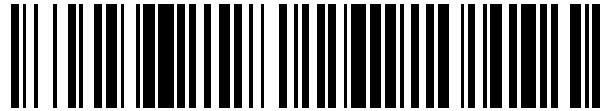


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 774 018**

51 Int. Cl.:

**H04B 17/30** (2015.01)

**H04M 9/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.06.2014 PCT/CN2014/081156**

87 Fecha y número de publicación internacional: **23.12.2015 WO15192395**

96 Fecha de presentación y número de la solicitud europea: **30.06.2014 E 14895085 (0)**

97 Fecha y número de publicación de la concesión europea: **06.11.2019 EP 3166239**

54 Título: **Método y sistema para evaluar la calidad de sonido de una voz humana**

30 Prioridad:

**17.06.2014 CN 201410269839**

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

**16.07.2020**

73 Titular/es:

**THE THIRD RESEARCH INSTITUTE OF MINISTRY OF PUBLIC SECURITY (50.0%)  
No. 1 Building Room 717, No. 76 Yueyang Rd.,  
Xuhui District  
Shanghai 200031, CN y  
SHANGHAI JINGHUI ELECTRONICS EQUIPMENT  
MINHANG CO., LTD. (50.0%)**

72 Inventor/es:

**RONG, LING;  
ZHAO, PENG;  
BAO, YIMING;  
TANG, DECHENG y  
HU, RONG**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 774 018 T3

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

**DESCRIPCIÓN**

Método y sistema para evaluar la calidad de sonido de una voz humana

**Antecedentes****Campo técnico**

- 5 La presente invención está relacionada con el campo de transmisión y detección de voz, y más particularmente con un método y sistema para evaluar la calidad de sonido de una voz humana.

**Técnica relacionada**

10 En la técnica anterior, requiere que la detección (el examen) o la prueba de rendimiento de transmisión de voz y de transmisión de video se realicen en el dispositivo utilizado por la transmisión de voz o video, con el objetivo de determinar que cumple con los requisitos de los indicadores. Cuando el resultado obtenido por estas detecciones o pruebas cumple con los requisitos de los indicadores, se puede asegurar que estos dispositivos de voz o video estén disponibles y sean fáciles de utilizar y tengan calidad garantizada, con el objetivo de realizar la producción y la aplicación práctica de estos dispositivos. Con el progreso continuo de cada técnica de dispositivo de transmisión de voz y video, el requisito de rendimiento del dispositivo y cada aspecto que coincide con el dispositivo cuando se utiliza el dispositivo también se actualiza y progresa continuamente. Por lo tanto, para la técnica de detección del dispositivo (examen), también se mejora continuamente, con el objetivo de mejorar la exactitud y precisión del resultado de detección y examen. El sistema de intercomunicación del edificio utilizado para controlar las entradas del edificio residencial se toma como ejemplo, la detección (examen) o la prueba de la propiedad de transmisión de frecuencia de voz completa del dispositivo y sistema utilizados es la parte más importante para la detección del sistema de intercomunicación completo. El método de prueba de la calidad de transmisión de frecuencia de voz del sistema de intercomunicación del edificio conocido es un método de prueba de conjunto completo de pruebas de sonido para evaluar la calidad de transmisión de voz del producto del sistema de intercomunicación del edificio, que incluye la prueba y el cálculo de cinco parámetros técnicos (tales como valor de evaluación de volumen, respuesta de frecuencia, distorsión, relación señal a ruido y valor de evaluación de enmascaramiento de tono lateral). En el método de prueba existente, la prueba de la señal de audio de la extremidad de salida se centra en la evaluación de la propiedad técnica de la respuesta de frecuencia, la distorsión y la propiedad de amplitud de la misma. Por ejemplo, como se muestra en la figura 3, una fuente de señal 301, tal como fuente de sonido/fuente de voz, genera una señal de audio, la señal de audio se pasa a través del dispositivo 302 amplificador de potencia y el trayecto de frecuencia de voz del sistema bajo prueba y se emite mediante una extremidad 303 de salida de frecuencia de voz (tal como un altavoz o auriculares bajo prueba) del sistema bajo prueba; el puerto de salida de la extremidad 303 de salida está fijado en un deflector de aislamiento acústico, el deflector de aislamiento acústico actúa como un cuerpo de pared instalado y utilizado para el dispositivo de simulación, con el objetivo de asegurar que el resultado de la prueba sea más exacto, un micrófono (tal como MIC) 304 se configura directamente a la dirección del puerto de salida, y una unidad central de procesamiento CPU incorporada en el micrófono 304 recoge la señal de salida y puede transformar la señal de audio en una señal electrónica; el micrófono 304 transmite la señal electrónica al dispositivo de medición para medir a través de un dispositivo 305 amplificador (tal como un amplificador de señal), y el analizador 306 de espectro emite un resultado, en el que la evaluación de parámetros técnicos correspondiente realizada por el dispositivo de medición para la señal de salida recogida incluye:

40 prueba de respuesta de frecuencia: una frecuencia de la fuente de señal de audio se modula a un intervalo de 200-4000Hz y se registra un valor recogido por la CPU (la señal de respuesta de frecuencia de voz de la extremidad de salida) de la extremidad del MIC, con el objetivo de obtener el valor efectivo de la presión de sonido (tal como la potencia del sonido, la energía del sonido) de cada frecuencia modulada por la fuente de señal de audio, y el cálculo (tal como la transformada de Fourier, la detección y el cálculo del medidor electrónico de frecuencia) se realiza para obtener la respuesta de frecuencia de la extremidad de salida (tal como un altavoz);

45 prueba de distorsión: una frecuencia de la fuente de señal de audio se modula a una frecuencia que se ha de probar y se registra un valor recogido por la CPU de la extremidad del MIC, el diagrama del espectro de la señal del valor recogido se obtiene a través de una transformación del espectro (tal como la transformada de Fourier), y un valor de distorsión del altavoz bajo la frecuencia de la señal de entrada de la fuente de señal correspondiente a la señal se calcula a través del diagrama de espectro;

50 prueba de propiedad de amplitud: se cambia un margen de salida de la fuente de señal de audio en la misma frecuencia y se registra un valor recogido por la CPU de la extremidad del MIC, el espectro de la señal del valor recogido se obtiene a través de una transformación de espectro (tal como la transformada de Fourier), y se ve si el valor de salida bajo la señal que introduce con un margen diferente corresponde al valor de entrada, por ejemplo, si se cumple con una relación lineal, por ejemplo, la entrada aumenta 10db y la salida también aumenta 10db.

55 Además, solamente basándose en los parámetros anteriores, el grado de restauración de los mismos se determina indirectamente, por ejemplo, cuanto menor es la distorsión, más plana es la respuesta de frecuencia (es decir, la variación entre la entrada y la salida tiende a ser estable y lineal tanto como sea posible), de manera que el sonido emitido se determine como mejor. El grado de restauración se refiere a una consistencia de la forma espectral del

marco original de la entrada de la fuente de señal y la forma espectral del marco recogido de la extremidad de salida, es un parámetro técnico importante para evaluar todo el rendimiento, en particular al rendimiento de salida, del dispositivo y sistema en la medición de frecuencia de voz. Ya que la técnica anterior carece de las pruebas y la determinación realizadas directamente sobre el grado de restauración de la señal de salida de la extremidad de salida, hace que la evaluación del rendimiento, particularmente el rendimiento de salida, del dispositivo y sistema sea inexacta. El método de prueba existente incluye además los siguientes defectos: no se pueden probar todas las frecuencias continuas en cuestión, con el objetivo de hacer que el resultado de la prueba sea inexacto; cuando el dispositivo y el sistema se aplican prácticamente, la señal de voz que prácticamente existe es una señal de punto de tono múltiple (por ejemplo, la voz humana es una variedad de superposición de N puntos de frecuencia/frecuencias); sin embargo, la prueba de distorsión existente utiliza un único punto de frecuencia (es decir, 200Hz, 400 Hz), es inconsistente con la situación de distorsión del punto práctico de tono múltiple, con el objetivo de hacer que el grado de restauración determinado mediante el uso de la distorsión de la prueba también sea inexacto. Por lo tanto, finalmente causaría que el resultado de detección de la señal de audio existente sea inexacto o que la precisión no sea alta, con el objetivo de crear una diferencia mayor entre el resultado de la prueba y la evaluación subjetiva práctica.

El documento CN 103607669 A describe un método de detección y un sistema de detección para características de transmisión de frecuencia de audio de un sistema de intercomunicación del edificio. El método de detección comprende las etapas: una porción de fuente de sonido genera señales de voz específicas, las señales de voz específicas tomadas como señales de voz de entrada son una entrada a un acceso detectado, las señales de voz específicas son transmitidas en el acceso detectado, y se emiten como señales de voz de salida detectadas; las señales de voz específicas comprenden una señal de voz de simulación P.50 o una señal de voz humana P.501 de la organización de normalización de las telecomunicaciones – unión de telecomunicaciones internacional (ITU-T); la presión de sonido relacionada con las señales de voz de salida detectadas se detectan basándose en las señales de voz de salida detectadas, y los valores del parámetro de característica de frecuencia de audio se calculan dependiendo de la presión de sonido detectada relacionada con las señales de voz de salida detectadas con el fin de determinar las características de transmisión de frecuencia de audio del acceso detectado. Mejorando la medición de las fuentes de sonido, los tableros y la distorsión de frecuencia de audio, se mejoran la exactitud de detección y la precisión del resultado de detección de las características de transmisión de frecuencia de audio del sistema de intercomunicación del edificio de simulación (modo bus), y el resultado de detección está relativamente próximo al efecto de uso real.

El documento US 2013/315405 A1 describe un procesador de sonido, un método de procesamiento de sonido, y describe el siguiente contenido: un micrófono es un dispositivo de entrada de sonido que recoge sonido, las señales de audio emitidas desde el micrófono se suministran al procesador de sonido para realizar conversión analógica a digital sobre señales de audio suministradas desde el micrófono. Una CPU realiza procesamiento de transformada de rápida de Fourier (FFT) sobre cada uno de los datos de audio en la posición próxima y los datos de audio en la posición de escucha, y encuentra una característica de frecuencia.

El documento US 2007/036364 A1 describe un aparato de compensación de campo de sonido y un método de compensación de campo de sonido. El aparato incluye: una sección que genera una señal de prueba de regulación del volumen de sonido; una sección de accionamiento que acciona un altavoz; un micrófono que recibe la salida del altavoz; y que controla operaciones de las secciones respectivas. La señal de prueba de regulación del volumen de sonido es una señal de suma representativa de la suma de señales de ondas sinusoidales de una sola frecuencia cuyas frecuencias se establecen con respecto a una relación entera. La sección de control hace que la sección de accionamiento accione el altavoz utilizando la señal de prueba de regulación del volumen de sonido para detectar por ello los niveles de señal de los componentes de frecuencia de las señales de ondas sinusoidales desde las señales de salida del micrófono. Según un valor promedio de los niveles de las señales, la sección de control establece un volumen de sonido de medición cuando el altavoz se acciona utilizando la señal de prueba de medición.

### Compendio

Basándose en el defecto anterior existente en la técnica anterior, el propósito principal de la presente solicitud es proporcionar un método y sistema para evaluar la calidad de sonido de una voz humana, con el objetivo de mejorar la exactitud y precisión para probar el grado de restauración de la señal de audio. Además, se mejora toda la precisión para probar el rendimiento de la señal de audio del dispositivo y del sistema. Con el fin de resolver el defecto técnico en la técnica anterior, el propósito de la presente solicitud se logra a través del siguiente esquema técnico.

La presente solicitud proporciona un método para probar el grado de restauración de una señal de audio, según la reivindicación 1.

En donde el dispositivo de recolección de señal de audio incluye: transformar la señal de voz continua recogida a un flujo de señal digital correspondiente, y transmitir el flujo de señal digital correspondiente al dispositivo de análisis de señal de audio para segmentar y analizar el flujo de señal.

5 En donde el dispositivo de análisis de señal de audio incluye: obtener la señal de voz humana genuina como la señal de voz de entrada generada por la parte de fuente de sonido; segmentar y dividir el flujo de señal de la señal de voz de entrada en un intervalo de tiempo, con el objetivo de dividir el flujo de señal en segmentos con el intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representa los segmentos una o más señales de voz, realizando una transformación de espectro sobre cada uno de los grupos de señales de voz, con el objetivo de obtener un valor de característica de sonido correspondiente después de transformar cada uno de los grupos de señal de voz.

10 En donde el dispositivo de análisis de señal de audio incluye: sincronizar el flujo de señal de la señal de voz de salida recogida del dispositivo de recolección de señal de audio y el flujo de señal de la señal de voz de entrada desde la parte de fuente de sonido correspondiente al flujo de señal de la señal de voz de salida; tomar el intervalo de tiempo idéntico utilizado para segmentar el flujo de señal de la señal de voz de entrada para segmentar y dividir el flujo de señal de la señal de voz continua recogida en segmentos con el intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representa los segmentos una o más señales de voz, realizar la transformación del espectro sobre cada uno de los grupos de señales de voz para obtener un valor de característica de sonido correspondiente después de transformar cada uno de los grupos de señales de voz.

15 En donde el dispositivo de análisis de señal de audio incluye: extraer un valor de característica de sonido correspondiente a un segmento de un grupo de señales de voz que pertenece al flujo de señal de la señal de voz de entrada, y extrayendo un valor de característica de sonido correspondiente al segmento y correspondiente a un segmento que pertenece al flujo de señal de la señal de voz de salida, y calcular y analizar valores de evaluación de contraste de los dos segmentos correspondientes basándose en un principio de similitud; contando y/o calculando en promedio todos los valores de evaluación de contrato correspondientes de los segmentos de los grupos de señales de voz que pertenecen al flujo de señal de la señal de voz de entrada y al flujo de señal de la señal de voz de salida correspondiente para determinar el grado de restauración.

20 La presente solicitud proporciona además un método para evaluar la calidad de sonido de una voz humana según la reivindicación 6.

25 En donde el método incluye además: obtener la señal de voz humana genuina como la señal de voz de entrada generada por la parte de fuente de sonido; segmentar y dividir el flujo de señal de la señal de voz de entrada en un intervalo de tiempo, con el objetivo de dividir el flujo de señal en segmentos con el intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representa los segmentos una o más señales de voz, realizando una transformación de espectro sobre cada de los grupos de señales de voz, con el objetivo de obtener un valor de característica de sonido correspondiente después de transformar cada uno de los grupos de señales de voz.

30 En donde la segmentación y el análisis del flujo de señal de la señal de voz de salida recogida incluye: sincronizar el flujo de señal de la señal de voz de salida recogida desde el dispositivo de recolección de señal de audio y el flujo de señal de la señal de voz de entrada desde la parte de fuente de sonido correspondiente al flujo de señal de la señal de voz de salida; tomar el intervalo de tiempo idéntico utilizado para segmentar el flujo de señal de la señal de voz de entrada para segmentar y dividir el flujo de señal de la señal de voz continua recogida en segmentos con el intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representa los segmentos una o más señales de voz, realizando la transformación del espectro sobre cada uno de los grupos de señales de voz para obtener un valor de característica de sonido correspondiente después de la transformación de cada uno de los grupos de señales de voz.

35 En donde la segmentación y el análisis del flujo de señal de la señal de voz de salida recogida para determinar el grado de restauración incluye: extraer un valor de característica de sonido correspondiente a un segmento de un grupo de señales de voz que pertenece al flujo de señal de la señal de voz de entrada, y extraer un valor de característica de sonido valor correspondiente al segmento y correspondiente a un segmento que pertenece al flujo de señal de la señal de voz de salida, y calcular y analizar valores de evaluación de contraste de los dos segmentos correspondientes basándose en un principio de similitud; contando yo calculando en promedio todos los valores de evaluación de contrato correspondientes de los segmentos de los grupos de señales de voz que pertenecen al flujo de señal de la señal de voz de entrada y al flujo de señal de la señal de voz de salida correspondiente para determinar el grado de restauración.

40 En donde el método incluye, además: segmentar el flujo de señal de la señal de voz de entrada y el flujo de señal de la señal de voz de salida en el intervalo de tiempo idéntico está tomando un intervalo de tiempo de 20 ms para segmentar el flujo de señal en segmentos de los grupos de señales de voz con una o más señales de voz.

**Breve descripción de los dibujos**

55 Los aspectos, características y ventajas anteriores y otros ejemplos de ciertas realizaciones ejemplares de la presente invención serán más evidentes a partir de la siguiente descripción tomada junto con los dibujos adjuntos, en los que:

la figura 1 es un diagrama de bloques de estructura de un sistema para evaluar la calidad de sonido de una voz humana según una realización de la presente solicitud;

la figura 2 es un diagrama de flujo de un método para evaluar la calidad de sonido de una voz humana según una realización de la presente solicitud; y

- 5 la figura 3 es una vista esquemática de una realización para medir algunos parámetros de evaluación de una extremidad de salida en la prueba de calidad de transmisión de frecuencia de voz existente.

### Descripción detallada

10 La idea principal de la solicitud actual es que en la prueba, la fuente de sonido adopte una voz humana como una señal de voz de entrada, de tal manera que un rendimiento de la frecuencia de voz de la voz transmitida en la prueba pueda conformar más una situación de solicitud práctica del dispositivo o sistema bajo prueba, con el objetivo de detectar o probar una característica de salida de sonido del sistema bajo prueba (un dispositivo de comunicación y un sistema de comunicación que tenga requisitos de rendimiento seguro, tal como un sistema de intercomunicación del edificio) para obtener resultados de prueba más precisos y exactos; Además, se realiza un proceso de segmentación continuo en la señal de audio continua, y las similitudes de la señal de audio continua y la  
15 señal de voz humana se comparan para obtener la evaluación del grado de restauración, determinando por ello de manera más precisa y más exacta el rendimiento de salida de sonido del sistema bajo prueba.

20 Para hacer los objetos, las soluciones técnicas y las ventajas de las realizaciones de la presente invención más claramente, las soluciones técnicas de la presente invención se describirán clara y completamente a continuación con referencia a las realizaciones y dibujos de la presente invención. Aparentemente, las realizaciones descritas son realizaciones meramente parciales de la presente invención, en lugar de todas las realizaciones. Otras realizaciones derivadas por los expertos en la técnica de acuerdo con las realizaciones de la presente invención sin pasar a través de esfuerzos creativos caerán todas dentro del alcance de protección de la presente invención.

25 La siguiente descripción con referencia a los dibujos adjuntos se proporciona para explicar las realizaciones ejemplares de la invención. Obsérvese que en el caso de que no haya conflicto, las realizaciones de la presente invención y las características de las realizaciones pueden combinarse arbitrariamente entre sí.

30 Se refiere a la figura 1 que muestra un diagrama de bloques de estructura de un sistema para evaluar la calidad de sonido de una voz humana según una realización de la presente solicitud. En la realización, se toma como ejemplo el muestreo y la prueba de la propiedad de transmisión de frecuencia de voz de un intercomunicador del edificio, el rendimiento de salida de sonido del sistema bajo prueba se determina basándose en el análisis del espectro de la señal de muestreo y el análisis del grado de restauración para la señal de la fuente de sonido. En un sistema 100 de detección de la realización, puede incluir principalmente: una parte 110 de fuente de sonido, un sistema 120 bajo prueba, un dispositivo 130 de recolección de señal de audio (recolector) y un dispositivo 140 de análisis de señal de audio.

35 La parte 110 de fuente de sonido genera una señal de voz particular, la señal de voz particular puede ser una voz humana estándar, tal como una señal de voz genuina del Sector de Normalización de Telecomunicaciones ITU P.501. La señal de voz puede ser como una señal de voz de entrada (tal como una fuente de señal 301), introducida a una extremidad de entrada del sistema 120 bajo prueba desde la parte 110 de fuente de sonido, transmitida en el sistema 120 bajo prueba, y puede ser como una señal de voz de salida que se ha de emitir desde una extremidad de salida (tal como un altavoz o el auricular 303) del sistema bajo prueba.

40 En la realización, el sistema 120 bajo prueba puede ser un sistema de intercomunicación del edificio, recibe la señal de voz de entrada desde la parte 110 de fuente de sonido, la señal de voz de entrada se transmite a través de un dispositivo de amplificación de potencia, un trayecto bajo prueba, un dispositivo de amplificación de potencia para la extremidad de salida del sistema bajo prueba, y la señal de voz de entrada pasada a través del sistema 120 bajo prueba como una señal de voz de salida se emite desde una extremidad de salida del sistema bajo prueba. En donde el trayecto bajo prueba puede ser un trayecto de comunicación necesario para ser probada en el sistema bajo prueba (tal como el sistema de intercomunicación del edificio bajo prueba).

45 El dispositivo 130 de recolección de señal de audio recoge la señal de voz emitida por el sistema 120 bajo prueba, transforma la señal de voz recogida y transmite la señal de voz transformada al dispositivo 140 de análisis de señal de audio para procesar y analizar. Por ejemplo, un micrófono (tal como MIC304 está configurado en la extremidad de salida).

50 El dispositivo 130 de recolección de señal de audio puede incluir un MIC, un amplificador de potencia, un recolector de señal de audio, etc.

55 Por ejemplo, el MIC recibe las señales de voz que se transmiten por el sistema 120 bajo prueba reproducidas por el altavoz de la extremidad de salida del sistema 120 bajo prueba. Específicamente, después de que estas señales de voz continuas introducidas desde la extremidad de entrada pasen a través del sistema 120 bajo prueba, actúan como señales de voz continuas emitidas a través de la extremidad de salida y recibidas por el MIC; las señales de

voz continuas se transmiten al recolector de señales de audio a través del amplificador de potencia, y el recolector de señales de audio transmite estas señales de voz continuas recogidas al dispositivo de análisis de señales de audio.

5 Además, por ejemplo, MIC recibe las señales de voz que son transmitidas por el sistema 120 bajo prueba reproducidas por el altavoz de la extremidad de salida del sistema 120 bajo prueba. Las señales de voz emitidas pueden transformarse en señales electrónicas a través de MIC y formarse como señales digitales a través del proceso de transformación A/D realizado por el procesador, tal como una CPU, de MIC, y las señales digitales correspondientes a las señales de voz se transmiten al dispositivo 140 de análisis de señal de audio para realizar el proceso y análisis de las señales digitales. Ya que la señal de voz de entrada de la parte 110 de fuente de sonido es  
10 continua y la señal de voz emitida por el sistema 120 bajo prueba también es continua, la señal digital continua correspondiente a la señal de voz continua recogida por el dispositivo 130 de recolección de señal de audio puede transmitirse al dispositivo 140 de análisis de señal de audio.

15 El dispositivo 140 de análisis de señal de audio recibe la señal de voz continua transmitida por el recolector 130 de señal de audio, o la señal digital continua correspondiente transformada por la señal de voz continua para procesar y analizar, con el objetivo de determinar el grado de restauración de la señal de voz.

20 Específicamente, el dispositivo 140 de análisis de señal de audio puede incluir un procesador incorporado (tal como una CPU) o un ordenador personal que tiene análisis y rendimiento del proceso. Cuando recibe la señal de voz continua, la CPU transforma la señal continua para formar una señal digital continua o un flujo de señal digital vocal; cuando recibe la señal digital continua que se ha transformado, la señal digital continua es el flujo de la señal digital vocal. Aquí, el flujo de señal digital correspondiente a la señal de voz continua se refiere al flujo de señal de la señal de voz.

25 Además, el flujo de señal de la señal de voz se segmenta, por ejemplo, el flujo de señal se divide en N "segmentos" de N "tramas" (N es un número entero positivo mayor que o igual a 0). Basándose en estos segmentos o tramas, todas las señales continuas recogidas relacionadas con todas las pruebas se procesan y analizan. En una realización, basándose en un principio del efecto Haas y considerando que el oído humano no reconoce el orden de la frecuencia del nivel de señal en un período de tiempo de 20 milisegundos (ms), se selecciona un intervalo de tiempo para segmentar el flujo de señal, y el tiempo de cada segmento/trama es de 20 ms. Además, la transformación de espectro se realiza en cada segmento/trama, el espectro del segmento/trama transformado y el espectro (es decir, el espectro de trama correspondiente del flujo de señal de entrada) del segmento  
30 correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte 110 de fuente de sonido original, se compara y analiza, con el objetivo de obtener el resultado de prueba/detección del grado de restauración.

Combinado con un diagrama de flujo de un método para evaluar la calidad de sonido de una voz humana de una realización de la presente invención mostrada en la figura 2, el sistema y el método del mismo para evaluar la calidad de sonido de una voz humana se describe más específicamente como sigue.

35 En la etapa 210, una parte 110 de fuente de sonido genera una señal de voz particular, la señal de voz particular puede ser una voz humana estándar, tal como una señal de voz genuina del Sector de Normalización de Telecomunicaciones ITU P.501. La señal de voz es una señal de voz de transmisión para probar la propiedad de transmisión de frecuencia de voz de un sistema 120 bajo prueba. La señal de voz puede ser como una señal de voz de entrada a través de la parte 110 de fuente de sonido, introducida a una extremidad de entrada del sistema 120 bajo prueba, transmitida en el sistema 120 bajo prueba, y finalmente puede ser como una señal de voz de salida para ser emitida desde una extremidad de salida del sistema 120 bajo prueba. La señal de voz humana incluye todas las distorsiones de intermodulación, la señal de voz humana se adopta como señal de entrada, además se conforma con el entorno de uso del sistema bajo prueba y la prueba de la misma es más precisa y objetiva. El sistema 120 bajo prueba es, por ejemplo, un sistema de intercomunicación del edificio. La implementación específica de esta etapa puede referirse a la descripción sobre la parte 110 de fuente de sonido y el sistema 120 bajo prueba de la misma en el sistema.  
45

50 En la etapa 220, la señal de voz humana es recogida por el dispositivo 130 de recolección de señal de audio a través de la señal de voz continua emitida por el sistema 120 bajo prueba, y enviada al dispositivo 130 de análisis de señal de audio para analizar. La implementación específica de esta etapa puede referirse a la descripción sobre el dispositivo 130 de recolección de señal de audio en el sistema.

55 En la etapa 230, el dispositivo 140 de análisis de señal de audio segmenta la señal de audio recogida (señal de voz) y a continuación realiza una transformación de espectro, y compara y analiza la señal de audio recogida con una señal de voz particular generada por la señal de fuente de sonido, con el objetivo de obtener el resultado de la prueba del grado de restauración. La implementación específica de esta etapa puede referirse a la descripción sobre el dispositivo 140 de análisis de señal de audio en el sistema.

En una realización, el proceso de segmentación sobre la señal de audio se describe adicionalmente como sigue.

En primer lugar, el procesador (CPU) del dispositivo 140 de análisis de señal de audio segmenta el flujo de señal digital correspondiente a la señal de voz continua introducida por la señal digital de la señal de voz estándar, es

decir, generada por la parte 110 de fuente de sonido de la extremidad de entrada, en donde cada segmento es de 20 ms (es decir, "trama"), y el flujo de señal se segmenta como N grupos de señales de voz (es decir, N señales de trama). En donde cada grupo de señales de voz (N señales de trama) incluye una o más señales de voz (o referidas al parámetro de señal), N señales de trama son, por ejemplo, P1, P2, P3, ..., PN, estas tramas segmentados se almacenan, y cada trama, tal como P1, consta de una señal digital dentro del período de tiempo de 20 ms. Además, cada señal de trama (es decir, el flujo de señal) se transforma al espectro correspondiente y se almacena el espectro correspondiente. En donde cada trama tiene el valor de la característica de sonido correspondiente después de la transformación.

Posteriormente, cuando el dispositivo 140 de análisis de señal de audio recibe el flujo de señal recogido, es decir, recibe el flujo de señal recogido desde la extremidad de salida para la prueba y corresponde a la señal de voz continua introducida, sincroniza el flujo de señal recogido y el flujo de señal de entrada; de manera similar, una forma de 20 ms de cada segmento se lleva al flujo de señal a N grupos de señales de voz, es decir N tramas. En donde cada señal de trama, es decir, cada grupo de señales de voz, incluye una o más señales de voz (o referidas al parámetro de señal), N señales de trama son, por ejemplo, p1, p2, p3, ..., pN, estas tramas segmentadas se almacenan, y cada trama, tal como P1, consta de una señal digital dentro del intervalo de tiempo de 20 ms. Además, cada señal de trama se transforma en un espectro correspondiente y se almacena el espectro correspondiente. Donde cada trama tiene el valor de la característica de sonido correspondiente después de la transformación.

Cuando la señal de voz (tal como el flujo de señal digital de la señal de voz) puede lograr la transformación del espectro a través de la transformada de Fourier u otros procesos de señal de sonido y obtener el valor de la característica de sonido de cada segmento al mismo tiempo. Basándose en la señal de voz de salida correspondiente a la señal de voz de entrada y transmitida por el sistema 120 bajo prueba y al igual que la señal de voz estándar de la señal de voz de entrada, después de que ambas sean sincronizadas, se realiza la misma segmentación y espectro transformado, con el objetivo de obtener el valor de la característica de sonido de cada grupo de señales o de cada señal de trama.

A continuación, estos valores de característica de sonido, es decir, valores de característica de sonido correspondientes a cada trama en P1, P2, P3, ..., PN y valores de característica de sonido correspondientes a cada trama en p1, p2, p3, ..., pN, se extraen, un valor de característica de sonido correspondiente a un segmento de un grupo de señales de voz que pertenece al flujo de señal de la señal de voz de entrada, basándose en una manera, tal como un principio de similitud/cálculo de similitud, un cálculo de similitud o un análisis de principio de similitud se realiza en cada trama correspondiente, es decir, los valores de la característica de sonido P1 correspondiente y los valores de la característica de sonido correspondientes a p1, con el objetivo de determinar el grado de restauración. Por ejemplo, la similitud entre el valor de característica P1 y el valor de característica p1 se calculan para obtener un valor de similitud 0-1 (tal como similar o 100% similar). Con el objetivo de determinar claramente el resultado del análisis, el valor puede multiplicarse por un múltiplo, tal como 100, y se adopta el grado centesimal, tal como grados de 0-100, de tal manera que cada trama pueda obtener una evaluación de análisis comparativo.

Por ejemplo, se realiza una comparación de similitud de coincidencia entre los valores de característica A, B, C, D de P1 y el valor de característica a, b, c', d de p1, con el objetivo de obtener que C sea diferente de c', solamente tres características son similares, y la similitud es  $\frac{3}{4} * 100 = 75$  puntos.

Alternativamente, por ejemplo, los valores de la característica de P1 son una agrupación de una dimensión [A, B, C, D], los valores de características de p1 son una agrupación de una dimensión [a, b, c, d], las disposiciones de los mismos son una agrupación de una dimensión, y los números de serie A-D y a-d están ordenados de pequeño a mayor. La tendencia de la disposición y la tendencia se analizan como similares/iguales según el principio de similitud, de tal manera que la puntuación comparativa del espectro de P1 y p1 es  $(\frac{1}{2} + \frac{1}{2}) * 100 = 100$ .

Finalmente, las puntuaciones de todas las tramas se determinan según la puntuación de cada trama, con el objetivo de determinar el grado de restauración, es decir, el grado de restauración (y el grado similar de la entrada) del sonido de salida al sonido de entrada. Por ejemplo, las puntuaciones de todas las tramas se cuentan para obtener una puntuación promedio de la señal de voz de salida, y la puntuación promedio es la puntuación del grado de restauración del sistema 120 bajo prueba. Además, puede existir una trama vacía en el flujo de señal, ya que la trama vacía no incluye la información de voz, con el fin de reducir la interferencia, la puntuación de la trama vacía puede eliminarse y la puntuación de la trama de señal de voz efectiva es solamente considerada, y a continuación se cuenta la puntuación de la trama de la señal de voz efectiva para obtener la puntuación promedio de la señal de salida, en donde la puntuación promedio es la puntuación del grado de restauración del sistema 120 bajo prueba.

En el esquema de la presente solicitud, la señal de fuente de sonido para la prueba adopta la voz humana en el proceso de prueba, con el objetivo de asegurar que el sistema bajo prueba en un entorno de trabajo práctico durante el proceso de prueba, y la señal sea la voz humana y la distorsión de la misma incluye todas las distorsiones de intermodulación. La segmentación y el procesamiento de la señal incluyen la prueba sobre la frecuencia continua, conforma la situación de la señal de la voz humana y muestra más adecuadamente la propiedad de salida de sonido del sistema bajo prueba. Por lo tanto, la segmentación y prueba se realizan directamente en la señal continua del grado de restauración, puede obtener resultados de prueba más precisos y más exactos del sistema y dispositivo bajo prueba.

5 Debería observarse que los términos "incluir", "contener" y cualquier variación de los mismos están destinados a cubrir una inclusión no exclusiva. Por lo tanto, un proceso, método, objeto o dispositivo que incluye una serie de elementos no solamente incluye estos elementos, sino que también incluye otros elementos no especificados expresamente, o puede incluir elementos inherentes al proceso, método, objeto o dispositivo. Si no se hacen más limitaciones, un elemento limitado por "incluir un/uno..." no excluye otros mismos elementos existentes en el proceso, el método, el artículo o el dispositivo que incluye el elemento.

10 Las descripciones anteriores son solamente realizaciones de la presente invención, pero no limitan la presente invención. Para los expertos en la técnica, la presente invención puede tener una variedad de modificaciones y cambios. Cualquier modificación, el reemplazo equivalente o la mejora hecha dentro del principio de la presente invención debería dentro del alcance de las reivindicaciones de la presente invención.



**REIVINDICACIONES**

1.- Un sistema para evaluar la calidad de sonido de una voz humana, que comprende al menos:

una parte (110) de fuente de sonido, un sistema (120) bajo prueba, un dispositivo (130) de recolección de señal de audio y un dispositivo (140) de análisis de señal de audio;

5 estando configurado el sistema para:

que la parte (110) de fuente de sonido genere una señal de voz humana como una señal de voz de entrada, e introduce la señal de voz de entrada al sistema (120) bajo prueba a partir de la parte (110) de fuente de sonido;

transmitir la señal de voz de entrada en el sistema (120) bajo prueba y emitir la señal de voz de entrada como una señal de voz de salida a partir de una extremidad de salida del sistema (120) bajo prueba;

10 que el dispositivo (130) de recolección de señal de audio recoja la señal de voz de salida continua emitida por la extremidad de salida, y transmita la señal de voz de salida recogida al dispositivo (140) de análisis de señal de audio; y

que el dispositivo (140) de análisis de señal de audio segmente y analice el flujo de señal de la señal de voz de salida;

15 en donde, el dispositivo (140) de análisis de señal de audio está configurado adicionalmente para adquirir un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido;

20 en donde la segmentación y el análisis del flujo de señal consisten en segmentar el flujo de señal de la señal de voz de salida al menos en un segmento, realizar una transformación de espectro sobre cada segmento, y comparar y analizar un espectro del segmento transformado y el espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido, con el objetivo de obtener un resultado de prueba/detección de un grado de restauración de la señal de voz de salida con respecto a la señal de voz de entrada;

25 en donde el grado de restauración se determina basándose en puntuaciones al menos de todos los segmentos, y, cuando un segmento vacío que no incluye información de voz existe al menos en dicho segmento, una puntuación del segmento vacío se elimina y solamente las puntuaciones de los segmentos restantes al menos de dicho segmento que comprende la señal de voz efectiva se consideran para determinar el grado de restauración.

30 2.- El sistema según la reivindicación 1, que se caracteriza por que el dispositivo (130) de recolección de señal de audio comprende: transformar la señal de voz continua recogida en un flujo de señal digital correspondiente, y transmitir el flujo de señal digital correspondiente al dispositivo (140) de análisis de señal de audio para segmentar y analizar el flujo de señal.

3.- El sistema según la reivindicación 1 o 2, que se caracteriza por que la adquisición de un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido comprende:

35 obtener la señal de voz humana como la señal de voz de entrada generada por la parte (110) de fuente de sonido;

40 segmentar y dividir el flujo de señal de la señal de voz de entrada en un intervalo de tiempo, con el objetivo de dividir el flujo de señal en segmentos con el intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representan los segmentos una o más señales de voz, realizar una transformación de espectro sobre cada uno de los grupos de señales de voz, con el objetivo de obtener un valor de característica de sonido correspondiente después de transformar cada uno de los grupos de señales de voz.

4.- El sistema según la reivindicación 3, que se caracteriza por que el dispositivo (140) de análisis de señal de audio está configurado adicionalmente para, antes de comparar y analizar un espectro del segmento transformado y un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte de fuente de sonido:

45 sincronizar el flujo de señal de la señal de voz de salida recogida a partir del dispositivo de recolección de señal de audio y el flujo de señal de la señal de voz de entrada que procede de la parte de fuente de sonido correspondiente al flujo de señal de la señal de voz de salida; y

50 en donde la segmentación del flujo de señal de la señal de voz de salida comprende tomar el intervalo de tiempo idéntico utilizado para segmentar el flujo de señal de la señal de voz de entrada para segmentar y dividir el flujo de señal de la señal de voz continua recogida en segmentos con intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representa los segmentos una o más señales de voz, y

en donde la realización de una transformación de espectro sobre cada segmento comprende realizar la transformación de espectro sobre cada uno de los grupos de señales de voz para obtener un valor de característica de sonido correspondiente después de la transformación de cada uno de los grupos de señales de voz.

5 5.- El sistema según la reivindicación 4, que se caracteriza por que la comparación y el análisis de un espectro del segmento transformado y de un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido, con el objetivo de obtener un resultado de prueba/detección, comprende:

10 extraer un valor de característica de sonido correspondiente a un segmento de un grupo de señales de voz que pertenece al flujo de señal de la señal de voz de entrada, y extraer un valor de característica de sonido correspondiente al segmento y correspondiente a un segmento que pertenece al flujo de señal de la señal de voz de salida, y calcular y analizar los valores de puntuación de contraste de los dos segmentos correspondientes basándose en un principio de similitud; y

15 contar y/o hacer un cálculo promedio de todos los valores de puntuación de contrato correspondientes de los segmentos de los grupos de señales de voz que pertenecen al flujo de señal de la señal de voz de entrada y el flujo de señal de la señal de voz de salida correspondiente para determinar el grado de restauración.

6.- Un método para evaluar la calidad de sonido de una voz humana, caracterizado por que el método comprende:

una parte (110) de fuente de sonido que genera (210) una señal de voz humana como una señal de voz de entrada, y que introduce la señal de voz de entrada al sistema (120) bajo prueba desde la parte (110) de fuente de sonido;

20 transmitir la señal de voz de entrada en el sistema (120) bajo prueba y emitir la señal de voz de entrada como una señal de voz de salida desde una extremidad de salida del sistema (120) bajo prueba;

recoger (220) la señal de voz de salida continua emitida por la extremidad de salida;

segmentar y analizar (230) el flujo de señal de la señal de voz de salida recogida;

en donde, el método comprende además: adquirir un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido:

25 en donde la segmentación y el análisis del flujo de señal comprenden segmentar el flujo de señal de la señal de voz de salida hacia al menos un segmento, y la comprar y analizar un espectro del segmento transformado y el espectro del segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido, con el objetivo de obtener un resultado de prueba/detección de un grado de restauración de la señal de voz de salida con respecto a la señal de voz de entrada;

30 en donde el grado de restauración se determina basándose en puntuaciones de un segmento o la totalidad de segmentos, y, cuando un segmento vacío que no incluye información de voz existe al menos en dicho segmento, una puntuación del segmento vacío se elimina y solamente las puntuaciones de segmentos restantes del al menos dicho segmento que comprende señal de voz efectiva se consideran para determinar el grado de restauración.

35 7.- El método según la reivindicación 6, que se caracteriza por que la adquisición de un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido comprende:

obtener la señal de voz humana como la señal de voz de entrada generada por la parte de fuente de sonido;

40 segmentar y dividir el flujo de señal de la señal de voz de entrada en un intervalo de tiempo, con el objetivo de dividir el flujo de señal en segmentos con intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representan los segmentos una o más señales de voz, realizando una transformación de espectro en cada de los grupos de señales de voz, con el objetivo de obtener un valor de característica de sonido correspondiente después de transformar cada uno de los grupos de señales de voz.

45 8.- El método según la reivindicación 7, que se caracteriza por que el método comprende además, antes de comparar y analizar un espectro del segmento transformado y un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte de fuente de sonido:

sincronizar el flujo de señal de la señal de voz de salida recogida desde el dispositivo (130) de recolección de señal de audio y el flujo de señal de la señal de voz de entrada desde la parte (110) de fuente de sonido correspondiente al flujo de señal de la señal de voz de salida; y

50 en donde la segmentación del flujo de señal de la señal de voz de salida comprende tomar el intervalo de tiempo idéntico utilizado para segmentar el flujo de señal de la señal de voz de entrada para segmentar y dividir el flujo de señal de la señal de voz continua recogida en segmentos con intervalo de tiempo idéntico, comprendiendo cada uno de los grupos de señales de voz que representan los segmentos una o más señales de voz, y

en donde la realización de una transformación de espectro sobre cada segmento comprende realizar la transformación de espectro sobre cada uno de los grupos de señales de voz para obtener un valor de característica de sonido correspondiente después de transformar cada uno de los grupos de señales de voz.

5 9.- El método según la reivindicación 8, que se caracteriza por que la comparación y el análisis de un espectro del segmento transformado y un espectro de un segmento correspondiente al flujo de señal de la señal de voz de la extremidad de entrada de la parte (110) de fuente de sonido, con el objetivo de obtener un resultado de prueba/detección, comprende:

10 extraer un valor de característica de sonido correspondiente a un segmento de un grupo de señales de voz que pertenece al flujo de señal de la señal de voz de entrada, y extraer un valor de característica de sonido correspondiente al segmento y correspondiente a un segmento que pertenece al flujo de señal de la señal de voz de salida, y calcular y analizar valores de puntuación de contraste de los dos segmentos correspondientes basándose en un principio de similitud; y

15 contar y/o hacer el cálculo promedio de todos los valores de puntuación de contrato correspondientes de los segmentos de los grupos de señales de voz que pertenecen al flujo de señal de la señal de voz de entrada y el flujo de señal de la señal de voz de salida correspondiente para determinar el grado de restauración.

10.- El método según la reivindicación 8, que se caracteriza por que el método comprende además: segmentar el flujo de señal de la señal de voz de entrada y el flujo de señal de la señal de voz de salida en el intervalo de tiempo idéntico que se toma un intervalo de tiempo de 20 ms para segmentar la señal en segmentos de los grupos de señales de voz con una o más señales de voz.

20

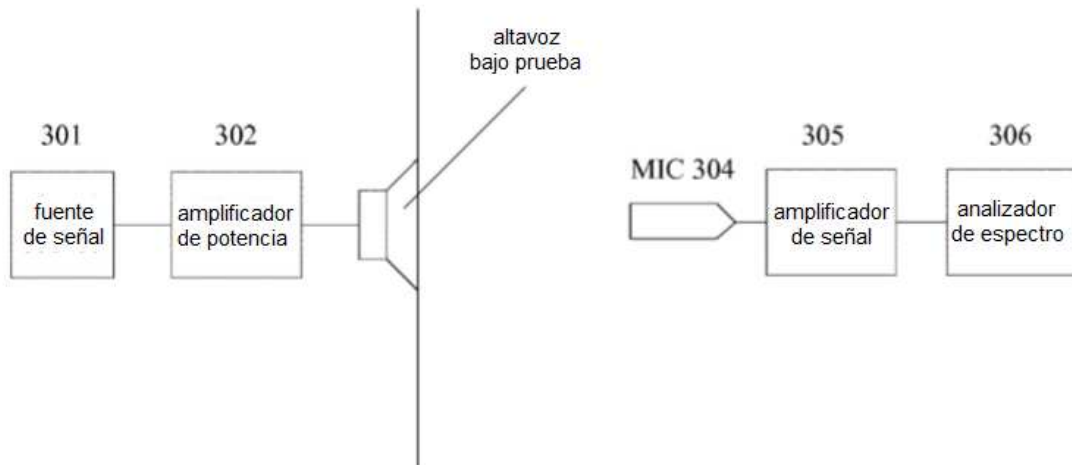


FIG. 3

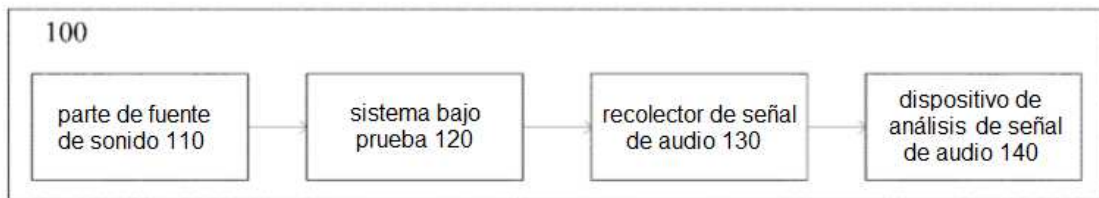


FIG. 2

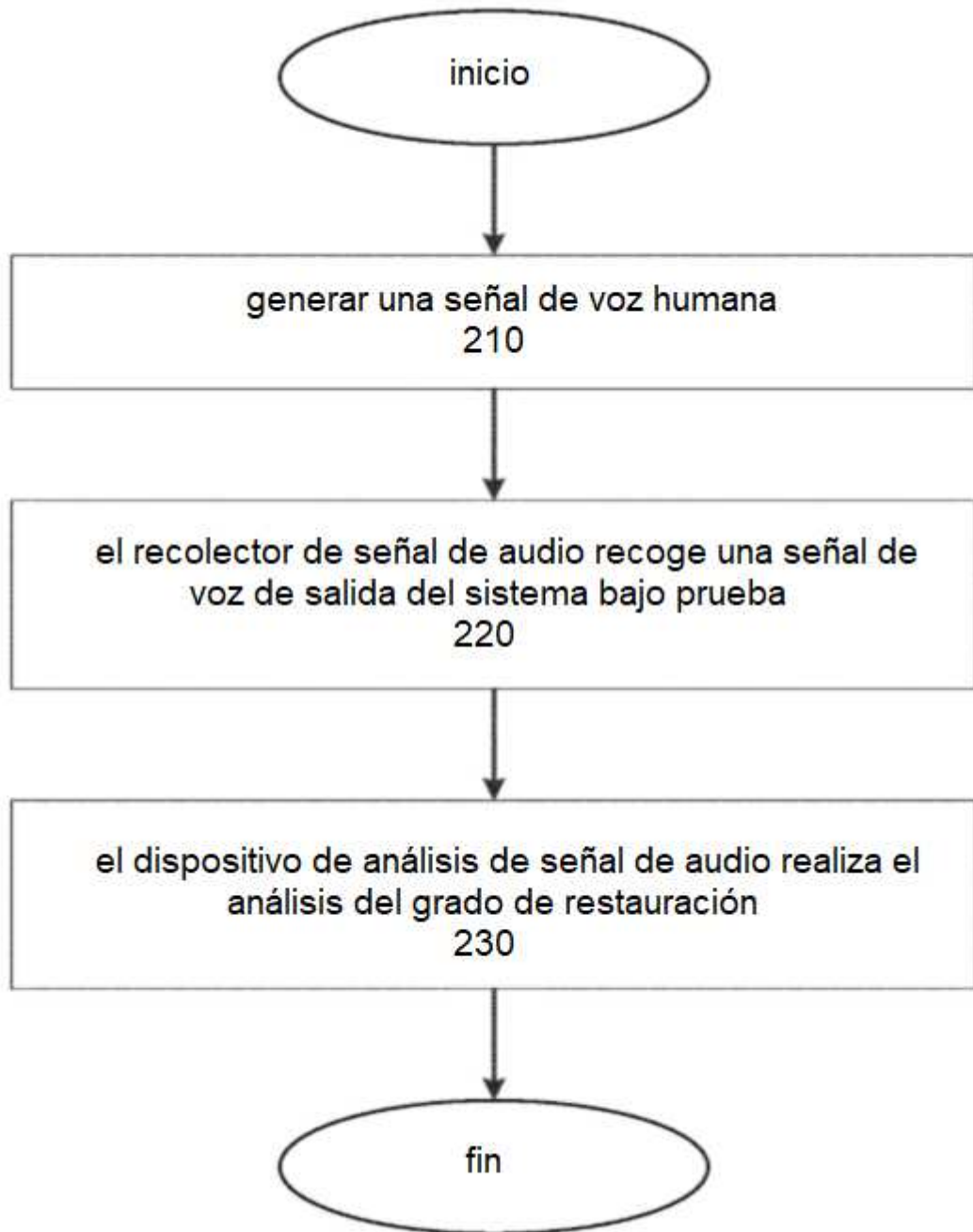


FIG. 1