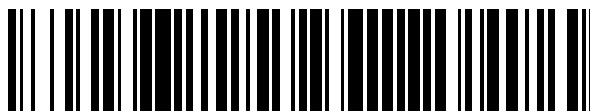


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 758 517**

51 Int. Cl.:

G10L 25/78 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.07.2015** **E 17202308 (7)**

97 Fecha y número de publicación de la concesión europea: **04.09.2019** **EP 3309784**

54 Título: **Estimación del ruido de fondo en las señales de audio**

30 Prioridad:

29.07.2014 US 201462030121 P

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

05.05.2020

73 Titular/es:

**TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)
(100.0%)
164 83 Stockholm, SE**

72 Inventor/es:

SEHLSTEDT, MARTIN

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 758 517 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

Estimación del ruido de fondo en las señales de audio

5 **Campo técnico**

Las realizaciones de la presente invención se refieren a procesamiento de señales de audio, y, en particular, a la estimación de ruido de fondo, por ejemplo, para soportar una decisión de actividad sonora.

10 **Antecedentes**

En los sistemas de comunicación que utilizan transmisión discontinua (DTX) es importante encontrar un equilibrio entre la eficiencia y el no reducir la calidad. En tales sistemas, se usa un detector de actividad para indicar señales activas, por ejemplo, de voz o música, que deben codificarse activamente, y segmentos con señales de fondo que pueden reemplazarse con ruido de confort generado en el lado del receptor. Si el detector de actividad es demasiado eficiente para detectar la falta de actividad, introducirá un recorte en la señal activa, que luego se percibe como degradación subjetiva de la calidad cuando el segmento activo recortado se reemplaza con ruido de confort. Al mismo tiempo, la eficiencia de la DTX se reduce si el detector de actividad no es lo suficientemente eficiente y clasifica los segmentos de ruido de fondo como activos y luego codifica activamente el ruido de fondo en lugar de entrar en un modo de DTX con ruido de confort. En la mayoría de los casos, el problema de recorte se considera peor.

La figura 1 muestra un diagrama de bloques general de un detector de actividad de sonido generalizado, SAD, o un detector de actividad de voz, VAD, que toma una señal de audio como entrada y produce una decisión de actividad como salida. La señal de entrada se divide en tramas de datos, es decir, en segmentos de señal de audio de, por ejemplo, 5-30 ms, dependiendo de la implantación, y se produce una decisión de actividad por trama como salida.

Una decisión primaria, "prim", es hecha por el detector primario ilustrado en la figura 1. La decisión principal es básicamente tan sólo una comparación de las características de una trama actual con las características de fondo, que se estiman a partir de tramas de entrada anteriores. Una diferencia entre las características de la trama actual y las características de fondo que es mayor que un umbral origina una decisión primaria activa. El bloque de adición de traspaso se usa para extender la decisión primaria en base a decisiones primarias pasadas para formar la decisión final, "bandera". La razón para usar el traspaso es principalmente para reducir/eliminar el riesgo de saturación de actividad en el medio y el extremo de fondo del recorte. Como se indica en la figura, un controlador de funcionamiento puede ajustar el/los umbral/e/s para el detector primario y la longitud de la adición del traspaso de acuerdo con las características de la señal de entrada. El bloque estimador de fondo se usa para estimar el ruido de fondo en la señal de entrada. El ruido de fondo también puede denominarse "el fondo" o "la característica de fondo" en el presente documento.

La estimación de la característica de fondo se puede hacer de acuerdo con dos principios básicamente diferentes, ya sea usando la decisión primaria, es decir, con la decisión o la retroalimentación métrica de la decisión, que se indica mediante una línea de puntos en la figura 1, o usando algunas otras características de la señal de entrada, es decir, sin retroalimentación de decisión. También es posible usar combinaciones de las dos estrategias.

Un ejemplo de un códec usando retroalimentación de decisión para la estimación de fondo es AMR-NB (banda estrecha de tasa múltiple adaptativa) y ejemplos de códecs en los que no se utiliza retroalimentación de decisión son el EVRC (códec de tasa variable mejorado) y el G.718.

Hay un número de diferentes características de la señal o rasgos que se pueden utilizar, pero una característica común utilizada en los VAD es la característica de frecuencia de la señal de entrada. Un tipo de características de frecuencia comúnmente utilizado es la energía de trama de subbanda, debido a su baja complejidad y a su funcionamiento fiable en baja SNR (relación señal/ruido). Por lo tanto, se asume que la señal de entrada se divide en diferentes subbandas de frecuencia y se estima el nivel de fondo para cada una de las subbandas. De esta manera, una de las características de ruido de fondo es el vector con los valores de energía para cada subbanda. Éstos son valores que caracterizan el ruido de fondo en la señal de entrada en el dominio de frecuencia.

Para conseguir el seguimiento del ruido de fondo, la actualización de la estimación de ruido de fondo real se puede hacer de, por lo menos, tres maneras diferentes. Una manera es usar un proceso de regresión automática, AR, por depósito de frecuencia para manipular la actualización. Ejemplos de tales códecs son el AMR-NB y el G.718. Básicamente, para este tipo de actualización, el tamaño del paso de la actualización es proporcional a la diferencia observada entre la entrada actual y la estimación de fondo actual. Otra manera es usar la escala multiplicativa de una estimación actual con la restricción de que la estimación nunca puede ser mayor que la entrada actual o menor que un valor mínimo. Esto significa que la estimación aumenta en cada trama hasta que sea más alta que la entrada actual. En esa situación, la entrada actual se utiliza como estimación. El EVRC es un ejemplo de códec que utiliza esta técnica para actualizar la estimación de fondo para la función del VAD. Obsérvese que el EVRC utiliza diferentes estimaciones de fondo para el VAD y la supresión de ruido. Cabe señalar que un VAD se puede utilizar en

otros contextos que no sean el de DTX. Por ejemplo, en códecs de velocidad variable, como en el EVRC, el VAD puede usarse como parte de una función de determinación de tasa.

5 Una tercera forma es utilizar una técnica denominada de mínimos en la que la estimación es el valor mínimo durante una ventana de tiempo deslizante de tramas anteriores. Básicamente, esto proporciona una estimación mínima que se escala, utilizando un factor de compensación, para obtener y aproximarse a una estimación promedio del ruido estacionario.

10 En los casos de alta SNR, donde el nivel de señal de la señal activa es mucho mayor que la señal de fondo, puede ser bastante fácil tomar la decisión de si una señal de entrada de audio está o no activa. Sin embargo, es muy difícil separar las señales activas y no activas en casos de baja SNR, y, en particular, cuando el fondo no es estacionario, o incluso es similar en características a la señal activa.

15 El rendimiento del VAD depende de la capacidad del estimador de ruido de fondo para realizar un seguimiento de las características del fondo - en particular cuando se trata de fondos no estacionarios. Con un mejor seguimiento, es posible hacer que el VAD sea más eficiente sin aumentar el riesgo de recorte de voz.

20 Mientras que la correlación es una característica importante que se utiliza para detectar el habla, principalmente la parte sonora de la voz, también hay señales de ruido que muestran alta correlación. En estos casos, el ruido con correlación impedirá la actualización de las estimaciones de ruido de fondo. El resultado es una gran actividad, ya que tanto el ruido de voz como el de fondo se codifican como contenido activo. Mientras que para altos SNR (aproximadamente > 20dB) sería posible reducir el problema utilizando la detección de pausa en base a energía, esto no es fiable para el rango de SNR de 20dB a 10dB o posiblemente de 5dB. Es en este rango en el que la solución descrita en el presente documento marca la diferencia.

25 M. Jelinek y R. Salami, en "Noise reduction method for wideband speech coding" 2004, 12ª conferencia europea de procesamiento de señales, pp. 1959-1962, enseña un método para la estimación del ruido de fondo donde la presencia de pausas, durante las cuales se estima dicho ruido, se determina en base al cociente entre un residual de una predicción lineal de segundo orden y un residual de una predicción lineal de 16º orden.

30 Sumario

35 Sería deseable conseguir una estimación mejorada del ruido de fondo en las señales de audio. "Mejorada" puede implicar, en el presente documento, tomar una decisión más correcta con respecto a si una señal de audio comprende voz o música activa o no, y, de este modo, estimar más a menudo, por ejemplo, actualizando una estimación anterior, estando en realidad libre de contenido activo, tal como de voz y/o de música, el ruido de fondo en los segmentos de señal de audio. En el presente documento, se proporciona un método mejorado para generar una estimación de ruido de fondo, que puede permitir, por ejemplo, un detector de actividad de sonido para tomar decisiones más adecuadas.

40 Para el fondo de estimación del ruido en las señales de audio, es importante ser capaz de encontrar funciones fiables para identificar las características de una señal de ruido de fondo también cuando una señal de entrada comprende una mezcla desconocida de señales activas y de fondo, donde las señales activas pueden comprender voz y/o música.

45 El inventor se ha dado cuenta de que las características relacionadas con las energías residuales para diferentes órdenes del modelo de predicción lineal pueden utilizarse para detectar pausas en señales de audio. Estas energías residuales pueden extraerse, por ejemplo, de un análisis de predicción lineal, que es común en los códecs de voz. Las características se pueden filtrar y combinar para crear un conjunto de características o parámetros que se pueden usar para detectar el ruido de fondo, lo que hace que la solución sea adecuada para su uso en la estimación del ruido. La solución descrita en el presente documento es particularmente eficiente para las condiciones cuando una SNR está en el rango de 10 a 20 dB.

50 Otra característica proporcionada en el presente documento es una medida de cercanía espectral para el fondo, que puede estar hecho por ejemplo mediante el uso de las energías de subbanda de dominio de frecuencia que se utilizan por ejemplo en una subbanda de SAD. La medida de cercanía espectral también puede usarse para tomar una decisión sobre si una señal de audio comprende o no una pausa.

55 De acuerdo con un primer aspecto, se proporciona un método para la estimación de ruido de fondo. El método comprende obtener al menos un parámetro asociado con un segmento de señal de audio, tal como una trama o parte de una trama, en base a una primera ganancia de predicción lineal, calculada como el cociente entre una energía de la señal de entrada y una energía de señal residual de una primera predicción lineal para el segmento de señal de audio; y, una segunda ganancia de predicción lineal calculada como el cociente entre la energía de señal residual de la primera predicción lineal y una energía de señal residual de una segunda predicción lineal para el segmento de señal de audio. El método comprende adicionalmente determinar si el segmento de señal de audio comprende una pausa en base al menos en el al menos un parámetro; y, actualizar una estimación de ruido de

fondo en base al segmento de señal de audio si se determina que el segmento de señal de audio comprende una pausa.

5 De acuerdo con un segundo aspecto, se proporciona un aparato para la estimación de ruido de fondo en una señal de audio. El aparato está configurado para obtener al menos un parámetro en base a una primera ganancia de predicción lineal, calculada como el cociente entre una energía de un segmento de señal de audio y una energía de señal residual de una primera predicción lineal para el segmento de señal de audio; y, una segunda ganancia de predicción lineal calculada como el cociente entre la energía de señal residual de la primera predicción lineal y una energía de señal residual de una segunda predicción lineal para el segmento de señal de audio. El estimador de ruido de fondo está configurado adicionalmente para determinar si el segmento de señal de audio comprende una pausa en base al menos al menos un parámetro; y, para actualizar una estimación de ruido de fondo en base al segmento de señal de audio si se determina que el segmento de señal de audio comprende una pausa.

15 De acuerdo con un tercer aspecto, se proporciona un códec de audio, que comprende el aparato de acuerdo con el segundo aspecto.

De acuerdo con un cuarto aspecto, se proporciona un dispositivo de comunicación, que comprende el aparato de acuerdo con el segundo aspecto.

20 Breve descripción de los dibujos

Los anteriores y otros objetos, características y ventajas de la tecnología descrita en el presente documento serán evidentes a partir de la siguiente descripción más particular de realizaciones, como se ilustra en los dibujos que se acompañan. Los dibujos no están necesariamente a escala, sino que se pone el énfasis en ilustrar los principios de la tecnología en el presente documento descrita.

La figura 1 es un diagrama de bloques que ilustra un detector de actividad y una lógica de determinación del traspaso.

30 La figura 2 es un diagrama de flujo que ilustra un método para la estimación del ruido de fondo, de acuerdo con una realización a modo de ejemplo.

La figura 3 es un diagrama de bloques que ilustra el cálculo de características relacionadas con las energías residuales para la predicción lineal de orden 0 y 2 de acuerdo con una realización a modo de ejemplo.

35 La figura 4 es un diagrama de bloques que ilustra el cálculo de características relacionadas con las energías residuales para la predicción lineal de orden 2 y 16 de acuerdo con una realización a modo de ejemplo.

40 La figura 5 es un diagrama de bloques que ilustra el cálculo de características relacionadas con una medida de cercanía espectral de acuerdo con una realización ejemplar.

La figura 6 es un diagrama de bloques que ilustra un estimador de fondo de energía de subbanda.

45 La figura 7 es un diagrama de flujo que ilustra una lógica de decisión de actualización de fondo de la solución descrita en el Anexo A.

Las figuras 8-10 son diagramas que ilustran el comportamiento de diferentes parámetros presentados en el presente documento cuando se calculan para una señal de audio que comprende dos ráfagas de voz.

50 Las figuras 11a-11c y 12-13 son diagramas de bloques que ilustran diferentes implantaciones de un estimador de ruido de fondo de acuerdo con realizaciones ejemplares.

Las figuras A2-A9 en las páginas de figuras marcadas como "Anexo A" están asociadas con el Anexo A, y se mencionan en dicho Anexo A con el número que sigue a la letra "A", es decir, 2-9.

55 Descripción detallada

La solución descrita en el presente documento se refiere a la estimación de ruido de fondo en señales de audio. En el detector de actividad generalizado ilustrado en la figura 1, la función de estimar el ruido de fondo la realiza el bloque denominado "estimador de fondo". Algunas realizaciones de la solución descrita en el presente documento pueden verse en relación con soluciones descritas anteriormente en WO2011/049514, WO2011/049515, y también en el Anexo A (Apéndice A). La solución descrita en el presente documento se comparará con las implantaciones de estas soluciones descritas anteriormente. Aunque las soluciones descritas en los documentos WO2011/049514, WO2011/049515 y el Anexo A son buenas soluciones, la solución presentada en el presente documento todavía tiene ventajas en relación con estas soluciones. Por ejemplo, la solución presentada en el presente documento es incluso más adecuada en su seguimiento del ruido de fondo.

El rendimiento de un VAD depende de la capacidad del estimador de ruido de fondo para realizar un seguimiento de las características del fondo - en particular cuando se trata de fondos no estacionarias. Con un mejor seguimiento, es posible hacer que el VAD sea más eficiente sin aumentar el riesgo de recorte de voz.

5 Un problema con los métodos de estimación de ruido actuales es que para conseguir un buen seguimiento del ruido de fondo en baja SNR, se necesita un detector de pausa fiable. Para voz sólo de entrada, es posible utilizar la tasa silábica o el hecho de que una persona no puede hablar todo el tiempo para encontrar las pausas en el habla. Tales soluciones podrían implicar que después de un tiempo suficiente de no hacer actualizaciones de fondo, los requisitos para la detección de pausa son "relajados", de tal manera que es más probable detectar una pausa en el habla. Esto permite responder a cambios abruptos en las características o en el nivel de ruido. Algunos ejemplos de tales lógicas de recuperación de ruido son: 1) Como las declaraciones de voz contienen segmentos con alta correlación, es normalmente seguro asumir que hay una pausa en el habla después de un número suficiente de tramas sin correlación. 2) Cuando la relación señal/ruido, $SNR > 0$, la energía del habla es más alta que el ruido de fondo, por lo que si la energía de trama está cerca de la energía mínima durante un tiempo más largo, por ejemplo, 1-5 segundos, también es seguro asumir que uno está en una pausa de voz. Si bien las técnicas anteriores funcionan bien con la entrada sólo de voz, no son suficientes cuando la música se considera una entrada activa. En la música puede haber segmentos largos con baja correlación que sin embargo son música. Además, la dinámica de la energía en la música también puede activar una detección de pausa falsa, lo que puede dar lugar a actualizaciones erróneas no deseadas de la estimación del ruido de fondo.

Idealmente, una función inversa de un detector de actividad, o lo que se llamaría un "detector de ocurrencia de pausa", sería necesario que controlara la estimación de ruido. Esto garantizaría que la actualización de las características de ruido de fondo se haga sólo cuando no haya una señal activa en la trama actual. Sin embargo, como se indicó anteriormente, no es una tarea fácil determinar si un segmento de señal de audio comprende una señal activa o no.

Tradicionalmente, cuando se sabía que la señal activa era una señal de voz, el detector de actividad se llamaba detector de actividad de voz (VAD). El término VAD para detectores de actividad a menudo se usa también cuando la señal de entrada puede comprender música. Sin embargo, en los códecs modernos, también es común referirse al detector de actividad como un detector de actividad de sonido (SAD) cuando también se detecta música como señal activa.

El estimador de fondo ilustrado en la figura 1 utiliza retroalimentación desde el detector primario y/o el bloque de traspaso para localizar segmentos de señal de audio inactivos. Al desarrollar la tecnología descrita en el presente documento, ha sido un deseo quitar, o al menos reducir, la dependencia de tal retroalimentación. Por lo tanto, para la estimación de fondo descrita en el presente documento, el inventor ha identificado que es importante poder encontrar características fiables para identificar los rasgos de las señales de fondo cuando sólo está disponible una señal de entrada con una mezcla desconocida de señal activa y de fondo. El inventor se ha dado cuenta adicionalmente de que no se puede asumir que la señal de entrada comience con un segmento de ruido, o incluso que la señal de entrada sea una voz mezclada con ruido, ya que puede ser que la señal activa sea música.

Un aspecto es que incluso aunque de la trama actual puede tener el mismo nivel de energía que la estimación de ruido actual, las características de frecuencia pueden ser muy diferentes, lo que hace indeseable realizar una actualización de la estimación de ruido usando la trama actual. La actualización de ruido de fondo relativa de la característica de cercanía introducida se puede utilizar para impedir actualizaciones en estos casos.

Además, durante la inicialización es deseable permitir que la estimación de ruido se inicie tan pronto como sea posible, evitando decisiones equivocadas, ya que esto podría potencialmente resultar en el recorte del SAD si la actualización de ruido de fondo se hace con contenido activo. El uso de una versión específica de inicialización de la característica de cercanía durante la inicialización puede resolver este problema, al menos en parte.

La solución descrita en el presente documento se refiere a un método para la estimación de ruido de fondo, en particular, a un método para detectar pausas en una señal de audio que rinde bien en situaciones de SNR difíciles. La solución se describirá en adelante con referencia a las figuras 2-5.

En el campo de la codificación de voz, es común el uso de la denominada predicción lineal para analizar la forma espectral de una señal de entrada. El análisis se hace típicamente dos veces por trama, y para mejorar la precisión temporal los resultados se interpolan después de tal manera que hay un filtro generado para cada bloque de 5 ms de la señal de entrada.

La predicción lineal es una operación matemática, donde los valores futuros de una señal discreta en el tiempo se estiman como una función lineal de muestras anteriores. En el procesamiento de señales digitales, la predicción lineal a menudo se denomina codificación predictiva lineal (LPC) y, de este modo, puede verse como un subconjunto de la teoría de filtros. En la predicción lineal, en un codificador de voz, se aplica un filtro $A(z)$ de predicción lineal a una señal de voz de entrada. $A(z)$ es un filtro totalmente cero que, cuando se aplica a la señal de entrada, elimina la

redundancia que se puede modelar usando el filtro $A(z)$ de la señal de entrada. Por lo tanto, la señal de salida del filtro tiene una energía menor que la señal de entrada cuando el filtro tiene éxito al modelar algún aspecto o algunos aspectos de la señal de entrada. Esta señal de salida se denomina "el residual", "la energía residual" o "la señal residual". Tales filtros de predicción lineal, alternativamente denominados filtros residuales, pueden ser de diferente orden de modelo con diferentes números de coeficientes de filtro. Por ejemplo, con el fin de modelar propiamente el habla, puede que se requiera un filtro de predicción lineal de orden 16 del modelo. De este modo, en un codificador de voz, se puede usar un filtro $A(z)$ de predicción lineal de orden 16 de modelo.

El inventor se ha dado cuenta de que las características relacionadas con la predicción lineal se pueden utilizar para detectar pausas en señales de audio en un intervalo de SNR de 20 dB a 10 dB o posiblemente de 5 dB. De acuerdo con realizaciones de la solución descrita en el presente documento, una relación entre energías residuales para diferentes órdenes de modelo para una señal de audio se utiliza para detectar pausas en la señal de audio. La relación utilizada es el cociente entre la energía residual de un orden de modelo inferior y un orden de modelo superior. El cociente entre energías residuales puede denominarse "ganancia de predicción lineal", ya que es un indicador de cuánta energía de señal ha podido modelar, o eliminar, el filtro de predicción lineal entre un orden de modelo y otro orden de modelo.

La energía residual dependerá del orden M del modelo del filtro $A(z)$ de predicción lineal. Una forma común de calcular los coeficientes de filtro para un filtro de predicción lineal es el algoritmo de Levinson-Durbin. Este algoritmo es recursivo, y, en el proceso de crear un filtro de predicción $A(z)$ de orden M también, como un "subproducto", producirá las energías residuales de los órdenes de modelos inferiores. Este hecho puede utilizarse de acuerdo con realizaciones de la invención.

La figura 2 muestra un método general a modo de ejemplo para la estimación del ruido de fondo en una señal de audio. El método puede ser realizado por un estimador de ruido de fondo. El método comprende obtener 201 al menos un parámetro asociado con un segmento de señal de audio, como una trama o parte de una trama, en base a una primera ganancia de predicción lineal, calculada como el cociente entre una señal residual proveniente de una predicción lineal de orden 0 y una señal residual proveniente de una predicción lineal de segundo orden para el segmento de señal de audio; y a una segunda ganancia de predicción lineal, calculada como el cociente entre una señal residual proveniente de una predicción lineal de segundo orden y una señal residual proveniente de una predicción lineal de 16° orden para el segmento de señal de audio.

El método comprende adicionalmente determinar 202 si el segmento de señal de audio comprende una pausa, es decir, si está libre de contenido activo tal como el habla y la música, en base al menos al al menos un parámetro obtenido; y, actualizar 203 una estimación de ruido de fondo en base al segmento de señal de audio cuando el segmento de señal de audio comprenda una pausa. Esto es, que el método comprende la actualización de una estimación de ruido de fondo cuando se detecta una pausa en el segmento de señal de audio en base al al menos un parámetro obtenido.

Las ganancias de predicción lineal podrían describirse como una primera ganancia de predicción lineal relacionada con pasar de predicción lineal de orden 0 a orden segundo para el segmento de señal de audio; y una segunda ganancia de predicción lineal relacionada con pasar de predicción lineal de orden segundo al orden 16° para el segmento de audio. Además, la obtención del al menos un parámetro podría describirse alternativamente como la determinación, el cálculo, la derivación o la creación. Las energías residuales relacionadas con las predicciones lineales de orden 0, 2 y 16 del modelo pueden obtenerse, recibirse o recuperarse de, es decir, de alguna manera proporcionarse por, una parte del codificador donde la predicción lineal se realiza como parte de un proceso de codificación regular. Por ello, la complejidad computacional de la solución descrita en el presente documento puede reducirse, en comparación a cuando las energías residuales necesitan ser derivadas especialmente para la estimación de ruido de fondo.

El al menos un parámetro obtenido sobre la base de las características de predicción lineal pueden proporcionar un análisis independiente nivel de la señal de entrada que mejora la decisión de si se debe realizar una actualización de ruido de fondo o no. La solución es particularmente útil en el rango SNR de 10 a 20dB, donde los SAD basados en energía tienen un rendimiento limitado debido al rango dinámico normal de las señales de voz.

En el presente documento, entre otras, las variables $E(0), \dots, E(m), \dots, E(M)$ representan las energías residuales para los órdenes de modelo 0 a M de los filtros $M+1$ $A_m(z)$. Obsérvese que $E(0)$ es simplemente la energía de entrada. Un análisis de señal de audio de acuerdo con la solución descrita en el presente documento proporciona varias características o parámetros nuevos al analizar la ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de orden 0 y una señal residual proveniente de una predicción lineal de segundo orden, y la ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de segundo orden y una señal residual proveniente de una predicción lineal de orden 16°. Esto es, que la ganancia de predicción lineal para pasar de predicción lineal de orden 0 a orden segundo es lo mismo que la "energía residual" $E(0)$ (para un orden 0 de modelo) dividida por la energía residual $E(2)$ (para un segundo orden de modelo). En consecuencia, la ganancia de predicción lineal para pasar de predicción lineal de orden segundo a orden 16° es la misma que la energía residual $E(2)$ (para un segundo orden de modelo)

dividida por la energía residual $E(16)$ (para un orden 16° de modelo). Los ejemplos de parámetros y la determinación de parámetros en base a las ganancias de predicción se describirán con más detalle a continuación. El al menos un parámetro obtenido de acuerdo con la realización general descrita anteriormente puede formar parte de un criterio de decisión utilizado para evaluar, ya se actualice o no la estimación del ruido de fondo.

5 Con el fin de mejorar una estabilidad a largo plazo del al menos un parámetro o rasgo, se puede calcular una versión limitada de la ganancia de predicciones. Esto es, la obtención del al menos un parámetro puede comprender limitar las ganancias de la predicción lineal, relacionadas con pasar de predicción lineal de orden 0 a orden segundo y de orden 2° a orden 16° , para tomar valores en un intervalo predefinido. Por ejemplo, las ganancias de predicción lineal pueden limitarse a tomar valores entre 0 y 8, como se ilustra, por ejemplo, en la Ec. 1 y la Ec. 6 más adelante.

15 La obtención del al menos un parámetro puede comprender adicionalmente la creación de al menos una estimación a largo plazo de cada ganancia primera y segunda de predicción lineal, por ejemplo, por medio de filtrado de paso bajo. Tal al menos una estimación a largo plazo se basaría después en las ganancias de predicción lineal correspondientes asociadas con al menos un segmento de señal de audio anterior. Podría ser creada más de una estimación a largo plazo, si, por ejemplo, unas estimaciones a largo plazo primera y segunda, relacionadas con una ganancia de predicción lineal, reaccionan de manera diferente a los cambios en la señal de audio. Por ejemplo, una primera estimación a largo plazo puede reaccionar más rápido a los cambios que una segunda estimación a largo plazo. Una tal primera estimación a largo plazo se puede denotar alternativamente como una estimación a corto plazo.

20 La obtención del al menos un parámetro puede comprender adicionalmente la determinación de una diferencia, tal como la diferencia absoluta Gd_{0_2} (Ec. 3) descrita más adelante, entre una de las ganancias de predicción lineal, asociada con el segmento de señal de audio, y una estimación a largo plazo de dicha ganancia de predicción lineal. Alternativamente o adicionalmente, se podría determinar una diferencia entre dos estimaciones a largo plazo, tal como en la Ec. 9 de más adelante. El término "determinación" podría intercambiarse alternativamente por el de cálculo, creación o derivación.

30 La obtención del al menos un parámetro puede, como se indicó anteriormente, comprender filtrado de paso bajo de las ganancias de predicción lineal, derivándose, de este modo, de las estimaciones a largo plazo, entre las cuales, algunas, alternativamente, pueden estar denotadas como estimaciones a corto plazo, dependiendo de cuántos segmentos se tomen en consideración en la estimación. Los coeficientes de filtro de al menos un filtro de paso bajo pueden depender de una relación entre una ganancia de predicción lineal, relacionada, por ejemplo, sólo con el segmento de señal de audio actual, y un promedio, denotado, por ejemplo, como promedio a largo plazo o estimación a largo plazo, de una ganancia de predicción correspondiente obtenida en base a una pluralidad de segmentos de señal de audio anteriores. Esto puede realizarse para crear, por ejemplo, adicionalmente, estimaciones a largo plazo de las ganancias de predicción. El filtrado de paso bajo se puede realizar en dos o más pasos, donde cada paso puede dar como resultado un parámetro, o una estimación, que se utiliza para tomar una decisión con respecto a la presencia de una pausa en el segmento de señal de audio. Por ejemplo, pueden analizarse o compararse diferentes estimaciones a largo plazo (tales como $G1_{0_2}$ (Ec. 2) y Gad_{0_2} (Ec. 4), y/o, $G1_{2_16}$ (Ec. 7), $G2_{2_16}$ (Ec. 8) y Gad_{2_16} (Ec. 10) descritas más adelante) que reflejan los cambios en la señal de audio de diferentes maneras, para detectar una pausa en un segmento de señal de audio actual.

45 La determinación 202 de si el segmento de señal de audio comprende o no una pausa puede adicionalmente estar basado en una medida de la cercanía espectral asociada con el segmento de señal de audio. La medida de cercanía espectral indicará lo cerca que esté el nivel de energía "por banda de frecuencia", del segmento de señal de audio procesado actualmente, del nivel de energía "por banda de frecuencia" de la estimación de ruido de fondo actual, por ejemplo, con un valor inicial o una estimación que es el resultado de una actualización anterior realizada antes del análisis del segmento de señal de audio actual. A continuación se muestra un ejemplo de determinación o derivación de una medida de cercanía espectral en las ecuaciones Ec. 12 y Ec. 13. La medida de cercanía espectral se puede utilizar para impedir actualizaciones de ruido basadas en tramas de baja energía con una gran diferencia en las características de frecuencia, en comparación con la estimación de fondo actual. Por ejemplo, la energía promedio sobre las bandas de frecuencia podría ser igualmente baja para el segmento de señal actual y la estimación de ruido de fondo actual, pero la medida de cercanía espectral revelaría si la energía se distribuye de manera diferente sobre las bandas de frecuencia. Tal diferencia en la distribución de energía podría sugerir que el segmento de señal actual, por ejemplo, la trama, puede ser contenido activo de bajo nivel, y una actualización de la estimación del ruido de fondo en base a la trama podría, por ejemplo, impedir la detección de tramas futuras con contenido similar. Como la SNR de subbanda es más sensible a los aumentos de energía usando incluso contenido activo de bajo nivel, puede resultar una gran actualización de la estimación de fondo si ese rango de frecuencia particular no existe en el ruido de fondo, tal como la parte de alta frecuencia de voz comparada con ruido de automóvil de baja frecuencia. Después de tal actualización, será más difícil detectar la voz.

65 Como ya se ha sugerido anteriormente, la medida de cercanía espectral puede ser derivada, obtenida o calculada en base a las energías para un conjunto de bandas de frecuencia, alternativamente denotadas subbandas, del segmento de señal de audio analizado actualmente y las estimaciones de ruido de fondo actuales correspondientes al conjunto de bandas de frecuencia. Esto también se ejemplificará y describirá adicionalmente con más detalle más

adelante, y se ilustra en la figura 5.

Como se indicó anteriormente, la medida de cercanía espectral se puede derivar obtener o calcular mediante la comparación de un nivel de energía de banda actual por frecuencia del segmento de señal de audio actualmente procesado con un nivel de energía de banda por frecuencia de una estimación de ruido de fondo actual. Sin embargo, para empezar, es decir, durante un primer período o un primer número de tramas al comienzo del análisis de una señal de audio, puede que no haya una estimación fiable del ruido de fondo, por ejemplo, dado que todavía no se ha realizado una actualización fiable de una estimación del ruido de fondo. Por lo tanto, se puede aplicar un período de inicialización para determinar el valor de cercanía espectral. Durante tal período de inicialización, los niveles de energía de banda por frecuencia del segmento de señal de audio actual serán, en cambio, comparados con una estimación de fondo inicial, que puede ser, por ejemplo un valor constante configurable. En los ejemplos más abajo, esta estimación inicial de ruido de fondo se establece en el valor de ejemplo $E_{\min} = 0,0035$. Después del período de inicialización, el procedimiento puede conmutarse al funcionamiento normal y comparar el nivel de energía de banda actual por frecuencia del segmento de señal de audio procesado actualmente con un nivel de energía de banda por frecuencia de una estimación de ruido de fondo actual. La longitud del período de inicialización se puede configurar por ejemplo, en base a simulaciones o pruebas que indiquen el tiempo necesario antes de que se proporcione una estimación, por ejemplo fiable y/o satisfactoria, del ruido de fondo. Un ejemplo usado más adelante, la comparación con una estimación inicial de ruido de fondo (en lugar de con una estimación "real" derivada en base a la señal de audio actual) se realiza durante las primeras 150 tramas.

El al menos un parámetro puede ser el parámetro ejemplificado en código más adelante, denotado NEW_POS_BG, y/o uno o más perteneciente/s a la pluralidad de parámetros que se describen más adelante, que conduce/n a la formación de un criterio de decisión o de un componente de un criterio de decisión para la detección de pausa. En otras palabras, el al menos un parámetro, o característica, obtenido 201 en base a las ganancias de predicción lineal puede ser uno o más de los parámetros descritos a continuación, puede comprender uno o más de los parámetros descritos a continuación y/o estar basado en uno o más de los parámetros descritos a continuación.

Características o parámetros relacionados con las energías residuales E(0) y E(2)

La figura 3 muestra un diagrama de bloques de visión general de la derivación de características o parámetros relacionados con E(0) y E(2), de acuerdo con un ejemplo de realización. Como se puede ver en la figura 3, la ganancia de predicción se calcula primero como E(0)/E(2). Una versión limitada de la ganancia de predicciones se calcula como

$$G_{0_2} = \max(0, \min(8, E(0)/E(2))) \quad (\text{Ec. 1})$$

donde E(0) representa la energía de la señal de entrada y E(2) es la energía residual después de una predicción lineal de segundo orden. La expresión en la Ec. 1 limita la ganancia de predicción a un intervalo entre 0 y 8. La ganancia de predicción en casos normales debería ser mayor que cero, pero pueden ocurrir anomalías, por ejemplo, para valores cercanos a cero y, por lo tanto, puede ser útil una limitación "mayor que cero" ($0 <$). La razón para limitar la ganancia de predicción a un máximo de 8 es que, para los fines de la solución descrita en el presente documento, es suficiente saber que la ganancia de predicción es aproximadamente 8 o mayor que 8, lo que indica una ganancia de predicción lineal significativa. Cabe señalar que cuando no haya diferencia entre la energía residual entre dos órdenes de modelo diferentes, la ganancia de predicción lineal será 1, lo que indica que el filtro de un orden de modelo superior no tiene más éxito en el modelado de la señal de audio que el filtro de un modelo de orden inferior. Además, si la ganancia G_{0_2} de predicción tomara valores demasiado grandes en las siguientes expresiones, podría arriesgar la estabilidad de los parámetros derivados. Nótese que 8 es sólo un valor de ejemplo que se ha seleccionado para una realización específica. El parámetro G_{0_2} se podría denotar alternativamente, por ejemplo, epsP_{0_2} o $g_{LP_{0_2}}$.

La ganancia de predicción limitada se filtra después en dos pasos para crear estimaciones a largo plazo de esta ganancia. El primer filtrado de paso bajo y, de este modo, la derivación de una primera característica o primer parámetro a largo plazo se realiza como:

$$G1_{0_2} = 0,85 G_{0_2} + 0,15 G_{0_2}, \quad (\text{Ec. 2})$$

donde el segundo " $G1_{0_2}$ " en la expresión debe leerse como el valor de un segmento de señal de audio anterior. Este parámetro será típicamente o bien 0 o bien 8, dependiendo del tipo de ruido de fondo en la entrada una vez que haya un segmento de entrada de sólo fondo. El parámetro $G1_{0_2}$ se podría denotar alternativamente, por ejemplo, $\text{epsP}_{0_2_lp}$ o $g_{LP_{0_2}}$. Se puede crear o calcular después otra característica o parámetro utilizando la diferencia entre la primera característica $G1_{0_2}$ a largo plazo y la trama mediante ganancia G_{0_2} de predicción limitada por trama, de acuerdo con:

$$Gd_{0_2} = \text{abs}(G1_{0_2} - G_{0_2}) \quad (\text{Ec. 3})$$

Esto le dará una indicación de la ganancia de predicción de la trama actual, en comparación con la estimación a

largo plazo de la ganancia de predicción. El parámetro Gd_{0_2} también se puede denotar, alternativamente, por ejemplo, $epsP_{0_2_ad}$ o $g_{ad_{0_2}}$. En la figura 3, esta diferencia se utiliza para crear una segunda estimación a largo plazo o característica Gad_{0_2} . Esto se hace usando un filtro que aplica diferentes coeficientes de filtro dependiendo de si la diferencia a largo plazo es mayor o menor que la diferencia promedio estimada actualmente de acuerdo con:

$$Gad_{0_2} = (1-a) Gad_{0_2} + a Gd_{0_2} \quad (\text{Ec. 4})$$

donde, si $Gd_{0_2} < Gad_{0_2}$ entonces $a = 0,1$ más $a = 0,2$

donde el segundo " Gad_{0_2} " en la expresión debe leerse como el valor de un segmento de señal de audio anterior.

El parámetro Gad_{0_2} se puede denotar, alternativamente, por ejemplo, Glp_{0_2} , $epsP_{0_2_ad_lp}$ o $g_{ad_{0_2}}$. Con el fin de evitar que el filtrado enmascare diferencias ocasionales de trama alta, se puede derivar otro parámetro, que no se muestra en la figura. Es decir, que la segunda característica a largo plazo Gad_{0_2} se puede combinar con la diferencia de trama con el fin de impedir dicho enmascaramiento. Este parámetro puede derivarse tomando el máximo de la versión Gd_{0_2} de trama y la versión Gad_{0_2} a largo plazo de la característica de ganancia de predicción, como:

$$Gmax_{0_2} = \max(Gad_{0_2}, Gd_{0_2}) \quad (\text{Ec. 5})$$

El parámetro $Gmax_{0_2}$ podría, alternativamente, denotarse, por ejemplo, $epsP_{0_2_ad_lp_max}$ o $g_{max_{0_2}}$.

Características o parámetros relacionados con las energías residuales E(2) y E(16)

La figura 4 muestra un diagrama de bloques general de la derivación de características o parámetros relacionados con E(2) y E(16), de acuerdo con una realización a modo de ejemplo. Como se puede ver en la figura 4, la ganancia de predicción se calcula primero como E(2)/E(16). Las características o parámetros creados utilizando la diferencia o relación entre la energía residual de 2º orden y la energía residual de 16º orden se derivan de manera ligeramente diferente a las descritas anteriormente en relación con la relación entre las energías residuales de 0º y 2º orden.

En el presente documento, también, una ganancia de predicción limitada se calcula como

$$G_{2_16} = \max(0, \min(8, E(2)/E(16))) \quad (\text{Ec.- 6})$$

donde E(2) representa la energía residual después de una predicción lineal de segundo orden y E(16) representa la energía residual después de una predicción lineal de orden 16º. El parámetro G_{2_16} se puede denotar, alternativamente, por ejemplo, $epsP_{2_16}$ o $g_{LP_{2_16}}$. Esta ganancia de predicción limitada se usa entonces para crear dos estimaciones a largo plazo de esta ganancia: una en la que el coeficiente del filtro difiere si la estimación a largo plazo se va o no a aumentar, como se muestra en:

$$G1_{2_16} = (1-a) G1_{2_16} + a G_{2_16} \quad (\text{Ec.- 7})$$

donde si $G_{2_16} > G1_{2_16}$ entonces $a = 0,2$ más $a = 0,03$.

El parámetro $G1_{2_16}$ podría, alternativamente, denotarse, por ejemplo, $epsP_{2_16_lp}$ o $g_{LP_{2_16}}$.

La segunda estimación a largo plazo utiliza un coeficiente de filtro constante como de acuerdo con:

$$G2_{2_16} = (1-b) G2_{2_16} + b G_{2_16}, \text{ donde } b=0,02$$

El parámetro $G2_{2_16}$ podría, alternativamente, denotarse, por ejemplo, $epsP_{2_16_lp2}$ o $g_{LP2_{0_2}}$.

Para la mayoría de tipos de señales de fondo, tanto $G1_{2_16}$ como $G2_{2_16}$ estará cerca de 0, pero tendrán diferentes respuestas al contenido donde se necesita la predicción lineal de orden 16º, que es típicamente para voz y otro contenido activo. La primera estimación a largo plazo, $G1_{2_16}$, será normalmente más alta que la segunda estimación a largo plazo $G2_{2_16}$. Esta diferencia entre las características a largo plazo se mide de acuerdo con:

$$Gd_{2_16} = G1_{2_16} - G2_{2_16} \quad (\text{Ec. 9})$$

El parámetro Gd_{2_16} podría, alternativamente, denotarse $epsP_{2_16_dlp}$ o $g_{ad_{2_16}}$.

Gd_{2_16} puede usarse entonces como una entrada a un filtro que crea una tercera característica a largo plazo de acuerdo con:

$$Gad_{2_16} = (1-C) Gad_{2_16} + C Gd_{2_16} \quad (\text{Ec. 10})$$

donde si $Gd_{2_16} < Gad_{2_16}$ entonces $c=0,02$ más $c=0,05$

5 Este filtro aplica diferentes coeficientes de filtro dependiendo de si la tercera señal a largo plazo se va aumentar o no. El parámetro Gad_{2_16} se puede denotarse, alternativamente, por ejemplo, $epsP_{2_16_dlp_lp2}$ o $g_{ad_{2_16}}$. También aquí, la señal a largo plazo Gad_{2_16} puede combinarse con la señal de entrada de filtro Gd_{2_16} para impedir que el filtrado enmascare entradas altas ocasionales para la trama actual. El parámetro final es entonces el máximo de la trama o segmento y la versión a largo plazo de la característica.

10 $Gmax_{2_16} = \max(Gad_{2_16}, Gd_{2_16})$ (Ec. 11)

El parámetro $Gmax_{2_16}$ podría, alternativamente, denotarse, por ejemplo, $epsP_{2_16_dlp_max}$ o $g_{max_{0_2}}$

Medida de cercanía espectral/diferencia

15 La característica de cercanía espectral utiliza el análisis de frecuencia de la trama o segmento de entrada actual en el que la energía de subbanda se calcula y se compara con la estimación de fondo de subbanda. Se puede usar un parámetro o característica de cercanía espectral en combinación con un parámetro relacionado con las ganancias de predicción lineal descritas anteriormente, por ejemplo, para asegurarse de que el segmento o la trama actual esté relativamente cerca, o al menos no muy lejos, de una estimación de fondo anterior.

20 La figura 5 muestra un diagrama de bloques del cálculo de una medida de diferencia o cercanía espectral. Durante el período de inicialización, por ejemplo, las primeras 150 tramas, la comparación se realiza con una constante correspondiente a la estimación de fondo inicial. Después de la inicialización, se pasa al funcionamiento normal y se compara con la estimación de fondo. Obsérvese que, mientras que el análisis espectral produce energías de subbandas para 20 subbandas, aquí el cálculo de $nonstaB$ sólo usa subbandas $i = 2, \dots, 16$, ya que es principalmente en estas bandas donde se ubica la energía de la voz. Aquí $nonstaB$ refleja la no estacionariedad.

25 Entonces, durante la inicialización, se calcula $nonstaB$ usando un $Emin$, que en el presente documento se establece en $Emin = 0,0035$ como:

30 $nonstaB = \sum(\text{abs}(\log(Ecb(i)+1)) - \log(Emin+1))$ (Ec. 12)

donde la suma se hace sobre $i = 2 \dots 16$.

35 Esto se hace para reducir el efecto de los errores de decisión en la estimación del ruido de fondo durante la inicialización. Después del período de inicialización, el cálculo se hace utilizando la estimación de ruido de fondo actual de la subbanda respectiva, de acuerdo con:

40 $nonstaB = \sum(\text{abs}(\log(Ecb(i)+1)) - \log(Ncb(i)+1))$ (Ec. 13)

donde la suma se hace sobre $i = 2 \dots 16$

45 La adición de la constante de 1 a cada energía de subbanda antes del logaritmo reduce la sensibilidad para la diferencia espectral para tramas de baja energía. El parámetro de $nonstaB$ podría, alternativamente, denotarse, por ejemplo, no_staB o $nonstat_B$.

50 Un diagrama de bloques que ilustra un ejemplo de realización de un estimador de fondo se muestra en la figura 6. La realización de la figura 6 comprende un bloque para entramado 601 de entrada, que divide la señal de audio de entrada en tramas o segmentos de longitud adecuada, por ejemplo, de 5-30 ms. La realización comprende adicionalmente un bloque para la extracción 602 de características que calcula las características, también denotadas parámetros en el presente documento, para cada trama o segmento de la señal de entrada. La realización comprende adicionalmente un bloque para lógica 603 de decisión de actualización, para determinar si una estimación de fondo puede o no actualizarse en base a la señal en la trama actual, es decir, si el segmento de señal está libre de contenido activo tal como voz y música. La realización comprende adicionalmente un actualizador 604 de fondo, para actualizar la estimación del ruido de fondo cuando la lógica de decisión de actualización indica que es adecuado hacer la actualización. En la realización ilustrada, se puede derivar una estimación de ruido de fondo por subbanda, es decir, para varias bandas de frecuencia.

60 La solución en el presente documento descrita puede utilizarse para mejorar una solución anterior para la estimación de ruido de fondo, que se describe en el anexo A en el presente documento, y también en el documento WO2011/049514. En adelante, la solución descrita en el presente documento se describirá en el contexto de esta solución descrita anteriormente. Se darán ejemplos de código de una implantación de código de una realización de un estimador de ruido de fondo. En adelante, se describen detalles de implantación reales para una realización de la invención en un codificador en base al G.718. Esta implantación utiliza muchas de las características de energía descritas en la solución en el Anexo A y el documento WO2011/049514. Para detalles adicionales a los presentados

en adelante, consulte el Anexo A y el documento WO2011/049514.

Las siguientes características de energía se definen en el documento WO2011/049514:

```

Etot;
Etot_l_lp;
Etot_v_h;
totalNoise;
sign_dyn_lp;

```

5

Las siguientes características de correlación se definen en el documento WO2011/049514:

```

aEn;
harm_cor_cnt
act_pred
cor_est

```

10

Se definen las siguientes características en la solución dada en el Anexo A:

```

Etot_v_h;
lt_cor_est = 0.01f*cor_est + 0.99f*lt_cor_est;
lt_tn_track = 0.03f* (Etot - totalNoise < 10) + 0.97f*lt_tn_track;
lt_tn_dist = 0.03f* (Etot - totalNoise) + 0.97f*lt_tn_dist;
    lt_Ellp_dist=0.03f* (Etot- Etot_l_lp) + 0.97f*lt_Ellp_dist;
    harm_cor_cnt
    low_tn_track_cnt

```

- 15 La lógica de actualización de ruido de la solución dada en el anexo A se muestra en la figura 7. Las mejoras, en relación con la solución descrita en el presente documento, del estimador de ruido del Anexo A están relacionadas principalmente con la parte 701, donde se calculan características; la parte 702, donde las decisiones de pausa se toman en base a diferentes parámetros; y además con la parte 703, donde se toman diferentes acciones en base a si se detecta o no una pausa. Además, las mejoras pueden tener un efecto en la actualización 704 de la estimación del ruido de fondo, que podría actualizarse, por ejemplo, cuando se detecta una pausa en base a las nuevas características, que no se habrían detectado antes de introducir la solución descrita en el presente documento. En la implantación de ejemplo descrita en el presente documento, las nuevas características introducidas en el presente documento se calculan de la siguiente manera, comenzando con `no_staB`, que se determina utilizando las energías `enr[i]` de subbanda de la trama actual, que corresponde a `Ecb(i)` anterior y en la figura 6, y la estimación `bckr [i]` de ruido de fondo actual, que corresponde a `Ncb(i)` anterior y en la figura 6. La primera parte de la primera sección de código a continuación está relacionada con un procedimiento inicial especial para las primeras 150 tramas de una señal de audio, antes de que se haya derivado una propia estimación de fondo.
- 20
- 25

```

/* calculate non-stationarity feature relative background (spectral closeness feature non_staB */
if (ini_frame < 150)
{
  /* During init don't include updates */
  if ( i >= 2 && i <= 16 )
  {
    non_staB += (float)fabs(log(enr[i] + 1.0f) -
                           log(E_MIN + 1.0f));
  }
}
else
{
  /* After init compare with background estimate */
  if ( i >= 2 && i <= 16 )
  {
    non_staB += (float)fabs(log(enr[i] + 1.0f) -
                           log(bckr[i] + 1.0f));
  }
}

if (non_staB >= 128)
{
  non_staB = 32767.0/256.0f;
}

```

Las secciones de código siguientes muestran cómo se calculan las nuevas características de las energías residuales de predicción lineal, es decir, aquellas para la ganancia de predicción lineal. En el presente documento las energías residuales se denominan epsP [m] (cf. E(m) usado anteriormente).

5

```

/*-----*
 * Linear prediction efficiency 0 to 2 order
 *(linear prediction gain going from 0th to 2nd order model of linear prediction filter)
 *-----*/
epsP_0_2 = max(0 , min(8, epsP[0] / epsP[2]));
epsP_0_2_lp = 0.15f * epsP_0_2 + (1.0f-0.15f) * st->epsP_0_2_lp;
epsP_0_2_ad = (float) fabs(epsP_0_2 - epsP_0_2_lp);
if (epsP_0_2_ad < epsP_0_2_ad_lp)
{
    epsP_0_2_ad_lp = 0.1f * epsP_0_2_ad + (1.0f - 0.1f) * epsP_0_2_ad_lp;
}
else
{
    epsP_0_2_ad_lp = 0.2f * epsP_0_2_ad + (1.0f - 0.2f) * epsP_0_2_ad_lp;
}
epsP_0_2_ad_lp_max = max(epsP_0_2_ad,st->epsP_0_2_ad_lp);

/*-----*
 * Linear prediction efficiency 2 to 16 order
 *(linear prediction gain going from 2nd to 16th order model of linear prediction filter)
 *-----*/

epsP_2_16 = max(0 , min(8, epsP[2] / epsP[16]));
if (epsP_2_16 > epsP_2_16_lp)
{
    epsP_2_16_lp = 0.2f * epsP_2_16 + (1.0f-0.2f) * epsP_2_16_lp;
}
else
{
    epsP_2_16_lp = 0.03f * epsP_2_16 + (1.0f-0.03f) * epsP_2_16_lp;
}
epsP_2_16_lp2 = 0.02f * epsP_2_16 + (1.0f-0.02f) * epsP_2_16_lp2;

epsP_2_16_dlp = epsP_2_16_lp-epsP_2_16_lp2;

if (epsP_2_16_dlp < epsP_2_16_dlp_lp2 )
{
    epsP_2_16_dlp_lp2 = 0.02f * epsP_2_16_dlp + (1.0f-0.02f) * epsP_2_16_dlp_lp2;
}
else
{
    epsP_2_16_dlp_lp2 = 0.05f * epsP_2_16_dlp + (1.0f-0.05f) * epsP_2_16_dlp_lp2;
}

    epsP_2_16_dlp_max = max(epsP_2_16_dlp,epsP_2_16_dlp_lp2);

```

- 5 El código siguiente ilustra la creación de métricas combinadas, umbrales y banderas utilizadas para la decisión de actualización real, es decir, la determinación de si se debe actualizar o no la estimación del ruido de fondo. Al menos algunos de los parámetros relacionados con las ganancias de predicción lineal y/o la cercanía espectral se indican en **negrita**.

```
comb_ahc_epsP = max(max(act_pred,lt_haco_ev),epsP_2_16_dlp);
comb_hcm_epsP = max(max(lt_haco_ev,epsP_2_16_dlp_max),epsP_0_2_ad_lp_max);
```

```
haco_ev_max = max(st_harm_cor_cnt==0,>lt_haco_ev);
Etot_l_lp_thr = st->Etot_l_lp + (1.5f + 1.5f * (Etot_lp<50.0f))*Etot_v_h2;
```

```
enr_bgd = Etot < Etot_l_lp_thr;
cns_bgd = (epsP_0_2 > 7.95f) && (non_sta < 1e3f);
lp_bgd = epsP_2_16_dlp_max < 0.10f;
ns_mask = non_sta < 1e5f;
lt_haco_mask = lt_haco_ev < 0.5f;
bg_haco_mask = haco_ev_max < 0.4f;
```

```
SD_1 = ((epsP_0_2_ad > 0.5f) && (epsP_0_2 > 7.95f));
```

```
bg_bgd3 = enr_bgd || ((cns_bgd || lp_bgd) && ns_mask && lt_haco_mask && SD_1==0);
```

```
PD_1 = (epsP_2_16_dlp_max < 0.10f);
PD_2 = (epsP_0_2_ad_lp_max < 0.10f);
PD_3 = (comb_ahc_epsP < 0.85f);
PD_4 = comb_ahc_epsP < 0.15f;
PD_5 = comb_hcm_epsP < 0.30f;
```

```
BG_1 = ((SD_1==0) || (Etot < Etot_l_lp_thr)) && bg_haco_mask && (act_pred < 0.85f) && (Etot_lp < 50.0f);
```

```
PAU = (aEn==0) || ((Etot < 55.0f) && (SD_1==0) && ((PD_3 && (PD_1 || PD_2)) || (PD_4 || PD_5)));
```

```
NEW_POS_BG = (PAU | BG_1) & bg_bgd3;
```

```
/* Original silence detector works in most cases */
aE_bgd = aEn == 0;
```

```
/* When the signal dynamics is high and the energy is close to the background estimate */
sd1_bgd = (st->sign_dyn_lp > 15) && (Etot - st->Etot_l_lp) < 2*st->Etot_v_h2 && st->harm_cor_cnt > 20;
```

```
/* init conditions steadily dropping act_pred and/or lt_haco_ev */
tn_ini = ini_frame < 150 && harm_cor_cnt > 5 &&
    ((st->act_pred < 0.59f && st->lt_haco_ev < 0.23f) ||
    st->act_pred < 0.38f ||
    st->lt_haco_ev < 0.15f ||
    non_staB < 50.0f ||
    aE_bgd);
```

```
/* Energy close to the background estimate serves as a mask for other background detectors */
bg_bgd2 = Etot < Etot_l_lp_thr || tn_ini;
```

- 5 Como es importante no hacer una actualización de la estimación de ruido de fondo cuando una trama o segmento actual comprende contenido activo, se evalúan varias condiciones con el fin de decidir si se va a hacer una actualización. El paso de decisión principal en la lógica de actualización de ruido es si se debe hacer una actualización o no, y esto se forma mediante la evaluación de una expresión lógica, que se subraya a continuación. El nuevo parámetro NEW_POS_BG (nuevo en relación a la solución del Anexo A y del documento WO2011/049514)
- 10 es un detector de pausa, y se obtiene en base a las ganancias de predicción lineal que van de orden 0 a 2º y de 2º a 16º del modelo de un filtro de predicción lineal, y tn_ini se obtiene en función de las características relacionadas con la cercanía espectral. Aquí sigue una lógica de decisión usando las nuevas características, de acuerdo con la realización a modo de ejemplo.

```

updt_step=0.0f;
if ((bg_bgd2 && (aE_bgd || sd1_bgd || lt_tn_track > 0.90f || NEW POS BG)) ||
tn_ini)
{
  if( ( (act_pred < 0.85f) &&
    aE_bgd &&
    (lt_Ellp_dist < 10 || sd1_bgd) && lt_tn_dist < 40 &&
    ( (Etot - totalNoise) < 10.0f) ) ||
    (st->first_noise_updt == 0 && st->harm_cor_cnt > 80 && aE_bgd && st->lt_aEn_zero > 0.5f) ||
    (tn_ini && (aE_bgd || non_staB < 10.0 || st->harm_cor_cnt > 80) )
  )
  {
    updt_step=1.0f;
    st->first_noise_updt = 1;
    for( i=0; i< NB_BANDS; i++)
    {
      st->bckr[i] = tmpN[i];
    }
  }
}
else if ( ( (st->act_pred < 0.80f) && (aE_bgd || PAU) && st->lt_haco_ev < 0.10f) ||

```

```

    ( ( st->act_pred < 0.70f ) && ( aE_bgd || non_staB < 17.0f ) && PAU && st->lt_haco_ev < 0.15f ) ||
    ( st->harm_cor_cnt > 80 && st->totalNoise > 5.0f && Etot < max(1.0f, Etot_l_lp + 1.5f* st->Etot_v_h2) ) ||
    ( st->harm_cor_cnt > 50 && st->first_noise_updt > 30 && aE_bgd && st->lt_aEn_zero > 0.5f ) ||
    tn_ini
  )
}
updt_step=0.1f;
if ( !aE_bgd &&
    st->harm_cor_cnt < 50 &&
    ( st->act_pred > 0.6f ||
      ( !tn_ini && Etot_l_lp - st->totalNoise < 10.0f && non_staB > 8.0f ) ) )
{
  updt_step=0.01f;
}
if (updt_step > 0.0f )
{
  st->first_noise_updt = 1;
  for( i=0; i< NB_BANDS; i++ )
  {
    st->bckr[i] = st->bckr[i] + updt_step * (tmpN[i]-st->bckr[i]);
  }
}
else if (aE_bgd || st->harm_cor_cnt > 100 )
{
  ( st->first_noise_updt ) += 1;
}
}
else
{
  /* If in music lower bckr to drop further */
  if ( st->low_tn_track_cnt > 300 && st->lt_haco_ev > 0.9f && st->totalNoise > 0.0f )
  {
    updt_step=-0.02f;
    for( i=0; i< NB_BANDS; i++ )
    {
      if (st->bckr[i] > 2*E_MIN)
      {
        st->bckr[i] = 0.98f*st->bckr[i];
      }
    }
  }
}
}
st->lt_aEn_zero = 0.2f * (st->aEn==0) + (1-0.2f)*st->lt_aEn_zero;

```

5 Como se ha indicado anteriormente, las características de la predicción lineal proporcionan análisis independiente de nivel de la señal de entrada que mejora la decisión para la actualización de ruido de fondo que es particularmente útil en el intervalo de SNR de 10 a 20 dB, donde la energía en base al SAD ha limitado el rendimiento debido a el rango dinámico normal de las señales de voz.

10 Las características de cercanía de fondo mejoran también la estimación de ruido de fondo, ya que pueden ser utilizadas tanto para la inicialización como para el funcionamiento normal. Durante la inicialización, pueden permitir una inicialización rápida para el ruido de fondo (nivel inferior) con contenido principalmente de baja frecuencia, común para el ruido del automóvil. Además, las características se pueden usar para impedir actualizaciones de ruido al usar tramas de baja energía con una gran diferencia en las características de frecuencia en comparación con la estimación de fondo actual, lo que sugiere que la trama actual puede ser contenido activo de bajo nivel y una actualización podría impedir la detección de futuras tramas con contenido similar.

15 Las figuras 8-10 muestran cómo los parámetros o medidas respectivos se comportan para la voz con fondo de ruido de automóvil de SNR de 10 dB. En las figuras 8-10 los puntos, "*", representan cada uno la energía de trama. Para las figuras 8 y 9a-c, la energía se ha dividido entre 10 para ser más comparable para las características basadas en G_{0_2} y G_{2_16} . Los diagramas corresponden a una señal de audio que comprende dos expresiones, donde la posición aproximada para la primera expresión es en las tramas 1310 a 1420, y para la segunda en las tramas 1500

20

a 1610.

5 La figura 8 muestra la energía de trama (/10) (punto, "*") y las características G_0_2 (círculo, "o") y Gmax_0_2 (más, "+"), para voz de SNR de 10 dB con ruido de automóvil. Obsérvese que G_0_2 es 8 durante el ruido del automóvil, ya que existe cierta correlación en la señal que se puede modelar utilizando la predicción lineal con el orden 2 del modelo. Durante las expresiones, la característica Gmax_0_2 se convierte en más de 1,5 (en este caso), y, después de la ráfaga de voz, cae a 0. En una implantación específica de una lógica de decisión, Gmax_0_2 tiene que estar por debajo de 0,1 para permitir actualizaciones de ruido usando esta característica.

10 La figura 9a muestra la energía de trama (/10) (punto, "*") y las características G_2_16 (círculo, "o"), G1_2_16 (cruz, "x"), G2_2_16 (más, "+"). La figura 9b muestra la energía de trama (/10) (punto, "*") y las características G_2_16 (círculo, "o"), Gd_2_16 (cruz, "x") y Gad_2_16 (más, "+"). La figura 9c muestra la energía de trama (/10) (punto, "*") y las características G_2_16 (círculo, "o") y Gmax_2_16 (más, "+"). Los diagramas que se muestran en las figuras 9a-c también se refieren a voz de SNR de 10dB con ruido de coche. Las características se muestran en tres diagramas para que sea más fácil ver cada parámetro. Obsérvese que G_2_16 (círculo, "o") está justo por encima de 1 durante el ruido del automóvil (es decir, expresiones externas), lo que indica que la ganancia del orden superior de modelo es baja para este tipo de ruido. Durante las expresiones, la característica Gmax_2_16 (más, "+" de la figura 9c) aumenta, y luego comienza a caer nuevamente a 0. En una implantación específica de una lógica de decisión, la característica Gmax_2_16 también tiene que ser inferior a 0,1 para permitir actualizaciones de ruido. En esta muestra de señal de audio en particular, esto no ocurre.

25 La figura 10 muestra la energía de trama (punto, "*") (no dividida por 10 esta vez) y la característica nonstaB (más, "+") para voz de SNR de 10dB con ruido de automóvil. La característica nonstaB está en el intervalo de 0-10 durante segmentos de sólo ruido, y para las expresiones se vuelve mucho más grande (ya que las características de frecuencia son diferentes para la voz). Sin embargo, debe tenerse en cuenta que, incluso durante las expresiones, hay tramas en las que la característica nonstaB cae en el rango de 0 a 10. Para estas tramas podría haber la posibilidad de hacer actualizaciones de ruido de fondo y, por ello, hacer un mejor seguimiento del ruido de fondo.

30 La solución descrita en el presente documento también se refiere a un estimador de ruido de fondo implantado en equipo físico informático (hardware) y/o equipo lógico informático (software).

Estimador de ruido de fondo, figuras 11a-11c

35 Un ejemplo de realización de un estimador de ruido de fondo se ilustra de manera general en la figura 11a. Por estimador de ruido de fondo se entiende un módulo o entidad configurado para estimar el ruido de fondo en señales de audio que comprenden, por ejemplo, voz y/o música. El codificador 1100 está configurado para realizar al menos un método correspondiente a los métodos descritos anteriormente con referencia, por ejemplo, a las figuras 2 y 7. El codificador 1100 está asociado con las mismas características técnicas, objetos y ventajas que las realizaciones de métodos descritas anteriormente. El estimador de ruido de fondo se describirá brevemente para evitar repeticiones innecesarias.

El estimador de ruido de fondo se puede implantar y/o describir como sigue:

45 El estimador 1100 de ruido de fondo está configurado para la estimación de un ruido de fondo de una señal de audio. El estimador 1100 de ruido de fondo comprende circuitería de procesamiento o medios de procesamiento 1101 y una interfaz 1102 de comunicación. La circuitería 1101 de procesamiento está configurada para hacer que el codificador 1100 obtenga, por ejemplo, determine o calcule, al menos un parámetro, por ejemplo NEW_POS_BG, en base a una primera ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de orden 0 y una señal residual proveniente de una predicción lineal de segundo orden para el segmento de señal de audio; y una segunda ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de segundo orden y una señal residual proveniente de una predicción lineal de 16º orden para el segmento de señal de audio.

55 La circuitería 1101 de procesamiento está configurada adicionalmente para provocar que el estimador de ruido de fondo determine si el segmento de señal de audio comprende una pausa, es decir, está libre de contenido activo tal como voz y música, en base al al menos un parámetro. La circuitería 1101 de procesamiento está configurada adicionalmente para hacer que el estimador de ruido de fondo actualice una estimación de ruido de fondo en base al segmento de señal de audio cuando el segmento de señal de audio comprende una pausa.

60 La interfaz 1102 de comunicación, que también puede ser denotada, por ejemplo, interfaz de entrada/salida (I/O), incluye una interfaz para enviar datos a y recibir datos de otras entidades o módulos. Por ejemplo, las señales residuales, relacionadas con los órdenes 0, 2 y 16 del modelo de predicción lineal, se pueden obtener, por ejemplo, recibidas, mediante la interfaz de I/O desde un codificador de señal de audio que realiza una codificación predictiva lineal.

65 La circuitería 1101 de procesamiento podría, como se ilustra en la figura 11b, comprender medios de procesamiento, tales como un procesador 1103, por ejemplo una CPU, y una memoria 1104 para almacenar o mantener

instrucciones. La memoria comprendería, entonces, instrucciones, por ejemplo, en forma de un programa informático 1105, que, cuando fuera ejecutado por los medios 1103 de procesamiento hace que el codificador 1100 realice las acciones descritas anteriormente.

5 Una implantación alternativa de la circuitería de procesamiento 1101 se muestra en la figura 11c. La circuitería de procesamiento del presente documento comprende una unidad o módulo 1106 de obtención o determinación, configurado para hacer que el estimador 1100 de ruido de fondo obtenga, por ejemplo, determine o calcule, al menos un parámetro, por ejemplo NEW_POS_BG, en base a una primera ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de orden 0 y una señal residual
 10 proveniente de una predicción lineal de segundo orden para el segmento de señal de audio; y una segunda ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de segundo orden y una señal residual proveniente de una predicción lineal de 16° orden para el segmento de señal de audio. La circuitería de procesamiento comprende adicionalmente una unidad o módulo 1107 de determinación, configurado para hacer que el estimador 1100 de ruido de fondo determine si el segmento de señal de audio comprende una pausa, es decir, está libre de contenido activo tal como voz y música, en base al al menos un parámetro. La circuitería 1101 de procesamiento comprende adicionalmente una unidad o módulo 1110 de actualización o estimación, configurado para hacer que el estimador de ruido de fondo actualice una estimación de ruido de fondo en base al segmento de señal de audio cuando el segmento de señal de audio comprende una
 15 pausa.

20 La circuitería 1101 de procesamiento podría comprender más unidades, tales como una unidad o módulo de filtro configurado para provocar que el estimador de ruido de fondo filtre de paso bajo las ganancias de predicción lineal, creando, de este modo, una o más estimaciones a largo plazo de las acciones de ganancias de predicción lineal tal que el filtrado de paso bajo pueda realizarse de otro modo, por ejemplo, mediante la unidad o módulo 1107 de determinación.
 25

30 Las realizaciones de un estimador de ruido de fondo que se han descrito anteriormente se podrían configurar para las diferentes realizaciones del método descritas en el presente documento, tal como limitando y filtrando de paso bajo las ganancias de predicción lineal; determinando una diferencia entre ganancias de predicción lineal y estimaciones a largo plazo y entre estimaciones a largo plazo; y/u obteniendo y usando una medida de cercanía espectral, etc.

35 Se puede asumir que el estimador 1100 de ruido de fondo comprenda una funcionalidad adicional, para llevar a cabo la estimación de ruido de fondo, tal como, por ejemplo, la funcionalidad ejemplificada en el Apéndice A.

40 La figura 12 ilustra un estimador 1200 de ruido de fondo de acuerdo con una realización a modo de ejemplo. El estimador 1200 de ruido de fondo comprende una unidad de entrada, por ejemplo, para recibir energías residuales para los modelos de órdenes 0, 2 y 16. El estimador de ruido de fondo comprende adicionalmente un procesador y una memoria, conteniendo, dicha memoria, instrucciones ejecutables por dicho procesador, por lo que dicho estimador de ruido de fondo está operativo para: realizar un método de acuerdo con una realización descrita en el presente documento.

45 En consecuencia, el estimador de ruido de fondo puede comprender, como se ilustra en la figura 13, una unidad 1301 de entrada/salida, una calculadora 1302 para calcular los dos primeros conjuntos de características de las energías residuales para los modelos de órdenes 0, 2 y 16 y un analizador 1303 de frecuencia para calcular la característica de cercanía espectral.

50 Un estimador de ruido de fondo como los descritos anteriormente puede estar comprendido, por ejemplo, en un VAD o SAD, un codificador y/o un decodificador, es decir, un códec, y/o en un dispositivo, tal como un dispositivo de comunicación. El dispositivo de comunicación puede ser un equipo de usuario (UE) en forma de teléfono móvil, cámara de vídeo, grabadora de sonido, tableta, ordenador de escritorio, ordenador portátil, decodificador de TV o servidor doméstico/pasarela doméstica/punto de acceso doméstico/enrutador doméstico. El dispositivo de comunicación puede en algunas realizaciones ser un dispositivo de red de comunicaciones adaptado para codificar y/o transcodificar señales de audio. Ejemplos de tales dispositivos de red de comunicaciones son los servidores, tales como servidores de medios, servidores de aplicaciones, enrutadores, pasarelas y estaciones base de radio. El
 55 dispositivo de comunicación también puede adaptarse para colocarse, es decir, empotrarse, en una nave tal como un barco, un avión no tripulado volador, o un avión, y en un vehículo de carretera tal como un automóvil, un autobús o un camión. Tal dispositivo empotrado pertenecería típicamente a una unidad telemática del vehículo o al sistema de información y entretenimiento del vehículo.

60 Los pasos, funciones, procedimientos, módulos, unidades y/o bloques descritos en el presente documento pueden implantarse en hardware usando cualquier tecnología convencional, tales como circuito discreto o tecnología de circuito integrado, incluyendo tanto circuitería electrónica de fines generales como circuitería de aplicación específica.

65 Los ejemplos particulares incluyen uno o más procesadores de señal digital y otros circuitos electrónicos conocidos adecuadamente configurados, por ejemplo puertas lógicas discretas interconectadas para realizar una función

especializada o circuitos integrados específicos (ASIC).

5 Alternativamente, al menos algunos de los pasos, funciones, procedimientos, módulos, unidades y/o bloques descritos anteriormente pueden implantarse en software tal como un programa informático para su ejecución por circuitería de procesamiento adecuada, incluyendo una o más unidades de procesamiento. El software podría ser transportado por un operador, como una señal electrónica, una señal óptica, una señal de radio o un medio de almacenamiento legible por ordenador antes y/o durante el uso del programa informático en los nodos de la red.

10 El diagrama o los diagramas de flujo presentados en el presente documento pueden considerarse como un diagrama o diagramas de flujo informático, cuando se realiza/n por uno o más procesadores. Un aparato correspondiente puede definirse como un grupo de módulos de función, donde cada paso realizado por el procesador corresponde a un módulo de función. En este caso, los módulos de función se implantan como un programa informático que se ejecuta en el procesador.

15 Los ejemplos de circuitería de procesamiento incluyen, pero no se limitan a, uno o más microprocesadores, uno o más procesadores de señal digital, DSP, una o más unidades centrales de procesamiento, CPU, y/o cualquier circuitería lógica programable adecuada tal como una o más matrices de puertas programables de campo, FPGA, o uno o más controladores lógicos programables, PLC. Es decir, que las unidades o módulos de las disposiciones en los diferentes nodos descritos anteriormente podrían implantarse mediante una combinación de circuitos analógicos y digitales, y/o uno o más procesadores configurados con software y/o soporte lógico inalterable (firmware), por ejemplo almacenados en una memoria. Uno o más de estos procesadores, así como el otro hardware digital, pueden incluirse en una sola circuitería integrada de aplicación específica, ASIC, o en varios procesadores, y diversos hardware digitales pueden estar distribuidos entre varios componentes separados, ya sea empaquetados individualmente o ensamblados en un sistema en un chip, SoC.

25 También debe entenderse que puede ser posible reutilizar las capacidades generales de procesamiento de cualquier dispositivo o unidad convencional en el que se implante la tecnología propuesta. También es posible reutilizar el software existente, por ejemplo, reprogramando el software existente o añadiendo nuevos componentes de software.

30 Las realizaciones descritas anteriormente se dan simplemente como ejemplos, y debe entenderse que la tecnología propuesta no está limitada a los mismos. El experto en la técnica entenderá que pueden realizarse diversas modificaciones, combinaciones y cambios en las realizaciones sin apartarse del presente alcance. En particular, se pueden combinar diferentes soluciones parciales en las diferentes realizaciones en otras configuraciones, cuando sea técnicamente posible.

35 Cuando se use la palabra "comprender" o "que comprende" se interpretará como no limitativa, es decir, como que significa "consistir al menos en".

40 También se debe observar que en algunas implantaciones alternativas, las funciones/actos anotados en los bloques pueden ocurrir fuera del orden anotado en los diagramas de flujo. Por ejemplo, dos bloques mostrados en sucesión pueden de hecho ejecutarse de manera substancialmente concurrente o los bloques a veces pueden ejecutarse en el orden inverso, dependiendo de la funcionalidad/los actos involucrados. Lo que es más, la funcionalidad de un bloque dado de los diagramas de flujo y/o diagramas de bloques puede separarse en múltiples bloques y/o la funcionalidad de dos o más bloques de los diagramas de flujo y/o diagramas de bloques puede estar al menos parcialmente integrada. Finalmente, pueden añadirse/insertarse otros bloques entre los bloques que se ilustran, y/o bloques/operaciones pueden omitirse sin apartarse del alcance de los conceptos inventivos.

50 Debe entenderse que la elección de las unidades que interactúan, así como la denominación de las unidades dentro de esta divulgación, son sólo para fines de ejemplificación, y los nodos adecuados para ejecutar cualquiera de los métodos descritos anteriormente pueden configurarse en una pluralidad de maneras alternativas con el fin de poder ejecutar las acciones de procedimiento sugeridas.

También debe observarse que las unidades descritas en esta divulgación deben considerarse como entidades lógicas y no necesariamente como entidades físicas separadas.

55 La referencia a un elemento en singular no está destinada a significar "uno/a y sólo uno/a", a menos que se indique explícitamente, sino "uno/a o más". Lo que es más, no es necesario que un dispositivo o método aborde todos y cada uno de los problemas que se intentan resolver con la tecnología divulgada en el presente documento descrita, para que se abarque por la presente.

60 En algunos casos en el presente documento, se omiten descripciones detalladas de dispositivos, circuitería y métodos bien conocidos de modo que no se oscurezca la descripción de la tecnología divulgada con detalles innecesarios. Todas las afirmaciones del presente documento, que mencionan los principios, aspectos y realizaciones de la tecnología divulgada, así como ejemplos específicos de la misma, están destinados a abarcar tanto sus equivalentes estructurales como funcionales. Además, se pretende que tales equivalentes incluyan tanto los equivalentes conocidos actualmente como los equivalentes desarrollados en el futuro, por ejemplo, cualquier

elemento desarrollado que realice la misma función, independientemente de la estructura.

APÉNDICE

- 5 Se proporciona un método para un estimador de ruido de fondo para estimación de ruido de fondo en una señal de audio, en el que la señal de audio comprende una pluralidad de segmentos de señal de audio, comprendiendo el método:
- 10 1) obtener (201) al menos un parámetro asociado con un segmento de señal de audio, en base a:
- una primera ganancia de predicción lineal calculada como el cociente entre una señal residual ($E(0)$) de una predicción lineal de orden 0 y una señal residual ($E(2)$) de una predicción lineal de segundo orden para el segmento de señal de audio; y
 - 15 - una segunda ganancia de predicción lineal calculada como el cociente entre una señal residual ($E(2)$) de una predicción lineal de segundo orden y una señal residual ($E(16)$) de una predicción lineal de 16º orden para el segmento de señal de audio;
- 20 2) determinar (202) si el segmento de señal de audio comprende una pausa, es decir, está libre de contenido activo tal como voz y música, en base al menos en el al menos un parámetro obtenido; y:
- cuando el segmento de señal de audio comprende una pausa:
- actualizar (203) una estimación de ruido de fondo en base al segmento de señal de audio.
- 25 La obtención del al menos un parámetro puede comprender limitar las ganancias de predicción lineal primera y segunda, para tomar valores en un intervalo predefinido.
- 30 La obtención del al menos un parámetro puede comprender la creación de al menos una estimación a largo plazo de cada una de las ganancias de predicción lineal primera y segunda, por ejemplo, por medio de filtrado de paso bajo, en el que la estimación a largo plazo se basa adicionalmente en las correspondientes ganancias de predicción lineal asociadas con al menos un segmento de señal de audio anterior.
- 35 La obtención del al menos un parámetro puede comprender determinar una diferencia entre una de las ganancias de predicción lineal asociadas con el segmento de señal de audio y una estimación a largo plazo de dicha ganancia de predicción lineal y/o entre dos estimaciones diferentes a largo plazo asociadas con una ganancia de predicción lineal.
- 40 La obtención del al menos un parámetro puede comprender un filtrado de paso bajo de las ganancias primera y segunda de predicción lineal.
- 45 Los coeficientes del filtro de al menos un filtro de paso bajo pueden depender de una relación entre una ganancia de predicción lineal asociada con el segmento de señal de audio y un promedio de una ganancia de predicción lineal correspondiente obtenida sobre la base de una pluralidad de segmentos anteriores de señal de audio.
- 50 La determinación de si el segmento de señal de audio comprende una pausa puede basarse adicionalmente en una medida de cercanía espectral asociada con el segmento de señal de audio.
- El método puede comprender adicionalmente la obtención de la medida de la cercanía espectral en base a las energías para un conjunto de bandas de frecuencia del segmento de señal de audio y estimaciones de ruido de fondo correspondientes al conjunto de bandas de frecuencia. Durante un período de inicialización, se puede usar un valor inicial, E_{min} como las estimaciones de ruido de fondo en base a las cuales se obtiene la medida de cercanía espectral.
- 55 Se proporciona adicionalmente un estimador (1100) de ruido de fondo, para estimar el ruido de fondo en una señal de audio que comprende una pluralidad de segmentos de señal de audio, estando configurado el estimador de ruido de fondo para:
- 60 1) obtener al menos un parámetro en base a:
- una primera ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de orden 0 y una señal residual proveniente de una predicción lineal de segundo orden para el segmento de señal de audio; y
 - 65 - una segunda ganancia de predicción lineal calculada como el cociente entre una señal residual proveniente de una predicción lineal de segundo orden y una señal residual proveniente de una predicción lineal de 16º orden para el

segmento de señal de audio;

2) determinar si el segmento de señal de audio comprende una pausa, es decir, si está libre de contenido activo tal como voz y música, basándose al menos en el al menos un parámetro; y

5 cuando el segmento de señal de audio comprende una pausa:

- actualizar una estimación de ruido de fondo en base al segmento de señal de audio.

10 El estimador de ruido de fondo de acuerdo con la reivindicación 10, en el que la obtención del al menos un parámetro comprende limitar las ganancias primera y segunda de predicción lineal para tomar valores en un intervalo predefinido.

15 En el estimador de ruido de fondo, la obtención del al menos un parámetro puede comprender: crear al menos una estimación a largo plazo de cada una de las ganancias de predicción lineal primera y segunda, por ejemplo, por medio de filtrado de paso bajo, en el que la estimación a largo plazo se basa adicionalmente en las ganancias de predicción lineal correspondientes asociadas con al menos un segmento de señal de audio anterior.

20 En el estimador de ruido de fondo, la obtención de al menos un parámetro puede comprender: determinar una diferencia entre una de las ganancias de predicción lineal asociadas con el segmento de señal de audio y una estimación a largo plazo de dicha ganancia de predicción lineal y/o entre dos estimaciones diferentes a largo plazo asociadas con una ganancia de predicción lineal.

25 En el estimador de ruido de fondo, la obtención del al menos un parámetro puede comprender un filtro de paso bajo de las ganancias primera y segunda de predicción lineal.

30 En el estimador de ruido de fondo, los coeficientes de filtro del al menos un filtro de paso bajo pueden depender de una relación entre una ganancia de predicción lineal asociada con el segmento de señal de audio y un promedio de una ganancia de predicción lineal correspondiente obtenida en base a una pluralidad de segmentos anteriores de señal de audio

35 El estimador de ruido de fondo puede configurarse para basar adicionalmente la determinación de si el segmento de señal de audio comprende una pausa en una medida de cercanía espectral asociada con el segmento de señal de audio.

El estimador de ruido de fondo puede estar configurado para obtener la medida de la cercanía espectral en base a las energías para un conjunto de bandas de frecuencia del segmento de señal de audio y de las estimaciones de ruido de fondo correspondientes al conjunto de bandas de frecuencia.

40 El estimador de ruido de fondo puede configurarse para usar un valor inicial, E_{min} , como las estimaciones de ruido de fondo en base a las cuales se obtiene la medida de cercanía espectral, durante un período de inicialización.

Se proporciona adicionalmente un detector de actividad de sonido, SAD, que comprende el estimador de ruido de fondo como se describió anteriormente.

45 Se proporciona adicionalmente un códec que comprende un estimador de ruido de fondo como se describió anteriormente.

50 Se proporciona adicionalmente un dispositivo inalámbrico que comprende un estimador de ruido de fondo como se describió anteriormente.

Se proporciona adicionalmente un nodo de red que comprende un estimador de ruido de fondo como se describió anteriormente.

55 Se proporciona adicionalmente un programa informático, que comprende instrucciones que, cuando se ejecutan en al menos un procesador, hacen que el al menos un procesador lleve a cabo el método como se describió anteriormente. También se proporciona un soporte que contiene dicho programa informático, en el que el soporte es un elemento de entre una señal electrónica, una señal óptica, una señal de radio o un medio de almacenamiento legible por ordenador.

60 **ANEXO A**

Las referencias a las figuras en el texto a continuación son referencias a las figuras A2-A9, de modo que la "figura 2" en adelante corresponde a la figura A2 en los dibujos.

65 La figura 2 es un diagrama de flujo que ilustra una realización a modo de ejemplo de un método para la estimación

del ruido de fondo de acuerdo con la tecnología propuesta en el presente documento. El método está destinado a ser realizado por un estimador de ruido de fondo, que puede ser parte de un SAD. El estimador de ruido de fondo, y el SAD, pueden estar comprendidos adicionalmente en un codificador de audio, que a su vez puede estar comprendido en un dispositivo inalámbrico o un nodo de red. Para el estimador de ruido de fondo descrito, ajustar la estimación de ruido hacia abajo no está restringido. Para cada trama se calcula una posible nueva estimación de ruido de subbanda, independientemente de si la trama es de fondo o contenido activo, si el nuevo valor es más bajo que el actual, se usa directamente, ya que lo más probable es que sea de una trama de fondo. La siguiente lógica de estimación de ruido es un segundo paso en el que se decide si se puede aumentar la estimación de ruido de subbanda y, si es así, cuánto, el aumento se basa en la posible nueva estimación de ruido de subbanda calculada anteriormente. Básicamente, esta lógica forma la decisión de que la trama actual es una trama de fondo, y, si esto no es seguro, puede permitir un aumento menor en comparación con lo que se estimó originalmente.

El método ilustrado en la figura 2 comprende: cuando un nivel de energía de un segmento de señal de audio es mayor que un umbral más alto 202:1 que un nivel de energía mínimo a largo plazo, It_{min} , o, cuando el nivel de energía del segmento de señal de audio es menor que un umbral más alto 202:2 que la It_{min} , pero no se detecta ninguna pausa 204:1 en el segmento de señal de audio:

- reducir 206 una estimación de ruido de fondo actual cuando se determina que el segmento de señal de audio 203:2 comprende música, y la estimación de ruido de fondo actual excede un valor mínimo 205:1, denotado "T" en la figura 2, y ejemplificado adicionalmente, por ejemplo, como $2 * E_{MIN}$ en el código a continuación.

Al realizar lo anterior y proporcionar la estimación del ruido de fondo a un SAD, el SAD está habilitado para realizar una detección de actividad de sonido más adecuada. Además, se habilita la recuperación de actualizaciones de estimación de ruido de fondo erróneas.

El nivel de energía del segmento de señal de audio utilizado en el método descrito anteriormente puede denominarse alternativamente, por ejemplo, como la energía de trama actual, $Etot$, o como la energía del segmento de señal, o trama, que puede calcularse sumando las energías de subbanda para el segmento de señal actual.

La otra característica de energía utilizada en el método anterior, es decir, el nivel de energía mínima a largo plazo, It_{min} , es una estimación, que se determina sobre una pluralidad de segmentos o tramas anteriores de señal de audio. Alternativamente, la It_{min} podría denotarse, por ejemplo, $Etot_l_lp$. Una manera básica de derivar la It_{min} sería utilizar el valor mínimo del historial de la energía de trama actual sobre cierto número de tramas pasadas. Si el valor calculado como: "energía de trama actual - estimación mínima a largo plazo" está por debajo de un valor umbral, denotado, por ejemplo, THR1, se dice, aquí, que la energía de trama actual está cerca de la energía mínima a largo plazo, o que está próxima a la energía mínima largo plazo. Es decir, que cuando $(Etot - It_{min}) < THR1$, la energía de trama actual, $Etot$, se puede determinar 202 para que esté cerca de la energía mínima a largo plazo It_{min} . El caso en el que $(Etot - It_{min}) = THR1$ puede referirse a cualquiera de las decisiones, 202:1 o 202:2, dependiendo de la implantación. La numeración 202:1 en la figura 2 indica la decisión de que la energía de trama actual no esté cerca de la It_{min} , mientras que 202:2 indica la decisión de que la energía de trama actual está cerca de la It_{min} . Otra numeración en la figura 2, con la forma XXX:Y, indica las decisiones correspondientes. La característica It_{min} se describirá con más detalle a continuación.

El valor mínimo, que la estimación de ruido de fondo actual va a exceder, con el fin de ser reducido, puede asumirse que es cero o un valor positivo bajo. Por ejemplo, como se ejemplificará en el código a continuación, una energía total actual de la estimación de fondo, que puede denotarse como "totalNoise" (ruido total) y determinarse, por ejemplo, como $10 * \log_{10} \Sigma backr [i]$, puede requerirse que exceda un valor mínimo de cero con el fin de que se cuestione la reducción. Alternativamente, o adicionalmente, cada entrada en un vector $backr[i]$, que comprende las estimaciones de fondo de subbanda, se puede comparar con un valor mínimo, E_{MIN} , con el fin de que se realice la reducción. En el siguiente ejemplo de código, E_{MIN} es un valor positivo bajo.

Cabe señalar que, de acuerdo con una realización preferida de la solución sugerida en el presente documento, la decisión de si el nivel de energía del segmento de señal de audio es más alto que un umbral que la It_{min} se basa únicamente en la información derivada de la señal de audio de entrada, es decir, que no se basa en la retroalimentación de una decisión del detector de actividad de sonido.

La determinación 204 de si una trama actual comprende una pausa o no, puede realizarse de diferentes maneras en base a uno o más criterios. Un criterio de pausa también puede denominarse detector de pausa. Se podría aplicar un sólo detector de pausa o una combinación de diferentes detectores de pausa. Con una combinación de detectores de pausa, cada uno se puede utilizar para detectar pausas en diferentes condiciones. Un indicador de que una trama actual puede comprender una pausa, o inactividad, es que una característica de correlación para la trama sea baja, y que varias tramas anteriores también hayan tenido características de baja correlación. Si la energía actual está cerca de la energía mínima a largo plazo y se detecta una pausa, el ruido de fondo se puede actualizar de acuerdo con la entrada actual, como se ilustra en la figura 2. Se puede considerar que se detecta una pausa cuando, además de eso, el nivel de energía del segmento de señal de audio es inferior a un umbral superior a la It_{min} : se ha determinado que un número predefinido de segmentos de señal de audio anteriores consecutivos no

comprende una señal activa y/o una dinámica de la señal de audio supera un umbral. Esto también se ilustra en el ejemplo de código adicionalmente más adelante.

5 La reducción 206 de la estimación de ruido de fondo permite la manipulación de situaciones en las que la estimación de ruido de fondo se ha convertido en "demasiado alta", es decir, en relación con un ruido de fondo cierto. Esto también podría expresarse, por ejemplo, como que la estimación del ruido de fondo se desvía del ruido de fondo real. Una estimación de ruido de fondo demasiado alta puede dar lugar a decisiones inadecuadas por parte del SAD, donde se determina que el segmento de señal actual esté inactivo incluso aunque comprenda voz o música activa. Una razón para que la estimación de ruido de fondo sea demasiado alta es, por ejemplo, que existan actualizaciones de ruido de fondo erróneas o no deseadas en la música, donde la estimación de ruido ha confundido la música con el fondo y ha permitido aumentar la estimación de ruido. El método divulgado permite que tal estimación de ruido de fondo actualizada erróneamente se ajuste, por ejemplo, cuando se determine que la siguiente trama de la señal de entrada comprenderá música. Este ajuste se realiza mediante una reducción forzada de la estimación del ruido de fondo, donde la estimación del ruido desciende, incluso si la energía del segmento de la señal de entrada actual es mayor que la estimación del ruido de fondo actual, por ejemplo, en una subbanda. Obsérvese que la lógica descrita anteriormente para la estimación del ruido de fondo se utiliza para controlar el aumento de la energía de la subbanda de fondo. Siempre se permite reducir la energía de la subbanda cuando la energía de la subbanda de la trama actual es inferior que la estimación del ruido de fondo. Esta función no se muestra explícitamente en la figura 2. Tal disminución generalmente tiene una configuración fija para el tamaño del paso. Sin embargo, la estimación del ruido de fondo sólo debería aumentarse en asociación con la lógica de decisión de acuerdo con el método descrito anteriormente. Cuando se detecte una pausa, las características de energía y correlación también se pueden usar para decidir cómo de grande debería ser el tamaño del paso de ajuste para el aumento de la estimación de fondo antes de que se haga la actualización real del ruido de fondo.

25 Como se mencionó anteriormente, algunos segmentos de música pueden ser difíciles de separar del ruido de fondo, debido a que son muy ruidosos. De este modo, la lógica de actualización de ruido puede permitir accidentalmente aumentar las estimaciones de energía de subbanda, incluso aunque la señal de entrada fuera una señal activa. Esto puede causar problemas, ya que la estimación del ruido puede ser más alta de lo que debería ser.

30 En los estimadores de ruido de fondo de la técnica anterior, las estimaciones de energía de subbanda sólo podían reducirse cuando una energía de subbanda de entrada estaba por debajo de una estimación de ruido actual. Sin embargo, dado que algunos segmentos de música pueden ser difíciles de separar del ruido de fondo, debido a que son muy ruidosos, los inventores se han dado cuenta de que se necesita una estrategia de recuperación para la música. En las realizaciones descritas en el presente documento, tal recuperación se puede realizar mediante reducción forzada de la estimación del ruido cuando la señal de entrada vuelve a características similares a la música. Esto es, cuando la energía y la lógica de pausa descritas anteriormente impiden, 202:1, 204:1, que se aumente la estimación de ruido, se prueba 203 si se sospecha que la entrada es música, y, si es así 203:2, las energías de banda se reducen 206 en una pequeña cantidad por cada trama hasta que las estimaciones de ruido alcancen el nivel más bajo 205:2.

40 Un estimador de fondo como los descritos anteriormente puede estar comprendido o implantado en un VAD o SAD y/o en un codificador y/o un decodificador, donde el codificador y/o decodificador puede implantarse en un dispositivo de usuario, tal como un teléfono móvil, un ordenador portátil, una tableta, etc. El estimador de fondo también podría estar comprendido en un nodo de red, tal como una pasarela de medios, por ejemplo, como parte de un códec.

50 La figura 5 es un diagrama de bloques que ilustra esquemáticamente una implantación de un estimador de fondo de acuerdo con una realización a modo de ejemplo. Un bloque 51 de entramado de entrada divide primero la señal de entrada en tramas de longitud adecuada, por ejemplo, 5-30 ms. Para cada trama, un extractor 52 de características calcula al menos las siguientes características de la entrada: 1) El extractor de características analiza la trama en el dominio de frecuencia y se calcula la energía para un conjunto de subbandas. Las subbandas son las mismas subbandas que se van a usar para la estimación de fondo. 2) El extractor de características analiza adicionalmente la trama en el dominio del tiempo y calcula una correlación denominada, por ejemplo, cor_est y/o lt_cor_est , que se utiliza para determinar si la trama comprende contenido activo o no. 3) El extractor de características utiliza adicionalmente la energía total de la trama actual, por ejemplo, denominada E_{tot} , para actualizar las características del historial de energía de las tramas de entrada actuales y anteriores, tal como la energía mínima a largo plazo, lt_min . Las características de correlación y energía se suministran después al bloque 53 de lógica de decisión de actualización.

60 Aquí, se implanta una lógica de decisión, de acuerdo con la solución descrita en el presente documento, en el bloque 53 de lógica de decisión de actualización, donde las características de correlación y energía se usan para tomar decisiones sobre si la energía de trama actual está cerca o no de una energía mínima a largo plazo; sobre si la trama actual es parte o no de una pausa (señal no activa); y si la trama actual es parte de la música o no. La solución de acuerdo con las realizaciones descritas en el presente documento implica cómo se usan estas características y decisiones para actualizar la estimación del ruido de fondo de una manera contundente.

65

A continuación, se describirán algunos detalles de implantación de realizaciones de la solución divulgada en el presente documento. Los detalles de implantación a continuación se toman de una realización en un codificador basado en el G.718. Esta realización utiliza algunas de las características descritas en WO2011/049514 y WO2011/049515,.

5

Las siguientes características se definen en el G.718 modificado descrito en el documento WO2011/09514

Etot;	La energía total para la trama de entrada actual
Etot_l	Rastrea la envoltura de energía mínima
Etot_l_lp;	Una versión suavizada de la envoltura de energía mínima Etot_l
totalNoise;	La energía total actual de la estimación de fondo.
bckr[i];	El vector con las estimaciones de fondo de subbanda
tmpN[i];	Una nueva estimación potencial precalculada de fondo
aEn;	Un detector de fondo que utiliza múltiples funciones (un contador)
harm_cor_cnt	Cuenta las tramas desde la última trama con correlación o evento armónico
act_pred	Una predicción de la actividad sólo a partir de las características de la trama de entrada
cor[i]	Vector con estimaciones de correlación para, i=0 final de la trama actual, i=1 inicio de la trama actual, i=2 final de la trama anterior

10

Las siguientes características se definen en el G.718 modificado descrito en el documento WO2011/09515

Etot_h	Rastrea la envoltura de energía máxima
sign_dyn_lp;	Una dinámica de señal de entrada suavizada

También la característica Etot_v_h fue definida en el documento WO2011/049514, pero en esta realización se ha modificado y ahora se implanta de la siguiente manera:

```

Etot_v = (float) fabs(*Etot_last - Etot);
if(Etot_v < 7.0f) /*note that no VAD flag or similar is used here*/
{
    *Etot_v_h -= 0.01f;
    if (Etot_v > *Etot_v_h)
    {if ((*Etot_v - *Etot_v_h) > 0.2f)
        {
            *Etot_v_h = *Etot_v_h + 0.2f;
        }
        else
        {
            *Etot_v_h = Etot_v; }}}

```

15

La Etot_v mide la variación de energía absoluta entre tramas, es decir, el valor absoluto de la variación de energía instantánea entre tramas. En el ejemplo anterior, se determina que la variación de energía entre dos tramas es "baja" cuando la diferencia entre las tramas última y actual es menor a 7 unidades. Esto se utiliza como un indicador de que la trama actual (y la trama anterior) puede ser parte de una pausa, es decir, comprender sólo ruido de fondo. Sin embargo, una variación baja tal podría encontrarse alternativamente, por ejemplo, en medio de una ráfaga de voz. La variable Etot_last es el nivel de energía de trama anterior.

20

Los pasos anteriores descritos en el código se pueden realizar como parte de los pasos "correlación calcular/actualizar y energía" en el diagrama de flujo en la figura 2, es decir, como parte de las acciones 201. En la implantación WO2011/049514, se utilizó un indicador de VAD para determinar si el segmento de señal de audio actual comprendía ruido de fondo o no. Los inventores se han dado cuenta de que la dependencia de la información de retroalimentación puede ser problemática. En la solución divulgada en el presente documento, la decisión de actualizar o no la estimación del ruido de fondo no depende de una decisión del VAD (ni del SAD).

25

30

Además, en la solución descrita en el presente documento, las siguientes características, que no son parte de la implantación WO2011 049514, se pueden calcular/actualizar como parte de los mismos pasos, es decir, los pasos correlación calcular/actualización y energía que se ilustran en la figura 2. Estas características también se utilizan en la lógica de decisión de si actualizar o no la estimación de fondo.

Con el fin de lograr una estimación de ruido de fondo más adecuada, se definen en adelante una serie de características. Por ejemplo, se definen las nuevas funciones relacionadas con la correlación, `cor_est` y `lt_cor_est`. La característica `cor_est` es una estimación de la correlación en la trama actual, y `cor_est` se usa también para producir `lt_cor_est`, que es una estimación suavizada a largo plazo de la correlación.

```
cor_est = (cor[0] + cor[1] + cor[2]) / 3.0f ;
```

```
st->lt_cor_est = 0.01f*cor_est + 0.99f * st->lt_cor_est;
```

Como se definió anteriormente, `cor[i]` es un vector que comprende estimaciones de correlación, y `cor[0]` representa el final de la trama actual, `cor[1]` representa el comienzo de la trama actual y `cor[2]` representa el final de una trama anterior.

Además, se calcula una nueva característica, `lt_tn_track`, que proporciona una estimación a largo plazo de la frecuencia con la que las estimaciones de fondo están cerca de la energía de trama actual. Cuando la energía de trama actual está lo suficientemente cerca de la estimación de fondo actual, esto se registra por una condición que señala (1/0) si el fondo está cerca o no. Esta señal se utiliza para formar el `lt_tn_track` de medida a largo plazo.

```
st->lt_tn_track = 0,03f* (Etot - st->totalNoise < 10) + 0.97f*st->lt_tn_track;
```

En este ejemplo, se añade 0,03 cuando la energía de trama actual está cerca de la estimación de ruido de fondo, y, de otro modo, el único término restante es 0,97 veces el valor anterior. En este ejemplo, "cerca" se define como que la diferencia entre la energía de trama actual, `Etot`, y la estimación del ruido de fondo, `totalNoise`, es inferior a 10 unidades. También son posibles otras definiciones de "cerca".

Además, la distancia entre la estimación de fondo actual, `Etot`, y la energía de trama actual, `totalNoise`, se usa para determinar una característica, `lt_tn_dist`, que proporciona una estimación a largo plazo de esta distancia. Se crea una característica similar, `lt_Ellp_dist`, para la distancia entre la energía mínima a largo plazo, `Etot_l_lp`, y la energía de trama actual, `Etot`.

```
st->lt_tn_dist = 0.03f* (Etot - st->totalNoise) + 0.97f*st->lt_tn_dist;
```

```
st->lt_Ellp_dist = 0.03f* (Etot - st->Etot_l_lp) + 0.97f*st->lt_Ellp_dist;
```

La característica `harm_cor_cnt`, introducida anteriormente, se usa para contar el número de tramas desde la última trama que tiene una correlación o un evento armónico, es decir, desde una trama de cumplimiento de ciertos criterios relacionados con la actividad. Esto es, cuando la condición `harm_cor_cnt==0`, esto implica que la trama actual es, probablemente, una trama activa, ya que muestra correlación o un evento armónico. Esto se utiliza para formar una estimación suavizada a largo plazo, `lt_haco_ev`, de la frecuencia con la que ocurren tales eventos. En este caso, la actualización no es simétrica, es decir, que se usan constantes de tiempo diferentes si la estimación aumenta o disminuye, como se puede ver a continuación.

```
if (st->harm_cor_cnt == 0) /*when probably active*/
{
    st->lt_haco_ev = 0,03f + 0.97f*st->lt_haco_ev; /*increase long term estimate*/
} else
{
    st->lt_haco_ev = 0.99f*st->lt_haco_ev; /*decrease long term estimate */ }

```

Un valor bajo de la característica `lt_tn_track`, presentado anteriormente, indica que la energía de trama de entrada no ha estado cerca de la energía de fondo para algunas tramas. Esto se debe a que el `lt_tn_track` se reduce para cada trama donde la energía de trama actual no está cerca de la estimación de energía de fondo. El `lt_tn_track` aumenta sólo cuando la energía de trama actual está cerca de la estimación de energía de fondo como se mostró anteriormente. Para obtener una mejor estimación de cuánto tiempo ha durado este "no seguimiento", es decir, que la energía de trama está lejos de la estimación de fondo, un contador, `low_tn_track_cnt`, para el número de tramas con esta ausencia de seguimiento, se forma como:

```
if (st->lt_tn_track<0.05f) /*when lt_tn_track is low */
{
    st->low_tn_track_cnt++; /*add 1 to counter */
} else
{
    st->low_tn_track_cnt=0; /*reset counter */
}

```

En el ejemplo anterior, "bajo" se define como por debajo del valor 0,05. Esto debería verse como un valor de ejemplo, que podría seleccionarse de manera diferente.

- 5 Para el paso "Formar pausa y decisiones de música" ilustrado en la figura 2, las siguientes tres expresiones de código se usan para formar la detección de pausa, también denotada detección de fondo. En otras realizaciones e implantaciones, también podrían añadirse otros criterios para la detección de pausa. La decisión real de música se forma en el código utilizando características de correlación y energía.

10 1.-

```
bg_bgd = Etot < Etot_l_lp + 0.6f * st->Etot_v_h;
```

15 El bg_bgd se convertirá en "1" o "verdadero" cuando la Etot esté cerca de la estimación del ruido de fondo. El bg_bgd sirve como una máscara para otros detectores de fondo. Es decir, si el bg_bgd no es "verdadero", no es necesario evaluar los detectores 2 y 3 de fondo a continuación. La Etot_v_h es una estimación de variación de ruido, que alternativamente podría denotarse N_{var} . La Etot_v_h se deriva de la energía total de entrada (en el dominio de log) usando la Etot_v, que mide la variación de energía absoluta entre tramas. Obsérvese que la característica Etot_v_h se limita a aumentar sólo un máximo de un valor constante bajo, por ejemplo, 0,2 para cada trama. La Etot_l_lp es una versión suavizada de la envoltura de energía mínima Etot_l.

20

2.-

```
aE_bgd = st->aEn == 0;
```

25

Cuando la aEn es cero, la aE_bgd se convierte en "1" o "verdadera". La aEn es un contador que aumenta cuando se determina que una señal activa está presente en una trama actual, y disminuye cuando se determina que la trama actual no comprende una señal activa. La aEn no puede aumentarse más que hasta un cierto número, por ejemplo, 6, y no puede reducirse a menos de cero. Después de varias tramas consecutivas, por ejemplo, 6, sin una señal activa, la aEn será igual a cero.

30

3.-

```
sd1_bgd = (st->sign_dyn_lp > 15) && (Etot - st->Etot_l_lp) < st->Etot_v_h && st->harm_cor_cnt > 20;
```

35

En el presente documento, la sd1_bgd será "1" o "verdadera" cuando tres condiciones diferentes son verdaderas: la dinámica de la señal, sign_dyn_lp es alta, en este ejemplo más de 15; la energía de trama actual está cerca de la estimación de fondo; y: cierto número de tramas han pasado sin correlación o eventos armónicos, en este ejemplo 20 tramas.

40

La función del bg_bgd es ser un indicador para detectar que la energía de trama actual está cerca de la energía mínima a largo plazo. Los dos últimos, el aE_bgd y el sd1_bgd representan pausa o detección de fondo en diferentes condiciones. El aE_bgd es el detector más general de los dos, mientras que el sd1_bgd detecta principalmente pausas de voz en una SNR alta. Una nueva lógica de decisión de acuerdo con una realización de la tecnología divulgada en el presente documento, se construye como sigue en el código a continuación. La lógica de decisión comprende la condición del bg_bgd de enmascaramiento y los dos detectores aE_bgd y sd1_bgd de pausa. También podría haber un tercer detector de pausa, que evalúa las estadísticas a largo plazo lo bien que el TotalNoise rastrea la estimación de energía mínima. Las condiciones evaluadas si la primera línea es verdadera es la lógica de decisión sobre lo grande que debería ser el tamaño del paso, updt_step y la actualización de estimación de ruido real es la asignación de valor a "st->bckr [i]=". Obsérvese que el tmpN[i] es un nivel de ruido potencialmente nuevo anteriormente calculado, calculado de acuerdo con la solución descrita en el documento WO2011/049514. La lógica de decisión a continuación sigue la parte 209 de la figura 2, que se indica en parte en conexión con el siguiente Código

50

ES 2 758 517 T3

```
if (bg_bgd && (aE_bgd II sdl_bgd II st->lt_tn_track >0.90f)) /*if 202:2 and 204:2*/
{
    if( (st->act_pred < 0.85f II (aE_bgd && st->lt_haco_ev < 0.05f)) &&
        (st->lt_Ellp_dist < 10 II sdl_bgd ) && st->it_tn_dist<40 &&
        ((Etot - st->totalNoise) < 15.0f II st->lt_haco_ev < 0.10f)) /*207*/
    {
        st->first_noise_updt = 1;
        for(i=0; i< NB_BANDS; i++)
        {
            st->bckr[i] = tmpN[i] /*208*/
        }
    }
    else if (aE_bgd && st->lt_haco_ev < 0.15f)
    {
        updt_step=0.1f;
        if (st->act_pred > 0.85f)
        {
            updt_step=0.01f /*207*/ }
        if (updt_step > 0.0f)

        {
            st->first_noise_updt = 1;
            for(i=0; i< NB_BANDS; i++)
            {
                st->bckr[i] = st->bckr[i] + updt_step * (tmpN[i]-st->bckr[i]); /*208*/
            }
        }
    }
    else
    {
        (st->first_noise_updt) +=1;
    }
}
else
{
    /* If in music lower bckr to drop further *//*if 203:2 and 205:1*/
    If (st->low_tn_track_cnt > 300 && st->lt_haco_ev > 0.9f && st-> totalNoise > 0.0f)
    { For (i=0; i< NB_BANDS; i++)
      { If (st->bckr[i] > 2 * E_MIN
        {
            St->bckr[i] = 0.98f * st->bckr[i]; /*206*/
        } } }
    Else
    {
        (st->first_noise_updt) += 1;
    }
}
```

- 5 El segmento de código en el último bloque de código que comienza con `/* If in music ... */` contiene el descenso forzado de la estimación de fondo que se usa si se sospecha que la entrada actual es música. Esto se decide como una función: largo período de ruido de fondo de seguimiento pobre en comparación con la estimación de energía mínima, AND, ocurrencias frecuentes de eventos armónicos o de correlación, AND, la última condición `"totalNoise>0"` es una comprobación de que la energía total actual de la estimación de fondo es mayor que cero, lo
- 10 que implica que se puede considerar una reducción de la estimación de fondo. Además, se determina si `"bckr[i]>2 * E_MIN"`, donde `E_MIN` es un valor positivo bajo. Esta es una comprobación de cada entrada en un vector que comprende las estimaciones de fondo de subbanda, de tal manera que una entrada tiene que exceder la `E_MIN` con el fin de poder reducirse (en el ejemplo, al multiplicarse por 0,98). Estas comprobaciones se hacen con el fin de evitar reducir las estimaciones de fondo en valores demasiado bajos.
- 15 Las realizaciones mejoran la estimación del ruido de fondo, la cual permite un rendimiento mejorado del SAD/VAD para lograr una solución de DTX de alta eficiencia y evitar la degradación de la calidad del habla o de la música causada por el recorte.
- 20 Con la eliminación de la retroalimentación de decisión, descrita en el documento WO2011/09514, a partir de la `Etot_v_h`, hay una mejor separación entre la estimación de ruido y el SAD. Esto tiene beneficios ya que la estimación del ruido no cambia si/cuando se cambia la función/el ajuste del SAD. Es decir, la determinación de una estimación de ruido de fondo se vuelve independiente de la función del SAD. Además, el ajuste de la lógica de estimación de ruido se vuelve más fácil ya que no se ve afectada por los efectos secundarios del SAD cuando se modifican las
- 25 estimaciones de fondo.

REIVINDICACIONES

1. Un método para estimar el ruido de fondo en una señal de audio, comprendiendo, el método:
- 5 a) obtener (201) al menos un parámetro asociado con un segmento de señal de audio de entrada en base a:
- una primera ganancia de predicción lineal calculada como cociente entre una energía de la señal de entrada y una energía de señal residual de una primera predicción lineal para el segmento de señal de audio; y
 - 10 - una segunda ganancia de predicción lineal calculada como el cociente entre la energía de señal residual de la primera predicción lineal y una energía de señal residual de una segunda predicción lineal para el segmento de señal de audio;
- 15 b) determinar (202) si el segmento de señal de audio comprende una pausa libre de voz y música, en base al menos en el al menos un parámetro; y:
- si se determina que el segmento de señal de audio comprende una pausa:
- actualizar (203) una estimación de ruido de fondo en base al segmento de señal de audio.
- 20 2. El método de acuerdo con la reivindicación 1, en el que la primera predicción lineal es una predicción lineal de segundo orden y la segunda predicción lineal es una predicción lineal de 16º orden.
- 25 3. El método de acuerdo con la reivindicación 1 o 2, en el que la obtención del al menos un parámetro comprende: limitar las ganancias de predicción lineal primera y segunda para tomar valores en un intervalo predefinido.
- 30 4. El método de acuerdo con cualquiera de las reivindicaciones 1 a 3, en el que la obtención del al menos un parámetro comprende: crear al menos una estimación a largo plazo de cada una de las ganancias de predicción lineal primera y segunda, en el que la estimación a largo plazo se basa adicionalmente en las ganancias de predicción lineal correspondientes asociadas con al menos un segmento de señal de audio anterior.
- 35 5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que la obtención del al menos un parámetro comprende: determinar una diferencia entre una de las ganancias de predicción lineal asociadas con el segmento de señal de audio y una estimación a largo plazo de dicha ganancia de predicción lineal.
- 40 6. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que la obtención del al menos un parámetro comprende: determinar una diferencia entre dos estimaciones a largo plazo asociadas con una de las ganancias de predicción lineal.
- 45 7. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en el que la obtención del al menos un parámetro comprende el filtrado de paso bajo de las ganancias de predicción lineal primera y segunda.
- 50 8. El método de acuerdo con la reivindicación 7, en el que los coeficientes de filtro de al menos un filtro de paso bajo dependen de una relación entre una ganancia de predicción lineal asociada con el segmento de señal de audio y un promedio de una ganancia de predicción lineal correspondiente obtenida en base a una pluralidad de segmentos anteriores de señal de audio.
- 55 9. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que la determinación de si el segmento de señal de audio comprende una pausa se basa adicionalmente en una medida de cercanía espectral asociada con el segmento de señal de audio.
- 60 10. El método de acuerdo con la reivindicación 9, que comprende adicionalmente obtener la medida de cercanía espectral en base a energías para un conjunto de bandas de frecuencia del segmento de señal de audio y estimaciones de ruido de fondo correspondientes al conjunto de bandas de frecuencia.
11. El método de acuerdo con la reivindicación 10, en el que, durante un período de inicialización, se usa un valor inicial E_{min} como las estimaciones de ruido de fondo en base a las cuales se obtiene la medida de cercanía espectral.
- 65 12. Un aparato (1100), para estimar el ruido de fondo en una señal de audio que comprende una pluralidad de segmentos de señal de audio, estando el aparato configurado para:
- a) obtener al menos un parámetro en base a:

- una primera ganancia de predicción lineal calculada como el cociente entre una energía del segmento de señal de audio y una energía de señal residual de una primera predicción lineal para el segmento de señal de audio; y
 - una segunda ganancia de predicción lineal calculada como el cociente entre la energía de señal residual de la primera predicción lineal y una energía de señal residual de una segunda predicción lineal para el segmento de señal de audio;
- 5
- b) determinar si el segmento de señal de audio comprende una pausa libre de voz y música, en base al menos en el al menos un parámetro; y
- 10
- si se determina que el segmento de señal de audio comprende una pausa:
- c) actualizar una estimación de ruido de fondo en base al segmento de señal de audio.
- 15
- 13.** El aparato de acuerdo con la reivindicación 12, en el que el aparato está configurado adicionalmente para realizar el método de acuerdo con cualquiera de las reivindicaciones 2 a 11.
- 14.** Un códec de audio que comprende el aparato de acuerdo con la reivindicación 12 o 13.
- 20
- 15.** Un dispositivo de comunicación que comprende el aparato de acuerdo con la reivindicación 12 o 13.

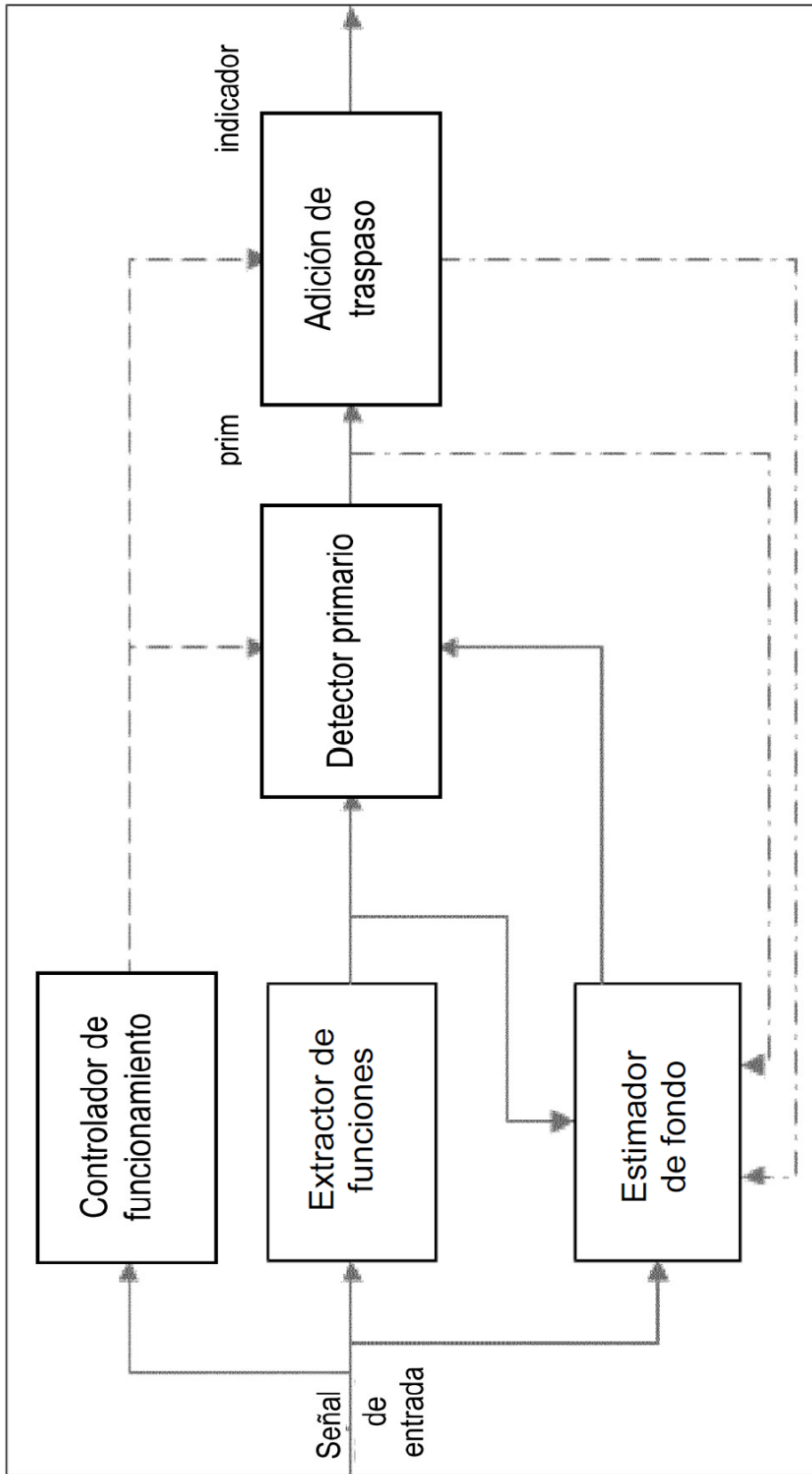


Figura 1

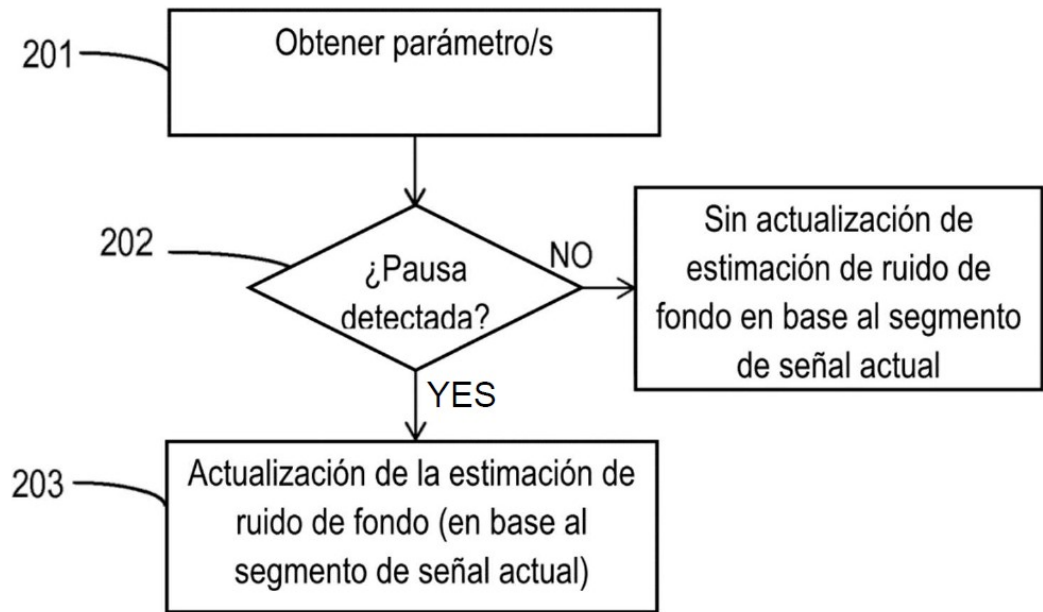


Figura 2

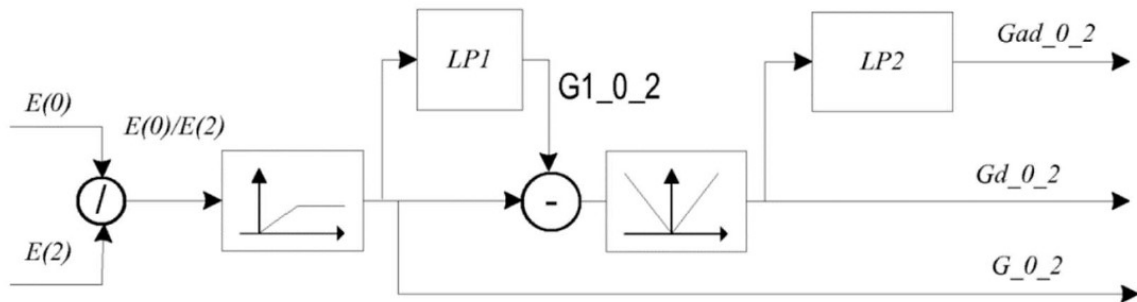


Figura 3

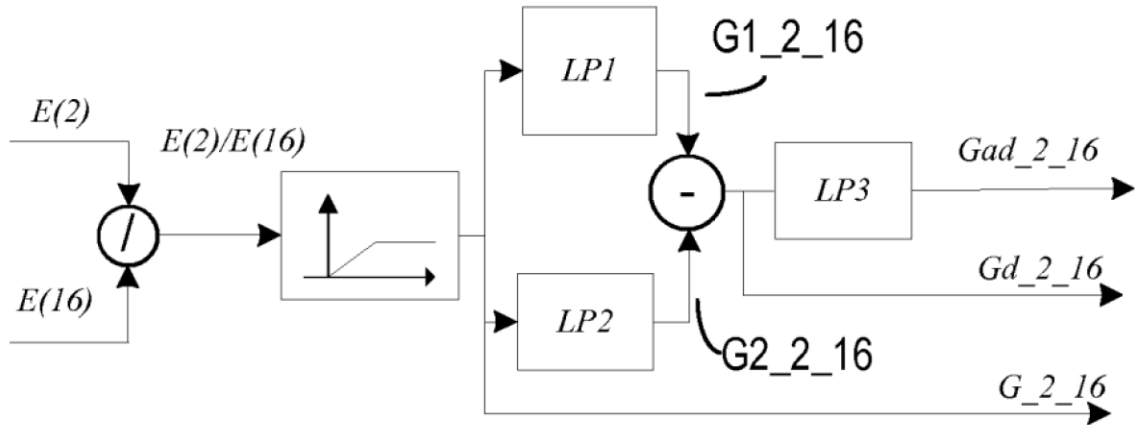


Figura 4

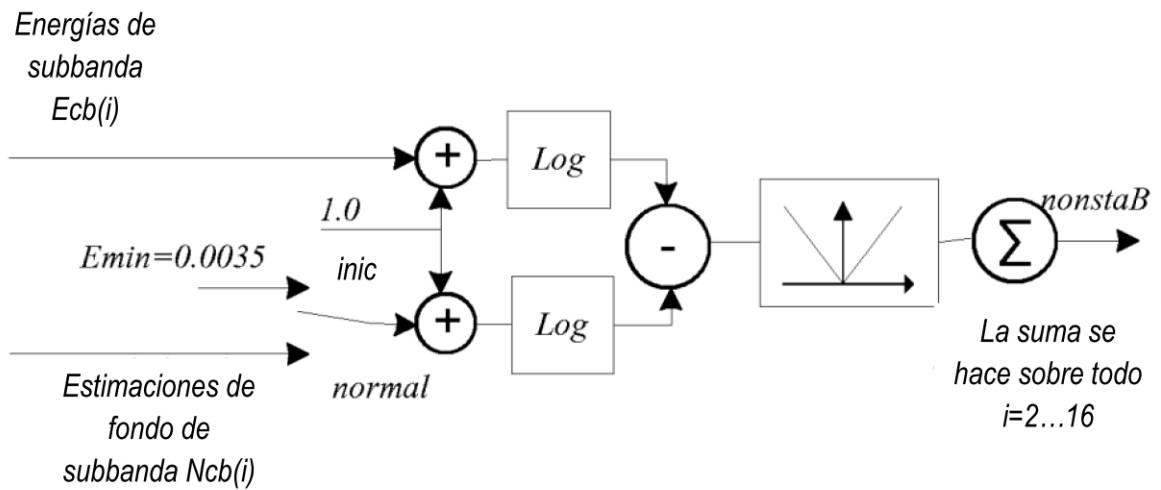


Figura 5

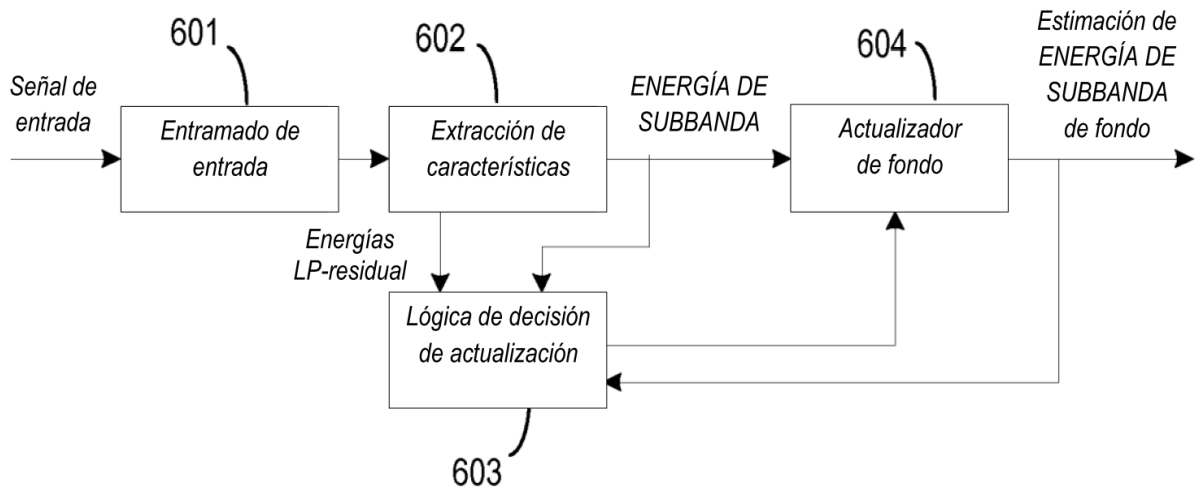


Figura 6

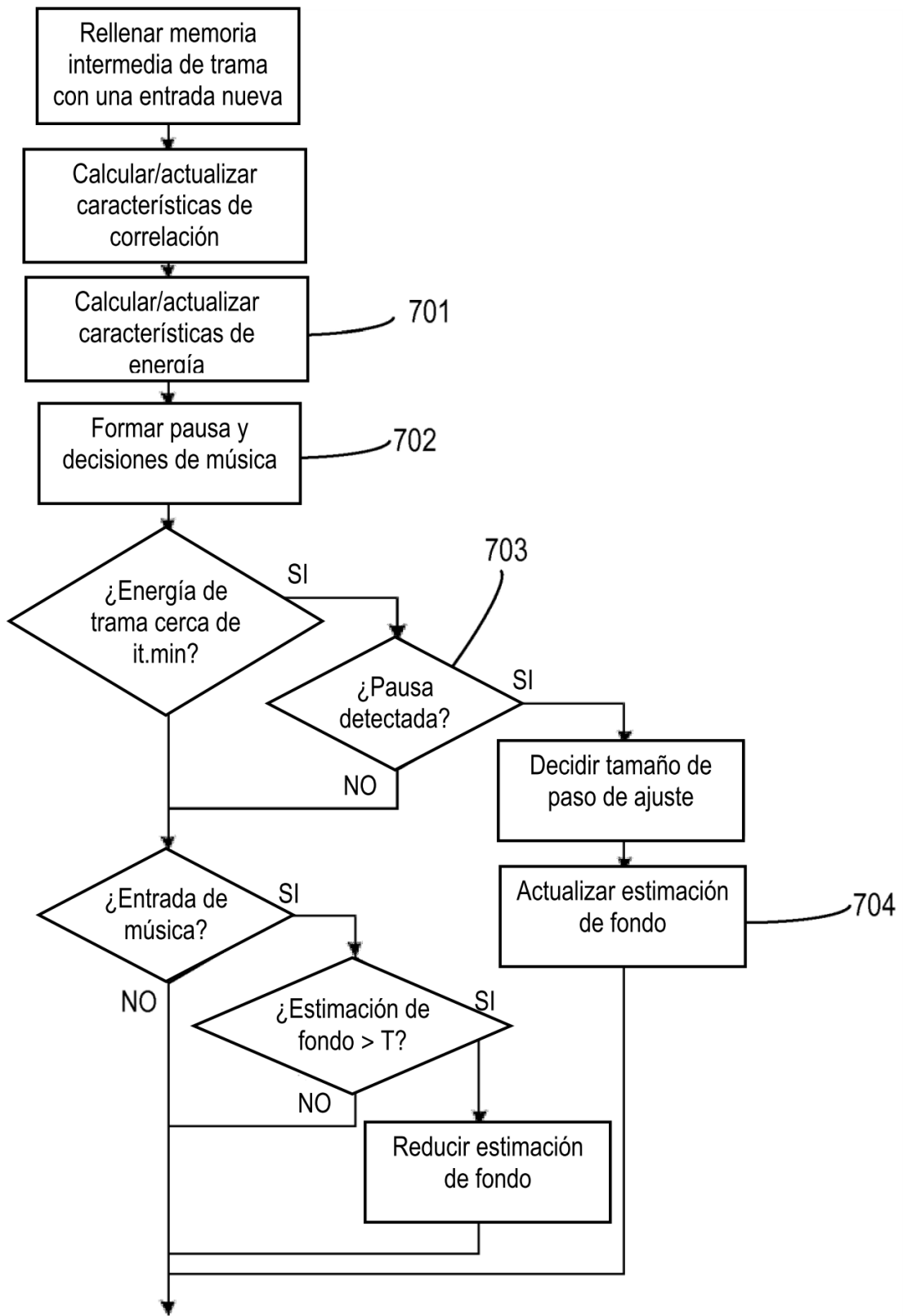


Figura 7

Función de características E(0) y E(2)

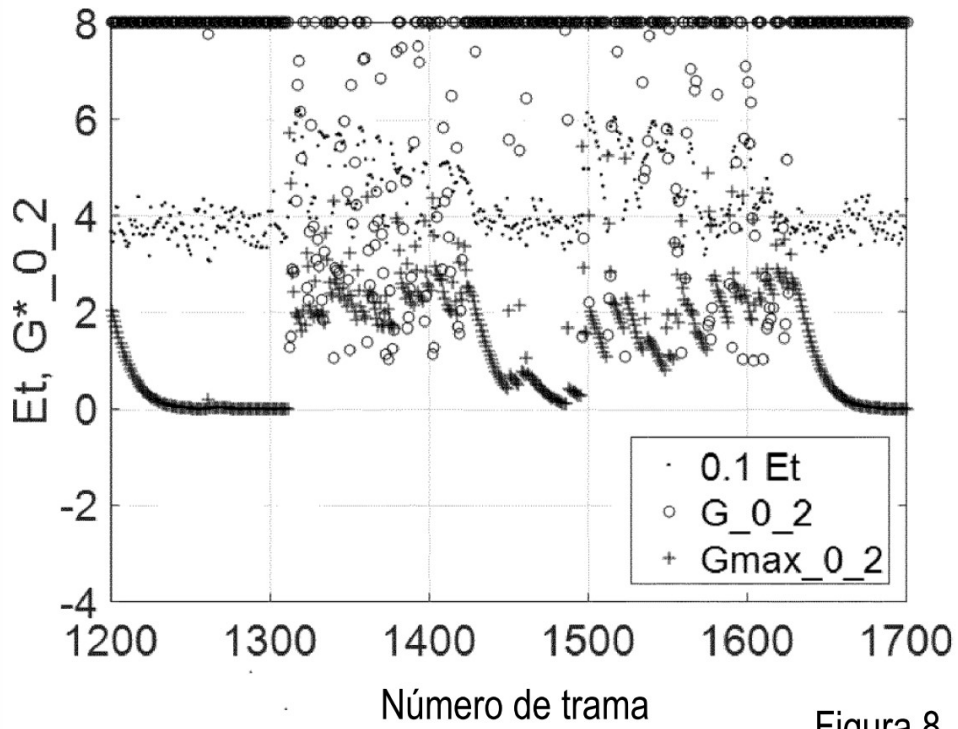


Figura 8

Función de características 1 E(2) y E(16)

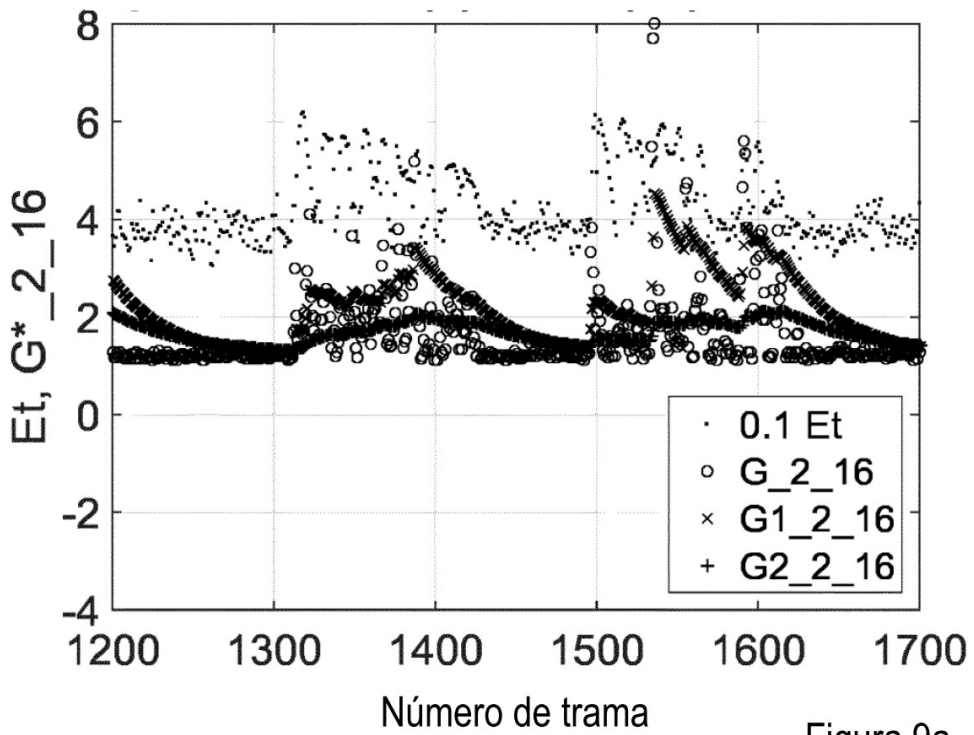
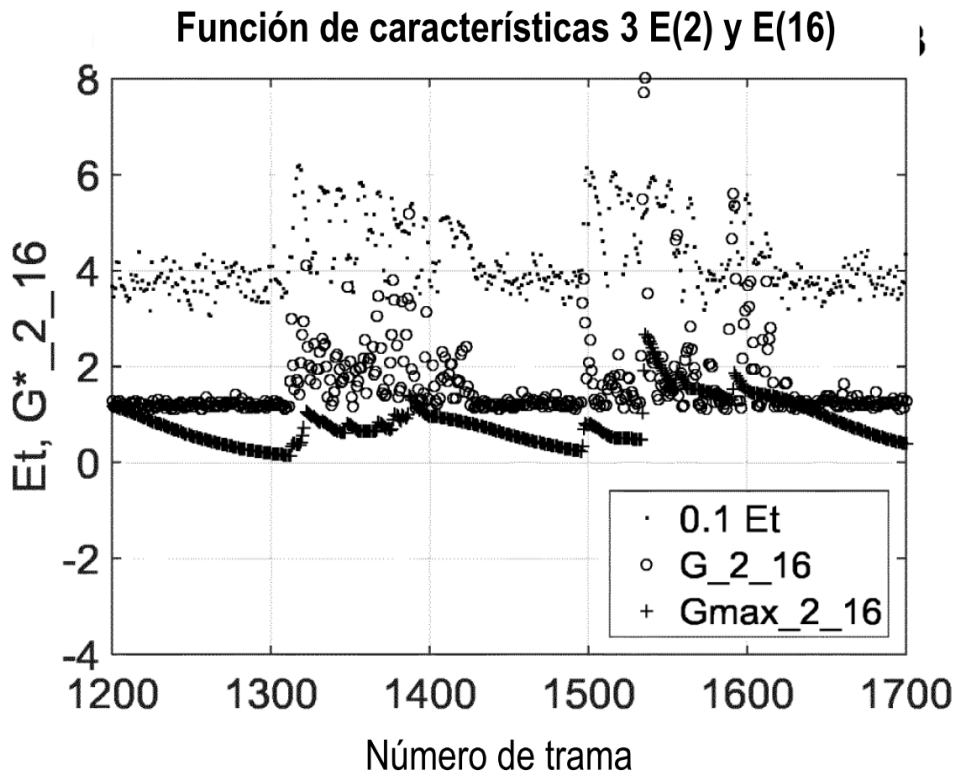
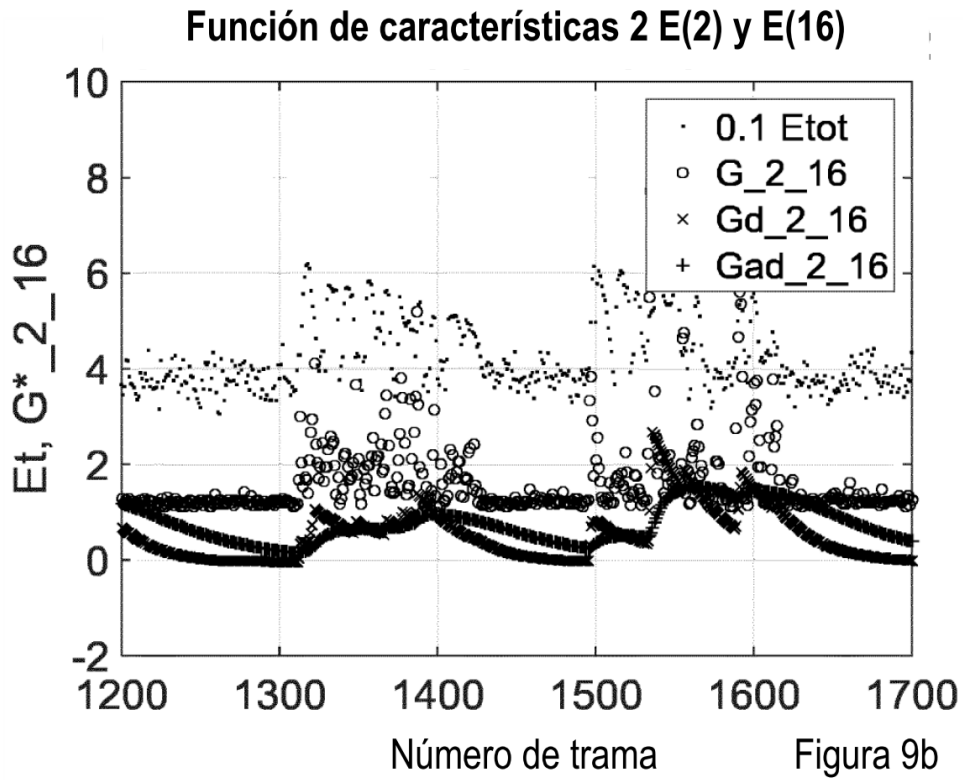


Figura 9a



Función de características non_staB

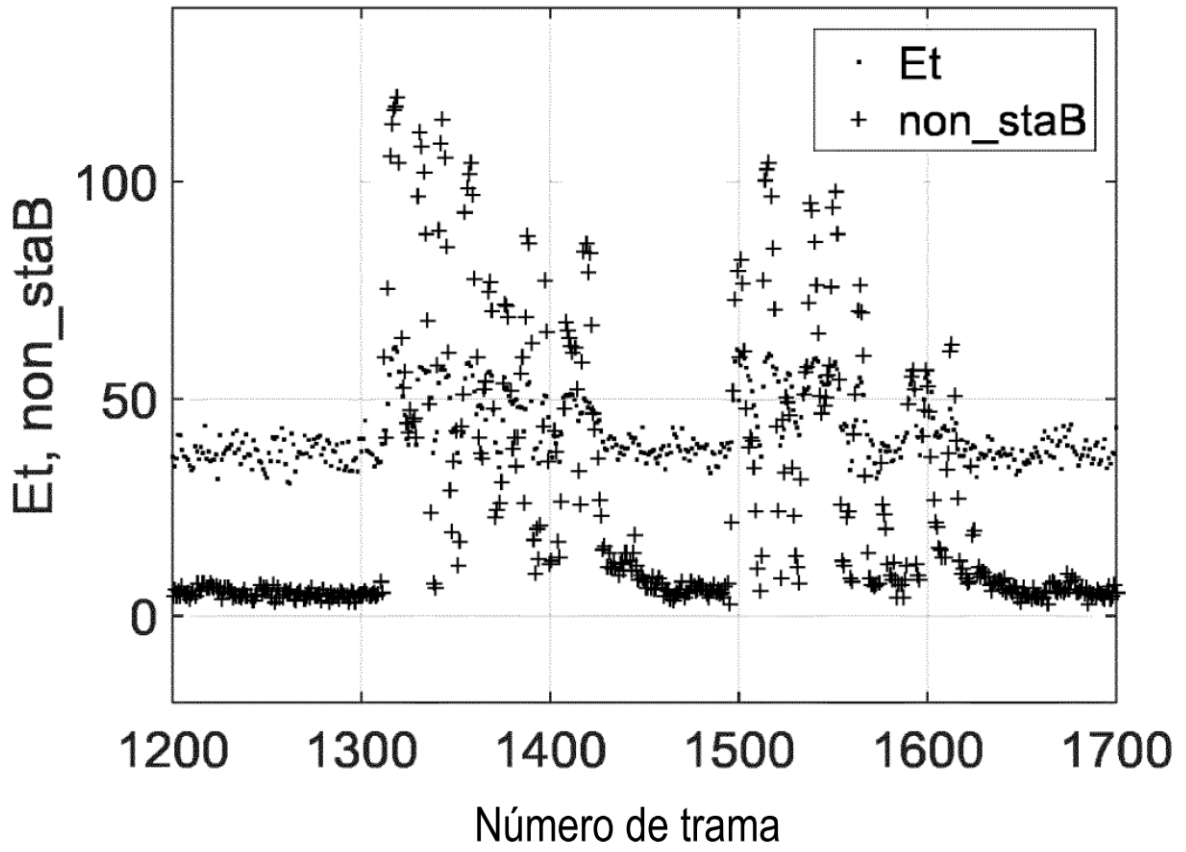


Figura 10

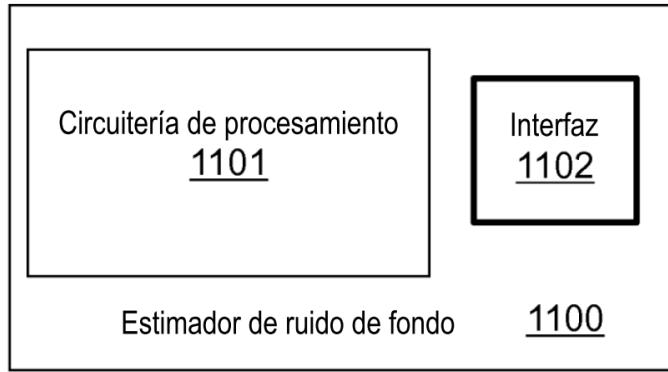


Figura 11a

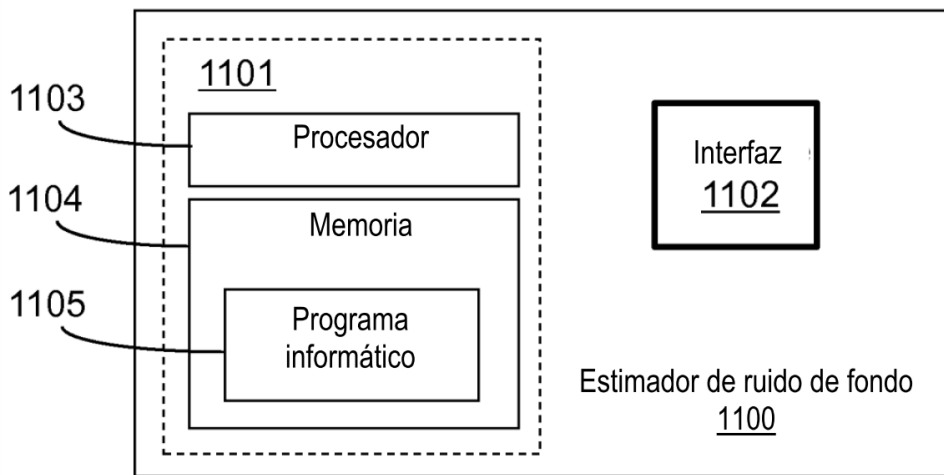


Figura 11b

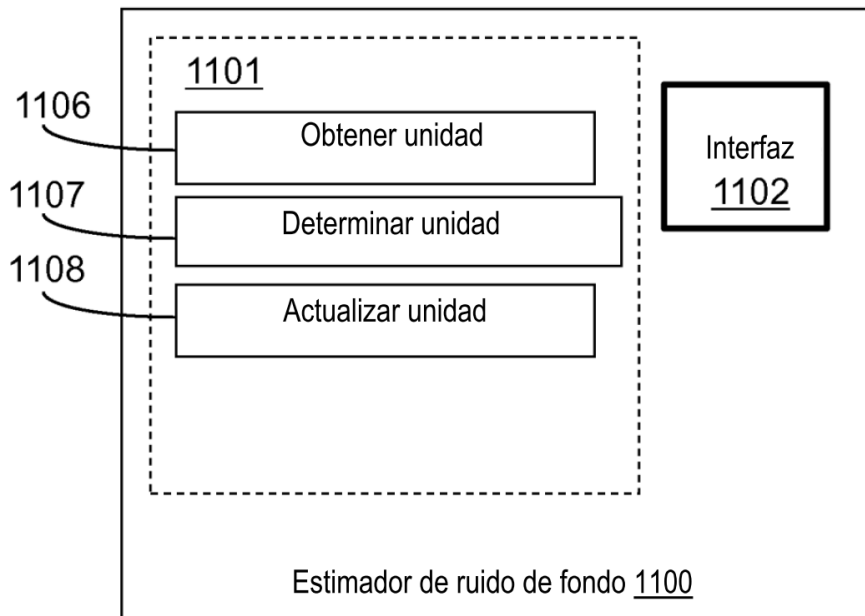


Figura 11c

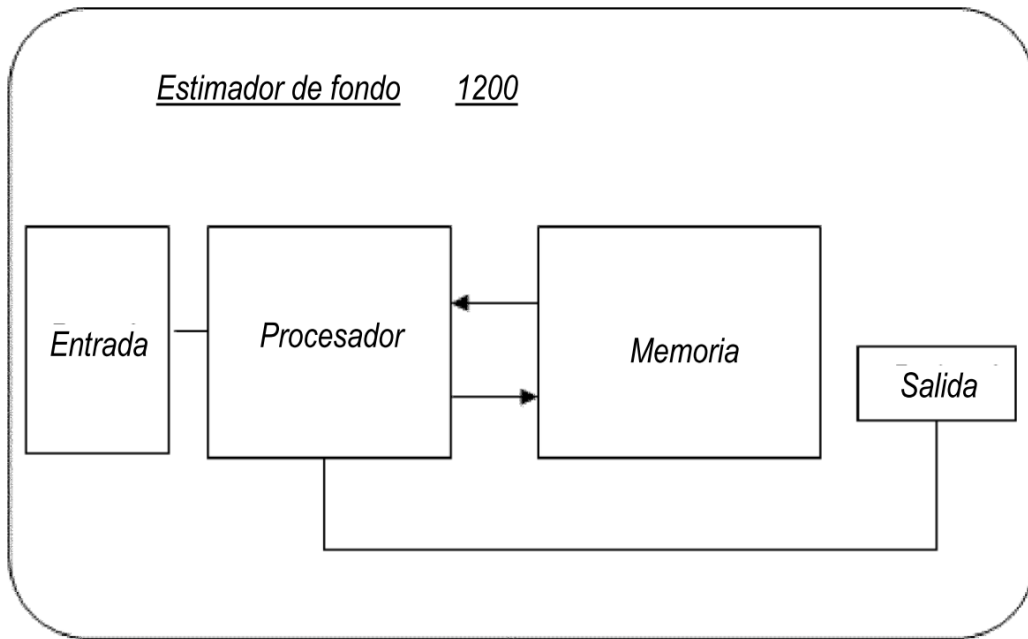


Figura 12

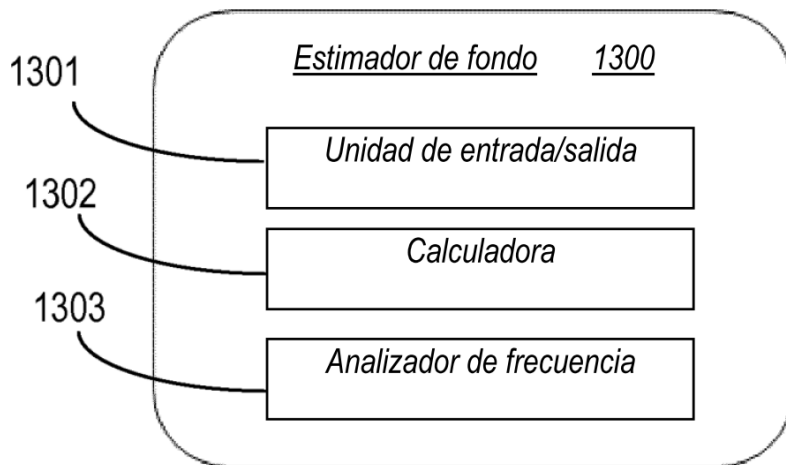


Figura 13

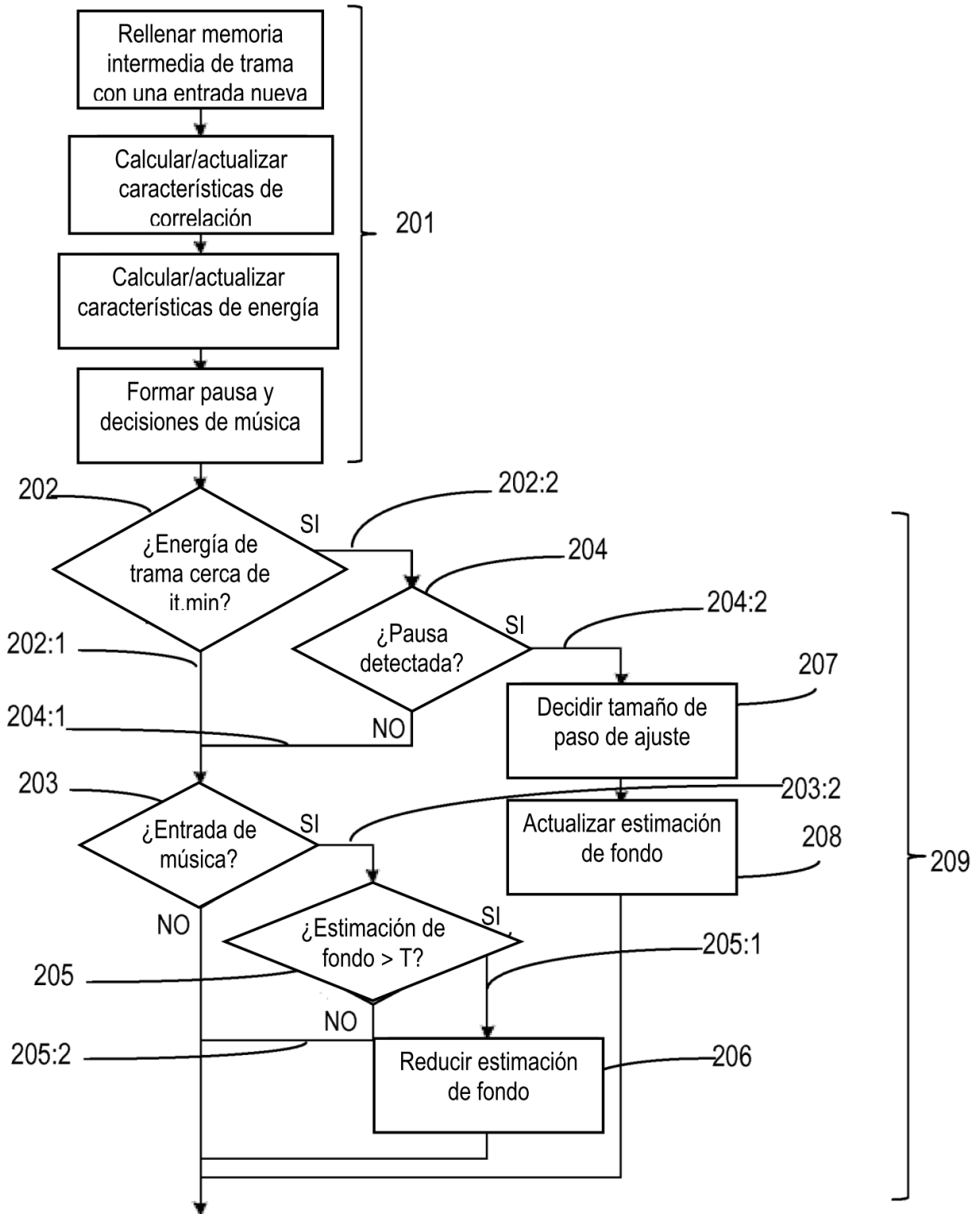


Figura A2

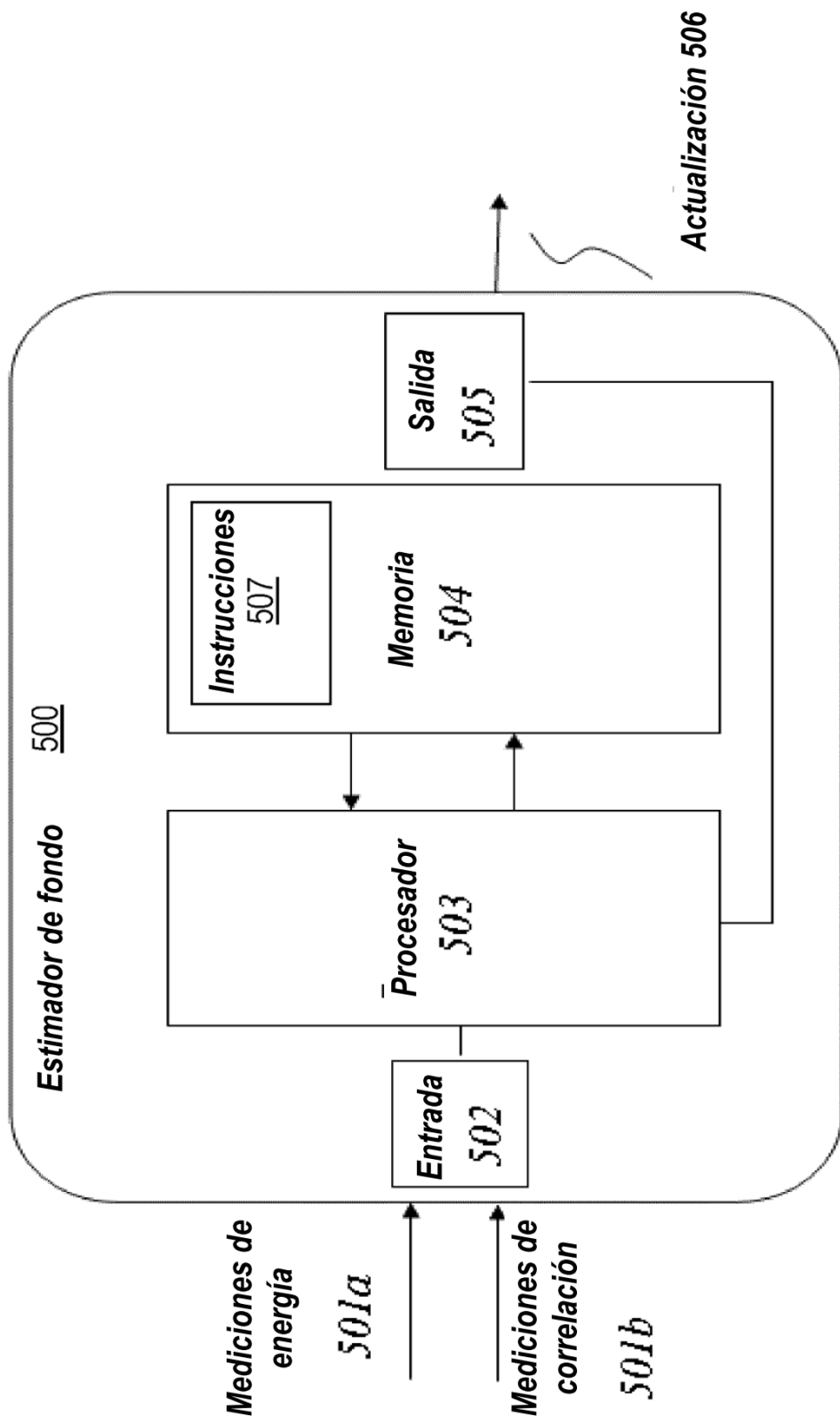


Figura A3

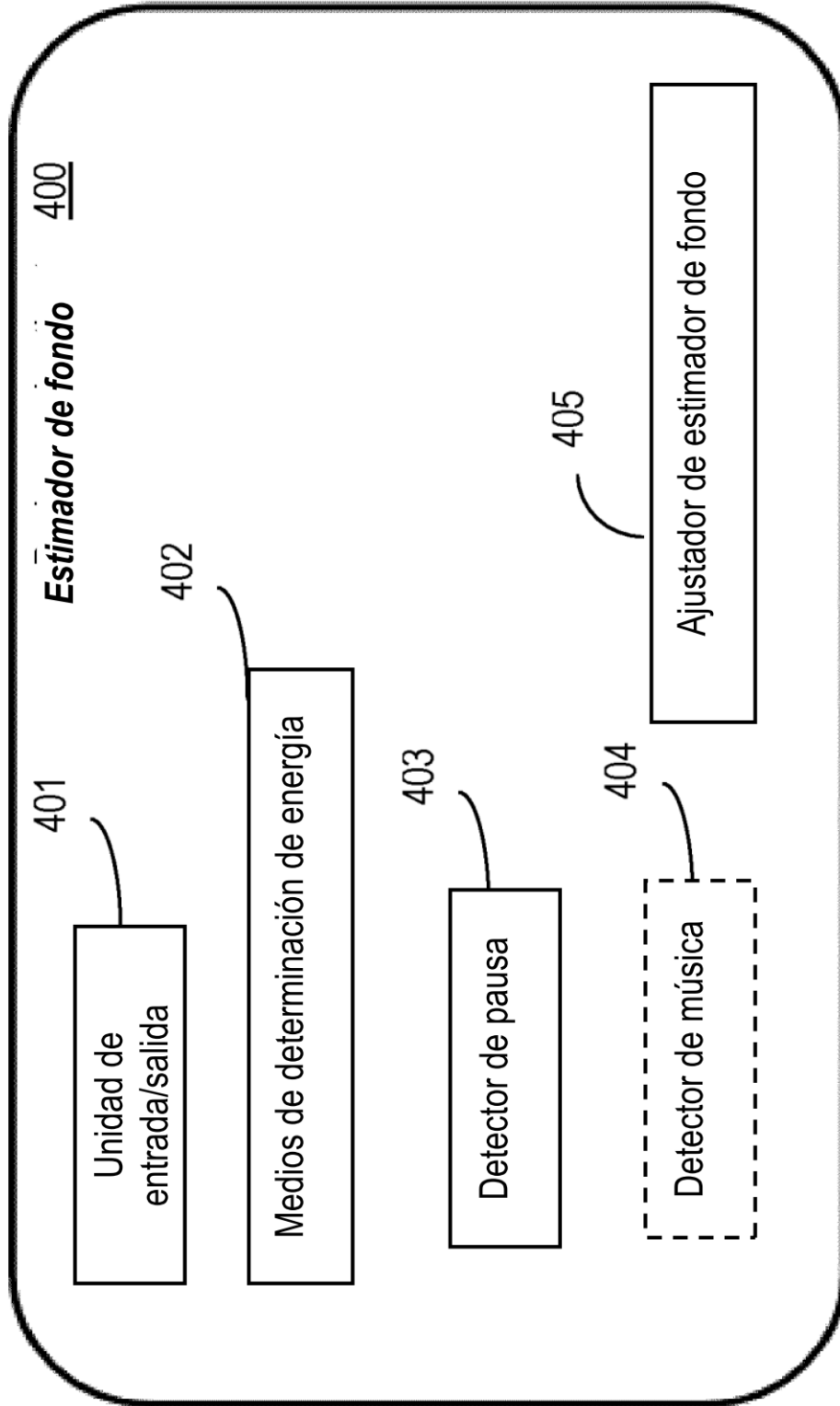


Figura A4

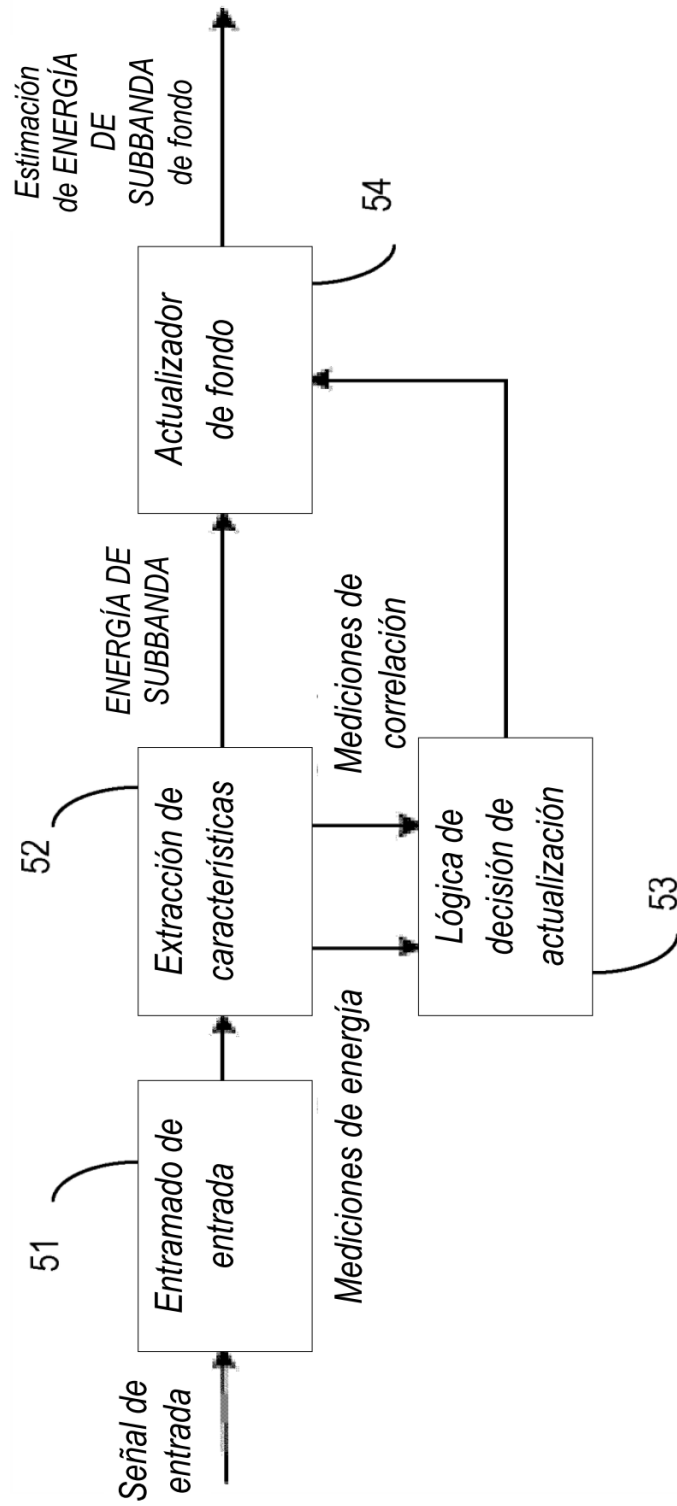


Figura A5

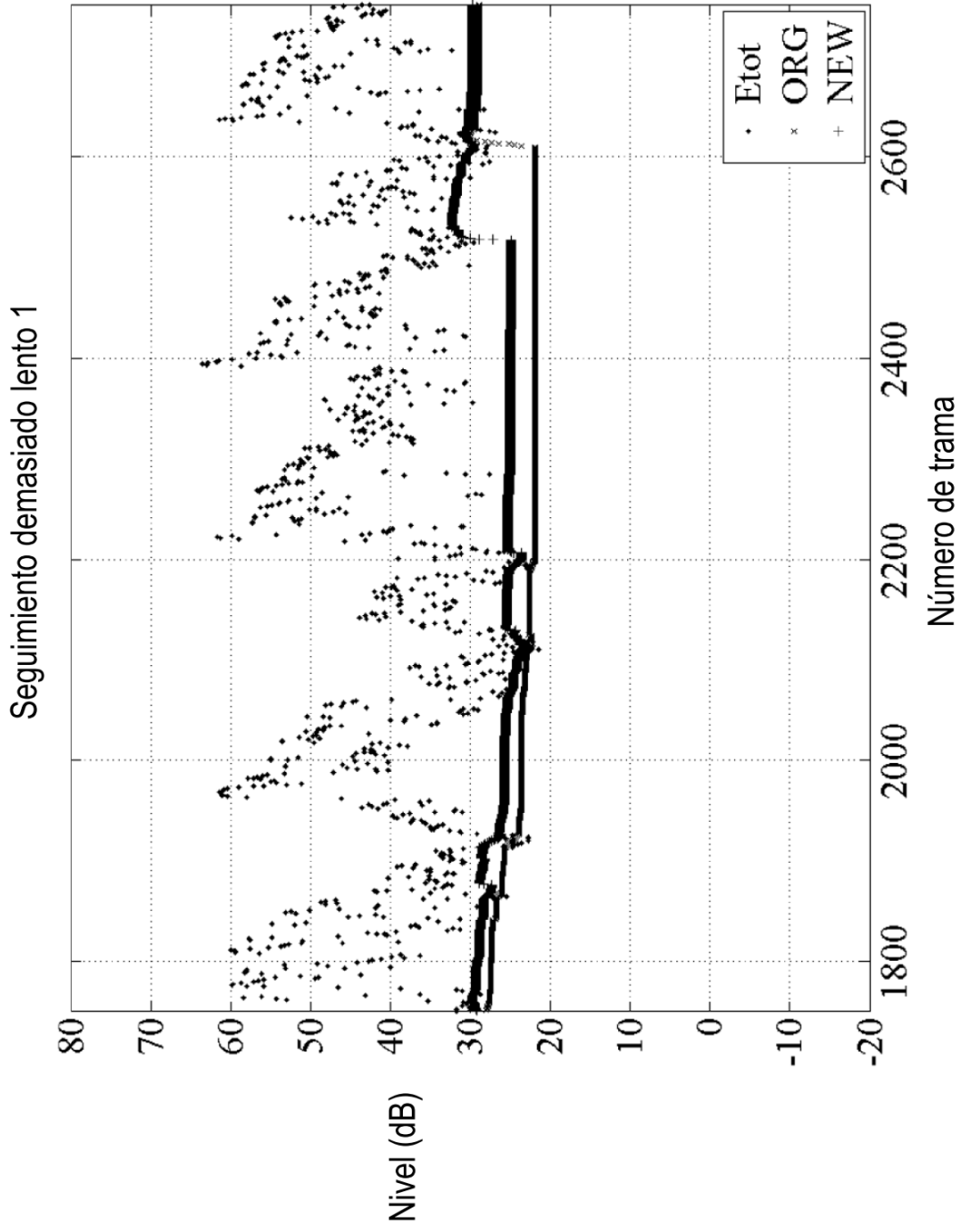


Figura A6

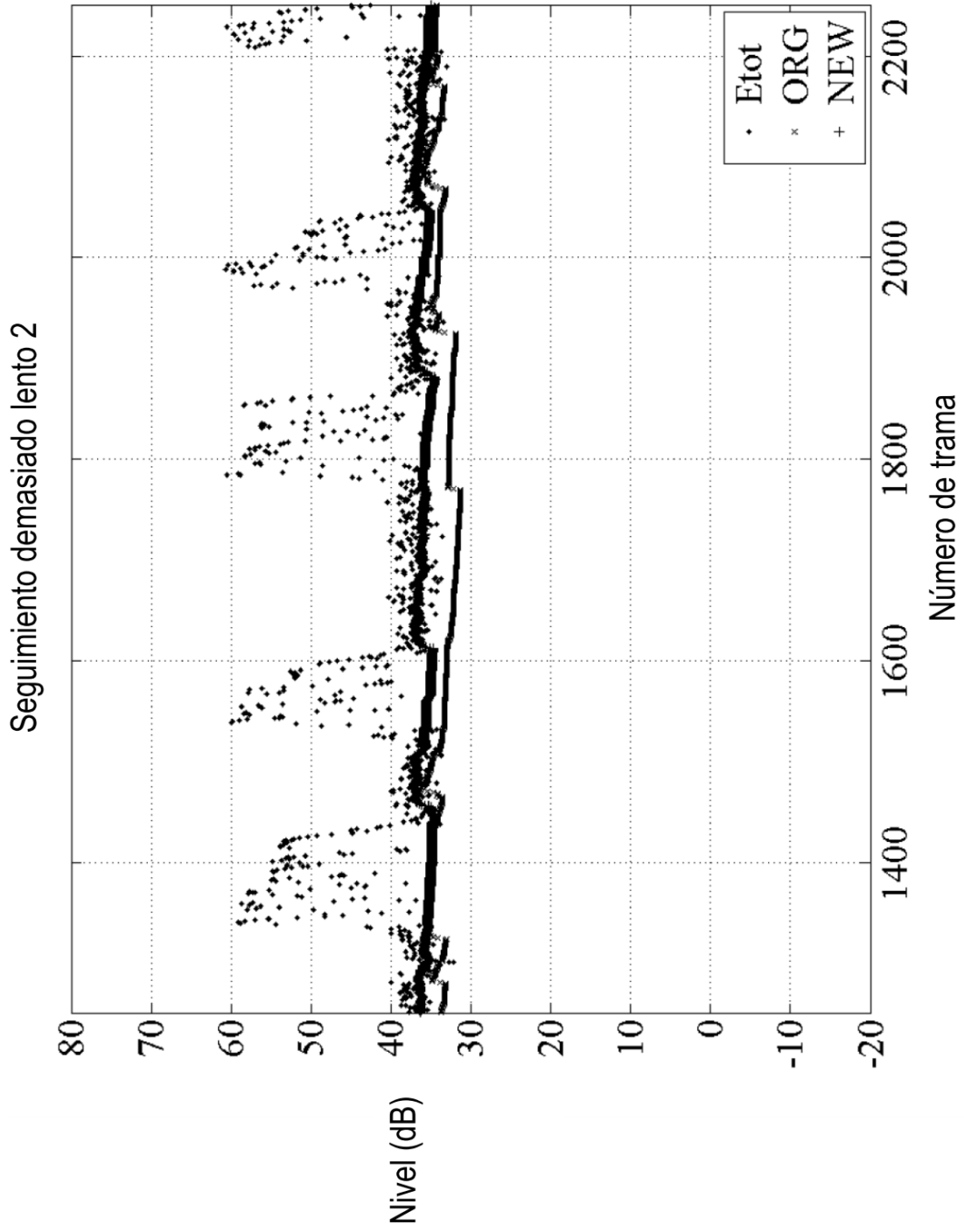


Figura A7

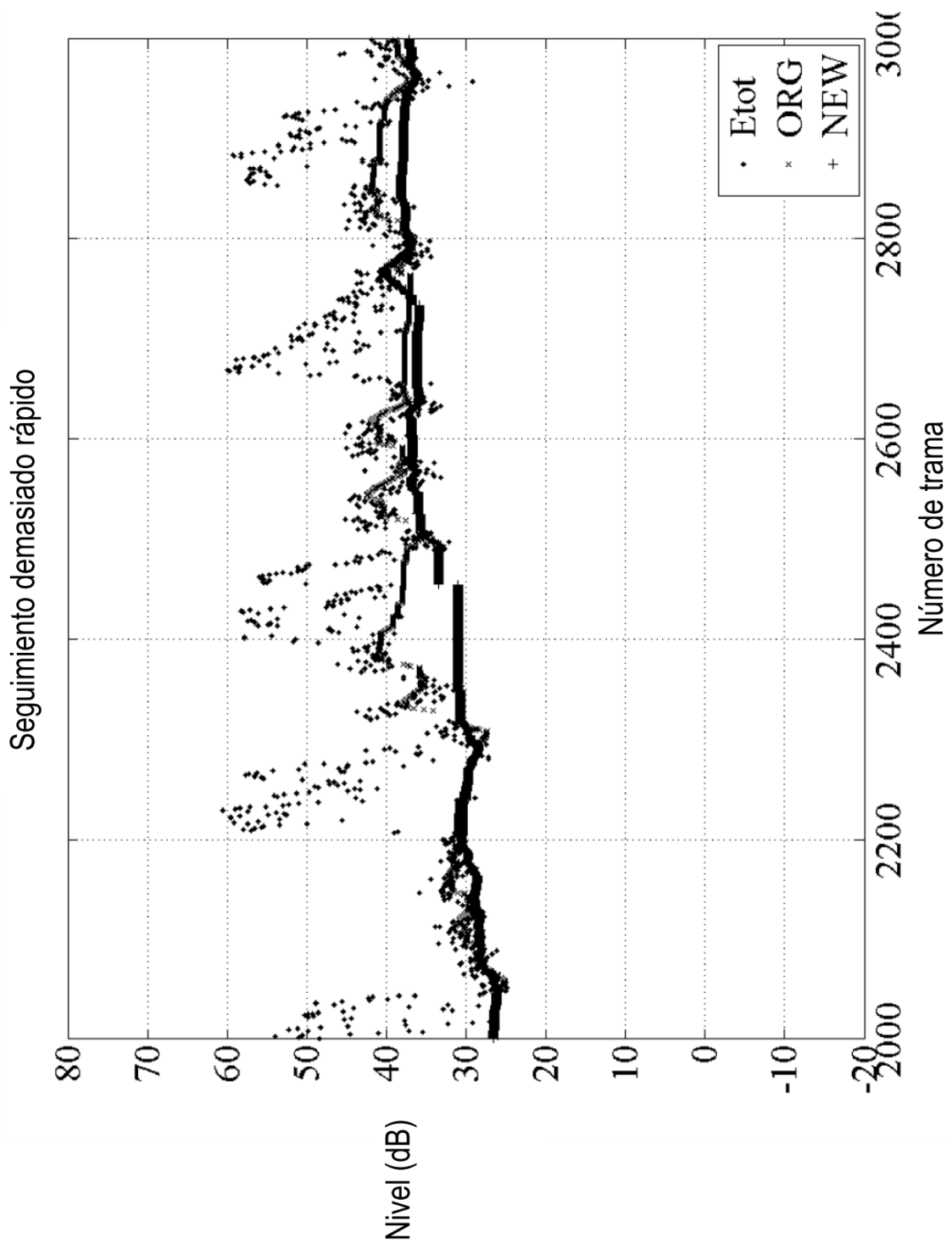


Figura A8

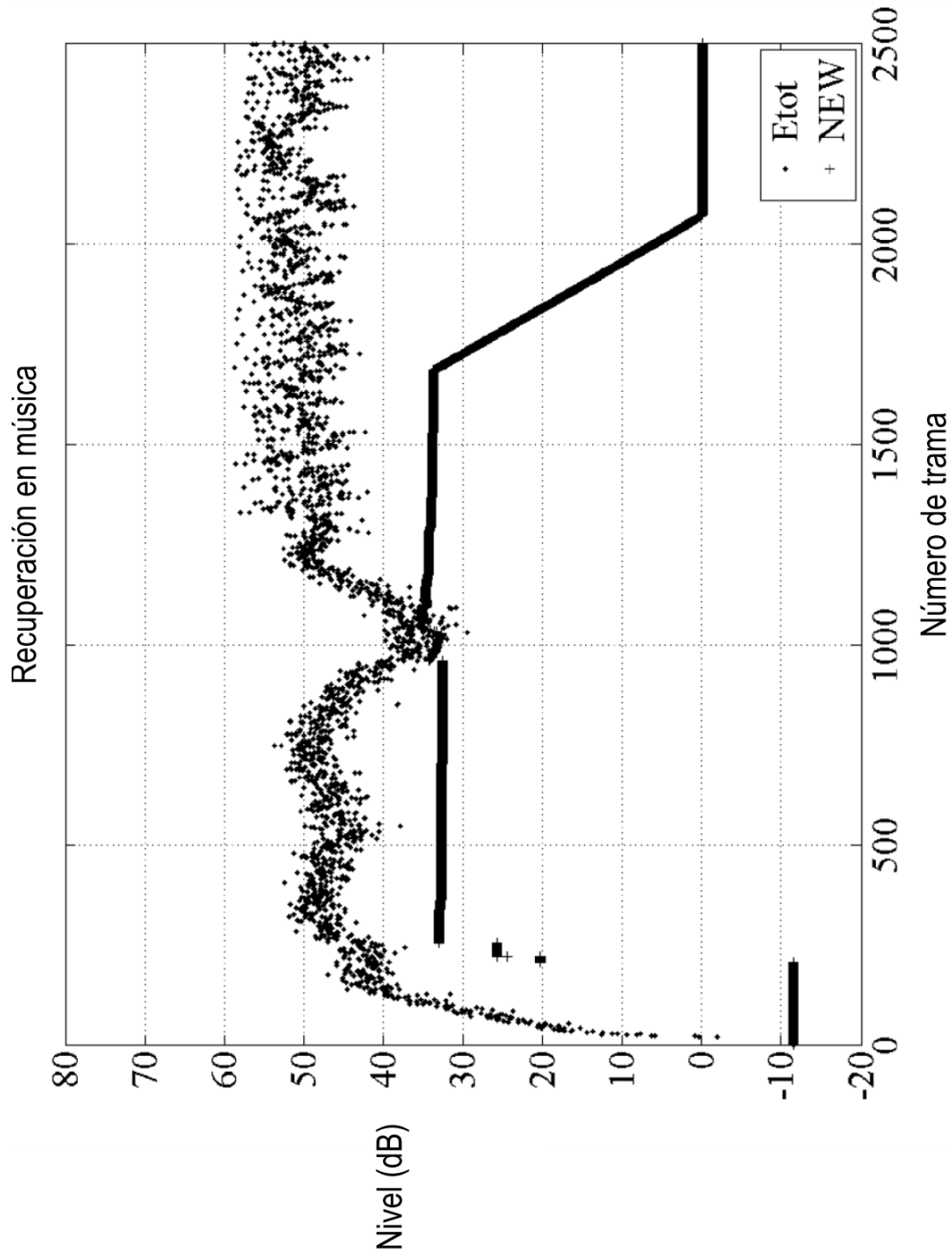


Figura A9