

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 800 348**

51 Int. Cl.:

G10L 17/06 (2013.01)

G10L 25/78 (2013.01)

G10L 17/02 (2013.01)

G10L 25/18 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.06.2017 PCT/CN2017/088073**

87 Fecha y número de publicación internacional: **20.12.2018 WO18227381**

96 Fecha de presentación y número de la solicitud europea: **13.06.2017 E 17829582 (0)**

97 Fecha y número de publicación de la concesión europea: **20.05.2020 EP 3433854**

54 Título: **Método y sistema para verificación de orador**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.12.2020

73 Titular/es:
**BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD. (100.0%)
Building 34, No. 8 Dongbeiwang West Road,
Haidian District
Beijing 100193, CN**

72 Inventor/es:
**CHEN, JIE;
SU, DAN;
FU, TIANXIAO y
HU, NA**

74 Agente/Representante:
UNGRÍA LÓPEZ, Javier

ES 2 800 348 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Método y sistema para verificación de orador

5 Campo técnico

La presente divulgación se refiere en general a tecnología de verificación de orador, y más particularmente, a métodos, sistemas y aparatos para la verificación de orador fiable y eficiente.

10 Antecedentes

La tecnología de verificación de orador implica la verificación de un orador en función de las características de su voz o huellas de voz. Dichas características pueden estar relacionadas con las características anatómicas del orador, como el tamaño y la forma de la laringe, la boca y la nariz, y patrones de comportamiento, tal como el tono de voz y el lenguaje. Las tecnologías de verificación de orador generalmente se dividen en dos categorías: verificación del orador dependiente del texto y verificación independiente del texto. Se han desarrollado y aplicado varias técnicas de verificación de orador para verificar la identificación de un orador, particularmente como mecanismo de seguridad.

Muchos servicios y productos ahora se ofrecen y procesan a través de la comunicación remota entre un terminal de usuario y un servidor de proveedor de servicios. Por ejemplo, cada vez es más popular para los pasajeros reserven un servicio de taxi a través de una aplicación en sus terminales móviles. A continuación, la información puede transmitirse a través de Internet a un sistema de gestión de servicio de taxis, que puede procesar la información del viaje, localizar un conductor disponible registrado en el sistema y organizar una recogida en consecuencia. Sin embargo, hay situaciones en las que usuarios no autorizados utilizan incorrectamente cuentas de conductores registrados, lo que causa problemas de seguridad y problemas en la gestión de registros de conductores. Por ejemplo, un conductor registrado puede compartir su cuenta con otros conductores sin autorización, o una cuenta de conductor o el terminal móvil del conductor pueden ser pirateados o robados. Tales usos no autorizados de una cuenta de conductor no solo causan pérdida de ingresos del proveedor, sino que también plantean preocupaciones de seguridad sobre la incapacidad de rastrear la identidad real de los conductores. Por lo tanto, se necesitan sistemas fiables para verificar la identificación del conductor para mejorar la seguridad de los servicios de taxi y una gestión eficiente del registro del conductor.

En vista de los problemas anteriores, esta divulgación proporciona métodos, dispositivos y sistemas de verificación de orador para mejorar la seguridad de aplicaciones tales como el servicio de llamadas de taxis y la gestión eficiente de las cuentas de usuario registradas.

El documento EP 0424071 divulga un sistema de reconocimiento de oradores para la verificación de orador para verificar una identidad reivindicada previa. El documento US 2017/084295 divulga una tecnología basada en aprendizaje automático que analiza una entrada de audio y proporciona predicciones de estado de oradores en respuesta a la entrada de audio.

Sumario

Un aspecto de la presente divulgación está dirigido a un método de verificación de orador. El método puede incluir: adquirir una grabación de audio; extraer señales de voz de la grabación de audio; extraer características de las señales de voz extraídas; y determinar si las señales de voz extraídas representan la voz de un orador predeterminado basándose en las características extraídas y un modelo de orador entrenado con datos de voz de referencia del orador predeterminado.

Otro aspecto de la presente divulgación está dirigido a un sistema de verificación de orador. El sistema puede incluir una memoria que incluye instrucciones y un procesador configurado para ejecutar las instrucciones para: recibir una grabación de audio; extraer señales de voz de la grabación de audio; extraer características de las señales de voz extraídas; y determinar si las señales de voz extraídas representan la voz de un orador predeterminado basándose en las características extraídas y un modelo de orador entrenado con datos de voz de referencia del orador predeterminado.

Otro aspecto de la presente divulgación está dirigido a un medio de almacenamiento no transitorio legible por ordenador que almacena instrucciones que, cuando son ejecutadas por uno o más procesadores, hacen que los procesadores realicen un método de verificación de orador. El método incluye: adquirir una grabación de audio; extraer señales de voz de la grabación de audio; extraer características de las señales de voz extraídas; y determinar si las señales de voz extraídas representan la voz de un orador predeterminado basándose en las características extraídas y un modelo de orador entrenado con datos de voz de referencia del orador predeterminado.

Otro aspecto de la presente divulgación está dirigido a un método de verificación de orador. El método puede incluir: adquirir una pluralidad de grabaciones de audio desde un terminal; extraer señales de voz de la pluralidad de

grabaciones de audio; extraer características de las señales de voz extraídas; clasificar las características extraídas en una o más clases; y cuando las características extraídas se clasifican en más de una clase, determinar que la pluralidad de grabaciones de audio incluye voces de uno o más oradores diferentes de un orador predeterminado.

- 5 Otro aspecto de la presente divulgación está dirigido a un sistema de verificación de orador. El sistema puede incluir una memoria que incluye instrucciones y un procesador configurado para ejecutar las instrucciones para: recibir una pluralidad de grabaciones de audio desde un terminal; extraer señales de voz de la pluralidad de grabaciones de audio; extraer características de las señales de voz extraídas; clasificar las características extraídas en una o más clases; y cuando las características extraídas se clasifican en más de una clase, determinar que la pluralidad de grabaciones de audio incluye voces de uno o más oradores diferentes de un orador predeterminado.

- 15 Otro aspecto adicional de la presente divulgación está dirigido a un medio de almacenamiento no transitorio legible por ordenador que almacena instrucciones que, cuando son ejecutadas por uno o más procesadores, hacen que los procesadores realicen un método de verificación de orador. El método incluye: adquirir una pluralidad de grabaciones de audio desde un terminal; extraer señales de voz de la pluralidad de grabaciones de audio; extraer características de las señales de voz extraídas; clasificar las características extraídas en una o más clases; y cuando las características extraídas se clasifican en más de una clase, determinar que la pluralidad de grabaciones de audio incluye voces de uno o más oradores diferentes de un orador predeterminado.

20 Breve descripción de los dibujos

- La figura 1 es un diagrama de bloques de un sistema de verificación de orador 100 utilizado en el área de gestión de servicio de taxis de acuerdo con una realización ejemplar.
- 25 La figura 2 es un diagrama de bloques de un verificador de orador 200, de acuerdo con una realización ejemplar aplicada en el área de gestión del servicio de taxis.
- La figura 3 es un diagrama de flujo que ilustra un método de preprocesamiento de señal de audio 300, de acuerdo con una realización ejemplar.
- La figura 4 es un diagrama de flujo que ilustra un método de preprocesamiento de señal de audio 400, de acuerdo con otra realización ejemplar.
- 30 La figura 5 es un diagrama de flujo que ilustra un proceso de verificación de orador 500, de acuerdo con una realización ejemplar.
- La figura 6 es un diagrama de flujo que ilustra un proceso de inscripción o registro de orador 600, de acuerdo con una realización ejemplar.
- 35 La figura 7 es un diagrama de flujo que ilustra un método de verificación de orador 700 en el que no se ha almacenado ningún modelo de orador asociado con una cuenta de conductor registrada, de acuerdo con una realización ejemplar.

Descripción detallada

- 40 Esta divulgación generalmente se refiere a la verificación de un orador en función de las características de voz extraídas de los datos de voz del orador. Los datos de voz se pueden obtener a partir de la entrada de grabación de audio, incluidas las señales de voz del orador, tal como la grabación de voz de un orador que lee ciertos textos, o una grabación de una conversación entre el orador y otra persona. Se puede obtener en tiempo real, o mediante la entrada de grabaciones almacenadas previamente. Además, se pueden usar diferentes terminales de usuario para recibir una entrada de grabación de audio, por ejemplo, PDA, tabletas, ordenadores tales como ordenadores portátiles, teléfonos móviles, relojes inteligentes, etc.

- En relación con el servicio de transporte, por ejemplo, un proveedor de servicios de transporte (por ejemplo, un proveedor de transporte en taxi, transporte privado en automóvil, transporte compartido, etc.) puede usar los datos de voz para verificar la identidad de un conductor o pasajero. Específicamente, los pasajeros que necesitan servicio de transporte, por ejemplo, servicio de taxi, pueden tener una aplicación de servicio de taxi instalada en sus terminales de usuario, tal como sus teléfonos móviles. Los conductores que brindan servicio de taxi también pueden tener las aplicaciones correspondientes instaladas en sus terminales, para recibir pedidos y/o recibir instrucciones del proveedor de servicios de transporte para procesar un pedido y recoger a un pasajero determinado. En un escenario típico, un pasajero puede ingresar información sobre un próximo viaje a través de una aplicación de teléfono móvil del lado del conductor y solicitar el servicio de taxi. El proveedor de servicios de transporte puede recibir dicha solicitud de servicio a través de una red, procesar el pedido en consecuencia y enviar un conductor disponible para organizar la recogida. Antes de la recogida del pasajero, un conductor puede contactar al pasajero de acuerdo con la información de contacto del pasajero, tal como un número de teléfono u otra información de cuenta asociada con el pasajero. La comunicación puede implicar una conversación directa entre el terminal del conductor y el terminal del pasajero. Para facilitar la gestión y garantizar la seguridad del servicio de taxis, la conversación puede grabarse y la grabación de audio puede proporcionarse a un sistema de verificación para su posterior procesamiento y verificación del orador, por ejemplo, para verificar la identidad del conductor de acuerdo con la información del conductor registrado.

- 65 La figura 1 es un diagrama de bloques de un sistema de verificación de orador 100 utilizado en la gestión del servicio

de taxis de acuerdo con una realización ejemplar. Con referencia a la figura 1, el sistema 100 puede incluir un verificador de orador 110, una red 120, una pluralidad de terminales de pasajeros 130 y una pluralidad de terminales de conductor 140, un sistema de gestión de servicio de taxis 150 y un panel de control 160.

5 El verificador de orador 110 puede configurarse para verificar un orador de acuerdo con las características de voz de un orador. Como se muestra en la figura 1, el verificador de orador 110 puede incluir, entre otras cosas, una memoria 111, un procesador 113, un almacenamiento 115, una interfaz de entrada/salida (E/S) 117, y una interfaz de comunicación 119. Al menos algunos de estos componentes del verificador de orador 110 pueden configurarse para transferir datos y enviar o recibir instrucciones entre los mismos. En algunas realizaciones, el verificador de
10 orador 110 puede instalarse como parte de una plataforma de gestión de servicios de taxis operada por el proveedor de servicios de transporte, o sus funciones también pueden incorporarse como parte de una aplicación de servicio de taxis (por ejemplo, una aplicación instalada en la terminal de pasajero 130), permitiendo que un usuario verifique la identidad del conductor que proporciona el servicio.

15 El procesador 113 puede incluir cualquier tipo apropiado de microprocesador de propósito general o especial, procesador de señal digital o microcontrolador. El procesador 113 puede configurarse como un módulo de procesador separado dedicado a verificar un orador y emitir el resultado de la verificación. Alternativamente, el procesador 113 puede configurarse como un módulo de procesador compartido para realizar otras funciones no relacionadas con la verificación del orador.

20 El procesador 113 puede configurarse para recibir datos y/o señales de otros componentes del sistema 110 para realizar la verificación del orador. Por ejemplo, el procesador 113 puede intercambiar información con el sistema de gestión de servicio de taxis 150 e instrucciones desde el panel de control 160 a través de, por ejemplo, la interfaz de E/S 117. Después de la verificación, el procesador 113 puede proporcionar el resultado de la verificación al sistema
25 de gestión del servicio de taxis 150 y al panel de control 160, que puede proporcionar instrucciones adicionales tales como almacenar los resultados de la verificación, actualizar el registro de la cuenta del conductor/usuario o enviar el resultado de la verificación a un correspondiente terminal de pasajero 130 o un terminal de conductor 140 correspondiente.

30 El procesador 113 también puede acceder a la información transmitida a través de la interfaz de comunicación 119. Por ejemplo, el procesador 113 puede recibir una grabación de audio de una conversación que incluye señales de audio desde un terminal de pasajero 130 y un terminal de conductor 140, transmitidos a través de la red 120. Después de la verificación, el procesador 113 puede, al recibir instrucciones del panel de control 160, transmitir el resultado de la verificación al terminal de pasajero 130 correspondiente y al terminal de conductor 140
35 correspondiente a través de la red 120.

El procesador 113 puede ejecutar instrucciones informáticas (por ejemplo, códigos de programa) almacenadas en la memoria 111 y/o el almacenamiento 115, y puede realizar funciones de acuerdo con técnicas ejemplares descritas en esta divulgación. Más funciones ejemplares del procesador 113 se describirán más adelante en relación con las
40 figuras 2-7.

La memoria 111 y el almacenamiento 115 pueden incluir cualquier tipo apropiado de almacenamiento masivo proporcionado para almacenar cualquier tipo de información que el procesador 113 pueda necesitar para operar. La memoria 111 y el almacenamiento 115 pueden manifestarse en un dispositivo de almacenamiento volátil o no volátil,
45 magnético, semiconductor, de cinta, óptico, extraíble, no extraíble u otro tipo de medio tangible (es decir, no transitorio) legible por ordenador, incluyendo, pero no limitado a, una memoria de solo lectura (ROM), una memoria flash, una memoria dinámica de acceso aleatorio (RAM) y una RAM estática. La memoria 111 y/o el almacenamiento 115 pueden configurarse para almacenar uno o más programas informáticos que pueden ser ejecutados por el procesador 113 para realizar los métodos descritos en el presente documento para la verificación del orador.

50 La memoria 111 y/o el almacenamiento 115 pueden configurarse adicionalmente para almacenar información y datos utilizados por el procesador 113. Por ejemplo, la memoria 111 y/o el almacenamiento 115 pueden almacenar las grabaciones de audio recibidas desde los terminales de pasajero 130 y/o los terminales de conductor 140, los datos generados durante el proceso de verificación del orador y el resultado final de la verificación.

55 La interfaz de E/S 117 puede configurarse para facilitar la comunicación entre el verificador de orador 110, el sistema de gestión de servicio de taxis 150 y el panel de control 160. Por ejemplo, el verificador de orador 110 puede proporcionar el resultado de verificación del orador al sistema de gestión del servicio de taxis 150 a través de la interfaz de E/S 117, para actualizar la información de la cuenta del conductor y el registro del servicio de taxis. El verificador de orador 110 también puede recibir instrucciones del panel de control 160 con respecto a la realización
60 de la verificación del orador, o para enviar el resultado de la verificación al terminal de pasajero 130 y al terminal de conductor 140 correspondientes.

65 La interfaz de comunicación 119 puede configurarse para comunicarse con los terminales de pasajero 130 y los terminales de conductor 140 a través de la red 120. La red 120 puede ser cualquier tipo de red cableada o inalámbrica que permita la transmisión y la recepción de datos. Por ejemplo, la red 120 puede ser una red cableada,

una red inalámbrica local (por ejemplo, Bluetooth™, WiFi, comunicaciones de campo cercano (NFC), etc.), una red celular, Internet o similar, o una combinación de los mismos. También se contemplan otros métodos de comunicación conocidos que proporcionan un medio para transmitir datos.

5 El terminal de pasajero 130 puede ser cualquier dispositivo capaz de recibir y transmitir información a través de la red 120, por ejemplo, un teléfono móvil, un ordenador tal como un ordenador portátil o de escritorio, PDA, tabletas, relojes inteligentes, etc., que llevan una aplicación de servicio de taxis. Se puede configurar para recibir y grabar señales de audio y transmitir información a través de la red 120. Por ejemplo, un pasajero puede ingresar información a través de una aplicación en un teléfono móvil, iniciar o participar en una conversación con un terminal de conductor 140. El teléfono móvil del pasajero funciona como un terminal de pasajero 130 que recibe una grabación de audio de la conversación, que luego puede transmitirse al verificador de orador 110 para la verificación del orador.

15 El terminal de conductor 140 puede ser cualquier dispositivo capaz de recibir y transmitir información a través de la red 120, similar al terminal de pasajero 130. Por ejemplo, el terminal de conductor 140 puede ser un teléfono móvil, un ordenador tal como un ordenador portátil o de escritorio, PDA, tabletas, relojes inteligentes, etc., que llevan una aplicación de servicio de taxis. También puede ser un dispositivo o módulo instalado como parte de un sistema de control del vehículo. Por ejemplo, un conductor puede ingresar información a través de una aplicación en un teléfono móvil, iniciar o participar en una conversación con un terminal de pasajero 130. El teléfono móvil del conductor funciona como un terminal de conductor 140 que recibe una grabación de audio de la conversación, que luego puede transmitirse al verificador de orador 110 para la verificación del orador. Además, durante el registro de la cuenta del conductor, el terminal de conductor 140 puede configurarse para recibir el registro del conductor o la información de configuración de la cuenta, tal como número de teléfono, sexo, cuenta de correo electrónico y grabaciones de voz. El terminal de conductor 140 puede transmitir la información de registro del conductor a través de la red 120 al verificador de orador 110.

30 El terminal de pasajero 130 y/o el terminal de conductor 140 pueden incluir además una pantalla que incluye una pantalla LCD, un LED, una pantalla de plasma o cualquier otro tipo de pantalla, y proporcionar una interfaz gráfica de usuario (GUI) presentada en la pantalla para la entrada del usuario y mostrar los datos. Por ejemplo, la información de la cuenta del usuario y del conductor se puede mostrar a través de la pantalla, y también se puede mostrar el resultado de la verificación del orador.

35 El sistema de gestión del servicio de taxis 150 puede incluir además una pluralidad de componentes para gestionar el registro de cuentas de conductor y la prestación del servicio de taxis. Puede comunicarse con el verificador de orador 110 a través de la interfaz de E/S 117, para transmitir información de la cuenta del conductor y recibir el resultado de verificación desde el verificador de orador 110, etc. Además puede incluir otros componentes, tal como una base de datos que almacena el registro del conductor asociado con los resultados de verificación del orador, un procesador para evaluar el rendimiento del conductor basado en el registro del conductor, y el procesador puede configurarse además para enviar un conductor disponible enviando una notificación de pedido al terminal de conductor 140 correspondiente.

45 Aunque se muestra como módulos separados en la figura 1, se contempla que el verificador de orador 110, el sistema de gestión de servicio de taxis 150 y el panel de control 160 también puedan integrarse como un solo sistema. Por ejemplo, en una realización, el verificador de orador 110 y el panel de control 160 pueden integrarse en el sistema de gestión de servicio de taxis 150, que es operado por el proveedor de servicio de transporte. Otros componentes del sistema 100 y sus funciones que no están directamente relacionadas con la verificación del orador no se describen en detalle en el presente documento.

50 La figura 2 es un diagrama de bloques de un verificador de orador 200, de acuerdo con una realización ejemplar aplicada en el área de gestión del servicio de taxis. Con referencia a la figura 2, el verificador de orador 200 puede incluir una sección de registro del conductor 210 y una sección de verificación del conductor 220. La sección de registro del conductor 210 puede incluir además un módulo de preprocesamiento 211, un módulo de extracción de características 213, un módulo de construcción de modelo de orador 215 y una base de datos de modelo de orador 217. La sección de verificación del conductor 220 puede incluir además un módulo de preprocesamiento 221, un módulo de extracción de características 223 y un módulo de verificación de orador 225. Los módulos de preprocesamiento 211 y 221 pueden estar incorporados en el mismo módulo funcional, realizando funciones de preprocesamiento tanto en un proceso de registro de conductor como en un proceso de verificación de conductor. De manera similar, los módulos de extracción de características 213 y 223 también pueden estar incorporados en el mismo módulo funcional, realizando funciones de extracción de características, tanto para el proceso de registro del conductor como para el proceso de verificación del conductor.

65 Durante el registro del conductor, el módulo de preprocesamiento 211 está configurado para preprocesar las grabaciones de audio del conductor recibidas transmitidas desde el terminal de conductor 140. Las etapas de preprocesamiento incluyen la extracción de señales de voz, la eliminación del ruido ambiental o las señales de ruido de fondo, y la extracción de datos de voz del conductor, para su posterior procesamiento mediante el módulo de extracción de características de voz 213. El preprocesamiento de las grabaciones de audio puede incluir además

seleccionar los datos de voz correspondientes a un orador consistente con el género asociado con el conductor, si hay señales de voz correspondientes a más de un género. Detalles de las etapas de preprocesamiento se describirán más adelante con referencia a las figuras 3 y 4.

5 El módulo de extracción de características 213 está configurado para extraer características de voz de los datos de voz preprocesados. Se pueden usar varias técnicas de extracción de características existentes utilizadas en tecnologías de verificación de voz o del orador. Por ejemplo, el módulo de extracción de características 213 puede configurarse para extraer Coeficientes Cepstrales de Frecuencia de Mel (MFCC), que fue introducido por Davis y Mermelstein en la década de 1980 y ampliamente utilizado en reconocimiento de voz, otros pueden incluir
10 Coeficientes de Predicción Lineal (LPC) o Coeficientes Cepstrales de Predicción Lineal (LPCC).

El módulo de construcción del modelo de orador 215 puede configurarse para utilizar las características extraídas por el módulo de extracción de características 213, y construir un modelo acústico de orador que represente las características de voz del conductor que se está registrando. El módulo de construcción de modelos de oradores 215
15 puede usar varias técnicas de construcción de modelos de oradores. Por ejemplo, redes neuronales profundas (DNN) se pueden entrenar utilizando las características extraídas de los datos de voz del conductor, y se puede construir un modelo de orador extrayendo un vector de características que represente las características de voz del orador. Una DNN es una red neuronal artificial (ANN) con múltiples capas ocultas entre las capas de entrada y salida. Otros modelos de oradores que pueden utilizarse incluyen el Modelo de mezcla gaussiana (GMM) o un
20 modelo de Modelo oculto de Markov (HMM). También hay combinaciones de diferentes enfoques para entrenar modelos de oradores. Esta divulgación actual no se limita a ningún enfoque de capacitación de modelo particular.

Después de registrarse en el sistema de servicio de taxis, un conductor tendrá un modelo de orador capacitado con sus características de voz únicas. Todos los modelos de oradores pueden almacenarse en la base de datos de
25 modelos de oradores 217, lo que significa que cada conductor registrado tendrá un modelo de orador asociado con su cuenta. Los modelos de oradores se utilizarán en el proceso de verificación del conductor.

La sección de verificación del conductor 220 incluye el módulo de preprocesamiento 221, el módulo de extracción de características 223 y el módulo de verificación de orador 225, que verifica un orador basado en los modelos de
30 oradores guardados en la base de datos de modelos de oradores 217.

El módulo de preprocesamiento 221 puede configurarse para preprocesar una entrada de grabación de audio, que puede tener la forma de una grabación de audio de una conversación entre un conductor y un pasajero. Similar a las etapas de preprocesamiento realizadas por el módulo de preprocesamiento 211 durante el registro del conductor, el preprocesamiento durante la verificación puede incluir la selección de señales transmitidas desde el terminal de
35 conductor 140, la extracción de señales de voz, la eliminación de ruido ambiental o señales de ruido de fondo, y la extracción de datos de voz del conductor, para su posterior procesamiento mediante el módulo de extracción de características de voz 223. El preprocesamiento de las grabaciones de audio puede incluir además seleccionar los datos de voz correspondientes a un orador consistente con el género asociado con el conductor registrado correspondiente, si están presentes señales de voz correspondientes a más de un género. Detalles del preprocesamiento se describirán más adelante con referencia a las figuras 3 y 4.
40

El módulo de extracción de características 223 está configurado para extraer funciones de voz de los datos de voz del orador preprocesados. El módulo de extracción de características 223 puede usar varias técnicas de extracción de características existentes usadas en tecnologías de reconocimiento de voz o de orador. Por ejemplo, el módulo de extracción de características 223 puede configurarse para extraer características de Coeficientes Cepstrales de Frecuencia de Mel (MFCC), que fue introducido por Davis y Mermelstein en la década de 1980 y ampliamente utilizado en reconocimiento de voz, otros pueden incluir Coeficientes de Predicción Lineal (LPC) o Coeficientes Cepstrales de Predicción Lineal (LPCC).
45
50

El módulo de verificación de orador 225 puede configurarse para verificar si el orador es el conductor registrado, en base a las características de voz de orador extraídas y al modelo de orador asociado con el conductor registrado. La verificación puede incluir verificar las características extraídas con el modelo de orador correspondiente al conductor registrado y calcular una puntuación de similitud. La puntuación de similitud se comparará con un umbral de similitud predeterminado. Detalles del proceso de verificación se describen a continuación con referencia a la figura 5.
55

La figura 3 es un diagrama de flujo que ilustra un método de preprocesamiento de señal de audio 300, de acuerdo con una realización ejemplar. El método 300 puede realizarse mediante un módulo de preprocesamiento de un dispositivo de verificación de orador, tal como el módulo de preprocesamiento 211 y/o 221 que se muestra en la figura 2. Como ejemplo, el método 300 ilustrado en el presente documento está relacionado con el preprocesamiento durante la verificación del orador/conductor utilizado en el contexto del servicio de taxis. Como se señaló anteriormente, se pueden realizar etapas de preprocesamiento similares durante el proceso de registro de usuario/conductor.
60

65 Etapa 301, el módulo de preprocesamiento 221 recibe una entrada de grabación de audio, que incluye una conversación entre un pasajero y un conductor. La grabación de audio puede transmitirse desde uno o ambos del

terminal de pasajero 130 y el terminal de conductor 140 a través de una red, como se muestra en la figura 1.

Etapa 303, el módulo de preprocesamiento 221 selecciona señales de audio transmitidas desde un terminal predeterminado. Esto puede realizarse seleccionando las señales de audio transmitidas desde el terminal de conductor 140 asociado con la cuenta que se está utilizando. Cuando se utiliza una cuenta registrada, el módulo de preprocesamiento 221 puede verificar la información del terminal asociada con la cuenta, y las señales de audio transmitidas desde el terminal correspondiente pueden seleccionarse en consecuencia. Por ejemplo, esto también puede realizarse basándose en la información del canal de las señales de audio transmitidas desde el terminal de conductor 140 y el terminal de pasajero 130.

Etapa 305, el módulo de preprocesamiento 221 elimina el ruido ambiental de las señales de audio extraídas, o extrayendo señales de voz. Las señales de audio seleccionadas desde el terminal de conductor 140 predeterminado pueden incluir diversos ruidos de fondo o de entorno, tales como sonidos de otros reproductores multimedia, sonidos de otros terminales cercanos, ruido generado por el funcionamiento del vehículo y ruido del entorno fuera del vehículo. Este ruido ambiental puede eliminarse, y las señales de voz pueden extraerse. Esto puede realizarse mediante diversas técnicas existentes, tal como clasificar diferentes señales de sonido de acuerdo con sus frecuencias y eliminar las señales correspondientes a las frecuencias que no son de voz. En algunas realizaciones, esto también puede realizarse usando el espectro de imagen de frecuencia de tiempo de resonancia (RTFI), produciendo un espectro de energía de frecuencia de tiempo y seleccionando las señales de voz en función de sus niveles de energía. Por ejemplo, el módulo de preprocesamiento 221 puede realizar análisis RTFI en las señales de audio extraídas para obtener un espectro de energía de frecuencia de tiempo, que luego se promedia en un dominio de tiempo para obtener un espectro de energía promedio. El módulo de preprocesamiento 221 puede entonces usar métodos de segmentación basados en el cambio de energía para determinar los puntos de inicio y finalización de las señales de voz y las señales de no voz. Los componentes de señal que no son de voz y los componentes de ruido de fondo de la señal de audio de entrada pueden eliminarse, en función de sus niveles de energía diferentes de los de las señales de voz.

Etapa 307, el módulo de preprocesamiento 221 determina si las señales de voz incluyen discursos de múltiples oradores de diferentes géneros. Por ejemplo, las señales de voz transmitidas desde un terminal 140 del conductor pueden incluir la voz del conductor, y también voz de otra persona cercana, tal como un pasajero en el vehículo que puede ser de un género diferente. Las señales de voz de oradores de diferentes géneros pueden clasificarse en función de la diferencia en las frecuencias fundamentales características de diferentes géneros. La frecuencia fundamental representa la frecuencia de la vibración de las cuerdas vocales y se correlaciona con los cambios en la tensión de las cuerdas vocales, el tamaño y el espesor de las cuerdas vocales y la presión de aire subglótica. La frecuencia fundamental varía con factores como el género y la edad. Se correlaciona con el tono y muestra qué tan alta o baja suena la voz de una persona. Por ejemplo, un hombre adulto típico tiene una frecuencia fundamental de 85 a 180 Hz, mientras que una mujer adulta típica tiene una frecuencia fundamental de 165 a 255 Hz. Se han desarrollado muchos algoritmos para estimar la frecuencia fundamental de una señal de audio, como el algoritmo YIN ampliamente utilizado, y las modificaciones de YIN, como el algoritmo YIN Probabilístico (PYIN).

Etapa 309, si se determina que las señales de voz incluyen voces de múltiples oradores de diferentes géneros, el módulo de preprocesamiento 221 extrae datos de voz asociados con un género predeterminado asociado con la cuenta del conductor. Se pueden eliminar las señales de voz de un orador de un género diferente al asociado con la cuenta del conductor. Por ejemplo, si el género asociado con la cuenta del conductor es masculino, y las señales de voz extraídas incluyen señales de voz de una pasajera, que pueden determinarse en función de las frecuencias fundamentales características correspondientes a una mujer, las señales de voz de la pasajera pueden eliminarse.

Etapa 311, el módulo de preprocesamiento 221 transmite los datos de voz extraídos del orador al módulo de extracción de características 223 para su posterior procesamiento.

La figura 4 es un diagrama de flujo que ilustra un método de preprocesamiento de señal de audio 400, de acuerdo con otra realización ejemplar. De manera similar al método 300, el método 400 puede estar relacionado con el preprocesamiento durante el registro del orador/conductor o los procesos de verificación utilizados en el contexto del servicio de taxis. Por ejemplo, el método 400 puede realizarse mediante un módulo de preprocesamiento de un dispositivo de verificación de orador, tal como el módulo de preprocesamiento 211 y/o 221 que se muestra en la figura 2. Solo con fines ilustrativos, el método 400 se describirá como realizado por el módulo de preprocesamiento 221 en un proceso de verificación de orador.

Con referencia a la figura 4, las etapas 401-405 son similares a las etapas 301-305, cuyos detalles no se repiten en el presente documento. A continuación, en la etapa 407, el módulo de preprocesamiento 221 determina los géneros del orador de las señales de voz extraídas. Por ejemplo, como se describió anteriormente en relación con la etapa 307, el módulo de preprocesamiento 221 puede determinar los géneros de los oradores en función de las frecuencias fundamentales de las señales de voz. Cuando se determina que las señales de voz representan (es decir, la grabación de audio incluye) discursos de oradores cuyos géneros son diferentes del género de un usuario predeterminado, tal como el conductor de un vehículo, el módulo de preprocesamiento 221 continúa con la etapa 409 y selecciona las señales de voz cuyos géneros de orador son diferentes del género del usuario predeterminado.

En la etapa 411, el módulo de preprocesamiento 221 calcula una relación de la duración de tiempo de una señal de voz seleccionada a lo largo de la duración de la grabación de audio, para determinar si la voz representada por la señal de voz seleccionada representa una porción significativa de la grabación de audio. Si la relación excede un umbral predeterminado, el módulo de preprocesamiento 221 concluye directamente que el orador en la grabación de audio no es el usuario predeterminado (etapa 413). Sin embargo, si la relación es igual o inferior al umbral predeterminado, el módulo de preprocesamiento 221 elimina la señal de voz seleccionada de las señales de voz extraídas (etapa 415), y luego vuelve a la etapa 411 para verificar la relación para la siguiente señal de voz seleccionada. En la etapa 417, el módulo de preprocesamiento 221 transmite las señales de voz extraídas restantes al módulo de extracción de características 223 para un procesamiento adicional (etapa 417).

Según el método 400, durante la fase de preprocesamiento de la grabación de audio, el dispositivo de verificación del orador identifica las señales de voz cuyos géneros de orador no coinciden con el género del usuario predeterminado. Si una o más de las señales de voz identificadas constituyen una porción significativa de la grabación de audio, el dispositivo de verificación de orador concluye que el orador en la grabación de audio no es el usuario predeterminado, sin la necesidad de procesar más la grabación de audio. De lo contrario, el dispositivo de verificación de orador filtra las señales de voz identificadas de la grabación de audio, para reducir la cantidad de datos que necesitan procesamiento adicional. De esta manera, se mejora la precisión de la verificación de orador mientras se reduce el costo de cálculo.

La figura 5 es un diagrama de flujo que ilustra un proceso de verificación de orador 500, de acuerdo con una realización ejemplar. Esto puede realizarse mediante la sección de verificación de conductor 220 del verificador de orador 200 que se muestra en la figura 2. Por ejemplo, en un contexto de gestión del servicio de taxis, el método de verificación de orador 500 de esta realización puede usarse para verificar si el orador es el conductor registrado.

Etapa 501, el módulo de extracción de características 223 recibe los datos de voz del orador transmitidos desde el módulo de preprocesamiento 221.

Etapa 503, el módulo de extracción de características 223 extrae las características de los datos de voz del orador. Como se describió anteriormente, se pueden usar varias técnicas de extracción de características existentes usadas en tecnologías de reconocimiento de voz u orador. Por ejemplo, el módulo de extracción de características puede configurarse para extraer características de Coeficientes Cepstrales de Frecuencia de Mel (MFCC), que fue introducido por Davis y Mermelstein en la década de 1980 y ampliamente utilizado en reconocimiento de voz, otros pueden incluir Coeficientes de Predicción Lineal (LPC) o Coeficientes Cepstrales de Predicción Lineal (LPCC).

A continuación, las etapas 505 - 517 ilustran el proceso de verificar si el orador es el mismo que el conductor registrado en función de las características extraídas de los datos de voz del orador, y un modelo de orador correspondiente al conductor registrado. Estas etapas pueden realizarse mediante el módulo de verificación de orador 225. Además, los modelos de oradores de todos los conductores registrados pueden construirse en función de las características de voz de los conductores respectivos, y guardarse durante el registro del conductor, cuyo proceso se ilustra más adelante con referencia a la figura 6.

Etapa 505, el módulo de verificación de orador 225 compara las características de los datos de voz del orador con las características de referencia asociadas con un usuario predeterminado. En el contexto del servicio de taxis, el usuario predeterminado corresponde al conductor de registro, cuyo modelo de orador se ha construido y almacenado en una base de datos de modelo de orador durante el registro del conductor. Las características de referencia corresponden a las características del conductor registrado.

Etapa 507, el módulo de verificación de orador 225 calcula una puntuación de similitud que mide la similitud entre las características de los datos de voz del orador y las características de referencia del conductor registrado. Se puede calcular una puntuación de similitud verificando las características de los datos de voz del orador contra el modelo de orador correspondiente al conductor registrado. La puntuación de similitud representa el grado de similitud entre las características de los datos de voz del orador y las características de referencia del conductor registrado. Por ejemplo, un vector de características que representa las características de los datos de voz del orador puede generarse y compararse con el vector de características que representa las características asociadas con el conductor registrado.

En algunas realizaciones, se puede obtener un vector de características que representa las características asociadas con el conductor registrado después de entrenar una DNN con una pluralidad de grabaciones de audio del conductor registrado, y tomar un vector de característica promedio basado en los vectores de característica de la pluralidad de grabaciones de audio. La puntuación de similitud puede obtenerse calculando la diferencia entre el vector de características obtenido de los datos de voz del orador y el vector de características de promedio correspondiente al conductor registrado. Además, en algunas realizaciones, la puntuación del análisis discriminante lineal probabilístico (PLDA) puede usarse para el cálculo de la puntuación de similitud.

Etapa 511, el módulo de verificación de orador 225 compara la puntuación de similitud con un umbral

predeterminado. La puntuación de similitud calculada se comparará con un umbral predeterminado. El umbral predeterminado representa el estándar mínimo cuando la verificación es exitosa. Por ejemplo, si el valor de similitud se manifiesta como una distancia entre el vector de características de los datos de voz del orador y el vector de características correspondiente al conductor registrado, la distancia calculada en la etapa 507 se comparará con una distancia umbral predeterminada. En aplicaciones prácticas, el umbral puede manifestarse de otras maneras o como otros valores, y puede variar según los requisitos del sistema.

Etapa 513, si la puntuación de similitud satisface el umbral predeterminado, el módulo de verificación de orador 225 concluye que el orador es el conductor de registro, es decir, la verificación es exitosa.

Etapa 515, si la puntuación de similitud no satisface el umbral predeterminado, el módulo de verificación de orador 225 concluye que el orador no es el conductor de registro, es decir, la verificación no es exitosa.

Etapa 517, el módulo de verificación de orador 225 genera el resultado de la determinación. Como se señaló anteriormente, tomando el sistema descrito en la figura 1 como ejemplo, el resultado de la determinación puede enviarse al sistema de gestión del servicio de la cabina 150 a través de la interfaz de E/S 117, para actualizar el registro del servicio de la cabina o el registro de la cuenta del conductor. También puede enviarse al panel de control 160, que luego puede enviar instrucciones solicitando que se proporcione el resultado de la determinación a una terminal de pasajero 130 o terminal de conductor 140 correspondiente, a través de la red 120.

La figura 6 es un diagrama de flujo que ilustra un proceso de inscripción o registro de orador 600, de acuerdo con una realización ejemplar. En el contexto del servicio de taxis, el proceso 600 representa un proceso en el que se registra y almacena una nueva cuenta de conductor, y se puede construir y almacenar un modelo de orador del conductor para fines de verificación futura. Por ejemplo, el proceso 600 puede realizarse mediante la sección de registro del conductor 210 del verificador de orador 200 que se muestra en la figura 2.

Etapa 601, un conductor configura una cuenta de usuario con el sistema de gestión de servicio de taxis 150. Un conductor puede registrar una nueva cuenta de usuario a través de una aplicación de terminal de conductor. La aplicación de terminal del conductor puede transmitir la información de registro requerida a un sistema de gestión del servicio de taxis, tal como el sistema de gestión del servicio de taxis 150, que procesa la solicitud de registro y configura una cuenta de usuario para el conductor.

Etapa 603, el sistema de gestión de servicio de taxis 150 almacena información de registro de usuario. Cuando se configura una nueva cuenta de conductor, la información de registro del conductor puede almacenarse, incluyendo la información asociada con el terminal de conductor 140 correspondiente, la información personal del conductor, tal como género, edad, experiencia de conducción, información de licencia e información de contacto, tal como el número de teléfono del conductor u otra información de cuenta asociada con el conductor.

Etapa 605, el módulo de preprocesamiento 211 recibe entrada de grabación de audio desde un terminal de usuario, tal como un terminal de conductor 140. Se le puede solicitar al conductor recién registrado que ingrese una o más de sus grabaciones de audio, esto puede ser en la forma de leer ciertos textos designados o enunciados de voz aleatorios. Como se muestra en la figura 1, las grabaciones de audio desde el terminal de conductor 140 pueden transmitirse a través de la red 120 al módulo de preprocesamiento 211.

Etapa 607, el módulo de preprocesamiento 211 elimina el ruido ambiental de las señales de audio o extrae las señales de voz de las señales de audio. Las entradas de grabaciones de audio en un terminal de conductor 140 pueden incluir ruido de fondo o sonidos no relacionados con la voz. Estos sonidos pueden eliminarse de las señales de audio antes de la extracción de características. En las realizaciones descritas, la etapa 507 es similar a la etapa 305, cuyos detalles no se repiten en el presente documento.

Etapa 609, el módulo de preprocesamiento 211 determina si las señales de voz incluyen voces de oradores de diferentes géneros. Esta etapa puede ser opcional, dependiendo de la calidad y de la pureza de la entrada de grabación de audio. De manera similar a la etapa 307 descrita en relación con la figura 3, si las señales de voz incluyen voces de oradores de diferentes géneros puede determinarse, por ejemplo, en función de las diferentes características de frecuencias fundamentales correspondientes a diferentes géneros.

Etapa 611, el módulo de preprocesamiento 211 extrae los datos de voz del usuario asociados con el género que se registró previamente. Durante la etapa 603, la información de género del conductor recién registrado se ha ingresado y almacenado. Si las voces de oradores de diferentes géneros están presentes en las señales de voz, las señales de voz correspondientes al género registrado pueden extraerse para su posterior procesamiento. En las realizaciones descritas, la etapa 611 es similar a la etapa 309 descrita en relación con la figura 3, cuyos detalles no se repiten en el presente documento.

Etapa 613, el módulo de extracción de características 213 extrae características de los datos de voz del usuario recibidas en el módulo de extracción de características. Esto incluye etapas similares al proceso de extracción de características descrito con referencia a la figura 5, cuyos detalles no se repiten en el presente documento.

Etapa 615, el módulo de construcción del modelo de orador 215 construye un modelo de orador basado en las características de referencia. Como se describió anteriormente con referencia a la figura 2, el módulo de construcción de modelos de oradores 215 puede usar diversas técnicas de construcción de modelos de oradores. Por ejemplo, redes neuronales profundas (DNN) se pueden entrenar utilizando las características extraídas de los datos de voz del conductor, y se puede construir un modelo de orador extrayendo un vector de características que represente las características de voz del orador. Otros modelos de oradores que pueden utilizarse incluyen el Modelo de mezcla gaussiana (GMM) y un modelo de Modelo oculto de Markov (HMM). También hay combinaciones de diferentes enfoques para construir modelos de oradores, la presente divulgación no se limita a ningún enfoque de construcción de modelo de orador particular.

Etapa 617, después de construir un modelo de orador correspondiente al conductor registrado, el modelo de orador puede almacenarse en la base de datos del modelo de orador 217. Cada conductor registrado puede tener un modelo de orador construido y almacenado para verificación futura.

En algunas realizaciones, una determinada cuenta de conductor registrada puede no tener un modelo de orador almacenado correspondiente al conductor. De acuerdo con una realización de verificación de orador, el verificador de orador puede recopilar una pluralidad de grabaciones de audio de conversaciones conductor-pasajero durante un período de tiempo, para determinar si una cuenta de conductor registrada es utilizada por múltiples usuarios, como se describe más adelante con referencia a la figura 6.

La figura 7 es un diagrama de flujo que ilustra un método de verificación de orador 700 en el que no se ha almacenado ningún modelo de orador asociado con una cuenta de conductor registrada ejemplar. Por ejemplo, el método 700 puede ser realizado por el sistema 100.

Etapa 701, el verificador de orador 110 recibe múltiples grabaciones de audio desde un terminal de usuario. Por ejemplo, cuando varios conductores han utilizado un terminal de conductor 140 asociado con una cuenta de conductor registrada para proporcionar servicio de taxis, el verificador de orador 110 puede recopilar múltiples grabaciones durante un período de tiempo.

Etapa 703, el verificador de orador 110 procesa previamente cada grabación de audio para extraer los datos de voz del orador transmitidos desde el terminal de conductor 140. En las realizaciones descritas, la etapa 603 es similar a los métodos 300 y/o 400, cuyos detalles no se repiten en el presente documento.

Etapa 705, el verificador de orador 110 extrae características de los datos de voz del orador. En las realizaciones descritas, la etapa 705 es similar a la etapa 503 o 613, cuyos detalles no se repiten en el presente documento.

Etapa 707, el verificador de orador 110 clasifica las características extraídas. Después de que las características de los datos de voz del orador se extraen con respecto a cada entrada de grabación de audio desde el terminal de conductor 140, las características se clasifican para determinar si corresponden a la misma persona. Se pueden usar diferentes clasificadores de características en la clasificación, por ejemplo, el clasificador Naive Bayes o el clasificador de máquina de vectores de soporte (SVM). También se pueden usar otros clasificadores en las tecnologías de reconocimiento de oradores existentes, y se puede elegir en función de las características extraídas. Además, la clasificación puede no necesitar usar todas las características extraídas.

Etapa 709, el verificador de orador 110 determina si las características extraídas corresponden al mismo orador. Dependiendo de si la clasificación en la etapa 707 muestra múltiples categorías de características, el verificador de orador 110 puede determinar si hay múltiples usuarios que usan la misma cuenta de conductor. Si las características pertenecen a más de una categoría, entonces se puede determinar que hay varios usuarios que usan la misma cuenta de conductor. El resultado de la determinación puede luego enviarse al sistema de gestión del servicio de taxis 150 para acciones adicionales.

Otro aspecto de la divulgación está dirigido a un medio no transitorio, legible por ordenador, que almacena instrucciones que, cuando se ejecutan, hacen que uno o más procesadores realicen los métodos de verificación del orador, como se discutió anteriormente. El medio legible por ordenador puede incluir medios volátiles o no volátiles, magnéticos, semiconductores, cintas, ópticos, extraíbles, no extraíbles u otros tipos de medios legibles por ordenador o dispositivos de almacenamiento legibles por ordenador. Por ejemplo, el medio legible por ordenador puede ser la unidad de almacenamiento o el módulo de memoria que tiene las instrucciones del ordenador almacenadas en el mismo, como se describe. En algunas realizaciones, el medio legible por ordenador puede ser un disco o una unidad flash que tiene las instrucciones de la computadora almacenadas en el mismo.

Será evidente para los expertos en la materia que se pueden realizar diversas modificaciones y variaciones en el sistema de verificación de orador descrito y los métodos relacionados. Otras realizaciones serán evidentes para los expertos en la materia a partir de la consideración de la memoria descriptiva y la práctica del sistema de verificación de orador descrito y los métodos relacionados. Por ejemplo, aunque los sistemas y métodos de verificación de orador descritos se describen en relación con un servicio de llamadas de taxis, se pueden adaptar e implementar

para otras aplicaciones en las que se puede verificar la identidad de una persona en función de la voz, tal como los centros de llamadas de servicio al cliente, sistemas de pedidos de entrega de alimentos, sistemas de pago a través de teléfono, etc. Se pretende que la memoria descriptiva y los ejemplos se consideren solo a modo de ejemplo, con un verdadero alcance indicado por las siguientes reivindicaciones.

5

REIVINDICACIONES

1. Un método de verificación de orador, que comprende:

5 adquirir una grabación de audio;
extraer señales de voz de la grabación de audio;
extraer características de las señales de voz extraídas; y
determinar si las señales de voz extraídas representan la voz de un orador predeterminado basándose en las
características extraídas y un modelo de orador entrenado con datos de voz de referencia del orador
10 predeterminado
en el que extraer las señales de voz comprende:

15 determinar los niveles de energía de la grabación de audio que comprende determinar un espectro de imagen
de resonancia de tiempo-frecuencia (RTFI) de la grabación de audio; y
extraer las señales de voz basadas en los niveles de energía.

2. El método de la reivindicación 1, en el que la grabación de audio se transmite desde un terminal (130, 140)
asociado con el orador predeterminado, o la grabación de audio es de una llamada telefónica entre un conductor de
un vehículo y un pasajero del mismo vehículo.

20 3. El método de la reivindicación 1, en el que la extracción de las señales de voz comprende:

determinar si la grabación de audio incluye voces de oradores de diferentes géneros; y
cuando se determina que la grabación de audio incluye voces de oradores de diferentes géneros, extraer señales
de voz correspondientes al género del orador predeterminado.
25

4. El método de la reivindicación 3, en el que determinar si la grabación de audio incluye voces de oradores de
diferentes géneros comprende:
detectar si la grabación de audio incluye frecuencias fundamentales características de diferentes géneros.

30 5. El método de la reivindicación 1, en el que:

las señales de voz extraídas incluyen una primera señal de voz; y
extraer las señales de voz comprende:
35 determinar el género del orador de la primera señal de voz;
cuando el género del orador de la primera señal de voz es diferente del género del orador predeterminado,
determinar una relación de una duración de tiempo de la primera señal de voz sobre una duración de tiempo
de la grabación de audio;
40 cuando la relación excede un umbral predeterminado, concluir que la grabación de audio no incluye la voz del
orador predeterminado; y
cuando la relación es igual o inferior al umbral predeterminado, eliminar la primera señal de voz de las
señales de voz extraídas.

45 6. El método de la reivindicación 1, en el que las características extraídas comprenden Coeficientes Cepstrales de
Frecuencia de Mel (MFCC) de las señales de voz extraídas.

50 7. El método de la reivindicación 1, en el que determinar si las señales de voz extraídas representan la voz del
orador predeterminado comprende además:

extraer características de referencia asociadas con el orador predeterminado de los datos de voz de referencia; y
capacitar el modelo de orador basado en las características de referencia.

55 8. El método de la reivindicación 7, en el que el modelo de orador es al menos uno de un modelo de mezcla
gaussiana (GMM), un modelo de Markov oculto (HMM) o un modelo de red neuronal profunda (DNN).

9. El método de la reivindicación 7, que comprende adicionalmente:

determinar una similitud entre las características extraídas y las características de referencia; y
determinar si las señales de voz extraídas representan la voz del orador predeterminado en función de la
similitud.
60

10. Un sistema de verificación de orador (100), que comprende:

una memoria (111) que incluye instrucciones; y
un procesador (113) configurado para ejecutar las instrucciones para:
65

recibir una grabación de audio;
extraer señales de voz de la grabación de audio, que comprende:

5 determinar los niveles de energía de la grabación de audio que comprende
 determinar un espectro de imagen de frecuencia de tiempo de resonador (RTFI) de la grabación de audio;
 y
 extraer las señales de voz en función de los niveles de energía;

10 extraer características de las señales de voz extraídas; y

 determinar si las señales de voz extraídas representan la voz de un orador predeterminado basándose en las características extraídas y un modelo de orador entrenado con datos de voz de referencia del orador predeterminado.

15 11. El sistema (100) de la reivindicación 10, en el que la grabación de audio se transmite desde un terminal (130, 140) asociado con el orador predeterminado, o la grabación de audio es de una llamada telefónica entre un conductor de un vehículo y un pasajero del mismo vehículo.

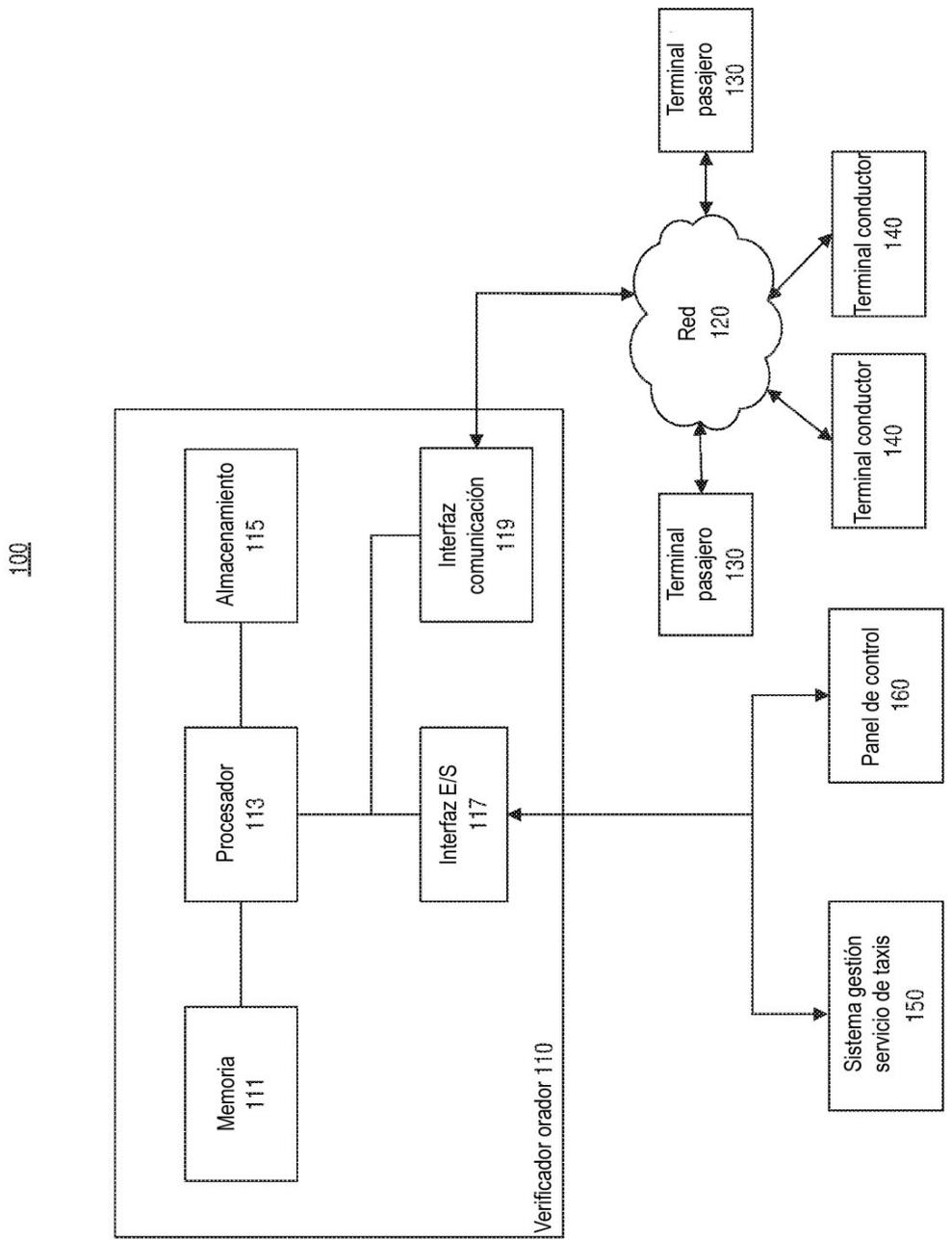


Fig. 1

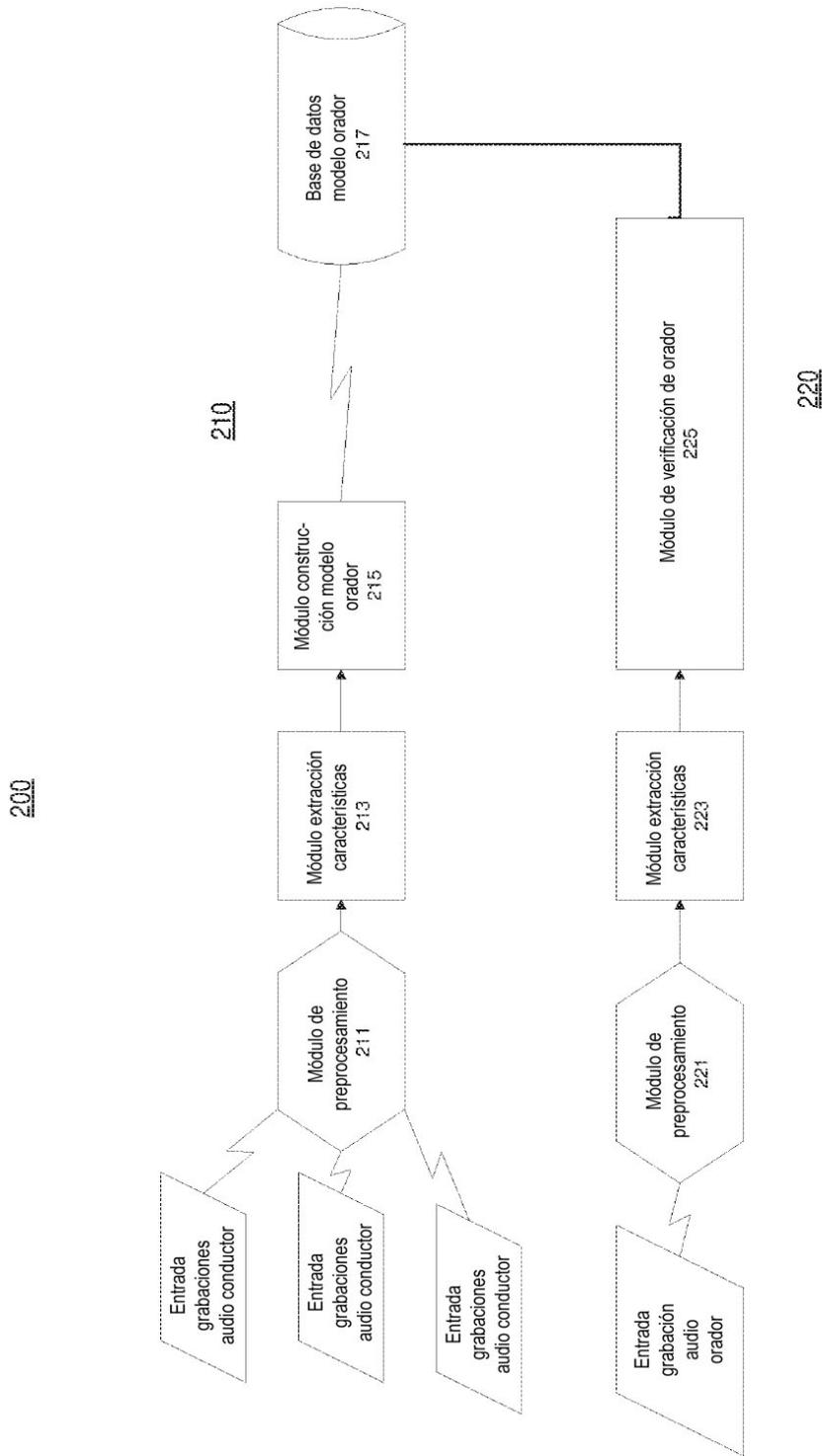


Fig. 2

300

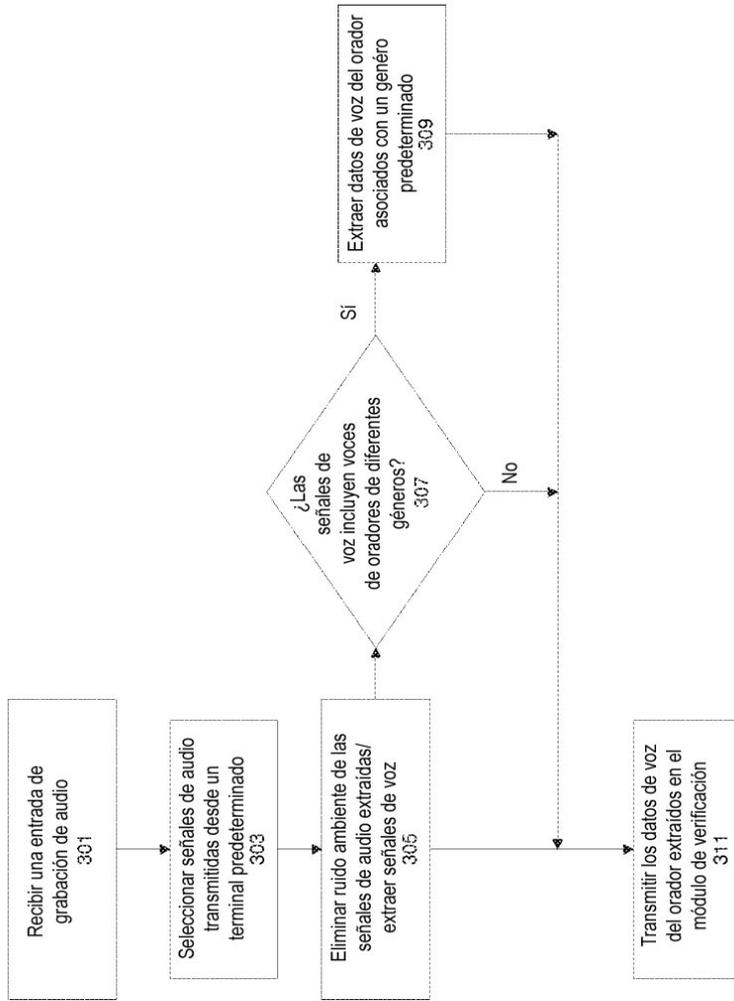


Fig. 3

400

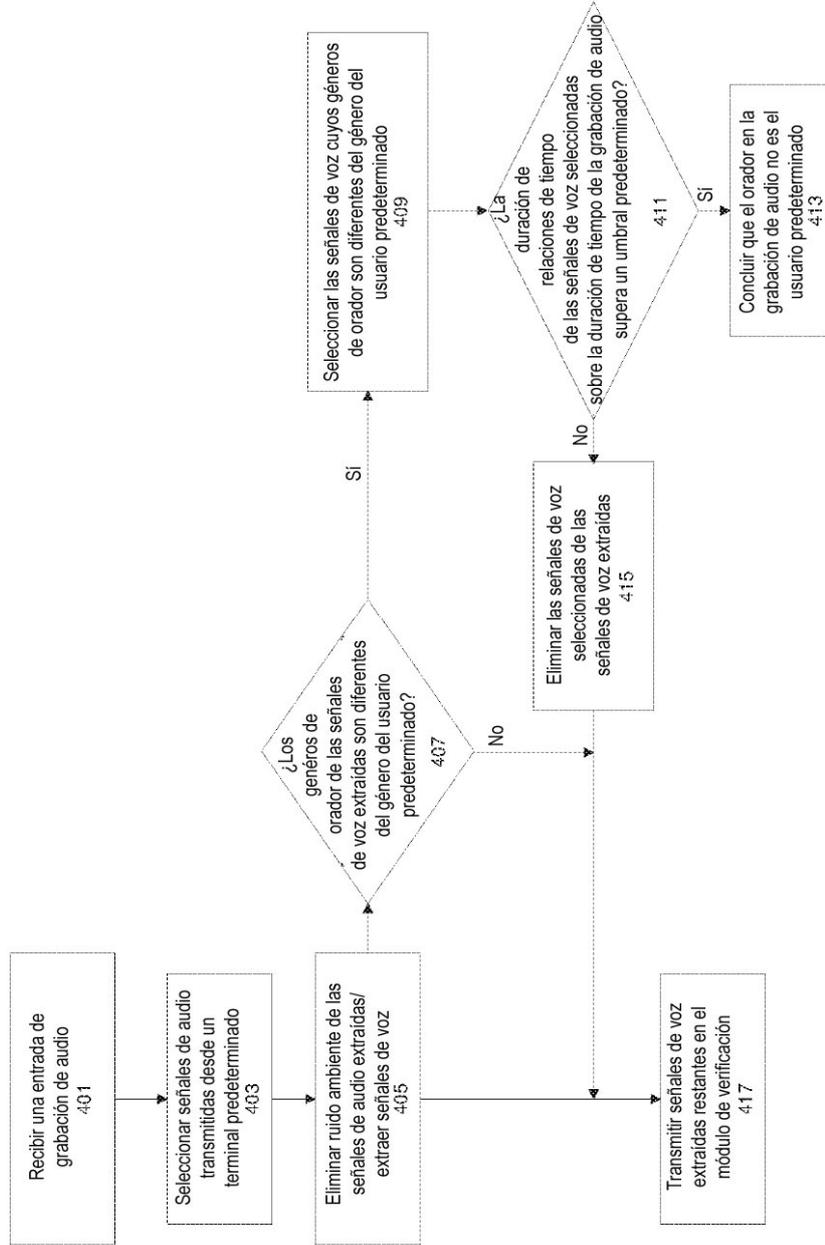


Fig. 4

500

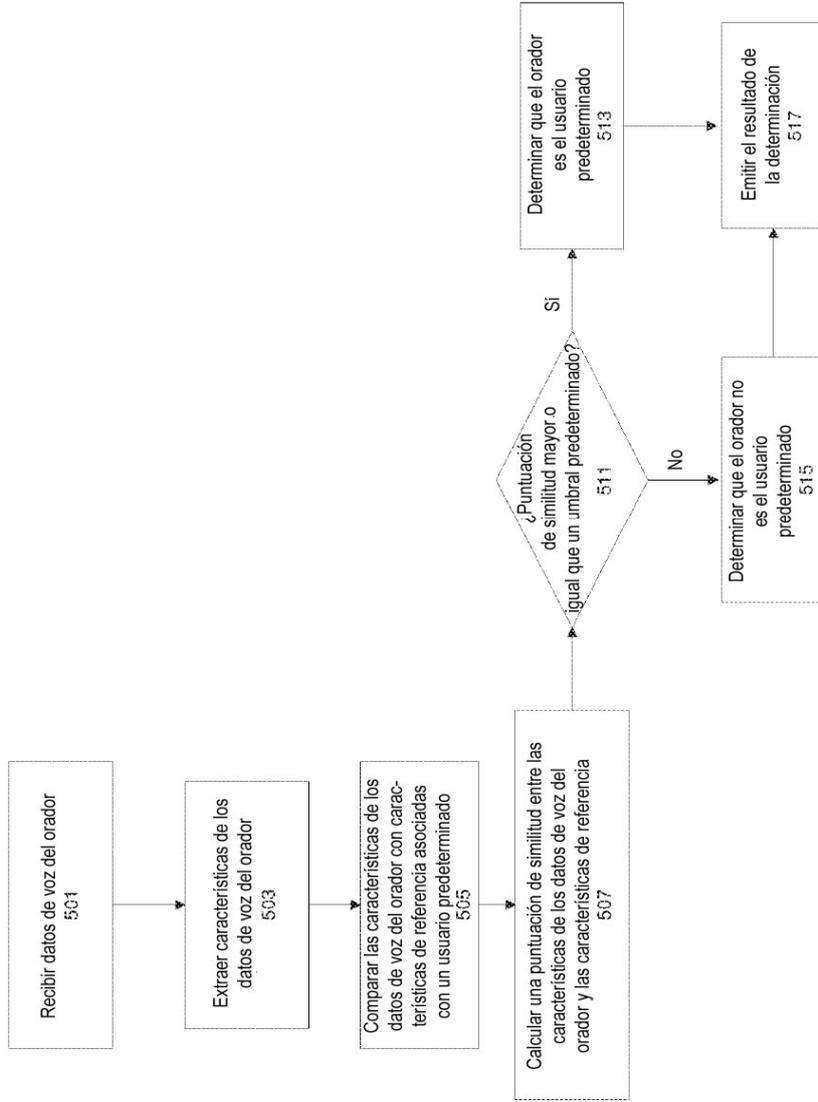


Fig. 5

600

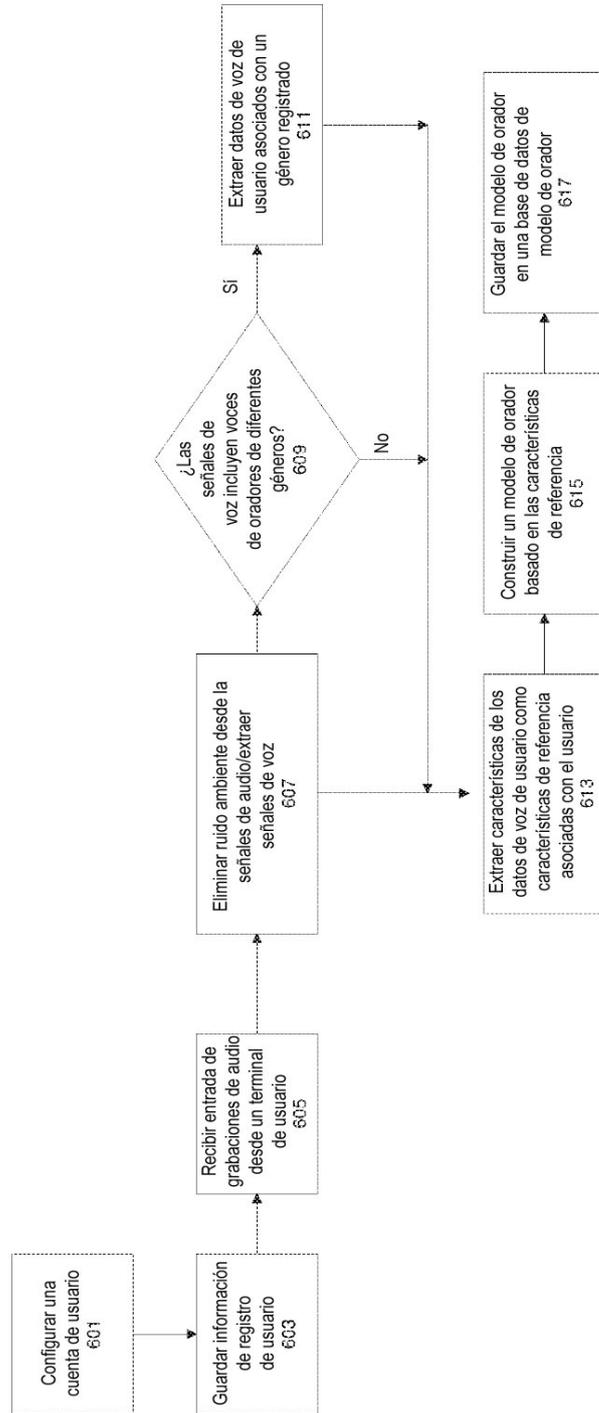


Fig. 6

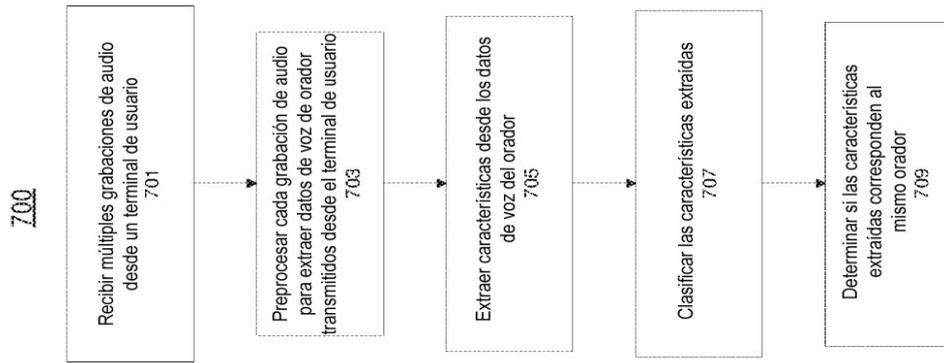


Fig. 7