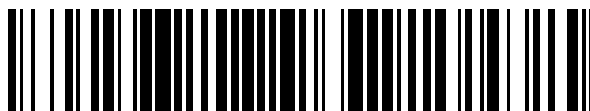


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 389 768**

51 Int. Cl.:
G10L 19/00 (2006.01)
H04M 3/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **04739953 .0**
96 Fecha de presentación: **16.06.2004**
97 Número de publicación de la solicitud: **1634277**
97 Fecha de publicación de la solicitud: **15.03.2006**

54 Título: **Extracción de secciones de señal de prueba para medir la calidad de una señal de audio**

30 Prioridad:
17.06.2003 DE 10327239

45 Fecha de publicación de la mención BOPI:
31.10.2012

45 Fecha de la publicación del folleto de la patente:
31.10.2012

73 Titular/es:
**OPTICOM, DIPL.-ING. MICHAEL KEYHL GMBH
(100.0%)
Nägelsbachstrasse 38
91052 Erlangen, DE**

72 Inventor/es:
**KEYHL, MICHAEL;
SCHMIDMER, CHRISTIAN y
BITTO, ROLAND**

74 Agente/Representante:
PONTI SALES, Adelaida

ES 2 389 768 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

Extracción de secciones de señal de prueba para medir la calidad de una señal de audio.

[0001] La presente invención se refiere a pruebas de audición para valorar la calidad de señales codificadas de audio y voz o para valorar la calidad de una conexión telefónica tal como, por ejemplo, una conexión telefónica por cable o inalámbrica. En particular, la presente invención se refiere a la facilitación de secciones de señal de prueba para la realización de las denominadas mediciones subjetivas y/u objetivas para valorar la calidad.

[0002] Para la valoración mediante mediciones técnicas de la calidad de señales codificadas de audio y voz se emplean hoy en día procedimientos de medición estandarizados basados en la percepción (Perceptual Measurement, medición perceptual). Procedimientos conocidos son el denominado procedimiento PESQ (PESQ = Perceptual Evaluation of Speech Quality = valoración perceptual de la calidad de la voz), que se describe en el documento de estandarización ITU-T P.862 (02/2001). Otro procedimiento de medición conocido para la valoración de la calidad es el denominado procedimiento PEAQ (PEAQ = Objective Measurements of Perceived Audio Quality = mediciones objetivas de la calidad de audio percibida), que se muestra en el documento de estandarización Rec. ITU-R BS. 1387-1 (1998-2001). Estos procedimientos u otros procedimientos para la valoración de la calidad tienen en común el hecho de que una señal que ha de comprobarse (“señal de prueba”), que normalmente es la señal de salida de un sistema o una red o, en general, un elemento que ha de analizarse (DUT), se compara con una señal original o también una señal de referencia, que normalmente es la señal de entrada al elemento DUT que ha de comprobarse.

[0003] En la figura 6 se muestra un “setting” (ajuste) general de este tipo. La señal de audio original que se alimenta a un DUT 600 representa en este caso la señal de referencia o la señal de entrada, mientras que la señal de salida tras el DUT 600 se utiliza o bien para realizar una prueba de audición con sujetos de prueba, tal como se indica mediante un sujeto 602, o bien para realizar un procedimiento de valoración de la calidad tal como, por ejemplo, PESQ o PEAQ, tal como se muestra mediante un modelo 604. Con ello, mediante la alimentación de la señal de salida procedente del DUT 600 al sujeto 602 puede realizarse una prueba de audición subjetiva que normalmente se realiza con varios sujetos de prueba en salas estandarizadas. Mediante la alimentación al modelo 604 de la señal de audio original antes del DUT 600, es decir, la señal de referencia, y la señal de audio distorsionada por el DUT, puede realizarse una comprobación objetiva, es decir, una evaluación algorítmica sin sujetos de prueba subjetivos.

[0004] El DUT 600 es normalmente un sistema cuya influencia en la calidad del audio debe valorarse. Un sistema de este tipo es, por ejemplo, una conexión de telecomunicaciones y, en particular, una conexión telefónica, que puede ser inalámbrica o por cable. Un DUT 600 alternativo es, por ejemplo, un tramo de codificador/decodificador para valorar el perjuicio de la calidad de un concepto de codificador con concepto de decodificador conectado posteriormente. La función del modelo, cuando el modelo opera en el marco previsto, debe ser una predicción de la calidad percibida que sujetos de prueba indicarían subjetivamente en una escala cuando escuchan la señal de salida del DUT 600.

[0005] En el caso del procedimiento PESQ, por ejemplo, la señal de audio original, es decir, la señal de audio antes del DUT 600, que es la señal de referencia, se compara, considerando un retardo temporal (delay), con la señal de audio distorsionada por el DUT 600, empleándose para ello un modelo psico-acústico. En particular, tanto la señal de audio original antes del DUT 600 como también la señal de audio distorsionada tras el DUT 600 se transforman en una denominada representación interna que es análoga a la representación psicofísica de señales de audio en el sistema auditivo humano, considerándose particularmente parámetros como la escala Bark y la intensidad sonora (sonido), tal como se conocen en la técnica. La representación psicofísica interna de la señal de audio original se compara entonces con la representación psicofísica interna de la señal de audio distorsionada para calcular, en función del modelo, uno o varios parámetros de error que permitan realizar una declaración cuantitativa de la calidad.

[0006] Un procedimiento de valoración de la calidad mostrado mediante la figura 6 se denomina también procedimiento “intrusivo” dado que es necesario alimentar la señal de referencia, es decir, la señal de audio original, al sistema que ha de comprobarse (DUT 600). Entonces, en la salida del DUT se obtiene, tal como ya se ha expuesto, la señal de prueba que ha de valorarse, que en la figura 6 también se denomina señal de audio distorsionada o, en general, señal de audio. La salida del DUT 600 puede ser, por ejemplo, el extremo distante de una conexión telefónica de dos abonados, alimentándose la señal de audio original en el extremo cercano como señal de referencia. En este caso, el procedimiento de medición tal como, por ejemplo, PESQ, caracterizaría la calidad de la voz de una conexión telefónica.

[0007] Tal como ya se ha expuesto, los procedimientos de medición algorítmicos se basan en una

combinación de hallazgos psico-acústicos y cognitivos sobre la percepción auditiva humana. El experimento que sirve de base a este procedimiento consiste, en primer lugar, en que se realiza una prueba de audición subjetiva en la que se presenta a un número estadísticamente suficiente de oyentes de prueba (“sujetos”) una serie de secuencias de voz o audio para su valoración. Los sujetos de prueba valoran estas secuencias mediante una escala de calidad discreta o continua que en la técnica también se denomina “escala de opinión” y varía, por ejemplo, de 1 (“bad” = mala) a 5 (“excellent” = excelente). Estas pruebas de audición subjetivas se muestran, por ejemplo, en el documento de estandarización ITU-T P.800 (08/1996).

[0008] Se ha mostrado que sujetos de prueba reales solo pueden valorar cualitativamente secuencias cortas. Si se presenta a los sujetos de prueba una secuencia más larga, es decir, una sección de señal de prueba más larga, entonces se establece una cierta “media estadística”. Dicho de otro modo, el proceso cognitivo del olvido de perturbaciones escuchadas conduce a un falseamiento de las declaraciones de los sujetos de la prueba, siendo este falseamiento inmanente al sistema debido al hecho de que los sujetos de la prueba son personas.

[0009] En consecuencia, por tanto, en procedimientos de prueba estandarizados tales como, por ejemplo, en el documento de estandarización Rec. ITU-R BS.1116-1 o Rec. ITU-R BS.1534, se prescriben secuencias de prueba que tienen una duración de normalmente entre 8 y 12 segundos, pero cuya longitud máxima no supera los 20 segundos. Estas secuencias de prueba son señales reales, sin embargo, no proceden estocástica o aleatoriamente de un escenario real, sino que son secuencias de prueba estandarizadas predeterminadas que pueden alimentarse en un experimento al DUT que ha de analizarse para obtener la señal de prueba de entrada, es decir, la señal de audio distorsionada por el DUT.

[0010] En los últimos tiempos se han presentado desarrollos que permiten realizar también pruebas no intrusivas que, por tanto, deben posibilitar una estimación de la calidad de la voz basándose exclusivamente en un análisis de la señal de prueba en el lado de recepción, es decir, sin alimentación de una señal de referencia al lado de emisión. Este tipo de desarrollos son especialmente ventajosos para realizaciones prácticas dado que, por ejemplo, permiten realizar una declaración sobre la calidad de la voz de una conexión de telefonía móvil únicamente en el terminal, sin que sea necesario ningún tipo de disposiciones técnicas de medición o precauciones o manipulaciones en la red telefónica, en cierto modo, para la alimentación de una señal de referencia. Cualquier conversación telefónica real debería poder someterse a una valoración de la calidad con un concepto no intrusivo de este tipo.

[0011] Este nuevo concepto no intrusivo está desarrollándose en estos momentos. Se parte de que, por motivos de comparación con conceptos de medición intrusivos, también se prescriben para el concepto de medición no intrusivo longitudes de las secuencias de prueba que son similares a las longitudes de las secuencias de prueba de las pruebas intrusivas, las cuales, por tanto, se eligen de modo que, en el oyente de la prueba no se produzca, debido a una secuencia muy larga, ninguna denominada “media estadística” o un olvido de un fallo, y las cuales, por otra parte, sean suficientemente largas para que pueda realizarse una declaración razonable. Tal como ya se ha expuesto, la duración de las secuencias de prueba normalmente es de entre 8 y 12 segundos, permitiéndose a veces también secuencias de prueba, es decir, secciones de señal de prueba, de un máximo de 20 segundos.

[0012] En particular en el caso de valoraciones no intrusivas de la calidad de una señal de audio distorsionada o en el caso de la valoración de la influencia de, por ejemplo, un canal de transmisión, 600 en la figura 6, sobre la señal de audio, ya no puede trabajarse sin más con secciones predefinidas de señal de prueba. En lugar de ello, debe recurrirse a señales de audio reales para la valoración de la calidad. Sin embargo, debe garantizarse la posibilidad de comparación de los resultados de la medición dado que esta es precisamente una ventaja fundamental de los procedimientos estandarizados de valoración de la calidad, es decir, el hecho de que los resultados de diferentes pruebas deben poderse comparar.

[0013] A continuación, se muestra mediante la figura 5 la problemática que surge en este sentido. La figura 5 muestra un diagrama de tiempo de una señal transmitida a través de una conexión telefónica, es decir, una señal de audio que ha sido distorsionada por la transmisión a través de una conexión telefónica. En el diagrama de tiempo de la figura 5 se representa, a lo largo de la ordenada, una amplitud normalizada, mientras que, a lo largo de la abscisa, se representa el tiempo. La señal mostrada en la figura 5 muestra claramente la característica de una señal de voz en el sentido de que, por una parte, están presentes secciones portadoras de información tales como, por ejemplo, la sección entre un segundo y nuevos segundos, y en el sentido de que las secciones portadoras de información están separadas entre sí por secciones no portadoras de información que se designan como pausas. La sección no portadora de información que sigue a la primera sección portadora de información se extiende desde aproximadamente 9 segundos a aproximadamente 10,8 segundos. A continuación, sigue nuevamente una sección portadora de información más larga de 10,8 segundos a aproximadamente 20,2 segundos. A esta segunda sección portadora de información sigue nuevamente una pausa entre 20,3 segundos aproximadamente y 21,3 segundos. A la segunda pausa le sigue nuevamente una sección portadora de información que se extiende aproximadamente hasta 23,7 segundos, a continuación de la cual sigue nuevamente una pausa.

[0014] La posibilidad más sencilla para la extracción de secciones de señal de prueba consistiría en descomponer la señal de audio mostrada en la figura 5 en secciones contiguas entre sí de igual longitud. Un tipo de fragmentación para obtener secciones de señal de prueba con una duración de aproximadamente 10 segundos se muestra mediante b(1), b(2), etc. Otro tipo de fragmentación de la señal de audio mostrada en la figura 5 para obtener 5 secciones de señal de prueba con una duración de, por ejemplo, 7,5 segundos se muestra mediante a(1), a(2), a(3), etc.

[0015] La fragmentación de la señal de audio en secciones de longitud constante es problemática en el sentido de que ya no puede calcularse qué tamaño tiene la sección portadora de información en una sección de señal de prueba y qué tamaño tiene la sección no portadora de información en una sección de señal de prueba, es decir, qué tamaño tiene la ponderación información/pausa. Además, particularmente en el caso de conversaciones telefónicas, puede suceder que entre los interlocutores se produzcan pausas más largas. Esto conduciría a que una sección de señal de prueba consistiera, por ejemplo, solo en una pausa. Es evidente sin más que basándose únicamente en una pausa no es posible ninguna valoración de la calidad.

[0016] El procedimiento mostrado en la figura 5 solo es “adecuado” si cualquier conversación telefónica es, por ejemplo, siempre más breve de 20 segundos, de modo que toda la conversación telefónica podría tomarse como sección de señal de prueba. Sin embargo, si este no es el caso, entonces la división en secciones temporales constantes, tal como se ha mostrado mediante la figura 5, no proporciona ningún tipo de posibilidad de comparación con un resultado de prueba de audición subjetiva. Además, los periodos de medición de diferente duración conducen a resultados cuando menos diferentes, si no inutilizables. En particular para la medición en redes de telefonía móvil desde el vehículo en circulación mediante las denominadas “Drive Test Tools”, es deseable una duración de la medición lo más breve posible o la fragmentación de conversaciones de prueba reales en intervalos temporales o periodos de medición más breves, tal como se indica en la figura 5 en a(1), a(2), a(3). Estas duraciones más breves de la medición son deseables, particularmente en el caso de redes de telefonía móvil, para correlacionar entonces los periodos de medición con datos geográficos con el objeto de obtener una afirmación geográficamente detallada en la calidad de un sistema de telefonía móvil.

[0017] Como ya se ha indicado, la figura 5 muestra la representación gráfica de la señal de tiempo de una señal de voz obtenida de una conversación telefónica real. Pueden observarse claramente las partes de modulación con actividad de voz, es decir, las secciones portadoras de información de la señal, aquí, frases pronunciadas, así como las pausas de voz en medio, es decir, las secciones no portadoras de información. Cabe indicar que la señal mostrada en la figura 5 se ha grabado en el lado del auricular de uno de los extremos de la comunicación actual. Tal como se ha indicado, en una conversación se producen pausas claramente más largas en las que habla la otra persona. Estas se han omitido en la figura 5 para mayor claridad.

[0018] En la figura 5 se muestran dos posibles fragmentaciones que se basan en una división en secciones temporales fijas. Puede observarse claramente que una sección temporal en medio de la modulación, es decir, una palabra o frase, puede comenzar (a(2), b(2)) o puede finalizar (a(1), a(2), ..., b(1)).

[0019] Además, también puede suceder, y sucederá particularmente en el caso de un diálogo, que una sección de señal de prueba pueda estar compuesta en una parte fundamental o esté formada totalmente por una pausa, tal como puede observarse parcialmente, por ejemplo, mediante la sección de señal de prueba a(2) que ya está formada en una tercera parte por una pausa.

[0020] Por tanto, la división en secciones temporales fijas de una señal de audio que ha de valorarse no satisface los requisitos impuestos a secuencias adecuadas para pruebas de audición, es decir, ejemplos de voz con normalmente dos frases de una duración máxima de 20 segundos. Además, de forma ideal, es deseable que este tipo de secuencias adecuadas para pruebas de audición comiencen con pausas, terminen con pausas y, en particular, cuando se consideran secciones de señal de prueba sucesivas, también estén separadas por pausas.

[0021] Asimismo, la conexión y desconexión “dura” en partes de modulación tales como, por ejemplo, la desconexión dura de la sección portadora de información en la sección de señal de prueba a(1) conduce a ruidos perturbadores que también se consideran ruidos perturbadores espectrales o “chasquidos”. En teoría de señales, el corte duro de una parte de modulación significa el plegado de la señal con una función de salto. Estos ruidos perturbadores o artefactos serían valorados como interferencia en un procedimiento de medición, lo cual conduciría directamente a que, por ejemplo, una conexión de comunicación se valore peor de lo que es.

[0022] El documento EP1271470A1 da a conocer un procedimiento y un dispositivo para determinar un empeoramiento de la calidad de voz de una señal sin que se utilice ninguna referencia o señal especial. Según un algoritmo de segmentación, se segmenta la señal que va a analizarse para identificar segmentos de voz, segmentos de pausa o segmentos de ruido de la señal. Mediante los segmentos de ruido o los segmentos de pausa se realiza

entonces una primera estimación de la calidad de la voz midiendo el nivel de ruido en estos segmentos. Los segmentos de voz se someten entonces a una medición de la calidad.

5 **[0023]** El documento WO02/065456A1 da a conocer un sistema y un procedimiento para la medición de la calidad de la voz. Primero se realiza un tratamiento previo de la señal de audio y, en concreto, utilizando un circuito para atenuar el ruido de fondo y utilizando un detector de actividad de voz que emplea algoritmos habituales en telefonía. El detector de actividad de voz genera una señal que solo contiene voz y ninguna pausa. Esta señal se alimenta entonces al algoritmo de valoración de la calidad.

[0024] El objetivo de la presente invención consiste en crear un concepto mejorado para la extracción de una sección de señal de prueba a partir de una señal de audio.

10 **[0025]** Este objetivo se alcanza gracias a un dispositivo para la extracción según la reivindicación 1, un procedimiento para la extracción según la reivindicación 16, un dispositivo para la medición de la calidad según la reivindicación 17, un procedimiento para la medición de la calidad según la reivindicación 19 o un programa informático según la reivindicación 20.

15 **[0026]** La presente invención se basa en el conocimiento de que para la extracción de una sección de señal de prueba primero debe analizarse la estructura temporal de la señal de audio para diferenciar una sección portadora de información de la señal de audio de una sección precedente no portadora de información de la señal de audio y una sección subsiguiente no portadora de información de la señal de audio. Basándose en el análisis de la señal de audio en relación con la detección de las secciones portadoras de información, se genera entonces una sección de señal de prueba basándose en la sección portadora de información de la señal de audio. Con ello se aleja del modo
20 de proceder de la división fija en secciones de señal contiguas. Ahora se obtienen secciones de señal de prueba según la invención en el sentido de que la señal de audio se somete a un análisis de señal en relación con su estructura temporal y en relación con su contenido de información para, basándose en los conocimientos así obtenidos, obtener para el tratamiento adicional secciones de señal, es decir, secciones de señal de prueba, que se correspondan en gran medida con las secuencias de prueba adecuadas para la prueba de audición. Por tanto, la
25 fragmentación según la invención de la señal de audio en secciones de señal de prueba no se realiza de forma independiente de la señal sino de forma adaptada a la señal.

[0027] Una ventaja de la presente invención consiste en que la extracción adaptada a la señal de audio de una sección de señal de prueba conduce a que se eviten artefactos inmanentes al sistema. En lugar de ello, se obtienen secciones de señal de prueba adecuadas para la prueba de audición que principalmente solo permiten la aplicación
30 y el uso de conceptos de medición no intrusivos.

[0028] Otra ventaja de la presente invención consiste en que no son necesarias modificaciones DUT o señales de referencia, sino que el concepto según la invención genera, a partir de señales de audio reales, secciones de señal de prueba que pueden manipularse en amplios límites en relación con los criterios predeterminados normalmente mediante pruebas de audición.

35 **[0029]** En un ejemplo de realización preferido de la presente invención, el análisis de la señal de audio tiene lugar mediante una detección de la actividad de voz, una detección de pausas o una detección de ruido o un reconocimiento de voz posterior.

[0030] La sección de señal de prueba puede, si las longitudes temporales son suficientes, contener directamente una sección completa portadora de información de la señal de audio. Sin embargo, en función de la forma de
40 realización, puede realizarse también una manipulación a una sección portadora de información de la señal de audio para, por ejemplo, añadir pausas al principio y al final de una sección portadora de información con el objeto de generar una relación predefinida de, por ejemplo, modulación de voz a, por ejemplo, pausa.

[0031] Mediante la facilitación de un valor mínimo predeterminado para la longitud temporal de la sección de señal de prueba y un valor máximo predeterminado para la longitud temporal de una sección de señal de prueba es
45 posible, en un ejemplo de realización preferido de la presente invención, generar también, a partir de secciones portadoras de información más largas, preferiblemente mediante incremento y disminución lentos del volumen de sonido (fade-in y fade-out), secciones de señal de prueba adecuadas para la prueba de audición, las cuales están fundamentalmente libres de artefactos dado que se oculta la conexión y desconexión rápidas y no naturales de una

sección portadora de información.

[0032] La presente invención es especialmente ventajosa en el sentido de que transforma una señal de audio cualquiera que normalmente contiene pausas largas en una secuencia de secciones de señal de prueba, cada una de las cuales está compuesta, en una proporción mínima especificable, por una sección portadora de información de la señal de audio. Con ello se eliminan, en cierta medida de forma automática, las largas pausas habituales. Una valoración de la calidad del canal de transmisión del que procede la señal de audio realiza entonces esta valoración de la calidad solo con secciones de señal de prueba útiles y no malgasta recursos sin sentido intentando en vano una valoración de la calidad de pausas de un abonado en, por ejemplo, una conversación telefónica.

[0033] A continuación, se explican de forma detallada ejemplos de realización preferidos de la presente invención haciendo referencia a los dibujos adjuntos. Muestran:

la fig. 1, un diagrama de bloques de un dispositivo para la extracción de una sección de señal de prueba según un ejemplo de realización preferido de la presente invención;

la fig. 2, una representación esquemática del dispositivo para el análisis de la figura 1 según un ejemplo de realización preferido de la presente invención;

15 la fig. 3, una representación detallada del dispositivo para la generación de la figura 1 según un ejemplo de realización preferido de la presente invención;

la fig. 4, una representación de la fragmentación de una señal de audio tal como se consigue gracias a la presente invención;

la fig. 5, una fragmentación de la señal de audio en secciones de señal de prueba de igual longitud; y

20 la fig. 6, un diagrama que muestra una vista general para explicar el modo de funcionamiento básico de una prueba de audición intrusiva para la valoración de la calidad de un sistema que va a comprobarse.

[0034] La figura 1 muestra un ejemplo de realización preferido de un dispositivo según la invención para extraer una sección de señal de prueba de una señal de audio. La señal de audio se alimenta a un dispositivo 10 para analizar una estructura temporal de la señal de audio. El dispositivo 10 para analizar la estructura temporal de la señal de audio está operativo para diferenciar una sección portadora de información de la señal de audio de una sección precedente no portadora de información de la señal de audio y una sección subsiguiente no portadora de información de la señal de audio. Después del dispositivo 10 está conectado un dispositivo 12 para generar la sección de señal de prueba, que, preferiblemente, es adecuada para la prueba de audición, basándose en la sección portadora de información de la señal de audio, la cual, en una alternativa, se facilita a través de una línea de conexión 14 por el dispositivo 10. De forma alternativa, el dispositivo 10 también puede estar configurado para emitir una indicación acerca de dónde comienza una sección portadora de información en una señal de audio y dónde finaliza, por ejemplo, mediante la indicación de las muestras si la señal de audio se presenta de forma discreta en el tiempo o mediante la indicación de puntos temporales absolutos. En un caso de este tipo, el dispositivo 12 para generar la sección de señal de prueba basándose en la sección portadora de información de la señal de audio está operativo para, utilizando las indicaciones correspondientes en la línea 14, extraer la sección portadora de información correspondiente, o al menos una parte de la sección portadora de información, directamente de la señal de audio que se facilita a través de una entrada 16, tal como se muestra mediante una línea de conexión desde la entrada 16 al dispositivo 12, que en la figura 1 se indica con 18. Tal como ya se ha indicado, el dispositivo 12 genera en una salida 20 del dispositivo mostrado en la figura 1 una secuencia de secciones de señal de prueba.

40 **[0035]** La figura 4 muestra el modo de funcionamiento del concepto según la invención mostrado mediante la figura 1. El dispositivo 10 para el análisis está operativo para detectar la sección portadora de información que se

extiende aproximadamente de 1,3 segundos a 8,8 segundos. Para analizar la señal de audio en relación con una sección portadora de información y una sección no portadora de información, por ejemplo, una pausa o un ruido, existen en la técnica muchas medidas conocidas tales como, por ejemplo, una detección de actividad de voz, procedimientos de predicción, procedimientos de detección de pausas, detecciones de nivel, procedimientos de gradientes, etc. Todos estos procedimientos se basan en analizar una rápida modificación de la amplitud de señal de una modificación lenta de la amplitud de señal considerando la modificación absoluta durante un determinado intervalo de tiempo. Las rápidas modificaciones que tienen lugar adicionalmente a una determinada amplitud, es decir, a un determinado nivel de intensidad sonora, indican las denominadas partes de modulación con actividad de voz si la señal es una señal de voz tal como se muestra en la figura 4. Por el contrario, las modificaciones lentas indican un reducido nivel o rápidos cambios en un nivel de ruido relativamente constante indican pausas o ruidos, es decir, secciones no portadoras de información de la señal de audio. Otros procedimientos diferencian, por ejemplo, ruido de voz mediante análisis espectral y correlación.

[0036] Ahora, el dispositivo 12 para generar las secciones de señal de prueba está operativo para, por ejemplo, realizar la fragmentación de la señal de audio en secciones de señal de prueba $m(1)$, $m(2)$, $m(3)$, ..., de modo que a una sección portadora de información le preceda y le siga una sección no portadora de información, tal como puede observarse, por ejemplo, mediante las secciones de señal de prueba $m(1)$, $m(2)$, $m(3)$ de la figura 4. En particular, el dispositivo para generar la sección de señal de prueba tal como se muestra en 12 en la figura 1 está operativo para extraer una sección portadora de información de la señal de audio directamente de la señal de audio y añadir al menos una parte de la sección no portadora de información precedente de la señal de audio y una parte de la sección no portadora de información siguiente con el objeto de obtener una sección de señal de prueba con una determinada longitud predefinida.

[0037] Esta forma de proceder es viable cuando la sección portadora de información de la señal de audio es más corta que una longitud máxima predeterminada de una sección de señal de prueba, es decir, por ejemplo, 12 segundos o hasta 20 segundos. Esta condición marginal debería darse, en particular, en el caso de señales de voz tales como las que se presentan en conversaciones telefónicas.

[0038] Sin embargo, si la señal de audio es una señal de música, entonces puede producirse perfectamente el caso de que una sección portadora de información de la señal de música, es decir, una sección con una modulación por encima de un determinado valor umbral de modulación, sea más larga que la longitud máxima predeterminada. Si este es el caso, entonces el dispositivo 12 para generar la sección de señal de prueba está operativo para generar una sección de señal de prueba en el sentido de que primero, partiendo de una situación de pausa, se incrementa poco a poco el volumen de sonido de la sección portadora de información en el sentido de que se reduce una atenuación de 1 a 0. Después, la sección portadora de información se toma directamente de la señal de audio y, en concreto, hasta un instante predeterminado en el que entonces tiene lugar nuevamente una disminución gradual del volumen de sonido al incrementarse nuevamente un factor de atenuación de 0 a 1 para, finalmente, generar artificialmente, es decir, sintetizar, al final de la sección de señal de prueba nuevamente una situación de pausa.

[0039] A continuación, se proporciona una representación detallada del dispositivo 10 para el análisis haciendo referencia a la figura 2. Nuevamente, el dispositivo 10 para el análisis de la señal de audio obtiene la señal de audio de una entrada de señales de audio 10. En el lado de salida, el dispositivo de análisis de señales 10 proporciona una indicación de la sección portadora de información o la propia sección portadora de información. De forma alternativa o adicional, el dispositivo 10 proporciona una indicación de una sección no portadora de información o la propia sección no portadora de información, tal como se muestra mediante una línea de salida 22. El dispositivo de análisis de señales 10 realiza, en ejemplos de realización preferidos de la presente invención, un análisis de la señal mediante una detección de actividad de voz, una detección de pausas/ruidos, una detección de nivel, una detección de intensidad sonora, una detección de modulación, etc. Todos estos conceptos se basan en que una sección portadora de información está correlacionada con la característica que ha de detectarse en el sentido de que la característica puede detectarse si la sección de la señal de audio porta información útil, y la característica que ha de detectarse no está presente cuando la sección de la señal de audio que se considera no porta información útil correspondiente, o a la inversa. Si la señal de audio se presenta, por ejemplo, como señal de voz, entonces el dispositivo para el análisis de la señal está operativo para, mediante una detección de actividad de voz ("Voice Activity Detection"), determinar el comienzo y el final de una sección portadora de información de la señal de audio, es decir, por ejemplo, una parte de modulación con actividad de voz. De forma alternativa o adicional, el dispositivo de análisis de señales 10 puede realizar una detección de pausas o, en redes reales, una detección de ruidos ("Noise Detection") para determinar la posición y longitud de las pausas de voz.

[0040] Una sección portadora de información es entonces la sección entre dos pausas de voz, aunque no se detecte directamente sino, en cierto modo, de forma indirecta al determinar la pausa de voz precedente y siguiente. Por tanto, una detección de pausas únicamente, es decir, una detección de una sección no portadora de información precedente y una siguiente en relación con una sección portadora de información considerada, proporciona también una diferenciación de la sección portadora de información de la señal de audio respecto de una sección no portadora de información precedente de la señal de audio y una sección no portadora de información subsiguiente de la señal de audio.

[0041] De forma alternativa o adicional, el dispositivo para el análisis de la señal de audio puede estar configurado para, mediante una detección de voz posterior que en la técnica también se conoce como ASR o “Automatic Speech Recognition” (reconocimiento automático de voz), analizar la voz así como las relaciones de frases para, por ejemplo, cuando se requiera, extraer siempre una sección de señal de prueba con un número predeterminado de palabras o un número predeterminado de frases. Esta funcionalidad también puede asumirla, tal como se muestra mediante la figura 3, el dispositivo 12 para generar las secciones de señal de prueba que se muestra en la figura 3 según un ejemplo de realización preferido de la presente invención. A través de líneas de transmisión 14 o 12 en la figura 2, el dispositivo 12 para generar una sección de señal de prueba obtiene, por ejemplo, una indicación de una sección portadora de información, es decir, una modulación, o una indicación de una pausa, es decir, una sección no portadora de información de la señal de audio que se facilita al dispositivo 12 a través de la línea de transmisión 18. Además, el dispositivo 12 obtiene información sobre una longitud máxima predeterminada a través de otra entrada 24, así como información sobre una longitud mínima predeterminada a través de otra entrada 26. Asimismo, en un ejemplo de realización preferido de la presente invención, se pretende una relación pausas/modulación predeterminada en una sección de señal de prueba. La información cuya relación pausas/modulación está predeterminada pueden alimentarse a través de otra entrada 28 al dispositivo 12 para generar las secciones de señal de prueba.

[0042] El dispositivo 12 para generar una sección de señal de prueba está operativo para añadir pausas al comienzo y al final de una sección portadora de información identificada con el objeto de generar una relación definida de modulación de voz a pausa tal como, por ejemplo, 40% de modulación de voz y 60% de pausa. Si una sección portadora de información fuera demasiado larga, entonces, en un ejemplo de realización preferido de la presente invención, el dispositivo 12 está operativo para crear una funcionalidad de incremento y disminución progresivos del volumen de sonido para incrementar o reducir de forma suave el volumen de sonido de la sección portadora de información, lo cual también se conoce en la técnica como “fade-in” y “fade-out”. Si una sección portadora de información de la señal de audio no es tan larga como la longitud máxima predeterminada, pero es más larga que la predeterminada por la relación pausas/modulación en una sección de señal de prueba de modo que la funcionalidad de incremento y disminución progresivos del volumen de sonido también puede realizarse con el objeto de “acortar” la sección portadora de información en aras de una sección no portadora de información más larga.

[0043] En un ejemplo de realización preferido de la presente invención, se realiza además un tratamiento recursivo de los pasos de la detección de actividad de voz, la detección de pausas, la detección de voz posterior y la adición de pausas para formar con diferente longitud secciones de señal de prueba adecuadas para la prueba de audición que, por ejemplo, pueden convertirse en secuencias de voz cuya duración correspondiente, no obstante, se sitúa dentro de la longitud mínima t_{\min} predeterminada y la longitud máxima t_{\max} predeterminada.

[0044] Por tanto, el concepto según la invención es efectivo para generar para cada señal de audio de duración t una serie de i secciones de señal de prueba, donde se cumple:

$$0 \leq i \leq \frac{t}{t_{\min}}$$

donde t_{\min} define la duración mínima predeterminada para una secuencia.

[0045] Las secciones de señal de prueba o fragmentos de la señal de audio generados de esta manera, que, tal como se muestra en la figura 4, pueden tener solapamientos en un ejemplo de realización preferido de la presente invención, se alimentan ahora a otro proceso, por ejemplo, un procedimiento de medición basado en la percepción para determinar la calidad de la voz según ITU-T P.862 (PESQ). Al mismo tiempo, los fragmentos generados también pueden utilizarse para una prueba de audición subjetiva, por ejemplo, según ITU-T P.800.

[0046] Ha de observarse que las secciones de señal de prueba generadas según la invención, tal como se muestran en la figura 4, a diferencia de las secciones de señal de prueba de longitud constante, ya no contienen secuencias que solo comprendan una pausa, es decir, “señales nulas”, de modo que el número de las secuencias se minimiza en función de la modulación para una señal de audio determinada y, en el caso extremo, puede ser incluso 5 0 si la señal de audio no presenta ninguna sección portadora de información, es decir, se compone única y exclusivamente de una pausa o ruidos. Este resultado coincide con el requisito de que un procedimiento de medición no puede valorar ninguna señal que esté compuesta exclusivamente por ruidos o pausa.

[0047] A continuación, se explica una fragmentación de una señal de audio en secuencias de prueba adecuadas para una prueba de audición según ITU-R BS.111.6 o BS.1534.

10 **[0048]** De forma análoga a la fragmentación según la invención de una señal de voz en secuencias de prueba adecuadas para una prueba de audición según ITU-T P.800, se fragmenta según la invención una señal de música también en secuencias de una longitud de aproximadamente 10 a 20 segundos. En un ejemplo de realización preferido, el dispositivo 10 para el análisis está configurado para realizar, en caso de una señal de música, una 15 detección de nivel, una detección de intensidad sonora o una detección de modulación con el objeto de determinar el principio y el final de partes de modulación, es decir, una sección portadora de información. Además, el dispositivo 10 está configurado para, mediante una detección de pausas o, en caso de redes reales, una detección de ruidos, determinar la posición y longitud de las pausas que en la técnica también se denominan intervalos de silencio.

20 **[0049]** Nuevamente, en caso de una modulación demasiado larga, es decir, una sección portadora de información demasiado larga, se prefiere también incrementar o reducir de forma suave el volumen de sonido de la señal de música mediante un automatismo de incremento y disminución progresivos del volumen de sonido. Además, se prefiere generar una relación definida de modulación a pausa, por ejemplo, 40:60, añadiendo pausas al principio y al final de una sección portadora de información identificada.

25 **[0050]** De nuevo, en un ejemplo de realización preferido de la presente invención, se realiza un uso recursivo de los pasos de detección de nivel, detección de intensidad sonora o detección de modulación, detección de pausas, el automatismo de incremento y disminución progresivos del volumen de sonido y la adición de pausas para formar secuencias de audio de diferente longitud cuya duración se sitúe dentro de longitudes mínimas y máximas t_{\min} y t_{\max} predeterminadas.

30 **[0051]** En relación con la definición a modo de ejemplo de la intensidad sonora y la modulación, se remite al documento de estandarización Rec. ITU-R BS.1387-1, sección 3.2 en relación con la modulación y sección 3.3 en relación con la intensidad sonora. Estas secciones se asumen por referencia en el presente documento.

[0052] Como resultado se obtiene nuevamente para cada señal de audio de duración t una serie de i secuencias de medición en donde se cumple lo siguiente:

$$0 \leq i \leq \frac{t}{t_{\min}}$$

donde t_{\min} define la duración mínima predeterminada para una secuencia.

35 **[0053]** Los fragmentos o secciones de señal de prueba así obtenidos o tratados pueden alimentarse ahora a otro proceso, por ejemplo, a un procedimiento de medición basado en la percepción para determinar la calidad del audio según ITU-R BS.1387-1 PEAQ. Al mismo tiempo, los fragmentos generados también pueden emplearse para una prueba de audición subjetiva.

40 **[0054]** En un ejemplo de realización alternativo de la presente invención, se prefiere analizar las secciones de señal de prueba sucesivas obtenidas de una medición periódica de la intensidad sonora o corrección de la intensidad sonora, por ejemplo, utilizando procedimientos conocidos tales como se definen en ITU-R WP6P cuestión 2/6 “Audio Metering Characteristics suitable for use in Digital Sound Production”. Este concepto sirve especialmente para crear una adaptación de la intensidad sonora de señales de audio, por ejemplo, en televisión para luchar contra

el problema de oscilaciones de nivel en señales de audio conocido con el nombre "publicidad demasiado alta". El concepto según la invención es ventajoso en este caso especialmente en el sentido de que puede controlarse de forma precisa según la presente invención la relación pausas/modulación de las secciones de señal de prueba basándose en la extracción adaptada a la señal de las secciones de señal de prueba, es decir, no se producen 5 influencias erróneas en la intensidad sonora del sonido debido a una fragmentación imprecisa de la señal de audio.

[0055] En función de las particularidades, el procedimiento según la invención para la extracción de una sección de señal de prueba se implementa en hardware o software. La implementación puede realizarse en un medio de almacenamiento digital, en particular, un disquete o CD, con señales de control que pueden leerse electrónicamente y que así pueden actuar conjuntamente con un sistema informático programable en el que se ejecuta el 10 procedimiento. Por tanto, en general, la invención consiste también en un producto de programa informático con un código de programación almacenado en un soporte legible por máquina para la realización del procedimiento según la invención cuando el producto de programa informático se ejecuta en un ordenador. Por tanto, dicho de otro modo, la invención puede realizarse como un programa informático con un código de programación para la implementación del procedimiento para la extracción de una sección de señal de prueba a partir de una señal de audio cuando el 15 programa informático se ejecuta en un ordenador.

REIVINDICACIONES

1. Dispositivo para extraer una sección de señal de prueba de una señal de audio que comprende las siguientes características: un dispositivo (10) para analizar una estructura temporal de la señal de audio con el objeto de distinguir una sección portadora de información de la señal de audio de una parte de pausa precedente no portadora de información de la señal de audio y una parte de pausa subsiguiente no portadora de información de la señal de audio; un dispositivo (12) para generar una secuencia de secciones de señal de prueba basándose en la sección portadora de información de la señal de prueba; y un dispositivo (20) para transmitir la sección de señal de prueba a un dispositivo de medición para la valoración de la calidad de un sistema de transmisión (600) desde el cual puede obtenerse la señal de audio, estando configurado el dispositivo (12) para la generación para añadir una o varias partes de pausa a la sección portadora de información de la señal de audio de modo que la sección de señal de prueba comprenda una relación predeterminada de longitud temporal de la sección portadora de información y longitud temporal de la sección o las secciones de pausa.
2. Dispositivo según la reivindicación 1, en el que la señal de audio presenta partes de voz y partes de pausa entre las partes de voz, y en el que el dispositivo (10) para el análisis está configurado para llevar a cabo una detección de actividad de voz con el objeto de detectar una parte de voz como una sección portadora de información dispuesta, en relación temporal, entre dos partes de pausa como secciones no portadoras de información.
3. Dispositivo según la reivindicación 1 o 2, en el que la señal de audio presenta partes de señal que tienen una potencia mínima predeterminada y partes de pausa que tienen una potencia inferior a la potencia de señal predeterminada, y en el que el dispositivo (10) para el análisis está configurado para llevar a cabo una detección de pausas para detectar dos partes de pausa sucesivas en el tiempo entre las cuales se dispone una parte de señal como una sección portadora de información.
4. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que la señal de audio presenta partes de ruido y partes de señal en medio, y en el que el dispositivo (10) para el análisis está configurado para llevar a cabo una detección de ruido con el objeto de detectar dos partes de ruido sucesivas en el tiempo para determinar una parte de señal que se disponga entre las partes de ruido como una sección portadora de información.
5. Dispositivo según la reivindicación 2, en el que el dispositivo (10) para el análisis está configurado adicionalmente para someter la parte de voz a una detección de voz y para analizar la sección portadora de información en relación con la voz y/o una relación de frases, y en el que el dispositivo (12) para la generación está configurado para considerar la voz y/o la relación de frases al generar la sección de señal de prueba.
6. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que el dispositivo (12) para la generación está configurado para generar la sección de señal de prueba de modo que la misma tenga una longitud temporal superior o igual a una longitud mínima predeterminada e inferior o igual a una longitud máxima predeterminada.
7. Dispositivo según la reivindicación 6, en el que la longitud mínima predeterminada se sitúa entre 2 y 12 segundos, y en el que la longitud máxima predeterminada se sitúa entre 12 y 25 segundos.
8. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que el dispositivo (12) para la generación está configurado para tomar al menos una parte de la sección portadora de información de la señal de audio sin modificar en la sección de señal de prueba.
9. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que el dispositivo (12) para la generación está configurado para detectar una longitud temporal de la sección portadora de información con el objeto de determinar si la longitud temporal supera un valor máximo predeterminado, y, en el caso de superar el valor máximo predeterminado, disminuir progresivamente el volumen de sonido de una parte al comienzo o al final de la sección portadora de información para obtener una sección portadora de información modificada cuya longitud temporal sea menor que el valor máximo predeterminado, y, en un intervalo de transición entre una parte en la que se ha disminuido el volumen de sonido y una parte no modificada, obtener una modificación continua del nivel de la sección portadora de información modificada, y con el objeto de generar la sección de señal de prueba de modo que la misma incluya el intervalo de transición y la sección no modificada.
10. Dispositivo según cualquiera de las reivindicaciones 1 a 8, en el que el dispositivo (12) para la generación está configurado para, en el caso de que la longitud temporal de la sección portadora de información sea menor que una longitud máxima predeterminada, tomar toda la sección portadora de información en la sección de señal de prueba.

11. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que el dispositivo (12) para la generación está configurado para generar dos secciones de señal de prueba sucesivas, situándose un punto final de una primera sección de señal de prueba, en relación con la señal de audio, temporalmente tras un punto de inicio de una segunda sección de señal de prueba sucesiva, de modo que tanto la primera sección de señal de prueba como también la segunda sección de señal de prueba tengan en común al menos una parte de una sección no portadora de información de la señal de audio.
12. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que el dispositivo (10) para el análisis y el dispositivo (12) para la generación están configurados para transformar la señal de audio en una secuencia de secciones de señal de prueba de diferente longitud, siendo cada sección de señal de prueba más larga o igual que una longitud mínima predeterminada y más corta o igual que una longitud máxima predeterminada.
13. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que el dispositivo (10) para el análisis está configurado para emitir un mensaje cuando en la señal de audio no pueda detectarse ninguna sección portadora de información.
14. Dispositivo según la reivindicación 1, en el que la señal de audio presenta una parte de música, y en el que el dispositivo (10) para el análisis está configurado para realizar una detección de nivel, una detección de intensidad sonora o una detección de modulación con el objeto de diferenciar una sección no portadora de información de una sección portadora de información.
15. Dispositivo según cualquiera de las reivindicaciones anteriores, en el que el dispositivo (12) para la generación está configurado para generar una secuencia de secciones de señal de prueba y presenta además un dispositivo para la medición de la intensidad sonora para obtener, para cada sección de señal de prueba, un valor de intensidad sonora, pudiendo emplearse los valores de intensidad sonora de las secciones de señal de prueba de una retroalimentación de intensidad sonora para controlar, en función de la retroalimentación, un desarrollo de la intensidad sonora de la señal de audio.
16. Procedimiento para extraer una sección de señal de prueba de una señal de audio con los siguientes pasos: analizar (10) una estructura temporal de la señal de audio con el objeto de diferenciar una sección portadora de información de la señal de audio de una parte de pausa precedente no portadora de información de la señal de audio y una parte de pausa subsiguiente no portadora de información de la señal de audio; generar (12) la sección de señal de prueba basándose en la sección portadora de información de la señal de audio; y transmitir la sección portadora de información a un dispositivo de medición para la valoración de la calidad de un sistema de transmisión (600) a partir del cual puede obtenerse la señal de audio, presentando la generación una adición de una o varias secciones de pausa a la sección portadora de información de la señal de audio de modo que la sección de señal de prueba presente una relación predeterminada de longitud temporal de la sección portadora de información y longitud temporal de la sección o las secciones de pausa.
17. Dispositivo para la medición de la calidad de un canal de transmisión con las siguientes características: un dispositivo para la recepción de una señal de audio del canal de transmisión; un dispositivo para la extracción de una o varias secciones de señal de prueba según cualquiera de las reivindicaciones 1 a 15; y un dispositivo para la valoración de la calidad del canal de transmisión basándose en la una sección o las varias secciones de señal de prueba.
18. Dispositivo según la reivindicación 17, en el que el dispositivo para la extracción está configurado para transformar la señal de audio en una secuencia de secciones de señal de prueba de diferente longitud, siendo cada una de las secciones de señal de prueba más larga o igual que una longitud mínima predeterminada y más corta o igual que una longitud máxima predeterminada, estando compuesta cada una de las secciones de señal de prueba, al menos en una proporción predeterminada, por una sección portadora de información de la señal de audio, y estando configurado el dispositivo para la valoración de la calidad para generar un valor de medición de la calidad para secciones de señal de prueba de la sucesión de secciones de señal de prueba.
19. Procedimiento para la medición de la calidad de un canal de transmisión con los siguientes pasos: recepción de una señal de audio del canal de transmisión; extracción de una o varias secciones de señal de prueba utilizando el procedimiento según la reivindicación 16; y valoración de la calidad del canal de transmisión basándose en la una sección o las varias secciones de señal de prueba.
20. Programa informático con un código de programación para la implementación del procedimiento para la extracción de una sección de señal de prueba según la reivindicación 16, o para la implementación del procedimiento para la medición de la calidad según la reivindicación 19 cuando el programa informático se ejecuta en un ordenador.

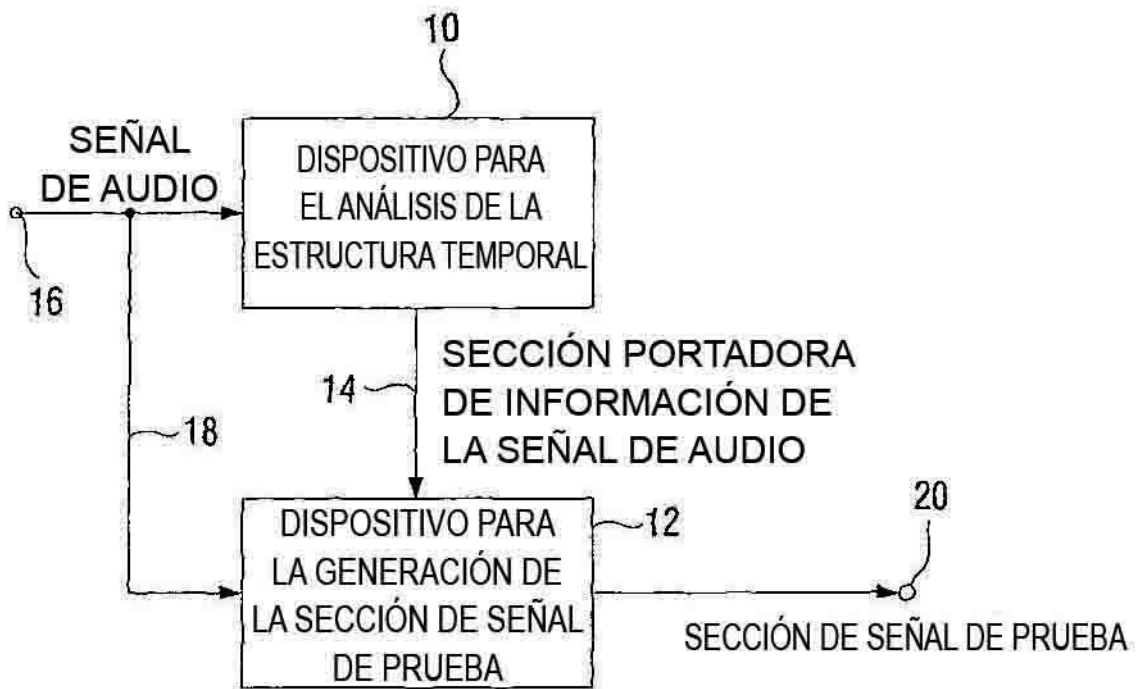


FIG. 1

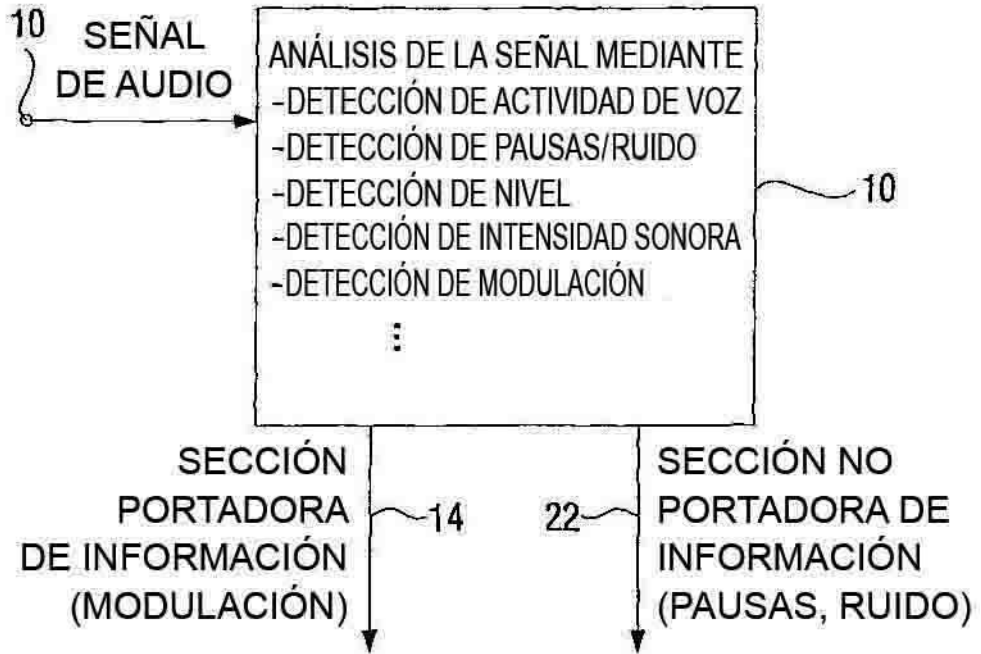


FIG. 2

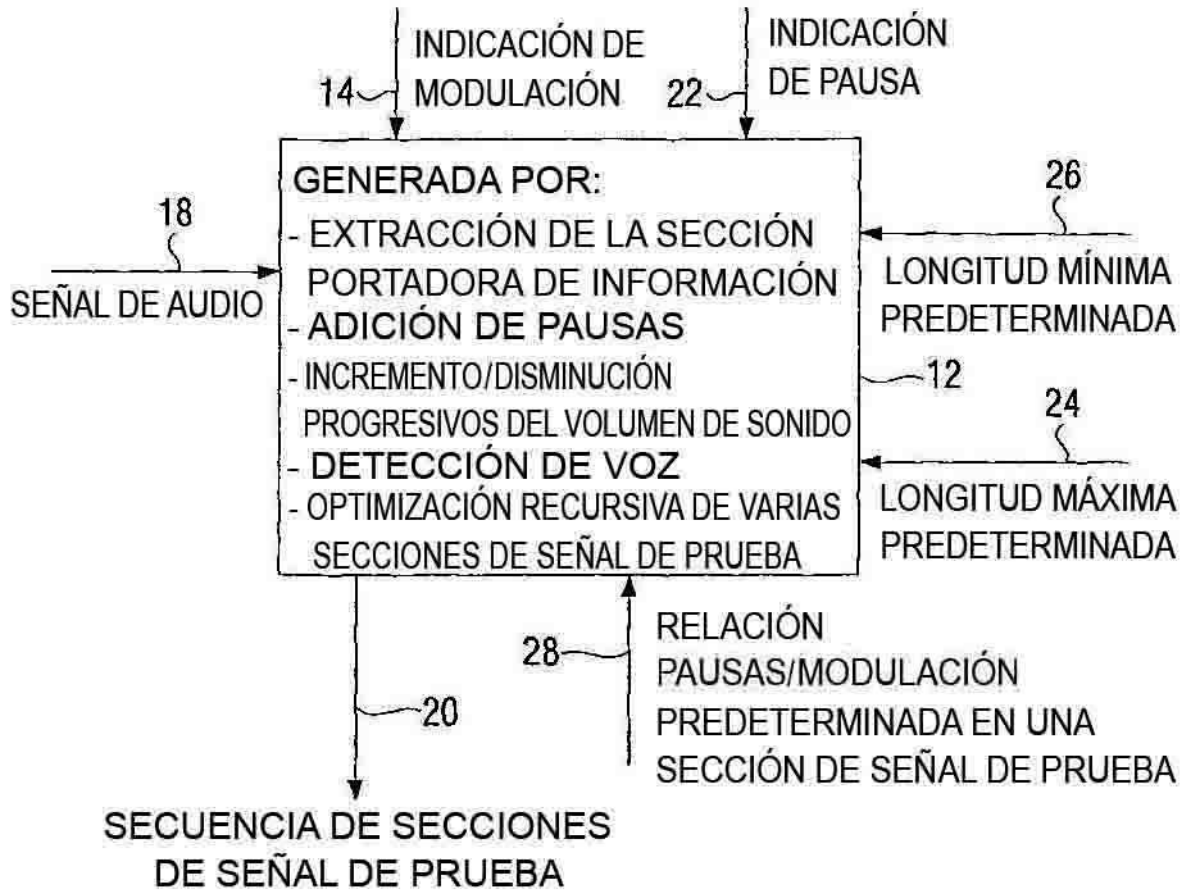


FIG. 3

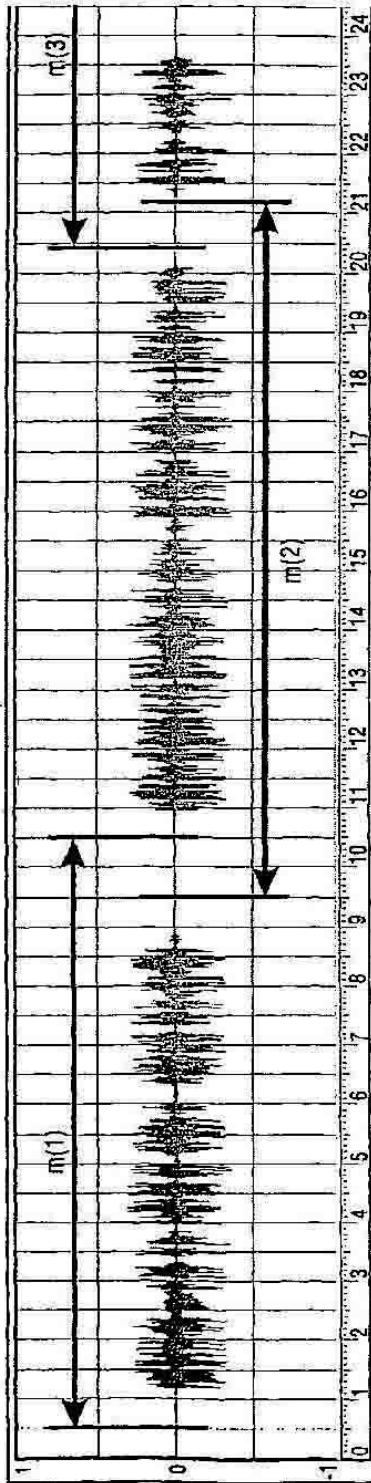


FIG. 4 (INVENCIÓN)

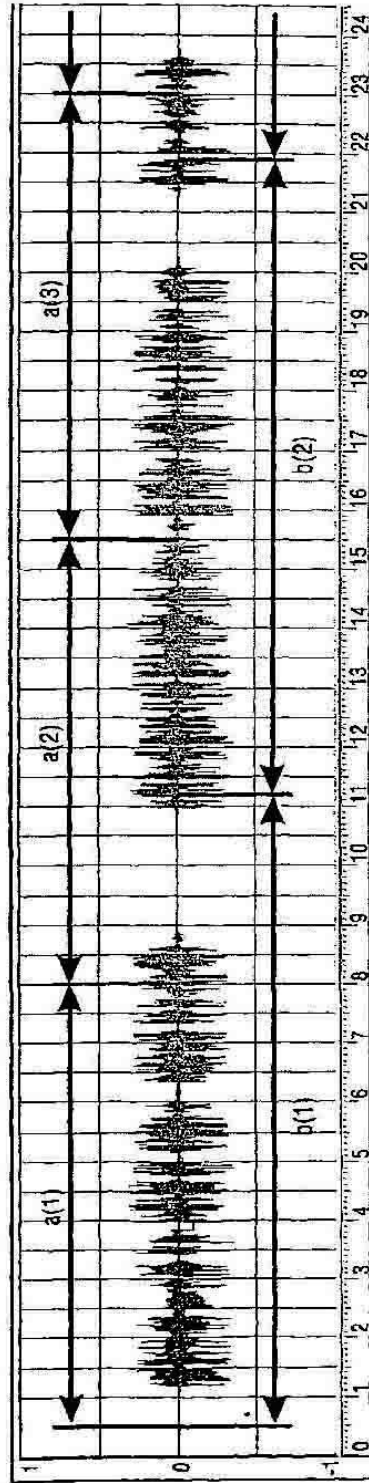


FIG. 5

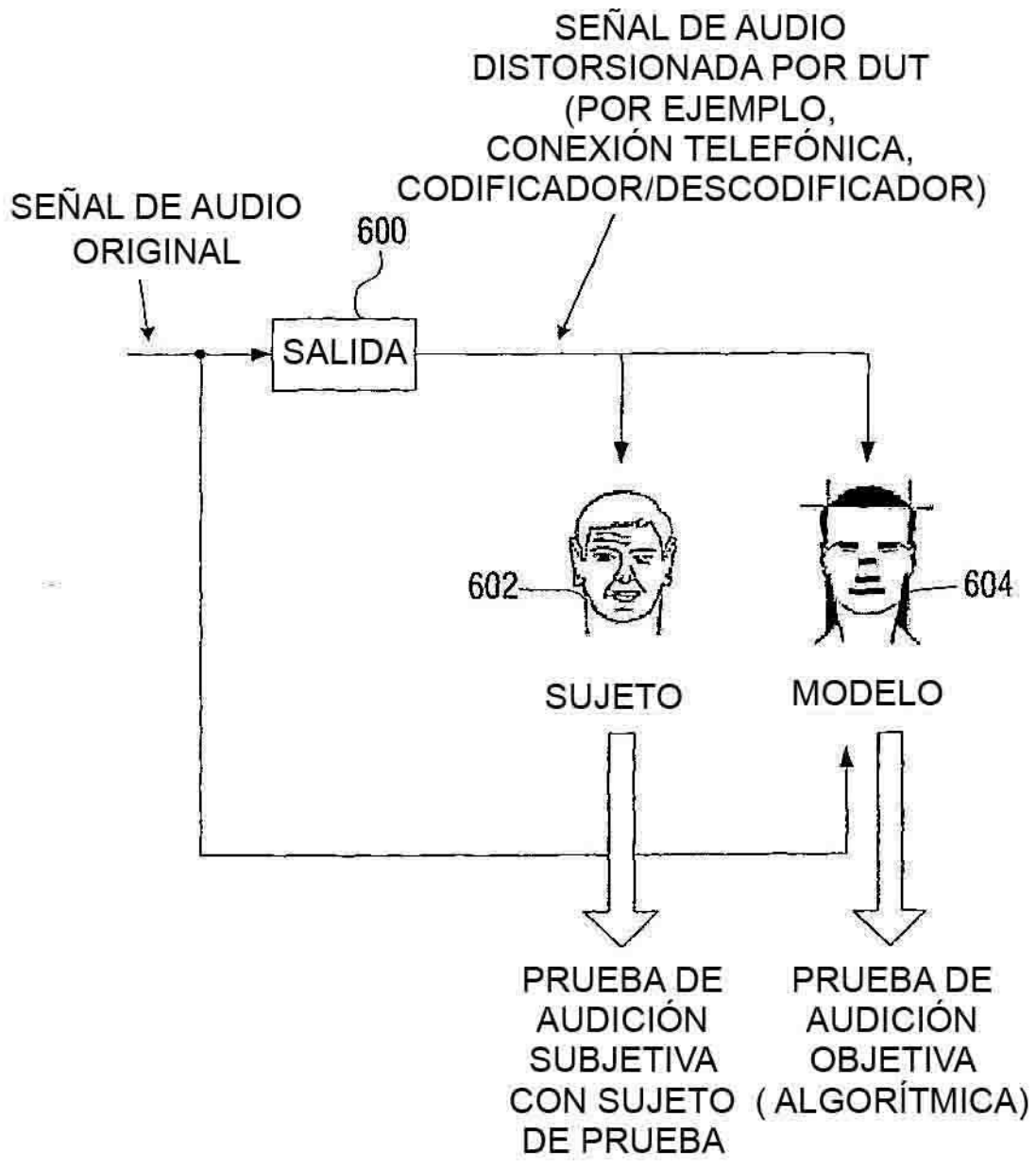


FIG. 6