usuario candidato que pronunció la frase durante la fase de prueba es efectivamente un usuario autorizado y, por lo tanto, autentificarlo, **en la medida en que** el resultado de comparación (21) es una distancia calculada entre las dos series de símbolos, siendo la distancia calculada (21) la distancia de Levenshtein, y **en la medida en que** el resultado de comparación (21) se compara con dos umbrales predeterminados, para detectar repeticiones mediante la reproducción de un registro del usuario autorizado capturado sin su conocimiento, durante la transmisión o el procesamiento de la frase de referencia.

2. Procedimiento de autentificación por biometría vocal de un usuario, según la reivindicación 1, **caracterizado porque** la transformación estadística (13, 18) es una transformación cuyo aprendizaje se realiza de manera no supervisada.

3. Procedimiento de autentificación por biometría vocal de un usuario según la reivindicación 1 o 2, **caracterizado porque** la transformación estadística (13, 18) utiliza modelos de Markov ocultos.

4. Procedimiento de autentificación por biometría vocal de un usuario, según cualquiera de las reivindicaciones anteriores, **caracterizado por que** el usuario es llevado a repetir la frase de referencia al menos dos veces seguidas y se compara el resultado de la comparación de las sucesivas series de símbolos correspondientes, con al menos un umbral predeterminado para detectar problemas de entorno sonoro.

5. Sistema de autentificación por biometría vocal de un usuario, para la aplicación de un procedimiento según cualquiera de las reivindicaciones anteriores, que comprende medios de referenciación previa de un usuario autorizado, en los que este usuario pronuncia al menos una vez una frase de referencia, **caracterizada porque** esta frase se transforma en una serie de símbolos de referencia mediante una transformación estadística común a todos los usuarios a los que se hace referencia, y medios de prueba de autenticación, que comprenden los primeros medios en los que un usuario candidato pronuncia al menos una vez la frase de referencia y esta frase pronunciada se transforma de la misma manera que la frase de referencia en la fase previa, utilizando la misma transformación, en una serie de símbolos candidata, y los segundos medios por los que la serie de símbolos candidata se compara a continuación con la serie de símbolos de referencia para determinar un resultado de comparación y este resultado se compara con al menos un umbral predeterminado, para decidir si el usuario candidato que ha pronunciado la frase de prueba es efectivamente un usuario autorizado y, por lo tanto, autentificarlo, el resultado de la comparación (21) siendo una distancia calculada entre las dos series de símbolos, siendo la distancia calculada (21) la distancia de Levenshtein, comparándose el resultado de la comparación (21) con dos umbrales predeterminados, con el fin de detectar las repeticiones por reproducción de una grabación del usuario autorizado captada sin su conocimiento, durante la transmisión o el tratamiento de la frase de referencia.

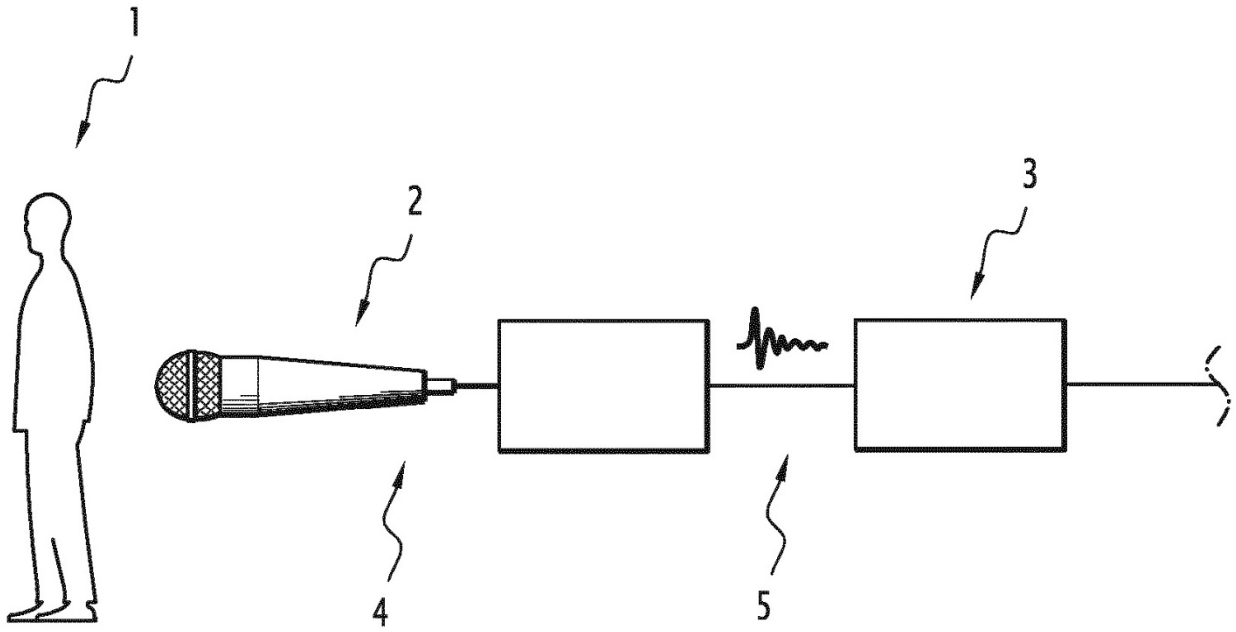

FIG.1

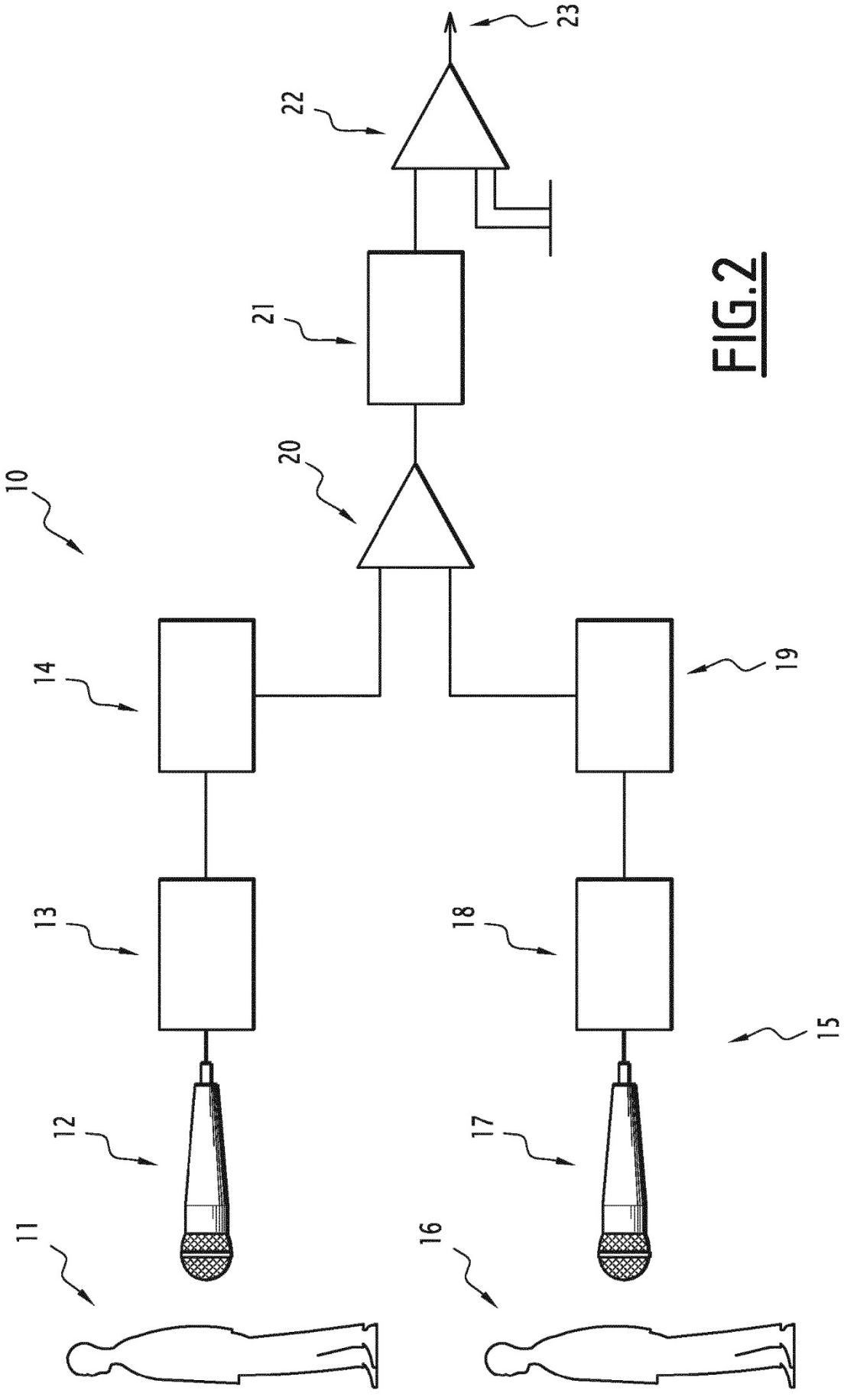

FIG.2