

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 984 637**

51 Int. Cl.:

G10L 25/66 (2013.01)

A61B 5/00 (2006.01)

G10L 15/06 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.02.2020 E 20158058 (6)**

97 Fecha y número de publicación de la concesión europea: **03.07.2024 EP 3709300**

54 Título: **Técnicas de diagnóstico basadas en alineación de muestras de voz**

30 Prioridad:

12.03.2019 US 201916299178

12.03.2019 US 201916299186

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

30.10.2024

73 Titular/es:

CORDIO MEDICAL LTD. (100.0%)

6 Yehonatan Netanyahu Street

6037604 Or-Yehuda, IL

72 Inventor/es:

SHALLOM, ILAN D.

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 984 637 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

Técnicas de diagnóstico basadas en alineación de muestras de voz

La presente invención se refiere en general a diagnósticos médicos, particularmente con respecto a afecciones fisiológicas que afectan a la voz de un sujeto.

5 Sakoe y Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition" (Optimización del algoritmo de programación dinámica para el reconocimiento de palabras habladas), transacciones IEEE sobre acústica, voz y procesamiento de señales 26.2 (1978): 43-49, informa de un algoritmo de normalización temporal basado en programación dinámica óptima (DP) para el reconocimiento de palabras habladas. En primer lugar, se da un principio general de normalización temporal utilizando una función del alineamiento temporal. A continuación, se derivan del principio dos definiciones de distancia normalizadas en el tiempo, denominadas formas simétricas y asimétricas. Estas dos formas se comparan entre sí a través de descripciones teóricas y estudios experimentales. Se establece la superioridad del algoritmo de forma simétrica. Se introduce una técnica, llamada restricción de pendiente, en donde la pendiente de la función de alineamiento está restringida para mejorar la discriminación entre palabras en diferentes categorías.

15 Rabiner, Lawrence R., "A tutorial on hidden Markov models and selected applications in speech recognition" (un tutorial sobre modelos ocultos de Markov y aplicaciones seleccionadas en reconocimiento de voz), que procede del IEEE 77.2 (1989): 257-286 revisa aspectos teóricos de tipos de modelos estadísticos, y muestra cómo se han aplicado a problemas seleccionados en el reconocimiento de voz por máquina.

20 El documento US7457753 describe un sistema para la evaluación remota de un usuario. El sistema comprende software de aplicación residente en un servidor y dispuesto para interactuar a través de una red con un usuario que opera un dispositivo cliente para obtener una o más señales de muestra de la voz del usuario. Un almacén de datos está dispuesto para almacenar las muestras de voz del usuario en asociación con detalles del usuario. Un motor de extracción de características está dispuesto para extraer una o más primeras características de las muestras de voz respectivas. Un comparador está dispuesto para comparar las primeras características extraídas de una muestra de voz con las segundas características extraídas de una o más muestras de referencia y para proporcionar una medida de cualquier diferencia entre las características primera y segunda para la evaluación del usuario.

25 El documento US2009/0099848 describe un sistema y un procedimiento para el diagnóstico pasivo de demencias. Los indicadores clínicos y psicométricos de demencias se identifican automáticamente mediante mediciones estadísticas longitudinales, y se utilizan métodos matemáticos para rastrear la naturaleza del cambio de idioma y/o las características de audio del paciente. El sistema y el método descritos incluyen unidades de procesamiento multicapa en donde el procesamiento inicial de los datos de audio grabados se procesa en una unidad local. Los datos sin procesar procesados y requeridos también se transfieren a una unidad central que realiza análisis en profundidad de los datos de audio.

35 El documento US2015/0216448 describe un método para medir la capacidad pulmonar y la estamina de un usuario, para detectar insuficiencia cardíaca crónica, EPOC o asma. El método incluye proporcionar una aplicación cliente en el dispositivo de comunicación móvil del usuario, incluyendo dicha aplicación cliente código informático ejecutable para: dar instrucciones al usuario para llenar sus pulmones con sonidos vocales del aire y absolutos dentro de un cierto intervalo de sonoridad (decibelios) mientras está exhalando; recibir y registrar por el dispositivo de comunicación móvil dichos sonidos vocales del usuario; detener el registro de los sonidos vocales; medir la longitud del tiempo de recepción de los sonidos vocales dentro de dicho intervalo de sonoridad; y visualizar la longitud en la pantalla del dispositivo de comunicación móvil. El documento WO 2017/068582 describe sistemas y métodos para monitorizar y determinar una afección médica de un usuario y/o paciente. Más específicamente, se determina un estado o progreso de una enfermedad basado en el análisis de la voz utilizando una comparación con un perfil de usuario que representa un estado estable/conocido del paciente.

45 Los aspectos de la presente invención se exponen en las reivindicaciones independientes adjuntas. Las características de diversas realizaciones se exponen en las reivindicaciones dependientes adjuntas.

50 También se describen en la presente memoria ejemplos de un método que comprende: obtener al menos un modelo de voz construido a partir de una o más muestras de voz de referencia, producidas por un sujeto en un primer momento mientras en que un estado fisiológico del sujeto era conocido. El modelo de voz incluye (i) uno o más estados acústicos exhibidos en las muestras de voz de referencia, estando los estados acústicos asociados con las respectivas funciones de distancia local de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El método incluye además recibir al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido; y calcular una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de prueba. El método incluye además, basándose en las funciones de distancia

- 5 local y en las transiciones permitidas, asignar la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de la muestra de prueba a los estados acústicos respectivos de manera que se minimice una distancia total entre los vectores de características de la muestra de prueba y los estados acústicos respectivos, basándose la distancia total en las distancias locales respectivas entre los vectores de características de la muestra de prueba y los estados acústicos respectivos. El método incluye además, en respuesta a la asignación de la muestra de voz de prueba a la secuencia de distancia mínima de los estados acústicos, generar una salida que indique el estado fisiológico del sujeto en el segundo momento.
- En algunos ejemplos, el método incluye además recibir las muestras de voz de referencia, y obtener el modelo de voz incluye obtener el modelo de voz construyendo el modelo de voz a partir de las muestras de voz de referencia.
- 10 En algunos ejemplos, la distancia total se basa en una suma de las respectivas distancias locales.
- En algunos ejemplos, la distancia total es la suma de las respectivas distancias locales.
- En algunos ejemplos, la suma es una primera suma, el modelo define además las respectivas distancias de transición para las transiciones permitidas, y la distancia total es una segunda suma de (i) la primera suma, y (ii) las distancias de transición para aquellas transiciones permitidas que se incluyen en la secuencia de distancia mínima de los estados acústicos.
- 15 En algunos ejemplos, generar la salida incluye: comparar la distancia total con un umbral predeterminado; y generar la salida en respuesta a la comparación.
- En algunos ejemplos, la función de distancia local de cada estado acústico devuelve un valor que depende de un registro negativo de una probabilidad estimada de que el vector de característica acústica dado corresponda al estado acústico.
- 20 En algunos ejemplos, las muestras de voz de referencia producidas mientras que el estado fisiológico del sujeto era estable con respecto a una condición fisiológica particular.
- En algunos ejemplos, las muestras de voz de referencia son primeras muestras de voz de referencia, el modelo de voz es un primer modelo de voz, los estados acústicos son primeros estados acústicos, la secuencia de distancia mínima es una primera secuencia de distancia mínima, y la distancia total es una primera distancia total, el método incluye además: recibir una o más segundas muestras de voz de referencia producidas por el sujeto mientras que el estado fisiológico del sujeto era inestable con respecto a la condición fisiológica particular; basándose en las segundas muestras de voz de referencia, construir al menos un segundo modelo de voz que incluye uno o más segundos estados acústicos exhibidos en las segundas muestras de voz de referencia; asignar la muestra de voz de prueba a una segunda secuencia de distancia mínima de los segundos estados acústicos, asignando los vectores característicos de muestra de prueba a los respectivos de los segundos estados acústicos de tal manera que se minimice una segunda distancia total entre los vectores característicos de muestra de prueba y los respectivos de los segundos estados acústicos; y comparar la segunda distancia total con la primera distancia total, y generar la salida incluye generar la salida en respuesta a comparar la segunda distancia total con la primera distancia total.
- 25 En algunos ejemplos, las muestras de voz de referencia producidas mientras que el estado fisiológico del sujeto era inestable con respecto a una condición fisiológica particular.
- 30 En algunos ejemplos, las muestras de voz de referencia y la muestra de voz de prueba incluyen la misma expresión predeterminada.
- En algunos ejemplos, las muestras de voz de referencia incluyen voz libre del sujeto, construir el al menos un modelo de voz incluye: identificar múltiples unidades de voz diferentes en la voz libre; construir modelos de unidades de voz respectivos para las unidades de voz identificadas; y construir el al menos un modelo de voz concatenando los modelos de unidades de voz, de tal manera que el modelo de voz represente una concatenación particular de las unidades de voz identificadas, y la muestra de voz de prueba incluya la concatenación particular.
- 40 En algunos ejemplos, la distancia total es una primera distancia total, y generar la salida incluye: calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, siendo la segunda distancia total diferente de la primera distancia total; y generar la salida en respuesta a la segunda distancia total.
- 45 En algunos ejemplos, calcular la segunda distancia total incluye: ponderar las respectivas distancias locales por ponderaciones respectivas, siendo al menos dos de las ponderaciones diferentes entre sí; y calcular la segunda distancia total sumando las distancias locales ponderadas.
- 50 En algunos ejemplos, las respectivas distancias locales son primeras distancias locales respectivas, y calcular la segunda distancia total incluye: modificar las funciones de distancia local de los respectivos estados acústicos; utilizar las funciones de distancia local modificadas, calcular las segundas distancias locales respectivas entre los vectores de características de muestra de prueba y los respectivos estados acústicos; y calcular la segunda

distancia total sumando las segundas distancias locales.

En algunos ejemplos, modificar las funciones de distancia local incluye modificar las funciones de distancia local para dar mayor peso a al menos una de las características acústicas que a al menos otra de las características acústicas.

5 Se dan a conocer adicionalmente en la presente memoria ejemplos de un aparato que incluye una interfaz de red y un procesador. El procesador está configurado para obtener al menos un modelo de voz construido a partir de una o más muestras de voz de referencia, producidas por un sujeto en un primer momento mientras que un estado fisiológico del sujeto era conocido. El modelo de voz incluye (i) uno o más estados acústicos exhibidos en las muestras de voz de referencia, estando los estados acústicos asociados con las respectivas funciones de distancia local de tal manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El procesador está configurado además para recibir, a través de la interfaz de red, al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, y para calcular una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de prueba. El procesador está configurado además para, basándose en las funciones de distancia local y en las transiciones permitidas, asignar la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de muestra de prueba a los estados acústicos respectivos de manera que se minimice una distancia total entre los vectores de características de muestra de prueba y los estados acústicos respectivos, basándose la distancia total en las distancias locales respectivas entre los vectores de características de muestra de prueba y los estados acústicos respectivos. El procesador está configurado además para, en respuesta a la asignación de la muestra de voz de prueba a la secuencia de distancia mínima de los estados acústicos, generar una salida que indica el estado fisiológico del sujeto en el segundo momento.

25 En la presente memoria se dan a conocer ejemplos de un sistema que incluye circuitos y uno o más procesadores. El procesador está configurado para llevar a cabo de manera cooperativa un proceso que incluye obtener al menos un modelo de voz construido a partir de una o más muestras de voz de referencia, producidas por un sujeto en un primer momento mientras que un estado fisiológico del sujeto era conocido. El modelo de voz incluye (i) uno o más estados acústicos exhibidos en las muestras de voz de referencia, estando los estados acústicos asociados con las respectivas funciones de distancia local de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El proceso incluye además recibir, a través de los circuitos, al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, y calcular una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de prueba. El proceso incluye además, basándose en las funciones de distancia local y en las transiciones permitidas, asignar la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de la muestra de prueba a los respectivos estados acústicos de manera que se minimice una distancia total entre los vectores de características de la muestra de prueba y los respectivos estados acústicos, basándose la distancia total en las respectivas distancias locales entre los vectores de características de la muestra de prueba y los respectivos estados acústicos. El proceso incluye además, en respuesta a la asignación de la muestra de voz de prueba a la secuencia de distancia mínima de los estados acústicos, generar una salida que indique el estado fisiológico del sujeto en el segundo momento.

En algunos ejemplos, los circuitos incluyen un convertidor analógico-digital (A/D).

En algunos ejemplos, los circuitos incluyen una interfaz de red.

50 En la presente memoria se dan a conocer ejemplos de un producto de software informático que incluye un medio legible por ordenador no transitorio tangible en donde se almacenan instrucciones de programa. Las instrucciones, cuando se leen por un procesador, hacen que el procesador obtenga al menos un modelo de voz construido a partir de una o más muestras de voz de referencia, producidas por un sujeto en un primer momento mientras que un estado fisiológico del sujeto era conocido. El modelo de voz incluye (i) uno o más estados acústicos exhibidos en las muestras de voz de referencia, estando los estados acústicos asociados con las respectivas funciones de distancia local de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. Las instrucciones hacen además que el procesador reciba al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, y calcule una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de prueba. Las instrucciones hacen además que el procesador,

basándose en las funciones de distancia local y en las transiciones permitidas, asigne la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de muestra de prueba a los respectivos estados acústicos de manera que se minimice una distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, estando basada la distancia total en las respectivas distancias locales entre los vectores de características de muestra de prueba y los respectivos estados acústicos. Las instrucciones hacen además que el procesador, en respuesta a la asignación de la muestra de voz de prueba a la secuencia de distancia mínima de los estados acústicos, genere una salida que indique el estado fisiológico del sujeto en el segundo momento.

En la presente memoria se dan a conocer ejemplos de otro método que incluye obtener múltiples modelos de voz contruidos a partir la voz libre de un sujeto, producidos por primera vez mientras que un estado fisiológico del sujeto era conocido. Cada uno de los modelos de voz incluye, para una respectiva diferente de múltiples unidades de voz diferentes en la voz libre, (i) uno o más estados acústicos exhibidos en la unidad de voz, estando los estados acústicos asociados con funciones de distancia local respectivas de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El método incluye además recibir al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, e identificar, en la muestra de voz de prueba, una o más partes de muestra de prueba que incluyen las unidades de voz identificadas, respectivamente. El método incluye además asignar las partes de muestra de prueba a los respectivos modelos de voz, mediante, para cada una de las partes de muestra de prueba, el cálculo de una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la parte de muestra de prueba, la identificación del modelo de voz que se construyó para la unidad de voz incluida en la parte de muestra de prueba, y, basándose en las funciones de distancia local y en las transiciones permitidas incluidas en el modelo de voz identificado, la asignación de la parte de muestra de prueba al modelo de voz identificado, asignando los vectores de características de muestra de prueba a los respectivos estados acústicos incluidos en el modelo de voz identificado de manera que se minimice una distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, basándose la distancia total en distancias locales respectivas entre los vectores de características de muestra de prueba y los respectivos estados acústicos. El método incluye además, en respuesta a la asignación de las partes de muestra de prueba a los respectivos modelos de voz, generar una salida que indica el estado fisiológico del sujeto en el segundo momento.

En algunos ejemplos, el método incluye además recibir la voz libre, y obtener los modelos de voz incluye obtener los modelos de voz: identificando las unidades de voz en la voz libre, y basándose en las unidades de voz, construir los modelos de voz.

En algunos ejemplos, la distancia total se basa en una suma de las respectivas distancias locales.

En algunos ejemplos, la muestra de voz de prueba incluye una expresión predeterminada que incluye al menos una de las unidades de voz identificadas.

En algunos ejemplos, la voz libre es voz libre de referencia, y la muestra de voz de prueba incluye voz libre de prueba.

Se dan a conocer adicionalmente en la presente memoria ejemplos de otro aparato que incluye una interfaz de red y un procesador. El procesador está configurado para obtener múltiples modelos de voz contruidos a partir de la voz libre de un sujeto, producidos por primera vez mientras que un estado fisiológico del sujeto era conocido. Cada uno de los modelos de voz incluye, para una respectiva diferente de múltiples unidades de voz diferentes en la voz libre, (i) uno o más estados acústicos exhibidos en la unidad de voz, estando los estados acústicos asociados con funciones de distancia local respectivas de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El procesador está configurado además para recibir, a través de la interfaz de red, al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, e identificar, en la muestra de voz de prueba, una o más partes de muestra de prueba que incluyen las unidades de voz identificadas, respectivamente. El procesador está configurado además para asignar las partes de muestra de prueba a los respectivos modelos de voz, mediante, para cada una de las partes de muestra de prueba, el cálculo de una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la parte de muestra de prueba, la identificación del modelo de voz que se ha construido para la unidad de voz incluida en la parte de muestra de prueba, y, basándose en las funciones de distancia local y en las transiciones permitidas incluidas en el modelo de voz identificado, la asignación de la parte de muestra de prueba al modelo de voz identificado, asignando los vectores de características de muestra de prueba a los respectivos estados acústicos incluidos en el modelo de voz identificado de manera que se minimice una distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, basándose la distancia total en distancias locales respectivas entre los vectores de características de

muestra de prueba y los respectivos estados acústicos. El procesador está configurado además para, en respuesta a la asignación de las partes de muestra de prueba a los respectivos modelos de voz, generar una salida que indica el estado fisiológico del sujeto en el segundo momento.

5 En la presente memoria se dan a conocer adicionalmente ejemplos de otro sistema que incluye circuitos y uno o más procesadores. Los procesadores están configurados para llevar a cabo de manera cooperativa un proceso que incluye obtener múltiples modelos de voz contruidos a partir de la voz libre de un sujeto, producidos por primera vez mientras que un estado fisiológico del sujeto era conocido. Cada uno de los modelos de voz incluye, para una respectiva diferente de múltiples unidades de voz diferentes en la voz libre, (i) uno o más estados acústicos exhibidos en la unidad de voz, estando los estados acústicos asociados con funciones de distancia local respectivas de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El proceso incluye además recibir, a través de los circuitos, al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, e identificar, en la muestra de voz de prueba, una o más partes de muestra de prueba que incluyen las unidades de voz identificadas, respectivamente. El proceso incluye además asignar las partes de muestra de prueba a los respectivos modelos de voz, mediante, para cada una de las partes de muestra de prueba, el cálculo de una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la parte de muestra de prueba, la identificación del modelo de voz que se ha construido para la unidad de voz incluida en la parte de muestra de prueba, y, basándose en las funciones de distancia local y en las transiciones permitidas incluidas en el modelo de voz identificado, la asignación de la parte de muestra de prueba al modelo de voz identificado, asignando los vectores de características de muestra de prueba a los respectivos estados acústicos incluidos en el modelo de voz identificado de manera que se minimice una distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, basándose la distancia total en distancias locales respectivas entre los vectores de características de muestra de prueba y los respectivos estados acústicos. El proceso incluye, además, en respuesta a la asignación de las partes de muestra de prueba a los respectivos modelos de voz, generar una salida que indica el estado fisiológico del sujeto en el segundo momento.

30 En la presente memoria se dan a conocer adicionalmente ejemplos de otro producto de software informático que incluye un medio legible por ordenador no transitorio tangible en donde se almacenan instrucciones de programa. Las instrucciones, cuando se leen por un procesador, hacen que el procesador obtenga múltiples modelos de voz contruidos a partir de la voz libre de un sujeto, producidos en un primer momento mientras que un estado fisiológico del sujeto era conocido. Cada uno de los modelos de voz incluye, para una respectiva diferente de múltiples unidades de voz diferentes en la voz libre, (i) uno o más estados acústicos exhibidos en la unidad de voz, estando los estados acústicos asociados con funciones de distancia local respectivas de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. Las instrucciones hacen además que el procesador reciba al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, e identifique, en la muestra de voz de prueba, una o más partes de muestra de prueba que incluyen las unidades de voz identificadas, respectivamente. Las instrucciones hacen además que el procesador asigne las partes de muestra de prueba a los respectivos modelos de voz, mediante, para cada una de las partes de muestra de prueba, el cálculo de una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la parte de muestra de prueba, la identificación del modelo de voz que se ha construido para la unidad de voz incluida en la parte de muestra de prueba, y, basándose en las funciones de distancia local y en las transiciones permitidas incluidas en el modelo de voz identificado, la asignación de la parte de muestra de prueba al modelo de voz identificado, asignando los vectores de características de muestra de prueba a los respectivos estados acústicos incluidos en el modelo de voz identificado de manera que se minimiza una distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, basándose la distancia total en distancias locales respectivas entre los vectores de características de muestra de prueba y los respectivos estados acústicos. Las instrucciones hacen además que el procesador, en respuesta a la asignación de las partes de muestra de prueba a los modelos de voz respectivos, genere una salida que indique el estado fisiológico del sujeto en el segundo momento.

55 En la presente memoria se dan a conocer adicionalmente ejemplos de otro método más que incluye obtener al menos un modelo de voz que incluye (i) uno o más estados acústicos exhibidos en una o más muestras de voz de referencia, estando los estados acústicos asociados con respectivas funciones de distancia local de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El método incluye además recibir al menos una muestra de voz de prueba producida por un sujeto, y calcular una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la

muestra de voz de prueba. El método incluye además, basándose en las funciones de distancia local y en las transiciones permitidas, asignar la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de muestra de prueba a los estados acústicos respectivos de manera que se minimice una primera distancia total entre los vectores de características de muestra de prueba y los estados acústicos respectivos, basándose la primera distancia total en las distancias locales respectivas entre los vectores de características de muestra de prueba y los estados acústicos respectivos. El método incluye además calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, siendo la segunda distancia total diferente de la primera distancia total y, en respuesta a la segunda distancia total, generar una salida que indica un estado fisiológico del sujeto.

En la presente memoria se dan a conocer adicionalmente ejemplos de otro aparato más que incluye una interfaz de red y un procesador. El procesador está configurado para obtener al menos un modelo de voz que incluye (i) uno o más estados acústicos exhibidos en una o más muestras de voz de referencia, estando los estados acústicos asociados con respectivas funciones de distancia local de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El procesador está configurado además para recibir, a través de la interfaz de red, al menos una muestra de voz de prueba producida por un sujeto, y para calcular una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de prueba. El procesador está configurado además para, basándose en las funciones de distancia local y en las transiciones permitidas, asignar la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de la muestra de prueba a los respectivos estados acústicos de manera que se minimice una primera distancia total entre los vectores de características de la muestra de prueba y los respectivos estados acústicos, la primera distancia total se basa en las respectivas distancias locales entre los vectores de características de la muestra de prueba y los respectivos estados acústicos. El procesador está configurado además para calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, siendo la segunda distancia total diferente de la primera distancia total y, en respuesta a la segunda distancia total, generar una salida que indica un estado fisiológico del sujeto.

En la presente memoria se dan a conocer adicionalmente ejemplos de otro sistema más que incluye circuitos y uno o más procesadores. Los procesadores están configurados para llevar a cabo de manera cooperativa un proceso que incluye obtener al menos un modelo de voz que incluye (i) uno o más estados acústicos exhibidos en una o más muestras de voz de referencia, estando los estados acústicos asociados con respectivas funciones de distancia local de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. El proceso incluye además recibir, a través de los circuitos, al menos una muestra de voz de prueba producida por un sujeto, y calcular una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de prueba. El proceso incluye además, basándose en las funciones de distancia local y en las transiciones permitidas, asignar la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de la muestra de prueba a los respectivos estados acústicos de manera que se minimice una primera distancia total entre los vectores de características de la muestra de prueba y los respectivos estados acústicos, la primera distancia total se basa en las respectivas distancias locales entre los vectores de características de la muestra de prueba y los respectivos estados acústicos. El proceso incluye además calcular una segunda distancia total entre los vectores de características de la muestra de prueba y los respectivos estados acústicos, siendo la segunda distancia total diferente de la primera distancia total y, en respuesta a la segunda distancia total, generar una salida que indica un estado fisiológico del sujeto.

En la presente memoria se dan a conocer adicionalmente ejemplos de otro producto de software informático que incluye un medio legible por ordenador no transitorio tangible en donde se almacenan instrucciones de programa. Las instrucciones, cuando son leídas por un procesador, hacen que el procesador obtenga al menos un modelo de voz que incluye (i) uno o más estados acústicos exhibidos en una o más muestras de voz de referencia, estando los estados acústicos asociados con respectivas funciones de distancia local de manera que, dado cualquier vector de característica acústica dentro de un dominio de las funciones de distancia local, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico, y (ii) siempre que el modelo de voz incluya múltiples estados acústicos, transiciones permitidas entre los estados acústicos. Las instrucciones hacen además que el procesador reciba al menos una muestra de voz de prueba producida por un sujeto, y calcule una pluralidad de vectores de características de muestra de prueba que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de prueba. Las instrucciones hacen además que el procesador, basándose en base funciones de distancia local y en las transiciones permitidas, asigne la muestra de voz de prueba a una secuencia de distancia mínima de los estados acústicos, asignando los vectores de características de muestra de prueba a los respectivos

- estados acústicos de manera que se minimice una primera distancia total entre los vectores de características de muestra de prueba y los respectivos estados acústicos, la primera distancia total se basa en las respectivas distancias locales entre los vectores de características de muestra de prueba y los respectivos estados acústicos. Las instrucciones hacen además que el procesador calcule una segunda distancia total entre los vectores de características de muestra de prueba y los estados acústicos respectivos, siendo la segunda distancia total diferente de la primera distancia total, y, en respuesta a la segunda distancia total, genere una salida que indica un estado fisiológico del sujeto.
- En la presente memoria se dan a conocer adicionalmente ejemplos de otro método más que incluye obtener una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra de voz de referencia, producida por un sujeto en un primer momento mientras que un estado fisiológico del sujeto era conocido; recibir al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que se desconocía el estado fisiológico del sujeto; calcular una pluralidad de vectores de características de muestra de prueba que cuantifican las características acústicas de diferentes partes respectivas de la muestra de voz de prueba; asignar la muestra de voz de prueba a la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, de manera que se minimiza una distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia; y en respuesta a la asignación de la muestra de voz de prueba a la muestra de voz de referencia, generar una salida que indica el estado fisiológico del sujeto en el segundo momento.
- En algunos ejemplos, el procedimiento incluye además recibir la muestra de voz de referencia, y obtener los vectores de características de muestra de referencia incluye obtener los vectores de características de muestra de referencia calculando los vectores de características de muestra de referencia basándose en la muestra de voz de referencia.
- En algunos ejemplos, la distancia total se deriva de las respectivas distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia.
- En algunos ejemplos, la distancia total es una suma ponderada de las distancias locales.
- En algunos ejemplos, asignar la muestra de voz de prueba a la muestra de voz de referencia incluye asignar la muestra de voz de prueba a la muestra de voz de referencia utilizando un algoritmo de alineamiento temporal dinámico (DTW).
- En algunos ejemplos, generar la salida incluye: comparar la distancia total con un umbral predeterminado; y generar la salida en respuesta a la comparación.
- En algunos ejemplos, la muestra de voz de referencia se ha producido mientras que el estado fisiológico del sujeto era estable con respecto a una condición fisiológica particular.
- En algunos ejemplos, la muestra de voz de referencia es una primera muestra de voz de referencia, los vectores de características de muestra de referencia son primeros vectores de características de muestra de referencia, y la distancia total es una primera distancia total, el método incluye además: recibir al menos una segunda muestra de voz de referencia producida por el sujeto mientras que el estado fisiológico del sujeto era inestable con respecto a la condición fisiológica particular; calcular una pluralidad de segundos vectores de características de muestra de referencia que cuantifican las características acústicas de diferentes partes respectivas de la segunda muestra de voz de referencia; asignar la muestra de voz de prueba a la segunda muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos de los segundos vectores de características de muestra de referencia, bajo las restricciones predefinidas, de manera que se minimiza una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los segundos vectores de características de muestra de referencia; y comparar la segunda distancia total con la primera distancia total, y generar la salida incluye generar la salida en respuesta a comparar la segunda distancia total con la primera distancia total.
- En algunos ejemplos, las muestras de voz de referencia se han producido mientras que el estado fisiológico del sujeto era inestable con respecto a una condición fisiológica particular.
- En algunos ejemplos, la muestra de voz de referencia y la muestra de voz de prueba incluyen la misma expresión predeterminada.
- En algunos ejemplos, la muestra de voz de referencia incluye voz libre del sujeto, y la muestra de voz de prueba incluye una pluralidad de unidades de voz que se incluyen en la voz libre.
- En algunos ejemplos, la distancia total es una primera distancia total, y generar la salida incluye: calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia, siendo la segunda distancia total diferente de la primera distancia total; y generar la salida en respuesta a la segunda distancia total.
- En algunos ejemplos, la primera distancia total es una primera suma ponderada de las respectivas distancias locales

- entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia, en donde la primera suma ponderada de las distancias locales se ponderan por las respectivas primeras ponderaciones, y la segunda distancia total es una segunda suma ponderada de las respectivas distancias locales en donde las distancias locales se ponderan por las respectivas segundas ponderaciones, siendo al menos una de las segundas ponderaciones diferente de una correspondiente de las primeras ponderaciones.
- 5 En algunos ejemplos, el método incluye, además: asociar los vectores de características de muestra de referencia con las respectivas unidades fonéticas acústicas (APU); y seleccionar las segundas ponderaciones en respuesta a las APU.
- 10 En algunos ejemplos, asociar los vectores de características de muestra de referencia con las APU incluye asociar los vectores de características de muestra de referencia con las APU aplicando un algoritmo de reconocimiento de voz a la muestra de voz de referencia.
- 15 En algunos ejemplos, la primera distancia total se basa en las respectivas primeras distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia, y la segunda distancia total se basa en las respectivas segundas distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia, siendo al menos una de las segundas distancias locales diferente de una correspondiente de las primeras distancias locales.
- 20 En algunos ejemplos, asignar la muestra de voz de prueba a la muestra de voz de referencia incluye calcular las primeras distancias locales utilizando una primera medida de distancia, y calcular la segunda distancia total incluye calcular las segundas distancias locales utilizando una segunda medida de distancia que es diferente de la primera medida de distancia.
- 25 En algunos ejemplos, calcular la segunda distancia total incluye calcular las segundas distancias locales basándose en al menos una de las características acústicas que no han contribuido a las primeras distancias locales.
- 30 En la presente memoria se dan a conocer adicionalmente ejemplos de otro aparato más que incluye una interfaz de red y un procesador. El procesador está configurado para obtener una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra de voz de referencia, producida por un sujeto en un primer momento mientras que un estado fisiológico del sujeto era conocido. El procesador está configurado además para recibir, a través de la interfaz de red, al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, y para calcular una pluralidad de vectores de características de muestra de prueba que cuantifican las características acústicas de diferentes partes respectivas de la muestra de voz de prueba. El procesador está configurado además para asignar la muestra de voz de prueba a la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, de manera que se minimiza una distancia total entre los vectores de características de muestra de prueba y los respectivos vectores de características de muestra de referencia. El procesador está configurado además para, en respuesta a la asignación de la muestra de voz de prueba a la muestra de voz de referencia, generar una salida que indica el estado fisiológico del sujeto en el segundo momento.
- 35 En la presente memoria se dan a conocer adicionalmente ejemplos de otro sistema más que incluye circuitos; y uno o más procesadores, configurados para llevar a cabo de manera cooperativa un proceso que incluye: obtener una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra de voz de referencia, producida por un sujeto en un primer momento mientras que un estado fisiológico del sujeto era conocido, recibir, a través de los circuitos, al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, calcular una pluralidad de vectores de características de muestra de prueba que cuantifican las características acústicas de diferentes partes respectivas de la muestra de voz de prueba, asignar la muestra de voz de prueba a la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos de los vectores de características de muestra de referencia, bajo restricciones predefinidas, de manera que se minimiza una distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia, y en respuesta a la asignación de la muestra de voz de prueba a la muestra de voz de referencia, generar una salida que indica el estado fisiológico del sujeto en el segundo momento.
- 40 En la presente memoria se dan a conocer ejemplos de otro producto de software informático que incluye un medio legible por ordenador no transitorio tangible en donde se almacenan instrucciones de programa, instrucciones que, cuando son leídas por un procesador, hacen que el procesador: obtenga una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra de voz de referencia, producida por un sujeto en un primer momento mientras que un estado fisiológico del sujeto era conocido, reciba al menos una muestra de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido, calcule una pluralidad de vectores de características de muestra de prueba que cuantifican las características acústicas de diferentes partes
- 45
- 50
- 55

respectivas de la muestra de voz de prueba, asigne la muestra de voz de prueba a la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos de los vectores de características de muestra de referencia, bajo restricciones predefinidas, de manera que se minimice una distancia total entre los vectores de características de muestra de prueba y los respectivos vectores de características de muestra de referencia, y en respuesta a la asignación de la muestra de voz de prueba a la muestra de voz de referencia, genere una salida que indica el estado fisiológico del sujeto en el segundo momento.

La presente invención se comprenderá más completamente a partir de la siguiente descripción detallada de realizaciones de la misma, tomada junto con los dibujos, en donde:

La Fig. 1 es una ilustración esquemática de un sistema para evaluar el estado fisiológico de un sujeto;

La Fig. 2 es una ilustración esquemática de una construcción de un modelo de voz;

La Fig. 3 es una ilustración esquemática de una asignación de una muestra de voz de prueba a un modelo de voz;

La Fig. 4 es una ilustración esquemática de una técnica para construir un modelo de voz a partir de múltiples modelos de unidades de voz;

La Fig. 5 es una ilustración esquemática de una asignación de una muestra de voz de prueba a una muestra de voz de referencia; y

La Fig. 6 es un diagrama de flujo para un algoritmo de ejemplo para evaluar una muestra de voz de prueba de un sujeto.

Las realizaciones de la presente invención incluyen un sistema para evaluar el estado fisiológico de un sujeto analizando la voz del sujeto. Por ejemplo, analizando la voz del sujeto, el sistema puede identificar un inicio de, o un deterioro con respecto a, una afección fisiológica tal como insuficiencia cardíaca congestiva (CHF), cardiopatía coronaria, fibrilación auricular o cualquier otro tipo de arritmia, enfermedad pulmonar obstructiva crónica (EPOC), asma, enfermedad pulmonar intersticial, edema pulmonar, derrame pleural, enfermedad de Parkinson o depresión. En respuesta a la evaluación, el sistema puede generar una salida, tal como una alerta al sujeto, al médico del sujeto y/o a un servicio de monitorización.

Para evaluar el estado fisiológico del sujeto, el sistema adquiere primero una o más muestras de voz de referencia (o "referencia") del sujeto cuando se considera que el estado fisiológico del sujeto es estable. Por ejemplo, las muestras de referencia pueden adquirirse después de una indicación del médico del sujeto de que el estado fisiológico del sujeto es estable. Como otro ejemplo, para un sujeto que padece edema pulmonar, el sistema puede adquirir las muestras de voz de referencia después del tratamiento del sujeto para estabilizar la respiración del sujeto. Posteriormente a la obtención de cada muestra de voz de referencia, el sistema extrae una secuencia de vectores de características acústicas de la muestra. Cada vector de características corresponde a un punto temporal respectivo diferente en la muestra, en virtud de la cuantificación de las propiedades acústicas de la muestra en las proximidades temporales del punto temporal.

Posteriormente a (por ejemplo, varios días después) adquirir las muestras de referencia, cuando se desconoce el estado del sujeto, el sistema adquiere al menos otra muestra de voz del sujeto, denominada en lo sucesivo "muestra de voz de prueba", y extrae los respectivos vectores de características de la muestra. Posteriormente, basándose en los vectores de características de la muestra de prueba y las muestras de referencia, el sistema calcula al menos un valor de distancia que cuantifica la desviación de la muestra de prueba de las muestras de referencia, como se describe en detalle a continuación. En respuesta a que esta distancia satisface uno o más criterios predefinidos (por ejemplo, en respuesta a que la distancia supera un umbral predefinido), el sistema puede generar una alerta y/u otra salida.

Más particularmente, en algunas realizaciones, basándose en los vectores de características extraídos de las muestras de referencia, el sistema construye un modelo estadístico paramétrico específico del sujeto, que representa la voz del sujeto mientras que el estado fisiológico del sujeto se considera estable. En particular, la voz del sujeto está representada por múltiples estados acústicos, que corresponden implícitamente a los estados físicos respectivos del sistema de producción de voz del sujeto. El modelo define además las transiciones permitidas entre los estados, y puede incluir además respectivas distancias de transición (o "costes") para las transiciones.

Los estados acústicos están asociados con respectivas funciones de distancia local paramétrica, que se definen para un dominio particular de vectores. Dado cualquier vector de características particular dentro del dominio, cada función de distancia local, cuando se aplica al vector de características, devuelve un valor que indica un grado de correspondencia entre el vector de características y el estado acústico con el que está asociada la función. En la presente especificación, este valor se denomina "distancia local" entre el vector de características y el estado acústico.

En algunas realizaciones, cada estado acústico está asociado con una función de densidad de probabilidad (PDF) respectiva, y la distancia local entre el estado acústico y un vector de características es el negativo del logaritmo de la PDF aplicada al vector de características. De manera similar, cada transición puede estar asociada con una probabilidad de transición respectiva, y el coste para la transición puede ser el negativo del logaritmo de la

probabilidad de transición. Al menos algunos modelos que tienen estas propiedades se conocen como modelos ocultos de Markov (HMM).

Posteriormente a la construcción del modelo, para analizar la muestra de voz de prueba, el sistema asigna la muestra de prueba al modelo, asignando cada uno de los vectores de características de la muestra de prueba (es decir, los vectores de características extraídos de la muestra de prueba) a uno respectivo de los estados acústicos pertenecientes al modelo. En particular, el sistema selecciona, de entre todas las asignaciones posibles, la asignación que proporciona una secuencia de estados que tienen la distancia total mínima, dadas las transiciones de estado permitidas. Esta distancia total puede calcularse como la suma de las respectivas distancias locales entre los vectores de características de muestra de prueba y los estados acústicos a los que se asignan; opcionalmente, la suma de las distancias de transición incluidas en la secuencia puede añadirse a esta suma. En respuesta a la distancia total entre la muestra y el modelo, el sistema puede generar una alerta y/u otra salida.

En algunas realizaciones, cada una de las muestras de referencia incluye la misma expresión particular, es decir, la misma secuencia de unidades de voz. Por ejemplo, el teléfono móvil del sujeto puede solicitar al sujeto que produzca las muestras de referencia repitiendo una o más frases, palabras o sílabas designadas, que pueden contener cualquier número de fonemas, dífonos, trífonos designados y/u otras unidades fonéticas acústicas (APU). A medida que el sujeto produce las muestras de referencia, un micrófono que pertenece al teléfono móvil puede registrar las muestras. Posteriormente, un procesador perteneciente al teléfono móvil o a un servidor remoto puede construir, a partir de las muestras, un modelo que representa la expresión particular. Posteriormente, para adquirir la muestra de prueba, el sistema solicita al sujeto que repita la expresión.

En otras realizaciones, las muestras de referencia se adquieren a partir de la voz libre del sujeto. Por ejemplo, el teléfono móvil del sujeto puede solicitar al sujeto que responda una o más preguntas, y las respuestas del sujeto a las preguntas pueden registrarse a continuación. Alternativamente, puede grabarse la voz del sujeto durante una conversación normal. Posteriormente a la adquisición de las muestras de referencia, el sistema utiliza un algoritmo de reconocimiento de voz adecuado para identificar diversas unidades de voz en las muestras de referencia. Por ejemplo, el sistema puede identificar diversas palabras, APU (tales como fonemas, sílabas, trífonos o dífonos) o unidades acústicas sintéticas tales como estados de HMM únicos. El sistema construye, a continuación, modelos respectivos, denominados en la presente memoria "modelos de unidades de voz", para estas unidades de voz. (En el caso de una unidad acústica sintética que incluye un único estado de HMM, el modelo de unidad de voz incluye un HMM de único estado).

Posteriormente a la construcción de los modelos de unidades de voz, el sistema puede concatenar los modelos de unidades de voz en un modelo combinado que representa una expresión particular, basándose en el orden en donde aparecen las unidades de voz en la expresión. (Para concatenar dos modelos de unidades de voz cualesquiera, el sistema añade una transición desde el estado final de un modelo al estado inicial del otro modelo, y, si se utilizan distancias de transición, asigna una distancia de transición a esta transición). El sistema puede adquirir, a continuación una muestra de prueba que incluye esta expresión particular, y asignar la muestra de prueba al modelo combinado.

Alternativamente, en lugar de concatenar los modelos de unidades de voz, el sistema puede solicitar al sujeto que produzca, para la muestra de prueba, cualquier expresión particular que incluya las unidades de voz para las que se han construido los modelos de unidades de voz. El sistema puede, a continuación, identificar estas unidades de voz en la muestra de prueba, y calcular la respectiva "distancia de unidad de voz" entre cada unidad de voz y el modelo de unidad de voz correspondiente. Basándose en las distancias de la unidad de voz, el sistema puede calcular una distancia total entre la muestra de prueba y las muestras de referencia. Por ejemplo, el sistema puede calcular la distancia total sumando las distancias de la unidad de voz.

Como otra alternativa más, la muestra de prueba puede adquirirse a partir de la voz libre del sujeto. Como el sistema identifica el contenido verbal de la muestra de prueba, el sistema puede calcular una distancia de unidad de voz respectiva para cada unidad de voz en la muestra de prueba que tiene un modelo de unidad de voz correspondiente. El sistema puede, a continuación, calcular la distancia total desde las distancias de la unidad de voz, como se ha descrito anteriormente.

En otras realizaciones, el sistema no construye un modelo a partir de las muestras de referencia, sino que, en su lugar, compara directamente la muestra de voz de prueba con cada una de las muestras de referencia individuales que se han adquirido previamente. Por ejemplo, para adquirir una muestra de referencia, el sistema puede solicitar al sujeto que pronuncie una expresión particular. Posteriormente, para adquirir la muestra de prueba, el sistema puede solicitar al sujeto que suprima la misma expresión y las dos muestras pueden compararse entre sí. Alternativamente, el sistema puede registrar la voz libre del sujeto, y extraer una muestra de referencia a partir de la voz libre, utilizando un algoritmo de reconocimiento automático de voz (ASR) para identificar el contenido verbal de la muestra de referencia. Posteriormente, para adquirir la muestra de prueba, el sistema puede solicitar al sujeto que produzca el mismo contenido verbal.

Para realizar la comparación entre las muestras de prueba y de referencia, el sistema utiliza un algoritmo de alineamiento, tal como el algoritmo de alineamiento temporal dinámico (DTW) mencionado anteriormente en los antecedentes, para alinear la muestra de prueba con la muestra de referencia, es decir, para encontrar una

correspondencia entre cada vector de características de muestra de prueba y un vector de características de muestra de referencia respectivo. (Por alineamiento, múltiples vectores de características de muestra de prueba consecutivos pueden corresponder a un único vector de características de muestra de referencia; de manera similar, múltiples vectores de características de muestra de referencia consecutivos pueden corresponder a un único vector de características de muestra de prueba). Al realizar la alineación, el sistema calcula una distancia D entre las dos muestras. Posteriormente, el sistema puede generar una alerta, y/o cualquier otra salida adecuada, en respuesta a D. (el alineamiento mencionado anteriormente también se denomina a continuación "asignación", en donde la muestra de prueba se asigna a la muestra de referencia).

En algunas realizaciones, se obtienen una o más muestras de voz de referencia cuando se considera que el estado fisiológico del sujeto es inestable, por ejemplo, debido al inicio de un deterioro con respecto a una enfermedad particular. (En el contexto de la presente solicitud, incluyendo las reivindicaciones, se dice que el estado fisiológico de un sujeto es "inestable" si la salud del sujeto se deteriora de cualquier manera, incluso si el sujeto no siente ningún síntoma del deterioro). Basándose en estas muestras, el sistema puede construir un modelo estadístico paramétrico que representa la voz del sujeto en el estado inestable. El sistema puede comparar, a continuación la muestra de prueba tanto con el "modelo estable" como con el "modelo inestable" y generar una alerta, por ejemplo, si la muestra de prueba está más cerca del modelo inestable que del modelo estable. Alternativamente, incluso sin construir un modelo estable, el sistema puede comparar la muestra de prueba con el modelo inestable, y generar una alerta en respuesta a la comparación, por ejemplo, en respuesta a que la distancia entre la muestra de prueba y el modelo sea menor que un umbral predefinido.

De manera similar, el sistema puede, utilizando una técnica de alineamiento como se ha descrito anteriormente, comparar la muestra de prueba directamente con una muestra de referencia "inestable", alternativa o adicionalmente a comparar la muestra de prueba con una muestra de referencia "estable". En respuesta a esta comparación, el sistema puede generar una alerta.

En algunas realizaciones, se obtienen múltiples muestras de voz de referencia de otros sujetos, normalmente mientras que estos sujetos están en un estado inestable con respecto a la afección particular que padece el sujeto. Basándose en estas muestras (y/o muestras que se adquirieron del sujeto), se construye un modelo de voz general (es decir, no específico del sujeto). Posteriormente, las muestras de prueba del sujeto pueden asignarse al modelo general. Ventajosamente, esta técnica puede obviar la necesidad de adquirir un número significativo de muestras de referencia del sujeto, lo que puede ser particularmente difícil de hacer mientras el estado del sujeto es inestable.

En algunas realizaciones, las secuencias de vectores de características de muestra de referencia se etiquetan como correspondientes a unidades de voz respectivas, tales como palabras o fonemas respectivos. Por ejemplo, cada muestra de referencia puede asignarse a un HMM independiente del hablante en donde grupos de uno o más estados corresponden a unidades de voz conocidas respectivas. (Como se ha observado anteriormente, tal asignación se realiza en cualquier caso en el evento de que la muestra de referencia se obtenga a partir de la voz libre del sujeto). Alternativamente, por ejemplo, la muestra de referencia puede ser etiquetada por un experto. Si se construye un modelo a partir de las muestras de referencia, el sistema también etiqueta secuencias de estados en el modelo, basándose en el etiquetado de las muestras de referencia.

En tales realizaciones, posteriormente a asignar la muestra de prueba al modelo o a una de las muestras de referencia, el sistema puede recalcular la distancia entre la muestra de prueba y el modelo o la muestra de referencia, dando mayor peso a una o más unidades de voz que se sabe que son más indicativas que otras con respecto a la condición fisiológica particular que se está evaluando. El sistema puede decidir, a continuación si generar una alerta en respuesta a la distancia recalculada, en lugar de decidir en respuesta a la distancia original que se ha calculado durante la asignación. Al recalcular la distancia, el sistema no cambia la asignación original, es decir, cada vector de características de muestra de prueba permanece asignado al mismo vector de características de muestra de referencia o estado de modelo.

Alternativa o adicionalmente, posteriormente a la asignación de la muestra de prueba al modelo o a una de las muestras de referencia, el sistema puede recalcular la distancia entre la muestra de prueba y el modelo o la muestra de referencia, utilizando diferentes funciones de distancia local de las que se utilizaron para la asignación. También en este caso, el sistema no cambia la asignación original, sino que, más bien, solo recalcula la distancia.

Por ejemplo, el sistema puede modificar las funciones de distancia local para tener en cuenta una o más características que no se han utilizado para realizar la asignación, o para dar mayor peso a ciertas características. En la invención tal como se define en las reivindicaciones adjuntas, las características que se enfatizan por el sistema son aquellas que se sabe que son más indicativas que otras con respecto a la condición fisiológica particular que se está evaluando. (Un ejemplo de una característica más indicativa es la variación del tono, que tiende a disminuir con el inicio de, o un deterioro con respecto a, ciertas enfermedades). El sistema también modifica las funciones de distancia local de manera que una o más características tienen menos peso, o no contribuyen en absoluto a la distancia local.

Descripción del sistema

Se hace referencia inicialmente a la Fig. 1, que es una ilustración esquemática de un sistema 20 para evaluar el estado fisiológico de un sujeto 22, según algunas realizaciones de la presente invención.

5 El sistema 20 comprende un dispositivo 32 de recepción de audio, tal como un teléfono móvil, una tableta, un ordenador portátil, un ordenador de escritorio, un asistente personal controlado por voz (tal como un dispositivo Eco de Amazon™ o Google Home™), o un dispositivo de altavoz inteligente, que es utilizado por el sujeto 22. El dispositivo 32 comprende un sensor 38 de audio (por ejemplo, un micrófono), que convierte ondas sonoras en señales eléctricas analógicas. El dispositivo 32 comprende además un procesador 36 y otros circuitos que comprenden, por ejemplo, un convertidor 42 analógico-digital (A/D) y/o una interfaz de red, tal como un controlador 10 34 de interfaz de red (NIC). Típicamente, el dispositivo 32 comprende además una memoria digital (o "dispositivo de almacenamiento"), una pantalla (por ejemplo, una pantalla táctil) y/u otros componentes de interfaz de usuario, tal como un teclado. En algunas realizaciones, el sensor 38 de audio (y, opcionalmente, el convertidor 42 A/D) pertenece a una unidad que es externa al dispositivo 32. Por ejemplo, el sensor 38 de audio puede pertenecer a unos auriculares que se conectan al dispositivo 32 mediante una conexión cableada o inalámbrica, tal como una 15 conexión Bluetooth.

El sistema 20 comprende además un servidor 40, que comprende un procesador 28, una memoria 30 digital (o "dispositivo de almacenamiento"), tal como un disco duro o un disco flash, y/u otros circuitos que comprenden, por ejemplo, un convertidor A/D y/o una interfaz de red, tal como un controlador 26 de interfaz de red (NIC). El servidor 20 40 puede comprender además una pantalla, un teclado y/o cualquier otro componente de interfaz de usuario adecuado. Típicamente, el servidor 40 está situado a distancia del dispositivo 32, por ejemplo, en un centro de control, y el servidor 40 y el dispositivo 32 se comunican entre sí, a través de sus respectivas interfaces de red, a través de una red 24, que puede incluir una red celular y/o Internet.

El sistema 20 está configurado para evaluar el estado fisiológico del sujeto procesando una o más señales de voz (también denominadas en la presente memoria "muestras de voz") recibidas del sujeto, como se describe en detalle 25 a continuación. Típicamente, el procesador 36 del dispositivo 32 y el procesador 28 del servidor 40 realizan cooperativamente la recepción y el procesamiento de al menos algunas de las muestras de voz. Por ejemplo, cuando el sujeto habla al dispositivo 32, las ondas sonoras de voz del sujeto pueden convertirse en una señal analógica mediante el sensor 38 de audio, que a su vez puede muestrearse y digitalizarse mediante el convertidor 42 A/D. (En general, la voz del sujeto puede muestrearse a cualquier velocidad adecuada, tal como una velocidad de 30 entre 8 y 45 kHz). La señal de voz digital resultante puede ser recibida por el procesador 36. El procesador 36 puede comunicar, a continuación la señal de voz, a través de la NIC 34, al servidor 40, de tal manera que el procesador 28 reciba la señal de voz a través de la NIC 26. Posteriormente, el procesador 28 puede procesar la señal de voz.

Típicamente, en el procesamiento de la voz del sujeto, el procesador 28 compara una muestra de prueba, producida por el sujeto mientras que el estado fisiológico del sujeto era desconocido, con una muestra de referencia, producida 35 mientras que el estado fisiológico del sujeto era conocido (por ejemplo, se ha considerado por un médico que era estable), o con un modelo construido a partir de múltiples de tales muestras de referencia. Por ejemplo, el procesador 28 puede calcular una distancia entre la muestra de prueba y la muestra de referencia o el modelo.

Basándose en el procesamiento de las muestras de voz del sujeto, el procesador 28 puede generar una salida que indica el estado fisiológico del sujeto. Por ejemplo, el procesador 28 puede comparar la distancia mencionada 40 anteriormente con un umbral y, en respuesta a esta comparación, generar una alerta, tal como una alerta de audio o visual, que indica un deterioro en la condición fisiológica del sujeto. Opcionalmente, tal alerta puede incluir una descripción del estado del sujeto; por ejemplo, la alerta puede indicar que los pulmones del sujeto están "húmedos", es decir, parcialmente llenos de fluido. Alternativamente, si las muestras de voz del sujeto indican que el estado del sujeto es estable, el procesador 28 puede generar una salida que indica que el estado del sujeto es estable.

45 Para generar la salida, el procesador 28 puede realizar una llamada o enviar un mensaje (por ejemplo, un mensaje de texto) al sujeto, al médico del sujeto y/o a un centro de monitorización. Alternativa o adicionalmente, el procesador 28 puede comunicar la salida al procesador 36, y el procesador 36 puede comunicar, a continuación la salida al sujeto, por ejemplo, mostrando un mensaje en la pantalla del dispositivo 32.

50 En otras realizaciones, el procesador 36 y el procesador 28 realizan cooperativamente el procesamiento de señales de voz mencionado anteriormente. Por ejemplo, el procesador 36 puede extraer vectores de características acústicas de las muestras de voz (como se describe adicionalmente a continuación), y comunicar estos vectores al procesador 28. El procesador 28 puede procesar, a continuación los vectores como se describe en la presente memoria. Alternativamente, el procesador 28 puede recibir (del procesador 36, de uno o más procesadores distintos, y/o directamente) una o más muestras de voz de referencia producidas por el sujeto 22 y/o por uno o más sujetos 55 distintos. Basándose en estas muestras, el procesador 28 puede calcular al menos un modelo de voz, o una pluralidad de vectores de características de muestra de referencia. El procesador 28 puede comunicar, a continuación el modelo, o los vectores de características de muestra de referencia, al procesador 36. Basándose en estos datos obtenidos del procesador 28, el procesador 36 puede procesar las muestras de prueba del sujeto 22 como se describe en la presente memoria. (Opcionalmente, el procesador 36 puede comunicar la distancia

mencionada anteriormente al procesador 28. El procesador 28 puede comparar, a continuación la distancia con el umbral mencionado anteriormente y, si es apropiado, generar una alerta). Como otra alternativa más, la técnica de diagnóstico completa descrita en la presente memoria puede ser realizada por el procesador 36, de manera que el sistema 20 no necesariamente necesita incluir el servidor 40.

5 A pesar de lo anterior, el resto de la presente descripción, por simplicidad, supone generalmente que el procesador 28, también denominado a continuación simplemente "el procesador", realiza todo el procesamiento.

10 En algunas realizaciones, el dispositivo 32 comprende un teléfono analógico que no incluye un convertidor A/D o un procesador. En tales realizaciones, el dispositivo 32 envía la señal de audio analógica desde el sensor 38 de audio al servidor 40 a través de una red telefónica. Típicamente, en la red telefónica, la señal de audio se digitaliza, se comunica digitalmente, y después se convierte de nuevo a analógica antes de llegar al servidor 40. Por consiguiente, el servidor 40 puede comprender un convertidor A/D, que convierte la señal de audio analógica entrante - recibida a través de una interfaz de red telefónica adecuada - en una señal de voz digital. El procesador 28 recibe la señal de voz digital desde el convertidor A/D, y, a continuación, procesa la señal como se describe en la presente memoria. Alternativamente, el servidor 40 puede recibir la señal de la red telefónica antes de que la señal se convierta de nuevo a analógica, de manera que el servidor no necesariamente necesita incluir un convertidor A/D.

15 Típicamente, el servidor 40 está configurado para comunicarse con múltiples dispositivos que pertenecen a múltiples sujetos diferentes, y para procesar las señales de voz de estos múltiples sujetos. Típicamente, la memoria 30 almacena una base de datos en donde los datos relevantes para el procesamiento de muestras de voz descrito en la presente memoria (por ejemplo, una o más muestras de voz de referencia o vectores de características extraídos de las mismas, uno o más modelos de voz, y/o una o más distancias umbral) se almacenan para los sujetos. La memoria 30 puede ser interna al servidor 40, como se muestra en la Fig. 1, o externa al servidor 40. Para realizaciones en donde el procesador 36 procesa la voz del sujeto, una memoria que pertenece al dispositivo 32 puede almacenar los datos relevantes para el sujeto.

20 El procesador 28 puede estar incorporado como un único procesador, o como un conjunto de procesadores en red o agrupados cooperativamente. Por ejemplo, un centro de control puede incluir una pluralidad de servidores interconectados que comprenden procesadores respectivos, que realizan cooperativamente las técnicas descritas en la presente memoria. En algunas realizaciones, el procesador 28 pertenece a una máquina virtual.

25 En algunas realizaciones, la funcionalidad del procesador 28 y/o del procesador 36, como se describe en la presente memoria, se implementa únicamente en hardware, por ejemplo, utilizando uno o más circuitos integrados de aplicación específica (ASIC) o matrices de puertas programables en campo (FPGA). En otras realizaciones, la funcionalidad del procesador 28 y del procesador 36 se implementa al menos parcialmente en software. Por ejemplo, en algunas realizaciones, el procesador 28 y/o el procesador 36 se realizan como un dispositivo informático digital programado que comprende al menos una unidad central de procesamiento (CPU) y una memoria de acceso aleatorio (RAM). El código de programa, que incluye programas de software, y/o datos se cargan en la RAM para su ejecución y procesamiento por la CPU. El código y/o datos de programa pueden descargarse al procesador de forma electrónica, a través de una red, por ejemplo. Alternativa o adicionalmente, el código y/o datos de programa pueden proporcionarse y/o almacenarse en medios tangibles no transitorios, tales como memoria magnética, óptica o electrónica. Dicho código y/o datos de programa, cuando se proporcionan al procesador, producen una máquina u ordenador de propósito especial, configurado para realizar las tareas descritas en la presente memoria.

30 40 Construcción de un modelo estadístico paramétrico

Se hace referencia ahora a la Fig. 2, que es una ilustración esquemática de una construcción de un modelo 46 de voz, según algunas realizaciones de la presente invención.

45 En algunas realizaciones, el procesador 28 (Fig. 1) construye al menos un modelo 46 estadístico paramétrico a partir de una o más muestras 44 de voz de referencia que se han adquirido del sujeto 22. El procesador utiliza, a continuación el modelo 46 para evaluar la voz subsiguiente del sujeto.

50 En particular, el procesador recibe primero muestras 44, por ejemplo, a través del dispositivo 32, como se ha descrito anteriormente con referencia a la Fig. 1. En general, las muestras de voz de referencia se producen por el sujeto mientras que el estado fisiológico del sujeto es conocido. Por ejemplo, las muestras de voz de referencia pueden producirse mientras el estado fisiológico del sujeto se considera, por un médico, que es estable con respecto a una afección fisiológica particular. Como ejemplo particular, para un sujeto que padece una afección fisiológica tal como edema pulmonar o derrame pleural, las muestras de referencia pueden producirse mientras que se considera que los pulmones del sujeto están libres de fluido. Alternativamente, las muestras de voz de referencia pueden producirse mientras el estado fisiológico del sujeto es inestable con respecto a una condición fisiológica particular, por ejemplo, mientras los pulmones del sujeto están húmedos.

55 A continuación, basándose en las muestras recibidas, el procesador construye el modelo 46. En particular, el procesador extrae típicamente vectores de características acústicas de las muestras de referencia (como se describe a continuación con referencia a la Fig. 3 para la muestra de prueba), y, a continuación, construye el modelo 46 a partir de los vectores. El modelo puede almacenarse, por ejemplo, en la memoria 30 (Fig. 1).

El modelo 46 incluye uno o más estados 48 acústicos (por ejemplo, APU y/o unidades acústicas sintéticas) que se presentan en las muestras de voz de referencia. Los estados 48 acústicos están asociados con las respectivas funciones 50 de distancia local. Dado cualquier vector de característica acústica "v" dentro del dominio de las funciones 50, la función de distancia local de cada estado acústico devuelve una distancia local que indica un grado de correspondencia entre el vector de característica acústica dado y el estado acústico. El modelo 46 incluye además las transiciones 52 entre los estados acústicos que se presentan en las muestras de voz de referencia; estas transiciones se denominan en la presente memoria "transiciones permitidas". En algunas realizaciones, el modelo 46 define además las respectivas distancias 54 de transición para las transiciones.

Por ejemplo, la Fig. 2 muestra un fragmento de ejemplo de un modelo de voz, que incluye (i) un primer estado acústico s_1 , que tiene una primera función de distancia local $d_1(v)$, (ii) un segundo estado acústico s_2 , que tiene una segunda función de distancia local $d_2(v)$ y (iii) un tercer estado acústico s_3 , que tiene una tercera función de distancia local $d_3(v)$. s_1 pasa a s_2 con una distancia de transición t_{12} y a s_3 con una distancia de transición t_{13} . s_3 pasa a s_1 con una distancia de transición t_{31} .

Como ejemplo simplificado específico, si el fragmento mostrado en la Fig. 2 representa la palabra "Bobby" tal como hablado por el sujeto en las muestras de voz de referencia, s_1 puede corresponder al fonema "b", s_3 puede corresponder al fonema "aw", y s_2 puede corresponder al fonema "ee". (Se observa que típicamente, en la práctica, al menos algunos fonemas están representados por una secuencia de múltiples estados).

En algunas realizaciones, cada uno de los estados acústicos está asociado con una respectiva función de densidad de probabilidad (PDF) multidimensional, de la que se deriva implícitamente la distancia local entre el vector de características "v" dado y el estado acústico. En particular, la PDF proporciona una probabilidad estimada de que el vector de características acústicas dado corresponda al estado acústico (es decir, que el vector de características dado se derive de la voz que se ha producido mientras el sistema de producción de voz del sujeto estaba en el estado físico correspondiente al estado acústico), y la distancia local se deriva de esta probabilidad estimada. Por ejemplo, la función de distancia local de cada estado acústico puede devolver un valor que depende del logaritmo negativo de la probabilidad estimada. Este valor puede ser, por ejemplo, el propio registro negativo, o un múltiplo del registro negativo.

Como ejemplo específico, cada estado acústico puede estar asociado con una PDF Gaussiana, de manera que la distancia local, cuando se calcula como una probabilidad logarítmica negativa, es la suma de los cuadrados de las diferencias entre los componentes del vector de características y los componentes correspondientes de la media de la distribución, ponderada por las inversas de las varianzas correspondientes de la distribución.

En otras realizaciones, las distancias locales se derivan de consideraciones teóricas de la información; un ejemplo de una medida de distancia que se basa en tales consideraciones es la medida de distancia de Itakura-Saito, que se menciona a continuación con referencia a la Fig. 5. Alternativamente, para realizaciones en donde se construyen tanto un modelo estable como un modelo inestable, las distancias locales pueden derivarse de consideraciones de discriminación de clase, en que las distancias locales pueden seleccionarse para discriminar mejor entre las muestras de referencia estables e inestables. Alternativamente, las distancias locales pueden derivarse de consideraciones heurísticas.

Típicamente, las distancias 54 de transición se basan en probabilidades de transición respectivas, como se estima a partir de las muestras de voz de referencia; por ejemplo, cada distancia de transición puede ser el logaritmo negativo de una probabilidad de transición respectiva.

En general, los parámetros del modelo (por ejemplo, los parámetros de las PDF mencionadas anteriormente) y las probabilidades de transición pueden estimarse a partir de las muestras de voz de referencia utilizando cualquier técnica adecuada, tal como el algoritmo de Baum-Welch, que se describe, por ejemplo, en la sección 6.4.3 de L. Rabiner y B-H. Juang, *Fundamentals of Speech Recognition* (Fundamentos del reconocimiento de voz), Prentice Hall, 1993.

Asignación de una muestra de prueba al modelo

Se hace referencia ahora a la Fig. 3, que es una ilustración esquemática de una asignación de una muestra 56 de voz de prueba a un modelo de voz, según algunas realizaciones de la presente invención.

Después de la adquisición de las muestras de referencia, en un momento posterior, cuando se desconoce el estado fisiológico del sujeto, el procesador utiliza el modelo 46 para evaluar el estado fisiológico del sujeto.

En particular, el procesador recibe primero al menos una muestra 56 de voz de prueba producida por el sujeto mientras que el estado fisiológico del sujeto era desconocido. A continuación, el procesador calcula una pluralidad de vectores 60 de características de muestra de prueba que cuantifican características acústicas de diferentes partes 58 respectivas de la muestra 56. Las características acústicas pueden incluir, por ejemplo, una representación de la envolvente espectral de la parte 58, que incluye, por ejemplo, coeficientes de predicción lineal y/o coeficientes cepstrales. Los vectores 60 pueden incluir cualquier número adecuado de características; a modo de ejemplo, la Fig. 3 muestra un vector v de cinco dimensiones.

En general, cada parte 58 puede ser de cualquier duración adecuada, tal como, por ejemplo, entre 10 y 100 ms. (Típicamente, las partes son de igual duración, aunque algunas realizaciones pueden utilizar análisis síncrono de tono con partes de duración variable). En algunas realizaciones, las partes 58 se solapan entre sí. Por ejemplo, los vectores 60 pueden corresponder a respectivos puntos de tiempo "t", donde cada vector describe las características acústicas de la parte de la señal que ocupa el período de tiempo [t-T, t+T], donde T está, por ejemplo, entre 5 y 50 ms. Los puntos de tiempo sucesivos pueden estar separados entre sí entre 10 y 30 ms, por ejemplo.

Posteriormente al cálculo de los vectores de características, basándose en las funciones de distancia local y en las transiciones permitidas que se definen por el modelo 46, el procesador asigna la muestra de voz de prueba a una secuencia de distancia mínima de estados acústicos pertenecientes al modelo, asignando los vectores de características de muestra de prueba a los estados acústicos respectivos de manera que se minimice la distancia total entre los vectores de características de muestra de prueba y los estados acústicos respectivos. La distancia total se basa en las respectivas distancias locales entre los vectores de características de muestra de prueba y los estados acústicos a los que se asignan los vectores de características; por ejemplo, la distancia total puede basarse en la suma de las respectivas distancias locales.

Para explicar adicionalmente, como se ilustra en la Fig. 3, cada asignación de la muestra de voz de prueba al modelo asigna cada índice "j" de los vectores de características a un índice m(j) de los estados acústicos, de manera que el j-ésimo vector v_j se asigna al estado acústico s_{m(j)}. (s_{m(j)} puede ser cualquier estado acústico al que hay una transición permitida de s_{m(j-1)}.) La asignación de v_j a s_{m(j)} da lugar a una distancia local d_j = d_{m(j)}(v_j) entre v_j y s_{m(j)}. Por lo tanto, suponiendo N vectores de características de muestra de prueba, la muestra de prueba se asigna a una

secuencia de N estados, y la suma de las distancias locales para esta asignación es $\sum_{j=1}^N d_j$. La distancia total para la asignación se basa en $\sum_{j=1}^N d_j$. Por ejemplo, la distancia total puede definirse como $\sum_{j=1}^N d_j$, o, si se incluyen

distancias de transición en el modelo, como $\sum_{j=1}^N d_j + \sum_{j=1}^{N-1} t_{j(j+1)}$ donde t_{j(j+1)} es la distancia de transición de j-ésimo estado a j+1er estado. El procesador encuentra la secuencia de estados para los que esta distancia total se minimiza.

A modo de ejemplo, haciendo referencia de nuevo a la Fig. 2, y suponiendo que el procesador extrae una secuencia de seis vectores de características {v₁,v₂,v₃,v₄,v₅,v₆} a partir de la muestra de prueba, el procesador puede asignar la muestra de prueba a la secuencia de estado de distancia mínima {s₁,s₃,s₁,s₂,s₂,s₃}. La distancia total para esta asignación puede calcularse como d₁(v₁)+t₁₃+d₃(v₂)+t₃₁+d₁(v₃)+t₁₂+d₂(v₄)+t₂₂+d₂(v₅)+t₂₃+d₃(v₆).

En algunas realizaciones, para encontrar la asignación óptima de la muestra de prueba al modelo, el sistema utiliza el algoritmo de Viterbi, que se describe en la sección 6.4.2 de la referencia mencionada anteriormente a Rabiner y Juang.

Posteriormente, en respuesta a la asignación de la muestra de voz de prueba a la secuencia de distancia mínima de estados acústicos, el procesador genera una salida que indica el estado fisiológico del sujeto en el momento en donde se ha producido la muestra de prueba.

Por ejemplo, el procesador puede comparar la distancia total para la asignación óptima con un umbral predeterminado, y luego generar la salida en respuesta a la comparación. En particular, si las muestras de voz de referencia se adquirieron mientras el estado del sujeto era estable, puede generarse una alerta en respuesta a que la distancia total supera el umbral; a la inversa, si las muestras de voz de referencia se han adquirido mientras el estado del sujeto era inestable, puede generarse una alerta en respuesta a que la distancia total es menor que el umbral.

En algunas realizaciones, el procesador determina el umbral basándose en la distribución estadística de la distancia total sobre un número adecuado de asignaciones, que pueden realizarse para un único sujeto (en cuyo caso el umbral puede ser específico del sujeto), o para múltiples sujetos respectivos. En particular, si las asignaciones se realizan cuando se sabe que el estado del sujeto o sujetos es estable, el umbral puede establecerse de manera que la distancia total sea menor que el umbral en un porcentaje suficientemente grande (por ejemplo, más del 98 %) de las asignaciones. Por el contrario, si las asignaciones se realizan cuando se sabe que el estado del sujeto o sujetos es inestable, el umbral puede establecerse de manera que la distancia total exceda el umbral en un porcentaje suficientemente grande de las asignaciones.

Alternativamente, el procesador puede construir dos modelos de voz: uno que utiliza muestras de voz de referencia adquiridas mientras que el estado del sujeto era estable, y otro que utiliza muestras adquiridas mientras que el estado del sujeto era inestable. La muestra de prueba puede, a continuación asignarse a una secuencia de estados de distancia mínima respectiva en cada uno de los modelos. Las respectivas distancias totales entre la muestra de prueba y los dos modelos pueden compararse entre sí, y puede generarse una salida en respuesta a la comparación. Por ejemplo, si la distancia entre la muestra de prueba y el modelo de estado estable supera la distancia entre la muestra de prueba y el modelo de estado inestable, puede generarse una alerta.

En algunas realizaciones, el sistema calcula las distancias totales respectivas, con referencia al mismo modelo o a diferentes modelos respectivos, para múltiples muestras de prueba. El sistema puede generar, a continuación una alerta en respuesta a las distancias, por ejemplo, en respuesta a una o más de las distancias que exceden un umbral.

5 En algunas realizaciones, las muestras de voz de referencia y la muestra de voz de prueba incluyen la misma expresión predeterminada. Por ejemplo, para adquirir las muestras de referencia, el dispositivo 32 (Fig. 1) puede (por ejemplo, en respuesta a instrucciones del servidor 40) solicitar al sujeto que pronuncie repetidamente una expresión particular. Posteriormente, para adquirir la muestra de prueba, se puede solicitar al sujeto de manera similar que suprima la misma expresión. Para solicitar al sujeto, el dispositivo puede reproducir la expresión, y solicitar (a través de un mensaje escrito o de audio) que el sujeto repita la expresión que se ha reproducido.
10 Alternativamente, por ejemplo, el contenido verbal de la expresión puede ser presentado en la pantalla del dispositivo, y se puede solicitar al sujeto que lea el contenido verbal.

15 En otras realizaciones, las muestras de voz de referencia incluyen voz libre del sujeto, es decir, voz cuyo contenido verbal no estaba predeterminado por el sistema 20. Por ejemplo, las muestras de voz de referencia pueden incluir voz conversacional normal del sujeto. A este respecto, se hace referencia ahora a la Fig. 4, que es una ilustración esquemática de una técnica para construir un modelo de voz a partir de múltiples modelos 64 de unidades de voz, según algunas realizaciones de la presente invención.

20 La Fig. 4 representa una muestra 61 de referencia, que incluye voz libre del sujeto. En algunas realizaciones, dada una muestra de este tipo, el procesador construye el modelo 46 identificando múltiples unidades 62 de voz diferentes en la voz libre, construyendo modelos 64 de unidades de voz respectivos para las unidades de voz identificadas (tal como se ha descrito anteriormente con referencia a la Fig. 2 para el modelo 46), y, a continuación, construyendo el modelo 46 mediante la concatenación de modelos 64 de unidades de voz, de tal manera que el modelo de voz representa una concatenación particular de las unidades de voz identificadas. Cada unidad de voz puede incluir una o más palabras, APU y/o unidades acústicas sintéticas.

25 Por ejemplo, suponiendo que la muestra de referencia incluye la frase "Yo estoy intentando todo el día alcanzarla, pero su línea está ocupada", el procesador puede identificar las unidades de voz "intentar", "alcanzar" y "línea", y construir modelos de unidades de voz respectivos para estas unidades de voz. Posteriormente, el procesador puede construir el modelo 46 concatenando los modelos de unidad de voz, de manera que, por ejemplo, el modelo representa la expresión "intentar alcanzar línea".

30 Para identificar las unidades 62 de voz, el procesador puede utilizar cualquiera de los algoritmos para el reconocimiento de voz conectado de vocabulario grande, independiente del hablante, descritos en los capítulos 7-8 de la referencia mencionada anteriormente a Rabiner y Juang. Un ejemplo de tal algoritmo es el algoritmo de programación dinámica de una etapa, descrito en la sección 7.5 de Rabiner y Juang, y descrito adicionalmente en Ney, Hermann, "The use of a one-stage dynamic programming algorithm for connected word recognition" (El uso de un algoritmo de programación dinámica de una etapa para el reconocimiento de palabras conectadas.), Transacciones IEEE en procesamientos acústicos, de voz y de señales 32.2 (1984) : 263-271. Para identificar fonemas u otras subpalabras, estos algoritmos pueden utilizarse en combinación con técnicas para el reconocimiento de subpalabras, tales como las descritas en las Secciones 8.2-8.4 de Rabiner y Juang. Un modelo de lenguaje, descrito en las Secciones 8.5-8.7 de Rabiner y Juang, puede utilizarse para facilitar este reconocimiento de subpalabras.

35 Posteriormente, para adquirir la muestra de prueba, se puede solicitar al sujeto que pronuncie la expresión particular que está representada por el modelo 46. Por ejemplo, continuando con el ejemplo anterior, se puede solicitar al sujeto que pronuncie "intentar alcanzar la línea".

40 En otras realizaciones, los modelos de unidad de voz permanecen separados entre sí, es decir, no se realiza concatenación. En algunas de tales realizaciones, se le pide al sujeto que suprima cualquier expresión predeterminada que incluya al menos una de las unidades de voz para las que se han construido los modelos de unidades de voz. El procesador identifica cada una de esas unidades de voz en la expresión, y, a continuación, procesa cada unidad de voz por separado. (Típicamente, el procesador identifica cada una de las unidades de voz utilizando los modelos de unidades de voz en combinación con un HMM de voz general, que representa toda la voz aparte de las unidades de voz para las que se han construido los modelos de voz).

45 En otras de tales realizaciones, el procesador recibe voz libre del sujeto para la muestra de prueba. El procesador identifica, además, en la muestra de prueba, una o más partes que incluyen unidades 62 de voz, respectivamente. Por ejemplo, si la muestra de prueba incluye la frase "Alinear y dejar de intentar alcanzar la parte delantera", el procesador puede identificar las partes de la muestra de prueba que incluyen "intentar", "alcanzar" y "línea". (Para identificar el contenido verbal de la voz libre de muestras de prueba, el procesador puede utilizar cualquiera de los algoritmos independientes del hablante descritos anteriormente).

50 Posteriormente, el procesador asigna las partes de muestra de prueba a los modelos respectivos de las unidades de voz, identificando, para cada una de las partes, el modelo de unidades de voz que se ha construido para la unidad de voz incluida en la parte, y, a continuación, realizando una asignación de distancia mínima de la parte al modelo de unidades de voz correspondiente. Por ejemplo, el procesador puede asignar la parte de muestra de prueba

"intentar" al modelo que se ha construido para la unidad de voz "intentar", "alcanzar" al modelo que se ha construido para "alcanzar" y "línea" al modelo que se ha construido para "línea".

Posteriormente, en respuesta a la asignación de las partes de muestra de prueba a los modelos de unidad de voz, el procesador genera una salida que indica el estado fisiológico del sujeto. Por ejemplo, el procesador puede calcular la suma de las distancias respectivas para las asignaciones, y, a continuación, generar una salida en respuesta a esta distancia. Por ejemplo, si el procesador calcula las distancias q_1 , q_2 , y q_3 para "intentar", "alcanzar" y "línea", respectivamente, el procesador puede generar una salida en respuesta a $q_1+q_2+q_3$.

Uso de una distancia total diferente para el diagnóstico

En algunas realizaciones, el procesador genera la salida no en respuesta a la distancia total que se ha minimizado en la asignación, sino más bien, a una distancia total diferente entre los vectores de características de muestra de prueba y los estados acústicos respectivos a los que se asignan los vectores. En otras palabras, el procesador puede asignar la muestra de prueba al modelo minimizando una primera distancia total, pero, a continuación, generar la salida en respuesta a una segunda distancia total que es diferente de la primera distancia total.

En algunas realizaciones, el procesador calcula la segunda distancia total ponderando las respectivas distancias locales por las respectivas ponderaciones, siendo al menos dos de las ponderaciones diferentes entre sí, y sumando después las distancias locales ponderadas. Por ejemplo, volviendo al ejemplo descrito anteriormente con referencia a la Fig. 2, en donde $\{v_1, v_2, v_3, v_4, v_5, v_6\}$ se asigna a $\{s_1, s_3, s_1, s_2, s_2, s_3\}$, el procesador puede calcular la segunda distancia total como $w_1*d_1(v_1)+t_{13}+w_3*d_3(v_2)+t_{31}+w_1*d_1(v_3)+t_{12}+w_2*d_2(v_4)+t_{22}+w_2*d_2(v_5)+t_{23}+w_3*d_3(v_6)$, en donde al menos dos de las ponderaciones $\{w_1, w_2, w_3\}$ son diferentes entre sí. Como ejemplo específico, si el estado acústico s_1 tiene más relevancia para el estado fisiológico del sujeto que los otros dos estados, w_1 puede ser mayor que cada uno de w_2 y w_3 .

Alternativa o adicionalmente, el procesador puede modificar las funciones de distancia local de los estados acústicos respectivos a los que se asignan los vectores de características. Utilizando las funciones de distancia local modificadas, el procesador puede calcular diferentes distancias locales entre los vectores de características de muestra de prueba y los estados acústicos respectivos a los que se asignan los vectores. El procesador puede, a continuación, calcular la segunda distancia total sumando estas nuevas distancias locales. Por ejemplo, para la asignación ejemplar descrita anteriormente, el procesador puede calcular la segunda distancia total como $d'_1(v_1)+t_{13}+d'_3(v_2)+\dots+d'_2(v_5)+t_{23}+d'_3(v_6)$, donde la notación "d'" indica una función de distancia local modificada.

Típicamente, las funciones de distancia local se modifican para dar un mayor peso a al menos una de las características acústicas cuantificadas en los vectores. Típicamente, las características acústicas seleccionadas para una mayor ponderación son aquellas que se sabe que son más relevantes para la condición fisiológica del sujeto que otras características.

Por ejemplo, la función de distancia local original puede devolver, para cualquier vector $[z_1 z_2.. z_K]$, el valor $\sum_{i=1}^K b_i$, donde $b_i = s_i(z_i - r_i)^2$ en donde cada r_i es una cantidad de referencia adecuada, y cada s_i es un peso, que puede ser 0 para algunos índices. En tales realizaciones, la función de distancia local modificada puede retornar $\sum_{i=1}^K c_i$ donde $c_i = s'_i(z_i - r_i)^2$, donde $\{s'_i\}$ son ponderaciones adecuadas que difieren de s_i para al menos algunos de los índices. Utilizando $\{s'_i\}$ que difieren de $\{s_i\}$, el procesador puede ajustar las ponderaciones relativas de las características. En algunos casos, la función modificada puede incluir un s'_i distinto de cero (y por lo tanto, un c_i distinto de cero) para al menos un índice para el que s_i (y por lo tanto, b_i) es cero, de manera que el procesador, al calcular la segunda distancia total, tiene en cuenta al menos una característica que no se ha utilizado en absoluto para realizar la asignación. (Se observa que, para eficiencia, el cálculo real de $\sum_{i=1}^K b_i$ y $\sum_{i=1}^K c_i$ puede saltar sobre cualquier término de valor cero).

En algunas realizaciones, la muestra de prueba del sujeto se asigna en un modelo no específico del sujeto, que se construye típicamente a partir de múltiples muestras de referencia producidas por otros sujetos que son inestables con respecto a la condición fisiológica del sujeto. (Opcionalmente, también se pueden utilizar una o más muestras de estado inestable del sujeto para construir el modelo). Posteriormente, se calcula una segunda distancia total entre la muestra de prueba y el modelo, como se ha descrito anteriormente. A continuación, el procesador puede generar una salida en respuesta a la segunda distancia total. Por ejemplo, si el modelo se construye a partir de muestras de referencia de estado inestable como se ha descrito anteriormente, el procesador puede generar una alerta en respuesta a que la segunda distancia total sea menor que un umbral.

Comparación directa

Como se ha observado anteriormente en la descripción general, en algunas realizaciones, el procesador compara directamente la muestra de voz de prueba con una muestra de referencia.

En particular, el procesador recibe primero la muestra de referencia, que, como se ha observado anteriormente, es producida por el sujeto mientras que el estado fisiológico del sujeto es conocido. Posteriormente, el procesador calcula una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de la muestra de voz de referencia, tal como se ha descrito anteriormente con referencia a la Fig. 3 para la muestra de prueba. Estas características pueden almacenarse en la memoria 30 (Fig. 1).

A continuación, en un momento posterior, el procesador recibe la muestra de prueba, que, como se ha observado anteriormente, es producida por el sujeto mientras que el estado fisiológico del sujeto es desconocido. El procesador extrae, a continuación vectores de características de muestra de prueba de la muestra de prueba, como se ha descrito anteriormente con referencia a la Fig. 3. Posteriormente, el procesador asigna la muestra de voz de prueba a la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia de manera que se minimiza una distancia total entre los vectores de características de muestra de prueba y los respectivos vectores de características de muestra de referencia bajo restricciones predefinidas.

Para más detalles con respecto a esta asignación, se hace referencia ahora a la Fig. 5, que es una ilustración esquemática de una asignación de una muestra de voz de prueba a una muestra de voz de referencia, según algunas realizaciones de la presente invención.

A modo de introducción, se observa que cualquier asignación de la muestra de prueba a la muestra de referencia, también denominado "alineamiento" de la muestra de prueba con la muestra de referencia, puede representarse por una secuencia de N pares de índices $\{(t_1, r_1), \dots, (t_N, r_N)\}$, donde cada índice t_i es el índice de un vector de características en la muestra de prueba, cada índice r_i es el índice de un vector de características en la muestra de referencia y, por lo tanto, cada par de índices (t_i, r_i) representa una correspondencia entre un vector de características de muestra de prueba y un vector de características de muestra de referencia. Por ejemplo, la correspondencia entre el décimo vector de características de muestra de prueba y el undécimo vector de características de muestra de referencia está representada por el par de índices (10, 11).

Típicamente, la secuencia de pares de índices debe satisfacer algunas restricciones predefinidas para que el alineamiento sea válido. Ejemplos de tales restricciones incluyen:

- monotonicidad y continuidad: $t_i \leq t_{i+1}$ y $r_i \leq r_{i+1}$, y $0 < (t_{i+1} + t_i) - (r_i + r_{i+1}) \leq 2$, para $i = 1, \dots, N-1$
- una pendiente restringida: $1 \leq t_{i+2} - t_i \leq 2$ y $1 \leq r_{i+2} - r_i \leq 2$, para $i = 1, \dots, N-2$
- condiciones límite: $t_1 = 1$, $r_1 = 1$, $t_N = M$, y $r_N = L$, donde la muestra de prueba incluye M vectores de características y la muestra de referencia incluye L vectores de características

Dado cualquier alineamiento particular, la distancia total D entre la muestra de prueba y la muestra de referencia

puede definirse como
$$D = \sum_{i=1}^N d(v_{t_i}^T, v_{r_i}^R) w_i$$
, donde $v_{t_i}^T$ es el t_i -ésimo vector característico de la muestra de

prueba, $v_{r_i}^R$ es el vector de características r_i -ésimo de la muestra de referencia, d es una distancia local entre los dos vectores de características que pueden utilizar cualquier medida de distancia adecuada (por ejemplo, la medida de distancia L1 o L2), y cada w_i es un peso que se aplica a d. En algunas realizaciones, $w_1 = 2$ y $w_i = (r_i + t_i) - (r_{i-1} + t_{i-1})$ para $i = 2, \dots, N$ de tal manera que la suma de las ponderaciones es M+L para cada alineamiento, eliminando así cualquier a priori tendencia entre los diferentes alineamientos. Alternativamente, la distancia total D puede derivarse de las distancias locales de cualquier otra manera adecuada.

Se ha observado que en el contexto de la presente solicitud, incluyendo las reivindicaciones, la "distancia" entre dos vectores puede definirse para incluir cualquier tipo de desviación, o alineamiento, de uno de los vectores en relación con el otro. Por lo tanto, la función de distancia local no devuelve necesariamente una distancia en el sentido

geométrico. Por ejemplo, puede no ser necesariamente cierto que $d(v_{t_i}^T, v_{r_i}^R) = d(v_{r_i}^R, v_{t_i}^T)$ y/o puede no ser necesariamente cierto que para tres vectores de características cualesquiera v_1, v_2 , y v_3 , $d(v_1, v_3) \leq d(v_1, v_2) + d(v_2, v_3)$. Un ejemplo de una medida de distancia no geométrica que puede utilizarse en realizaciones de la presente invención es la medida de distancia de Itakura-Saito entre vectores de coeficientes de predicción lineal (LPC), que se describe en la sección 4.5.4 de la referencia mencionada anteriormente a Rabiner y Juang.

Además de la introducción anterior, la Fig. 5 ilustra un alineamiento de la muestra de prueba con la muestra de referencia, que puede ser realizada por el procesador, por ejemplo, utilizando el algoritmo de alineamiento temporal dinámico (DTW), que se describe en la referencia mencionada anteriormente a Sakoe y Chiba. En particular, la Fig. 5 muestra una correspondencia, entre algunos de los vectores de características de muestra de prueba y los correspondientes de los vectores de características de muestra de referencia, resultante del alineamiento. Cada par

de vectores de características correspondientes tiene una distancia local asociada donde $d_i = d(v_{t_i}^T, v_{r_i}^R)$. De entre todos los alineamientos posibles, el procesador selecciona el alineamiento que minimiza la distancia D, por ejemplo,

utilizando un algoritmo de programación dinámica descrito en la sección 4.7 de la referencia mencionada anteriormente a Rabiner y Juang. (Se observa que el algoritmo DTW incluye un algoritmo de programación dinámica para encontrar el alineamiento óptimo).

5 (Para evitar cualquier confusión, se observa que los cuatro vectores de características de muestra de referencia mostrados en la Fig. 5 no son necesariamente los primeros cuatro vectores de características que pertenecen a la muestra de referencia. Por ejemplo, r_2 puede ser 2 y r_3 puede ser 4, de tal manera que el tercer vector de características de muestra de referencia no se asigna. De manera similar, los cuatro vectores de características de muestra de prueba mostrados en la Fig. 5 no son necesariamente los primeros cuatro vectores de características pertenecientes a la muestra de prueba).

10 En respuesta a la asignación de la muestra de voz de prueba a la muestra de voz de referencia, el procesador puede generar una salida que indica el estado fisiológico del sujeto en el momento en donde se ha adquirido la muestra de voz de prueba. Por ejemplo, el procesador puede comparar la distancia total D con un umbral predefinido adecuado, y generar una salida en respuesta a la comparación.

15 En algunas realizaciones, como se ha descrito anteriormente con referencia a la Fig. 2, la muestra de voz de referencia se produce mientras que el estado fisiológico del sujeto se considera estable con respecto a una condición fisiológica particular. En otras realizaciones, la muestra de voz de referencia se ha producido mientras que el estado fisiológico del sujeto se considera inestable. En otras realizaciones más, el procesador recibe dos muestras de voz de referencia: una muestra de voz de estado estable y una muestra de voz de estado inestable. El procesador asigna, a continuación la muestra de prueba a cada una de las muestras de voz de referencia, produciendo de este modo una
20 primera distancia a la muestra de voz de estado estable, y una segunda distancia a la muestra de voz de estado inestable. El procesador compara, a continuación las dos distancias entre sí, y genera una salida en respuesta a las mismas. Por ejemplo, si la segunda distancia es menor que la primera distancia, lo que indica que la muestra de prueba es más similar a la muestra de referencia de estado inestable, el procesador puede generar una alerta.

25 En algunas realizaciones, la muestra de voz de referencia y la muestra de voz de prueba incluyen la misma expresión predeterminada, como se ha descrito anteriormente con referencia a la Fig. 3. En otras realizaciones, la muestra de voz de referencia incluye voz libre del sujeto, y la muestra de voz de prueba incluye una pluralidad de unidades de voz que están incluidas en la voz libre. Por ejemplo, utilizando las técnicas descritas anteriormente con referencia a la Fig. 4, el procesador puede identificar múltiples unidades de voz diferentes en la voz libre del sujeto. El procesador puede, a continuación construir una expresión a partir de estas unidades de voz, y, a continuación
30 solicitar al sujeto que produzca la muestra de prueba pronunciando la expresión.

En algunas realizaciones, el sistema calcula múltiples distancias, con respecto a diferentes muestras de referencia respectivas, para muestras de prueba respectivas; el sistema puede generar, a continuación una alerta en respuesta a las múltiples distancias, por ejemplo, en respuesta a una o más de las distancias que exceden un umbral.

Uso de una distancia total diferente para el diagnóstico

35 En la invención, tal como se define en las reivindicaciones adjuntas, el procesador, posteriormente a la realización de la asignación de la muestra de prueba a la muestra de referencia, calcula otra distancia total diferente entre los vectores de características de la muestra de prueba y los vectores de características de la muestra de referencia a los que se asignan. El procesador genera, a continuación una salida en respuesta a esta otra distancia total.

El procesador selecciona primero la asignación que minimiza $\sum_{i=1}^N d(v_{t_i}^T, v_{r_i}^R)w_i$ como se ha descrito
40 anteriormente. Posteriormente, el procesador puede (sin cambiar la asignación) calcular $\sum_{i=1}^N d(v_{t_i}^T, v_{r_i}^R)u_i$, en donde al menos una de las ponderaciones nuevas u_i es diferente de la ponderación original w_i correspondiente. Dicho de otro modo, el procesador calcula otra suma ponderada de las distancias locales en donde las distancias locales son ponderadas por un nuevo conjunto de ponderaciones $\{u_i\}$ que difiere del conjunto original de ponderaciones $\{w_i\}$ en que, para al menos un índice i , u_i es diferente de w_i .

45 Las nuevas ponderaciones se seleccionan asociando los vectores de características de muestra de referencia con las respectivas APU, y seleccionando, a continuación las nuevas ponderaciones en respuesta a las APU. (En este contexto, se dice que un vector está asociado con una APU por el procesador si el procesador considera que el

vector ha sido extraído de la voz que está incluida en la APU). Por ejemplo, en respuesta a $v_{r_2}^R$ y $v_{r_3}^R$ estando asociada con una APU particular que se sabe que es más relevante que otras APU para la condición fisiológica del
50 sujeto, el procesador puede asignar un valor más alto a u_2 y u_3 , con respecto a las otras ponderaciones nuevas.

Para asociar los vectores de características de la muestra de referencia con las respectivas APU, el procesador puede aplicar cualquier algoritmo de reconocimiento de voz adecuado a la muestra de voz de referencia. Por ejemplo, el procesador puede utilizar cualquiera de los algoritmos para el reconocimiento de voz conectado de vocabulario grande, independiente del hablante, descritos en los capítulos 7-8 de la referencia mencionada

anteriormente a Rabiner y Juang, tal como el algoritmo de programación dinámica de una etapa.

En la invención tal como se define en las reivindicaciones adjuntas, al calcular la nueva distancia total, el procesador puede (sin cambiar la asignación) utilizar diferentes distancias locales. En otras palabras, el procesador puede

5 calcular la nueva distancia total como $\sum_{i=1}^N d'(v_{t_i}^T, v_{r_i}^R)w_i$ (o $\sum_{i=1}^N d'(v_{t_i}^T, v_{r_i}^R)u_i$, en donde d' es una función de distancia local diferente de la función original, de manera que al menos una de las nuevas distancias locales difiere de la distancia local original correspondiente, es decir, $d'(v_{t_i}^T, v_{r_i}^R)$ es diferente $d(v_{t_i}^T, v_{r_i}^R)$ para al menos un índice i .

10 Por ejemplo, para las nuevas distancias locales, el procesador puede utilizar una nueva medida de distancia que es diferente de la medida de distancia original. (Por ejemplo, el procesador puede utilizar la medida de distancia L1 en lugar de la medida de distancia L2). Alternativa o adicionalmente, el procesador puede calcular las nuevas distancias locales basándose en al menos una característica acústica que no ha contribuido a las primeras distancias locales. Por ejemplo, si la distancia local original no depende de los terceros elementos respectivos de los vectores (que pueden cuantificar cualquier característica acústica particular), el procesador puede modificar la función de distancia local de manera que la salida de la función depende de estos elementos.

Algoritmo ejemplar

15 Se hace referencia ahora a la Fig. 6, que es un diagrama de flujo para un algoritmo 66 ejemplar para evaluar una muestra de voz de prueba de un sujeto, según algunas realizaciones de la presente invención.

20 El algoritmo 66 comienza en una etapa 68 de recepción, en donde el procesador recibe una muestra de voz de prueba del sujeto. Después de la recepción de la muestra, el procesador extrae vectores de características de muestra de prueba de la muestra, en una etapa 70 de extracción. A continuación, el procesador comprueba, en una etapa 72 de comprobación, si está disponible un modelo de referencia adecuado. (Como se ha observado anteriormente con referencia a la Fig. 4, dicho modelo puede construirse a partir de muestras de referencia que se han recibido del sujeto, y/o a partir de muestras de referencia que se han recibido de otros múltiples sujetos). Por ejemplo, el procesador puede buscar un modelo adecuado consultando una base de datos que está almacenada en la memoria 30 (Fig. 1).

25 Posteriormente, si el procesador es capaz de encontrar un modelo de referencia adecuado, el procesador, en una primera etapa 78 de asignación, asigna los vectores de características de muestra de prueba a una secuencia de estados en el modelo de referencia de manera que se minimiza una primera distancia total entre los vectores y los estados, como se ha descrito anteriormente con referencia a la Fig. 3. Alternativamente, si el procesador no puede encontrar un modelo de referencia adecuado, el procesador, en una etapa 74 de recuperación, recupera una secuencia de vectores de características de muestra de referencia, que se han extraído previamente de una muestra de referencia del sujeto. Posteriormente, en una segunda etapa 76 de asignación, el procesador asigna los vectores de características de muestra de prueba a los vectores de características de muestra de referencia, de manera que se minimiza una primera distancia total entre las secuencias de vectores, tal como se ha descrito anteriormente con referencia a la Fig. 5.

35 Después de la primera etapa 78 de asignación o la segunda etapa 76 de asignación, el procesador, en una etapa 80 de cálculo de distancia, calcula una segunda distancia total entre (i) los vectores de características de muestra de prueba y (ii) el modelo de referencia o los vectores de características de muestra de referencia. Por ejemplo, como se ha descrito anteriormente con referencia a las figs. 4-5, el procesador puede, al calcular la segunda distancia total, cambiar las ponderaciones relativas de las distancias locales, y/o cambiar las propias distancias locales.

40 Posteriormente, en una etapa 82 de comparación, el procesador compara la segunda distancia total con un umbral. Si la segunda distancia total es mayor que (o, en algunos casos, tal como cuando las muestras de referencia corresponden a un estado inestable, menor que) el umbral, el procesador genera una alerta, en una etapa 84 de generación de alertas. De lo contrario, el algoritmo 66 puede terminar sin ninguna actividad adicional; alternativamente, el procesador puede generar una salida que indica que el estado del sujeto es estable.

45 Las realizaciones de la presente invención se han descrito con referencia particular a los ejemplos ilustrados. Sin embargo, se apreciará que se pueden realizar variaciones y modificaciones a los ejemplos descritos dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1.- Un método para indicar un estado fisiológico de un sujeto, que comprende:

5 obtener una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra (44) de voz de referencia, producida por un sujeto (22) por primera vez mientras que un estado fisiológico del sujeto, con respecto a una condición fisiológica, era conocido;

recibir al menos una muestra (56) de voz de prueba producida por el sujeto por segunda vez, mientras que el estado fisiológico del sujeto era desconocido;

10 calcular una pluralidad de vectores (60) de características de muestra de prueba que cuantifican las características acústicas de diferentes partes (58) respectivas de la muestra de voz de prueba; y

asignar la muestra de voz de prueba a la muestra de voz de referencia de manera que se minimice una primera distancia total entre la muestra de voz de prueba y la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, para minimizar la primera distancia total;

15 caracterizado por que el procedimiento comprende, además:

calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo la segunda distancia total diferente de la primera distancia total, y

20 en respuesta a la segunda distancia total, generar una salida que indica el estado fisiológico del sujeto en el segundo momento;

por que la primera distancia total es una primera suma ponderada de respectivas distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, en donde la primera suma ponderada de las distancias locales se ponderan por respectivas primeras ponderaciones;

25 por que el método comprende, además:

asociar los vectores de características de la muestra de referencia con las respectivas unidades fonéticas acústicas, y

30 seleccionar segundas ponderaciones en respuesta a una relevancia de las unidades fonéticas acústicas para la condición fisiológica, siendo al menos una de las segundas ponderaciones diferente de una correspondiente de las primeras ponderaciones; y

por que la segunda distancia total es una segunda suma ponderada de las respectivas distancias locales en donde las distancias locales son ponderadas por las segundas ponderaciones, respectivamente.

35 2.- El método según la reivindicación 1, en donde asignar la muestra de voz de prueba a la muestra de voz de referencia comprende asignar la muestra de voz de prueba a la muestra de voz de referencia utilizando un algoritmo de alineamiento temporal dinámico.

3.- El método según la reivindicación 1, en donde la muestra de voz de referencia se ha producido mientras que el estado fisiológico del sujeto era estable con respecto a la condición fisiológica.

40 4.- El método según la reivindicación 1, en donde asociar los vectores de características de muestra de referencia con las unidades fonéticas acústicas comprende asociar los vectores de características de muestra de referencia con las unidades fonéticas acústicas aplicando un algoritmo de reconocimiento de voz a la muestra de voz de referencia.

5.- Un método para indicar un estado fisiológico de un sujeto, que comprende:

45 obtener una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra (44) de voz de referencia, producida por un sujeto (22) por primera vez mientras que un estado fisiológico del sujeto, con respecto a una condición fisiológica, era conocido;

recibir al menos una muestra (56) de voz de prueba producida por el sujeto por segunda vez, mientras que el estado fisiológico del sujeto era desconocido;

calcular una pluralidad de vectores (60) de características de muestra de prueba que cuantifican las características acústicas de diferentes partes (58) respectivas de la muestra de voz de prueba; y

asignar la muestra de voz de prueba a la muestra de voz de referencia de manera que se minimice una primera distancia total entre la muestra de voz de prueba y la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, para minimizar la primera distancia total;

5 caracterizado por que el procedimiento comprende, además:

calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo la segunda distancia total diferente de la primera distancia total, y

10 en respuesta a la segunda distancia total, generar una salida que indica el estado fisiológico del sujeto en el segundo momento;

por que la primera distancia total se basa en las respectivas primeras distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba; y

15 por que la segunda distancia total se basa en segundas distancias locales respectivas entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo al menos una de las segundas distancias locales diferente de una correspondiente de las primeras distancias locales para enfatizar una o más de las características acústicas que son más indicativas que otras con respecto a la condición fisiológica.

6.- El método según la reivindicación 5,

20 en donde asignar la muestra de voz de prueba a la muestra de voz de referencia comprende calcular las primeras distancias locales utilizando una primera medida de distancia, y

en donde calcular la segunda distancia total comprende calcular las segundas distancias locales utilizando una segunda medida de distancia que es diferente de la primera medida de distancia.

25 7.- El método según la reivindicación 5, en donde calcular la segunda distancia total comprende calcular las segundas distancias locales basándose en al menos una de las características acústicas que no han contribuido a las primeras distancias locales.

8.- Un sistema para indicar un estado fisiológico de un sujeto, que comprende:

circuitos (26, 42); y

uno o más procesadores (28, 36), configurados para llevar a cabo de manera cooperativa un proceso que incluye:

30 obtener una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra (44) de voz de referencia, producida por un sujeto (22) por primera vez mientras que un estado fisiológico del sujeto, con respecto a una condición fisiológica, era desconocido,

35 recibir, a través de los circuitos, al menos una muestra (56) de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido,

calcular una pluralidad de vectores de características de muestra de prueba (60) que cuantifican las características acústicas de diferentes partes respectivas (58) de la muestra de voz de prueba, y

40 asignar la muestra de voz de prueba a la muestra de voz de referencia de manera que se minimice una primera distancia total entre la muestra de voz de prueba y la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, para minimizar la primera distancia total;

caracterizado por que el procedimiento comprende, además:

45 calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo la segunda distancia total diferente de la primera distancia total, y

en respuesta a la segunda distancia total, generar una salida que indica el estado fisiológico del sujeto en el segundo momento;

50 por que la primera distancia total es una primera suma ponderada de respectivas distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, en donde la primera

suma ponderada de las distancias locales se ponderan por respectivas primeras ponderaciones;

por que el proceso comprende, además:

asociar los vectores de características de la muestra de referencia con las respectivas unidades fonéticas acústicas, y

5 seleccionar segundas ponderaciones en respuesta a una relevancia de las unidades fonéticas acústicas para la condición fisiológica, siendo al menos una de las segundas ponderaciones diferente de una correspondiente de las primeras ponderaciones; y

por que la segunda distancia total es una segunda suma ponderada de las respectivas distancias locales en donde las distancias locales son ponderadas por los segundas ponderaciones, respectivamente.

10 9.- El sistema según la reivindicación 8, en donde asignar la muestra de voz de prueba a la muestra de voz de referencia incluye asignar la muestra de voz de prueba a la muestra de voz de referencia utilizando un algoritmo de alineamiento temporal dinámico.

15 10.- El sistema según la reivindicación 8, en donde asociar los vectores de características de muestra de referencia con las unidades fonéticas acústicas incluye asociar los vectores de características de muestra de referencia con las unidades fonéticas acústicas aplicando un algoritmo de reconocimiento de voz a la muestra de voz de referencia.

11.- Un sistema para indicar un estado fisiológico de un sujeto, que comprende:

circuitos (26, 42); y

uno o más procesadores (28, 36), configurados para llevar a cabo de manera cooperativa un proceso que incluye:

20 obtener una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra (44) de voz de referencia, producida por un sujeto (22) por primera vez mientras que un estado fisiológico del sujeto, con respecto a una condición fisiológica, era conocido

recibir, a través de los circuitos, al menos una muestra (56) de voz de prueba producida por el sujeto en un segundo momento, mientras que el estado fisiológico del sujeto era desconocido,

25 calcular una pluralidad de vectores (60) de características de muestra de prueba que cuantifican las características acústicas de diferentes partes (58) respectivas de la muestra de voz de prueba, y

asignar la muestra de voz de prueba a la muestra de voz de referencia de manera que se minimice una primera distancia total entre la muestra de voz de prueba y la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, para minimizar la primera distancia total;

30 caracterizado por que el procedimiento comprende, además:

calcular una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo la segunda distancia total diferente de la primera distancia total, y

35 en respuesta a la segunda distancia total, generar una salida que indica el estado fisiológico del sujeto en el segundo momento;

por que la primera distancia total se basa en las respectivas primeras distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba; y

40 por que la segunda distancia total se basa en segundas distancias locales respectivas entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo al menos una de las segundas distancias locales diferente de una correspondiente de las primeras distancias locales para enfatizar una o más de las características acústicas que son más indicativas que otras con respecto a la condición fisiológica.

45 12.- El sistema según la reivindicación 11, en donde asignar la muestra de voz de prueba a la muestra de voz de referencia incluye calcular las primeras distancias locales utilizando una primera medida de distancia, y en donde calcular la segunda distancia total incluye calcular las segundas distancias locales utilizando una segunda medida de distancia que es diferente de la primera medida de distancia.

50 13.- El sistema según la reivindicación 11, en donde calcular la segunda distancia total incluye calcular las segundas distancias locales basándose en al menos una de las características acústicas que no han contribuido a las primeras

distancias locales.

14.- Un producto de software informático para indicar un estado fisiológico de un sujeto, que comprende un medio tangible no transitorio legible por ordenador en donde se almacenan instrucciones de programa, instrucciones que, cuando son leídas por un procesador (28), hacen que el procesador:

5 obtenga una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra (44) de voz de referencia, producida por un sujeto (22) por primera vez mientras que un estado fisiológico del sujeto era conocido,

reciba al menos una muestra (56) de voz de prueba producida por el sujeto por segunda vez, mientras que el estado fisiológico del sujeto, con respecto a una condición fisiológica, era desconocido,

10 calcule una pluralidad de vectores (60) de características de muestra de prueba que cuantifican las características acústicas de diferentes partes (58) respectivas de la muestra de voz de prueba,

asigne la muestra de voz de prueba a la muestra de voz de referencia de manera que se minimice una primera distancia total entre la muestra de voz de prueba y la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, para minimizar la primera distancia total;

15 caracterizado por que las instrucciones hacen además que el procesador:

calcule una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo la segunda distancia total diferente de la primera distancia total, y calcule una segunda distancia total entre los vectores de características de muestra de referencia, y calcule una segunda distancia total, y

20 en respuesta a la segunda distancia total, genere una salida que indica el estado fisiológico del sujeto en el segundo momento;

por que la primera distancia total es una primera suma ponderada de respectivas distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, en donde la primera suma ponderada de las distancias locales se ponderan por respectivas primeras ponderaciones;

25 por que las instrucciones hacen además que el procesador:

asocie los vectores de características de la muestra de referencia con las respectivas unidades fonéticas acústicas, y

30 seleccione segundas ponderaciones en respuesta a una relevancia de las unidades fonéticas acústicas para la condición fisiológica, siendo al menos una de las segundas ponderaciones diferente de una correspondiente de las primeras ponderaciones; y

por que la segunda distancia total es una segunda suma ponderada de las respectivas distancias locales en donde las distancias locales son ponderadas por las segundas ponderaciones, respectivamente.

35 15.- Un producto de software informático para indicar un estado fisiológico de un sujeto, que comprende un medio tangible no transitorio legible por ordenador en donde se almacenan instrucciones de programa, instrucciones que, cuando son leídas por un procesador (28), hacen que el procesador:

40 obtenga una pluralidad de vectores de características de muestra de referencia que cuantifican características acústicas de diferentes partes respectivas de al menos una muestra (44) de voz de referencia, producida por un sujeto (22) por primera vez mientras que un estado fisiológico del sujeto, con respecto a una condición fisiológica, era conocido,

reciba al menos una muestra (56) de voz de prueba producida por el sujeto por segunda vez, mientras que el estado fisiológico del sujeto era desconocido,

45 calcule una pluralidad de vectores (60) de características de muestra de prueba que cuantifican las características acústicas de diferentes partes (58) respectivas de la muestra de voz de prueba, y

asigne la muestra de voz de prueba a la muestra de voz de referencia de manera que se minimice una primera distancia total entre la muestra de voz de prueba y la muestra de voz de referencia, asignando los vectores de características de muestra de prueba a los respectivos vectores de características de muestra de referencia, bajo restricciones predefinidas, para minimizar la primera distancia total;

50 caracterizado por que las instrucciones hacen además que el procesador:

calcule una segunda distancia total entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo la segunda distancia total diferente de la primera distancia total, y calcule una segunda distancia total entre los vectores de características de muestra de referencia, y calcule una segunda distancia total, y

5 en respuesta a la segunda distancia total, genere una salida que indica el estado fisiológico del sujeto en el segundo momento;

por que la primera distancia total se basa en las respectivas primeras distancias locales entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba; y

10 por que la segunda distancia total se basa en segundas distancias locales respectivas entre los vectores de características de muestra de prueba y los respectivos de los vectores de características de muestra de referencia a los que se asignan los vectores de características de muestra de prueba, siendo al menos una de las segundas distancias locales diferente de una correspondiente de las primeras distancias locales para enfatizar una o más de las características acústicas que son más indicativas que otras con respecto a la condición fisiológica.

15

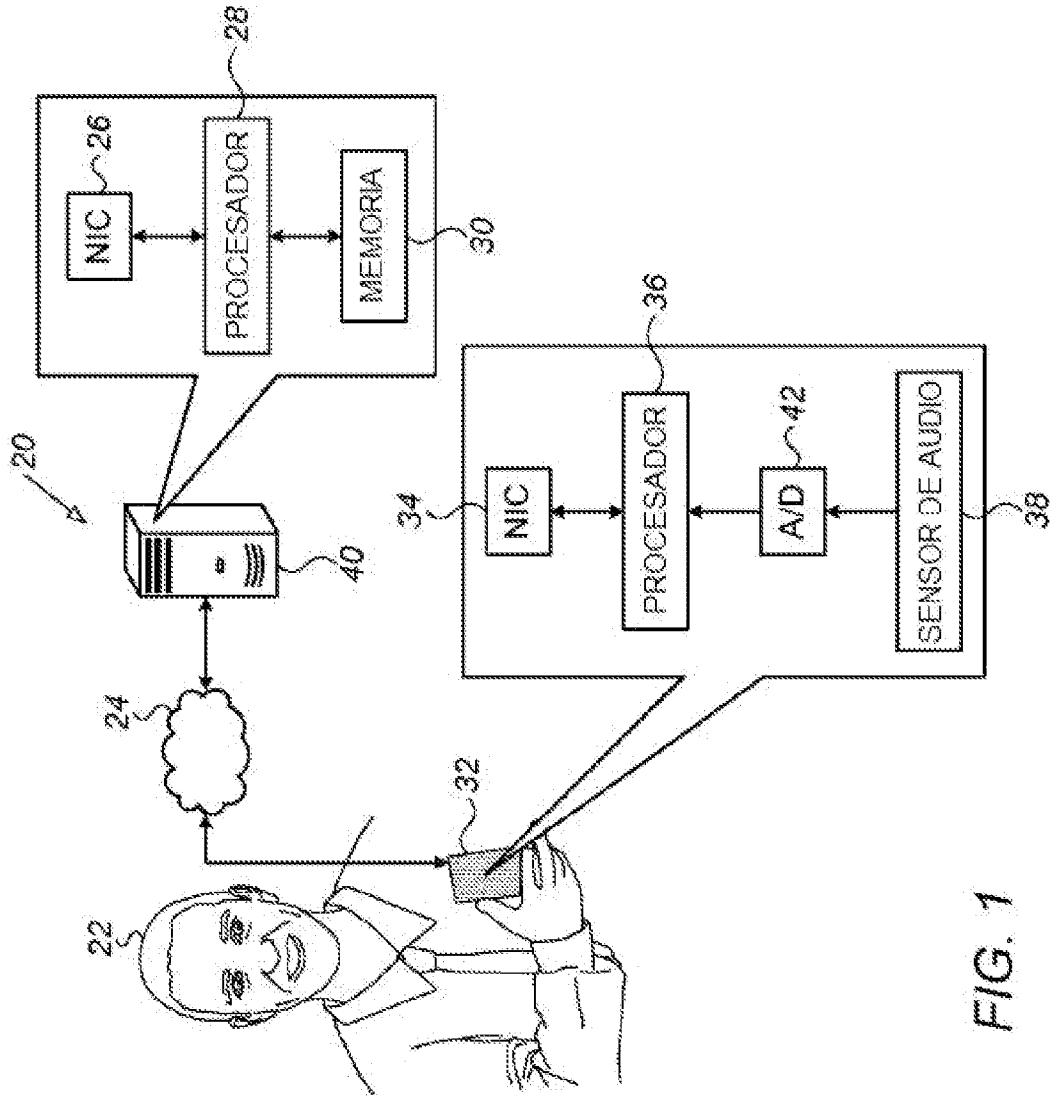
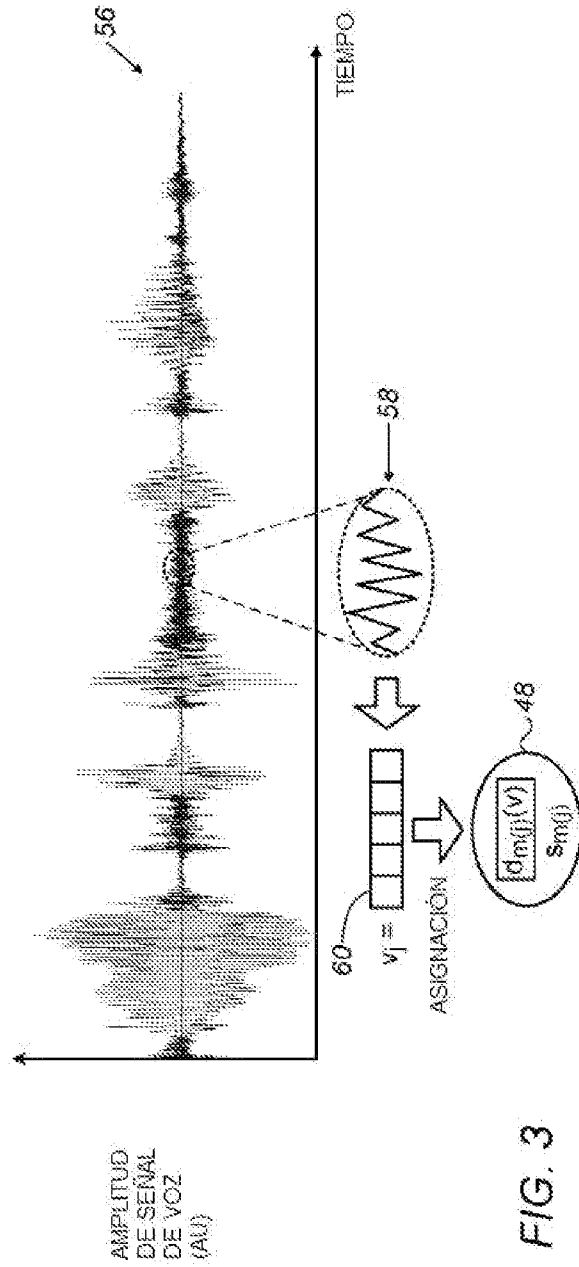
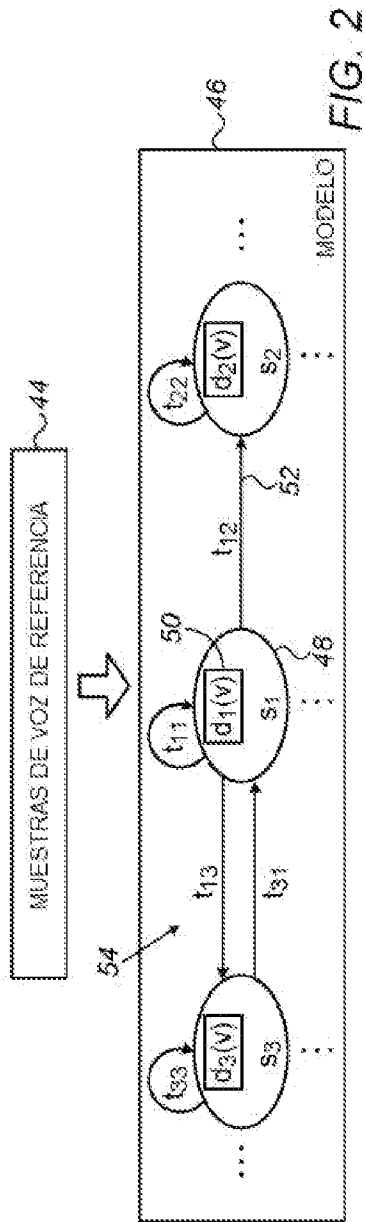


FIG. 1



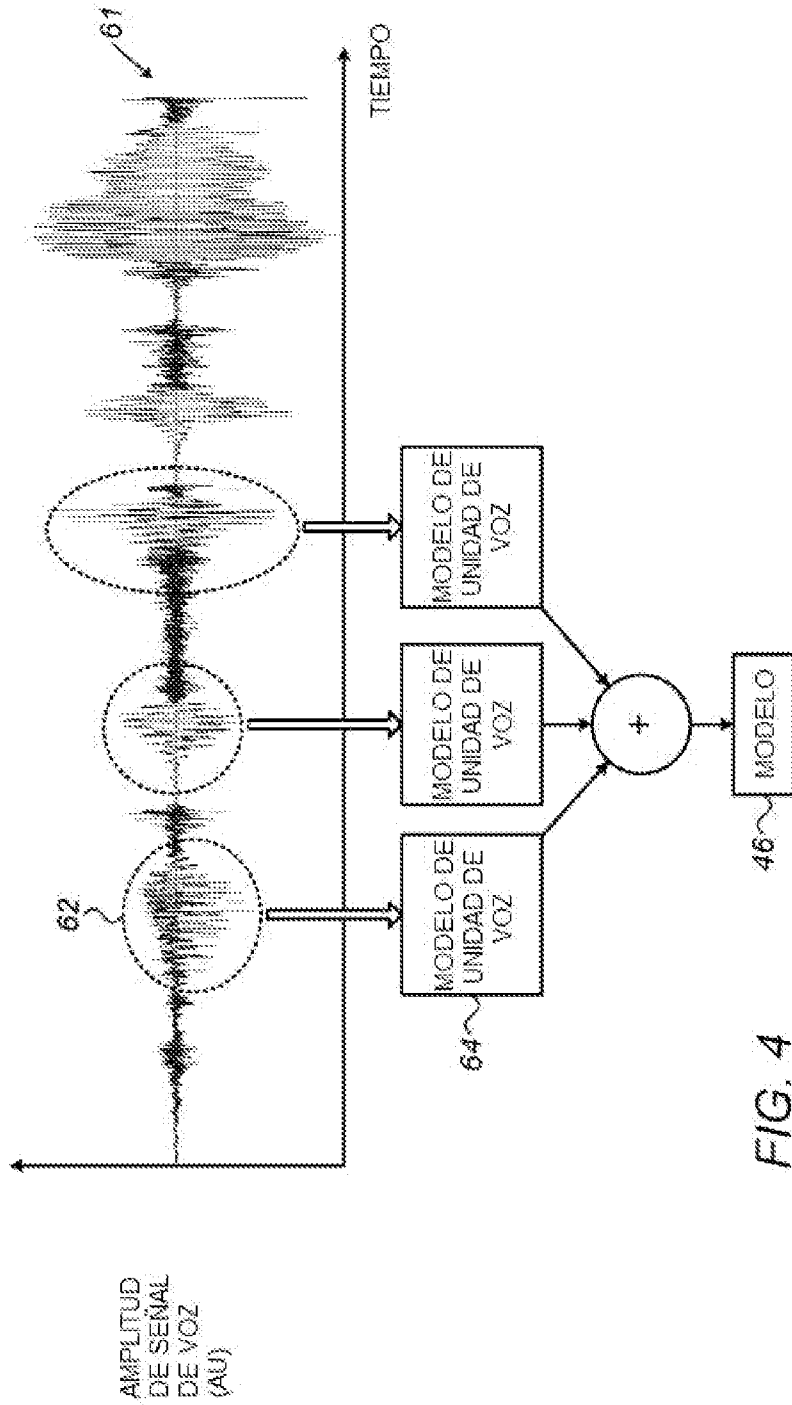


FIG. 4

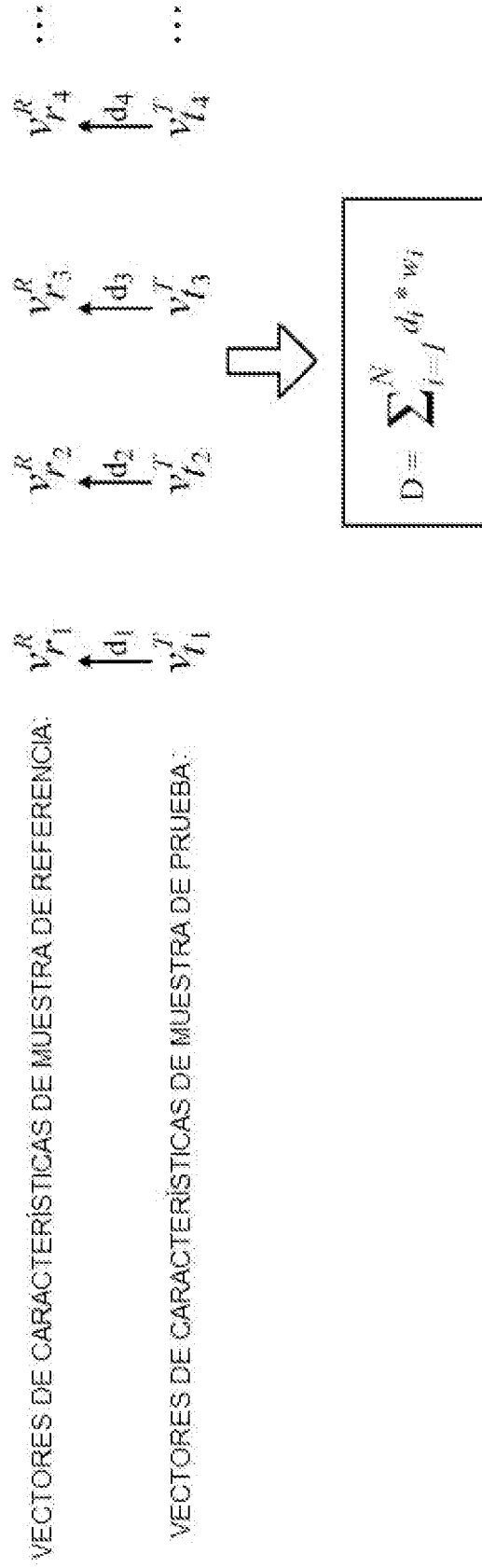


FIG. 5

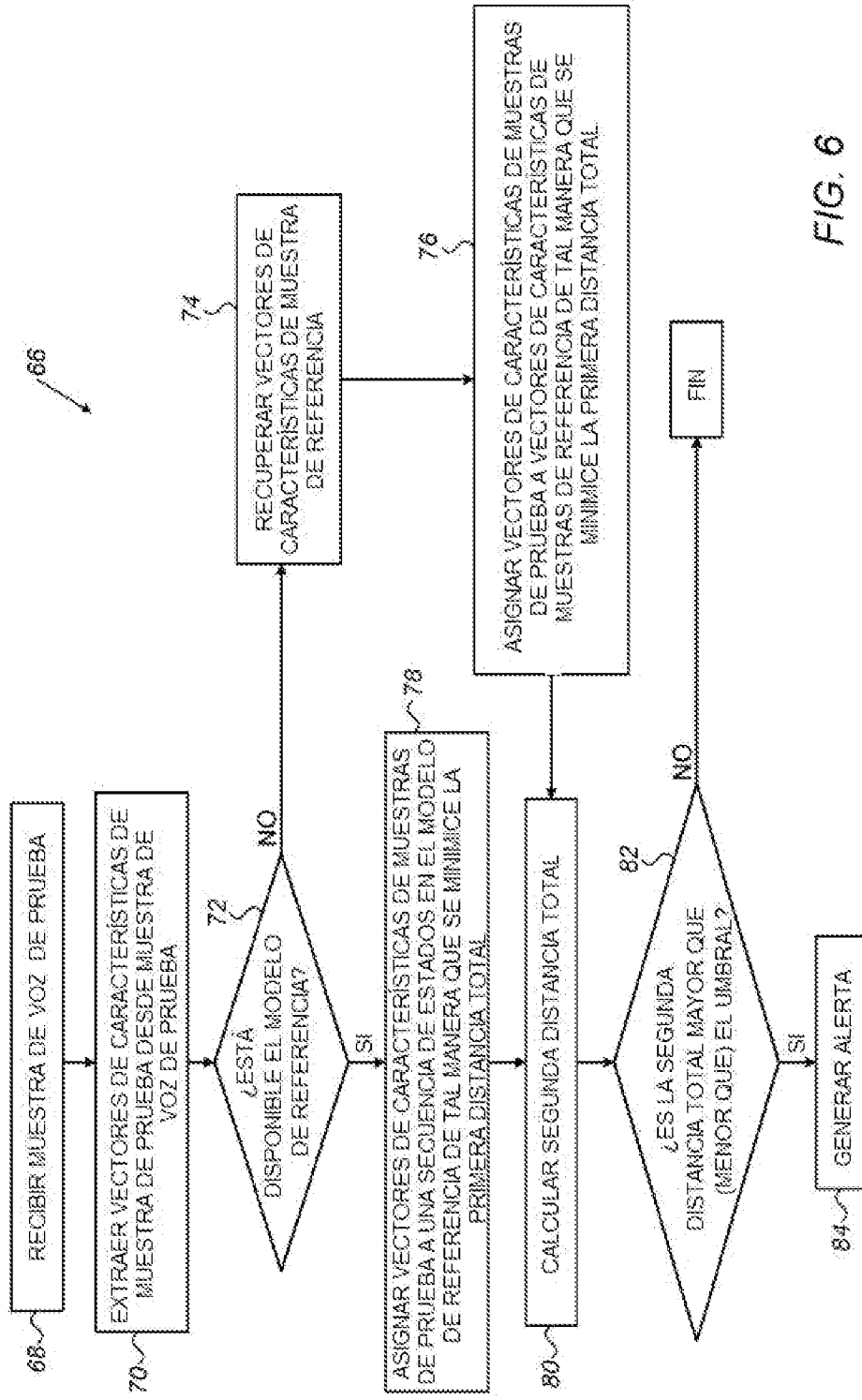


FIG. 6