(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 531 556**

(51) Int. CI.:

G10L 25/69 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **09.08.2010 E 10751828 (4)**

(97) Fecha y número de publicación de la concesión europea: **24.12.2014 EP 2465113**

(54) Título: **Método, producto de programa de ordenador y sistema para determinar una calidad percibida de un sistema de audio**

(30) Prioridad:

14.08.2009 EP 09010501
04.05.2010 EP 10161830

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
17.03.2015

(73) Titular/es:

KONINKLIJKE KPN N.V. (50.0%)
Maanplein 55
2516 CK The Hague, NL y
NEDERLANDSE ORGANISATIE VOOR TOEGEPAST -NATUURWETENSCHAPPELIJK ONDERZOEK TNO (50.0%)

(72) Inventor/es:

BEERENDS, JOHN y
VAN VUGT, JEROEN

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

DESCRIPCIÓN

Método, producto de programa de ordenador y sistema para determinar una calidad percibida de un sistema de audio

Campo de la invención

La invención está relacionada con un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio con respecto a una señal de referencia. La invención está relacionada, además, con un producto de programa de ordenador que comprende un código ejecutable de ordenador, por ejemplo almacenado en un medio legible por un ordenador, adaptado para poner en práctica dicho método cuando es ejecutado por un procesador. Por último, la invención está relacionada con un sistema para determinar un indicador de calidad que represente una calidad percibida de una señal de salida de un sistema de audio con respecto a una señal de entrada del sistema de audio que sirve como señal de referencia.

Antecedentes de la invención

La calidad de un dispositivo de audio se puede determinar de forma subjetiva u objetiva. Las pruebas subjetivas consumen tiempo, son caras y difíciles de reproducir. Por lo tanto, se han desarrollado varios métodos para medir de forma objetiva la calidad de una señal de salida, en particular una señal de voz, de un dispositivo de audio. En dichos métodos, se determina la calidad de la voz de una señal de salida tal como se recibe desde un sistema de procesamiento de señales de voz mediante la comparación con una señal de referencia.

Un método actual que se utiliza ampliamente para este propósito es el método descrito en la Recomendación P.862 de la ITU-T titulada “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs (Evaluación de percepción de la calidad de voz (PESQ): Un método objetivo para la evaluación de la calidad de voz extremo a extremo de redes de telefonía de banda estrecha y codecs de voz)”. En la recomendación P.862 de la ITU-T, se debe determinar la calidad de una señal de salida desde un sistema de procesamiento de señales de voz, cuya señal en general está distorsionada. La señal de salida y una señal de referencia, por ejemplo la señal de entrada del sistema de procesamiento de señales de voz, se mapean sobre señales de representación en función de un modelo de percepción psicofísico del sistema de audición humana. En función de estas señales, se determina una señal diferencia que representa la distorsión de la señal de salida comparada con la señal de referencia. En general se define un indicador de calidad que representa una calidad percibida de una señal de salida como un indicador que muestra una alta correlación con la calidad de voz percibida de forma subjetiva. El indicador de calidad se denomina comúnmente Puntuación de Opinión Media (MOS) ya que se determina mediante una prueba subjetiva en la que los sujetos (humanos) expresan su opinión sobre una escala de calidad. En general el indicador de calidad se deriva de una comparación de la representación interna de la señal de salida de un dispositivo sometido a evaluación con la representación interna de la señal de entrada al dispositivo sometido a evaluación. La representación interna se puede calcular transformando la señal desde el dominio externo físico al dominio interno psicofísico. En la recomendación P.862 de la ITU-T el núcleo del algoritmo que se utiliza en el cálculo de la representación psicofísica de la señal está compuesto por las siguientes operaciones principales, escalado a un nivel fijo, alineación de tiempo, transformación del dominio amplitud tiempo al dominio potencia tiempo frecuencia, ajuste de la escala de potencia y frecuencia. Las operaciones dan como resultado una representación interna en términos de intensidad sonora, tiempo y tono a partir del cual se pueden calcular las funciones diferencia. Estas funciones diferencia se utilizan a continuación para obtener un indicador único de calidad. Para cada archivo de voz se puede obtener de este modo una puntuación MOS y una puntuación de indicador de calidad que deberían tener la mayor correlación posible entre ellas. Como ejemplo se puede determinar la calidad de un codec de voz comparando las representaciones internas de la salida del codec con las representaciones internas de la entrada del codec. Para cada archivo de voz que se codifica utilizando el codec el indicador de calidad producirá un número que debería tener una correlación alta con la puntuación MOS determinada de forma subjetiva para el archivo de voz codificado/decodificado. La señal diferencia se procesa a continuación de acuerdo con un modelo cognitivo, en el que se han modelado ciertas propiedades de la percepción auditiva humana basándose en pruebas, con el fin de obtener una calidad de señal que sea una medida de la calidad de la percepción auditiva de la señal de salida.

Tal como ha indicado claramente la recomendación P.862 de la ITU-T, se sabe que PESQ proporciona predicciones inexactas cuando se utilizan niveles auditivos variables. PESQ asume un nivel de audición estándar de 79 dB SPL (Nivel de Presión Sónica) y compensa la señal de entrada cuando existen niveles de señal no óptimos. Por lo tanto no se tiene en cuenta el efecto subjetivo de la desviación respecto a los niveles de audición óptima. En los sistemas actuales de telecomunicaciones, en particular utilizando sistemas de Voz Sobre IP (VOIP) y tecnologías similares, se producen muy a menudo niveles de audición no óptima. En consecuencia, PESQ en general no proporciona predicciones óptimas de la percepción de las señales de voz procesadas en dichos sistemas de telecomunicación, los cuales se están convirtiendo cada vez en más habituales.

La publicación “Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech

quality assessment part II-Psychoacoustic model (Evaluación perceptiva de la Calidad de Voz (PESQ) El Nuevo Estándar de la ITU para el modelo Psicoacústico de evaluación de la calidad de Voz Extremo a Extremo parte II (por Beerends J G y otros, Revista de la sociedad de ingeniería acústica, sociedad de ingeniería acústica, Nueva York, NY, EE.UU., vol. 50, núm. 10) muestra un nuevo modelo para la evaluación perceptiva de la calidad de voz (PESQ) que fue estandarizado por la Unión Internacional de Telecomunicaciones como Recomendación P.862. A diferencia de los modelos de evaluación de codecs anteriores, como por ejemplo PSQM y MNB (ITU-T P.861), PESQ permite predecir la calidad subjetiva con una buena correlación en un muy amplio rango de condiciones, las cuales pueden incluir distorsiones de codificación, errores, ruido, filtrado, retardo, y retardo variable.

Resumen de la invención

Es deseable disponer de un método para determinar la calidad de las transmisiones de un sistema de audio que proporcione una correlación mejorada entre la calidad de la voz como la determinada por una medición objetiva y la calidad de la voz como la determinada en una prueba subjetiva. Para este propósito, un modo de realización de la invención está relacionado con un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio de acuerdo con la reivindicación 1. En otro modo de realización de acuerdo con la reivindicación 3, la separación de las acciones de escalado local permite una implementación por separado y/o una manipulación de las variaciones de nivel debido a cortes y pulsos de tiempo.

En un modo de realización adicional de acuerdo con la reivindicación 4, el nivel de ruido predeterminado se puede corresponder con un nivel de ruido que se considere que sea un nivel de ruido bajo deseable con el fin de servir como una representación ideal para la señal de salida. En todavía otro modo de realización adicional de acuerdo con la reivindicación 5, la supresión de ruido de la señal de salida puede permitir la supresión del ruido hasta un nivel de ruido que represente la perturbación experimentada por el dispositivo sometido a evaluación.

Se ha comprobado que una supresión de ruido adicional después de un escalado global en un modo de realización de acuerdo con la reivindicación 6 mejora aún más la correlación entre una calidad de voz medida de forma objetiva y una calidad de voz tal como la obtenida en los experimentos de calidad de escucha subjetiva. En algunos modos de realización de la invención, la invención está relacionada, además, con un producto de programa de ordenador de acuerdo con la reivindicación 8. Por último, en algunos modos de realización de la invención, la invención está relacionada, además, con un sistema para determinar un indicador de calidad que represente una calidad percibida de una señal de salida Y(t) de un sistema de audio de acuerdo con la reivindicación 9.

Breve descripción de los dibujos

En los dibujos:

la FIG. 1 muestra de forma esquemática una configuración general que incluye un sistema para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio con respecto a una señal de referencia;

la FIG. 2 muestra de forma esquemática un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio, con respecto a una señal de referencia de acuerdo con PESQ;

la FIG. 3 muestra de forma esquemática un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio, con respecto a una señal de referencia de acuerdo con un modo de realización de la invención; y

la FIG. 4 muestra de forma esquemática un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio, con respecto a una señal de referencia de acuerdo con un modo de realización adicional de la invención.

Descripción detallada de los modos de realización de la invención

A continuación se realiza una descripción de ciertos modos de realización de la invención, proporcionados únicamente a modo de ejemplo.

A lo largo de la descripción, los términos “local” y “global” se utilizarán con respecto a una operación realizada sobre una señal. Una operación “local” se refiere a una operación realizada sobre una parte de la señal en el tiempo, por ejemplo una única trama. Una operación “global” se refiere a una operación realizada sobre toda la señal.

A lo largo de la descripción, los términos “salida” y “distorsionado” se pueden utilizar en relación con una señal que se origina en una salida de un sistema de audio, como un dispositivo de procesamiento de voz. A lo largo de la descripción, los términos “referencia” y “original” se pueden utilizar en relación con una señal ofrecida como una entrada al sistema de audio, siendo utilizada la señal, además, como una señal con la que se va a comparar la señal de salida o distorsionada.

La FIG. 1 muestra de forma esquemática una configuración general que incluye un sistema para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio, por ejemplo un dispositivo de procesamiento de voz, con respecto a una señal de referencia. Dicho método pretende obtener una medida objetiva de la calidad de la transmisión de un sistema de audio. La configuración incluye un sistema 10 de audio sometido a evaluación, por ejemplo una red de telecomunicaciones, un elemento de red o un dispositivo de procesamiento de voz en una red o estación móvil. La configuración también incluye un sistema 20 para medir la calidad de la transmisión del sistema de audio, denominado de aquí en adelante como sistema 20 de medición de la calidad.

El sistema 20 de medición de la calidad se dispone para recibir dos señales de entrada. Una primera señal de entrada es una señal de voz *X(t)* que se proporciona directamente al sistema 20 de medición de la calidad (esto es, no se proporciona a través del sistema 10 de audio), y sirve como señal de referencia. Una segunda señal de entrada es una señal de voz *Y(t)* que se corresponde con la señal de voz *X(t)* que se ve afectada por el sistema 10 de audio. El sistema 20 de medición de la calidad proporciona una señal Q de calidad de la salida que representa una estimación de la calidad perceptiva del enlace de voz a través del sistema 10 de audio.

En este modo de realización, el sistema 20 de medición de la calidad comprende una sección 20a de procesamiento preliminar, una sección 20b de procesamiento, y una sección 20c de combinación de señales para procesar las dos señales de entrada *X(t)* e *Y(t)* de modo que se puede proporcionar la señal Q de salida.

La sección 20a de procesamiento preliminar comprende un dispositivo 30 de procesamiento preliminar dispuesto para llevar a cabo una o más acciones de procesamiento preliminar como, por ejemplo, un escalado de nivel fijo y alineación de tiempo con el fin de obtener las señales preprocesadas $X_p(t)$ e $Y_p(t)$. Aunque la FIG. 1 muestra un único dispositivo 30 de procesamiento preliminar también es posible disponer de un dispositivo de procesamiento preliminar independiente para la señal de voz *X(t)* y otro para la señal de voz *Y(t)*.

La sección 20b de procesamiento del sistema 20 de medición de la calidad se dispone para mapear las señales preprocesadas sobre señales de representación de acuerdo con un modelo de percepción psicofísico del sistema de audición humano. La señal preprocesada $X_p(t)$ se procesa en un primer dispositivo 40a de procesamiento con el fin de obtener una señal de representación *R(X),* mientras que la señal preprocesada $Y_p(t)$ se procesa en un segundo dispositivo 40b de procesamiento con el fin de obtener una señal de representación *R(Y)*. El primer dispositivo 40a de procesamiento y el segundo dispositivo 40b de procesamiento se pueden acomodar en un único dispositivo de procesamiento.

La sección 20c de combinación de señales del sistema 20 de medición de la calidad se dispone para combinar las señales de representación *R(X), R(Y)* con el fin de obtener una señal diferencial *D* mediante la utilización de un dispositivo 50 de determinación de diferencias. Por último, un dispositivo 60 de modelado procesa la señal diferencial *D* de acuerdo con un modelo en el que se han modelado ciertas propiedades de los humanos con el fin de obtener la señal Q de calidad. Las propiedades humanas, por ejemplo propiedades cognitivas, se pueden obtener a través de pruebas de escucha subjetivas llevadas a cabo con una serie de sujetos humanos.

El dispositivo 30 de procesamiento preliminar, el primer dispositivo 40a de procesamiento y el segundo dispositivo 40b de procesamiento pueden formar un sistema de procesamiento que se puede utilizar para poner en práctica algunos modos de realización de la invención tal como se explicará con más detalle más adelante. El sistema de procesamiento o sus componentes pueden tomar la forma de un procesador hardware como, por ejemplo, un Circuito Integrado de Aplicación Específica (ASIC) o un dispositivo informático para ejecutar código ejecutable por un ordenador en forma de software o firmware. El dispositivo informático puede comprender, por ejemplo, un procesador y una memoria que se encuentra conectada al procesador a través de un medio de comunicación. Ejemplos de memoria incluyen, pero no se limitan a, Memoria de Solo Lectura (ROM), Memoria de Acceso Aleatorio (RAM), ROM Programable Borrable (EPROM), ROM Programable Borrable Eléctricamente (EEPROM), y una memoria flash (instantánea).

El dispositivo informático puede comprender, además, una interfaz de usuario con el fin de permitir la entrada de instrucciones o notificaciones por parte de usuarios externos. Ejemplos de una interfaz de usuario incluyen, pero no se limitan a, un ratón, un teclado y una pantalla táctil.

El dispositivo informático se pude disponer para cargar código ejecutable por un ordenador almacenado en un medio legible por un ordenador, por ejemplo una Memoria de Solo Lectura de Disco Compacto (CD ROM), un Disco de Vídeo Digital (DVD) o cualquier otro tipo de portador de datos conocido legible por un ordenador. Para este propósito el dispositivo informático puede comprender una unidad de lectura.

El código ejecutable por un ordenador almacenado en el medio legible por un ordenador, después de cargar el código en la memoria del dispositivo informático, se puede adaptar para poner en práctica algunos modos de realización de la invención que se describirán más adelante.

Alternativa o adicionalmente, dichos modos de realización de la invención pueden tomar la forma de un producto de

programa de ordenador que comprenda un código ejecutable por un ordenador con el fin de poner en práctica un método semejante cuando se ejecuta en un dispositivo informático. A continuación, después de cargar el código ejecutable por un ordenador en una memoria del dispositivo informático, un procesador del dispositivo informático puede poner en práctica el método.

De este modo, un método de medida perceptivo objetivo simula en un programa de ordenador la percepción del sonido de sujetos con el objeto de predecir la calidad de los sistemas de audio percibida subjetivamente como, por ejemplo, codificadores de voz, enlaces de telefonía y teléfonos móviles. Las señales físicas de entrada y salida del dispositivo sometido a evaluación se mapean sobre representaciones psicofísicas que se ajustan lo más posible a las representaciones internas dentro de la cabeza del ser humano. La calidad del dispositivo sometido a evaluación se evalúa en función de las diferencias en la representación interna. El mejor método de medida perceptiva objetiva disponible en la actualidad es PESQ (Evaluación Perceptiva de la Calidad de la Voz).

La FIG. 2 muestra de forma esquemática un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio con respecto a una señal de referencia de acuerdo con una PESQ tal como se establece en la Recomendación P.862 de la ITU-T, denominada de aquí en adelante como PESQ. La PESQ se puede utilizar en una configuración como la que se muestra de forma esquemática en la FIG. 1. En la PESQ, una señal de referencia $X(t)$ se compara con una señal de salida $Y(t)$ que es el resultado de pasar $X(t)$ a través de un sistema de audio, por ejemplo un sistema de procesamiento de voz como un sistema de comunicación. La señal de calidad de la salida de la PESQ, también denominada como puntuación PESQ, es una predicción de la calidad percibida que en una prueba de escucha subjetiva sería dada por los sujetos a $Y(t)$. La puntuación PESQ tiene la forma de la denominada puntuación de opinión media (MOS). Para este propósito la salida de la PESQ se mapea sobre una escala de tipo MOS, esto es, un único número en el intervalo de -0,5 a 4,5 aunque para la mayor parte de los casos el intervalo de salida será entre 1,0 y 4,5, lo cual es un intervalo normal de valores MOS encontrado en un experimento de calidad de escucha de Puntuación de Categoría Absoluta (ACR).

El procesamiento previo en la PESQ comprende el ajuste de niveles de ambas señales $X(t)$, $Y(t)$ con el fin de obtener, respectivamente, las señales $X_s(t)$, $Y_s(t)$ así como un filtrado del Sistema de Referencia Intermedio (IRS) con el fin de obtener, respectivamente, las señales $X_{IRSS}(t)$, $Y_{IRSS}(t)$. El ajuste de niveles supone escalar la intensidad a un nivel fijo, en 79 dB SPL de PESQ. El filtrado IRS se lleva a cabo con el fin de asegurar que el método de medición de la calidad de transmisión es relativamente insensible al filtrado de un elemento de los sistemas de telecomunicaciones, por ejemplo, un teléfono móvil o similar. Por último, se determina un retardo de tiempo entre la señal de referencia $X_{IRSS}(t)$, $Y_{IRSS}(t)$ para dar como resultado una señal $Y_{IRSS}'(t)$ de salida desplazada en el tiempo. De este modo se supone que la comparación entre la señal de referencia y la señal de salida tiene lugar con respecto al mismo tiempo.

El oído humano lleva a cabo una transformación de tiempo frecuencia. En la PESQ, esto se modela mediante la realización de una Transformada Rápida de Fourier (FFT) localizada con una ventana de Hanning sobre las señales de tiempo $X_{IRSS}(t)$, $Y_{IRSS}'(t)$. La ventana de Hanning tiene típicamente un tamaño de 32 ms. Las ventanas de tiempo adyacentes, denominadas de aquí en adelante tramas, se solapan típicamente un 50%. Se descarta la información de fase. La suma de las partes reales cuadráticas e imaginarias cuadráticas de los componentes complejos de la FFT, esto es, el espectro de potencia, se utilizan para obtener representaciones de potencia $PX_{WIRSS}(f)_n$ y $PY_{WIRSS}(f)_n$, donde n representa la trama bajo consideración. Las representaciones de potencia se dividen en bandas de frecuencia, denominadas de aquí en adelante bandas FFT.

El sistema auditivo humano tiene una mejor resolución en frecuencia en bajas frecuencias que en altas frecuencias. Una escala de tonos refleja este fenómeno, y por esta razón la PESQ ajusta las frecuencias a una escala de tonos, en este caso a una denominada escala Bark. La conversión del eje de frecuencias (discreto) supone agrupar las bandas FFT para formar bandas Bark, típicamente 24. Las señales resultantes se denominan densidades de potencia de tono o funciones de densidad de potencia de tono y se representan como $PPX_{WIRSS}(f)_n$ y $PPY_{WIRSS}(f)_n$. Las funciones de densidad de potencia de tono proporcionan una representación interna que es análoga a la representación psicofísica de las señales de audio en el sistema de audición humano, teniendo en cuenta la frecuencia perceptiva.

Para gestionar el filtrado en el sistema de audio sometido a evaluación, se promedian en el tiempo el espectro de potencia de la referencia y las densidades de potencia de tono de salida. Se calcula un factor de compensación parcial a partir de la relación del espectro de salida con el espectro de referencia. La densidad de potencia del tono de referencia $PPX_{WIRSS}(f)_n$ de cada trama n se multiplica a continuación con este factor de compensación parcial con el fin de ecualizar la referencia de la señal de salida. Esto da como resultado una densidad de potencia del tono de referencia $PPX'_{WIRSS}(f)_n$ filtrada inversamente. Esta compensación parcial se utiliza debido a que el filtrado suave es difícilmente perceptible mientras que el filtrado intenso puede estar molestando al oyente. La compensación se lleva a cabo sobre la señal de referencia debido a que la señal de salida es la que es evaluada por el sujeto en un experimento de escucha ACR.

Con el fin de compensar las variaciones de ganancia de corta duración se calcula un factor de escala local. El factor

de escala local se multiplica a continuación con la función de densidad de potencia de tono $PPY_{WIRSS}(f)_n$ de salida con el fin de obtener una función de densidad de potencia de tono $PPY'_{WIRSS}(f)_n$ escalada localmente.

Después de una compensación parcial para el filtrado llevada a cabo sobre la señal de referencia y una compensación parcial para las variaciones de ganancia de corta duración llevada a cabo sobre la señal de salida, las densidades de potencia de tono de referencia y degradada se transforman a una escala de intensidad sonora Sone utilizando la ley de Zwicker. Las secuencias de dos dimensiones $LX(f)_n$, y $LY(f)_n$ se denominan funciones de densidad de intensidad sonora para la señal de referencia y la señal de salida, respectivamente. Para $LX(f)_n$ esto quiere decir:

$$LX(f)_n = S_l\left(\frac{P_0(f)}{0.5}\right)^{\gamma}\cdot\left[\left(0.5+0.5\cdot\frac{PPX'_{WIRSS}(f)_n}{P_0(f)}\right)^{\gamma}-1\right] \qquad (1)$$

donde $P_0(f)$ es el umbral absoluto de audición, S_l el factor de escala de intensidad sonora, y γ, la denominada potencia de Zwicker, tiene un valor de aproximadamente 0,23. Las funciones de densidad de intensidad sonora simbolizan la representación interna psicofísica de las señales de audio en el sistema de audición humano teniendo en cuenta la percepción de intensidad sonora.

A continuación se realiza la diferencia entre las funciones de densidad de intensidad sonora de referencia y de salida $LX(f)_n$, $LY(f)_n$ dando como resultado una función diferencia de densidad de intensidad sonora $D(f)_n$. Después de la sustracción perceptiva se puede derivar una medida de la calidad percibida teniendo en cuenta tanto una medida D de la perturbación como una medida D_A de la perturbación asimétrica. En la recomendación P.862 de la ITU-T se pueden encontrar más detalles con respecto a la PESQ.

La FIG. 3 muestra esquemáticamente un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio con respecto a una señal de referencia de acuerdo con un modo de realización de la invención. Después de las acciones de procesamiento previo como el filtrado IRS y el retardo de tiempo, se transforman tanto la señal de referencia como la señal de salida de señales en el dominio del tiempo a señales en el dominio del tiempo frecuencia perceptivo.

Esto se puede llevar a cabo de un modo parecido al que se muestra en la FIG. 2 que hace referencia a la PESQ. Esto es, en primer lugar se ejecuta una función para establecer ventanas, por ejemplo una ventana de Hanning, con el fin de dividir la señal de referencia y la señal de salida en tramas de tiempo que se corresponden entre sí. A continuación se realiza una FFT sobre las tramas de tiempo con el fin de transformar las señales desde el dominio del tiempo al dominio del tiempo frecuencia. Después de la FFT, las señales se ajustan a una escala de tonos, por ejemplo una escala de frecuencias en Bark, con el fin de obtener una representación en el dominio del tiempo frecuencia perceptivo, denominado de aquí en adelante dominio de frecuencia perceptiva.

Al contrario que en la aproximación realizada en la PESQ tal como se muestra de forma esquemática en la FIG. 2, el método que se muestra de forma esquemática en la FIG. 3 tiene en cuenta las variaciones de nivel, en particular las denominadas variaciones de nivel global de emisión. Teniendo en cuenta el nivel global de emisión, se puede aumentar de forma considerable la precisión del indicador de calidad, especialmente en aquellos casos en los que el nivel de emisión no se ajusta al nivel de emisión estandarizado que se utiliza en los cálculos de acuerdo con la recomendación P.862 de la ITU-T. Esto es, la correlación entre el indicador de calidad obtenido de forma objetiva y una calidad obtenida de forma subjetiva mejora para aplicaciones en las que el nivel global de emisión es mayor o menor que el nivel estándar. Dicho nivel global de emisión diferente se utiliza a menudo en los sistemas de Voz sobre IP (VoIP), por ejemplo para prevenir la retroalimentación acústica.

Con el fin de poder tener en cuenta las variaciones de nivel de intensidad, en el procesamiento previo no se realiza ninguna acción de ajuste de nivel sobre la señal de salida. Sin embargo, como se aclarará más abajo, es deseable obtener información con respecto a la señal de referencia que sea independiente del nivel global de emisión. En otras palabras, con el fin de obtener dicha información, el nivel de intensidad total de la señal de referencia debería de ser el mismo para todas las pruebas subjetivas para las que se desee realizar predicciones de calidad.

Por esta razón, la señal de referencia se escala globalmente hacia un nivel de intensidad fijo. El escalado de la señal de referencia se puede llevar a cabo antes de la transformación, esto es, en el dominio del tiempo, tal como se muestra de forma esquemática en la FIG. 3. Alternativamente, la señal de referencia se puede escalar después de la transformación en el dominio del tiempo frecuencia (perceptivo).

Después del escalado de la señal de referencia a un nivel de intensidad fijo, se llevan a cabo las mediciones sobre tramas de tiempo dentro de la función de referencia escalada con el fin de obtener las características de la señal de referencia. En particular las características de la señal con respecto al nivel de intensidad de estas tramas de tiempo, por ejemplo su nivel de intensidad promedio o su nivel de intensidad pico, se determinan en función de las

mediciones realizadas.

Después de las mediciones de nivel de trama, también denominada detección de nivel de trama, la señal de referencia escalada se escala a un nivel de intensidad asociado con la señal de salida. Preferiblemente, este escalado utiliza únicamente bandas de frecuencia que estén dominadas por la señal de voz, por ejemplo las bandas entre 400 y 3500 Hz. Esta acción de escalado se realiza porque debido a un resultado del escalado anterior de la señal de referencia al nivel de intensidad fijo, la diferencia del nivel de intensidad entre la señal de referencia y la señal de salida puede ser tal que sea imposible obtener un indicador de calidad fiable. El escalado de la señal de referencia escalada tiene el objeto de crear una diferencia del nivel de intensidad entre la señal de referencia escalada y la señal de salida que permita una valoración del impacto del nivel global de emisión sobre la calidad percibida. De este modo la acción de escalado realizada compensa parcialmente la diferencia del nivel de intensidad entre la señal de referencia escalada y la señal de salida. Las diferencias de nivel que excedan un cierto valor umbral no se pueden compensar completamente permitiendo modelar el impacto del nivel global bajo de presentación, por ejemplo alguien configura el volumen de su dispositivo de reproducción a un nivel de intensidad bajo. La reproducción de voz a un nivel bajo se utiliza comúnmente en sistemas de VOIP, por ejemplo para tratar las interrupciones en el control acústico del eco.

El escalado puede utilizar un algoritmo de escalado suave, esto es un algoritmo que escale la señal a tratar de modo que se compensen las pequeñas desviaciones de potencia, preferiblemente por trama de tiempo, mientras que se compensan parcialmente las desviaciones más grandes, en función de una relación de potencia entre la señal de referencia y la señal de salida. Se pueden encontrar más detalles con respecto a la utilización del escalado suave en la solicitud de patente de los Estados Unidos 2005/159944, la patente de los Estados Unidos 7.313.517 y la patente de los Estados Unidos 7.315.812, todas ellas asignadas al solicitante e incorporadas a este documento mediante referencia.

Después de la acción de escalado global, la señal de referencia puede someterse a una compensación de frecuencia tal como se ha descrito haciendo referencia a la FIG. 2. De forma parecida, la señal de salida se puede someter a una acción de escalado local. El escalado local también se puede realizar con respecto a la señal de referencia tal como se muestra de forma esquemática en la FIG. 3. A continuación, tanto la señal de referencia como la señal de salida se someten a un ajuste de intensidad a la escala de intensidad de sonido tal como se ha descrito haciendo referencia a la PESQ que se muestra en la FIG. 2. La señal de referencia y la señal de salida se representan ahora en el dominio de intensidad de sonido perceptivo.

En el dominio de intensidad de sonido perceptivo, a diferencia de la PESQ que se muestra en la FIG. 2, tanto la señal de salida con la señal de referencia son objeto de una acción de escalado adicional. Hasta este punto, los niveles de señal de la señal de salida no se han cambiado de forma significativa, y unos niveles muy bajos de la señal de salida ahora únicamente provocarán diferencias sin importancia en la representación interna. Esto da como resultado errores en la estimación de calidad.

Para este propósito, en primer lugar, se escala la señal de salida a un nivel de intensidad de sonido fijo. La intensidad de sonido fijo se puede determinar mediante experimentos de calibración llevados a cabo en experimentos de calidad de escucha subjetivos. Si se utiliza una calibración inicial de nivel global para la señal de referencia tal como se describe en la recomendación P.861 y/o P.862 de la ITU-T, dicho nivel de intensidad de sonido fijo se encuentra alrededor de 20, un valor de escalado asociado a una intensidad de sonido interna adimensional.

Como resultado del escalado del nivel de intensidad de sonido de la señal de salida, la diferencia del nivel de intensidad de sonido entre la señal de salida y la señal de referencia es tal que no se puede determinar un indicador de calidad fiable. Para salvar esta posibilidad no deseada, también es necesario escalar el nivel de intensidad de sonido de la señal de referencia. Por lo tanto, después del escalado del nivel de intensidad de sonido de la señal de salida se escala el nivel de intensidad de sonido de la señal de referencia hacia un nivel de intensidad de sonido asociado con la señal de salida escalada. Ahora tanto la señal de referencia como la señal de salida tienen un nivel de intensidad de sonido que se puede utilizar para calcular las representaciones internas relevantes perceptivamente necesarias para obtener una medida objetiva de la calidad de la transmisión de un sistema de audio.

Dentro de las acciones de escalado global llevadas a cabo en el dominio de la intensidad de sonido perceptivo, se puede utilizar la intensidad de sonido promedio tanto de la señal de referencia como de la de salida. La intensidad de sonido promedio de estas señales se puede determinar sobre tramas de tiempo para las que el nivel de intensidad en la señal de referencia tal como se ha medido durante la detección de nivel de trama excede un valor umbral adicional, por ejemplo el valor del criterio de actividad de voz. El valor del criterio de actividad de voz se puede corresponder con un umbral de escucha absoluto. Si se utiliza el valor del criterio de actividad de voz, estas tramas se pueden denominar como tramas de voz. Para el cálculo, para la señal de salida se tienen en cuenta las tramas de tiempo correspondientes a tramas de tiempo para las que el nivel de intensidad excede el valor umbral adicional. De este modo, en un modo de realización que utiliza el valor del criterio de actividad de voz, la intensidad

de sonido promedio de la señal de referencia se determina con respecto a las tramas de voz, mientras que la intensidad de sonido promedio de la señal de salida se determina con respecto a las tramas de tiempo correspondientes a las tramas de voz dentro de la señal de referencia.

Por último, en la FIG. 3 se hace la diferencia perceptiva entre la señal de referencia y la señal de salida. Esto se puede hacer de una forma conocida a partir de la PESQ y descrita con referencia a la FIG. 2. Esto es, se determinan en paralelo un indicador representativo de la degradación global, D_n, y un indicador representativo de degradaciones añadidas, DA_n.

Tal como se muestra en la FIG. 3 el esquema permite un acercamiento diferente en relación con el cálculo de ambos indicadores D_n, DA_n. Es posible aplicar dos veces el método tal como se muestra en la FIG. 3, esto es, una vez para determinar un indicador de calidad que representa la calidad con respecto a la degradación global, la otra vez para determinar un indicador de calidad que representa la calidad con respecto a degradaciones añadidas en comparación con la señal de referencia. La aplicación del método dos veces permite la optimización de los cálculos con respecto a diferentes tipos de distorsiones Dicha optimización puede mejorar de forma considerable la correlación entre una calidad de voz medida de forma objetiva y una calidad de voz como la obtenida en los experimentos de calidad de escucha subjetivos.

En un modo de realización en el que el método se aplica dos veces, se pueden utilizar de forma diferente los resultados de la detección del nivel de las tramas. Por ejemplo, puede ser diferente la selección de las tramas de tiempo, por ejemplo basándose en diferentes valores umbral de actividad de la voz.

La FIG. 4 muestra de forma esquemática un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio con respecto a una señal de referencia de acuerdo con un modo de realización adicional de la invención. En este método, tanto la señal de referencia como la señal de salida se someten a pasos de procesamiento previo, por ejemplo filtrado IRS y retardo de tiempo como se conoce a partir de la PESQ y se describe haciendo referencia a la FIG. 2. Antes de obtener una representación de tiempo frecuencia de las señales mediante la realización de una transformada rápida de Fourier localizada en combinación con la utilización de una función para establecer ventanas, por ejemplo, una ventana de Hanning tal como se conoce a partir de la PESQ, la señal de referencia se escala de forma global a un nivel fijo. El escalado global a un nivel fijo es parecido al ajuste de niveles utilizado en la PESQ. Sin embargo, en este caso únicamente se escala de este modo la señal de referencia. En esta etapa no se escala la señal de salida. El nivel fijo coincide preferiblemente con un nivel de aproximadamente 73 dB SPL para un fragmento de voz presentado de forma diótica o dicótica y con un nivel de aproximadamente 79 dB SPL para un fragmento de voz presentado de forma monótica. La señal de salida se escala con un factor de modo que la representación interna se corresponde con el nivel acústico real utilizado en la prueba subjetiva.

Después de haber obtenido la representación de potencia frecuencia debida a la FFT llevada a cabo sobre una ventana de tiempo seleccionada mediante una función para establecer ventanas, por ejemplo una ventana de Hanning, la señal de referencia se escala hacia la señal de salida sobre un nivel global con un algoritmo que únicamente compensa de forma parcial la diferencia de nivel de intensidad entre la señal de referencia y la señal de salida. La diferencia resultante se puede utilizar para estimar el impacto del nivel de intensidad sobre la calidad de la transmisión percibida.

En un modo de realización, el escalado de la intensidad de la señal de referencia desde el nivel de intensidad fijo al nivel de intensidad asociado con la señal de salida puede estar basado en la multiplicación de la señal de referencia con un factor de escala. Dicho factor de escala se puede obtener a partir de la determinación del nivel de intensidad promedio de la señal para al menos una parte de las señales de referencia y de salida. A continuación se puede utilizar el nivel de intensidad promedio de la señal de referencia y el nivel de intensidad promedio de la señal de salida en un cálculo parcial para obtener un factor de escala preliminar. Por último, el factor de escala se puede determinar definiendo el factor de escala para que sea igual al factor de escala preliminar si el factor de escala preliminar es más pequeño que un valor umbral y, en caso contrario, para que sea igual al factor de escala preliminar incrementado con un valor dependiente del factor de escala preliminar adicional.

Después del escalado global al nivel de intensidad de la señal de salida, la señal de referencia es objeto de un escalado local en el dominio del tiempo frecuencia perceptivo y una compensación parcial de frecuencia utilizando el mismo método que el descrito haciendo referencia a la PESQ en la FIG. 2. Aunque en el modo de realización que se muestra en la FIG. 4 el escalado local se lleva a cabo con respecto a la señal de referencia, es igualmente posible aplicar este paso de escalado local con respecto a la señal de salida, por ejemplo de una forma tal como la que se muestra en la FIG. 2. El objeto de la acción de escalado local está relacionado con la compensación de variaciones de ganancia de corto plazo. La selección de la señal de referencia o de la señal de salida puede depender de la aplicación específica. En general se compensa la señal de referencia, debido a que la señal de referencia en general no se presenta a un sujeto de prueba en las mediciones de calidad subjetivas.

En un modo de realización, la primera compensación parcial en frecuencia puede utilizar el denominado algoritmo de

escalado suave. En el algoritmo de escalado suave, la señal a tratar, esto es, la señal de referencia o la señal de salida, se mejora escalándola de modo que se compensan pequeñas desviaciones de potencia, preferiblemente por trama de tiempo, mientras que desviaciones mayores se compensan parcialmente, dependiendo de una relación de potencia entre la señal de referencia y la señal de salida. Se pueden encontrar más detalles con respecto a la utilización del escalado suave en la solicitud de patente de los Estados Unidos 2005/159944, la patente de los Estados Unidos 7.313.517 y la patente de los Estados Unidos 7.315.812, todas ellas asignadas al solicitante e incorporadas a este documento mediante referencia.

Preferiblemente, a continuación se lleva a cabo un paso de excitación tanto en la señal de referencia como en la señal de salida con el fin de compensar el ruido de los componentes de frecuencia como resultado de la ejecución anterior de la transformada rápida de Fourier con una función para establecer ventanas, por ejemplo una ventana de Hanning, con respecto a estas señales. El paso de excitación se lleva a cabo utilizando una curva de anulación propia para acentuar la representación de ambas señales. Por ejemplo, en el artículo “A perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation (Una Medida de Calidad de Audio perceptual Basada en una Representación de Sonido Psicoacústico)”, por J.G. Beerends y J.A. Stemerdink, Revista de la Sociedad de Ingeniería de Audio, Vol. 40, Núm. 12 (1992) pp. 963-978 se pueden encontrar más detalles con respecto al cálculo de dicha curva de anulación propia. En este artículo, se calcula la excitación y se determina la calidad utilizando representaciones de excitación borrosas. En un modo de realización, la excitación calculada se utiliza a continuación para derivar una curva de anulación propia que a su vez se puede utilizar para obtener una representación en tiempo frecuencia más acentuada. En su forma más simple, la curva de anulación propia se corresponde con una fracción de la curva de excitación.

Después de un ajuste de la intensidad a una escala de intensidad de sonido tal como se utiliza en la PESQ, y se describe haciendo referencia a la FIG. 2, la señal de referencia y la señal de salida se escalan localmente en el dominio de la intensidad de sonido. En primer lugar, se escalan aquellas partes de la señal de referencia que sean más altas que la señal de salida. A continuación se escalan las porciones de la señal de salida que sean más altas que la señal de referencia.

La separación de estas acciones de escalado local permite una implementación y/o manipulación independientes de las variaciones de nivel debido a recortes y pulsos de tiempo. Si una porción de la señal de referencia es más alta que una porción correspondiente de la señal de salida, esta diferencia se puede deber a un recorte de tiempo, por ejemplo provocado por una trama perdida. Con el fin de medir el impacto perceptivo del recorte de tiempo, la señal de referencia se escala hacia abajo a un nivel que se considere óptimo para el cálculo de diferencias de perturbación (asimétrica). Esta acción de escalado local sobre la señal de salida también elimina el ruido en la señal de salida hasta un nivel que sea más óptimo para el cálculo de diferencias de perturbación (asimétrica). El impacto del ruido sobre la calidad percibida de forma subjetiva se puede estimar con más exactitud combinando este escalado local con una acción de eliminación de ruido sobre la señal de salida.

A continuación, se puede realizar una segunda compensación parcial de frecuencia. Esta compensación de frecuencia se puede llevar a cabo de forma parecida a como se realiza en la PESQ, pero realizándola ahora en el dominio de la intensidad de sonido. En un modo de realización, la segunda compensación parcial de frecuencia utiliza un algoritmo de escalado suave tal como se ha descrito anteriormente haciendo referencia a la primera compensación parcial de frecuencia. Se ha encontrado que la utilización de una segunda compensación parcial de frecuencia mejora, además, la correlación entre una calidad de voz medida de forma objetiva y la calidad de voz que se obtiene en los experimentos de calidad de escucha subjetiva.

Tal como se ha descrito anteriormente, la primera compensación parcial de frecuencia y la segunda compensación parcial de frecuencia pueden ser parecidas a la compensación parcial de frecuencia utilizada en la PESQ, tal como se ha discutido haciendo referencia a la FIG. 2. De este modo, estas acciones de compensación de frecuencia pueden utilizar una operación promedio que comprende una estimación basada en la respuesta lineal a la frecuencia del sistema sometido a evaluación. En algunos modos de realización, la estimación únicamente se lleva a cabo sobre tramas para las que el valor de nivel de intensidad de la señal de referencia es mayor que un valor umbral, por ejemplo el valor del criterio de actividad de voz. Como será fácilmente entendible a partir del esquema de la FIG. 4, dicha selección de tramas de voz puede estar basada en los niveles detectados en la acción de detección del nivel de la trama.

En este momento, las bandas altas tanto de la señal de referencia como de la señal de salida se establecen preferiblemente a cero debido a que tienen una influencia mínima en la calidad de la transmisión percibida a determinar. Además, los niveles de intensidad de las bandas bajas de la señal de salida se escalan localmente hacia los niveles de intensidad de bandas parecidas de la señal de referencia. Por ejemplo, todas las bandas asociadas a Bark 23 y mayores se pueden poner a cero, mientras que se pueden escalar las bandas de Bark de la señal de salida asociadas a Bark 0 a 5. Por lo tanto, las bandas de Bark asociadas a Bark 0 – 22 en la señal de referencia y las bandas de Bark asociadas a Bark 6 a 22 en la señal de salida no son objeto de ninguna de estas operaciones.

Hasta este momento, los niveles de señal de la señal de salida no han cambiado de forma significativa, y unos

niveles muy bajos de la señal de salida provocarán ahora únicamente diferencias marginales en la representación interna. Esto provoca errores en la estimación de calidad. Por lo tanto, tanto la señal de referencia como la señal de salida se escalan globalmente hacia un nivel que pueda ser utilizado para calcular las representaciones internas relevantes para la percepción necesarias para obtener una medida objetiva de la calidad de la transmisión de un sistema de audio. En primer lugar, el nivel global de la señal de salida se escala hacia un nivel de intensidad de sonido interno fijo. Si se utiliza una calibración inicial del nivel global para la señal de referencia tal como se describe en las recomendaciones P.861 y/o P.862 de la ITU-T, dicho nivel interno global fijo se encuentra en torno a 20, un número de escalado adimensional asociado a la intensidad de sonido interna. En segundo lugar, los niveles de la señal de referencia se escalan a los niveles correspondientes a la señal de salida de modo similar y por las mismas razones que las descritas haciendo referencia a la FIG. 3.

Por último, de forma parecida al método descrito haciendo referencia a la FIG. 2, se obtiene la diferencia entre la señal de referencia y la señal de salida resultando una señal diferencia. Después de la sustracción perceptiva, se puede derivar una medición de la calidad percibida, por ejemplo de un modo como el que se muestra en la FIG. 2 y se describe en la recomendación P.862 de la ITU-T.

De forma alternativa, el método se aplica dos veces. Una vez para determinar un indicador de calidad representativo de la calidad con respecto a la degradación global en comparación con la señal de referencia, y la otra vez para determinar un indicador de calidad representativo de la calidad con respecto a las degradaciones añadidas en comparación con la señal de referencia.

En algunos modos de realización de la invención, el método incluye, además, uno o más pasos de eliminación de ruidos. El impacto del ruido en la calidad de transmisión de un sistema de audio, en particular la calidad de la voz, es dependiente de los cambios en el nivel local y/o el espectro local. En la PESQ no se tiene en cuenta este efecto de forma correcta. La PESQ únicamente utiliza el nivel de potencia local por trama para eliminar el ruido a un nivel que cuantifica de forma aproximada el impacto del ruido. El uno o más pasos de eliminación de ruido pueden proporcionar una mejora significativa para predecir la calidad de transmisión de un sistema de audio.

En un modo de realización, dicha eliminación de ruido se lleva a cabo sobre la señal de referencia después del ajuste de la intensidad a la escala Sone de intensidad de sonido. Esta acción de eliminación de ruido se puede realizar para eliminar el ruido hasta un nivel de ruido determinado previamente. El nivel de ruido determinado previamente se puede entonces corresponder con un nivel de ruido que se considere que sea un nivel de ruido bajo deseable para servir como una representación ideal para la señal de salida.

De forma parecida, en un modo de realización, dicha eliminación de ruido se lleva a cabo sobre la señal de salida después del ajuste de la intensidad a la escala Sone de intensidad de sonido. En este caso, la acción de eliminación de ruido se puede configurar para eliminar el ruido hasta un nivel de ruido representativo de la perturbación experimentada por el dispositivo sometido a evaluación, por ejemplo el sistema 10 de audio de la FIG. 1.

En algunos otros modos de realización, la señal de referencia y la señal de salida son sometidas además a una acción adicional de eliminación de ruido después del escalado global tal como se muestra de forma esquemática en la FIG. 3 mediante una línea discontinua. Se ha encontrado que dicha eliminación de ruido adicional después de un escalado global mejora más aún la correlación entre una calidad de voz medida de forma objetiva y la calidad de voz tal como se obtiene en los experimentos de calidad de escucha subjetivos.

En algunos modos de realización que utilizan uno o más pasos de eliminación de ruido, los parámetros del nivel de intensidad determinados de las tramas de tiempo dentro de la señal de referencia escalada se utilizan para seleccionar tramas de tiempo dentro de la señal de salida para ser incluidas en el uno o más cálculos de eliminación de ruido. Por ejemplo, las tramas de tiempo dentro de la señal de referencia escalada se pueden seleccionar para el cálculo en función de que su valor de intensidad esté por debajo de cierto valor umbral, por ejemplo el valor de criterio de silencio. Una trama de tiempo dentro de la señal de referencia escalada para la que el valor de intensidad se encuentre por debajo del valor del criterio de silencio se puede denominar trama de silencio. Las tramas de tiempo seleccionadas dentro de la señal de salida se corresponden entonces con tramas de silencio dentro de la señal de referencia escalada. Preferiblemente, dicho proceso de selección progresa identificando una serie de tramas de silencio consecutivas, por ejemplo, 8 tramas de silencio. Dicha serie de tramas de silencio consecutivas se puede denominar intervalo de silencio. El nivel de intensidad medido dentro de las tramas de silencio, y en particular las tramas de silencio dentro de un intervalo de silencio, expresa un nivel de ruido que se encuentra inherentemente presente en la señal de referencia bajo consideración. En otras palabras, no existe influencia del dispositivo sometido a evaluación.

La invención se ha descrito haciendo referencia a ciertos modos de realización descritos más arriba. Se reconocerá que estos modos de realización son susceptibles de varias modificaciones y formas alternativas bien conocidas por aquellos experimentados en la técnica.

REIVINDICACIONES

1. Un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un sistema de audio con respecto a una señal de referencia, donde la señal de referencia y la señal de salida se procesan y comparan, y el procesamiento incluye dividir la señal de referencia y la señal de salida en tramas de tiempo mutuamente correspondientes, en donde el procesamiento comprende, además:

- escalar la intensidad de la señal de referencia hacia un nivel de intensidad fijo;
- realizar mediciones sobre las tramas de tiempo dentro de la señal de referencia escalada para determinar las características de las tramas de tiempo de la señal de referencia; caracterizadas por que el método comprende además los pasos de:
- escalar la intensidad de la señal de referencia desde el nivel de intensidad fijo hacia el nivel de intensidad asociado a la señal de salida;
- escalar la intensidad de sonido de la señal de salida hacia un nivel de intensidad de sonido fijo en el dominio de intensidad de sonido perceptivo, utilizando el escalado de intensidad de sonido de la señal de salida las características de las tramas de tiempo de la señal de referencia; y
- escalar la intensidad de sonido de la señal de referencia desde un nivel de intensidad de sonido correspondiente a un nivel de intensidad asociado a la señal de salida hacia un nivel de intensidad de sonido asociado al nivel de intensidad de sonido de la señal de salida escalada en el dominio de intensidad de sonido perceptivo, utilizando la intensidad de sonido de la señal de referencia las características de las tramas de tiempo de la señal de referencia;
- realizar la sustracción perceptiva de la señal de referencia y la señal de salida para obtener una señal diferencia;
- y derivar el indicador de calidad a partir de la señal diferencia.

2. El método de la reivindicación 1, en donde el escalado de la intensidad de la señal de referencia desde el nivel de intensidad fijo hacia un nivel de intensidad asociado a la señal de salida se basa en la multiplicación de la señal de referencia por un factor de escala, estando definido el factor de escala por:

- la determinación de un nivel de intensidad promedio de la señal de referencia para un número de tramas de tiempo;
- la determinación de un nivel de intensidad promedio de la señal de salida para un número de tramas de tiempo correspondientes a las tramas de tiempo de la señal de referencia utilizadas para determinar el nivel de intensidad promedio de la señal de referencia;
- la derivación de un factor de escala preliminar mediante la determinación de una fracción basada en el nivel de intensidad promedio de la señal de referencia y el nivel de intensidad promedio de la señal de salida;
- la determinación de un factor de escala definiendo que el factor de escala sea igual al factor de escala preliminar si el factor de escala preliminar es más pequeño que un valor umbral, y en caso contrario que sea igual al factor de escala preliminar incrementado con un valor adicional dependiente del factor de escala preliminar.

3. El método de una cualquiera de las reivindicaciones precedentes, en donde el método, antes del escalado de la intensidad de sonido del nivel de salida hasta un nivel de intensidad de sonido fijo comprende, además:

- escalar localmente el nivel de intensidad de sonido de la señal de referencia hacia el nivel de intensidad de sonido de la señal de salida para las partes de la señal de referencia con un nivel de intensidad de sonido que sea más alto que el nivel de intensidad de sonido de la señal de salida; y
- escalar a continuación de forma local el nivel de intensidad de sonido de la señal de salida hacia el nivel de intensidad de sonido de la señal de referencia para las partes de la señal de salida con un nivel de intensidad de sonido que sea más alto que el nivel de intensidad de sonido de la señal de referencia.

4. El método de una cualquiera de las reivindicaciones precedentes, en el que la señal de referencia en el dominio de intensidad de sonido perceptivo, antes de ser escalada hacia un nivel de intensidad de sonido asociado al nivel de intensidad de sonido de la señal de salida en el dominio de intensidad de sonido perceptivo, es objeto de una acción de eliminación de ruido para eliminar el ruido hasta un nivel de ruido determinado previamente.

5. El método de una cualquiera de las reivindicaciones precedentes, en el que la señal de salida en el dominio

de intensidad de sonido perceptivo, antes de ser escalada hacia un nivel de intensidad de sonido fijo, es objeto de un algoritmo de eliminación de ruido para eliminar el ruido hasta un nivel de ruido representativo de la perturbación.

6. El método de una cualquiera de las reivindicaciones precedentes, en el que la señal de referencia y la señal de salida en el dominio de intensidad de sonido perceptivo, antes de la comparación, son objeto de una eliminación de ruido global.

7. El método de la reivindicación 1 en el que el sistema de audio es un dispositivo de procesamiento de voz.

8. Un producto de programa de ordenador que comprende código ejecutable por un ordenador, por ejemplo almacenado en un medio legible por un ordenador, adaptado para llevar a cabo, cuando es ejecutado por un procesador, el método tal como se ha descrito por una cualquiera de las reivindicaciones 1-7.

9. Un sistema (20) para determinar un indicador de calidad que representa una calidad percibida de una señal de salida Y(t) de un sistema (10) de audio, por ejemplo un dispositivo de procesamiento de voz, con respecto a una señal de entrada X(t) del sistema de audio que sirve como una señal de referencia, comprendiendo el sistema:

- un dispositivo (30) de procesamiento previo para procesar previamente la señal de referencia y la señal de salida;
- un primer dispositivo (40a) de procesamiento para procesar la señal de referencia, y un segundo dispositivo (40b) de procesamiento para procesar la señal de salida con el fin de obtener señales de representación R(X), R(Y) para la señal de referencia y la señal de salida, respectivamente;
- un dispositivo (50) de diferenciación para combinar las señales de representación de la señal de referencia y la señal de salida con el fin de obtener una señal diferencial D; y
- un dispositivo (60) de modelado para procesar la señal diferencial con el fin de obtener una señal Q de calidad que representa una estimación de la calidad perceptiva del sistema de procesamiento de la voz;

en donde el dispositivo de procesamiento previo, el primer dispositivo de procesamiento y el segundo dispositivo de procesamiento forman un sistema de procesamiento para llevar a cabo el método de una cualquiera de las reivindicaciones 1-7.

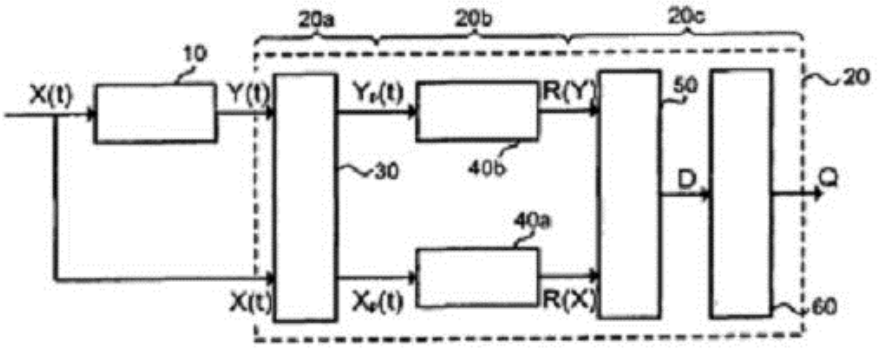

FIG. 1

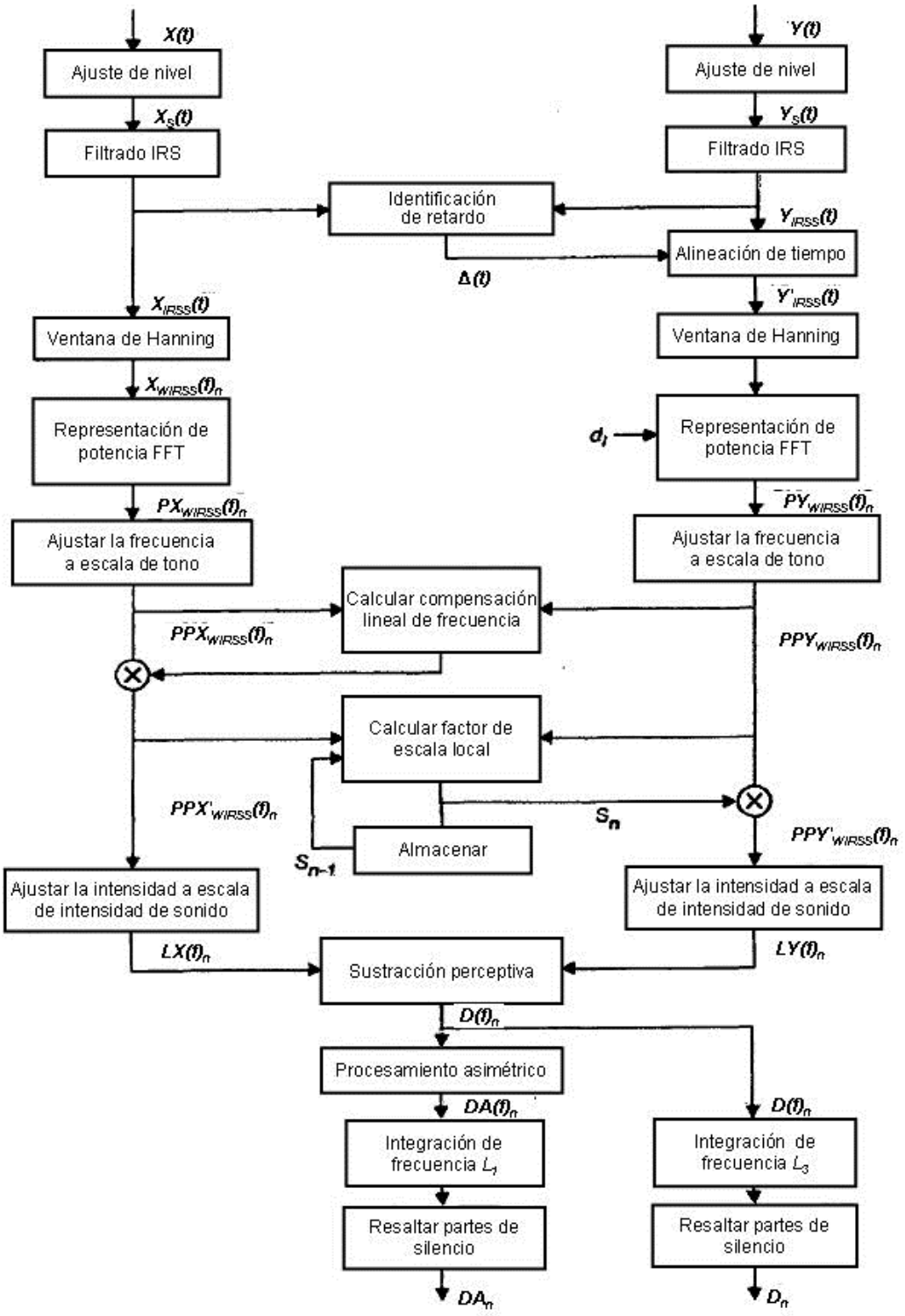
X(t)
Ajuste de nivel
$X_S(t)$
Filtrado IRS
Identificación de retardo
Δ(t)
$X_{IRSS}(t)$
Ventana de Hanning
$X_{WIRSS}(t)_n$
Representación de potencia FFT
$PX_{WIRSS}(t)_n$
Ajustar la frecuencia a escala de tono
$PPX_{WIRSS}(t)_n$
Calcular factor de escala local
$PPX'_{WIRSS}(t)_n$
S_{n-1}
Almacenar
Ajustar la intensidad a escala de intensidad de sonido
$LX(t)_n$
Y(t)
Ajuste de nivel
$Y_S(t)$
Filtrado IRS
$Y_{IRSS}(t)$
Alineación de tiempo
$Y'_{IRSS}(t)$
Ventana de Hanning
d_l
Representación de potencia FFT
$PY_{WIRSS}(t)_n$
Ajustar la frecuencia a escala de tono
Calcular compensación lineal de frecuencia
$PPY_{WIRSS}(t)_n$
S_n
$PPY'_{WIRSS}(t)_n$
Ajustar la intensidad a escala de intensidad de sonido
$LY(t)_n$
Sustracción perceptiva
$D(t)_n$
Procesamiento asimétrico
$DA(t)_n$
Integración de frecuencia L_1
Resaltar partes de silencio
DA_n
$D(t)_n$
Integración de frecuencia L_3
Resaltar partes de silencio
D_n

FIG. 2

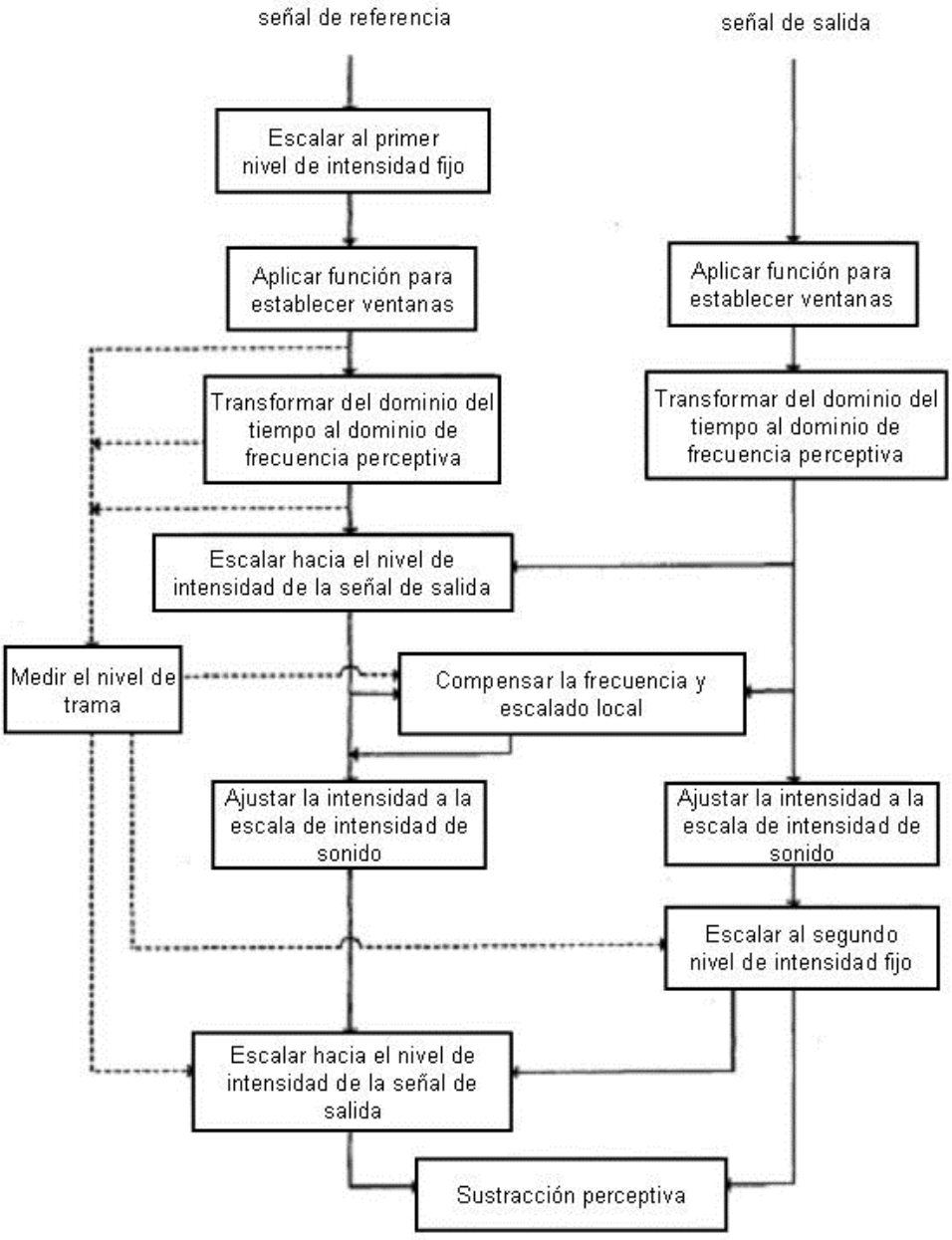
señal de referencia
señal de salida
Escalar al primer nivel de intensidad fijo
Aplicar función para establecer ventanas
Aplicar función para establecer ventanas
Transformar del dominio del tiempo al dominio de frecuencia perceptiva
Transformar del dominio del tiempo al dominio de frecuencia perceptiva
Escalar hacia el nivel de intensidad de la señal de salida
Medir el nivel de trama
Compensar la frecuencia y escalado local
Ajustar la intensidad a la escala de intensidad de sonido
Ajustar la intensidad a la escala de intensidad de sonido
Escalar al segundo nivel de intensidad fijo
Escalar hacia el nivel de intensidad de la señal de salida
Sustracción perceptiva

FIG. 3

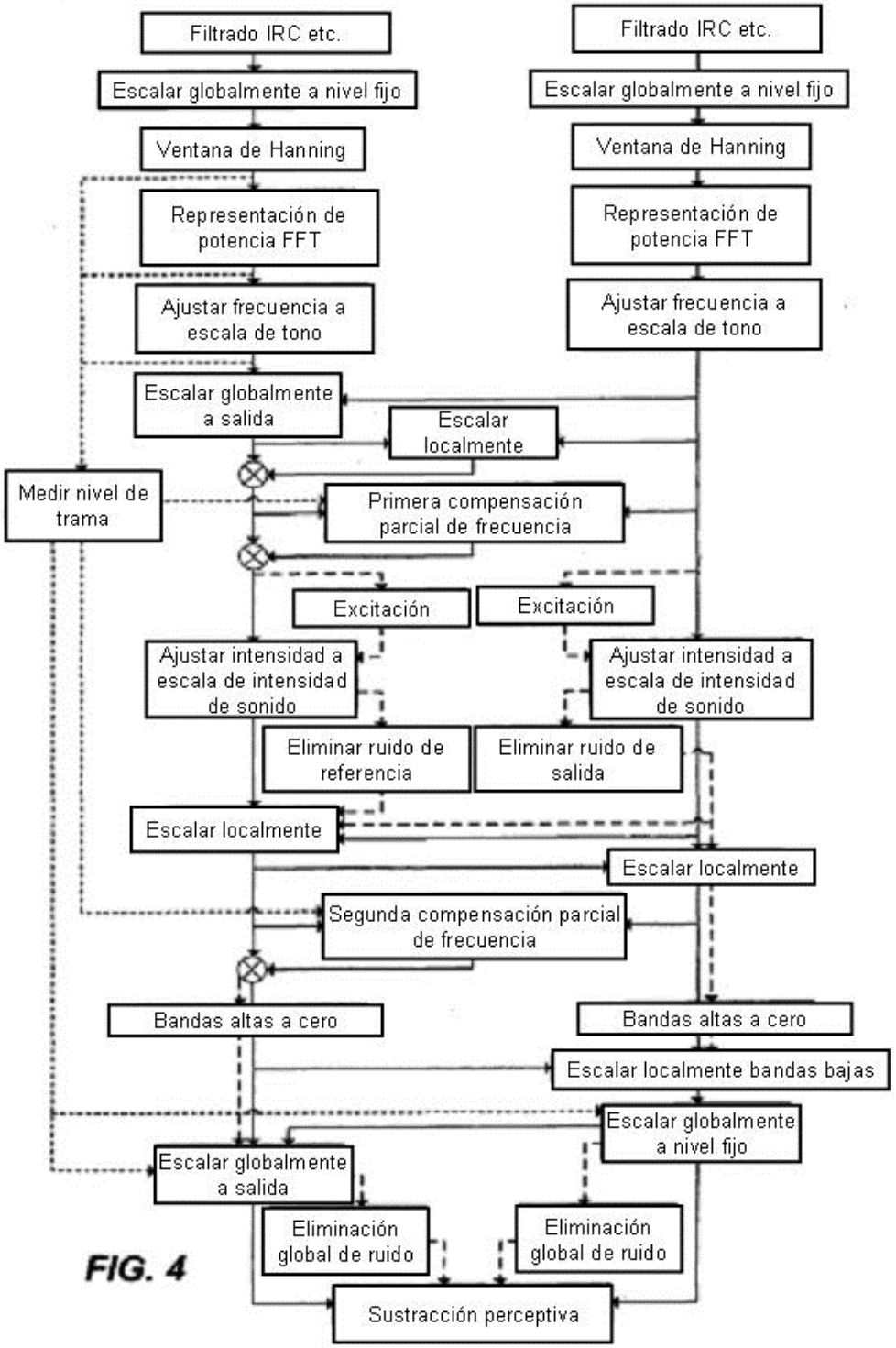
Filtrado IRC etc.
Escalar globalmente a nivel fijo
Ventana de Hanning
Representación de potencia FFT
Ajustar frecuencia a escala de tono
Escalar globalmente a salida
Escalar localmente
Medir nivel de trama
Primera compensación parcial de frecuencia
Excitación
Excitación
Ajustar intensidad a escala de intensidad de sonido
Ajustar intensidad a escala de intensidad de sonido
Eliminar ruido de referencia
Eliminar ruido de salida
Escalar localmente
Escalar localmente
Segunda compensación parcial de frecuencia
Bandas altas a cero
Bandas altas a cero
Escalar localmente bandas bajas
Escalar globalmente a nivel fijo
Escalar globalmente a salida
Eliminación global de ruido
Eliminación global de ruido
Sustracción perceptiva

FIG. 4