

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 768 719**

51 Int. Cl.:

G10L 25/03 (2013.01)

G10L 25/21 (2013.01)

G10L 19/012 (2013.01)

G10L 21/0216 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **21.07.2015 PCT/EP2015/066657**

87 Fecha y número de publicación internacional: **04.02.2016 WO16016051**

96 Fecha de presentación y número de la solicitud europea: **21.07.2015 E 15739587 (2)**

97 Fecha y número de publicación de la concesión europea: **20.11.2019 EP 3175457**

54 Título: **Método para estimar ruido en una señal de audio, estimador de ruido, codificador de audio, decodificador de audio, y sistema para transmitir señales de audio**

30 Prioridad:

28.07.2014 EP 14178779

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

23.06.2020

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**SCHUBERT, BENJAMIN;
JANDER, MANUEL;
LOMBARD, ANTHONY;
DIETZ, MARTIN y
MULTRUS, MARKUS**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 768 719 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 para estimar ruido en una señal de audio, estimador de ruido, codificador de audio, decodificador de audio, y sistema para transmitir señales de audio

5 La presente invención se refiere al campo de procesar señales de audio, más específicamente a un enfoque para estimar ruido en una señal de audio, por ejemplo, en una señal de audio a ser codificada o en una señal de audio que ha sido codificada. Hay realizaciones que describen un método para estimar ruido en una señal de audio, un estimador de ruido, un codificador de audio, un decodificador de audio y un sistema para transmitir señales de audio.

10 En el campo de procesar señales de audio, por ejemplo, para codificar señales de audio o para procesar señales de audio decodificadas, hay situaciones en que se desea estimar el ruido. Por ejemplo, los documentos PCT/EP2012/077525 y PCT/EP2012/077527 describen usar un estimador de ruido, por ejemplo un estimador de ruido estadístico mínimo, para estimar el espectro del ruido de fondo en el dominio de la frecuencia. La señal que es alimentada al algoritmo ha sido transformada en términos de bloques de datos al dominio de la frecuencia, por ejemplo, mediante una transformada rápida de Fourier (FFT) o cualquier otro banco de filtros adecuado. El entramado usualmente es idéntico al entramado del códec, esto es, se pueden volver a usar las transformadas ya existentes en el códec, por ejemplo, en un codificador EVS (servicios mejorados de voz) la FFT usada para el procesamiento. Con el propósito de la estimación de ruido, se computa el espectro de potencia de la FFT. Se agrupa el espectro en bandas psicoacústicamente motivadas y se acumulan las bandejas espectrales de potencia que están dentro de una banda para formar un valor de energía por banda. Finalmente, se alcanza un conjunto de valores de energía mediante este enfoque, el cual también con frecuencia se usa para procesar psicoacústicamente la señal de audio. Cada banda tiene su propio algoritmo de estimación de ruido, esto es, en cada trama se procesa el valor de energía de esa trama usando el algoritmo de estimación de ruido que analiza la señal a lo largo del tiempo y da un nivel de ruido estimado para banda en cualquier trama dada.

25 La resolución de muestra usada para señales de fragmento hablado y audio de alta calidad puede ser 16 bits, esto es la señal tiene una relación de señal a ruido (SNR) de 96dB. Computar el espectro de potencia significa transformar la señal al dominio de la frecuencia y calcular el cuadrado de cada bandeja de frecuencia. Debido a la función cuadrado, esto requiere un rango dinámico de 32 bits. La suma de varias bandejas del espectro de potencia en bandas requiere capacidad de aumento adicional para el rango dinámico porque la distribución de energía dentro de la banda realmente es desconocida. Como resultado, se necesita soportar un rango dinámico de más que 32 bits, típicamente alrededor de 40 bits, para ejecutar el estimador de ruido en un procesador.

35 En dispositivos que procesan señales de audio, los cuales funcionan sobre la base de energía recibida de una unidad de almacenamiento de energía, como una batería, por ejemplo, dispositivos portátiles como teléfonos móviles, para preservar energía, es esencial un procesamiento de las señales de audio eficiente en cuanto energía, para que la batería dure más. De acuerdo con enfoques conocidos, el procesamiento de señales de audio es realizado por procesadores de punto los cuales, típicamente, soportan procesamiento de datos en un formato de punto fijo de 16 o 32 bit. La más baja complejidad para el procesamiento se logra procesando datos de 16bit, mientras que procesar datos de 32bit ya requiere algo de sobrecarga. Procesar datos con rango dinámico de 40 bits requiere dividir los datos en dos, a saber, mantisa y exponente, ambos de los cuales deben ser tratados cuando se modifican los datos los cuales, a su vez, dan como resultado una complejidad computacional más elevada aún, y todavía más altas demandas de almacenamiento.

45 Se dan a conocer ejemplos adicionales de estimación de ruido en De Wet F et al., "Additive background noise as a source of non-linear mismatch in the cepstral and log-energy domain", XP004630841, y en Rotaru M et al., "An efficient GSC VSS-APA beamformer with integrated log-energy based VAD for noise reduction in speech reinforcement systems", XP032518224.

50 Empezando desde la técnica anterior discutida arriba, es un objetivo de la presente invención proveer un enfoque para estimar el ruido en una señal de audio de una manera eficiente usando un procesador de punto fijo que evita innecesaria sobrecarga computacional.

55 Este objetivo se alcanza mediante la materia del caso según se define en las reivindicaciones independientes.

La presente invención proporciona un método para estimar ruido en una señal de audio, como se expone en la reivindicación 1, y un estimador de ruido, como se expone en la reivindicación 8.

60 La estimación de ruido se puede llevar a cabo en base al algoritmo de estadística de cuadrados mínimos descrito por R. Martin en "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics", 2001.

Se pueden usar algoritmos de estimación de ruido alternativos, como el estimado de ruido basado en MMSE descrito por T. Gerkmann y R. C. Hendriks en "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay", 2012, o el algoritmo descrito por L. Lin, W. Holmes, y E. Ambikairajah, en "Adaptive noise estimation algorithm for speech enhancement", 2003.

5 La presente invención proporciona un producto de programa informático no transitorio, como se expone en la reivindicación 7.

10 La presente invención provee un codificador de audio, como se expone en la reivindicación 9.

La presente invención provee un decodificador de audio, como se expone en la reivindicación 10.

15 La presente invención proporciona un sistema para transmitir señales de audio, como se expone en la reivindicación 11.

Todas las siguientes incidencias de la palabra "realización/realizaciones", si hacen referencia a combinaciones de rasgos diferentes de los definidos por las reivindicaciones independientes, hacen referencia a ejemplos que se completaron originalmente pero que no representan realizaciones de la invención reivindicada actualmente; estos ejemplos todavía se muestran solo para fines ilustrativos.

20 La presente invención se basa en los hallazgos de los inventores de que, contrario a los enfoques convencionales en los cuales se ejecuta un algoritmo de estimación de ruido sobre datos de energía lineales, con el propósito de estimar niveles de ruido en material de audio/fragmento hablado, es posible ejecutar el algoritmo también sobre la base de datos de entrada logarítmicos. Para la estimación de ruido la demanda sobre la precisión de datos no es muy alta, por ejemplo, cuando se usan valores estimados para la generación de ruido de comodidad como se describe en el documento PCT/EP2012/077525 o PCT/EP2012/077527, se ha hallado que es suficiente estimar un nivel de ruido prácticamente correcto por banda, esto es, si se estima que el nivel de ruido es, por ejemplo, 0,1dB más alto o no, no será notable en la señal final. Así, mientras se pueden necesitar 40 bits para cubrir el rango dinámico de los datos, la precisión de los datos para señales de nivel medio/alto, en enfoques convencionales, es mucho mayor que la realmente necesaria. Sobre la base de estos hallazgos, de acuerdo con unas realizaciones, el elemento clave de la invención es convertir el valor de energía por banda al dominio logarítmico, preferiblemente el dominio del logaritmo en base 2 y para llevar a cabo la estimación de ruido, por ejemplo, sobre la base del algoritmo de estadística de cuadrados mínimos o cualquier otro algoritmo adecuado, directamente en un dominio logarítmico lo cual permite expresar los valores de energía en 16 bits lo cual, a su vez, permite un procesamiento más eficiente, por ejemplo usando un procesador de punto fijo.

En lo que sigue se describirán realizaciones de la presente invención haciendo referencia a los dibujos adjuntos, en los cuales:

40 la figura 1 muestra un diagrama de bloques simplificado de un sistema para transmitir señales de audio que implementa el enfoque inventivo para estimar ruido en una señal de audio a codificar o en una señal de audio decodificada,

45 la figura 2 muestra un diagrama de bloques simplificado de un estimador de ruido de acuerdo con una realización que se puede usar en un codificador de señal de audio y/o en un decodificador de señal de audio, y

la figura 3 muestra un diagrama de flujo que representa el enfoque inventivo para estimar ruido en una señal de audio de acuerdo con una realización.

50 En lo que sigue se describirán realizaciones del enfoque inventivo con más detalle y se hace notar que los elementos de dibujo acompañantes que tienen igual o similar funcionalidad, se denotan mediante los mismos signos de referencia.

55 La figura 1 muestra un diagrama de bloques simplificado de un sistema para transmitir señales de audio que implementa el enfoque inventivo en el lado del codificador y/o en el lado del decodificador. El sistema de la figura 1 comprende un codificador 100 que recibe en una entrada 102, una señal de audio 104. El codificador incluye un procesador 106 de codificación que recibe la señal 104 de audio y que genera una señal de audio codificada que es provista en una salida 108 del codificador. El procesador de codificación puede ser programado o compilado para procesar tramas de audio consecutivas de la señal de audio y para implementar el enfoque inventivo para estimar ruido en la señal 104 de audio a ser codificada. En otras realizaciones el codificador no necesita ser parte de un sistema de transmisión, como quiera que sea, puede ser un dispositivo independiente que genera señales de audio codificadas o puede ser parte de un transmisor de señal de audio. De acuerdo con una realización, el codificador 100 puede contener una antena 110 para permitir una transmisión inalámbrica de la señal de audio, como se indica

en 112. En otras realizaciones, el codificador 100 puede entregar la señal de audio codificada provista en la salida 108 usando una línea de conexión física, como se indica por ejemplo en el signo de referencia 114.

5 El sistema de la figura 1 comprende además un decodificador 150 que tiene una entrada 152 que recibe una señal de audio codificada a ser procesada por el decodificador 150, por ejemplo, vía la línea 114 física o vía una antena 154. El decodificador 150 comprende un procesador 156 de decodificación que opera sobre la señal codificada y que provee una señal 158 de audio decodificada en una salida 160. El procesador de decodificación puede ser programado o compilado para procesamiento para implementar el enfoque inventivo para estimar ruido en la señal 104 de audio decodificada. En otras realizaciones el decodificador no necesita ser parte de un sistema de 10 transmisión, más bien puede ser un dispositivo independiente para decodificar señales de audio codificadas o puede ser parte de un receptor de señal de audio.

La figura 2 muestra un diagrama de bloques simplificado de un estimador 170 de ruido de acuerdo con una realización. El estimador 170 de ruido se puede usar en un codificador de señal de audio y/o en un decodificador de 15 señal de audio mostrado en la figura 1. El estimador 170 de ruido incluye un detector 172 para determinar un valor 174 de energía para la señal 102 de audio, un convertidor 176 para convertir el valor 174 de energía al dominio logarítmico (ver valor 178 de energía convertido), y un estimador 180 para estimar un nivel 182 de ruido para la señal 102 de audio basado en el valor 178 de energía convertido. El estimador 170 puede ser implementado por un procesador común o por una pluralidad de procesadores programados o compilados para implementar la 20 funcionalidad del detector 172, el convertidor 176 y el estimador 180.

En lo que sigue se describirán en más detalle realizaciones del enfoque inventivo que pueden ser implementadas en por lo menos uno del procesador 106 de codificación y el procesador 156 de decodificación de la figura 1, o por el 25 estimador 170 de la figura 2.

La figura 3 muestra un diagrama de flujo del enfoque inventivo para estimar ruido en una señal de audio. Se recibe una señal de audio y, en un primer paso S100 se determina un valor 174 de energía para la señal de audio, el cual, luego, en el paso S102, es convertido al dominio logarítmico. Sobre la base del valor 178 de energía convertido, en el paso S104, se estima el ruido. De acuerdo con unas realizaciones, en el paso S106 se determina si el 30 procesamiento ulterior de los datos de ruido estimados, los cuales están representados por datos 182 logarítmicos, deben estar en el dominio logarítmico o no. En caso de que se desee ulterior procesamiento en el dominio logarítmico (sí en el paso S106), los datos logarítmicos que representan el ruido estimado son procesados en el paso S108, por ejemplo, los datos logarítmicos son convertidos en parámetros de transmisión en caso de que ocurra transmisión también en el dominio logarítmico. Si no (no en el paso S106), los datos 182 logarítmicos, son 35 convertidos nuevamente a datos lineales en el paso S110, y los datos lineales son procesados en el paso S112.

De acuerdo con unas realizaciones, en el paso S100, determinar el valor de energía para la señal de audio puede hacerse como en los enfoques convencionales. El espectro de potencia de la FFT, la cual ha sido aplicada a la señal de audio, es computado y agrupado en bandas psicoacústicamente motivadas. Las bandejas espectrales de 40 potencia de dentro de una banda son acumuladas para formar un valor de energía por banda de modo que se obtiene un conjunto de valores de energía. En otras realizaciones, se puede calcular el espectro de potencia en base a cualquier transformada espectral adecuada, como la MDCT (Transformada Discreta de Coseno Modificada), un CLDFB (Banco de filtros complejo de bajo retardo), o una combinación de varias transformadas que cubren diferentes partes del espectro. En el paso S100 se determina el valor 174 de energía para cada banda, y se convierte el valor 174 de energía para cada banda al dominio logarítmico en el paso S102, de acuerdo con unas 45 realizaciones, en el dominio del logaritmo en base 2. Las energías de banda pueden ser convertidas al dominio del logaritmo en base 2 como sigue:

$$E_{n_log} = \frac{\lfloor (\log_2(1 + E_{n_lin})) \cdot 2^N \rfloor}{2^N}$$

- 50 LX] suelo (x),
- E_{n_log} valor de energía de la banda n en el dominio del logaritmo en base 2,
- E_{n_lin} valor de energía de la banda n en el dominio lineal,
- N resolución/precisión.

55 De acuerdo con unas realizaciones, se realiza la conversión al dominio del logaritmo en base 2, lo cual es ventajoso por el hecho de que la función (ent)log2 usualmente puede ser calculada muy rápidamente, por ejemplo, en un ciclo, en procesadores de punto fijo usando la función "norma" la cual determina el número de ceros delanteros en un número de punto fijo. Algunas veces se necesita una mayor precisión que (ent)log2, lo cual se expresa en la fórmula de arriba mediante la constante N. Esta precisión levemente mayor se puede lograr con una simple tabla de 60 búsqueda que tenga los bits más significativos después de la instrucción norma y una aproximación, los cuales son

enfoques comunes para lograr cálculo logarítmico de baja complejidad cuando es aceptable una menor precisión en la fórmula de arriba, la constante "1" dentro de la función log2 es agregada para asegurar que las energías convertidas quedan positivas. De acuerdo con unas realizaciones, esto puede ser importante en el caso de que el estimador de ruido se base en un modelo estadístico de la energía de ruido, ya que realizar una estimación de ruido sobre valores negativos quebrantaría un modelo tal y daría como resultado un comportamiento inesperado del estimador.

De acuerdo con una realización, en la fórmula de arriba N se fija en 6, lo cual es equivalente a $2^6 = 64$ bits de rango dinámico. Este es mayor que el rango dinámico descrito arriba de 40 bits y por lo tanto, es suficiente. Para procesar los datos el objetivo es usar datos de 16bit, lo cual deja 9 bits para la mantisa y un bit para el signo. Un formato así comúnmente se denota como un formato "6Q9". Alternativamente, como sólo se pueden considerar valores positivos, el bit de signo puede ser evitado y puede ser usado para la mantisa dejando un total de 10 bits para la mantisa, a lo cual se refiere como un formato "6Q10".

En el documento "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics" de R. Martin de 2001 se puede hallar una descripción detallada del algoritmo de estadística de cuadrados mínimos. Esencialmente consiste en rastrear los mínimos de un espectro de potencia alisado sobre una ventana temporal deslizante de una longitud dada para cada banda espectral, típicamente sobre un par de segundos. El algoritmo también incluye una compensación de desviación para mejorar la exactitud de la estimación de ruido. Asimismo, para mejorar el rastreo de un ruido que varía con el tiempo, se pueden usar mínimos locales computados sobre una ventana temporal mucho más corta en lugar de los mínimos originales, ya que eso produce un aumento moderado de las energías de ruido estimadas. En el documento "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics" de R. Martin de 2001 está determinada la magnitud tolerada del incremento, mediante el parámetro noise_slope_max. De acuerdo con una realización, se usa un algoritmo de estimación de ruido estadístico mínimo el cual, convencionalmente, se ejecuta sobre datos de energía lineales. Sin embargo, de acuerdo con los hallazgos de los inventores, con el propósito de estimar niveles de ruido en material de audio o material de fragmento hablado, en cambio, el algoritmo puede ser alimentado en cambio con datos de entrada logarítmicos. Si bien el procesamiento de señal en sí mismo permanece sin modificar, sólo se requiere un mínimo de reajuste, el cual consiste en disminuir el parámetro noise_slope_max para hacer frente al rango dinámico de los datos logarítmicos comparado con los datos lineales. Hasta ahora se supuso que hay que ejecutar el algoritmo estadístico mínimo, u otras técnicas de estimación de ruido adecuadas, sobre datos lineales, esto es, se supuso que no eran adecuados datos que en realidad son una representación logarítmica. Contrario es esta suposición convencional, los inventores hallaron que la estimación de ruido en realidad puede ejecutarse sobre la base de datos logarítmicos, lo cual permite usar datos de entrada que se representan en sólo 16 bits lo cual, como consecuencia, provee una mucho menor complejidad en implementaciones de punto fijo, ya que la mayoría de las operaciones se pueden hacer en 16 bits y sólo algunas partes del algoritmo todavía requieren 32 bits. En el algoritmo de estadístico mínimo, por ejemplo, la compensación de desviación se basa en la variación de la potencia ingresada, por ende un estadístico de cuarto orden que típicamente todavía requiere una representación de 32 bits.

Como se describió arriba con relación a la figura 3, el resultado del proceso de estimación de ruido puede ser procesado ulteriormente de diferentes maneras. De acuerdo con unas realizaciones, una primera manera es usar los datos 182 logarítmicos directamente, como se muestra en el paso S108, por ejemplo, convirtiendo directamente los datos 182 logarítmicos en parámetros de transmisión si estos parámetros son transmitidos en el dominio logarítmico también, lo cual es con frecuencia el caso. Una segunda manera es procesar los datos 182 logarítmicos tal que sean convertidos de nuevo al dominio lineal para ulterior procesamiento, por ejemplo, usando funciones de traslación las cuales usualmente son muy rápidas y típicamente requieren sólo un ciclo en un procesador, junto con una tabla de búsqueda o usando una aproximación, por ejemplo:

$$E_{n_lin} = 2^{(E_{n_log} - 1)}$$

En lo que sigue, se describirá un ejemplo detallado para implementar el enfoque inventivo para estimar ruido sobre la base de datos logarítmicos, con referencia a un codificador, si bien, como se destacó arriba, el enfoque inventivo también puede ser aplicado a señales que han sido decodificadas en un decodificador, como se describe por ejemplo en los documentos PCT/EP2012/077525 o PCT/EP2012/077527. La siguiente realización describe una implementación del enfoque inventivo para estimar el ruido en una señal de audio en un codificador de audio, como el codificador 100 de la figura 1. Más específicamente, se dará una descripción de un algoritmo de procesamiento de señal de un codificador de Servicios Mejorados de Voz (codificador de EVS) para implementar el enfoque inventivo para estimar el ruido en una señal de audio recibida en el codificador de EVS.

Se suponen bloques de entrada de muestras de audio de 20 ms de longitud en el formato de PCM (Modulación por Código de Pulsos) uniforme. Se suponen cuatro tasas de muestreo, por ejemplo, 8 000, 16 000, 32 000 y 48 000 muestras/s y las velocidades de transmisión de bits para la serie de bits codificada pueden ser de 5,9; 7,2; 8,0; 9,6;

13,2; 16,4; 24,4; 32,0; 48,0; 64,0 o 128,0 kbit/s. También se puede proveer un códec de modo interoperable AMR-WB (Banda ancha de Velocidad Multiadaptativa) que opera a velocidades de transmisión de bits para la serie de bits codificada de 6,6; 8,85; 12,65; 14,85; 15,85; 18,25; 19,85; 23,05 o 23,85 kbit/s.

5 Con el propósito de la siguiente descripción, se aplican las siguientes convenciones a las expresiones matemáticas:

LXJ indica el entero más grande menor o igual que x: $L1.1J=1$, $L1.0J=1$ y $L-1.1J=-2$;

\sum indica una suma;

10 A menos que se indique de manera diferente, $\log(x)$ denota el logaritmo en base 10 a lo largo de la siguiente descripción.

15 El codificador acepta señales de banda completa (FB), banda superancha (SWB), banda ancha (WB) o banda angosta (NB) muestreadas a 48, 32, 16 o 8 kHz. De manera similar, la salida del decodificador puede ser 48, 32, 16 o 8 kHz, FB, SWB, WB o NB. El parámetro R (8, 16, 32 o 48) se usa para indicar la velocidad de muestreo de entrada en el codificador o la velocidad de muestreo de salida en el decodificador.

20 La señal de entrada es procesada usando tramas de 20 ms. El retardo del códec depende de la velocidad de muestreo de la entrada y la salida. Para entrada de WB y salida de WB, el retardo algorítmico global es de 42,875 ms. Consiste en una trama de 20 ms, 1,875 ms retardo de filtros de remuestreo de entrada y salida, 10 ms para la vista de anticipación del codificador, 1 ms retardo pos-filtrado, y 10 ms en el decodificador para encargarse de la operación de suma superpuesta de codificación de transformada de estrato superior. Para entrada NB y salida NB, no se usan estratos más altos, pero se usa el retardo del decodificador de 10 ms para mejorar el rendimiento del códec en presencia de enmendadura de trama y para señales de música. El retardo algorítmico global para entrada NB y salida NB es 43,875 ms – una trama de 20 ms, 2 ms para filtro de remuestreo de entrada, 10 ms para vista de anticipación del codificador, 1,875 ms para el filtro de remuestreo de salida, y retardo de 10 ms en el decodificador. Si la salida se limita al estrato 2, el retardo de códec puede reducirse en 10 ms.

30 La funcionalidad general del codificador comprende las siguientes secciones de procesamiento: procesamiento común, CELP (Predicción Lineal Excitada por Código) modo de codificación, MDCT (Transformada Discreta de Coseno Modificada) modo de codificación, modos de codificación mutantes, información lateral de encubrimiento de enmendadura de trama, operación de DTX/CNG (Generador de Transmisión Discontinua/Ruido de Comodidad), opción interoperable AMR-WB, y codificación consciente de canal.

35 De acuerdo con la presente realización, el enfoque inventivo es implementado en la sección de operación de DTX/CNG. El códec está equipado con un algoritmo de detección de actividad de señal (SAD) para clasificar cada trama de entrada como activa o inactiva. Soporta una operación de transmisión discontinua (DTX) en la cual se usa un módulo de generación de ruido de comodidad del dominio de la frecuencia (FD-CNG) para aproximar y actualizar el estadístico del ruido de fondo a una velocidad de transmisión de bits variable. Así, la velocidad de transmisión durante períodos de señal inactiva es variable y depende del nivel estimado del ruido de fondo. Sin embargo, también se puede fijar la velocidad de actualización de CNG por medio de un parámetro de línea de comando.

45 Para poder producir un ruido artificial que se parezca al ruido de fondo ingresado real en términos de características espectrotemporales, el módulo de FD-CNG hace uso de un algoritmo de estimación de ruido para rastrear la energía del ruido de fondo presente en la entrada del codificador. Luego las estimaciones de ruido son transmitidas como parámetros en forma de tramas SID (Descriptor de Inserción de Silencio) para actualizar la amplitud de las secuencias aleatorias generadas en cada banda de frecuencia en el lado del decodificador durante fases inactivas.

50 El estimador de ruido FD-CNG descansa en un enfoque de análisis espectral híbrido. Las bajas frecuencias correspondientes al ancho de banda núcleo son convertidas mediante análisis de FFT de alta resolución, mientras que las restantes frecuencias más altas son capturadas mediante un CLDFB el cual exhibe una resolución espectral significativamente menor, de 400Hz. Notar que el CLDFB también se usa como una herramienta de remuestreo para reducir el muestreo de la señal de entrada a la frecuencia de muestreo del núcleo.

55 El tamaño de la trama de SID, sin embargo, en la práctica, es limitado. Para reducir el número de parámetros que describen el ruido de fondo, se promedian las energías de entrada entre grupos de bandas espectrales llamadas particiones en lo que sigue.

60 1. Energías de Partición Espectral

Las energías de partición son computadas separadamente para las bandas de FFT y las de CLDFB. Las $L_{SID}^{[FFT]}$ energías correspondientes a las particiones de FFT y las $L_{SID}^{[CLDFB]}$ energías correspondientes a las particiones de CLDFB entonces son concatenadas formando un solo conjunto E_{FD-CNG} del tamaño $L_{SID} = L_{SID}^{[FFT]} + L_{SID}^{[CLDFB]}$ que servirá como entrada al estimador de ruido descrito abajo (ver "2. Estimación de Ruido con FD-CNG").

5

1.1 Computación de las Energías de Partición del FFT

Las energías de partición para las frecuencias que cubren el ancho de banda núcleo se obtienen como

$$E_{FD-CNG}(i) = \frac{E_{CB}^{[0]}(i) + E_{CB}^{[1]}(i)}{2} H_{de-emph}(i) \quad i = 0, \dots, L_{SID}^{[FFT]} - 1$$

10

Donde $E_{CB}^{[0]}(i)$ y $E_{CB}^{[1]}(i)$ son las energías promedio en banda crítica i para las ventanas primera y segunda de análisis, respectivamente. El número de particiones de FFT $L_{SID}^{[FFT]}$ que capturan los rangos de ancho de banda núcleo está en el rango entre 17 y 21, de acuerdo con la configuración usada (ver "1.3 Configuraciones de Codificador de FD-CNG"). Se usan los pesos espectrales de deénfasis $H_{de-emph}(i)$ para compensar un filtro pasa

15

$$\{H_{de-emph}(0), \dots, H_{de-emph}(L_{SID}^{[FFT]} - 1)\} = \{9.7461, 9.5182, 9.0262, 8.3493, 7.5764, 6.7838, 5.8377, 4.8502, 4.0346, 3.2788, 2.6283, 2.0920, 1.6304, 1.2850, 1.0108, 0.7916, 0.6268, 0.5011, 0.4119, 0.3637\}.$$

20

1.2 Computación de las Energías de Partición del CLDFB

Las energías de partición para frecuencias por arriba del ancho de banda núcleo son computadas como

$$E_{FD-CNG}(i) = \frac{1}{16} \frac{1}{8(A_{CLDFB})^2} \frac{\sum_{j=j_{min}(i)}^{j_{max}(i)} E_{CLDFB}(j)}{j_{max}(i) - j_{min}(i) + 1} \quad i = L_{SID}^{[FFT]}, \dots, L_{SID}^{[FFT]} + L_{SID}^{[CLDFB]} - 1$$

Donde $j_{min}(i)$ y $j_{max}(i)$ son los índices de las bandas primera y última del CLDFB en la i -ésima partición, respectivamente, $E_{CLDFB}(j)$ es la energía total de la j -ésima banda del CLDFB, y A_{CLDFB} es un factor de escala.

25

La constante 16 se refiere al número de ranuras de tiempo en el CLDFB. En número de particiones del L_{CLDFB} depende de la configuración usada, como se describe abajo.

1.3 Configuraciones de Codificador de FD-CNG

30

La siguiente tabla lista el número de particiones y sus límites superiores para las diferentes configuraciones del FD-CNG.

Tabla 1: Configuraciones de la estimación de ruido del FD-CNG en el codificador

Velocidad es de transmisión de bits [kbps]	$L_{SID}^{[FFT]}$	$L_{SID}^{[CLDFB]}$	$f_{max}(i),$ $i = 0, \dots, L_{SID}^{[FFT]} - 1$ [Hz]	$f_{max}(i),$ $i = L_{SID}^{[FFT]}, \dots, L_{SID} - 1$ [Hz]	
NB	•	17	0	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3975	×

WB	≤ 8	20	0	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375	×
	$8 < \bullet \leq 13.2$	20	1	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375	8000
	> 13.2	21	0	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375, 7975	×
SWB/ FB	≤ 13.2	20	4	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375	8000, 10000, 12000, 14000
	> 13.2	21	3	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375, 7975	10000, 12000, 16000

Para cada partición $i = 0, \dots, L_{\text{SID}} - 1$, $f_{\text{max}}(i)$ corresponde a la frecuencia de la última banda de i -ésima partición. Los índices $j_{\text{min}}(i)$ y $j_{\text{max}}(i)$ de las bandas primera y última en cada partición espectral pueden obtenerse como una función de la configuración del núcleo, como sigue:

$$j_{\text{max}}(i) = \begin{cases} f_{\text{max}}(i) \frac{\text{core_FFT_length}}{\text{core_sampling_rate}} & i = 0, \dots, L_{\text{SID}}^{\text{[FFT]}} - 1 \\ j_{\text{max}}(L_{\text{SID}}^{\text{[FFT]}} - 1) + \frac{2f_{\text{max}}(i) - \text{core_sampling_rate}}{800} & i = L_{\text{SID}}^{\text{[FFT]}}, \dots, L_{\text{SID}} - 1 \end{cases}$$

$$j_{\text{min}}(i) = \begin{cases} f_{\text{min}}(0) \frac{\text{core_sampling_rate}}{\text{core_FFT_length}} & i = 0 \\ j_{\text{max}}(i - 1) + 1 & i > 0 \end{cases}$$

donde $f_{\text{min}}(0) = 50\text{Hz}$ es la frecuencia de la primera banda de la partición espectral. Por ende, el FD-CNG genera algo de ruido de comodidad por arriba de 50 Hz solamente.

10 2. Estimación de Ruido del FD-CNG

El FD-CNG descansa en un estimado de ruido para rastrear la energía del ruido de fondo presente en el espectro de entrada. Esto se basa mayormente en el algoritmo estadístico mínimo descrito por R. Martin, en "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics" de 2001. Sin embargo, para reducir el rango dinámico de las energías de entrada $\{E_{\text{FD-CNG}}(0), \dots, E_{\text{FD-CNG}}(L_{\text{SID}} - 1)\}$ y por ende facilitar la implementación de punto fijo del algoritmo de estimación de ruido (ver "2.1 Compresión del rango dinámico para energías de entrada"). Luego se usa la transformada inversa sobre las estimaciones de ruido resultantes para recuperar el rango dinámico original (ver «2.3 Expansión del rango dinámico para las energías de ruido estimadas»).

20 2.1 Compresión de rango dinámico para las energías de entrada

Las energías de entrada son procesadas mediante una función no lineal y son cuantificadas con resolución de 9-bit como sigue:

$$E_{MS}(i) = \frac{\lfloor \log_2 \left((1 + E_{FD-CNG}(i)) 2^9 \right) \rfloor}{2^9} \quad i = 0, \dots, L_{SID} - 1$$

5

2.2 Rastreo de ruido

En el documento "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics" de R. Martin de 2001 se puede hallar una descripción detallada del algoritmo de estadística de cuadrados mínimos. Esencialmente consiste en rastrear los mínimos de un espectro de potencia alisado sobre una ventana temporal deslizante de una longitud dada para cada banda espectral, típicamente sobre un par de segundos. El algoritmo también incluye una compensación de desviación para mejorar la exactitud de la estimación de ruido. Asimismo, para mejorar el rastreo de un ruido que varía con el tiempo, se pueden usar mínimos locales computados sobre una ventana temporal mucho más corta en lugar de los mínimos originales, ya que eso produce un aumento moderado de las energías de ruido estimadas. En el documento "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics" de R. Martin de 2001 está determinada la magnitud tolerada del incremento, mediante el parámetro noise_slope_max.

Las salidas principales del rastreador de ruido son las estimaciones $N_{MS}(i), i = 0, \dots, L_{SID} - 1$. Para obtener transiciones más suaves en el ruido de comodidad, se puede aplicar un filtro recursivo de primer orden, esto es $\bar{N}_{MS}(i) = 0.95 \bar{N}_{MS}(i) + 0.05 N_{MS}(i)$.

Además, la energía de entrada $E_{MS}(i)$ es promediada sobre las 5 últimas tramas. Esto se usa para aplicar un límite superior sobre $\bar{N}_{MS}(i)$ en cada partición espectral.

25

2.3 Expansión de rango dinámico para las energías de ruido estimadas

Las energías de ruido estimadas son procesadas mediante una función no lineal para compensar la compresión de rango dinámico descrita arriba:

30

$$N_{FD-CNG}(i) = 2^{\bar{N}_{MS}(i) - 1} \quad i = 0, \dots, L_{SID} - 1.$$

De acuerdo con la presente invención, se describe un enfoque mejorado para estimar ruido en una señal de audio, el cual permite reducir la complejidad del estimador de ruido, especialmente para señales de audio/fragmento hablado las cuales son procesadas en procesadores usando aritmética de punto fijo. El enfoque inventivo permite reducir el rango dinámico usado para el estimador de ruido para procesar señales de audio/fragmento hablado, por ejemplo en un entorno descrito en el documento PCT/EP2012/077527, el cual se refiere a la generación de un ruido de comodidad con alta resolución espectral-temporal, o en el documento PCT/EP2012/077527, el cual se refiere a suma de ruido de comodidad para modelar ruido de fondo a baja velocidad de transmisión de bits. En los escenarios descritos, se usa un estimador de ruido que opera sobre la base del algoritmo de estadístico mínimo para mejorar la calidad del ruido de fondo o para una generación de ruido de comodidad para señales de fragmento hablado ruidosas, por ejemplo fragmento hablado en presencia de ruido de fondo lo cual es una situación muy común en llamada telefónica y una de las categorías probadas del códec de EVS. El códec de EVS, de acuerdo con la estandarización, usará un procesador con aritmética fija, y el enfoque inventivo permite reducir la complejidad de procesamiento reduciendo el rango dinámico de la señal que es usado por el estimador de ruido estadístico mínimo procesando el valor de energía para la señal de audio en el dominio logarítmico y ya no en el dominio lineal.

A pesar de que algunos aspectos del concepto descrito han sido descritos en el contexto de un aparato, está claro que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a un paso de método o a un rasgo de un paso de método. Análogamente, los aspectos descritos en el contexto de una etapa de método también representan una descripción de un correspondiente bloque o componente o rasgo de un correspondiente aparato.

Dependiendo de ciertos requerimientos de implementación, las realizaciones del invento pueden ser implementadas en hardware o en software. La implementación puede ser realizada usando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blue-Ray, un CD, un ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, teniendo señales de control legibles electrónicamente almacenadas en las mismas, las cuales

55

cooperan (o son capaces de cooperar) con un sistema informático programable tal que se ejecute el respectivo método. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

5 Algunas realizaciones de acuerdo con la invención comprenden un portador de datos que tiene señales de control legibles electrónicamente, las cuales son capaces de cooperar con un sistema informático programable, tal que uno de los métodos descrito en la presente sea ejecutado.

10 Generalmente, realizaciones de la presente invención pueden ser implementadas como un programa informático con un código de programa, siendo código de programa operativo para ejecutar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede ser almacenado, por ejemplo, sobre un portador legible por una máquina.

15 Otras realizaciones comprenden el programa informático para ejecutar uno de los métodos descritos en la presente, almacenado en un portador legible por una máquina.

En otras palabras, una realización del método inventivo es, por lo tanto, un programa informático que tiene un código de programa para ejecutar uno de los métodos descritos en la presente, cuando el programa informático se ejecuta en un ordenador.

20 Una realización adicional de los métodos inventivos es, por lo tanto, un portador de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para ejecutar uno de los métodos descritos en la presente.

25 Una realización adicional del método inventivo es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para ejecutar uno de los métodos descritos en la presente. El flujo de datos o la secuencia de señales pueden ser configuradas, por ejemplo, para ser transferidos vía una conexión de comunicación de datos, por ejemplo, vía Internet.

30 Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado para o adaptado para ejecutar uno de los métodos descritos en la presente.

Una realización adicional comprende un ordenador que tiene instalado en ella el programa informático para ejecutar uno de los métodos descritos en la presente.

35 En algunas realizaciones se puede usar un dispositivo de lógica programable (por ejemplo un campo de matriz de puertas programables) para realizar algunas o todas las funcionalidades de los métodos descritos en la presente. En algunas realizaciones, el campo de matriz de puertas programables puede cooperar con un microprocesador para realizar uno de los métodos descritos en la presente. Generalmente, los métodos preferiblemente son realizados mediante algún aparato de hardware.

40 Las realizaciones que se describieron más arriba son puramente ilustrativas para los principios de la presente invención. Se entiende que las modificaciones y variaciones posibles de las disposiciones y de los detalles descritos en la presente serán evidentes para los expertos en la materia. Por lo tanto, es la intención que la invención esté limitada sólo por el alcance de las siguientes reivindicaciones de patente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en la presente.

45

REIVINDICACIONES

1. Método para estimar ruido en una señal (102) de audio, comprendiendo el método:
 5 determinar (S100) un valor (174) de energía para la señal (102) de audio;
 convertir (S102) el valor (174) de energía en el dominio del logaritmo en base 2; y
 10 estimar (S104) un nivel (182) de ruido para la señal (102) de audio basado en el valor (178) de energía convertido directamente en el dominio del logaritmo en base 2,
 en el que el valor (174) de energía está convertido (S102) en el dominio del logaritmo en base 2 como sigue:

$$E_{n_log} = \frac{\lfloor (\log_2(1 + E_{n_lin})) \cdot 2^N \rfloor}{2^N}$$

- LX] suelo (x), que indica el número entero más grande menor que o igual a x,
 20 E_{n_log} valor de energía de banda n en el dominio del logaritmo en base 2,
 E_{n_lin} valor de energía de banda n en el dominio lineal,
 N resolución de cuantificación.
 25 2. Método de acuerdo con la reivindicación 1, en donde estimar (S104) el nivel de ruido comprende ejecutar un algoritmo de estimación de ruido predefinido, como el algoritmo de estadístico mínimo.
 3. Método de acuerdo con la reivindicación 1 o 2, en donde determinar (S100) el valor (174) de energía comprende obtener un espectro de potencia de la señal (102) de audio transformando la señal (102) de audio en el dominio de la frecuencia, agrupar el espectro de potencia en bandas psicoacústicamente motivadas, y acumular las bandejas espectrales de potencia dentro de una banda para formar un valor (174) de energía para cada banda, en donde el valor (174) de energía para cada banda es convertido en el dominio del logaritmo en base 2, y en donde se estima un nivel de ruido para cada banda basado en el correspondiente valor (174) de energía convertido.
 30 4. Método de acuerdo con la reivindicación 3, en donde la señal (102) de audio (102) comprende una pluralidad de tramas, y en donde para cada trama se determina el valor (174) de energía y se convierte en el dominio del logaritmo en base 2, y se estima el nivel de ruido para cada banda de una trama basado en el valor (174) de energía convertido.
 40 5. Método de acuerdo con una de las reivindicaciones 1 a 4, en donde estimar (S104) el nivel de ruido basado en el valor (178) de energía convertido produce datos logarítmicos, y en donde el método comprende además:
 45 usar (S108) los datos logarítmicos directamente para ulterior procesamiento, o
 convertir (S110, S112) los datos logarítmicos de nuevo en el dominio lineal para ulterior procesamiento.
 50 6. Método de acuerdo con la reivindicación 5, en donde
 los datos logarítmicos son convertidos (S108) directamente en datos de transmisión, en caso de que se haga una transmisión en el dominio logarítmico, y
 55 convertir (S110) los datos logarítmicos directamente en datos de transmisión usa una función traslación junto con una tabla de búsqueda o una aproximación, por ejemplo $E_{n_lin} = 2(E_{n_log})$
 7. Producto de programa informático no transitorio que comprende un medio legible por ordenador que almacena instrucciones que, cuando son ejecutadas en un ordenador, llevan a cabo el método de una de las reivindicaciones 1 a 6.
 60

8. Estimador (170) de ruido, que comprende:

un detector (172) configurado para determinar un valor (174) de energía para la señal (102) de audio;

5 un convertidor (176) configurado para convertir el valor (174) de energía en el dominio del logaritmo en base 2; y

un estimador (180) configurado para estimar un nivel (182) de ruido para la señal (102) de audio basado en el valor (178) de energía convertido directamente en el dominio del logaritmo en base 2,

10 en el que el valor (174) de energía está convertido (S102) en el dominio del logaritmo en base 2 como sigue:

$$E_{n_log} = \frac{\lfloor (\log_2(1 + E_{n_lin})) \cdot 2^N \rfloor}{2^N}$$

15 $\lfloor x \rfloor$ suelo (x), que indica el número entero más grande menor que o igual a x,

E_{n_log} valor de energía de banda n en el dominio del logaritmo en base 2,

20 E_{n_lin} valor de energía de banda n en el dominio lineal,

N resolución de cuantificación.

9. Codificador (100) de audio, que comprende el estimador de ruido de acuerdo con la reivindicación 8.

25 10. Decodificador (150) de audio, que comprende el estimador (170) de ruido de acuerdo con la reivindicación 8.

11. Sistema para transmitir señales (102) de audio, comprendiendo el sistema:

30 un codificador (100) de audio configurado para generar señal (102) de audio codificada basado en una señal (102) de audio recibida; y

35 un decodificador (150) de audio configurado para recibir la señal (102) de audio codificada, para decodificar la señal (102) de audio codificada, y para entregar la señal (102) de audio decodificada,

donde por lo menos uno del codificador de audio y el decodificador de audio comprende el estimador (170) de ruido de acuerdo con la reivindicación 8.

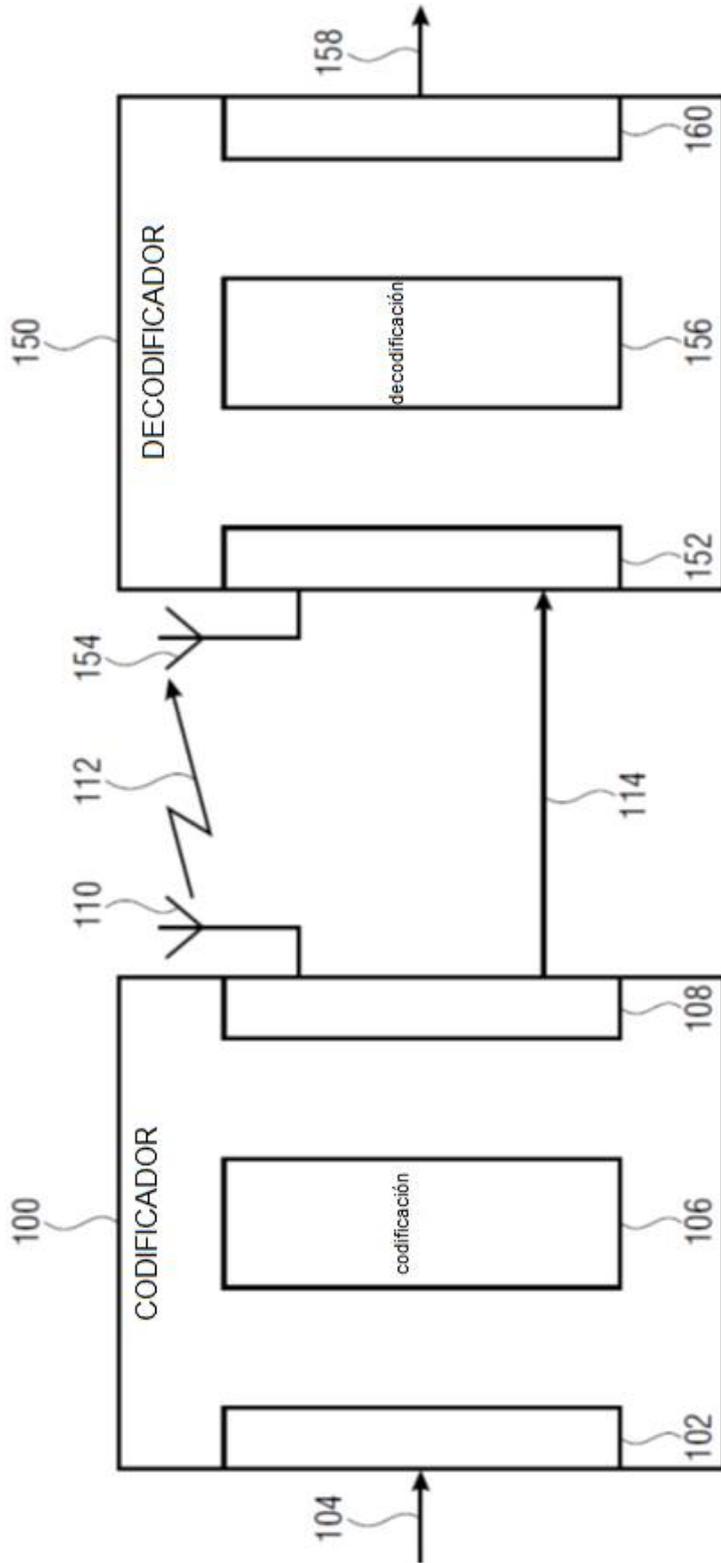


FIG 1

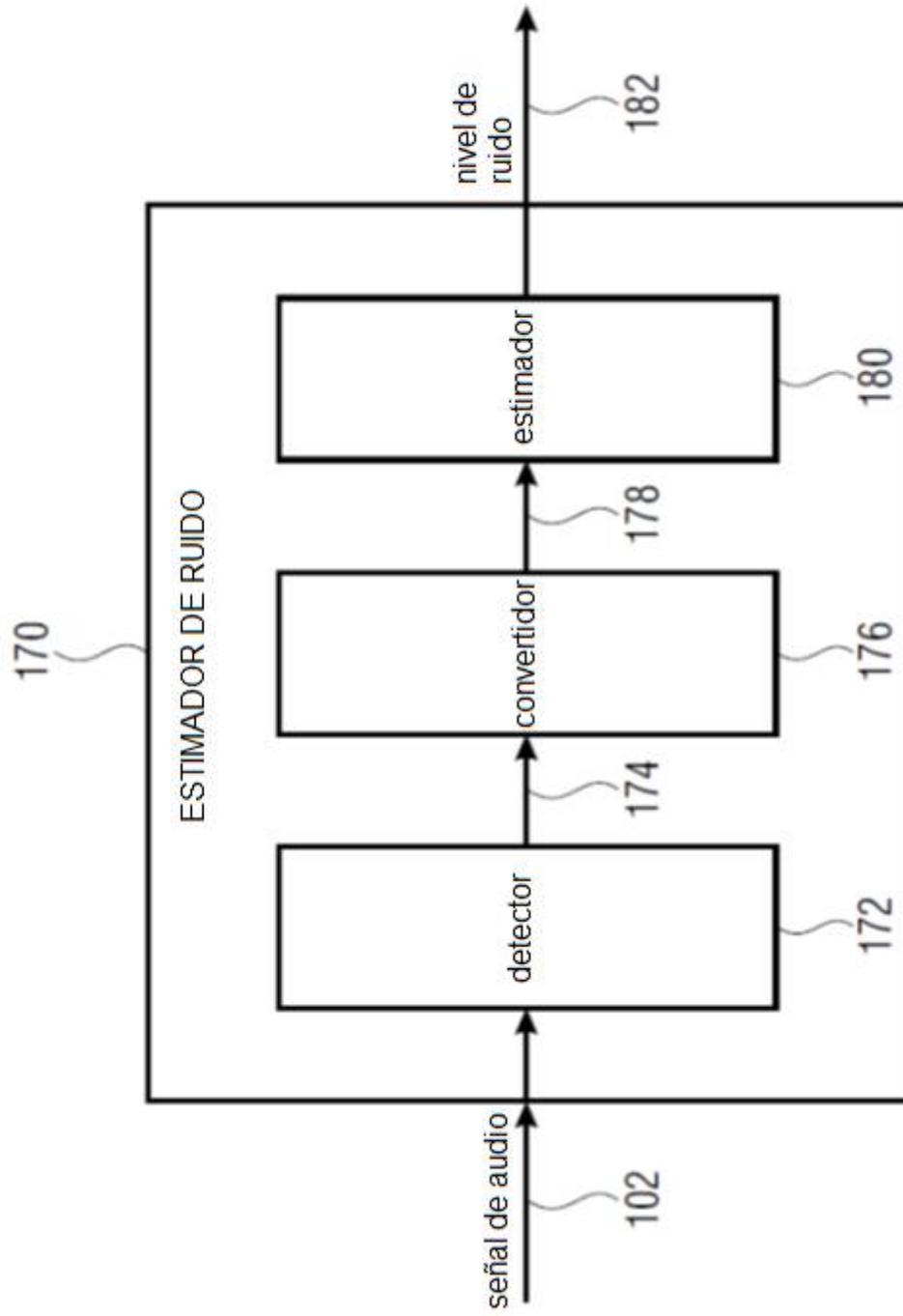


FIG 2

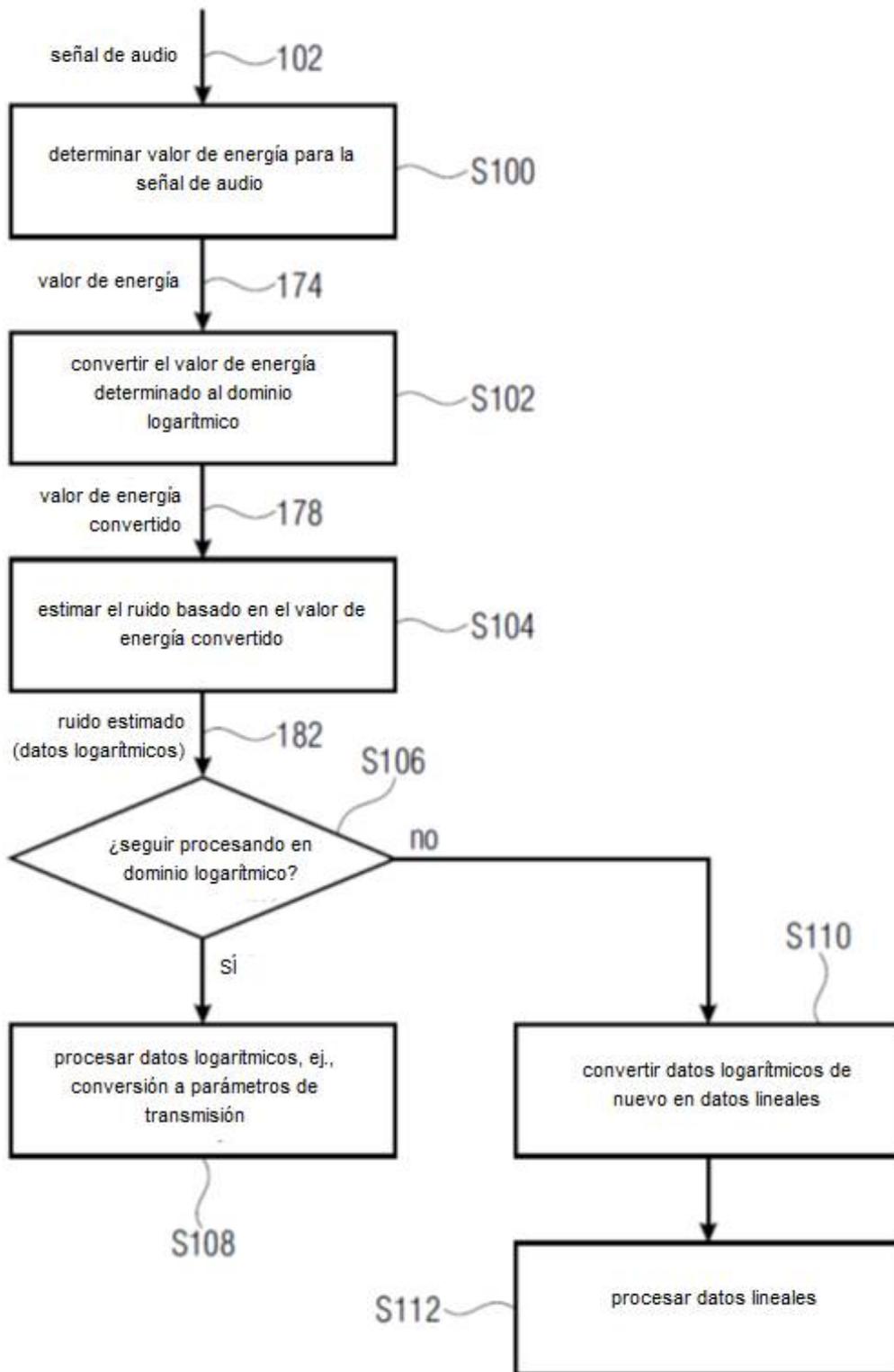


FIG 3