

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 526 126**

51 Int. Cl.:

**G10L 25/69** (2013.01)

**G10L 21/0208** (2013.01)

**H04M 3/22** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.08.2010 E 10740657 (1)**

97 Fecha y número de publicación de la concesión europea: **08.10.2014 EP 2465112**

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

30 Prioridad:

**14.08.2009 EP 09010500**

**30.03.2010 EP 10158410**

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

**07.01.2015**

73 Titular/es:

**KONINKLIJKE KPN N.V. (50.0%)**

**Maanplein 55**

**2516 CK The Hague, NL y**

**NEDERLANDSE ORGANISATIE VOOR**

**TOEGEPAST- NATUURWETENSCHAPPELIJK**

**ONDERZOEK TNO (50.0%)**

72 Inventor/es:

**BEERENDS, JOHN y**

**VAN VUGT, JEROEN**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

**ES 2 526 126 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, producto de programa informático y sistema para determinar una calidad percibida de un sistema de audio

**CAMPO DEL INVENTO**

5 El invento se refiere a un método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio, con respecto a una señal de referencia. El invento se refiere además a un producto de programa informático que comprende un código ejecutable por ordenador, por ejemplo almacenado en un medio legible por ordenador, adaptado para realizar, cuando es ejecutado por un procesador, tal método. Finalmente el invento se refiere a un sistema para determinar un indicador de calidad que representa una calidad percibida de una  
10 señal de salida de un sistema de audio con respecto a una señal de entrada del sistema de audio que sirve como una señal de referencia.

**ANTECEDENTES DEL INVENTO**

15 La calidad de un dispositivo de audio puede ser determinada bien subjetiva o bien objetivamente. Los ensayos subjetivos consumen tiempo, son caros, y difíciles de reproducir. Por ello, se han desarrollado varios métodos para medir la calidad de una señal de salida, en particular una señal de voz, de un dispositivo de audio de un modo objetivo. En tales métodos, la calidad de voz de una señal de salida como es recibida desde un sistema de tratamiento de señal de voz es determinada por comparación con una señal de referencia.

20 Un método corriente que es ampliamente utilizado para este propósito es el método descrito en la Recomendación P.862 de ITU-T titulada "Evaluación perceptiva de calidad de voz (PESQ): Un método objetivo para evaluación de la calidad de voz de final a final de redes telefónicas de banda estrecha y codec de voz". En la recomendación P.862 de ITU-T, la calidad de una señal de salida procedente de un sistema de tratamiento de señal de voz, cuya señal es generalmente distorsionada, ha de ser determinada. La señal de salida y una señal de referencia, por ejemplo la señal de entrada del sistema de tratamiento de señal, son hechas corresponder con señales de representación de acuerdo con un modelo de percepción psicofísico del sistema auditivo humano. Basado en estas señales, se determina una señal diferencial que es representativa de una distorsión dentro de la señal de salida cuando es comparada a la señal de referencia. La señal  
25 diferencial es tratada a continuación de acuerdo con un modelo cognitivo, en el que ciertas propiedades de percepción de audición humana basadas en ensayos han sido modeladas, para obtener una señal de calidad que es una medida de la calidad de la percepción auditiva de la señal de salida.

30 Generalmente, los dispositivos de medición objetivos como el PESQ interpretan el ruido en términos de una disminución de calidad. Sin embargo, esta interpretación es demasiado simplista. En los sistemas de telecomunicaciones actuales, en particular en sistemas que utilizan Voz-Sobre-IP (VOIP) y tecnologías similares, el impacto del ruido en la calidad de la voz varía dependiendo del nivel de señal local. Consecuentemente, el PESQ frecuentemente no proporciona predicciones óptimas de la percepción de las señales de voz tratadas en tales sistemas de telecomunicaciones, que están resultando crecientemente populares.

35 En la publicación "Descomposición Degradación de la Calidad Percibida de Señales de Voz sobre la Base de una Aproximación de Modelado Perceptivo", J. Audio, Eng. Soc, Vol. 55, Nº 12, págs. 1059-1076 (Diciembre de 2007) los autores describen el modo en cómo percibimos la calidad de una señal de voz y cómo diferentes degradaciones tales como el ruido y las distorsiones de respuesta de frecuencia contribuyen a la calidad de voz percibida total (audición).

**RESUMEN DEL INVENTO**

40 Se desea tener un método de determinación de la calidad de transmisión de un sistema de audio que proporcione una correlación mejorada entre la calidad de voz cuando es determinada por medición objetiva y la calidad de voz cuando es determinada en ensayos subjetivos. Con este propósito, una realización del invento se refiere a un método para determinar un indicador de calidad que represente una calidad percibida de una señal de salida de un dispositivo de audio de acuerdo a la reivindicación 1. El indicador de calidad proporciona una indicación objetiva de la calidad de transmisión de un sistema de audio que puede ser utilizada para comparar diferentes sistemas de audio entre sí.

45 El contraste de ruido se refiere a la variación del nivel de ruido, preferiblemente medida a lo largo del tiempo. El valor absoluto del nivel de ruido y la variancia del ruido han de ser tenidos en cuenta. La variación del nivel de ruido a lo largo del tiempo puede desempeñar un papel relevante en la determinación del parámetro de contraste de ruido. Las figs. 6A-6F muestran gráficos esquemáticos del nivel de ruido en función del tiempo para explicar adicionalmente el concepto de contraste de ruido.

50 La fig. 6A muestra un gráfico ejemplar del nivel de ruido en función del tiempo que demuestra el concepto de contraste de ruido. En este ejemplo, el nivel de ruido medio reside generalmente alrededor de un cierto nivel de base, por ejemplo, durante el período de tiempo  $T_2$ , y ocasionalmente tiene un nivel de ruido más elevado, por ejemplo durante el período de tiempo  $T_1$ . El contraste de ruido de la señal en la fig. 6A se refiere a la diferencia entre el nivel de ruido elevado cuando está presente durante el período  $T_1$  y el nivel de ruido regular cuando está presente durante el período  $T_2$ . El contraste de  
55 ruido es de influencia sobre la calidad percibida de la señal de salida de un dispositivo de audio. Si tal señal de salida

comprendiera un diseño de ruido como se ha mostrado en la fig. 6A, la presencia de las partes del nivel de ruido más elevado que excede de los niveles de ruido normales puede ser experimentada como muy perturbadora por un oyente o persona que escucha. En particular un cambio en el nivel de ruido focaliza al oyente sobre el ruido. Así los niveles de ruido cambiantes son generalmente experimentados como más perturbadores que los niveles de ruido constantes.

- 5 El contraste de ruido puede ser determinado comparando diferentes partes de la señal entre sí. Cada parte puede tener una cierta longitud. La influencia sobre la longitud del intervalo de tiempo de partes de señal utilizadas en la evaluación del contraste de ruido es demostrada con referencia a los gráficos mostrados en las figs. 6B, 6C.

Las figs. 6B, 6C muestran el nivel de ruido de dos señales. Para ambas señales, el ruido está distribuido normalmente. Los niveles de ruido de las señales tienen el mismo nivel de ruido medio, sin embargo, la varianza del ruido es diferente. El intervalo de confianza de  $3\sigma$  del ruido en la fig. 6B es mayor que el intervalo de confianza de  $3\sigma$  mostrado en la fig. 6C. Obsérvese que la variación del nivel de ruido regular mostrada dentro del intervalo de confianza no es probable que ocurra en la realidad y es simplemente utilizado para mostrar que la señal puede variar ampliamente dentro del intervalo de confianza.

- 10  
15 En las figs. 6B, 6C, si las partes de señal utilizadas para evaluar el contraste de ruido son suficientemente grandes, por ejemplo un período de tiempo  $T_3$ , el valor de nivel de ruido promedio dentro de estas partes de señal es igual al valor medio del nivel de ruido. Este valor será encontrado para todas las partes de señal que dan como resultado un contraste de ruido igual a cero.

- 20 Sin embargo, si el período de tiempo de partes utilizadas para evaluar el contraste de ruido es relativamente corto, por ejemplo una longitud  $T_4$  mostrada en las figs. 6A, 6B, el valor promedio del nivel de ruido dentro de cada parte puede diferir bien del valor del nivel de ruido medio a largo plazo, y de otras partes. El contraste de ruido será así entonces distinto de cero. En tal caso, la varianza del nivel de ruido puede tener una influencia significativa sobre el contraste de ruido. La varianza de ruido de la señal en la fig. 6B es mayor que la varianza de ruido de la señal en la fig. 6C. Debido a la dispersión más amplia de los valores de nivel de ruido en la señal de la fig. 6B, la diferencia entre el nivel de ruido promedio dentro de un período de tiempo corto seleccionado y el nivel de ruido medio a un plazo mayor será generalmente mayor que tal diferencia obtenida durante una evaluación similar de la señal mostrada en la fig. 6C. Consecuentemente, el contraste de ruido de la señal en la fig. 6B será mayor que el contraste de ruido de la señal en la fig. 6C si se utilizan períodos cortos de tiempo. El contraste de ruido como se ha descrito con referencia a las figs. 6B, 6C es menos indicativo para la perturbación experimentada por un oyente. Se prefiere por ello la selección de un período de tiempo suficientemente largo para la evaluación del contraste de ruido.

- 30 En general, un parámetro de contraste de ruido puede ser definida como una medida del contraste de ruido dentro de una parte de la señal. Más particularmente, el parámetro de contraste de ruido es una medida de la variación de nivel de ruido entre partes seleccionadas dentro de la señal de salida, por ejemplo tramas de tiempo seleccionadas. La selección de las partes de señal puede estar basada en el modelo de percepción psicofísico del sistema auditivo humano. Por ejemplo, la selección de partes de señal para medición de niveles de ruido y cálculo de un parámetro de contraste de ruido puede tener en cuenta que las variaciones del nivel de ruido pueden ser más importantes que el nivel de ruido absoluto medido. Esto abarca la situación en la que los oyentes consideran que las variaciones del nivel de ruido de fondo son más perturbadoras que el ruido con un nivel de ruido promedio ligeramente más elevado que tiene menos variaciones. Tales preferencias no han sido aún incluidas en los métodos actuales para determinar la calidad percibida de las propiedades de transmisión de un dispositivo de audio tal como el PESQ.

- 40 Las figs. 6D-6F demuestran además consideraciones que pueden ser tenidas en cuenta para calcular un parámetro de contraste de ruido basado en niveles de ruido medidos en diferentes partes de una señal.

La fig. 6D representa la variación del nivel de ruido de dos señales, es decir la señal I y la señal II, en función del tiempo. A lo largo del período de tiempo completo mostrado, el nivel de ruido de la señal I es mayor que el nivel de ruido de la señal H. Sin embargo, el contraste de ruido de ambas señales es el mismo.

- 45 El parámetro de contraste de ruido puede ser una medida del contraste de ruido en que el valor absoluto del nivel de ruido es tenido en cuenta o donde solamente es tenida en cuenta la diferencia absoluta en niveles de ruido. Las variaciones del nivel de ruido para la señal II son mayores en proporción al nivel de ruido absoluto. El parámetro de contraste de ruido puede ser calculado para tener esto en cuenta, de manera que el parámetro de contraste de ruido de la señal II puede ser por ello mayor que el parámetro de contraste de ruido de la señal I. El parámetro de contraste de ruido puede ahora ser construido para ser más susceptible al contraste de ruido a menores niveles de ruido que a contraste de ruido a mayores niveles de ruido utilizando niveles de ruido absolutos en el cálculo. El uso de este tipo de medida puede ser beneficioso para mejorar la determinación de la calidad percibida de una señal de audio, ya que los oyentes son típicamente más sensibles a las variaciones del nivel de ruido a bajos niveles de ruido que a las variaciones del nivel de ruido similares a niveles de ruido elevados. Además, un solo cambio rápido en un nivel de ruido puede focalizar al oyente sobre el ruido de fondo perturbador y así puede conducir a un mayor impacto de este nivel de ruido de fondo sobre la calidad de voz total percibida.

El parámetro de contraste de ruido puede además es definido de tal modo que sea diferente para diferentes tipos de

variaciones del nivel de ruido como será ilustrado con referencia a las figs. 6E, 6F. En estas figuras, solamente se ha mostrado el nivel de ruido medio de la señal.

5 La fig. 6E muestra un gráfico de nivel de ruido en función del tiempo en el que el nivel de ruido comienza a un nivel de ruido elevado y permanece a ese nivel durante un periodo de tiempo  $T_5$ . El nivel de ruido cae entonces gradualmente durante un periodo de tiempo  $T_6$  hacia un nivel de ruido inferior y permanece a este nivel durante el periodo de tiempo  $T_5$ .

La fig. 6F muestra un gráfico del nivel de ruido en función del tiempo en el que el nivel de ruido comienza en el nivel de ruido bajo y permanece a ese nivel durante un periodo de tiempo  $T_5$ . El nivel de ruido asciende entonces en el periodo de tiempo  $T_6$ , y permanece al nivel de ruido más elevado durante un periodo de tiempo  $T_5$ .

10 Durante el intervalo de tiempo mostrado es decir  $2xT_5 + T_6$ , los niveles de ruido promedio para las señales mostradas en las figs. 6E, 6F son los mismos. Sin embargo, incluso si el nivel de ruido promedio es el mismo, la percepción del nivel de ruido puede ser diferente para niveles de ruido que son crecientes a lo largo del tiempo en contraste con niveles de ruido que son decrecientes a lo largo del tiempo. En particular, si el intervalo de tiempo mostrado representa un intervalo del tiempo en que no hay voz, puede experimentarse un incremento de ruido durante tal periodo de silencio como más perturbador que una disminución de ruido.

15 El parámetro de contraste de ruido puede ahora ser definido de tal modo que este aspecto perceptivo sea tenido en cuenta. Por ejemplo, el parámetro de contraste de ruido puede ser definido de tal modo que partes de señal en un tiempo posterior dentro de un intervalo de tiempo que es evaluado tienen más peso que partes de la señal previas dentro del intervalo de tiempo. En caso de las situaciones representadas en las figs. 6E, 6F, el parámetro de contraste de ruido puede ser definido dividiendo el nivel de ruido medido en el periodo de tiempo  $T_5$  antes del cambio gradual del nivel de ruido sobre el nivel de ruido debido en el periodo de tiempo  $T_5$  después de ese cambio gradual. En tal caso, el parámetro de contraste de ruido de la variación del nivel de ruido mostrado en la fig. 6E es menor que 1, mientras el parámetro de contraste de ruido de la variación del nivel de ruido mostrado en la fig. 6F es mayor que 1.

20 El parámetro de contraste de ruido es así utilizado para acceder adicionalmente a la percepción subjetiva de ruido. Utilizar el contraste de ruido dentro de una señal mejora la correlación entre la calidad de voz obtenida por medición objetiva y la calidad de voz cuando es determinada en ensayos subjetivos.

En otra realización de acuerdo con la reivindicación 3, el parámetro de contraste de ruido puede ser establecido a un valor fijo predeterminado si la fracción comprimida está por debajo del valor fijo predeterminado.

30 En una realización de acuerdo con la reivindicación 4, aplicar la supresión de ruido sobre la señal de referencia comprende: calcular una densidad espectral de ruido promedio de referencia a lo largo de un número de tramas de tiempo que tienen un nivel de intensidad detectado por debajo del valor de umbral; determinar un factor de supresión de ruido de referencia basado en una correlación optimizada entre información de evaluación de calidad objetiva y subjetiva; y calcular un nivel de ruido deseado basado en el factor de supresión de ruido de referencia y en la densidad espectral de ruido promedio de referencia. La supresión de ruido de referencia realizada sobre la señal de referencia escalada puede así ser prevista para suprimir el ruido hasta un nivel de ruido que sea considerado como un nivel de ruido bajo deseable. Tal nivel de ruido bajo deseable puede ser determinado en experimentos de evaluación de calidad subjetivos.

35 En una realización de acuerdo con la reivindicación 5, aplicar la supresión de ruido sobre la señal de salida comprende: calcular una densidad espectral de ruido de salida promedio a lo largo de un número de tramas de tiempo correspondientes a tramas de tiempo dentro de la señal de referencia que tienen un nivel de intensidad detectado por debajo de un valor de umbral; determinar una constante de nivel de ruido bajo basado en una correlación optimizada entre información de evaluación de calidad objetiva y subjetiva; determinar un factor de supresión de ruido de salida basado en una correlación optimizada entre información de evaluación de calidad objetiva y subjetiva; y calcular un nivel de ruido representativo de perturbación en la señal de salida utilizando la constante de nivel de ruido bajo, el factor de supresión de ruido de salida y la densidad espectral de ruido de salida promedio. La supresión del ruido de salida realizada sobre la señal puede ser a continuación prevista para suprimir el ruido hasta un nivel de ruido representativo de la perturbación. La supresión de ruido de salida puede ser dependiente de la frecuencia.

40 En una realización de acuerdo con la reivindicación 6, el tratamiento comprende además: transformar la señal de referencia y la señal de salida desde el dominio de tiempo hacia el dominio de tiempo-frecuencia; derivar una función de densidad de energía de tono de referencia a partir de la señal de referencia y derivar una función de densidad de energía de tono de salida a partir de la señal de salida; escalar localmente la función de densidad de energía de tono de referencia para obtener una función de densidad de energía de tono de referencia localmente escalada; compensar parcialmente o bien la función de densidad de energía de tono de salida o bien la función de densidad de energía de tono de referencia con respecto a la frecuencia; y derivar una función de densidad de sonoridad de referencia y una función de densidad de sonoridad de salida.

45 En otra realización de acuerdo con la reivindicación 7, el tratamiento comprende además: escalar la intensidad de la función de densidad de energía de tono de referencia a partir del nivel de intensidad fijado hacia un nivel de intensidad relacionado con la función de densidad de energía de tono de salida para obtener una diferencia de nivel de intensidad en las funciones de densidad de energía de tono que permite la cuantificación del impacto del nivel de intensidad sobre la

5 calidad percibida; escalar la sonoridad de la función de densidad de sonoridad de salida hacia un nivel de sonoridad fijado en el dominio de sonoridad perceptivo; y escalar la sonoridad de la función de densidad de solidaridad de referencia desde un nivel de sonoridad correspondiente al nivel de intensidad relacionado de salida hacia un nivel de sonoridad relacionado al nivel de sonoridad de la función de densidad de sonoridad de salida en el dominio de sonoridad perceptivo, para obtener una diferencia de nivel de sonoridad entre la función de densidad de sonoridad de referencia y la función de densidad de sonoridad de salida que permita la cuantificación del impacto de la sonoridad sobre la calidad percibida de la señal de salida.

10 En una realización de acuerdo con la reivindicación 8, seleccionar las tramas de tiempo comprende seleccionar tramas de tiempo de la señal de salida correspondientes a tramas de tiempo de la señal de referencia escalada que satisfacen una condición predeterminada. Tal condición predeterminada, de acuerdo con la reivindicación 9, puede incluir que la trama de tiempo de la señal de referencia escalada, una trama de silencio tiene un valor de nivel de intensidad por debajo de un valor de umbral. En otra realización de acuerdo con la reivindicación 10, la condición predeterminada incluye además que la trama de tiempo de la señal de referencia escalada es parte de una serie de tramas de silencio consecutivas, un intervalo de silencio. El uso de tramas de silencio y/o intervalos de silencio mejora la fiabilidad del parámetro de contraste de ruido, y permite así una mejora adicional de la correlación entre la calidad de voz cuando es determinada por medición objetiva y la calidad de voz cuando es determinada en un ensayo subjetivo.

15 En una realización de acuerdo con la reivindicación 11, el invento se refiere a un producto de programa informático que comprende un código ejecutable por ordenador, por ejemplo almacenado en un medio legible por ordenador, adaptado para realizar, cuando es ejecutado por un procesador, cualquiera de las realizaciones de método antes mencionadas.

20 Finalmente, en otra realización de acuerdo con la reivindicación 12, el invento se refiere a un sistema para determinar un indicador de calidad que representa una calidad percibida de una señal de salida  $Y(t)$  de un sistema de audio, por ejemplo un dispositivo de tratamiento de voz, con respecto a una señal de entrada  $X(t)$  del sistema de audio que sirve como una señal de referencia comprendiendo el sistema: un dispositivo de tratamiento previo para tratar previamente la señal de referencia y la señal de salida; un primer dispositivo de tratamiento para tratar la señal de referencia, y un segundo dispositivo de tratamiento para tratar la señal de salida para obtener señales de representación  $R(X)$ ,  $R(Y)$  para la señal de referencia y la señal de salida respectivamente; un dispositivo de diferenciación para combinar las señales de representación de la señal de referencia y de la señal de salida de modo que se obtenga una señal diferencial  $D$ ; y un dispositivo de modelado para tratar la señal diferencial para obtener una señal de calidad  $Q$  que representa una estimación de la calidad perceptiva del sistema de tratamiento de voz; en que el dispositivo de tratamiento previo, el primer dispositivo de tratamiento, y el segundo dispositivo de tratamiento forman un sistema de tratamiento para realizar cualquiera de las realizaciones de método antes mencionadas.

#### BREVE DESCRIPCIÓN DE LOS DIBUJOS

En los dibujos:

35 La fig. 1 muestra esquemáticamente una configuración general que incluye un sistema para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio con respecto a una señal de referencia;

La fig. 2 muestra esquemáticamente un método de determinación de un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio con respecto a una señal de referencia de acuerdo a PESQ;

40 La fig. 3 muestra esquemáticamente un método de determinación de un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio con respecto a una señal de referencia de acuerdo con una realización del presente invento;

La fig. 4 muestra esquemáticamente una realización del presente invento aplicada al método mostrado en la fig. 2;

La fig. 5 muestra esquemáticamente una realización del presente invento aplicada a otro método;

Las figs. 6A-6F muestran diferentes gráficos del nivel de ruido en función del tiempo.

#### 45 DESCRIPCIÓN DETALLADA DE LOS DIBUJOS

Lo que sigue es una descripción de ciertas realizaciones del invento, dadas a modo de ejemplo solamente.

A lo largo de toda la descripción, los términos "local" y "global" serán utilizados con respecto a una operación realizada sobre una señal. Una operación "local" se refiere a una operación realizada sobre parte de la señal de tiempo, por ejemplo sobre una única trama. Una operación "global" se refiere a una operación realizada sobre la señal completa.

50 A lo largo de toda la descripción, los términos "emitida" y "distorsionada" pueden ser utilizados con relación a una señal que se origina a partir de una salida de un sistema de audio, como un dispositivo de tratamiento de voz. A lo largo de toda la descripción, los términos "referencia" y "original" pueden ser utilizados en relación a una señal ofrecida como una entrada al sistema de audio, siendo utilizada además la señal como una señal con la que la señal emitida o distorsionada

ha de ser comparada.

La fig. 1 muestra esquemáticamente una configuración general que incluye un sistema para determinar un indicador de calidad que representa una calidad percibida de una señal emitida de un dispositivo de audio, por ejemplo un sistema de tratamiento de voz, con respecto a una señal de referencia. La configuración incluye un sistema de audio 10 bajo investigación, por ejemplo una red de telecomunicación, elemento de red o dispositivo de tratamiento de voz en una red o estación móvil. La configuración también incluye un sistema 20 para medir la calidad de transmisión del sistema de audio, denominado en lo que sigue como un sistema 20 de medición de calidad.

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

En esta realización, el sistema 20 de medición de calidad comprende una sección 20a de tratamiento previo, una sección 20b de tratamiento, y una sección 20c de combinación de señal para tratar las dos señales de entrada  $X(t)$ ,  $Y(t)$  de tal modo que pueda ser proporcionada la señal de salida  $Q$ .

La sección 20a de tratamiento previo comprende un dispositivo 30 de tratamiento previo dispuesto para realizar una o más acciones de tratamiento previo tales como escalado de nivel fijo y alineación de tiempo para obtener señales tratadas previamente  $X_p(t)$ , e  $Y_p(t)$ .

La sección 20b de tratamiento del sistema 20 de medición de calidad está prevista para hacer corresponder las señales tratadas previamente sobre señales de representación de acuerdo a un modelo de percepción psicofísico del sistema auditivo humano. La señal tratada previamente  $X_p(t)$  es tratada en un primer dispositivo de tratamiento 40a para obtener señal de representación  $R(X)$ , mientras la señal tratada previamente  $Y_p(t)$  es tratada en un segundo dispositivo de tratamiento 40b para obtener la señal de representación  $R(Y)$ .

La sección 20c de combinación de señal del dispositivo 20 de medición de calidad está prevista para combinar las señales de representación  $R(X)$ ,  $R(Y)$  para obtener una señal diferencial  $D$  utilizando un dispositivo de diferenciación 50. Finalmente, un dispositivo de modelado 60 trata la señal diferencial  $D$  de acuerdo con un modelo en el que han sido modeladas ciertas propiedades de los seres humanos para obtener la señal de calidad  $Q$ . Las propiedades humanas, por ejemplo propiedades cognitivas, pueden ser obtenidas mediante ensayos de audición subjetivos realizados con un número de sujetos humanos.

El dispositivo 30 de tratamiento previo, el primer dispositivo de tratamiento 40a, y el segundo dispositivo de tratamiento 40b pueden formar un sistema de tratamiento que puede ser utilizado para llevar a cabo realizaciones del invento como se explicará con más detalle posteriormente. El sistema de tratamiento o componentes del mismo pueden tener la forma de un procesador de hardware tal como un Circuito Integrado Específico de Aplicación (ASIC) o un dispositivo informático para ejecutar un código ejecutable por ordenador en forma de software o firmware. El dispositivo informático puede comprender, por ejemplo un procesador y una memoria que está acoplada comunicativamente al procesador. Ejemplos de memoria incluyen, pero no están limitados a, Memoria Solo de Lectura (ROM), Memoria de Acceso Aleatorio (RAM), ROM Programable que se Puede Borrar (EPROM), ROM Programable que se Puede Borrar Eléctricamente (EEPROM), y memoria flash.

El dispositivo informático puede comprender además una interfaz de usuario para permitir la introducción de instrucciones o notificaciones por usuarios externos. Ejemplos de una interfaz de usuario incluyen, pero no están limitados a, un ratón, un teclado, y una pantalla táctil.

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

El código ejecutable por ordenador almacenado en el medio legible por ordenador, después de la carga del código en la memoria del dispositivo informático, puede ser adaptado para llevar a cabo realizaciones del invento que serán descritas posteriormente.

Alternativa o adicionalmente, tales realizaciones del invento pueden tomar la forma de un producto de programa informático que comprende un código ejecutable por ordenador para realizar tal método cuando es ejecutado sobre un dispositivo informático. El método puede ser realizado a continuación con un procesador del dispositivo informático después de cargar el código ejecutable por ordenador a una memoria del dispositivo informático.

Así, un método de medición perceptiva objetivo imita la percepción de sonido de sujetos en un programa informático con el objetivo de predecir la calidad percibida de modo subjetivo de sistemas de audio, tales como codec de voz, enlaces

telefónicos, y teléfonos móviles. Las señales físicas de entrada y salida del dispositivo bajo ensayo son hechas corresponder sobre las representaciones psicofísicas que se adaptan tanto como sea posible a las representaciones internas dentro de la cabeza de un ser humano. La calidad del dispositivo bajo ensayo es juzgada sobre la base de diferencias en la representación interna. El método de medición perceptivo objetivo mejor conocido disponible es el PESQ (Evaluación Perceptiva de Calidad de Voz).

La fig. 2 muestra esquemáticamente un método de determinación de un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio con respecto a una señal de referencia de acuerdo a PESQ como se ha descrito en la Recomendación P.862 de la ITU-T, en lo que sigue PESQ. La PESQ puede ser utilizado en una configuración como se ha mostrado esquemáticamente en la fig. 1. En la PESQ, una señal de referencia  $X(t)$  es comparada con una señal de salida  $Y(t)$  que es el resultado de hacer pasar  $X(t)$  a través de un sistema de audio, por ejemplo un sistema de tratamiento de voz como un sistema de comunicación. La calidad de la señal de salida de la PESQ, también denominada como calificación o nota de PESQ, es una predicción de la calidad percibida que se le daría a  $Y(t)$  por los sujetos en un ensayo de audición subjetivo. La calificación de PESQ tiene la forma de una así llamada calificación de opinión media (MOS). Con este propósito, la salida de la PESQ es hecha corresponder sobre una escala a modo de MOS, es decir un único número en el intervalo de -0,5 a 4,5, aunque para la mayor parte de los casos el intervalo de salida estará entre 1,0 y 4,5, que es el intervalo normal de los valores de MOS encontrados en un experimento de calidad de audición de Clasificación de Categoría Absoluta (ACR).

El tratamiento previo en la PESQ comprende la alineación de nivel de ambas señales  $X(t)$ ,  $Y(t)$  para obtener señales  $X_s(t)$ ,  $Y_s(t)$  respectivamente, así como un filtrado del Sistema de Referencia Intermedio (IRS) para obtener señales  $X_{IRSS}(t)$ ,  $Y_{IRSS}(t)$  respectivamente. La alineación de nivel implica escalar la intensidad hacia un nivel fijo, en PESQ 79 dB SPL. El filtrado de IRS es realizado para asegurar que el método de medir la calidad de transmisión es relativamente insensible al filtrado de un elemento del sistema de telecomunicación, por ejemplo un teléfono móvil o similar. Finalmente, se determina un retardo de tiempo entre la señal de referencia  $X_{IRSS}(t)$  e  $Y_{IRSS}(t)$  que conduce a una señal de salida desplazada en el tiempo  $Y_{IRSS}'(t)$ . La comparación entre la señal de referencia y la señal de salida es ahora asumida que tiene lugar con respecto al mismo tiempo.

El oído humano realiza una transformación tiempo-frecuencia. En la PESQ, esto es modelado realizando una transformación rápida de Fourier (FFT) a corto plazo con una ventana de Hanning sobre señales de tiempo  $X_{IRSS}(t)$  e  $Y_{IRSS}'(t)$ . La ventana de Hanning tiene típicamente un tamaño de 32 ms. Las ventanas de tiempo adyacentes, denominadas aquí como tramas, se solapan típicamente en un 50%. La información de fase es desechada. La suma de las partes real cuadrada e imaginaria cuadrada de los componentes de la FFT compleja, es decir los espectros de energía, son utilizados para obtener representaciones de energía  $PX_{WIRSS}(f)_n$  y  $PY_{WIRSS}(f)_n$ , donde  $n$  indica la trama en consideración. Las representaciones de energía son divididas en bandas de frecuencia, denominadas a continuación como bandas de FFT.

El sistema auditivo humano tiene una resolución de frecuencia más fina a bajas frecuencias que a altas frecuencias. Una escala de tono refleja este fenómeno, y por esta razón PESQ deforma las frecuencias a una escala de tono, en este caso una denominada escala de Bark. La conversión del eje de frecuencia (discreto) implica la clasificación ordenada de bandas de FFT para formar bandas de Bark, típicamente 24. Las señales resultantes son denominadas como densidades de energía de tono o funciones de densidad de energía de tono y denominadas como  $PPX_{WIRSS}(f)_n$  y  $PPY_{WIRSS}(f)_n$ . Las funciones de densidad de energía de tono proporcionan una representación interna que es análoga a la representación psicofísica de señales de audio en el sistema auditivo humano, teniendo en cuenta la frecuencia perceptiva.

Para tratar sobre el filtrado en el sistema de audio que ha de ser ensayado, el espectro de energía de las densidades de energía de tono de referencia y de salida son promediados a lo largo del tiempo. Un factor de compensación parcial es calculado a partir de la relación del espectro de salida al espectro de referencia. La densidad de energía de tono de referencia  $PPX_{WIRSS}(f)_n$  de cada trama  $n$  es multiplicada a continuación por este factor de compensación parcial para igualar la señal de referencia a la señal de salida. Esto da como resultado una densidad de energía de tono de referencia filtrada inversamente  $PPX'_{WIRSS}(f)_n$ . Esta compensación parcial es utilizada debido a que el filtrado suave es difícil de observar mientras el filtrado severo puede ser perturbador para el oyente. La compensación es llevada a cabo sobre la señal de referencia debido a que la señal de salida es la que es juzgada por el sujeto en un experimento de audición de ACR.

Con el fin de compensar las variaciones de ganancia a corto plazo, se calcula un factor de escalado local. El factor de escalado local es multiplicado a continuación por la función de densidad de energía de tono de salida  $PPY_{WIRSS}(f)_n$  para obtener una función de densidad de energía  $PPY'_{WIRSS}(f)_n$  escalada localmente.

Después de la compensación parcial para el filtrado realizado sobre la señal de referencia y la compensación parcial para las variaciones de ganancias a corto plazo realizada sobre la señal de salida, las densidades de energía de tono de referencia y degradadas son transformadas a una escala de sonoridad Sone utilizando la ley de Zwicker. Las agrupaciones bidimensionales resultantes  $LX(f)_n$  y  $LY(f)_n$  son denominadas como funciones de densidad de sonoridad para la señal de referencia y la señal de salida respectivamente. Para  $LX(f)_n$  esto significa:

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

donde  $P_0(f)$  es el umbral de audición absoluto,  $S_l$  el factor de escalado de sonoridad, y  $\gamma$ , la denominada energía de Zwicker, tiene un valor de aproximadamente 0,23. Las funciones de densidad de sonoridad representan la representación psicofísica interna de señales de audio en el sistema auditivo humano teniendo en cuenta la percepción de sonoridad.

Entonces las funciones de densidad de sonoridad de referencia y de salida  $LX(f)_n$ ,  $LY(f)_n$  son sustraídas dando como resultado una función de densidad de sonoridad diferencia  $D(f)_n$  a partir de la cual puede ser derivada una medida de calidad percibida, o indicador de calidad. Tal derivación puede ser hecha de un modo como se ha descrito más adelante. Otros detalles pueden ser encontrados en la Recomendación P.862 de ITU-T que está incluida aquí como referencia.

Como se ha mencionado antes, las funciones de densidad de sonoridad de referencia y de salida  $LX(f)_n$  y  $LY(f)_n$  son sustraídas, lo que da como resultado una diferencia con signo. Cuando esta diferencia es positiva, es decir  $LY(f)_n$  es mayor que  $LX(f)_n$ , se han añadido componentes tales como el ruido. Por otro lado, cuando esta diferencia es negativa, han sido omitidos componentes a partir de la señal original. La función de densidad de sonoridad diferencia forma una agrupación de diferencia que puede ser denominada como una densidad de perturbación en bruto. Simultáneamente, para cada tiempo y frecuencia el mínimo de las funciones de densidad de sonoridad de referencia y de salida es calculado y comprimido, por ejemplo por multiplicación por un factor de multiplicación, en la PESQ igual a 0,25. La minimización y compresión subsiguiente da como resultado una agrupación bidimensional denominada además como una agrupación de máscara. Si la densidad de perturbación en bruto es positiva y mayor que el valor respectivo correspondiente en la agrupación de máscara, el valor de la agrupación de máscara respectivo es sustraído de la densidad de perturbación en bruto. Si la densidad de perturbación en bruto se encuentra entre más y menos la magnitud del valor de agrupación de máscara, la densidad de perturbación es establecida a cero. Si la densidad de perturbación en bruto es más negativa que menos el valor de agrupación de máscara, el valor es añadido a la densidad de perturbación en bruto. El resultado de estas comparaciones y modificaciones subsiguientes se transforma en una densidad de perturbación en función del tiempo (ventana número  $n$ ) y frecuencia  $D(f)_n$ .

Si una señal de salida es distorsionada por la adición de componentes, tales como el ruido, la señal de salida puede ser descompuesta en la señal de entrada y la distorsión. Sin embargo, si la señal de salida es distorsionada por eliminación o retirada de componentes tal composición no puede ser realizada fácilmente en caso de que pueda serlo. Esta situación es denominada en la PESQ cómo el efecto de asimetría y es modelada calculando una densidad de perturbación asimétrica  $DA(f)_n$  por trama por multiplicación de la densidad de perturbación  $D(f)_n$  por un factor de asimetría. En la PESQ, este factor de asimetría es igual a la relación de las densidades de energía de tono de salida y de referencia elevada a la potencia de 1,2. Si el factor de asimetría es menor que un valor de umbral predeterminado, en la PESQ igual a 3, el factor de asimetría es establecido a cero. Si el factor de asimetría excede de otro valor de umbral predeterminado, en la PESQ igual a 12, el factor es limitado a ese valor de umbral. Así para la densidad de perturbación asimétrica  $DA(f)_n$  sólo los valores distintos de cero en la agrupación respectiva se refieren a celdas de agrupación para las que la densidad de energía de tono de salida ha excedido de la densidad de energía de tono original.

La densidad de perturbación  $D(f)_n$  y la densidad de perturbación asimétrica  $DA(f)_n$  son a continuación integradas, o sumadas, a lo largo del eje de frecuencias utilizando dos normas  $L_p$  diferentes así denominadas y una ponderación sobre tramas que tienen una baja sonoridad, es decir:

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

$$DA_n = M_n \sqrt{\sum_{f=1, \dots, \text{Número de bandas de Bark}} (|DA(f)_n| W_f)} \quad (4)$$

con  $M_n$  un factor de multiplicación previsto para enfatizar perturbaciones que ocurren durante silencios en la señal de entrada, y  $W_f$  una serie de constantes proporcional a la anchura de las clases de Bark que están siendo utilizadas. Los valores agregados,  $D_n$  y  $DA_n$ , obtenidos por las ecuaciones (3) y (4) son denominados como perturbaciones de trama. Las perturbaciones de trama son limitadas a un valor máximo, en la PESQ igual a 45. Si tramas consecutivas tienen una perturbación de trama por encima del umbral, pueden ser alineadas de nuevo.

A continuación, los valores de perturbación de trama y los valores de perturbación de trama asimétricos son agregados a lo largo de intervalos de un número limitado de tramas, en la PESQ igual a 20 tramas. Estos intervalos tienen un solapamiento, típicamente del 50%, y no se utiliza función de ventana. A continuación los valores agregados a lo largo del número limitado de tramas, tanto para perturbación como para perturbación asimétrica, son agregados a lo largo del intervalo activo del archivo de voz para formar un valor de perturbación promedio y un valor de perturbación asimétrica promedio respectivamente. La calificación final, es decir el indicador de calidad, es entonces una combinación lineal del valor de perturbación promedio y del valor de perturbación asimétrica promedio.

La fig. 3 muestra la determinación de un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio con respecto a una señal de referencia de acuerdo a una realización del presente invento. Después de las acciones de tratamiento previo tales como el filtrado por IRS y el retardo de tiempo, la señal de referencia y la señal de salida son transformadas desde el dominio de tiempo al dominio de frecuencia perceptiva. Esto puede conseguirse de un modo similar a como se ha mostrado en la fig. 2 con referencia a la PESQ. Es decir, en primer lugar una función de implementación de ventana es ejecutada en combinación con una FFT para llevar la señal desde el dominio del tiempo al dominio del tiempo-frecuencia. Después de la FFT, las señales son deformadas a una escala de tono, por ejemplo una escala de frecuencia en Bark, para obtener una representación en el dominio de frecuencia perceptivo.

Adicionalmente, la señal de referencia es escalada globalmente a un nivel de intensidad fija. Esta acción de escalado puede ser realizada antes de la transformación, como se ha mostrado en la fig. 3, pero puede también ser realizada en el dominio de tiempo-frecuencia (perceptivo).

Dentro del dominio de tiempo-frecuencia perceptivo, denominado además como dominio de frecuencia perceptivo, la señal de referencia escalada puede ser sometida a compensación de frecuencia como se ha descrito con referencia a la fig. 2. De manera similar, la señal de salida puede ser sometida a una acción de escalado local. La escalada local puede también ser realizada con respecto a la señal de referencia como se ha mostrado esquemáticamente en la fig. 3. Tanto la señal de referencia escalada como la señal de salida son a continuación sometidas a deformación de intensidad a la escala de sonoridad como se ha descrito con referencia a la PESQ mostrado en la fig. 2. La señal de referencia y la señal de salida son a continuación representadas en el dominio de sonoridad perceptiva.

Además de todas las acciones antes mencionadas, que pueden todas estar presentes de algún modo en un método basado en la PESQ como se ha mostrado en la fig. 2, el método de determinación de un indicador de calidad de acuerdo con realizaciones del invento incluye también la medición del nivel de intensidad de tramas de tiempo dentro del sistema de referencia escalado, a continuación denominado también como una detección del nivel de trama o medición de nivel de trama. En la acción de detección de nivel de trama, se determina un parámetro relacionado al nivel de intensidad de las tramas de tiempo dentro de la señal de referencia escalada. Tal parámetro de nivel de intensidad puede por ejemplo corresponder a un nivel de intensidad promedio por trama de tiempo o a un nivel de intensidad de pico medido por trama.

Los parámetros de nivel de intensidad determinados de las tramas de tiempo dentro de la señal de referencia escalada son utilizados para seleccionar tramas de tiempo dentro de la señal de salida que han de ser incluidas en uno o más cálculos con vistas a la supresión de ruido. En una realización de selección, las tramas de tiempo dentro de las señales de referencia escalada son identificadas para las que el valor de intensidad se encuentra por debajo de un cierto valor, denominado a continuación como un valor de criterio de silencio. Una trama de tiempo dentro de la señal de referencia escalada para la que el valor de intensidad se encuentra por debajo del valor de criterio de silencio será denominada como trama de silencio. Las tramas de tiempo seleccionadas dentro de la señal de salida corresponden a las tramas de silencio dentro de la señal de referencia escalada. Preferiblemente, el proceso de selección progresa identificando una serie de tramas de silencio consecutivas, por ejemplo 8 tramas de silencio. Tales series de tramas de silencio consecutivas serán denominadas a continuación como un intervalo de silencio. El nivel de intensidad medido dentro de las tramas de silencio, y en particular las tramas de silencio dentro de un intervalo de silencio, expresa un nivel de ruido que está presente de manera inherente en la señal de referencia en consideración. En otras palabras, no hay influencia del dispositivo bajo ensayo.

Los cálculos con vistas a la supresión de ruido incluyen el cálculo de un parámetro de contraste de ruido. El parámetro de contraste de ruido es una medida del contraste de ruido dentro de una parte de la señal. El parámetro de contraste de ruido puede ser utilizado para controlar una supresión de ruido de al menos una de las señales de referencia y de salida dentro del dominio de sonoridad perceptivo.

El cálculo del parámetro de contraste de ruido puede comprender seleccionar un par de tramas de tiempo dentro de la señal de salida que corresponde a tramas de silencio dentro de un intervalo de silencio. Por ejemplo, un par de tramas de tiempo dentro de la señal de salida puede ser seleccionado que corresponde a la primera y cuarta tramas de tiempo en un intervalo de silencio de 8 tramas de tiempo consecutivas. Después de esta selección, puede determinarse un parámetro de fracción basado en los niveles de intensidad de las dos tramas de tiempo seleccionadas.

En una realización, el valor del nivel de intensidad de una primera trama de tiempo seleccionada del par seleccionado de tramas de tiempo dentro de la señal de salida es dividido por la segunda trama de tiempo del par o, viceversa, para obtener el parámetro de fracción. En otra realización, el parámetro de fracción, FP, puede ser obtenido por una división similar seguida por una acción de compresión, por ejemplo utilizando una energía, a saber

$$FP = \left( \frac{\text{Trama1deenergíaseleccionada} + \text{delta}}{\text{Trama2deenergíaseleccionada} + \text{delta}} \right)^\alpha \quad (2)$$

El coeficiente  $\alpha$  de energía puede ser por ejemplo de aproximadamente 0,35 y el parámetro delta, que es utilizado para evitar una división por cero, puede ser por ejemplo de aproximadamente 0,2.

El parámetro de contraste de ruido puede ser definido a continuación, siendo sustancialmente igual al parámetro de fracción FP si Trama1deenergíaseleccionada representa el nivel de energía o de intensidad de la primera trama seleccionada. El parámetro de contraste de ruido cuantifica entonces saltos desde niveles inferiores a más elevados de ruido de fondo.

- 5 El ejemplo del cálculo de un parámetro de contraste de ruido de acuerdo con una realización del invento descrito anteriormente es clarificado adicionalmente por el código del programa C proporcionado a continuación:

```
noiseContrastMax = 1.0;
for (frame = start; frame <= stop; frame++) {
if (originalSilent[frame] && originalSilent[frame+1] && originalSilent[frame+2] &&
originalSilent[frame+3] && originalSilent[frame+4] && originalSilent[frame+5] &&
originalSilent[frame+6] && originalSilent[frame+7]) {
hulp1 = aDistortedLoudness [frame]+0.2;
hulp2 = aDistortedLoudness [frame+3]+0.2;
if (hulp1>1.5) hulp1 = 1.5;
if (hulp2>1.5) hulp2 = 1.5;
hulpRatio = (hulp2)/(hulp1);
hulp1 = pow(hulpRatio,0.35);
if (hulp1>noiseContrastMax) noiseContrastMax = hulp1; } }
```

10 En este programa, el parámetro de contraste de ruido se refiere al máximo contraste de ruido denominado como *noiseContrastMax*. Adicionalmente, *originalSilent[frame]* se refiere a una trama de silencio dentro de la señal de referencia (siendo la trama la primera trama de un intervalo de silencio de 8 tramas de silencio y siendo la trama +7 la última trama de silencio de este intervalo de silencio), y *aDistortedLoudness* se refiere al nivel de sonoridad de la señal de salida dentro de la trama seleccionada correspondiente a una trama dentro del intervalo de silencio definido con referencia a la señal de referencia. En este ejemplo, el coeficiente de energía denominado como  $\alpha$  en la ecuación (2) es igual a 0,35, mientras el parámetro delta en la ecuación (2) es igual a 0,2.

15 Además, en este ejemplo, el parámetro de contraste de ruido es igual a un valor fijo predeterminado (es decir 1,0) si la fracción comprimida (*hulp1* consumo es definida en una pero la última línea) está por debajo del valor fijo predeterminado. Si la fracción comprimida excede del valor fijo predeterminado, el parámetro de contraste de ruido es igual a la fracción comprimida. En una realización, se busca el contraste máximo a lo largo de todos los intervalos de silencio.

20 En comparación al estado de los métodos de la técnica para determinar un indicador de calidad, por ejemplo la PESQ mostrada esquemáticamente en la fig. 2, el método esquemáticamente mostrado en la fig. 3 incluye además una o más acciones de supresión de ruido. El algoritmo de supresión de ruido que es utilizado en tales acciones de supresión de ruido incluye el uso del parámetro de contraste de ruido. Las acciones de supresión de ruido pueden ser realizadas sobre la señal de referencia escalada así como sobre la señal de salida. En la realización mostrada en la fig. 3, en el dominio de sonoridad perceptiva, la señal de referencia escalada y la señal de salida son sometidas a una acción de supresión de ruido de referencia y a una acción de supresión de ruido de salida respectivamente.

25 La supresión de ruido de referencia puede comprender calcular una densidad espectral de ruido promedio de referencia a lo largo de un número de tramas de tiempo basado en las mediciones de nivel de intensidad realizadas en la acción de detección del nivel de trama. Como se ha mencionado anteriormente, preferiblemente los datos relacionados a las tramas de tiempo que pertenecen a un intervalo de silencio son utilizados para obtener la densidad espectral de ruido promedio de referencia. En caso de ensayar un dispositivo que maneja señales de voz, el uso de los intervalos de silencio asegura que hay una influencia mínima de voz sobre la densidad espectral de ruido promedio de referencia. Adicionalmente, un factor de supresión de ruido de referencia puede ser determinado optimizando la correlación entre los experimentos de evaluación de calidad objetivo y subjetivo. El factor de supresión de ruido de referencia así determinado se cree que representa un nivel ideal de reducción de ruido con el fin de obtener un nivel de ruido bajo ideal para construir una señal de referencia ideal. Generalmente, este nivel será muy bajo, por ejemplo de aproximadamente 20 a 30 dB SPL en el caso de estado estacionario con poco ruido. El factor de supresión de ruido de referencia tiene generalmente un valor comprendido entre 0,2 y 0,6.

30 La supresión de ruido de referencia realizada sobre la señal de referencia escalada permite la supresión del ruido hasta un nivel de ruido que es considerado como un nivel de ruido bajo deseable. En una realización, el factor de supresión de ruido de referencia, la densidad espectral de ruido de referencia promedio y el parámetro de contraste de ruido son utilizados para calcular el nivel de ruido bajo deseado. En una realización el nivel de intensidad de la señal de referencia escalada en el dominio de frecuencia perceptivo es reducido multiplicando la densidad espectral de ruido de referencia promedio por el factor de supresión de ruido de referencia dividido por el parámetro de contraste de ruido.

45 En un programa C, la supresión de ruido con respecto a la señal de referencia puede parecer similar a lo siguiente:

```

factorReference = factorReference/noiseContrastMax;
for (nu = 0; nu < aNumberOfBands; nu++) {
for (frame = start; frame <= stop; frame++) {
PitchPowerDensityReference[frame][nu] -=
    factorReference*AverageSpectrumReferenceOverSilentReference[nu];
if (PitchPowerDensityReference[frame][nu] < 0.0)
    PitchPowerDensityReference[frame][nu] = 0.0; } }

```

5 La supresión de ruido de salida puede comprender calcular una densidad espectral de ruido de salida promedio a lo largo de un número de tramas de tiempo. La densidad espectral de ruido de salida promedio es calculada preferiblemente a lo largo de tramas que corresponden a tramas de silencio que pertenecen a un intervalo de silencio en la señal de referencia escalada y alineada en el tiempo por razones similares como se ha descrito con referencia a la supresión del nivel de ruido de referencia.

10 Adicionalmente se determina una constante de nivel de ruido bajo y un factor de supresión de ruido de salida. La constante de nivel de ruido bajo corresponde a un nivel de ruido bajo que parece siempre estar presente y no influye en la percepción de calidad de una señal. Con el fin de cuantificar el impacto del ruido, este nivel de ruido bajo es excluido de la supresión de ruido. La función del factor de supresión de ruido de salida es similar a la función del factor de supresión de ruido de referencia descrito con referencia a la señal de referencia escalada. Tanto la constante de nivel de ruido bajo como el factor de supresión de ruido de salida pueden ser determinados mediante optimización de la correlación entre los experimentos de evaluación de calidad objetivo y subjetivo.

15 La supresión del ruido de salida realizada sobre la señal de salida puede ser prevista para suprimir el ruido hasta un nivel de ruido representativo de la perturbación. La supresión del ruido de salida puede ser dependiente de la frecuencia.

20 El nivel de ruido representativo de la perturbación en una señal de salida puede ser calculado utilizando la constante de nivel de ruido bajo, el factor de supresión de ruido de salida y la densidad espectral de ruido de salida promedio en combinación con el parámetro de contraste de ruido. El parámetro de contraste de ruido puede ser utilizado para adaptar el valor de la constante de nivel de ruido bajo para compensar diferencias de contraste de ruido, por ejemplo mediante multiplicación de la constante de nivel de ruido bajo por el parámetro de contraste de ruido. Similarmente, el factor de supresión de ruido de salida puede ser adaptado por división por el parámetro de contraste de ruido.

En una realización el nivel de intensidad de la señal de salida en el dominio de frecuencia perceptivo es reducido por la densidad espectral de ruido de referencia promedio menos la constante de nivel de ruido bajo adaptada multiplicado por el factor de supresión de ruido de salida adaptado.

25 Un programa C ejemplar que muestra tal realización del algoritmo de supresión de ruido con respecto a la señal de salida se ha dado a continuación:

```

constantDistorted = constantDistorted *noiseContrastMax;
factorDistorted = factorDistorted/noiseContrastMax;
bandLow = 2.0; bandHigh = 20.0;
for (nu = 0; nu < aNumberOfBands; nu++) {
hulp1 = 1.0; hulp2 = 1.0;
if (centreBark [nu]<bandLow) hulp1=(5.0-2.0*centreBark [nu]);
if (centreBark [nu]>bandHigh) hulp2=(1.0+0.07*(bandHigh-centreBark[nu]));
hulp = AverageInputSpectrumDistortedOverSilentReference[nu]-constantDistorted*hulp2;
if (hulp<0.0) hulp = 0.0;
hulpPow = 0.8*pow((AverageSpectrumDistortedOverSilentReference[nu]+1.0),0.03);
If (hulpPow>1.2) hulpPow=1.2;
for (frame = start; frame <= stop; frame++) {
PitchPowerDensity[frame][nu] -= factorDistorted*hulp*hulpPow*hulp1;
if (PitchPowerDensityDistorted[frame][nu]<0.0)
    PitchPowerDensityDistorted[frame][nu]=0.0; } }

```

30 Finalmente, la señal de referencia y la señal de salida pueden ser sustraídas perceptivamente. Esto puede hacerse de un modo conocido a partir de la PESQ y descrito con referencia a la fig. 2. Es decir, un indicador representativo de la degradación completa,  $D_n$  y un indicador representativo de las degradaciones añadidas,  $DA_n$ , son determinados en paralelo.

El esquema como se ha mostrado en la fig. 3 permite una aproximación diferente con relación al cálculo de ambos indicadores. Es posible realizar el método como se ha mostrado en la fig. 3 dos veces, es decir, una vez para determinar

un indicador de calidad que representa calidad con respecto a la degradación completa, la otra vez para determinar un indicador de calidad que representa calidad con respecto a las degradaciones añadidas en comparación a la señal de referencia. Realizar el método dos veces permite la optimización de cálculos con respecto a diferentes tipos de distorsiones. Tal optimización puede mejorar considerablemente la correlación entre una calidad de voz medida objetivamente y una calidad de voz tal y como es obtenida en experimentos de calidad de audición subjetiva.

Por ejemplo, intervalos de silencio utilizados para la determinación del indicador de calidad para la degradación completa pueden ser determinados de manera diferente que los intervalos de silencio utilizados para determinar el indicador de calidad para la degradación añadida. En particular, el valor de umbral utilizado para identificar intervalos de silencio para la degradación completa, que pueden ser denominados como un primer valor de umbral, pueden ser inferiores que el valor de umbral utilizado para identificar intervalos de silencio utilizados para la degradación añadida. El último valor de umbral puede ser denominado como un segundo valor de umbral. Las tramas de tiempo de intervalo de silencio calificadas con respecto al primer valor de umbral pueden ser denominadas como tramas de súper silencio, mientras que las tramas de tiempo de intervalos de silencio calificadas con respecto al segundo valor de umbral pueden ser denominadas como tramas de silencio.

Utilizar realizaciones de método del invento de acuerdo con el método mostrado en la fig. 3 proporciona un incremento significativo en la correlación entre una calidad de voz medida objetivamente y la calidad de voz como es obtenida en experimentos de calidad de audición subjetivos.

La fig. 4 muestra esquemáticamente una realización del presente invento aplicada al método mostrado en la fig. 2. Después del tratamiento previo, a saber alineación de nivel, filtrado por IRS, identificación de retardo de tiempo y adaptación, la señal de referencia y la señal de salida son sometidas a la función de implementación de ventana, transformación rápida de Fourier, y deformación de frecuencia a la escala de tono para conseguir la transformación desde el dominio de tiempo hacia una representación de trama de tiempo en el dominio de frecuencia perceptivo. La detección del nivel de trama con respecto a la señal de referencia como se ha descrito con referencia a la fig. 3 puede ser realizada sobre la señal de referencia en cualquier lugar a lo largo de esta transformación de dominio, mostrada esquemáticamente por las líneas de puntos. Como resultado de la detección del nivel de trama, las tramas de tiempo pueden ser identificadas con un valor de nivel de intensidad que excede de un cierto valor de umbral, denominado a continuación como un valor de criterio de actividad de voz. Estas tramas de tiempo pueden ser utilizadas en el cálculo de compensación de frecuencia lineal de la señal de referencia. Además, después de la deformación de intensidad de la función de densidad de energía de tono de referencia y de la función de densidad de energía de tono de salida a la escala de sonoridad para obtener una función de densidad de sonoridad de referencia y una función de densidad de sonoridad de salida respectivamente, ambas funciones de densidad de sonoridad pueden ser sometidas a una acción de supresión de ruido basada en un parámetro de contraste de ruido como se ha descrito con referencia a la fig. 3. Como resultado de la supresión de ruido, el impacto del ruido sobre la calidad de voz puede ser modelado más precisamente, y pueden mejorar las predicciones de la percepción de la señal que ha de ser tratada, por ejemplo, una señal de voz.

La fig. 5 muestra esquemáticamente una realización del presente invento aplicada a otro método de determinación de un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio con respecto a una señal de referencia. En esta realización, son tenidas en cuenta las variaciones del nivel de intensidad. Por esta razón, la señal de salida no esta sujeta a una acción de alineación de nivel.

La señal de referencia y la señal de salida deben sufrir operaciones de tratamiento previo como el filtrado por IRS y retardo de tiempo como es utilizado en PESQ. Además, con el fin de utilizar la detección de nivel de trama, la señal de referencia es sometida a una acción de escalado global, que escala la señal de referencia hacia un nivel de intensidad fijo. El nivel de intensidad fijo coincide preferiblemente con un nivel de aproximadamente 73 dB SPL para un fragmento de voz presentado dióticamente o dicóticamente y con un nivel de aproximadamente 79 dB SPL para un fragmento de voz presentado monóticamente.

A lo largo de las diferentes operaciones en el proceso de transformar la señal de referencia y la señal de salida desde el dominio del tiempo hacia una representación en tramas de tiempo en el dominio de frecuencia perceptivo, la detección de intensidad de las tramas de tiempo puede ser realizada de un modo como se ha descrito con respecto a la fig. 3 y similar a como se ha utilizado en la realización relacionada con PESQ mostrada en la fig. 4. Es decir, en primer lugar las señales son divididas en tramas de tiempo utilizando una función de implementación de ventana, por ejemplo una ventana de Hanning, seguida por una FFT hacia el dominio de tiempo-frecuencia.

Después de las mediciones del nivel de trama, también denominadas como detección del nivel de tramas, la señal de referencia escalada es escalada hacia un nivel de intensidad relacionado a la señal de salida. Esta acción de escalado global puede utilizar un algoritmo que sólo compensa parcialmente la diferencia de nivel de intensidad entre la señal de referencia y la señal de salida. La diferencia que es dejada puede ser utilizada para estimar el impacto del nivel de intensidad sobre la calidad de transmisión percibida.

Después del escalado global hacia el nivel de intensidad de la señal de salida, la señal de referencia es sometida a un escalado local en el dominio de frecuencia perceptivo y una compensación de frecuencia parcial utilizando la misma aproximación que se ha descrito con referencia a PESQ en la fig. 2. Aunque en la realización mostrada en la fig. 5 el

escalado local es realizado con referencia a la señal de referencia, es igualmente posible aplicar esta operación de escalado local con respecto a la señal de salida, por ejemplo de un modo como se ha mostrado en la fig. 2. El objeto de la acción de escalado local se refiere a la compensación de variaciones de ganancia a corto plazo. Si la señal de referencia o la señal de salida ha de ser seleccionada puede depender de la aplicación específica. En general, la señal de referencia es compensada, debido a que la señal de referencia nunca es presentada a un ensayo sometido en mediciones de calidad subjetivas.

En una realización, la primera compensación de frecuencia parcial utiliza un algoritmo de escalada así denominado suave. En el algoritmo de escalado suave, la señal que ha de ser tratada, a saber o bien la señal de referencia o bien la señal de salida, es mejorada escalándola de tal modo que se compensan pequeñas desviaciones de energía, preferiblemente por trama de tiempo, mientras que desviaciones mayores son compensadas parcialmente, dependiendo de una relación de energía entre la señal de referencia y la señal de salida. Más detalles con respecto al uso de escalado suave pueden ser encontrados en la solicitud de patente Norteamericana 2005/159944, patente Norteamericana nº 7.313.517, y patente norteamericana nº 7.315.812, todas cedidas a la solicitante e incorporadas aquí como referencia.

Preferiblemente, se realiza a continuación una operación de excitación tanto sobre la señal de referencia como sobre la señal de salida para compensar la dispersión de componentes de frecuencia como resultado de la ejecución previa de la transformada rápida de Fourier con función de implementación de ventana con respecto a estas señales. Más detalles con respecto a la excitación realizada sobre señales con este propósito pueden ser encontradas por ejemplo en el artículo "Una Medida de Calidad de Audio perceptiva Basada en una Representación de Sonido psíquico- acústico" por J.G. Deererends y J.A. Stemerding, J. Audio Eng. Soc. Vol. 40, Nº 12 (1992) págs. 963 -978. En este artículo, la excitación es calculada y la calidad es determinada utilizando representaciones de excitación dispersadas. En una realización, la excitación calculada es entonces utilizada para derivar una curva de auto-enmascaramiento que a su vez puede ser utilizada para obtener una representación de tiempo-frecuencia afilada. En su forma más simple, la curva de auto-enmascaramiento corresponde a una fracción de la curva de excitación.

Después de una deformación de intensidad a escala de sonoridad como es utilizada en la PESQ, y descrita con referencia a la fig. 2, la señal de referencia y la señal de salida con escaladas localmente en el dominio de sonoridad. En primer lugar, son escaladas aquellas partes de la señal de referencia que son más ruidosas que la señal de salida. A continuación son escaladas partes de la señal de salida que son más ruidosas que la señal de referencia.

La separación de estas acciones de escalado local permite la implementación y/o manipulación por separado de variaciones de nivel debido a limitación de tiempo e impulsos. Si una parte de la señal de referencia es más ruidosa que una parte correspondiente de la señal de salida, esta diferencia puede ser debida a la limitación del tiempo, por ejemplo causado por una trama errónea. Con el fin de cuantificar el impacto perceptivo de la limitación del tiempo, la señal de referencia es escalada hacia abajo a un nivel que es considerado óptimo para el cálculo de la diferencia de perturbación (asimétrico). Esta acción de escalado local sobre la señal de salida también suprime ruido en la señal de salida hasta un nivel que es más óptimo para el cálculo de la diferencia de perturbación (asimétrica). El impacto del ruido sobre la calidad percibida subjetivamente puede ser estimado más precisamente combinando este escalado local con una acción de supresión de ruido sobre la señal de salida.

A continuación, es llevada a cabo una segunda compensación de frecuencia parcial. La segunda compensación de frecuencia parcial puede ser realizada de un modo similar a como en la PESQ, sin embargo, siendo utilizada ahora en el dominio de sonoridad. En una realización, la segunda compensación de frecuencia parcial utiliza un algoritmo de escalado suave como se ha descrito anteriormente con referencia a la primera compensación de frecuencia parcial.

Tanto la primera compensación de frecuencia parcial como la segunda compensación de frecuencia parcial pueden utilizar resultados de la detección de nivel de trama que es llevada a cabo sobre la señal de referencia. En una realización, la primera y segunda compensaciones de frecuencia parcial pueden utilizar una estimación de respuesta de frecuencia lineal del sistema bajo ensayo basado en todas las tramas para las que el valor de la señal de referencia de entrada es mayor que un valor de umbral, por ejemplo un valor de criterio de actividad de voz. Este valor de umbral puede corresponder a un umbral de audición absoluto.

Preferiblemente, en este punto, bandas altas tanto de la señal de referencia como de la señal de salida son establecidas a cero debido a que se desconectan para tener una influencia despreciable sobre la calidad de transmisión percibida que ha de ser determinada. Adicionalmente, los niveles de intensidad de las bandas bajas de la señal de salida son escalados localmente hacia los niveles de intensidad de bandas similares de la señal de referencia. Por ejemplo, todas las bandas relacionadas con Bark 23 y más altas pueden ser establecidas a cero, mientras que las bandas de Bark en la señal de salida relacionadas con Bark 0 a 5 pueden ser escaladas. Las bandas Bark 0 - 22 en la señal de referencia y las bandas de Bark relacionadas a Bark 6 a 22 en la señal de salida no son entonces sometidas a ninguna de estas operaciones.

Hasta este punto, los niveles de señal de la señal de salida no han sido cambiados significativamente, y niveles muy bajos de la señal de salida causarán ahora sólo diferencias marginales en la representación interna. Esto conduce a errores en la estimación de calidad.

- 5 Con este propósito, en primer lugar el nivel de sonoridad es escalado a un nivel de sonoridad fijo, también denominado como el nivel de sonoridad interno fijo. Si se utiliza una calibración del nivel global de comienzo para la señal de referencia como se ha descrito en las Recomendaciones P.861 y/o P.862 de la ITU-T, tal nivel interno global fijo se encuentra alrededor de 20, número de escalado relacionado con la sonoridad interna sin dimensiones. En segundo lugar, los niveles de la señal de referencia son escalados hacia los niveles correspondientes de la señal de salida.
- 10 Como resultado del escalado del nivel de sonoridad de la señal de salida, la diferencia de nivel de sonoridad entre la señal de salida y la señal de referencia es tal que no puede determinarse un indicador de calidad fiable. Para superar esta expectativa indeseable, el nivel de sonoridad de la señal de referencia necesita ser escalado también. Por ello, después de escalar el nivel de sonoridad de la señal de salida, la señal de referencia es escalada globalmente hacia una sonoridad relacionada a la señal de salida escalada. A continuación tanto la señal de referencia como la señal de salida tienen un nivel de sonoridad que puede ser utilizado para calcular las representaciones internas relevantes perceptivamente necesarias para obtener una medida objetiva de la calidad de transmisión de un sistema de audio.
- 15 En algunas realizaciones, las acciones de escalado global en el dominio de sonoridad utilizan información obtenida en la acción de detección de nivel de trama. Por ejemplo, el escalado puede ser realizado solamente en tramas de tiempo para las que el nivel de intensidad excede de un cierto valor de umbral, por ejemplo el valor de criterio de actividad de voz.
- Finalmente, de manera similar al método descrito con referencia a la fig. 2, la señal de referencia y la señal de salida son sustraídas dando como resultado una señal diferencia. Después de la sustracción perceptiva, puede ser derivada una medida de calidad percibida, por ejemplo de un modo como se ha mostrado en la fig. 2 y descrito en la Recomendación P.862 de la ITU-T.
- 20 Sin embargo, en contraste a la PESQ como se ha descrito en la Recomendación P-862 de la ITU-T, algunas realizaciones del invento, incluyendo las realizaciones mostradas en las figs. 3-5, pueden ser realizadas dos veces. Una vez para determinar un indicador de calidad representativo de todas las degradaciones, y la otra vez para determinar un indicador de calidad representativo de las degradaciones añadidas en comparación a la señal de referencia. Tal procedimiento ha sido descrito con más detalle con referencia a la fig. 3.
- 25 En la realización mostrada en la fig. 5, la supresión de ruido de la señal de referencia y de la señal de salida es introducida en dos etapas diferentes en el método. En primer lugar, después de la deformación de intensidad a la escala de sonoridad Sone, la señal de referencia puede ser sometida a una supresión de ruido para suprimir ruido hasta un nivel de ruido predeterminado. De manera similar, la señal de salida puede ser sometida a una acción de supresión de ruido. En este caso, la supresión de ruido puede estar prevista para suprimir ruido hasta un nivel de ruido representativo de la perturbación experimentada por el usuario del dispositivo bajo ensayo, por ejemplo el sistema de audio 10 en la fig. 1. Más detalles con respecto a estas acciones de supresión de ruido han sido descritos con referencia a la fig. 3.
- 30 En segundo lugar, tanto la señal de referencia como la señal de salida pueden ser sometidas además a una acción de supresión de ruido adicional después del escalado global. Se ha encontrado que tal supresión de ruido adicional después del escalado global mejora además la correlación entre una calidad de voz medida objetivamente y la calidad de voz según es obtenida en experimentos de calidad de audición subjetivos.
- 35 Realizaciones de las acciones de supresión de ruido pueden ser similares a las acciones de supresión de ruido descritas anteriormente con referencia a la fig. 3. Los parámetros, en particular el factor de supresión de ruido de referencia, la constante de ruido de salida y el acoplamiento del parámetro de contraste de ruido pueden ser diferentes para obtener resultados óptimos.
- 40 El método de determinación de la calidad de transmisión de un sistema de audio, como se ha presentado con referencia a la fig. 5 proporciona predicciones más exactas en caso de que el sistema de audio funcione a niveles de intensidad de audición variables como resultado de la introducción de un número de operaciones de escalado global. Las operaciones de supresión de ruido, como se ha mencionado anteriormente, permiten una predicción mejorada del impacto del ruido sobre la calidad de voz cuando este impacto varía dependiendo del nivel de señal local.
- 45 El invento ha sido descrito con referencia a ciertas realizaciones descritas anteriormente. Se reconocerá que estas realizaciones son susceptibles de distintas modificaciones y formas alternativas bien conocidas para los expertos en la técnica.

**REIVINDICACIONES**

1. Método para determinar un indicador de calidad que representa una calidad percibida de una señal de salida de un dispositivo de audio, con respecto a una señal de referencia, en el que la señal de referencia y la señal de salida son tratadas y comparadas, y el tratamiento incluye dividir la señal de referencia y la señal de salida en tramas de tiempo mutuamente correspondientes, en el que el tratamiento comprende además:
- 5           – escalar la señal de referencia hacia un nivel de intensidad fijo;
  - seleccionar tramas de tiempo de la señal de salida basado en mediciones realizadas sobre la señal de referencia escalada;
  - 10          – calcular un parámetro de contraste de ruido basado en las tramas de tiempo seleccionadas de la señal de salida, siendo el parámetro de contraste de ruido una medida del contraste de ruido dentro de la señal;
  - aplicar una supresión de ruido en al menos una de la señal de referencia y de la señal de salida en el dominio de sonoridad perceptiva basado en el parámetro de contraste de ruido;
  - sustraer perceptivamente las señales de referencia y la señal de salida para formar una señal diferencia;
  - derivar el indicador de calidad desde la señal de diferencia;
  - 15          – caracterizado por que la operación de calcular el parámetro de contraste de ruido comprende:
    - determinar una fracción dividiendo el valor de intensidad de una primera trama de tiempo seleccionada dentro de la señal de salida por el valor de intensidad de una segunda trama de tiempo seleccionada dentro de la señal de salida;
    - comprimir la fracción utilizando una ley de energía con un coeficiente de compresión predeterminado de modo que se obtenga el parámetro de contraste de ruido.
  - 20
2. El método de la reivindicación 1, en el que el dispositivo de audio es un sistema de tratamiento de voz.
3. Método de la reivindicación 1, en el que el parámetro de contraste de ruido es establecido a un valor fijo predeterminado si la fracción comprimida está por debajo del valor fijo predeterminado.
4. Método de cualquiera de las reivindicaciones precedentes, en el que aplicar la supresión de ruido sobre la señal de referencia comprende:
- 25           – calcular una densidad espectral de ruido promedio a lo largo de un número de tramas de tiempo que tiene un nivel de intensidad detectado por debajo de un valor de umbral;
  - determinar un factor de supresión de ruido de referencia basado en una correlación optimizada entre información de evaluación de calidad objetiva y subjetiva; y
  - 30          – calcular un nivel de ruido deseado basado en el factor de supresión de ruido de referencia y en la densidad espectral de ruido promedio de referencia.
5. Método de cualquiera de las reivindicaciones precedentes, en el que aplicar la supresión de ruido sobre la señal de salida comprende:
- 35           – calcular una densidad espectral de ruido de salida promedio a lo largo de un número de tramas del tiempo correspondientes a tramas del tiempo dentro de la señal de referencia que tienen un nivel de intensidad detectado por debajo de un valor de umbral;
  - determinar una constante del nivel de ruido bajo basado en una correlación optimizada entre información de evaluación de calidad objetiva y subjetiva;
  - 40          – determinar un factor de supresión de ruido de salida basado en una correlación optimizada entre información de evaluación de calidad objetiva y subjetiva; y
  - calcular un nivel de ruido representativo de perturbación en la señal de salida utilizando la constante del nivel de ruido bajo, el factor de supresión de ruido de salida y la densidad espectral de ruido de salida promedio.
6. Método de cualquiera de las reivindicaciones precedentes, en el que el tratamiento comprende además:
- 45           – transformar la señal de referencia y la señal de salida desde el dominio del tiempo hacia el dominio de tiempo-frecuencia;

- derivar una función de densidad de energía de tono de referencia a partir de la señal de referencia y derivar una función de densidad de energía de tono de salida a partir de la señal de salida;
- escalar localmente la función de densidad de energía de tono de referencia para obtener una función de densidad de energía de tono de referencia escalada localmente;
- 5     – compensar parcialmente o bien la función de densidad de energía de tono de salida o bien la función de densidad de energía de tono de referencia escalada localmente con respecto a la frecuencia; y
- derivar una función de densidad de sonoridad de referencia y una función de densidad de sonoridad de salida.

7. Método de la reivindicación 6, en el que el tratamiento comprende además:

- 10     – escalar la intensidad de la función de densidad de energía de tono de referencia a partir del nivel de intensidad fijo hacia un nivel de intensidad relacionado a la función de densidad de energía de tono de salida para obtener una diferencia de nivel de intensidad entre las funciones de densidad de energía de tono que permiten la cuantificación del impacto del nivel de intensidad sobre la calidad percibida;
- escalar la sonoridad de la función de densidad de sonoridad de salida hacia un nivel de sonoridad fijo en el dominio de sonoridad perceptivo; y
- 15     – escalar la sonoridad de la función de densidad de sonoridad de referencia a partir de un nivel de sonoridad correspondiente al nivel de intensidad relacionado con la salida hacia un nivel de sonoridad relacionado con el nivel de sonoridad de la función de densidad de sonoridad de salida en el dominio de sonoridad perceptivo, para obtener una diferencia de nivel de sonoridad entre la función de densidad de sonoridad de referencia y la función de densidad de sonoridad de salida que permita la cuantificación del impacto de sonoridad sobre la
- 20     calidad percibida de la señal de salida.

8. Método de cualquiera de las reivindicaciones precedentes, en el que seleccionar las tramas de tiempo comprende seleccionar tramas de tiempo de la señal de salida correspondientes a tramas de tiempo de la señal de referencia escalada que satisfacen una condición predeterminada.

25     9. Método de la reivindicación 8, en el que dicha condición predeterminada incluye que la trama de tiempo de la señal de referencia escalada, una trama de silencio, tiene un valor del nivel de intensidad por debajo del valor de umbral.

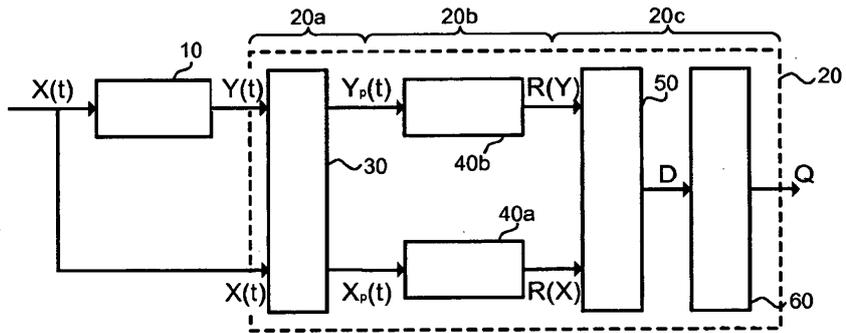
10. Método de la reivindicación 9, en el que dicha condición predeterminada incluye además que la trama de tiempo de la señal de referencia escalada es parte de una serie de tramas de silencio consecutivas, un intervalo de silencio.

30     11. Producto de programa informático que comprende un código ejecutable por ordenador, por ejemplo almacenado en un medio legible por ordenador, adaptado para realizar, cuando es ejecutado por un procesador, el método según ha sido definido por cualquiera de las reivindicaciones 1-10.

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

- un dispositivo (30) de tratamiento previo para tratar previamente la señal de referencia y la señal de salida;
- 35     – un primer dispositivo de tratamiento (40a) para tratar la señal de referencia, y un segundo dispositivo de tratamiento (40b) para tratar la señal de salida para obtener señales de representación  $R(X)$ ,  $R(Y)$  para la señal de referencia y la señal de salida respectivamente;
- un dispositivo de diferenciación (50) para combinar las señales de representación de la señal de referencia y de la señal de salida de modo que se obtenga una señal diferencial  $D$ ; y
- 40     – un dispositivo de modelado (60) para tratar la señal diferencial para obtener una señal de calidad  $Q$  que representa una estimación de la calidad efectiva del sistema de tratamiento de voz;

en el que el dispositivo de tratamiento previo, el primer dispositivo de tratamiento, y el segundo dispositivo de tratamiento forman un sistema de tratamiento para realizar el método de cualquiera de las reivindicaciones 1-10.



**FIG. 1**

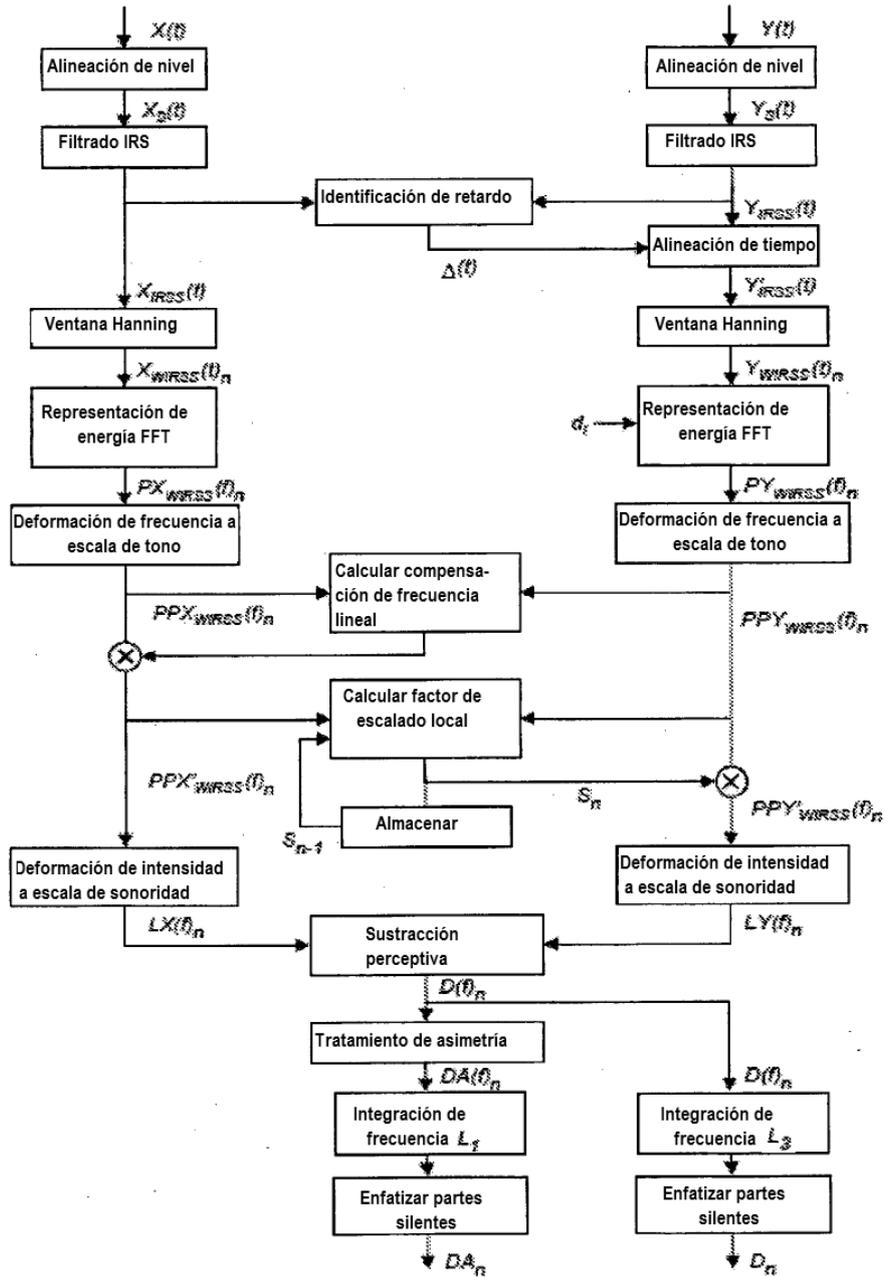
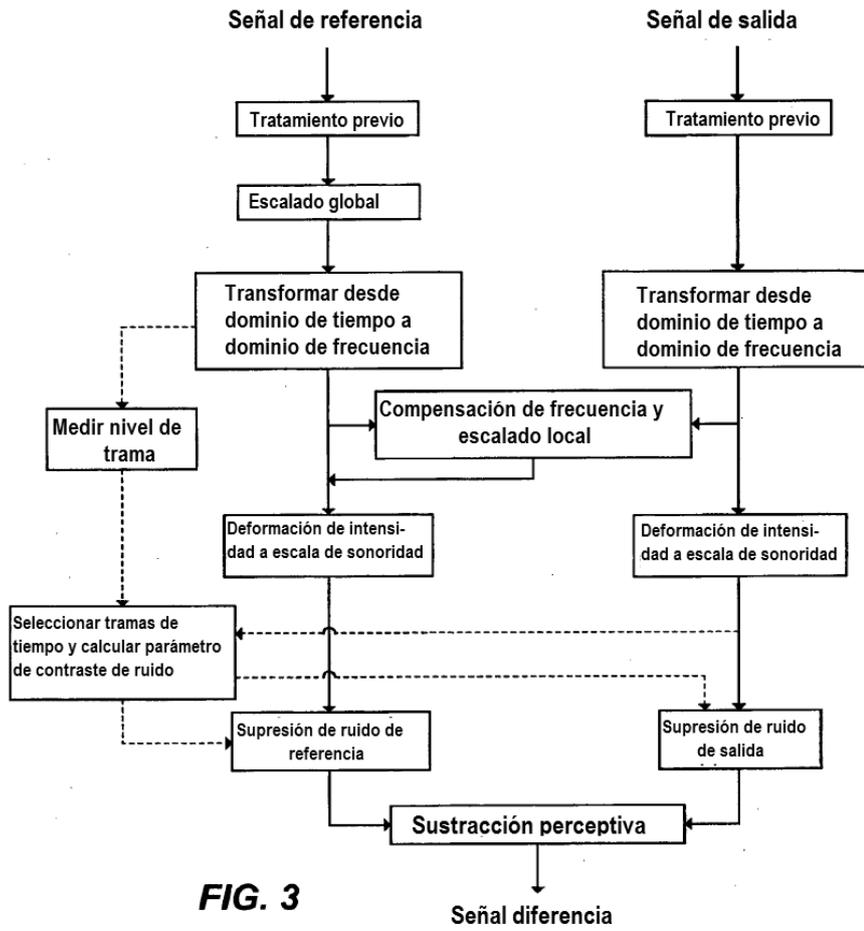


FIG. 2



**FIG. 3**

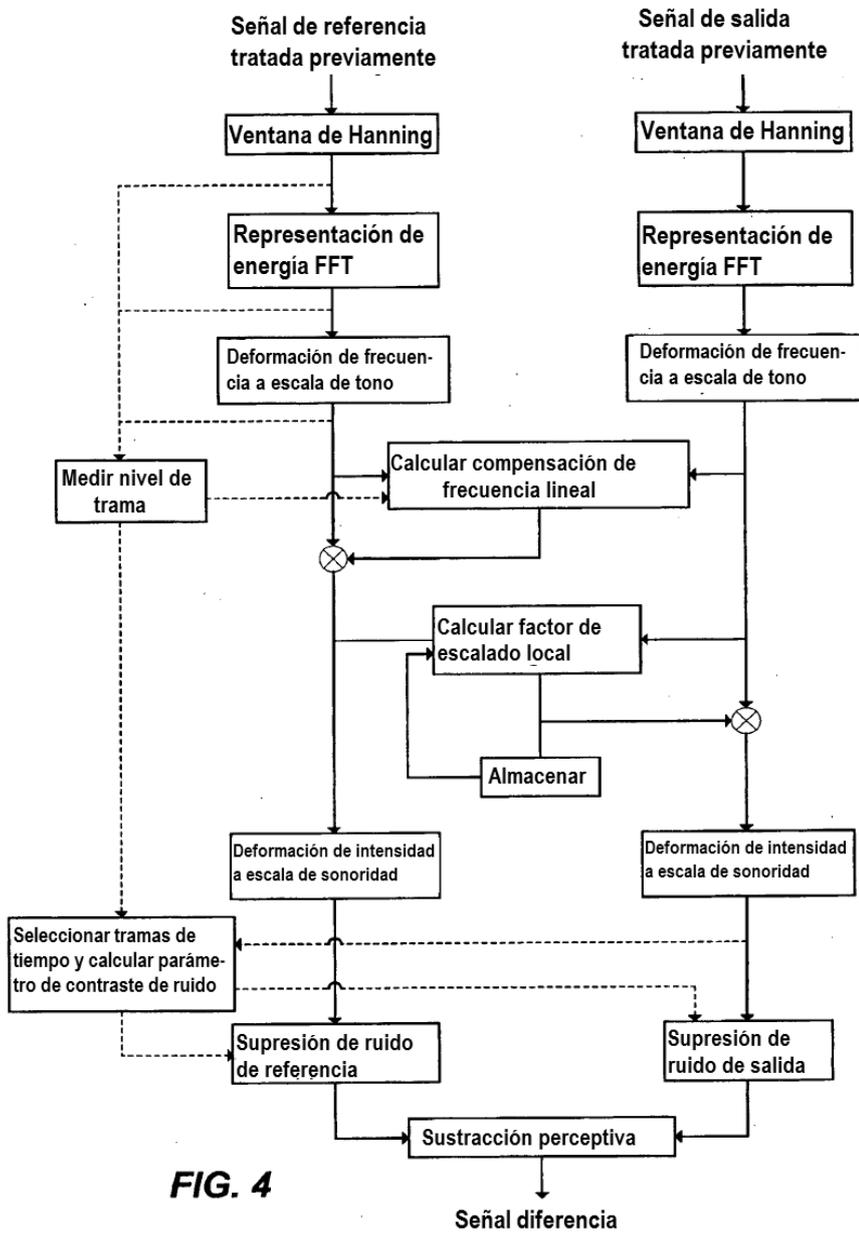


FIG. 4

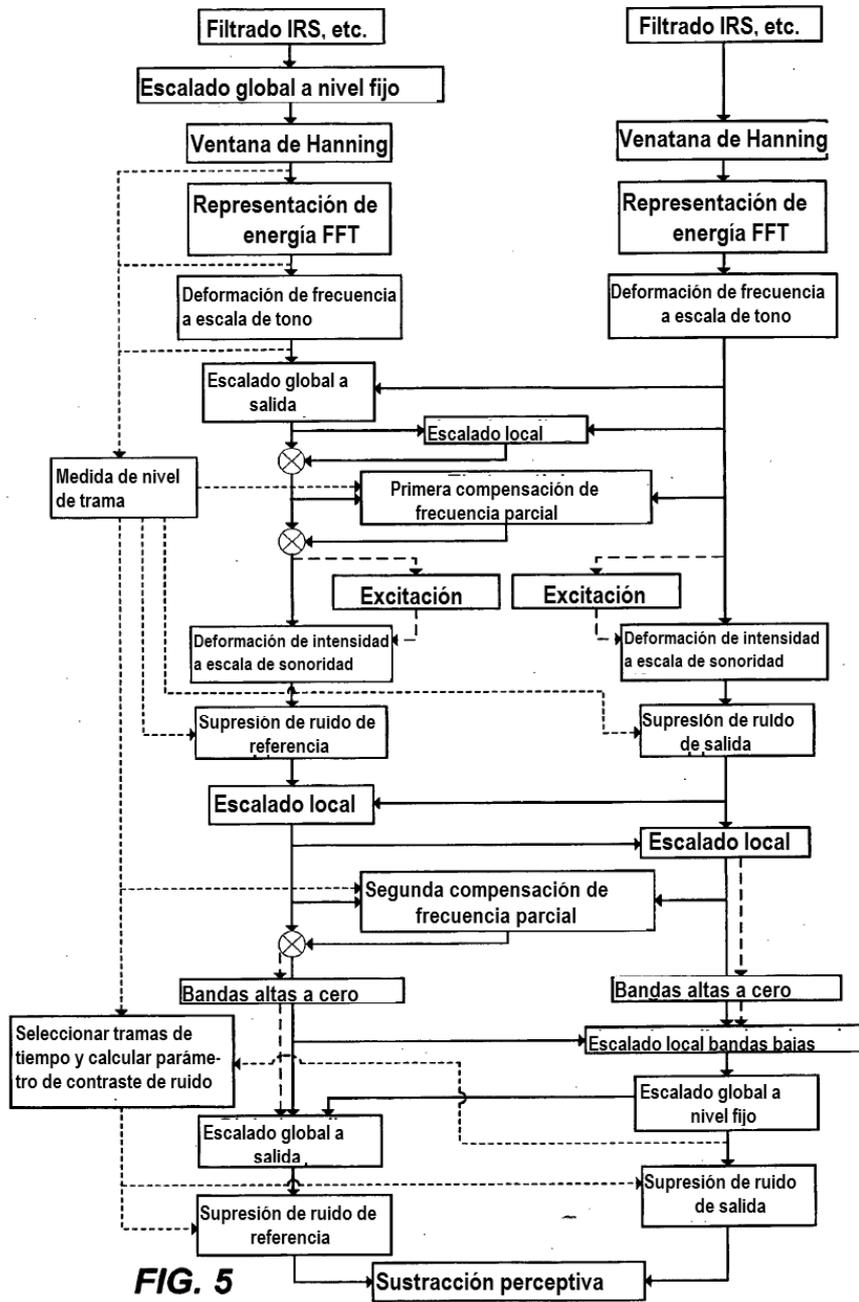
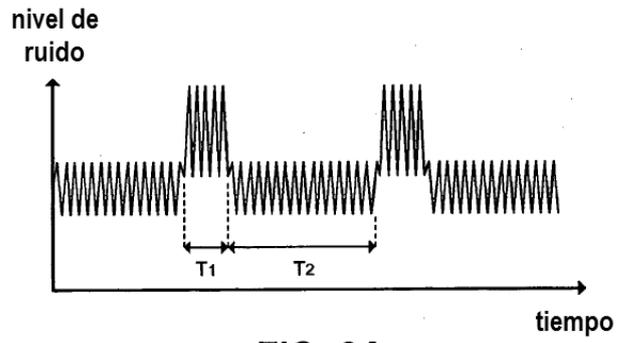
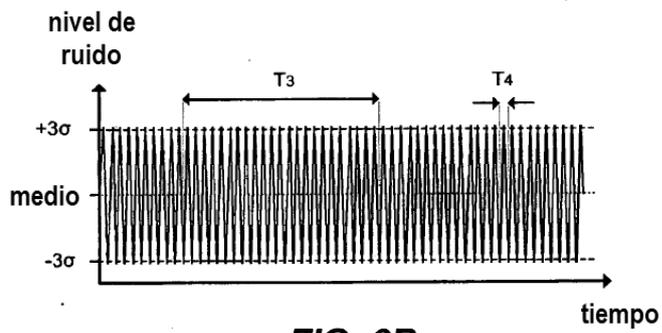


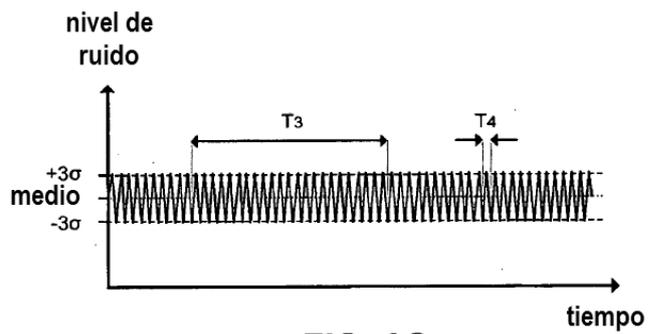
FIG. 5



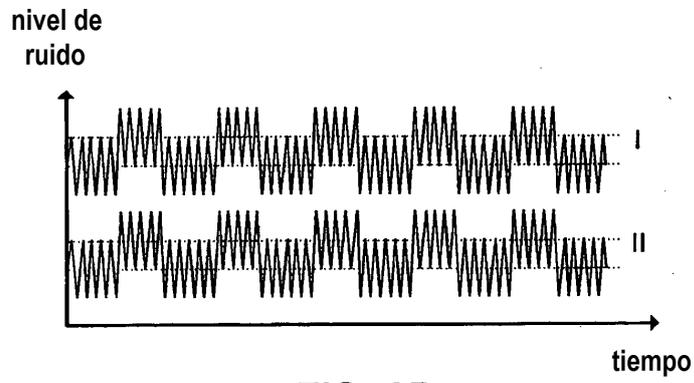
**FIG. 6A**



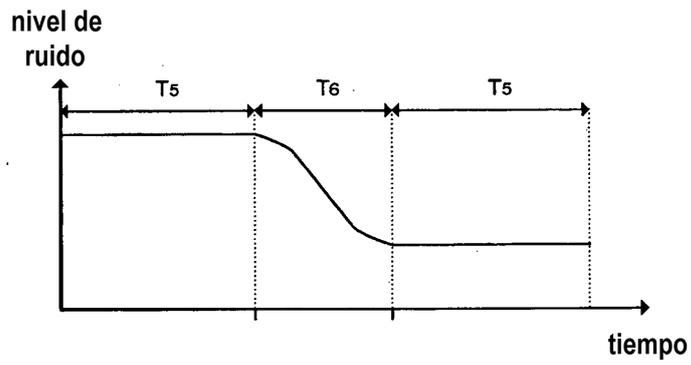
**FIG. 6B**



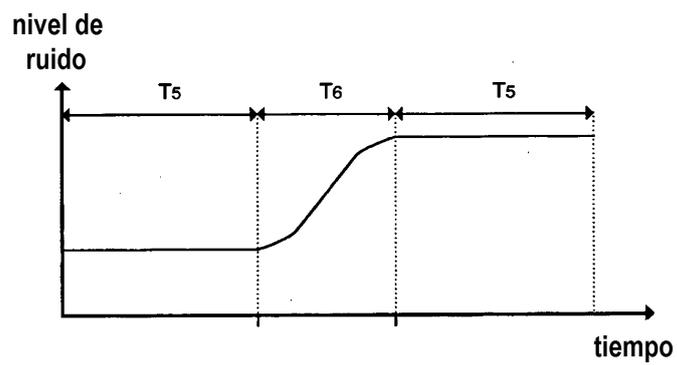
**FIG. 6C**



**FIG. 6D**



**FIG. 6E**



**FIG. 6F**