



①9



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

①1 Número de publicación: **2 298 725**

⑤1 Int. Cl.:
G10L 19/00 (2006.01)

①2

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑧6 Número de solicitud europea: **04714792 .1**

⑧6 Fecha de presentación : **26.02.2004**

⑧7 Número de publicación de la solicitud: **1611571**

⑧7 Fecha de publicación de la solicitud: **04.01.2006**

⑤4 Título: **Procedimiento y sistema para predicción de calidad de voz de un sistema de transmisión de audio.**

③0 Prioridad: **31.03.2003 EP 03075949**

④5 Fecha de publicación de la mención BOPI:
16.05.2008

④5 Fecha de la publicación del folleto de la patente:
16.05.2008

⑦3 Titular/es: **Koninklijke KPN N.V.**
Stationsplein 7
9726 AE Groningen, NL

⑦2 Inventor/es: **Beerends, John, Gerard y**
Van den Homberg, Marc, Jan, Christiaan

⑦4 Agente: **Arias Sanz, Juan**

ES 2 298 725 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

Procedimiento y sistema para predicción de calidad de voz de un sistema de transmisión de audio.

5 Campo de la invención

La presente invención se refiere a un procedimiento y a un sistema para medir la calidad de transmisión de un sistema en prueba, siendo procesadas y comparadas mutuamente una señal de entrada introducida en el sistema en prueba y una señal de salida que resulta del sistema en prueba.

10 Técnica anterior

Tal procedimiento y sistema se conocen a partir de la Recomendación P.862 de la ITU-T (Sector de Normalización de las Telecomunicaciones de la Unión Internacional de Telecomunicaciones), “Calidad de transmisión telefónica, instalaciones de telefónicas, redes de área local-Procedimientos para valoración objetiva y subjetiva de calidad-Evaluación perceptiva de calidad de voz (PESQ), un procedimiento objetivo para valoración de calidad de voz de extremo a extremo de las redes telefónicas de banda estrecha y códecs de voz”, ITU-T 02.2001[8].

También, el artículo de J. Beerends *et al.* “Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for end-to-end Speech Quality Assessment Part II Psychoacoustic Model”, J. Audio Eng. Soc., Vol. 50, nº 10, octubre de 2002, describe tal procedimiento y sistema [9].

En el procedimiento y sistema P.862 está presente una desventaja, ya que el procedimiento y sistema aplicados en la medición de calidad estándar no compensan correctamente las grandes variaciones de respuesta de frecuencia del sistema en prueba ni las grandes diferencias de potencia local entre la señal de entrada y la de salida. Esto puede tener como resultado una mala correlación entre las puntuaciones de la calidad de voz percibida tal como la proporciona el procedimiento y sistema y la calidad de voz percibida tal como la evalúan las personas de prueba.

Resumen de la invención

La presente invención trata de proporcionar una mejora de la correlación entre la calidad de voz percibida tal como es medida por el procedimiento y sistema P.862 y la calidad de voz real tal como es percibida por las personas de prueba.

Según la presente invención, se proporciona un procedimiento tal como se define en la Reivindicación 1.

La presente invención está basada en la comprensión de que en ciertas circunstancias (presencia de ruido, presencia de grandes desviaciones de respuesta de frecuencia en el sistema bajo prueba) el procedimiento estandarizado existente no mide correctamente la calidad de voz percibida.

Si se calcula una compensación de frecuencia en presencia de ruido surgirá una estimación errónea de la función de respuesta de frecuencia en regiones de frecuencia donde hay poca energía. Si se calcula un factor de escala temporal local sobre una señal que ha pasado por un sistema que muestra grandes desviaciones en la respuesta de frecuencia el factor de escala local no puede calcularse correctamente. Ambos efectos tienen que ser calculados correctamente para poder predecir la calidad percibida subjetivamente de las señales de voz.

Puede implementarse una corrección según la presente invención sustituyendo el cálculo de una compensación de frecuencia lineal y el cálculo de un factor de escala de potencia local por un cálculo iterativo de la compensación de frecuencia y el factor de escala local. Calculando en primer lugar una estimación aproximada de la compensación de frecuencia necesaria, es decir, no compensando a la cantidad que normalmente se llevaría a cabo, se obtiene una señal en el tiempo a partir de la cual pueden hacerse estimaciones mejores con respecto al factor de escala temporal local que es necesario para predecir correctamente la calidad percibida final. Después de este cálculo de escala local se obtiene una señal de tiempo a partir de la cual puede hacerse una estimación mejor para la compensación de frecuencia necesaria.

En términos generales, esto mejorará el resultado de la predicción de calidad de voz utilizando el procedimiento según la invención. Además, en otras circunstancias, esta adaptación del procedimiento y sistema estandarizados no tendrá una influencia negativa en otras circunstancias.

El cálculo del factor de escala de potencia local puede implementarse como se describe en la Recomendación P.862 de la ITU-T o en la solicitud de patente EP1343145.

En una realización ventajosa particular, el bucle iterativo comprende un cálculo de una primera compensación de frecuencia lineal parcial y la aplicación de la primera compensación de frecuencia lineal parcial a la densidad de potencia tonal de la señal de entrada, se guido de un cálculo de un factor de escala de potencia local y la aplicación del factor de escala de potencia local a la densidad de potencia tonal de la señal de salida, seguido por un cálculo de una segunda compensación de frecuencia lineal parcial y la aplicación de la compensación de frecuencia lineal a la densidad de potencia tonal parcialmente compensada de la señal de entrada. En una realización adicional, la aplicación

de las compensaciones a las densidades de potencia tonal de la señal de entrada y de salida están intercambiadas, es decir, la primera y segunda compensaciones de frecuencia lineal parcial se aplican a la densidad de potencia tonal de la señal de salida, y el factor de escala de potencia local se aplica a la densidad de potencia tonal de la señal de entrada. Estas realizaciones sólo requieren muy pocos cambios en el procedimiento P.862 estandarizado existente, mientras que mejoran su resultado.

En una realización adicional, la compensación de frecuencia lineal parcial es una primera estimación que es inferior a la compensación de frecuencia lineal que se usaría para la evaluación correcta de la distorsión lineal (como se prescribe, por ejemplo, en la Recomendación P.862 de la ITU-T), por ejemplo el 50% de la corrección de amplitud de la compensación de frecuencia lineal normal. Esta compensación parcial también puede llevarse a cabo dependiente de la frecuencia, por ejemplo teniendo intervalos de frecuencia limitados a lo largo de los cuales se lleva a cabo una compensación parcial mayor que a lo largo de otros intervalos de frecuencia. Por ejemplo, se pueden compensar sólo compensaciones de respuesta de frecuencia como las encontradas con técnicas de micrófono cercano que tienen como resultado una intensificación de baja frecuencia por debajo de aproximadamente 500 Hz.

En un segundo aspecto, la presente invención se refiere a un sistema para medir la calidad de transmisión de un sistema de transmisión de audio tal como se define en la Reivindicación 6. Este sistema, y los sistemas tales como se definen en las reivindicaciones dependientes, proporcionan ventajas comparables a las ventajas del procedimiento tal como se describe anteriormente.

Breve descripción de los dibujos

La presente invención se tratará más detalladamente más adelante, usando varias realizaciones ejemplares, con referencia a los dibujos adjuntos, en los que

la Fig. 1 muestra esquemáticamente un sistema PESQ de la técnica anterior, descrito en la Recomendación P.862 de la ITU-T;

la Fig. 2 muestra una vista de una implementación de modelo perceptivo como el usado en el sistema PESQ de la Fig. 1;

la Fig. 3 muestra la misma implementación PESQ que la Fig. 2 que, sin embargo, está modificada para que sea apropiada para ejecutar el procedimiento según una realización de la presente invención.

Descripción detallada de realizaciones ejemplares

La Fig. 1 muestra esquemáticamente un esquema conocido de una aplicación de una técnica de medición objetiva que está basada en un modelo de percepción y cognición auditiva humana, y que sigue la Recomendación P.862 [8] de la ITU-T, para estimar la calidad perceptiva de enlaces o codecs de voz. El acrónimo usado para esta técnica o dispositivo es PESQ (evaluación perceptiva de calidad de voz). Comprende un sistema o red de telecomunicaciones en prueba 10, denominado en lo sucesivo sistema 10 por brevedad, y un dispositivo de medición de calidad 11 para el análisis perceptivo de las señales de voz ofrecidas. Se usa una señal de voz $X_0(t)$, por una parte, como una señal de entrada del sistema 10 y, por otra parte, como una primera señal de entrada $X(t)$ del dispositivo 11. Una señal de salida $Y(t)$ del sistema 10, que de hecho es la señal de voz $X_0(t)$ afectada por el sistema 10, se usa como segunda señal de entrada del dispositivo 11. Una señal de salida Q del dispositivo 11 representa una estimación de la calidad perceptiva del enlace de voz a través del sistema 10. Como el extremo de entrada y el extremo de salida de un enlace de voz, particularmente en el caso de que se ejecute a través de una red de telecomunicaciones, son remotos, para las señales de entrada del dispositivo de medición de calidad 11 se hace uso en la mayoría de los casos de señales de voz $X(t)$ almacenadas en bases de datos. Aquí, como es habitual, se entiende que señal de voz significa cada sonido perceptible básicamente por el oído humano, como voz y tonos. Por supuesto, el sistema en prueba 10 también puede ser un sistema de simulación, que simula una red de telecomunicaciones. El dispositivo 11 lleva a cabo una etapa de proceso principal que comprende sucesivamente, en una sección de preproceso 11.1, una etapa de preproceso llevada a cabo por el medio de preproceso 12, en una sección de proceso 11.2, una etapa de proceso adicional llevada a cabo por un primer y un segundo medio de proceso de señales 13 y 14, y, en una sección de combinación de señales 11.3, una etapa de proceso de señales combinadas llevada a cabo por un medio diferenciador de señales 15 y un medio de modelización 16. En la etapa de preproceso las señales $X(t)$ e $Y(t)$ son preparadas para la etapa de proceso adicional en los medios 13 y 14, incluyendo el preproceso operaciones de escala del nivel de potencia y alineación de tiempo. La etapa de proceso adicional implica transformación de la señal de salida (degradada) $Y(t)$ y la señal de referencia $X(t)$ en señales de representación $R(Y)$ y $R(X)$ según un modelo de percepción psicofísica del sistema auditivo humano. Durante la etapa de proceso de señales combinadas se determina una señal diferencial o de perturbación D por el medio diferenciador 15 a partir de dichas señales de representación, que luego es procesada por el medio de modelización 16 de acuerdo con un modelo cognitivo, en el que han sido modeladas ciertas propiedades de evaluadores humanos, para obtener la señal de calidad Q .

En una primera etapa ejecutada por el sistema PESQ se computa una serie de retardos entre la entrada original y la salida degradada, uno para cada intervalo de tiempo para el que el retardo es significativamente diferente del intervalo de tiempo previo. Para cada uno de estos intervalos se calcula un punto de partida y de parada correspondiente. El algoritmo de alineación está basado en el principio de comparar la confianza de tener dos retardos en un cierto

intervalo de tiempo con la confianza de tener un único retardo para ese intervalo. El algoritmo puede ocuparse de cambios de retardo tanto durante los silencios como durante las partes de voz activa.

Basándose en el conjunto de retardos que se encuentran el sistema PESQ compara la señal (de entrada) original con la salida degradada alineada del dispositivo en prueba usando un modelo perceptivo. La clave para este procedimiento es la transformación tanto de la señal original como de la señal degradada en representaciones internas (LX, LY), análogas a la representación psicofísica de señales de audio en el sistema auditivo humano, teniendo en cuenta la frecuencia perceptiva (Bark) y la sonoridad (Sone). Esto se logra en varias fases: alineación de tiempo, alineación de nivel a un nivel de escucha calibrado, transformación de tiempo-frecuencia, distorsión de frecuencia, y cambio de escala de sonoridad compresiva.

La representación interna se procesa para tener en cuenta efectos como las variaciones locales de ganancia y el filtrado lineal que pueden -si no son muy severos- tener poca importancia perceptiva. Esto se logra limitando la cantidad de compensación y haciendo que la compensación se retrase detrás del efecto. Estas diferencias menores, de estado estacionario entre la original y la degradada son compensadas. Los efectos más severos, o las variaciones rápidas, sólo son compensados parcialmente de manera que queda un efecto residual y contribuye a la perturbación perceptiva total. Esto permite que se use un pequeño número de indicadores de calidad para modelar todos los efectos subjetivos. En el sistema PESQ, se computan dos parámetros de error en el modelo cognitivo; estos se combinan para dar una MOS (puntuación de opinión media) de calidad de escucha objetiva. Las ideas básicas usadas en el sistema PESQ se describen en las referencias [1] a [5] de la bibliografía.

El modelo perceptivo en el sistema PESQ de la técnica anterior

En la Fig. 2 se ilustra una parte de una implementación del dispositivo 11 (es decir, la parte del modelo perceptivo), que comprende en esencia el primer y segundo medios de proceso de señales 13 y 14, y el medio diferenciador 15 como se describieron anteriormente.

El modelo perceptivo de un sistema PESQ, mostrado en la figura 2, se usa para calcular una distancia entre la señal de voz original y la degradada ("Puntuación PESQ"). Este puede pasarse por una función monótona para obtener una predicción de una MOS subjetiva para una prueba subjetiva dada. La puntuación PESQ es transformada a una escala de tipo MOS.

Umbral de audición absoluta

El umbral de audición absoluta $P_0(f)$ se interpola para obtener los valores en el centro de las bandas Bark que se usan. Estos valores se almacenan en una matriz y se usan en la fórmula de sonoridad de Zwicker.

Los factores de escala de potencia y sonoridad

Existen constantes de ganancia arbitrarias que siguen la FFT para análisis de tiempo-frecuencia y en el cálculo de sonoridad sólo son para calibrar el sistema.

Filtrado IRS en recepción

Si se supone que las pruebas de escucha se llevaron a cabo usando IRS (sistema intermedio de referencia) en recepción o un IRS en recepción modificado característico en el auricular, se aplica el filtrado necesario a las señales de voz en el preproceso (sección 11.1 en la Fig. 1), obteniendo como resultado las señales $X_{\text{IRSS}}(t)$ e $Y_{\text{IRSS}}(t)$.

Cálculo del intervalo de tiempo de voz activa

Si el archivo de voz original y degradada empieza o termina con grandes intervalos de silencio, esto podría influir en el cálculo de ciertos valores de distorsión media a lo largo de los archivos. Por lo tanto, se hace una estimación de las partes silenciosas al principio y al final de estos archivos.

FFT o descomposición de tiempo-frecuencia a corto plazo

El oído humano realiza una transformación de tiempo-frecuencia. En el sistema PESQ esta se implementa mediante una FFT a corto plazo con superposición entre ventanas (marcos) temporales sucesivas. Los espectros de potencia -la suma de los cuadrados de las partes real e imaginaria de las componentes complejas de la FFT- se almacenan en matrices de valor real separadas para las señales original y degradada. La información de fase dentro de una ventana de Hanning individual se desecha en el sistema PESQ y todos los cálculos están basados sólo en las representaciones de potencia $PX_{\text{WIRSS}}(f)_n$ y $PY_{\text{WIRSS}}(f)_n$. Los puntos de partida de las ventanas en la señal degradada se desplazan sobre el retardo. El eje de tiempo de la señal de voz original se deja como está. Si el retardo aumenta, se omiten del proceso partes de la señal degradada, mientras que para disminuciones del retardo se repiten partes.

Cálculo de las densidades de potencia tonal

La escala de Bark refleja que, a bajas frecuencias, el sistema auditivo humano tiene una resolución de frecuencia más precisa que a altas frecuencias. Esto se implementa confinando las bandas de la FFT y sumando las potencias correspondientes de las bandas de la FFT con una normalización de las partes sumadas. La función de distorsión que transforma la escala de frecuencia en hercios a la escala de tono en Bark no sigue exactamente los valores dados en la bibliografía. Las señales resultantes se conocen como las densidades de potencia tonal $PPX_{WIRSS}(f)_n$ y $PPY_{WIRSS}(f)_n$.

Compensación de la densidad de potencia tonal original (compensación de respuesta de frecuencia lineal)

Para tratar el filtrado en el sistema en prueba, se promedian a lo largo del tiempo el espectro de potencia de las densidades de potencia tonal original y degradada. Este promedio se calcula a lo largo de tramas de voz activas que usan sólo célula de tiempo-frecuencia cuya potencia está una cierta fracción por encima del umbral absoluto de audición. Por cada Bark confinado modificado, se calcula un factor de compensación parcial a partir de la relación del espectro degradado al espectro original. La densidad de potencia tonal original $PPX_{WIRSS}(f)_n$ de cada trama n es multiplicada luego por su factor de compensación parcial para igualar la señal original a la degradada. Esto tiene como resultado una densidad de potencia tonal original filtrada inversamente $PPX'_{WIRSS}(f)_n$. Esta compensación parcial se usa porque el filtrado severo puede ser perturbador para el oyente. La compensación se lleva a cabo sobre la señal original porque la señal degradada es la que es evaluada por los sujetos en un experimento ACR.

Compensación de la densidad de potencia tonal distorsionada (compensación de ganancia variable en el tiempo)

Las variaciones de ganancia a corto plazo son compensadas parcialmente procesando las densidades de potencia tonal trama por trama (es decir, compensación local). Para las densidades de potencia tonal original y degradada, se computa en cada trama n la suma de todos los valores que exceden el umbral absoluto de audición. Se calcula la relación de la potencia en los archivos originales y degradados y se limita a un intervalo predeterminado. A esta relación se aplica un filtro de paso bajo de primer orden (a lo largo del eje temporal). La densidad de potencia tonal distorsionada en cada trama, n , se multiplica luego por esta relación, teniendo como resultado la densidad de potencia tonal distorsionada de ganancia compensada parcialmente $PPY'_{WIRSS}(f)_n$.

Esta compensación parcial o cálculo del factor de escala local puede implementarse usando la realización descrita en la solicitud de patente europea no publicada previamente pendiente de tramitación 02075973.4 del solicitante, que se incorpora en este documento por referencia (véase específicamente la Fig. 3).

Cálculo de las densidades de sonoridad

Después de la compensación por filtrado y variaciones de ganancia a corto plazo, las densidades de potencia tonal original y degradada son transformadas a una escala de sonoridad Sone usando la ley de Zwicker [7].

$$LX(f)_n = S_1 \cdot \left(\frac{P_0(f)}{0,5} \right)^\gamma \cdot \left[\left(0,5 + 0,5 \cdot \frac{PPX'_{WIRSS}(f)_n}{P_0(f)} \right)^\gamma - 1 \right]$$

con $P_0(f)$ el umbral absoluto y S_1 el factor de escala de sonoridad.

Por encima de 4 Bark, la potencia de Zwicker, γ , es 0,23, el valor dado en la bibliografía. Por debajo de 4 Bark, la potencia de Zwicker se incrementa ligeramente para tener en cuenta el denominado efecto de reclutamiento. Las matrices bidimensionales resultantes $LX(f)_n$ y $LY(f)_n$ se denominan densidades de sonoridad.

Cálculo de la densidad de perturbación

Se computa la diferencia señalada entre la densidad de sonoridad distorsionada y original. Cuando esta diferencia es positiva, se han añadido componentes como el ruido. Cuando esta diferencia es negativa, se han omitido componentes de la señal original. Esta diferencia se denomina densidad de perturbación bruta.

Se computa el mínimo de la densidad de sonoridad original y degradada para cada célula de tiempo-frecuencia. Estos mínimos se multiplican por 0,25. La matriz bidimensional correspondiente se denomina matriz de máscara. Se aplican las siguientes reglas en cada célula de tiempo-frecuencia:

- Si la densidad de perturbación bruta es positiva y mayor que el valor de máscara, el valor de máscara se resta de la perturbación bruta.
- Si la densidad de perturbación bruta está entre más y menos la magnitud del valor de máscara la densidad de perturbación se establece como cero.

- Si la densidad de perturbación bruta es más negativa que menos el valor de máscara, el valor de máscara se suma a la densidad de perturbación bruta.

El efecto neto es que las densidades de perturbación bruta tienden a desplazarse a cero. Esto representa una zona muerta antes de que una célula de tiempo-frecuencia real sea percibida como distorsionada. Lo anterior modela el proceso de pequeñas diferencias que son inaudibles en presencia de señales de gran sonoridad (enmascaramiento) en cada célula de tiempo-frecuencia. El resultado es una densidad de perturbación como función del tiempo (número de ventana n) y la frecuencia, $D(f)_n$.

Esta sustracción perceptiva de las densidades de sonoridad $LX(f)_n$ y $LY(f)_n$ que tiene como resultado la densidad de perturbación $D(f)_n$ puede implementarse como se describe con referencia a la Fig. 4 de la solicitud de patente europea no publicada previamente pendiente de tramitación 02075973.4 del solicitante, que se incorpora en este documento por referencia.

15 *Multiplicación célula a célula con un factor de asimetría*

El efecto de asimetría está causado por el hecho de que cuando un códec distorsiona la señal de entrada en general será muy difícil introducir una nueva componente de tiempo-frecuencia que se integre con la señal de entrada, y la señal de salida resultante estará descompuesta así en dos percepciones diferentes, la señal de entrada y la distorsión, conduciendo a distorsión claramente audible [2]. Cuando el códec omite una componente de tiempo-frecuencia la señal de salida resultante no puede ser descompuesta de la misma manera y la distorsión es menos inaceptable. Este efecto se modela calculando una densidad de perturbación asimétrica $DA(f)_n$ por trama mediante multiplicación de la densidad de perturbación $D(f)_n$ por un factor de asimetría. Este factor de asimetría es igual a la relación de las densidades de potencia tonal distorsionada y original elevada a la potencia de 1,2. Si el factor de asimetría es menor que 3 se establece como cero. Si excede de 12 se fija a ese valor. De este modo sólo quedan aquellas células de tiempo-frecuencia, como valores no nulos, para las que la densidad de potencia tonal degradada excedió la densidad de potencia tonal original.

30 *Agregación de las densidades de perturbación*

La densidad de perturbación $D(f)_n$ y la densidad de perturbación asimétrica $DA(f)_n$ se integran (se suman) a lo largo del eje de frecuencia usando dos normas L_p diferentes y una ponderación sobre tramas suaves (que tienen baja sonoridad):

$$D_n = M_n \sqrt[3]{\sum_{f=1, \dots, \text{Número de bandas de Bark}} (D(f)_n | W_f)^3}$$

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

con M_n un factor multiplicador, $1/(\text{potencia de la trama original más una constante})^{0.04}$, que tiene como resultado un énfasis de las perturbaciones que se producen durante los silencios en el fragmento de voz original, y W_f una serie de constantes proporcionales a la anchura de los confines Bark modificados. Después de esta multiplicación los valores de perturbación de trama están limitados a un máximo de 45. Estos valores agregados, D_n y DA_n , se denominan perturbaciones de trama.

Si la señal distorsionada contiene una disminución del retardo mayor que 16 ms (media ventana) se modifica la estrategia de repetición. Se descubrió que es mejor ignorar las perturbaciones de trama durante tales eventos en el cálculo de la calidad de voz objetiva. Como consecuencia se ponen a cero las perturbaciones de trama cuando esto ocurre. Las perturbaciones de trama resultantes se denominan D'_n y DA'_n .

55 *Realineación de intervalos incorrectos*

Las tramas consecutivas con una perturbación de trama superior a un umbral se denominan intervalos incorrectos. En una minoría de los casos la medida objetiva predice grandes distorsiones a lo largo de un número mínimo de tramas erróneas debido a retardos de tiempo incorrectos observados por el preproceso. Para los denominados intervalos incorrectos se estima un nuevo valor de retardo maximizando la correlación cruzada entre la señal original absoluta y la señal degradada absoluta ajustado según los retardos observados por el preproceso. Cuando la correlación cruzada máxima está por debajo de un umbral, se concluye que el intervalo está comparando ruido con ruido y el intervalo ya no se denomina incorrecto, y se detiene el proceso para ese intervalo. Si no, se vuelve a computar la perturbación de trama para las tramas durante los intervalos incorrectos y, si es menor, sustituye a la perturbación de trama original. El resultado son las perturbaciones de trama finales D''_n y DA''_n que se usan para calcular la calidad percibida.

Agregación de la perturbación dentro de intervalos de fracción de segundo

- A continuación, los valores de perturbación de trama y los valores de perturbación asimétrica de trama son agregados sobre intervalos de fracción de segundo de 20 tramas (teniendo en cuenta la superposición de tramas: aproximadamente 320 ms) usando normas L_6 , un valor p más alto que en la agregación a lo largo de la longitud del archivo de voz. Estos intervalos también se superponen el 50 por ciento y no se usa función ventana.

Agregación de la perturbación a lo largo de la duración de la señal

- Los valores de perturbación de fracción de segundo y los valores de perturbación asimétrica de fracción de segundo son agregados a lo largo del intervalo activo de los archivos de voz (las tramas correspondientes) usando ahora normas L_2 . El valor más alto de p para la agregación dentro de intervalos de fracción de segundo comparado con el valor más bajo de p de la agregación a lo largo del archivo de voz es debido al hecho de que cuando partes de las fracciones de segundo son distorsionadas esa fracción de segundo pierde significado, mientras que si una primera frase en un archivo de voz está distorsionada la calidad de las otras frases permanece intacta.

Cálculo de la puntuación PESQ

- La puntuación PESQ final es una combinación lineal del valor medio de la perturbación y el valor medio de la perturbación asimétrica.

- El procedimiento PESQ anteriormente descrito (tal como se prescribe en la Recomendación P.862 de la ITU-T) tiene la desventaja de que no puede tratar correctamente señales de voz con grandes diferencias en las variaciones de respuesta de frecuencia. La compensación de variación de respuesta de frecuencia y la compensación de escala de potencia local están siendo calculadas incorrectamente, teniendo como resultado un cálculo erróneo de la calidad de voz de un sistema 10.

- La presente invención está basada en la comprensión de que si se calcula una compensación de frecuencia en presencia de ruido se originará una estimación errónea de la función de respuesta de frecuencia en regiones de frecuencia donde hay poca energía. Si se calcula un factor de escala temporal local sobre una señal que ha pasado por el sistema que muestra grandes desviaciones en la respuesta de frecuencia no puede calcularse correctamente el factor de escala local. Ambos efectos tienen que ser calculados correctamente para poder predecir la calidad percibida subjetivamente de las señales de voz.

- En la Fig. 3 se ilustra una realización ventajosa particular de la parte del modelo perceptivo del procedimiento PESQ, que corresponde a la ilustración de la Fig. 2. Sin embargo, el cálculo de la compensación de frecuencia lineal y el cálculo del factor de escala de potencia local son diferentes.

- El cálculo de la compensación de respuesta de frecuencia lineal y el cálculo del factor de escala de potencia local están puestos en un bucle iterativo. En primer lugar, se calcula una estimación aproximada de la compensación de frecuencia necesaria. A continuación se calcula una compensación de frecuencia lineal parcial que es inferior a la compensación de frecuencia lineal que se usaría para evaluación correcta de la distorsión lineal, por ejemplo el 50% de la corrección de amplitud de la compensación de frecuencia lineal normal. Esta compensación parcial también puede llevarse a cabo teniendo intervalos de frecuencia limitados sobre los que se lleva a cabo una compensación parcial mayor que sobre otros intervalos de frecuencia. Por ejemplo, se pueden compensar sólo las variaciones de respuesta de frecuencia como las encontradas con técnicas de micrófonos cercanos que tienen como resultado una baja intensificación de frecuencia por debajo de aproximadamente 500 Hz.

- No compensando hasta la cantidad que normalmente se llevaría a cabo, se obtiene una señal en el tiempo $PPX'_{WIRSS}(f)_n$ a partir de la cual pueden hacerse mejores estimaciones con respecto al factor de escala temporal local que es necesario para predecir correctamente la calidad final percibida. Después de este cálculo de escala local, aplicado a la señal degradada $PPY_{WIRSS}(f)_n$ se obtiene una señal de tiempo $PPY'_{WIRSS}(f)_n$ a partir de la cual puede hacerse una mejor estimación para la compensación de frecuencia necesaria final. La compensación de frecuencia final (es decir, la compensación para las desviaciones de frecuencia restantes) aplicada a la señal compensada parcialmente $PPX'_{WIRSS}(f)_n$ tiene como resultado una señal final $PPX''_{WIRSS}(f)_n$. Las señales resultantes $PPY'_{WIRSS}(f)_n$ y $PPX''_{WIRSS}(f)_n$ son procesadas luego nuevamente tal como se ha descrito anteriormente (distorsión a escala de sonoridad y etapas posteriores).

- Para el experto en la materia, resultará claro que pueden hacerse más modificaciones a la presente realización. La cantidad de compensación parcial puede adaptarse al contexto experimental. También es posible calcular y aplicar primero una compensación de factor de escala de potencia local parcial, luego calcular y aplicar la compensación de respuesta de frecuencia lineal y, por último, calcular y aplicar un factor de escala de potencia local final. También entra dentro del alcance de la presente invención usar más de tres etapas secundarias en las etapas de cálculo iterativo.

Antecedentes

[1] **BEERENDS** (J.G.), **STEMERDINK** (J.A.): A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation, *J. Audio Eng. Soc.* Vol. 42, Nº 3, págs. 115-123, marzo de 1994.

[2] **BEERENDS** (J.G.): Modelling Cognitive Effects that Play a Role in the Perception of Speech Quality, *Speech Quality Assessment*, Workshop papers, Bochum, págs. 1-9, noviembre de 1994.

[3] **BEERENDS** (J.G.): Measuring the quality of speech and music codecs, an integrated psychoacoustic approach, *98th AES convention*, pre-print Nº 3945, 1995.

[4] **HOLLIER** (M.P.), **HAWKSFORD** (M.O.), **GUARD** (D.R.): Error activity and error entropy as a measure of psychoacoustic significance in the perceptual domain, *IEEE Proceedings - Vision, Image and Signal Processing*, 141 (3), 203-208, junio de 1994.

[5] **RIX** (A.W.), **REYNOLDS** (R.), **HOLLIER** (M.P.): Perceptual measurement of end-to-end speech quality over audio and packet-based networks, *106th AES Convention*, pre-print Nº 4873, mayo de 1999.

[6] **HOLLIER** (M.P.), **HAWKSFORD** (M.O.), **GUARD** (D.R.), Characterisation of communications systems using a speech-like test stimulus, *Journal of the AES*, 41 (12), 1008-1021, diciembre de 1993.

[7] **ZWICKER** (Feldtkeller): Das Ohr als Nachrichtenempfänger, *S. Hirzel Verlag*, Stuttgart, 1967.

[8] ITU-T recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", ITU-T 02.2001.

[9] **BEERENDS** (J.G.); **HEKSTRA** (A.P.); **RIX** (A.W.); **HOLLIER** (M.P.), Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment Part II - Psychoacoustic Model, *J. Audio Eng. Soc.*, Vol. 50, nº 10, octubre de 2002.

[10] Solicitud de patente europea EP1343145, Koninklijke KPN N.V.

REIVINDICACIONES

1. Procedimiento para medir la calidad de transmisión de un sistema de transmisión de audio (10), siendo introducida una señal de entrada (X) en el sistema (10), teniendo como resultado una señal de salida (Y), en el que son procesadas tanto la señal de entrada (X) como la señal de salida (Y), que comprende:

- preproceso de la señal de entrada (X) y la señal de salida (Y) para obtener densidades de potencia tonal ($PPX_{WIRSS}(f_n)$, $PPY_{WIRSS}(f_n)$) para las señales respectivas;

- compensación de respuesta de frecuencia lineal y ganancia variable en el tiempo para obtener densidades de potencia tonal compensadas ($PPX''_{WIRSS}(f_n)$, $PPY'_{WIRSS}(f_n)$), en la que la compensación de respuesta de frecuencia lineal y ganancia variable en el tiempo comprende un bucle iterativo que tiene al menos tres cálculos de compensaciones, comprendiendo un cálculo de una primera compensación parcial de un primer tipo, un cálculo de una compensación de un segundo tipo, y un cálculo de una segunda compensación parcial del primer tipo, comprendiendo el primer tipo de cálculo y el segundo tipo de cálculo un cálculo diferente de entre un cálculo de una compensación de respuesta de frecuencia lineal y un cálculo de un factor de escala de potencia local;

- cálculo de una puntuación (Q) indicativa de la calidad de transmisión del sistema (10) a partir de las densidades de potencia tonal compensadas ($PPX''_{WIRSS}(f_n)$, $PPY'_{WIRSS}(f_n)$).

2. Procedimiento según la reivindicación 1, en el que el bucle iterativo comprende un cálculo de una primera compensación de frecuencia lineal parcial y la aplicación de la primera compensación de frecuencia lineal parcial a la densidad de potencia tonal de la señal de entrada ($PPX_{WIRSS}(f_n)$), seguido por un cálculo de un factor de escala de potencia local y la aplicación del factor de escala de potencia local a la densidad de potencia tonal de la señal de salida ($PPY_{WIRSS}(f_n)$), seguido por un cálculo de una segunda compensación de frecuencia lineal parcial y la aplicación de la compensación de frecuencia lineal a la densidad de potencia tonal parcialmente compensada de la señal de entrada ($PPX'_{WIRSS}(f_n)$).

3. Procedimiento según la reivindicación 1, en el que el bucle iterativo comprende un cálculo de una primera compensación de frecuencia lineal parcial y la aplicación de la primera compensación de frecuencia lineal parcial a la densidad de potencia tonal de la señal de salida ($PPY_{WIRSS}(f_n)$), seguido por un cálculo de un factor de escala de potencia local y la aplicación del factor de escala de potencia local a la densidad de potencia tonal de la señal de entrada ($PPX_{WIRSS}(f_n)$), seguido por un cálculo de una segunda compensación de frecuencia lineal parcial y la aplicación de la compensación de frecuencia lineal a la densidad de potencia tonal parcialmente compensada de la señal de salida ($PPY'_{WIRSS}(f_n)$).

4. Procedimiento según la reivindicación 2 ó 3, en el que la primera compensación de frecuencia lineal parcial es una primera estimación que es inferior a una compensación de frecuencia lineal requerida para la evaluación correcta de la distorsión lineal.

5. Procedimiento según la reivindicación 4, en el que la primera compensación de frecuencia lineal parcial es una función dependiente de la frecuencia.

6. Sistema para medir la calidad de transmisión de un sistema de transmisión de audio (10), siendo introducida una señal de entrada (X) en el sistema (10), teniendo como resultado una señal de salida (Y), que comprende:

- medios de preproceso (12) para preproceso de la señal de entrada (X) y la señal de salida (Y) para obtener densidades de potencia tonal ($PPX_{WIRSS}(f_n)$, $PPY_{WIRSS}(f_n)$) para las señales respectivas;

- medios de compensación (13, 14) para compensación de respuesta de frecuencia lineal y ganancia variable en el tiempo para obtener densidades de potencia tonal compensadas ($PPX''_{WIRSS}(f_n)$, $PPY'_{WIRSS}(f_n)$), que comprende un bucle iterativo que tiene al menos tres cálculos de compensaciones, comprendiendo un cálculo de una primera compensación parcial de un primer tipo, un cálculo de una compensación de un segundo tipo, y un cálculo de una segunda compensación parcial del primer tipo, comprendiendo el primer tipo de cálculo y el segundo tipo de cálculo un cálculo diferente de entre un cálculo de una compensación de respuesta de frecuencia lineal y un cálculo de un factor de escala de potencia local; y

- medios de cálculo (15, 16) para cálculo de una puntuación (Q) indicativa de la calidad de transmisión del sistema (10) a partir de las densidades de potencia tonal compensadas ($PPX''_{WIRSS}(f_n)$, $PPY'_{WIRSS}(f_n)$).

7. Sistema según la reivindicación 6, en el que el bucle iterativo comprende un cálculo de una primera compensación de frecuencia lineal parcial y la aplicación de la primera compensación de frecuencia lineal parcial a la densidad de potencia tonal de la señal de entrada ($PPX_{WIRSS}(f_n)$), seguido por un cálculo de un factor de escala de potencia local y la aplicación del factor de escala de potencia local a la densidad de potencia tonal de la señal de salida ($PPY_{WIRSS}(f_n)$), seguido por un cálculo de una segunda compensación de frecuencia lineal parcial y la aplicación de la segunda compensación de frecuencia lineal parcial a la densidad de potencia tonal parcialmente compensada de la señal de entrada ($PPX'_{WIRSS}(f_n)$).

ES 2 298 725 T3

8. Sistema según la reivindicación 6, en el que el bucle iterativo comprende un cálculo de una primera compensación de frecuencia lineal parcial y la aplicación de la primera compensación de frecuencia lineal parcial a la densidad de potencia tonal de la señal de salida ($PPY_{\text{WIRSS}}(f_n)$), seguido por un cálculo de un factor de escala de potencia local y la aplicación del factor de escala de potencia local a la densidad de potencia tonal de la señal de entrada ($PPX_{\text{WIRSS}}(f_n)$),
5 seguido por un cálculo de una segunda compensación de frecuencia lineal parcial y la aplicación de la segunda compensación de frecuencia lineal parcial a la densidad de potencia tonal parcialmente compensada de la señal de salida ($PPY'_{\text{WIRSS}}(f_n)$).

9. Sistema según la reivindicación 7 u 8, en el que la primera compensación de frecuencia lineal parcial es una
10 primera estimación que es inferior a una compensación de frecuencia lineal requerida para la evaluación correcta de la distorsión lineal.

10. Sistema según la reivindicación 9, en el que la primera compensación de frecuencia lineal parcial es una función
15 dependiente de la frecuencia.

11. Producto de programa de software que comprende código de software ejecutable por ordenador, que cuando se ejecuta en un sistema procesador, permite al sistema procesador ejecutar el procedimiento según una de las
20 reivindicaciones 1 a 5.

Fig 1

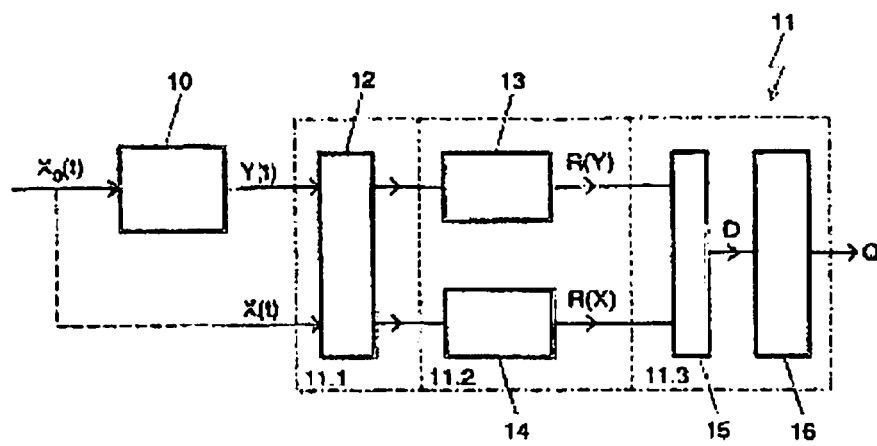


Fig 2

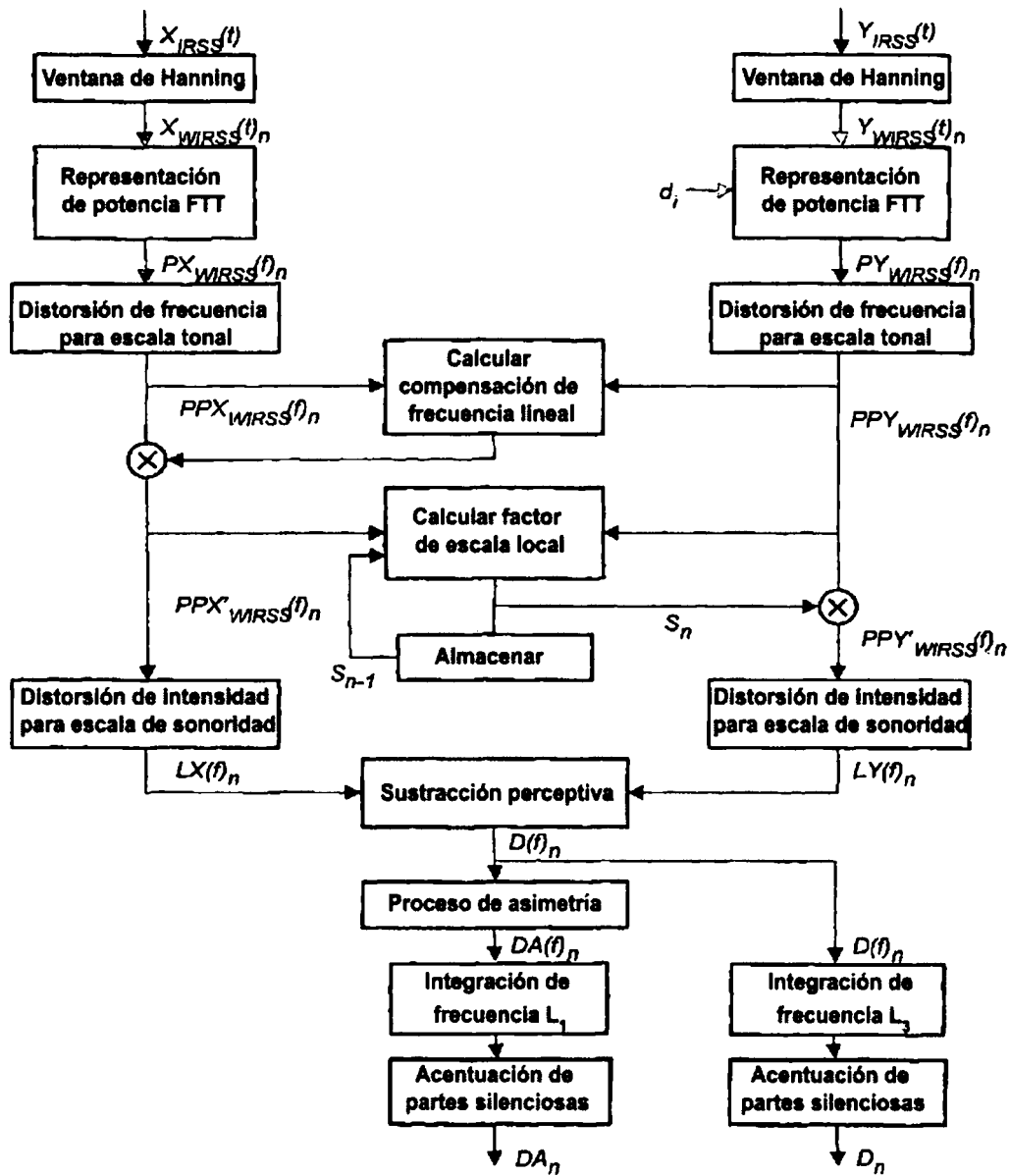


Fig 3

