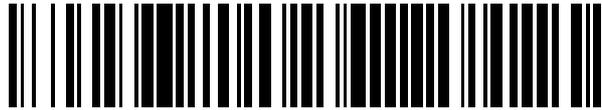


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 556 587**

51 Int. Cl.:

G10L 25/69 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **15.11.2012 E 12791582 (5)**

97 Fecha y número de publicación de la concesión europea: **16.09.2015 EP 2780910**

54 Título: **Método y aparato para evaluar la inteligibilidad de una señal de voz degradada**

30 Prioridad:

17.11.2011 EP 11189598

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

19.01.2016

73 Titular/es:

**NEDERLANDSE ORGANISATIE VOOR
TOEGEPAST- NATUURWETENSCHAPPELIJK
ONDERZOEK TNO (100.0%)
Anna van Buerenplein 1
2595 DA 's-Gravenhage, NL**

72 Inventor/es:

BEERENDS, JOHN GERARD

74 Agente/Representante:

LAZCANO GAINZA, Jesús

ES 2 556 587 T3

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

DESCRIPCIÓN

Método y aparato para evaluar la inteligibilidad de una señal de voz degradada

5 Campo de la Invención

10 La presente invención se relaciona con un método para evaluar la inteligibilidad de una señal de voz degradada recibida de un sistema de transmisión de audio, al transportar a través de dicho sistema de transmisión de audio una señal de voz de referencia tal como suministrar dicha señal de voz degradada, en donde el método comprende: muestrear dicha
 15 señal de voz de referencia en una pluralidad de marcos de señal de referencia y determinar para cada marco una representación de señal de referencia; muestrear dicha señal de voz degradada en una pluralidad de marcos de señal degradados y determinar para cada marco una representación de señal degradada; formar pares de marco al asociar cada marco de señal de referencia con un marco de señal degradado correspondiente, y suministrar para cada par de marcos una función de diferencia que representa una diferencia entre dicho marco de señal degradado y dicho marco de señal de referencia asociado.

La presente invención se relaciona además con un aparato para efectuar un método como se describió anteriormente, y con un producto de programa de ordenador.

20 Antecedentes

25 Durante las pasadas décadas los métodos de medición de calidad objetiva de voz se han desarrollado y desplegado utilizando una aproximación de medición perceptiva. En esta aproximación un algoritmo basado en la percepción simula el comportamiento de un sujeto que califica la calidad de un fragmento de audio en una prueba de escucha. Para la calidad de voz se utiliza principalmente la así llamada prueba de escucha absoluta con calificación de categoría, donde los sujetos juzgan la calidad del fragmento de voz degradado sin tener acceso a un fragmento de voz limpio de referencia. Las pruebas de escucha llevadas a cabo en la Unión de Telecomunicaciones Internacional (ITU) principalmente utilizan una calificación de categoría absoluta (ACR) de escala de opinión de 5 puntos, que es
 30 consecuentemente también utilizada en los métodos de medición de calidad de voz objetiva que fueron estandarizados por el ITU, Medición de Calidad Perceptiva de la Voz. (PSQM (ITU – T Rec. P. 861, 1996)), y su Evaluación Perceptiva de Seguimiento de la Calidad de Voz (PESQ (ITU – T Rec. P. 862, 2000)). El enfoque de estos estándares de medición está en la calidad de voz de banda estrecha (ancho de banda de audio 100-3500 Hz), aunque la extensión del ancho de banda (50 – 7000 Hz) fue ideado en el 2005. El PESQ suministra unas muy buenas correlaciones con las pruebas de escucha subjetivas en los datos de voz de banda estrecha y correlaciones aceptables para datos de banda ancha.

35 Unos nuevos servicios de voz de ancho de banda están siendo desarrollados por la industria de telecomunicaciones, la necesidad surgió de un estándar de medición avanzado de desempeño verificado, y capaz de mayores anchos de banda de audio. Por lo tanto el Grupo 12 de Estudio ITU – T (ITU –sector Telecom) inicio la estandarización de un nuevo algoritmo de evaluación de calidad de voz como una actualización de tecnología del PESQ. El nuevo estándar de medición, de tercera generación, POLQA (Evaluación de Calidad de Escucha Perceptivo Objetivo), soluciona los inconvenientes del estándar PESQ P. 862 tal como la evaluación incorrecta del impacto de las distorsiones de respuesta de frecuencia lineal, la compresión estiramiento del tiempo como se encuentra en Voz sobre IP, cierto tipo de distorsiones de códec y reverberaciones.

45 La solicitud de patente europea EP 2048657A1 describe un método y sistema para la medición de la inteligibilidad de la voz de un sistema de transmisión de audio adaptado para obtener una función de densidad de perturbación de una entrada de referencia y una entrada degradada. La función de densidad de perturbación se multiplica por una función de corrección derivada de un cálculo de correlación de las densidades de potencia de altura tonal compensadas asociadas con la señal de entrada y un marco previo independiente. La función de la densidad de perturbación corregida se agrega sobre la frecuencia y el tiempo para obtener una medición de la inteligibilidad de la voz.

50 Aunque el POLQA (p. 863) suministra un número de mejoras sobre los algoritmos de evaluación de calidad anterior PSQM (P. 861) y PESQ (P. 862), las presentes versiones del POLQA, como el PSQM y el PESQ, no manejan una condición de calidad perceptiva subjetiva elemental, a saber la inteligibilidad. A pesar también de ser dependiente de un número de parámetros de calidad de audio, la inteligibilidad está más cercanamente relacionada con la calidad de transferencia de información que con la calidad del sonido. En términos de algoritmos de evaluación de calidad, la naturaleza de la inteligibilidad opuesta a la calidad de sonido hace que los algoritmos produzcan una calificación de evaluación que desajusta la calificación que se habría asignado si la señal de voz hubiera sido evaluada por una persona o una audiencia. Sin perder de vista el objetivo de compartir información, un ser humano valorará una señal de voz inteligible por encima de una señal que sea menos inteligible, pero que sea similar en términos de calidad de
 60 sonido. Los algoritmos actualmente conocidos no han sido capaces de abordar correctamente esto al grado requerido.

Resumen de la Invención

Es un objeto de la presente invención buscar una solución para la desventaja anteriormente mencionada de la técnica anterior, y suministrar un algoritmo de evaluación de calidad para la evaluación de las señales de voz (degradada) que se adaptan para tener en cuenta la inteligibilidad de la señal de voz para la evaluación de la misma.

5 La presente invención logra esto y otros objetos porque se suministra un método para evaluar la inteligibilidad de la señal de voz degradada recibida de un sistema de transmisión de audio, al transportar a través de un sistema de transmisión de audio una señal de voz de referencia tal como suministrar dicha señal de voz degradada, en donde el método comprende: muestrear dicha señal de voz de referencia en una pluralidad de marcos de señal de referencia y determinar para cada marco una representación de señal de referencia, muestrear dicha señal de voz degradada en una pluralidad de marcos de señal degradado y determinar para cada marco una representación de señal degradada; formar pares de marco al asociar cada marco de señal de referencia con un marco de señal degradado correspondiente, y suministrar para cada par de marco una función de diferencia que representa una diferencia entre dicho marco de señal degradado y dicho marco de señal de referencia asociado; compensar dicha función de diferencia para uno o más tipos de perturbación tal como suministrar para cada par de marcos una función de densidad de perturbación que se adapta a un modelo de percepción auditiva humana; derivar de dichas funciones de densidad de perturbación de una pluralidad de pares marco un parámetro de calidad total, dicho parámetro de calidad es al menos indicativo de dicha inteligibilidad de dicha señal de voz degradada; en donde, dicho método comprende además las etapas de: determinar el valor de sonoridad para cada uno de dichos marcos de señal de referencia ; y determinar un valor de ponderación dependiente sobre dicho valor de sonoridad y dicho marco de señal de referencia; en donde dicha etapa de compensar dicha función de diferencia comprende una etapa de ponderar dicha función de diferencia utilizando dicho valor de ponderación dependiente de la sonoridad, para incorporar un impacto de perturbación sobre dicha inteligibilidad de dicha señal de voz degradada en dicha evaluación.

La presente invención maneja la inteligibilidad al reconocer que el ruido y otras perturbaciones son más destructivas para la comunicación cuando la información está particularmente siendo transmitida. En las comunicaciones de voz, esto es durante el tiempo cuando la señal de voz real lleva palabras habladas. Más aún, la invención de manera correcta tiene en cuenta la modulación y la naturaleza variable del lenguaje hablado, y suministra una manera de incorporar la naturaleza destructiva de las perturbaciones y su dependencia de esta modulación y de la naturaleza variable del lenguaje hablado. Al incluir un valor de ponderación dependiente del valor de sonoridad de la señal de referencia, el método de la presente invención permite ponderar la cantidad de perturbación dependiente de si o no la información está siendo realmente transportada en la señal de voz degradada.

De acuerdo con una realización de la invención, para determinar el valor de ponderación dependiente de la sonoridad, el método comprende una etapa de comparar dicho valor de sonoridad con un umbral, y hacer dicho valor de ponderación dependiente de si el valor de sonoridad excede dicho umbral. Como se apreciará, comparando el valor de sonoridad con un umbral se permite utilizar una aproximación diferente para la evaluación del ruido y las perturbaciones durante las pausas de la voz y durante las palabras habladas. El impacto de las perturbaciones será diferente durante las palabras habladas que durante los periodos silentes, y se puede tratar de manera diferente cuando se hace uso de un umbral.

De acuerdo con una realización adicional, el valor de ponderación se fija a un valor máximo cuando dicho valor de sonoridad para dicho marco de señal de referencia excede dicho umbral. Por ejemplo, por encima del umbral, el método de la presente invención puede simplemente aplicar un valor de ponderación de 1.0 para incluir completamente todas las perturbaciones durante las palabras habladas.

De acuerdo con una realización adicional, el valor de ponderación es una función que es dependiente del valor de sonoridad, por ejemplo cuando dicho valor de sonoridad para dicho marco de señal de referencia es más pequeño que dicho umbral. Tal función puede ser una dependencia lineal, u otra dependencia adecuada sobre el valor de sonoridad. De acuerdo con una realización específica y de acuerdo con los experimentos suministra buen valor el valor de ponderación puede ser igual al valor de sonoridad cuando el valor de sonoridad para el marco de señal de referencia es más pequeño que dicho umbral.

De acuerdo con una realización adicional, además de comparar el valor de sonoridad con un primer umbral, para determinar dicho valor de ponderación dependiente de la sonoridad, el método comprende la etapa de comparar el valor de sonoridad con un segundo umbral, en donde el valor de ponderación es más pequeño que el valor máximo cuando el valor de sonoridad para el marco de señal de referencia excede el segundo umbral. El segundo umbral en esta realización es más grande que el primer umbral, y adicionalmente permite ponderar la perturbación de manera diferente dependiendo de si la perturbación se encuentra durante la pronunciación de una vocal o una consonante en la señal de voz. Se ha observado que la perturbación durante la pronunciación de una consonante se experimenta como más molesto para un receptor que la perturbación durante una vocal. De acuerdo con una realización particular, cuando dicho valor de sonoridad para dicho marco de señal de referencia excede el segundo umbral, el valor de ponderación se hace inversamente dependiente de una cantidad con la cual el valor de sonoridad excede el segundo umbral.

El valor de sonoridad se puede determinar como un valor único para el marco completo, o se puede determinar de una manera dependiente de la frecuencia. En este último caso, el valor de ponderación se hace dependiente de dicho valor de sonoridad dependiente de la frecuencia. La sonoridad es un valor dependiente de la frecuencia, en la medida en que es un parámetro que indica que tan "fuerte" se percibe un sonido por el oído humano, y el oído humano se puede

considerar como un sensor de audio dependiente de la frecuencia. Eso también revela que las perturbaciones pueden afectar la inteligibilidad dependiendo de la frecuencia de tales perturbaciones.

5 La presente invención se puede aplicar a algoritmos de evaluación de calidad tales como el POLQA o el PESQ, o su
predecesor PSQM. Estos algoritmos son particularmente desarrollados para evaluar las señales de voz degradadas.
Dentro del POLQA (algoritmo de evaluación de calidad de escucha objetivo perceptivo), el último algoritmo de
evaluación de calidad que está actualmente bajo desarrollo, la señal de voz de referencia y la señal de voz degradada
están ambos representados al menos en términos de la altura tonal y la sonoridad. Determinar el valor de sonoridad de
10 un marco es por lo tanto directo en POLQA, haciendo la aplicación de la presente invención en particular útil para este
algoritmo (P. 863).

De acuerdo con un segundo aspecto, la invención está dirigida a un producto de programa de ordenador que
comprende un código ejecutable de ordenador para efectuar un método tal como se describió anteriormente cuando se
15 ejecuta por un ordenador.

De acuerdo con un tercer aspecto, la invención está dirigida a un aparato para efectuar un método como se describió
anteriormente, para evaluar la inteligibilidad de una señal de voz degradada, que comprende: una unidad de recepción
para recibir dicha señal de voz degradada de un sistema de transmisión de audio que transporta una señal de voz de
referencia, y para recibir dicha señal de voz de referencia; una unidad de muestreo para muestrear dicha señal de voz
20 de referencia en una pluralidad de marcos de señal de referencia, y para muestrear dicha señal de voz degradada en
una pluralidad de marcos de señal degradados; una unidad de procesamiento para determinar para cada marco de
señal de referencia una representación de señal de referencia, y para determinar para cada marco de señal degradado
una representación de señal degradada; una unidad de comparación para formar pares marco al asociar cada marco de
señal de referencia con un correspondiente marco de señal degradado, y para suministrar para cada par de marco una
25 función de diferencia que representa una diferencia entre dicho marco de señal degradado y dicho marco de señal de
referencia; una unidad compensadora para compensar dicha función de diferencia para uno o más tipos de perturbación
tales como suministrar para cada par de marco una función de densidad de perturbación que se adapta a un modelo de
percepción del auditorio humano; y dicha unidad de procesamiento esta además dispuesta para derivar de dichas
funciones de densidad de perturbación de una pluralidad de pares marco un parámetro de calidad total que es al menos
30 indicativo de dicha inteligibilidad de dicha señal de voz degradada; en donde, dicha unidad de procesamiento está
además dispuesta para: determinar un valor de sonoridad para cada uno de dichos marcos de señal de referencia; y
para determinar un valor de ponderación dependiente de dicho valor de sonoridad de dicho marco de señal de
referencia; en donde dicha unidad compensadora se conecta a dicha unidad de procesamiento, y esta además
dispuesta para ponderar dicha función de diferencia utilizando el valor de ponderación dependiente de la sonoridad
35 recibida de dicha unidad de procesamiento.

Breve descripción de los dibujos

La presente invención se explica adicionalmente por medio de las realizaciones específicas, con referencia a los dibujos
40 incluidos, en donde:

La Figura 1 suministra una revisión de una primera parte del modelo perceptivo POLQA en una realización de acuerdo
con la invención;

45 La Figura 2 suministra una revisión ilustrativa del alineamiento de frecuencia utilizado en el modelo perceptivo POLQA
en una realización de acuerdo con la invención;

La Figura 3 suministra una revisión de una segunda parte del modelo perceptivo POLQA que sigue la primera parte
ilustrada en la Figura 1, en una realización de acuerdo con la invención;

50 La Figura 4 es una revisión de una tercera parte del modelo perceptivo POLQA en una realización de acuerdo con la
invención;

La Figura 5 es una revisión esquemática de una aproximación de enmascaramiento utilizada en el modelo POLQA en
55 una realización de acuerdo con la invención;

La Figura 6 es una ilustración esquemática de la ponderación dependiente de la sonoridad de la perturbación de
acuerdo con la invención;

60 La Figura 7 es una ilustración esquemática de una realización adicional de la ponderación dependiente de la sonoridad
de la perturbación de acuerdo con la invención.

Descripción detallada

65 Modelo Perceptivo POLQA

La aproximación básica del POLQA (ITU - T rec. P. 863) es la misma que se utilizó en el PESQ (ITU – T rec. P. 862), es decir, una señal de entrada de referencia y una señal de voz de salida degradada son mapeadas en una representación interna utilizando un modelo de percepción humana. La diferencia entre dos representaciones internas se utiliza por un modelo cognitivo para predecir la calidad de voz percibida de la señal degradada. Una importante nueva idea ejecutada en el POLQA es la aproximación de idealización que retira los niveles bajos de ruido en la señal de entrada de referencia y optimiza el timbre. Cambios principales adicionales en el modelo perceptivo incluyen el modelamiento del impacto del nivel de reproducción sobre la calidad percibida y una división principal en el procesamiento de los niveles bajo y alto de distorsión.

Una revisión del modelo perceptivo utilizado en el POLQA es dado en la Fig. 1 a 4. La Fig. 1 suministra una primera parte del modelo perceptivo utilizado en el cálculo de la representación interna de la señal de entrada de referencia $X(t)$ y la señal $Y(t)$ de salida degradada. Ambos son escalados 17, 46 y las representaciones internas 13, 14 en términos del tiempo de sonoridad de la altura tonal se calculan en el número de etapas descritas adelante, después de lo cual se calcula la función 12 de diferencia, indicada en la Fig. 1 con el operador 7 de cálculo de diferencia. Se calculan dos diferentes sabores de la función de diferencia perceptiva, uno para la perturbación total introducida por el sistema que utiliza los operadores 7 y 8 bajo prueba y uno para las partes agregadas de la perturbación que utiliza los operadores 9 y 10. Este modela la asimetría en impacto entre las degradaciones causadas por los componentes tiempo – frecuencia de omisión de la señal de referencia comparada con las degradaciones originadas por la introducción de nuevos componentes de tiempo – frecuencia. En el POLQA ambos sabores se calculan en dos diferentes aproximaciones, una enfocada en el rango normal de degradaciones y una enfocada en las degradaciones fuertes que resultan en cuatro cálculos de función de diferencia 7, 8, 9, y 10 indicados en la Fig. 1.

Para las señales de salida degradadas con alabeo 49 de dominio de frecuencia se utiliza un algoritmo 52 de alineación dado en la Fig. 2. El procesamiento final para conseguir las calificaciones MOS – LQO se dan en la Fig. 3 y en la Fig. 4

El POLQA inicia con el cálculo de algunas configuraciones constantes básicas después de lo cual las densidades de potencia de altura tonal (potencia como función del tiempo y frecuencia) de referencia y degradada se derivan del tiempo y de las señales de tiempo alineadas con frecuencia. De las densidades de potencia de la altura tonal se derivan las representaciones internas de la referencia y degradadas en un número de etapas. Adicionalmente estas densidades también se utilizan para derivar los primeros tres indicadores de calidad POLQA para las distorsiones de respuesta de frecuencia (FREQ), ruido aditivo (RUIDO) y reverberaciones ambientales (REVERB). Estos tres indicadores 41, 42 y 43 de calidad se calculan de manera separada del indicador de perturbación principal con el fin de permitir un análisis de impacto balanceado en un amplio rango de diferentes tipos de distorsión. Estos indicadores también se pueden utilizar para un análisis más detallado del tipo de degradaciones que fueron encontrados en la señal de voz que utiliza una aproximación de descomposición de la degradación.

Como se estableció cuatro diferentes variantes de las representaciones internas de las representaciones de referencia y degradadas se calculan en 7, 8, 9 y 10; dos variantes enfocadas en las perturbaciones para las distorsiones normales y grandes, y dos enfocadas en las perturbaciones agregadas para las distorsiones normales y grandes. Estas cuatro diferentes variantes 7, 8, 9 y 10 son las entradas para el cálculo de las densidades de perturbación final.

Las representaciones internas de la referencia 3 se denominan como representaciones ideales por que los niveles bajos de ruido en la referencia son retirados (etapa 33) y las distorsiones de timbre como se encuentran en la señal degradada que pueden haber resultado de un timbre no óptimo de la referencia original de las grabaciones de la referencia original son parcialmente compensadas (etapa 35).

Las cuatro diferentes variantes de las representaciones ideal e interna degradada calculadas utilizando los operadores 7, 8, 9 y 10 se utilizan para calcular dos densidades 142 y 143, de perturbación final, una que representa la perturbación 142 final como una función del tiempo y la frecuencia enfocada en la degradación total y una que representa la perturbación 143 final como una función del tiempo y la frecuencia pero enfocadas en el procesamiento de la degradación agregada.

La Fig. 4 da una revisión del cálculo del MOS- LQO, la calificación MOS objetiva, de las dos densidades 142 y 143 de perturbación final y los indicadores FREQ. 41, RUIDO 42, REVERB 43.

Precómputo de las configuraciones constantes

Tamaño de la Ventana FFT que Depende de la Frecuencia de la Muestra.

El POLQA opera sobre tres diferentes velocidades 8, 16 y 48 kHz de muestra diferentes que muestrea para cual tamaño W de ventana se ajusta a respectivamente 256, 512 y 2048 muestras con el fin de hacer coincidir la ventana de análisis de tiempo del sistema de auditorio humano. El traslapo entre marcos sucesivos es el 50% utilizando una ventana Hann. El espectro de potencia, la suma de las partes reales cuadradas e imaginarias cuadradas de los componentes FFT complejos – se almacenan en disposiciones valoradas reales separadas para ambos, la señal de referencia y la degradada. La información de fase dentro del marco único es descartada en POLQA y todos los cálculos se basan en las representaciones de potencia, solamente.

Cálculo del punto de partida parada.

5 En pruebas subjetivas, el ruido usualmente inicia antes de comenzar la actividad de voz en la señal de referencia. Sin embargo uno puede esperar que el ruido de estado constante líder en una prueba subjetiva disminuye el impacto del ruido de estado constante mientras que en mediciones objetivas que tienen en cuenta el ruido líder se incrementará el impacto; por lo tanto se espera que la omisión de los ruidos líder y rezagado es la aproximación perceptiva correcta. Por lo tanto, después de haber verificado la expectativa en los datos de entrenamiento disponible, los puntos de partida y parada en el procesamiento POLQA se calculan desde el inicio y final del archivo de referencia. La suma de cinco valores de muestra absoluto sucesivos (utilizando el rango PCM de 16 bits normal +32.000) debe exceder 500 desde el inicio y el final del archivo de voz original con el fin de que esa posición sea designada como partida o final. El intervalo entre esta partida y final se define como el intervalo de procesamiento activo. Las distorsiones por fuera de este intervalo se ignoran en el procesamiento POLQA.

15 El Factor de escalamiento de potencia y sonoridad SP y SL

Para calibración del tiempo FFT a frecuencia se genera la transformación de una onda sinusoidal con una frecuencia de 1000 Hz y una amplitud de 40 dB SPL, utilizando una calibración de señal X (t) de referencia hacia 73 dB SPL. Esta onda sinusoidal es transformada al dominio de frecuencia utilizando una FFT con ventana en las etapas 18 y 49 con una longitud determinada por la frecuencia de muestra para X(t) y Y(t) respectivamente. Después de convertir el eje de frecuencia a la escala Bark en 21 y 54 la amplitud pico de la densidad de potencia de la altura tonal resultante es luego normalizada a un valor de potencia de 10^4 por la multiplicación con un factor de escalamiento de potencia SP 20 y 55 para X(t) y Y(t) respectivamente.

25 El mismo tono de referencia de 40 dB SPL se utiliza para calibrar la escala de sonoridad Sicoacústica (Sone). Después de alinear el eje de intensidad a una escala de sonoridad que utiliza la ley de Zwicker la integral de la densidad de la sonoridad sobre la escala de frecuencia Bark se normaliza en 30 y 58 a 1 Sone utilizando el factor de escalamiento de sonoridad SL 31 y 59 para X(t) y Y(t) respectivamente.

30 Escalamiento y cálculo de las densidades de potencia de altura tonal.

La señal Y(t) 5 degradada se multiplica por 46 mediante el factor C 47 de calibración, que tiene en cuenta el mapeo de la sobrecarga dB en el dominio digital a dB SPL en el dominio acústico, y luego es transformado 49 al dominio de tiempo – frecuencia con 50% de marcos FFT traslapantes. La señal X(t) 3 de referencia es escalada 17 hacia el nivel óptimo fijo predefinido de aproximadamente 73 dB SPL equivalente antes de que esta se transforme 18 al dominio de tiempo – frecuencia. Este procedimiento de calibración es fundamentalmente diferente de aquel utilizado en PESQ donde tanto la degradada como la referencia son escaladas hacia un nivel óptimo fijo predefinido. El PESQ presupone que todo desempeño llevado a cabo al mismo nivel de reproducción óptima mientras que en el POLQA se utilizan niveles de pruebas subjetivas entre 20 dB a + 6 con relación al nivel óptimo. En el modelo perceptivo POLQA uno puede así no utilizar un escalamiento hacia un nivel óptimo fijo predefinido.

Después del nivel de escalamiento se transforman 18, 49 la señal de referencia y degradada al dominio de tiempo – frecuencia utilizando la aproximación FFT con ventana. Para archivos donde el eje de la frecuencia de la señal degradada es alabeado cuando se compara con la señal de referencia un desalabeo en el dominio de frecuencia se lleva a cabo sobre los marcos FFT. En la primera etapa de este desalabeo tanto los espectros de potencia FFT de referencia como el degradado se preprocesan para reducir la influencia de ambas distorsiones de respuesta de frecuencia muy estrecha, así como también las diferencias en forma espectral total sobre los siguientes cálculos. El preprocesamiento 77 consiste en efectuar un promedio de ventana deslizante en 78 sobre ambos espectros de potencia, tomando el algoritmo 79, y efectuando una normalización de la ventana de deslizamiento en 80. Luego las alturas tonales de la referencia corriente y el marco degradado se computan utilizando un algoritmo de altura tonal subarmónico estocástico. La proporción 74 de la proporción de referencia de altura tonal degradada es luego utilizada para determinar (en la etapa 84) un rango de posibles factores de alabeo. Si es posible, este rango de búsqueda se extiende al utilizar las proporciones de altura tonal para el par de marcos precedente y sucesivo.

55 El algoritmo de alineación de frecuencia entonces se itera a través del rango de búsqueda y los alabeos 85 el espectro de potencia degradado con un factor de alabeo de la iteración corriente, y los procesos 88 del espectro de potencia alabeado tal como se describió anteriormente. La correlación de la referencia procesada y el espectro degradado alabeado procesado es luego computada (en la etapa 89) para receptáculos por debajo de 1500 Hz. Después de completar la iteración a través del rango de búsqueda, el “mejor” (es decir aquel que resulte en la correlación más alta) factor de alabeo es recuperado en la etapa 90. La correlación de la referencia procesada y el mejor espectro degradado alabeado es luego comparada contra la correlación de la referencia procesada original y el espectro degradado. El “mejor” factor de alabeo es luego mantenido 97 si la correlación se incrementa un umbral establecido. Si es necesario, el factor de alabeo se limita en 98 a un cambio relativo máximo al factor de alabeo determinado para el par de marcos previos.

65

Después del desalabeo que puede ser necesario para alinear el eje de frecuencia de referencia y degradado, la escala de frecuencia en Hz es alabeada en las etapas 21 y 54 hacia una escala de altura tonal en Bark que refleje que a bajas frecuencias, el sistema auditivo humano tiene una resolución de frecuencia más fina que a altas frecuencias. Esto se implementa al discretizar las bandas FFT y sumar las correspondientes potencias de las bandas FFT con una normalización de las partes sumadas. La función de alabeo que mapea la escala de frecuencia Hertz a la escala de altura tonal en Bark aproxima los valores dados en la literatura para este propósito, y que es conocida por el lector medianamente versado. La referencia resultante de las señales degradadas es conocida como las densidades de potencia de altura tonal $PPX(f)_n$ (no indicadas en la Fig. 1) y $PPY(f)_n$ 56 con f la frecuencia en Bark y el índice n que representa el índice de marco.

Cómputo de los marcos activo, silente y súper silente de voz (etapa 25)

El POLQA opera en tres clases de marcos, que se distinguen en la etapa 25:

los marcos activos de voz donde el nivel de marco de la señal de referencia está por encima de un nivel que es de aproximadamente 20 dB por debajo del promedio,

los marcos silentes donde el nivel de marco de la señal de referencia está por debajo del nivel que es aproximadamente 20 dB por debajo del promedio y

los marcos súper silente donde el nivel de marco de la señal de referencia está por debajo del nivel que es de aproximadamente 35 dB por debajo del nivel promedio.

Cálculo de los indicadores de frecuencia, ruido y reverberación.

El impacto global de las distorsiones de respuesta de frecuencia, ruido y reverberaciones ambientales se cuantifica separadamente en la etapa 40. Para el impacto de las distorsiones de respuesta de frecuencia global, se calcula un indicador 41 del espectro promedio de referencia y de las señales degradadas. Con el fin de hacer la estimación del impacto para las distorsiones de respuesta de frecuencia independientes del ruido aditivo, la densidad del espectro del ruido promedio del degradado sobre los marcos silentes de la señal de referencia se sustraen de la densidad de sonoridad de la altura tonal de la señal degradada. La densidad de la sonoridad de la altura tonal resultante de la densidad degradada y de la sonoridad de la altura tonal de la referencia son entonces promediados en cada banda Bark sobre todos los marcos activos de voz para el archivo de referencia y degradado. La diferencia en la densidad de sonoridad de la altura tonal entre estas dos densidades luego degradado sobre la altura tonal para derivar el indicador 41 para cuantificar el impacto de las distorsiones (FREC) de respuesta de frecuencia.

Para el impacto del ruido aditivo, se calcula un indicador 42 del espectro promedio de la señal degradada sobre los marcos silentes de la señal de referencia. La diferencia entre la densidad de sonoridad de la altura tonal promedio de los marcos degradados sobre los silentes y la densidad de sonoridad de la altura tonal de referencia cero determinan una función de la densidad de sonoridad de ruido que cuantifica el impacto del ruido aditivo. La función de densidad de sonoridad de ruido es luego integrada sobre la altura tonal para derivar un indicador 42 (RUIDO) de impacto de ruido promedio. Este indicador 42 es así calculado de un silencio ideal de tal manera que una cadena transparente que se mide utilizando una señal de referencia de ruido no suministrará la máxima calificación MOS en las mediciones de calidad de voz de extremo a extremo POLQA finales.

Para el impacto de las reverberaciones ambientales, se calcula la función de energía sobre tiempo (ETC) proveniente de la serie de tiempo de referencia y degradadas. El ETC representa la cubierta de la respuesta de impulso. En una primera etapa la reflexión más ruidosa se calcula al simplemente determinar el valor máximo de la curva ETC después del sonido directo. En el sonido directo el modelo POLQA se define como los sonidos que llegan dentro de 60 ms. Luego una segunda reflexión más ruidosa se determina sobre el intervalo sin el sonido directo y sin tener en cuenta las reflexiones que llegan dentro de los 100 ms desde la reflexión más fuerte. Luego se determina la tercera reflexión más fuerte sobre el intervalo sin el sonido directo y sin tener en cuenta las reflexiones que llegan dentro de los 100 ms desde la reflexión más fuerte y la segunda más fuerte. La energía de las tres reflexiones más fuertes se combina entonces en un indicador 43 de reverberación única (REVERB).

Escalamiento global y local de la señal de referencia hacia la señal (etapa 26) degradada.

La señal de referencia está ahora de acuerdo con la etapa 17 en el nivel ideal interno, es decir equivalente a aproximadamente 73 dB SPL, mientras que la señal degradada se representa a un nivel que coincide con el nivel de reproducción como resultado de 46. Antes de que se haga la comparación entre la señal de referencia y la degradada las diferencias de nivel global se compensan en la etapa 26. Adicionalmente, pequeños cambios en el nivel local son parcialmente compensados para contar el hecho de que variaciones de nivel suficientemente pequeñas no sean notorias a sujetos en una situación de solo escuchar. El igualamiento 26 de nivel global se lleva a cabo sobre la base de la potencia promedio de referencia y la señal degradada utilizando los componentes de frecuencia entre 400 y 3500 Hz. La señal de referencia es escalada globalmente hacia la señal degradada el impacto de la diferencia de nivel de reproducción global es mantenida así en este rango de procesamiento. De manera similar, para variar lentamente las

distorsiones de la ganancia se lleva a cabo un escalamiento local para cambios de nivel hasta de aproximadamente 3 dB utilizando el ancho de banda completo tanto del archivo de voz de referencia como el degradado.

5 Compensación parcial de la densidad de potencia de la altura tonal original para las distorsiones (etapa 27) de respuesta de frecuencia lineal

10 Con el fin de modelar correctamente el impacto de las distorsiones de respuesta de la frecuencia lineal, inducidas por el filtrado del sistema bajo prueba, una aproximación de compensación parcial se utiliza en la etapa 27. Para modelar la imperceptibilidad de las distorsiones de respuesta de frecuencia lineal moderadas en las pruebas subjetivas, la señal de referencia es parcialmente filtrada con las características de transferencia del sistema bajo prueba. Esto se lleva a cabo al calcular el espectro de potencia promedio de las densidades de potencia de altura tonal original y degradada sobre todos los marcos activos de voz. Por receptáculo Bark, se calcula 27 un factor de compensación parcial de la relación del espectro degradado al espectro original.

15 Modelamiento de los efectos de enmascaramiento, cálculo de la excitación de densidad de la sonoridad de la altura tonal

20 El enmascaramiento es modelado en las etapas 30 y 58 al calcular una representación remanente de las densidades de potencia de altura tonal. Tanto la remanencia del tiempo como el dominio de frecuencia son tomadas en cuenta de acuerdo con los principios ilustrados en la Fig. 5a a 5c. La remanencia del dominio de tiempo – frecuencia utiliza la aproximación de convolución. Para esta representación remanente, las representaciones de la referencia y la densidad de potencia de altura tonal degradada se recalculan suprimiendo los componentes de tiempo – frecuencia de baja amplitud, que son parcialmente enmascarados por los componentes ruidosos en la vecindad en el plano de tiempo – frecuencia. Esta supresión se ejecuta de dos diferentes maneras, una sustracción de la representación remanente de la representación no remanente y una división de la representación no remanente por la presentación remanente. Las representaciones resultantes claras de la densidad de potencia de la altura tonal son entonces transformadas a representaciones de densidad de sonoridad de altura tonal que utiliza una versión modificada de la ley de potencia de Zwicker:

$$LX(f)_n = SL * \left(\frac{P_0(f)}{0.5} \right)^{0.22 * f_B * P_n} * \left[\left(0.5 + 0.5 \frac{PPX(f)_n}{P_0(f)} \right)^{0.22 * f_B * P_n} - 1 \right]$$

30 Con SL el factor de escalamiento de sonoridad, P0(f) el umbral de escucha absoluto, fB y Pfn una corrección dependiente de la frecuencia y el nivel definido por

$$\begin{aligned} f_B &= -0.03 * f + 1.06 && \text{para } f < 2.0 \text{ Bark} \\ f_B &= 1.0 && \text{para } 2.0 \leq f \leq 22.0 \text{ Bark} \\ f_B &= -0.2 * (f - 22.0) + 1.0 && \text{para } f > 22.0 \text{ Bark} \\ P_{fn} &= (PPX(f)_n + 600)^{0.008} \end{aligned}$$

40 Con f representando la frecuencia en Bark, PPX(f)_n la densidad de potencia de altura tonal en la celda de tiempo de frecuencia f, n. Los dos arreglos dimensionales resultantes LX(f)_n y LY(f)_n son denominados densidades de sonoridad de altura tonal, en la salida de la etapa 30 para la señal X(t) de referencia y la etapa 58 de la señal Y(t) degradada respectivamente.

45 Supresión de ruido a nivel bajo global en las señales de referencia y degradadas.

50 Los niveles bajos de ruido en la señal de referencia, que no se afectan por el sistema bajo ensayo (por ejemplo un sistema transparente) se atribuirán al sistema bajo ensayo por los sujetos debido al procedimiento de prueba de calificación de categoría absoluta. Estos niveles bajo de ruido tienen así que ser suprimidos en el cálculo de la representación interna de la señal de referencia. Este "proceso de idealización" se lleva a cabo en la etapa 33 al calcular la densidad de sonoridad de ruido de estado constante promedio de la señal LX(f)_n de referencia sobre los marcos súper silentes como una función de la altura tonal. Esta densidad de sonoridad de ruido promedio es luego parcialmente sustraída de todos los marcos de densidad de sonoridad de altura tonal de la señal de referencia. El resultado es una representación interna idealizada de la señal de referencia, a la salida de la etapa 33.

55 El ruido de estado estable que es audible en la señal degradada tiene un menor impacto que el ruido de estado no estable. Este mantiene todos los niveles de ruido y el impacto de este efecto se puede modelar al retirar parcialmente el ruido de estado estable proveniente de la señal degradada. Esto se lleva a cabo en la etapa 60 al calcular la densidad de sonoridad de ruido de estado estable promedio de los marcos de la señal LY(f)_n degradada para los cuales los correspondientes marcos de la señal de referencia se clasifican como súper silentes, como una función de la altura tonal. Esta densidad de sonoridad de ruido promedio es luego parcialmente sustraída de todos los marcos de densidad de sonoridad de altura tonal de la señal degradada. La compensación parcial utiliza una estrategia diferente para bajos y altos niveles de ruido. Para los bajos niveles de ruido la compensación es solamente marginal mientras que la supresión

que se utiliza se vuelve más agresiva para el ruido aditivo fuerte. El resultado es una representación 61 interna de la señal degradada con un ruido auditivo que se adapta al impacto subjetivo tal como se observa en las pruebas de escucha utilizando una representación libre de ruido idealizada de la señal de referencia.

5 En la presente realización, en la etapa 33 anterior, además de efectuar la supresión de ruido a nivel bajo global, también el indicador 32 VOLUMEN se determina para cada uno de los marcos de señal de referencia, de acuerdo con la presente invención. El indicador VOLUMEN o el valor VOLUMEN se utilizará para determinar un factor ponderado dependiente de la sonoridad para ponderar tipos específicos de distorsiones. La ponderación misma se puede
10 implementar en las etapas 125 y 125' para cuatro representaciones de distorsiones suministradas por los operadores 7, 8, 9 y 10, luego de suministrar las densidades 142 y 143 de perturbación final.

Aquí, el indicador de nivel de sonoridad se ha determinado en la etapa 33, pero uno puede apreciar que el indicador de nivel de sonoridad se puede determinar para cada marco de señal de referencia en otra parte del método. En la etapa 33 determinar el indicador de nivel de sonoridad es posible debido al hecho de que ya la densidad fuerte de ruido de estado estable promedio se determinó para la señal $LX(f)_n$ de referencia de los marcos súper silentes, que son entonces
15 utilizados en la construcción de la señal de referencia libre de ruido para todos los marcos de referencia. Sin embargo, aunque es posible ejecutar este en la etapa 33, no es la manera más preferida de ejecución.

De manera alternativa, el indicador de nivel de sonoridad (VOLUMEN) se puede tomar de la señal de referencia en una etapa adicional que sigue a la etapa 35. Esta etapa adicional también se indica en la Figura 1 como una casilla 35' punteada con una salida 32' (VOLUMEN) de línea punteada. Si se ejecuta allí en la etapa 35', ya no es necesario tomar el indicador de nivel de sonoridad de la etapa 33, como una persona medianamente versada lo pudiera apreciar.

Escalamiento local de la densidad de sonoridad de la altura tonal distorsionada para ganancia con variación de tiempo entre la señal degradada y la de referencia (etapas 34 y 63)
25

Variaciones lentas en la ganancia son inaudibles y los pequeños cambios ya están compensados para el cálculo en la representación de señal de referencia. La compensación restante necesaria antes de que se pueda calcular la representación interna correcta se lleva a cabo en dos etapas; primero la referencia es compensada en la etapa 34 para los niveles de señal donde la sonoridad de señal degradada es menor que la sonoridad de la señal de referencia, y segundo la degradada se compensa en la etapa 63 para los niveles de señal donde la sonoridad de la señal de referencia es menor que la sonoridad de la señal degradada.
30

La primera compensación 34 escala la señal de referencia hacia un nivel inferior para las partes de la señal donde el degradado muestra una pérdida severa de señal tal como en situaciones de recorte de tiempo. El escalamiento es tal que la diferencia restante entre la referencia y la degradada representa el impacto del recorte de tiempo en la calidad de voz percibida local. Las partes donde la sonoridad de la señal de referencia es menor que la sonoridad de la señal degradada no están compensadas y así el ruido aditivo y el recorte de ruido no se compensan en esta primera etapa.
35

La segunda compensación 63 escala la señal degradada hacia un nivel inferior para las partes de la señal donde la señal degradada muestra pulsaciones y para las partes de la señal donde no existe ruido en los intervalos silentes. El escalamiento es tal que la diferencia restante entre la referencia y la degradada representa el impacto de las pulsaciones y cambia lentamente el ruido aditivo sobre la calidad de voz percibida local. Mientras que las pulsaciones son compensadas tanto en las partes activas silentes como de voz, el ruido es compensado solamente en las partes silentes.
40
45

Compensación parcial de la densidad de sonoridad de la altura tonal original para distorsiones de respuesta de frecuencia lineal (etapa 35)

50 Las distorsiones de la respuesta de frecuencia lineal imperceptibles ya fueron compensadas al filtrar parcialmente la señal de referencia en el dominio de densidad de potencia de altura tonal en la etapa 27. Con el fin de corregir adicionalmente el hecho de que las distorsiones lineales son menos inaceptables que las distorsiones no lineales, la señal de referencia es ahora parcialmente filtrada en la etapa 35 en el dominio de sonoridad de altura tonal. Esto se lleva a cabo al calcular el espectro de sonoridad promedio en las densidades original y de sonoridad de altura tonal degradadas sobre todos los marcos activos de voz. Para el receptáculo Bark, se calcula un factor de compensación parcial de la relación del espectro de sonoridad degradado al espectro de sonoridad original. Este factor de compensación parcial se utiliza para filtrar la señal de referencia con una versión suavizada, de menor amplitud de la respuesta de frecuencia del sistema bajo ensayo. Después de este filtrado, la diferencia entre las densidades de sonoridad de altura tonal de referencia y degradada que resultan de las distorsiones de respuesta de frecuencia lineal se disminuye a un nivel que representa el impacto de las distorsiones de la respuesta de frecuencia lineal sobre la calidad de voz percibida.
55
60

Escalamiento final y supresión de ruido de la densidad de sonoridad de altura tonal

65 Hasta este punto todos los cálculos sobre las señales son llevados a cabo a nivel de reproducción como se utiliza en el experimento subjetivo. Para los niveles bajos de reproducción, este dará como resultado una diferencia baja entre las

5 densidades de sonoridad de altura tonal de referencia y degradadas y en general en una estimación demasiado optimista de la calidad de voz que se escucha. Con el fin de compensar este efecto la señal degradada es ahora escalada hacia un nivel interno fijo "virtual" en la etapa 64. Después de este escalamiento, la señal de referencia es escalada en la etapa 36 hacia el nivel de señal degradado y tanto la señal de referencia como la degradada están listas para una operación de supresión de ruido final entre 37 y 65 respectivamente. Esta supresión de ruido tiene cuidado de las últimas partes de los niveles de ruido de estado estable en el dominio de sonoridad que aún tiene gran impacto en el cálculo de la calidad de voz. Las señales 13 y 14 resultantes son ahora un dominio de representación interna relevante perceptivo y de la altura tonal – sonoridad – tiempo ideal $LX_{ideal}(f)_n$ 13 y de la altura tonal – sonoridad – tiempo degradado $LY_{deg}(f)_n$ 14 se pueden calcular las densidades 142 y 143 de perturbación. Cuatro diferentes variantes de las funciones de altura tonal – sonoridad – tiempo ideal y degradada se calculan en 7, 8, 9 y 10, dos variantes, (7 y 8) enfocadas en las perturbaciones para las distorsiones normales y grandes, y dos (9 y 10) enfocadas en las perturbaciones agregadas para las distorsiones normales y grandes.

15 Cálculo de las densidades de perturbación final

Se calculan dos diferentes sabores de las densidades 142 y 143 de perturbación. La primera, la densidad de perturbación normal, se deriva en 7 y 8 de la diferencia entre la función de altura tonal – sonoridad – tiempo ideal $LX_{ideal}(f)_n$ y la función de altura tonal – sonoridad – tiempo degradado $LY_{deg}(f)_n$. La segunda se deriva en 9 y 10 de la función de altura tonal – sonoridad – tiempo ideal y altura tonal- sonoridad – tiempo degradado utilizando versiones que se optimizan con relación a las degradaciones introducidas y que se denominan perturbaciones agregadas. En este cálculo de la perturbación agregada, las partes de señal donde la densidad de potencia degradada es mayor que la densidad de potencia de referencia son ponderadas con un factor dependiente de la proporción o de la relación de potencia en cada celda de altura tonal – tiempo, el factor de asimetría.

25 Para poder tratar con un rango grande de distorsiones se llevan a cabo dos diferentes versiones de procesamiento, una enfocada en distorsiones pequeñas a medias basadas en 7 y 9 y una enfocada en distorsiones medias a grandes basadas en 8 y 10. La conmutación entre las dos se lleva a cabo sobre la base de una primera estimación de la perturbación enfocada en un nivel pequeño a medio de las distorsiones. Esta aproximación de procesamiento conduce a la necesidad de calcular cuatro diferentes funciones de altura tonal – sonoridad – tiempo ideales y cuatro diferentes funciones de altura tonal – sonoridad – tiempo degradadas con el fin de poder calcular una perturbación única y una función de perturbación agregada única (ver Fig. 3) que son entonces compensadas por un número de diferentes tipos de cantidades severas de distorsiones específicas.

35 Las desviaciones severas del nivel de escucha óptimo están cuantificadas en 127 y 127' por un indicador directamente derivado del nivel de señal de la señal degradada. Este indicador (LEVEL) global también se utiliza en el cálculo de MOS – LQO.

40 Las distorsiones severas introducidas por las repeticiones de los marcos se cuantifican 128 y 128' mediante un indicador derivado de una comparación de la correlación de los marcos consecutivos de la señal de referencia con la correlación de los marcos consecutivos de la señal degradada.

45 Las desviaciones severas del timbre "ideal" óptimo de la señal degradada se cuantifican 129 y 129' por un indicador derivado de la proporción de la sonoridad de la banda de frecuencia superior y la sonoridad de la banda de frecuencia inferior. Las compensaciones se llevan a cabo por marco y a nivel global. Esta compensación calcula la potencia en las bandas Bark inferior y superior (por debajo de 12 y por encima de 7 Bark, es decir, utilizando un traslape de 5 Bark) de la señal degradada y "castiga" cualquier desbalance severo sin importar el hecho de que este pudiera ser el resultado de un timbre de voz incorrecto del archivo de voz de referencia. Nótese que una cadena transparente que utiliza señales de referencia pobremente registradas, que contienen demasiado ruido y/o un timbre de voz incorrecto, no suministrarán así la calificación MOS máxima en una medición de calidad de voz extremo a extremo POLQA. Esta compensación también tiene un impacto cuando se mide la calidad de los dispositivos que son transparentes. Cuando se utilizan señales de referencia que muestran una desviación significativa del timbre "ideal" óptimo del sistema bajo ensayo se juzgaran como no transparentes aún si el sistema no introduce ninguna degradación en la señal de referencia.

55 El impacto de los picos severos en la perturbación se cuantifica en 130 y 130' en el indicador APLANADO que también se utiliza en el cálculo del MOS- LQO.

Las variaciones de nivel de ruido severas que enfocan la atención de los sujetos hacia el ruido se cuantifican en 131 y 131' mediante un indicador de contraste de ruido derivado de las partes silentes de la señal de referencia.

60 En las etapas 133 y 133', de acuerdo con la invención, se efectúa una operación de ponderación para ponderar las perturbaciones dependientes de si ellas coinciden o no con la voz hablada real. Con el fin de evaluar la inteligibilidad de la señal degradada, las perturbaciones que son percibidas durante periodos silentes no se consideran como de detrimento en la medida en que las perturbaciones son percibidas durante la voz hablada real. Por lo tanto, de acuerdo con la invención, con base en el indicador VOLUMEN determinado en la etapa 33 (o en la etapa 35' en la realización alternativa) proveniente de la señal de referencia, se determina un valor de ponderación para ponderar cualquier perturbación. El valor de ponderación se utiliza para ponderar la función de diferencia (es decir perturbaciones) para

incorporar el impacto de las perturbaciones sobre la inteligibilidad de la señal de voz degradada en la evaluación. En particular, en razón a que el valor de ponderación se determina con base en el indicador de VOLUMEN, el valor de ponderación se puede representar mediante una función dependiente de la sonoridad. En la presente realización, el valor de ponderación dependiente de la sonoridad se determina al comparar el valor de ponderación a un umbral. Si el indicador de sonoridad excede el umbral las perturbaciones percibidas son tomadas completamente en consideración cuando se efectúa la evaluación. De otro lado, si el valor de sonoridad es más pequeño que el umbral, el valor de ponderación se hace dependiente del indicador del nivel de sonoridad; es decir, en la realización presente el valor de ponderación es igual a el indicador del nivel de sonoridad (en el régimen donde VOLUMEN está por debajo del umbral). La ventaja es que para las partes débiles de la señal de voz, por ejemplo, en los extremos de las palabras habladas justo antes de una pausa o silencio, las perturbaciones son tomadas parcialmente en cuenta por tener efectos de detrimentos sobre la inteligibilidad. Como un ejemplo, uno puede apreciar que una cierta cantidad de ruido percibido mientras se pronuncia la letra "f" al final de una palabra, puede originar que el receptor perciba esto como la letra "s". Esto puede ir en detrimento de la inteligibilidad. De otro lado, las personas expertas pueden apreciar que también es posible (en una diferente realización) simplemente no tener en cuenta cualquier ruido durante el silencio o pausas, al cambiar el valor de ponderación a cero cuando el valor de sonoridad está por debajo del umbral anteriormente mencionado. El método de ponderar la perturbación de una manera dependiente de la sonoridad esta además descrito adelante en relación con la Figura 6.

Adicionalmente a lo anterior el método propuesto puede además ser extendido para tener en cuenta el hecho de que las perturbaciones que son percibidas durante la pronunciación de las vocales en una señal de voz no tienen efecto de detrimento en la medida en que las perturbaciones que no tienen detrimento como las perturbaciones que son percibidas durante las consonantes. El análisis de la cubierta de potencia de la señal de voz revela que generalmente, la sonoridad de la señal durante la pronunciación de las vocales representa un máximo local, mientras que durante la pronunciación de las consonantes la sonoridad esta usualmente a un nivel intermedio. Las perturbaciones durante la pronunciación de una consonante tienen más impacto sobre la inteligibilidad de la voz que las perturbaciones durante las vocales donde la potencia de señal es suficientemente fuerte para que el observador identifique la vocal. Por lo tanto, como una mejora adicional, el valor de sonoridad se puede comparar con dos umbrales. La comparación de la sonoridad con el primer umbral hará que el sistema opere como se indicó anteriormente; es decir, la sonoridad está por debajo del primer umbral haciendo más pequeño el valor de ponderación que un valor máximo y dependiente de la sonoridad, mientras que exceder el primer umbral origina que el valor de ponderación se establezca al máximo (por ejemplo 1.0 para tener en cuenta completamente la perturbación). La comparación de la sonoridad con el segundo umbral hará que el sistema opere como sigue. Si la sonoridad está por debajo del segundo umbral, el valor de ponderación será más pequeño que el valor máximo y dependiente de la sonoridad. Si la sonoridad excede el primer umbral, el valor de ponderación se ajusta a un valor máximo. Esta realización del método de ponderar la perturbación se ilustra en la Figura 7.

Procediendo de nuevo con la Figura 3, se detectan severos saltos en el alineamiento y el impacto se cuantifica en las etapas 136 y 136' mediante un factor de compensación.

Finalmente, la perturbación y las densidades de perturbación agregadas son recortadas en 137 y 137' a un nivel máximo y la varianza de la perturbación 138 y 138' y el impacto de los saltos 140 y 140' en la sonoridad de la señal de referencia se utilizan para compensar las estructuras de tiempo específica de las perturbaciones.

Esto produce la densidad de perturbación final de $D(f)_n$ 142 para la perturbación regular y la densidad $DA(f)_n$ 143 de perturbación final para la perturbación agregada.

Agregación de la perturbación sobre la altura tonal, los acelerones y el tiempo, mapeando la calificación MOS intermedia

La perturbación final $D(f)_n$ 142 y las densidades de perturbación $DA(f)_n$ agregadas 143 son integradas por marco sobre el eje de altura tonal que resulta en dos diferentes perturbaciones por marco, una derivada de la perturbación y una derivada de la perturbación agregada, utilizando la integración 153 y 159 L1 (ver Fig. 4):

$$D_n = \sum_{f=1, \dots, \text{Número de bandas Bark}} |D(f)_n| W_f$$

$$DA_n = \sum_{f=1, \dots, \text{Número de bandas Bark}} |DA(f)_n| W_f$$

Con W_f una serie de constantes proporcionales al ancho de los receptáculos Bark.

Luego estas dos perturbaciones por marco son promediadas sobre los acelerones de voz de seis marcos consecutivos con un L_4 155 y un L_1 160 ponderado para la perturbación y para la perturbación agregada, respectivamente.

$$DS_n = \sqrt[4]{\frac{1}{6} \sum_{m=n, n+6} D_m^4}$$

$$DAS_n = \frac{1}{6} \sum_{m=n, n+6} D_m$$

Finalmente, la perturbación y la perturbación agregada se calculan por archivo del L₂ 156 y 161 promediando durante el tiempo

$$D = \sqrt{\frac{1}{\text{número De Marcos}} \sum_{n=1, \dots, \text{número De Marcos}} DS_n^2}$$

$$DA = \sqrt{\frac{1}{\text{número De Marcos}} \sum_{n=1, \dots, \text{número De Marcos}} DAS_n^2}$$

5

La perturbación agregada se compensa en la etapa 161 para reverberaciones fuertes y ruido auditivo fuerte utilizando los indicadores REVERB 42 y RUIDO 43. Las dos perturbaciones son entonces combinadas 170 con un indicador 41 (FREC) de frecuencia para derivar un indicador interno que está linealizado con el tercer orden de regresión polinomial para conseguir un MOS como el indicador 171 intermedio.

10

Computación del POLQA MOS- LQO Final

La calificación POLQA bruta se deriva del MOS como un indicador intermedio que utiliza cuatro diferentes compensaciones todas en la etapa 175:

15

Dos compensaciones para las características de tiempo- frecuencia específicas de la perturbación, una calculada con una agregación L₅₁₁ sobre la frecuencia 148, acelerones 149 y tiempo 150, y uno calculado con una agregación L₃₁₃ sobre la frecuencia 145, acelerones 146 y el tiempo 147.

20

Una compensación para niveles de representación muy bajo que utilizan el indicador de NIVEL.

una compensación para distorsiones de timbre grandes utilizando el indicador de APLANADO

25

El entrenamiento de este mapeo se lleva a cabo en un conjunto grande de degradaciones, que incluye las degradaciones que no fueron parte del punto de referencia POLQA. Estas calificaciones MOS brutas 176 son en la mayor parte ya linealizadas por el tercer mapeo polinomial de orden utilizado en el cálculo del MOS como el indicador 171 intermedio.

30

Finalmente las calificaciones 176 POLQA MOS brutas se mapean en 180 hacia las calificaciones 181 MOS- LQO utilizando un tercer orden polinomial que se optimiza para las 62 bases de datos como estaban disponibles al final de la etapa de la estandarización POLQA. En el modo de banda estrecha la máxima calificación POLQA MOS- LQO es 4.5 mientras que en el modo de súper ancho de banda este punto es de 4.75. Una consecuencia importante del proceso de idealización es que bajo algunas circunstancias, cuando la señal de referencia contiene ruido o cuando el timbre de voz se distorsiona severamente, una cadena transparente no suministrará la calificación MOS máxima de 4.5 en el modo de ancho de banda de 4,75 en el modo súper ancho de banda.

35

La Fig. 6 ilustra una revisión de un método de ponderación de la perturbación o ruido con respecto al valor de sonoridad de acuerdo con la presente invención. Aunque el método como se ilustró en la Figura 6 solo se enfoca en las partes relevantes que se relacionan con determinar el valor de sonoridad y efectuar la ponderación de las perturbaciones, se apreciará que este método se puede incorporar como parte de un método de evaluación como se describió en este documento, o una alternativa de este.

40

En la etapa 222, se determina un valor de sonoridad para cada marco de la señal 220 de referencia. Esta etapa se puede implementar en la etapa 33 de la Figura 1, o como se describió anteriormente en la etapa 35' también descrita en la Figura 1 como una alternativa preferida. La persona experta puede apreciar que el valor de sonoridad se puede determinar de alguna otra manera en el método, siempre y cuando el valor de sonoridad esté disponible a tiempo al efectuar la ponderación.

45

En la etapa 225, el valor de sonoridad determinado en la etapa 222 se compara con un umbral 226. El resultado de esta comparación puede ser que el valor de sonoridad sea mayor que el umbral 226, en cuyo caso el método por vía de 228;

50

o que el valor de sonoridad puede ser más pequeño que el umbral 226, en cuyo caso el método continua hasta la senda 231.

Si el valor de sonoridad es mayor que el umbral (senda 228) en la etapa 230 se determina el factor de ponderación dependiente de la sonoridad. En la presente realización, el factor de ponderación se ajusta a 1.0 con el fin de tener completamente en cuenta la perturbación y la señal degradada. La persona experta apreciará que la situación donde el valor de sonoridad es mayor que el umbral corresponde a la señal de voz que lleva información en el tiempo presente (el marco de señal de referencia coincide con las palabras reales que son habladas). La invención no está limitada al factor de ponderación de 1.0 en la situación anteriormente mencionada; la persona experta puede optar por utilizar cualquier otro valor o dependencia considerada adecuada para una situación dada. La invención primariamente se enfoca en hacer una distinción entre las perturbaciones encontradas durante el habla y las perturbaciones encontradas durante (casi) los periodos silentes, al tratar las perturbaciones de manera diferente en ambos regímenes.

En el caso en que el valor de la sonoridad sea menor que el umbral y el método continúa a la senda 231, en la etapa 233 el valor de ponderación se determina al establecer el factor de ponderación por ser dependiente sobre el valor de sonoridad. Se han experimentado buenos resultados al utilizar directamente el valor de sonoridad como un factor de ponderación. Sin embargo se puede aplicar cualquier dependencia adecuada, es decir, lineal, cuadrática, una polinomial o cualquier orden adecuado, u otra dependencia. El factor de ponderación debe ser más pequeño de 1.0 como se apreciará.

Como una alternativa al factor de ponderación dependiente de la sonoridad descrita, también es posible incluir una dependencia de frecuencia de la sonoridad en el método de la presente invención. En ese caso, el factor de ponderación no solo será dependiente de la sonoridad, sino también de la frecuencia de la perturbación en la señal de voz.

El factor de ponderación determinado en una de las etapas 230 y 233 se utiliza como un valor 235 de entrada para ponderar la importancia de las perturbaciones en la etapa 240 como una función de si o no la señal degradada lleva de hecho voz hablada en el presente marco. En la etapa 240, la señal 238 de diferencia es recibida y se aplica el factor 235 de ponderación para suministrar la salida deseada (OUT).

La Fig. 7 ilustra una revisión de una realización adicional de un método de ponderar la perturbación o ruido con respecto al valor de sonoridad de acuerdo con la presente invención. En vista de las similitudes entre las Figuras 6 y 7, en la Figura 7 se han utilizado los mismos signos de referencia que en la Figura 6 para los elementos y etapas del método que son similares o equivalentes al método descrito en la Figura 6. De nuevo, el método como se ilustra en la Figura 7 solo se enfoca en las partes relevantes que se relacionan con determinar el valor de la sonoridad y efectuar la ponderación de las perturbaciones, pero se apreciará que este método se puede incorporar como parte de un método de evaluación como se describió en este documento, o una alternativa de este.

En la etapa 222, el valor de sonoridad se determina para cada marco de la señal 220 de referencia. Esta etapa se puede ejecutar en la etapa 33 de la Figura 1, o como se describió anteriormente en la etapa 35' también descrita en la Figura 1 como una alternativa preferida. La persona experta puede apreciar que el valor de sonoridad se puede determinar en alguna otra parte en el método, siempre y cuando el valor de sonoridad esté disponible a tiempo al efectuar la ponderación.

En la etapa 225, el valor de sonoridad determinado en la etapa 222 se compara con un primer umbral 226. El resultado de esta comparación puede ser que el valor de sonoridad sea mayor que el primer umbral 226, en cuyo caso el método continuado por vía del 228; o que el valor de sonoridad sea más pequeño que el primer umbral 226, en cuyo caso el método continua a través de la senda 231.

Si el valor de sonoridad es mayor que el primer umbral (senda 228), en la etapa 242, el valor de sonoridad se compara con un segundo umbral 243, el segundo umbral 243 es mayor que el primer umbral 226. El resultado de esta comparación puede ser que el valor de sonoridad sea mayor que el segundo umbral 243, en cuyo caso el método continúa por vía de 245; o que el valor de sonoridad pueda ser más pequeño que el umbral 243, en cuyo caso el método continúa a la senda 248.

Si el valor de sonoridad es más pequeño que el segundo umbral 243 (senda 248), en la etapa 249 se determina el factor de ponderación dependiente de la sonoridad. En la presente realización, el factor de ponderación se ajusta a 1.0 (un valor máximo) con el fin de tomar en cuenta completamente la perturbación en la señal degradada. La persona experta apreciará que la situación donde el valor de sonoridad es mayor que el umbral corresponde a la señal de voz durante la pronunciación de una vocal; es decir, un máximo local en la envoltura de potencia. La invención está limitada al factor de ponderación de 1.0 en la situación anteriormente mencionada; la persona experta puede optar por utilizar cualquier otro valor o dependencia considerada adecuada para una situación dada. En esta realización, la invención se enfoca en hacer una distinción entre las perturbaciones encontradas durante la voz y las perturbaciones encontradas durante (casi) los periodos silentes. Más aún, cuando la perturbación es encontrada durante la voz, esta realización se enfoca además en hacer una distinción entre la perturbación encontrada durante la pronunciación de las vocales y las

perturbaciones encontradas durante la pronunciación de las consonantes. Las perturbaciones son tratadas de manera diferente en cada uno de estos regímenes.

5 En caso de que el valor de sonoridad sea mayor que el segundo umbral 243 y el método continua hasta la senda 245, la etapa 246 se determina el valor de ponderación al ajustar el factor de ponderación como siendo dependiente del valor de sonoridad. Se han experimentado buenos resultados al hacer el factor de ponderación dependiente de la siguiente manera:

10 Valor de ponderación = (sonoridad –segundo umbral + 1.0)^{-1+q}

En donde el factor q de potencia se puede igualar a cualquier valor deseado. Se obtuvieron buenos resultados con q = 0,3

15 En lugar de la relación anterior, cualquier dependencia adecuada se puede aplicar, es decir, lineal, cuadrática, polinomial o cualquier orden adecuado, u otra dependencia. El factor de ponderación puede ser más pequeño que el valor máximo 1.0 como se apreciará.

20 Como una alternativa al factor de ponderación dependiente de la sonoridad descrito anteriormente, también es posible incluir la dependencia de la frecuencia de la sonoridad en el método de la presente invención. En ese caso, el factor de ponderación no será dependiente de la sonoridad, sino también de la frecuencia de la perturbación en la señal de voz.

25 El factor de ponderación determinado en una de las etapas 233, 246 o 249 se utiliza como un valor 235 de entrada para ponderar la importancia de las perturbaciones en la etapa 240 como una función de si o no la señal degradada realmente lleva voz hablada en el presente marco. En la etapa 240, se recibe la señal 238 de diferencia y se aplica el factor 235 de ponderación para suministrar la salida (OUT) deseada. La invención se puede practicar de manera diferente que la específicamente descrita aquí, y el alcance de la invención está limitado a las realizaciones específicas descritas anteriormente y a los dibujos anexos, sino que puede variar dentro del alcance como se define en las reivindicaciones anexas.

30 Signos de referencia

3 señal X(t) referencia

5 señal Y(t) degradada, amplitud-tiempo

7 cálculo de diferencia

35 8 primer variante de cálculo de diferencia

9 segunda variante de cálculo de diferencia

10 tercer variante de cálculo de diferencia

12 señal de diferencia

13 altura tonal-sonoridad-tiempo $LX_{ideal}^{(f)}_n$ ideal interna

40 14 altura tonal-sonoridad-tiempo $LY_{deg}^{(f)}_n$ degradado interno

17 escalamiento global hacia el nivel fijo

18 FFT de ventana

20 factor de escalamiento SP

21 alabeo a Bark

45 25 detecciones de marco (súper) silente

26 escalamiento global & local a nivel degradado

27 compensación de frecuencia parcial

30 excitación y alabeo a sone

31 factor SL de escalamiento de umbral absoluto

50 32 VOLUMEN

32` VOLUMEN (determinado de acuerdo a la etapa 35` alternativa)

33 supresión de ruido a nivel bajo global

34 escalamiento local si $Y < X$

35 compensación de frecuencia parcial

55 35` (alternativa) determinar sonoridad

36 escalamientos hacia nivel degradado

37 supresión de ruido a nivel bajo global

40 indicadores FREC RUIDO REVERB

41 indicador FREC

60 42 indicador RUIDO

43 indicador REVERB

44 indicador $PW_{R_{total}}$ (relación de potencia audio total entre señal de grad. Y señal de ref.)

45 indicador $PW_{R_{marco}}$ (por relación de potencia de audio de marco entre la señal de grad. y ref.)

46 escalamientos a nivel de reproducción

65 47 factor C de calibración

49 FFT de ventana

- 52 alineación de frecuencia
- 54 alabeo a Bark
- 55 factor SP de escalamiento
- 56 altura tonal-potencia-tiempo $PPY^{(f)}$ de señal degradada
- 5 58 excitación y alabeo a sone
- 59 factor SL de escalamiento de umbral absoluto
- 60 supresión de ruido a nivel alto global
- 61 altura tonal-sonoridad-tiempo de señal degradada
- 63 escalamiento local si $Y > X$
- 10 64 escalamiento hacia el nivel interno fijo
- 65 supresión de ruido a nivel alto global
- 70 espectro de referencia
- 72 espectro degradado
- 74 relación de altura tonal de ref. y deg. del marco corriente y +/-1 circundante
- 15 77 preprocesamiento
- 78 picos y caídos estrechas alisadas en el espectro FFT
- 79 tomar log del espectro, aplicar umbral para intensidad mínima
- 80 forma del espectro log total aplanado utilizando ventana de deslizamiento 83 ciclo de optimización
- 84 rango de factores de alabeo: relación de altura tonal $\min \leq 1 \leq \max$ relación altura tonal
- 20 85 espectro degradado de alabeo
- 88 aplicar preprocesamiento
- 89 computar correlación del espectro para receptáculos $< 1500\text{Hz}$
- 90 seguir el mejor factor de alabeo
- 93 espectro degradado de alabeo
- 25 94 aplicar preprocesamiento
- 95 computar correlación para espectro para receptáculo $< 3000\text{Hz}$
- 97 mantener el espectro degradado alabeado si la correlación es suficiente para reestablecer el original de otra manera
- 98 limitar el cambio del factor de alabeo de un marco al siguiente
- 100 regular ideal
- 30 101 regular degradado
- 104 distorsiones grandes ideales
- 105 distorsiones grandes degradadas
- 108 agregado ideal
- 109 agregado degradado
- 35 112 distorsiones grandes agregadas ideales
- 113 distorsiones grandes agregadas degradadas
- 116 selección regular de densidad de perturbación
- 117 selección de distorsiones grandes de densidad de perturbación
- 119 selección de densidad de perturbación agregada
- 40 120 selección de distorsiones grandes de densidad de perturbación agregada
- 121 entrada de PW_R_{total} a la función 123 de conmutación
- 122 entradas PW_R_{marco} a función 123 de conmutación
- 123 decisión de distorsión grande (conmutación)
- 125 factores de corrección para cantidades severas de distorsiones específicas
- 45 125` factores de corrección para cantidades severas de distorsiones específicas
- 127 nivel
- 127` nivel
- 128 repetición de marco
- 128` repetición de marco
- 50 129 timbre
- 129` timbre
- 130 aplanado espectral
- 130` aplanado espectral
- 131 contraste de ruido en periodo silente
- 55 131` contraste de ruido en periodo silente
- 133 ponderación de perturbación de pendiente de sonoridad
- 133` ponderación de perturbación de pendiente de sonoridad
- 134 sonoridad de la señal de referencia
- 134` sonoridad de la señal de referencia
- 60 136 alinear saltos
- 136` alinear saltos
- 137 recorte a degradación máxima
- 137` recorte a degradación máxima
- 138 varianza de perturbación
- 65 138` varianza de perturbación
- 140 saltos de sonoridad

- 140` saltos de sonoridad
- 142 densidad de $D_n^{(f)}$ de perturbación final
- 143 densidad de $DA_n^{(f)}$ de perturbación agregada final
- 145 integración de frecuencia L3
- 5 146 integración de acelerón L1
- 147 integración de tiempo L3
- 148 integración de frecuencia L5
- 149 integración de acelerón L1
- 150 integración de tiempo L1
- 10 153 integración de frecuencia L1
- 155 integración de acelerón L4
- 156 integración de tiempo L2
- 159 integración de frecuencia L1
- 160 integración de acelerón L1
- 15 161 integración de tiempo L2
- 170 mapeo a calificación MOS intermedia
- 171 MOS como indicador intermedio
- 175 compensaciones de escala MOS
- 176 calificaciones MOS bruta
- 20 180 mapeo a MOS-LQO
- 181 MOS LQO
- 185 intensidad sobre el tiempo para un tono sinusoidal corto
- 187 tono sinusoidal corto
- 188 umbral de enmascaramiento para un segundo tono sinusoidal corto
- 25 195 intensidad sobre frecuencia para tono sinusoidal corto
- 198 tono sinusoidal corto
- 199 hacer umbral para un segundo tono sinusoidal corto
- 205 intensidad sobre frecuencia y tiempo en grafica 3D
- 211 umbral de enmascaramiento utilizado como resistencia a la supresión que conduce a una representación interna
- 30 aguda
- 220 marcos de señal de referencia
- 222 determinar VOLUMEN
- 225 comparar VOLUMEN con UMBRAL
- 226 (PRIMER) UMBRAL
- 35 228 VOLUMEN > UMBRAL
- 230 FACTOR DE PONDERACIÓN = 1,0
- 231 VOLUMEN < UMBRAL
- 233 FACTOR DE PONDERACIÓN lineal dependiente de VOLUMEN
- 235 determinar valor para VALOR DE PONDERACIÓN
- 40 238 señal de diferencia perturbación
- 240 etapa de ponderación de perturbación
- 242 comparar VOLUMEN con SEGUNDO UMBRAL
- 243 SEGUNDO UMBRAL
- 245 VOLUMEN > SEGUNDO UMBRAL
- 45 246 FACTOR DE PONDERACIÓN lineal dependiente de VOLUMEN, por ejemplo:
 $VALOR DE PONDERACIÓN = (VOLUMEN - 2 \cdot UMBRAL + 1.0) \cdot 10^{-q}$
 Donde q puede ser igual a 0,3.
- 248 VOLUMEN < SEGUNDO UMBRAL
- 249 FACTOR DE PONDERACIÓN = 1,0
- 50

Reivindicaciones

- 5 1. Método para evaluar la inteligibilidad de una señal de voz degradada recibida de un sistema de transmisión de audio, al transportar a través de dicho sistema de transmisión de audio una señal de voz de referencia tal como para suministrar dicha señal de voz degradada, en donde el método comprende:
- 10 -muestrear dicha señal de voz de referencia en una pluralidad de marcos de señal de referencia y determinar para cada marco una representación de señal de referencia;
- 15 -muestrear dicha señal de voz degradada en una pluralidad de marcos de señal degradados y determinar para cada marco una representación de señal degradada;
- formar pares de marco al asociar dichos marcos de señal de referencia y dichos marcos de señal degradada el uno con el otro, y suministrar para cada par de marcos una función de diferencia que representa una diferencia entre dicho marco de señal degradada y dicho marco de señal de referencia asociado;
- el método se caracteriza por:
- 20 -compensar dicha función de diferencia para uno o más tipos de perturbación tal como suministrar para cada par de marcos una función de densidad de perturbación que se adapta a un modelo de percepción de auditorio humano;
- 25 -derivar desde dichas funciones de densidad de perturbación de una pluralidad de pares de marco un parámetro de calidad total, dicho parámetro de calidad es al menos significativo de dicha inteligibilidad de dicha señal de voz degradada;
- en donde, dicho método comprende además las etapas de:
- 30 determinar el valor de sonoridad para cada uno de dichos marcos de señal de referencia; y
- determinar el valor de ponderación dependiente de dicho valor de sonoridad de dicho marco de señal de referencia;
- donde dicha etapa de compensar dicha función de diferencia comprende una etapa de ponderar
- 35 dicha función de diferencia utilizando dicho valor de ponderación dependiente de sonoridad, para incorporar un impacto de perturbación sobre dicha inteligibilidad de dicha señal de voz degradada en dicha evaluación.
- 40 2. Método de acuerdo a la reivindicación 1, en donde para determinar dicho valor de ponderación dependientes de sonoridad, dicho método comprende la etapa de comparar dicho valor de sonoridad con un primer umbral, y hacer dicho valor de ponderación dependiente si el valor de sonoridad excede dicho primer umbral.
- 45 3. Método de acuerdo a la reivindicación 2, que comprende además fijar dicho valor de ponderación a un valor máximo cuando dicho valor de sonoridad para dicho marco de señal de referencia excede dicho primer umbral.
- 50 4. Método de acuerdo a cualquiera de las reivindicaciones 2 o 3, en donde dicho valor de ponderación se hace más pequeño que un valor máximo y dependiente de dicho valor de sonoridad cuando dicho valor de sonoridad para dicho marco de señal de referencia es más pequeño que dicho primer umbral.
- 55 5. Método de acuerdo a la reivindicación 4, en donde dicho valor de ponderación se hace igual a dicho valor de sonoridad cuando dicho valor de sonoridad para dicho marco de señal de referencia es más pequeño que dicho primer umbral.
- 60 6. Método de acuerdo a una cualquiera de las reivindicaciones previas, en donde para determinar dicho valor de ponderación dependiente de sonoridad, el método comprende una etapa de comparar el valor de sonoridad con un segundo umbral, y en donde el valor de ponderación se hace más pequeño que un valor máximo cuando el valor de sonoridad para el marco de señal de referencia excede el segundo umbral.
7. Método de acuerdo a la reivindicación 6, en donde dicho valor de sonoridad para dicho marco de señal de referencia excede el segundo umbral, el valor de ponderación se hace inversamente dependiente de una cantidad con la cual el valor de sonoridad excede el segundo umbral.
- 65 8. Método de acuerdo a una cualquiera de las reivindicaciones previas, en donde dicho valor de sonoridad se determina de una manera dependiente de frecuencia, y en donde dicho valor de ponderación se hace dependiente de dicho valor de sonoridad dependiente de frecuencia.

9. Método de acuerdo a una cualquiera de las reivindicaciones previas, en donde dicho método para evaluar la inteligibilidad de dicha señal de voz degradada se basa en un algoritmo de evaluación de calidad de escucha objetivo perceptivo (POLQA).

5 10. Producto de programa de ordenador que comprende un código ejecutable de ordenador para efectuar un método de acuerdo a una cualquiera de las reivindicaciones previas cuando se ejecuta mediante un ordenador.

11. Aparato para efectuar un método de acuerdo a una cualquiera de las reivindicaciones 1-9, para evaluar la inteligibilidad de una señal de voz degradada, que comprende:

10

-una unidad de recepción para recibir dicha señal de voz degradada de un sistema de transmisión de audio que transporta una señal de voz de referencia, y para recibir dicha señal de voz de referencia;

15

-una unidad de muestreo para muestrear dicha señal de voz de referencia en una pluralidad de marcos de señal de referencia, y para muestrear dicha señal de voz degradada en una pluralidad de marcos de señal degradados;

20

-una unidad de procesamiento para determinar para cada marco de señal de referencia una representación de señal de referencia, y para determinar para cada marco de señal degradado una representación de señal degradada;

-una unidad de comparación para formar pares de marco al asociar dichos marcos de señal de referencia y dichos marcos de señal degradados el uno con el otro, y

25

-para suministrar para cada par de marcos una función de diferencia que representa una diferencia entre dicho marco de señal degradado y de referencia;

el aparato se caracteriza por

30

-una unidad compensadora para compensar dicha función de diferencia para uno o más tipos de perturbación tal como suministrar para cada par de marcos una función de densidad de perturbación que se adapta al modelo de percepción de auditorio humano; y

35

-dicha unidad y procesamiento está dispuesta además para derivar de dichas funciones de densidad de perturbación de una pluralidad de pares de marco un parámetro de calidad total que es al menos indicativo de dicha inteligibilidad de dicha señal de voz degradada;

en donde, dicha unidad de procesamiento esta además dispuesta para:

40

-determinar el valor de sonoridad para cada uno de dichos marcos de señal de referencia; y para

-determinar un valor de ponderación dependiente de dicho valor de sonoridad de dicho marco de señal de referencia; en donde dicha unidad compensadora se conecta a dicha unidad de procesamiento, y esta además dispuesta para ponderar dicha función de diferencia utilizando dicho valor de ponderación dependiente de sonoridad recibido de dicha unidad de procesamiento.

45

12. Aparato de acuerdo a la reivindicación 11, en donde dicha unidad de procesamiento se dispone además para comparar dicho valor de sonoridad con un primer umbral, y hacer dicho valor de ponderación dependiente de si dicho valor de sonoridad excede dicho primer umbral.

50

13. Aparato de acuerdo con la reivindicación 12, en donde dicha unidad de procesamiento se dispone además para fijar dicho valor de ponderación a un valor máximo cuando dicho valor de sonoridad de dicho marco de señal de referencia excede dicho primer umbral.

55

14. Aparato de acuerdo a la reivindicación 12 o 13, en donde dicha unidad de procesamiento está dispuesta además para hacer dicho valor de ponderación igual a dicho valor de sonoridad cuando dicho valor de sonoridad para dicho marco de señal de referencia es más pequeño que dicho primer umbral.

60

15. Aparato de acuerdo a cualquiera de las reivindicaciones 11- 14, en donde la unidad de procesamiento está dispuesta además para comparar dicho valor de sonoridad con un segundo umbral, y hacer el valor de ponderación más pequeño que un valor máximo cuando el valor de sonoridad para el marco de señal de referencia excede el segundo umbral.

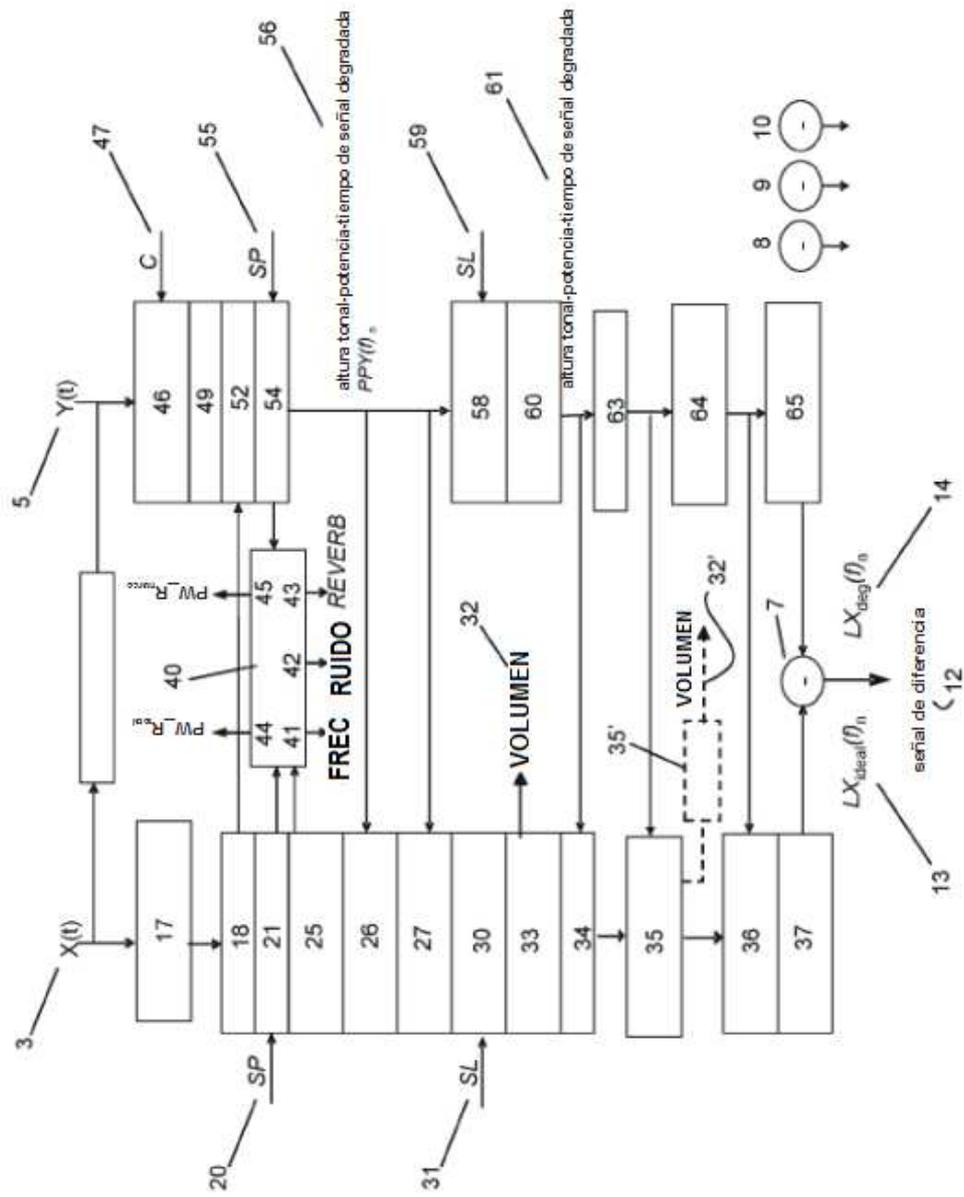


Fig. 1

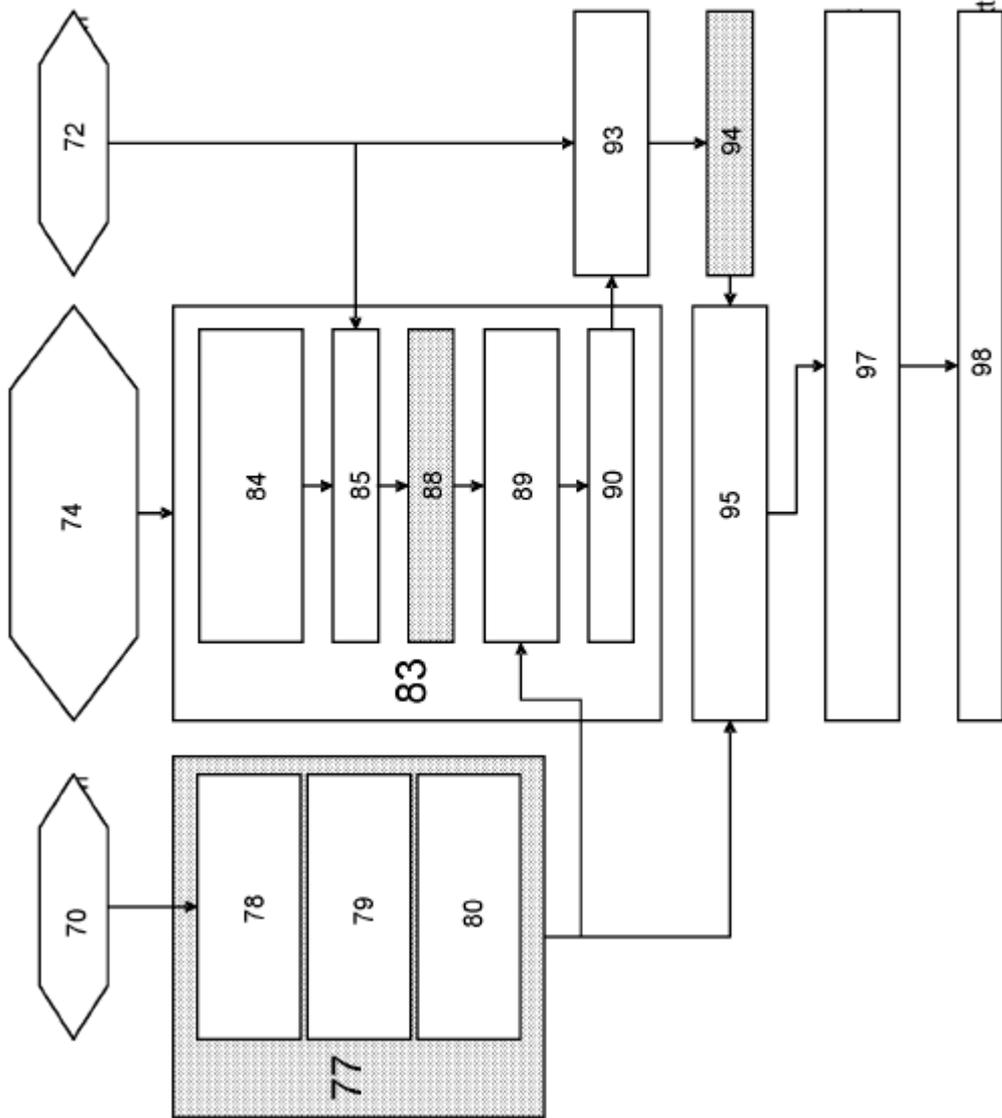


Fig. 2

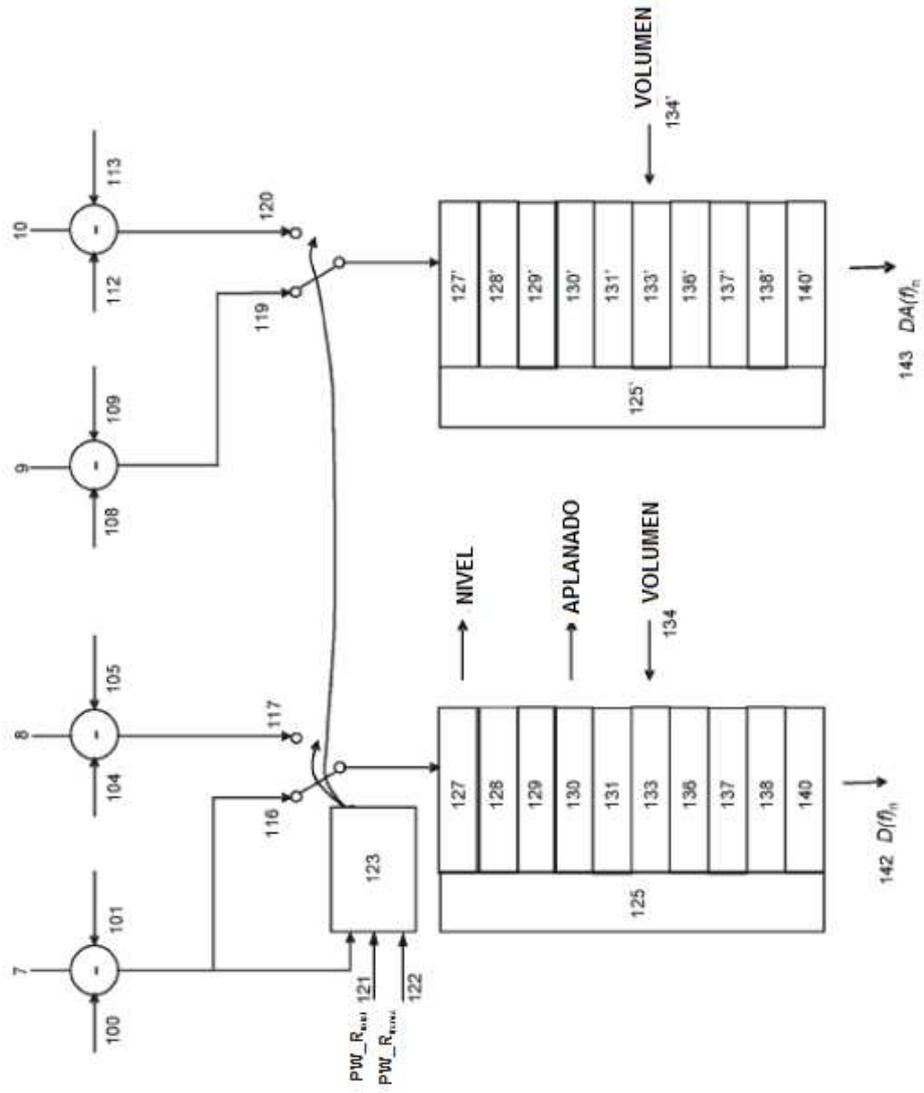


Fig. 3

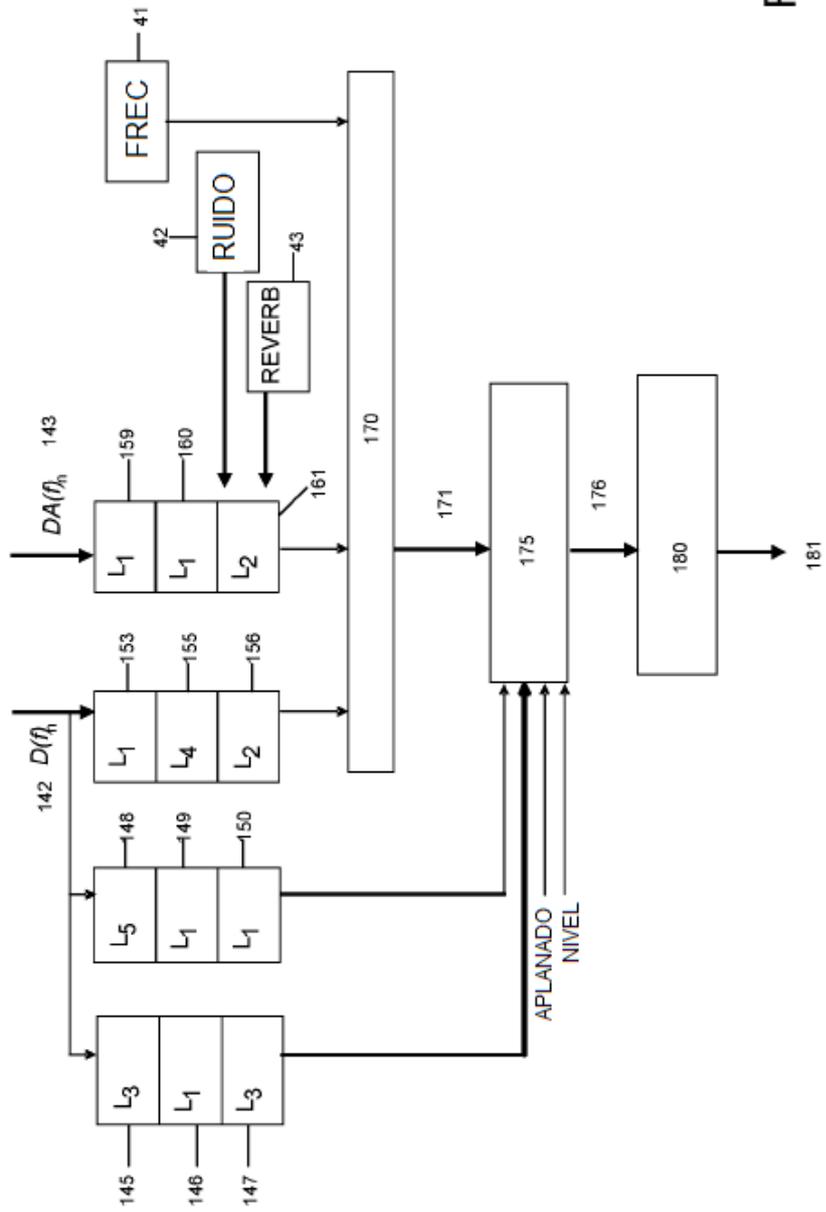
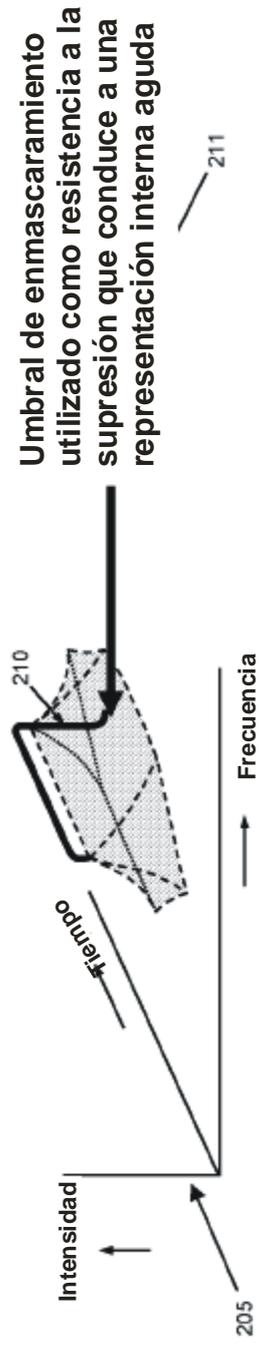
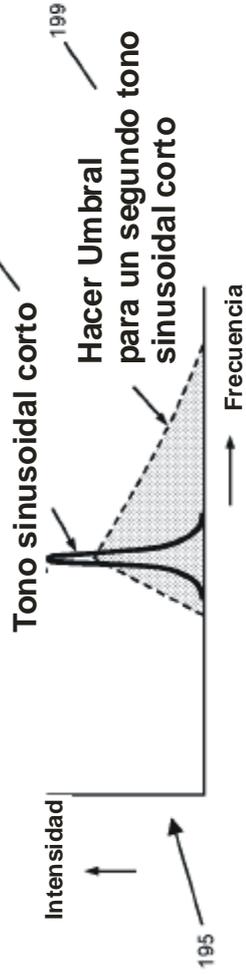
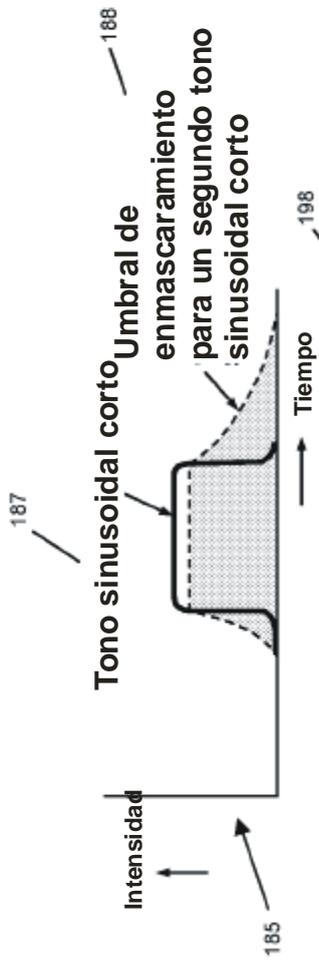


Fig. 4



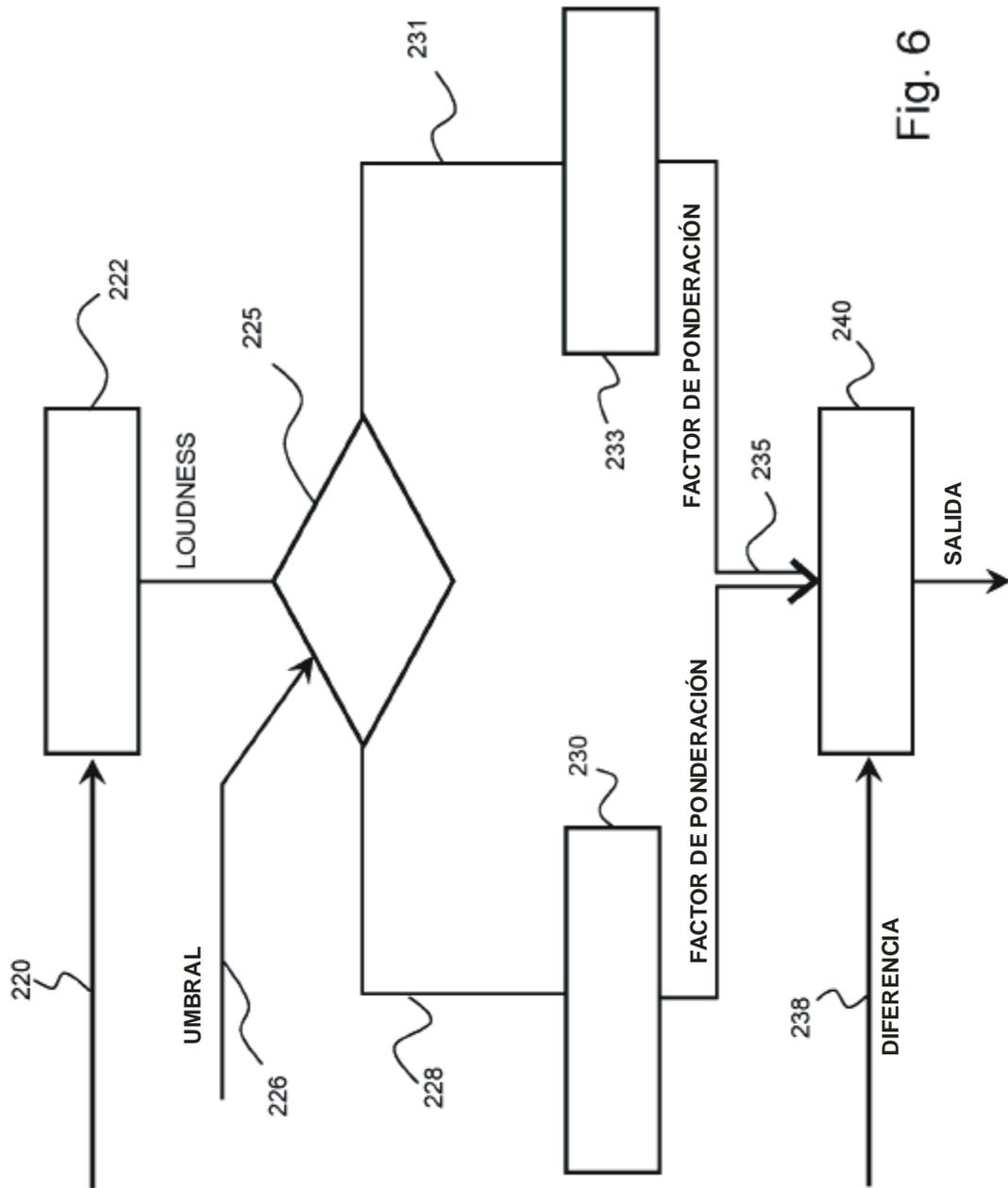


Fig. 6

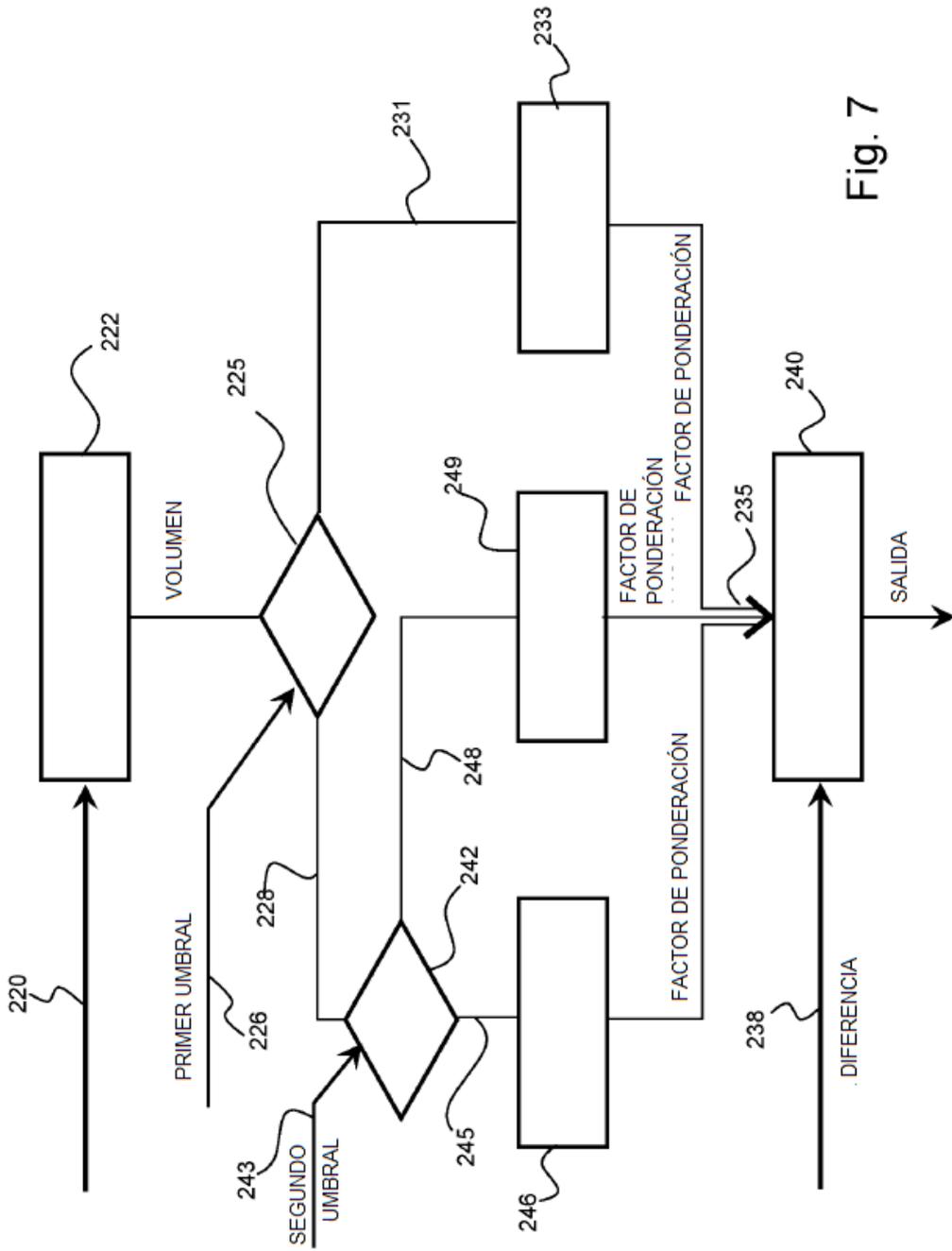


Fig. 7