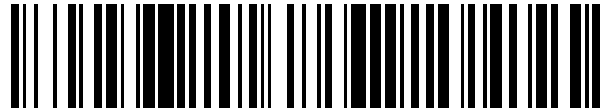


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 664 348**

51 Int. Cl.:

G10L 25/78 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.07.2015 PCT/SE2015/050770**

87 Fecha y número de publicación internacional: **04.02.2016 WO16018186**

96 Fecha de presentación y número de la solicitud europea: **01.07.2015 E 15739357 (0)**

97 Fecha y número de publicación de la concesión europea: **27.12.2017 EP 3175458**

54 Título: **Estimación de ruido de fondo en 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:
19.04.2018

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 2 664 348 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 de ruido de fondo en señales de audio

5 **Campo técnico**

Las realizaciones de la presente invención se refieren al procesamiento de señales de audio, y en particular a la estimación de ruido de fondo, por ejemplo para apoyar 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 la no reducción de la calidad. En tales sistemas, se usa un detector de actividad para indicar señales activas, por ejemplo voz o música, que se codificarán activamente, y segmentos con señales de fondo que pueden reemplazarse por 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 percibirá como degradación subjetiva de la calidad cuando el segmento activo recortado se reemplaza por ruido de confort. Al mismo tiempo, la eficiencia del 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 introducir un modo 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 de descripción de un detector de actividad de sonido generalizado, SAD o 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, segmentos de señal de audio de por ejemplo 5-30 ms, dependiendo de la implementación, y una decisión de actividad por trama se produce como salida.

Una decisión primaria, "prim", es realizada por el detector primario ilustrado en la figura 1. La decisión primaria es básicamente una comparación de las características de una trama actual con las características de fondo, que se estiman a partir de las tramas de entrada anteriores. Una diferencia entre las características de la trama actual y las características de fondo que es mayor que un umbral provoca una decisión primaria activa. El bloque de adición de bloqueo se usa para extender la decisión principal basándose en decisiones primarias anteriores para formar la decisión final, "indicador". La razón para usar el bloqueo es principalmente reducir/eliminar el riesgo de recortes a la mitad y al final de la ráfaga de actividad. Como se indica en la figura, un controlador de operación puede ajustar el umbral o umbrales para el detector primario y la duración de la adición de bloqueo de acuerdo con las características de la señal de entrada. El bloque de estimación de fondo se usa para estimar el ruido de fondo en la señal de entrada. El ruido de fondo también se puede denominar como "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 mediante la decisión primaria, es decir, con la retroalimentación de decisión o decisión métrica, que se indica mediante la línea de rayas y puntos en la figura 1, o mediante el uso de algunas otras características del señal de entrada, es decir, sin respuesta de decisión. También es posible usar combinaciones de las dos estrategias.

Un ejemplo de un códec que utiliza retroalimentación de decisión para la estimación de antecedentes es AMR-NB (banda estrecha de velocidad múltiple adaptable) y ejemplos de códecs en los que no se usa retroalimentación de decisión son EVRC (CODEC de velocidad variable mejorada) y G.718.

Hay varias características de señales diferentes que se pueden usar, pero una característica común utilizado en los VAD son las características de frecuencia de la señal de entrada. Un tipo comúnmente usado de características de frecuencia es la energía de trama de subbanda, debido a su baja complejidad y funcionamiento fiable en baja SNR. Por lo tanto, se supone que la señal de entrada se divide en diferentes subbandas de frecuencia y el nivel de fondo se estima para cada una de las subbandas. De esta forma, una de las características de ruido de fondo es el vector con los valores de energía para cada subbanda. Estos son valores que caracterizan el ruido de fondo en la señal de entrada en el dominio de la frecuencia.

Para lograr el seguimiento del ruido de fondo, la actualización de la estimación de ruido de fondo real se puede realizar al menos de tres formas diferentes. Una forma es usar un proceso autoregresivo, AR, por intervalo de frecuencia para manejar la actualización. Ejemplos de tales códecs son AMR-NB y 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 forma 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 se aumenta cada trama hasta que sea más alta que la entrada actual. En esa situación, la entrada actual se usa como estimación. EVRC es un ejemplo de códec que utiliza esta técnica para actualizar la estimación de fondo para la función VAD. Téngase en cuenta que EVRC usa diferentes estimaciones de fondo para VAD y supresión de ruido. Cabe señalar que un VAD se puede utilizar en otros contextos que DTX. Por ejemplo, en códecs de velocidad variable, tales como EVRC, el VAD se puede usar como parte de una función de

determinación de velocidad.

Una tercera forma es usar una técnica llamada mínima en la que la estimación es el valor mínimo durante una ventana de tiempo deslizante de tramas anteriores. Esto básicamente da una estimación mínima que se escala, usando un factor de compensación, para obtener una estimación media aproximada para el ruido estacionario.

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

El rendimiento del VAD depende de la capacidad del estimador de ruido de fondo para rastrear 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.

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

M. Jelinek y R. Salami "Método de reducción de ruido para codificación de discurso de banda ancha" 2004 duodécima conferencia europea de procesamiento de señal, p. 1959-1962, enseñan un método para la estimación de ruido de fondo donde la presencia de pausas, durante las cuales se estima dicho ruido, se determina basándose en un cociente entre un residuo de una predicción lineal de 2º orden y un residuo de una predicción lineal de 16º orden.

Sumario

Sería deseable lograr una estimación mejorada del ruido de fondo en las señales de audio. "Mejorada" puede implicar aquí tomar una decisión más correcta con respecto a si una señal de audio comprende voz o música activa o no, y por lo tanto estimando más a menudo, por ejemplo actualizando una estimación previa, el ruido de fondo en segmentos de señal de audio estando realmente libre de contenido activo, como voz y/o música. En el presente documento, es provisto 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.

Para la estimación de ruido de fondo en señales de audio, es importante poder encontrar características 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

El inventor se ha dado cuenta de que las características relacionadas con las energías residuales para diferentes órdenes de modelos de predicción lineal pueden utilizarse para detectar pausas en las señales de audio. Estas energías residuales se pueden extraer por ejemplo a partir 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 de ruido. La solución descrita en el presente documento es particularmente eficiente para las condiciones en que una SNR está en el rango de 10 a 20 dB.

Otra característica proporcionada en el presente documento es una medida de la cercanía espectral al fondo, que puede realizarse por ejemplo utilizando las energías de subbanda de dominio de la frecuencia que se usan, por ejemplo en una subbanda de SAD. La medida de cercanía espectral también se puede usar para tomar una decisión de si una señal de audio comprende una pausa o no.

De acuerdo con un primer aspecto, es provisto 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, basándose en una primera ganancia de predicción lineal, calculada como un cociente entre una señal residual de una predicción lineal de orden 0 y una señal residual de una predicción lineal de 2º orden para el segmento de señal de audio; y, una segunda ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de 2º orden y una señal residual de una predicción lineal de 16º orden para el segmento de señal de audio. El método comprende además determinar si el segmento de señal de audio comprende una pausa basándose al menos en al menos dicho parámetro obtenido; y, actualizar una estimación de ruido de fondo basándose en el segmento de señal de audio cuando el segmento de señal de audio comprende una pausa.

De acuerdo con un segundo aspecto, es provisto un estimador de ruido de fondo. El estimador de ruido de fondo está configurado para obtener al menos un parámetro asociado con un segmento de señal de audio basándose en una primera ganancia de predicción lineal, calculada como un cociente entre una señal residual de una predicción lineal de orden 0 y una señal residual de una predicción lineal de 2º orden para el segmento de señal de audio; y, una segunda ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de 2º orden y una señal residual de una predicción lineal de 16º orden para el segmento de señal de audio. El estimador de ruido de fondo está configurado además para determinar si el segmento de señal de audio comprende una pausa basándose al menos en al menos dicho parámetro obtenido; y, para actualizar una estimación de ruido de fondo basándose en el segmento de señal de audio cuando el segmento de señal de audio comprende una pausa.

De acuerdo con un tercer aspecto, es provisto un SAD, que comprende un estimador de ruido de fondo de acuerdo con el segundo aspecto.

De acuerdo con un cuarto aspecto, es provisto un códec, que comprende un estimador de ruido de fondo de acuerdo con el segundo aspecto.

De acuerdo con un quinto aspecto, es provisto un dispositivo de comunicación, que comprende un estimador de ruido de fondo de acuerdo con el segundo aspecto.

De acuerdo con un sexto aspecto, es provisto un nodo de red, que comprende un estimador de ruido de fondo de acuerdo con el segundo aspecto.

De acuerdo con un séptimo aspecto, es provisto un programa informático, que comprende instrucciones que, cuando se ejecutan en al menos un procesador, hacen que al menos dicho procesador lleve a cabo el método de acuerdo con el primer aspecto.

De acuerdo con un octavo aspecto, es provista una portadora, que contiene un programa informático de acuerdo con el séptimo aspecto.

Breve descripción de los dibujos

Los anteriores y otros objetos, características y ventajas de la tecnología divulgada en el presente documento serán evidentes a partir de la siguiente descripción más particular de las realizaciones tal como se ilustra en los dibujos que se acompañan. Los dibujos no son necesariamente a escala, sino que se hace hincapié en ilustrar los principios de la tecnología divulgada en el presente documento.

La figura 1 es un diagrama de bloques que ilustra un detector de actividad y bloqueo lógico de determinación.

La figura 2 es un diagrama de flujo que ilustra un método para la estimación de ruido de fondo, de acuerdo con una realización ejemplar.

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

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

La figura 5 es un diagrama de bloques que ilustra el cálculo de las 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.

La figura 7 es un diagrama de flujo que ilustra una lógica de decisión de actualización de fondo a partir 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.

Las figuras 11a-11c y 12-13 son diagramas de bloques que ilustran diferentes implementaciones 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.

Descripción detallada

La solución divulgada en el presente documento se refiere a la estimación de ruido de fondo en señales de audio. En el detector de actividad generalizada ilustrado en la figura 1, la función de estimar el ruido de fondo se realiza mediante el bloque denominado "estimador de fondo". Algunas realizaciones de la solución descrita en el presente documento se pueden ver en relación con las soluciones divulgadas previamente en los documentos WO 2011/049514, WO 2011/049515, y también en el Anexo A (Apéndice A). La solución divulgada en el presente documento se comparará con implementaciones de estas soluciones divulgadas previamente. Aunque las soluciones divulgadas en los documentos WO 2011/049514, WO 2011/049515 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 rastrear 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.

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

Idealmente, se necesitaría una función inversa de un detector de actividad, o lo que se llamaría "detector de ocurrencia de pausa", para controlar la estimación de ruido. Esto garantizaría que la actualización de las características de ruido de fondo se haga solo cuando no haya 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 o no una señal activa.

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 se usa a menudo 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 debe detectar música como una señal activa.

El estimador de fondo ilustrado en la figura 1 utiliza la retroalimentación desde el detector primario y/o el bloque de bloqueo para localizar segmentos de señal de audio inactivos. Al desarrollar la tecnología descrita en el presente documento, ha sido un deseo de eliminar, o al menos reducir la dependencia de tal retroalimentación. Para la estimación de antecedentes divulgada en el presente documento, el inventor ha identificado como importante poder encontrar características fiables para identificar las características de las señales de fondo cuando solo está disponible una señal de entrada con una mezcla desconocida de señal activa y de fondo. El inventor se ha dado cuenta además de que no se puede suponer que la señal de entrada comienza con un segmento de ruido, o incluso que la señal de entrada se mezcla con ruido, ya que puede ser que la señal activa sea música.

Un aspecto es que, aunque 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 que no sea deseable 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 usar para evitar actualizaciones en estos casos.

Además, durante la inicialización es deseable permitir que la estimación de ruido comience tan pronto como sea posible, evitando decisiones erróneas, ya que esto podría resultar en un recorte del SAD si la actualización de ruido de fondo se realiza usando 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 funciona bien en situaciones de SNR difíciles. La solución se describirá a continuación con referencia a las figuras 2-5.

En el campo de la codificación de voz, es común usar la llamada predicción lineal para analizar la forma espectral de una señal de entrada. El análisis generalmente se realiza dos veces por trama, y para una precisión temporal mejorada, los resultados son entonces interpolados de modo que haya 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 de tiempo discreto 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, por lo tanto, 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 de 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 menor energía que la señal de entrada cuando el filtro tiene éxito en el modelado de algún aspecto o aspectos de la señal de entrada. Esta señal de salida se denomina "residual", "energía residual" o "señal residual". Dichos filtros de predicción lineal, alternativamente denominados filtros residuales, pueden ser de diferente orden de modelo con diferente número de coeficientes de filtro. Por ejemplo, para modelar correctamente la voz, puede requerirse un filtro de predicción lineal de orden de modelo 16. Por lo tanto, en un codificador de voz, se puede usar un filtro $A(z)$ de predicción lineal de orden de modelo 16.

El inventor se ha dado cuenta de que las características relacionadas con la predicción lineal pueden usarse para detectar pausas en señales de audio en un rango SNR de 20 dB hasta 10 dB o posiblemente 5 dB. De acuerdo con las realizaciones de la solución descrita en el presente documento, se utiliza una relación entre las energías residuales para diferentes órdenes de modelo para una señal de audio 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 la cantidad de energía de señal que el filtro de predicción lineal ha sido capaz de modelar, o eliminar, entre un modelo de orden y otro modelo de orden.

La energía residual dependerá del orden M de 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 creación de un filtro $A(z)$ de predicción de orden M también, como un "subproducto", produce las energías residuales de los órdenes de modelos inferiores. Este hecho se puede utilizar de acuerdo con las realizaciones de la invención.

La figura 2 muestra un método general ejemplar para la estimación de 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, tal como una trama o parte de una trama, basándose en una primera ganancia de predicción lineal, calculada como un cociente entre una señal residual de una predicción lineal de orden 0 y una señal residual de una predicción lineal de 2º orden para el segmento de señal de audio; y, una segunda ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de 2º orden y una señal residual de una predicción lineal de 16º orden para el segmento de señal de audio.

El método comprende además determinar 202 si el segmento de señal de audio comprende una pausa, es decir, está libre de contenido activo tal como voz y música, basándose al menos en al menos dicho parámetro obtenido; y actualizar 203 una estimación de ruido de fondo basándose en el segmento de señal de audio cuando el segmento de señal de audio comprende una pausa. Es decir, el método comprende la actualización de una estimación de ruido de fondo cuando se detecta una pausa en el segmento de señal de audio basándose al menos en al menos dicho parámetro obtenido.

Las ganancias de predicción lineal podrían describirse como una primera ganancia de predicción lineal relacionada con ir desde la predicción lineal de orden 0 a la predicción lineal de 2º orden para el segmento de señal de audio; y una segunda ganancia de predicción lineal relacionada con ir de predicción lineal de 2º orden a decimosexto para el segmento de señal de audio. Además, la obtención de al menos dicho parámetro podría describirse alternativamente como determinar, calcular, derivar o crear. Las energías residuales relacionadas con las predicciones lineales del orden de modelo 0, 2 y 16 pueden obtenerse, recibirse o recuperarse, es decir, proporcionarse de alguna manera, de una parte del codificador donde la predicción lineal se realiza como parte de un proceso de codificación regular. De este modo, la complejidad computacional de la solución descrita en el presente documento puede reducirse, en comparación con cuando las energías residuales necesitan derivarse especialmente para la estimación de ruido de fondo.

Al menos dicho parámetro obtenido basándose en las características de predicción lineal puede proporcionar un análisis independiente de nivel de la señal de entrada que mejora la decisión de realizar una actualización de ruido de fondo o no. La solución es particularmente útil en el rango de SNR de 10 a 20 dB, 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 $E(0)$, ..., $E(m)$, ..., $E(M)$ representan las energías residuales

para los órdenes de modelo de 0 a M de los filtros $M + 1$ $A_m(z)$. Téngase en cuenta que $E(0)$ es solo 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 mediante el análisis de la ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de orden 0 y una señal residual de una predicción lineal de 2º orden, y la ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de 2º orden y una señal residual de una predicción lineal de 16º orden. Es decir, la ganancia de predicción lineal para pasar de la predicción lineal de orden 0 a la de 2º orden es lo mismo que la "energía residual" $E(0)$ (para un orden de modelo 0) dividida por la energía residual $E(2)$ (para un 2º orden de modelo). En consecuencia, la ganancia de predicción lineal para pasar de la predicción lineal de 2º orden a la predicción lineal de orden 16 es la misma que la energía residual $E(2)$ (para un 2º orden de modelo) dividida por la energía residual $E(16)$ (para un 16º orden de modelo). Los ejemplos de parámetros y la determinación de parámetros basados en las ganancias de predicción se describirán con más detalle más adelante. Al menos dicho parámetro obtenido de acuerdo con la realización general descrita anteriormente puede formar parte de un criterio de decisión usado para evaluar si actualizar o no la estimación de ruido de fondo.

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

La obtención de al menos un parámetro puede comprender además la creación de al menos una estimación a largo plazo de cada una de las ganancias de predicción lineal primera y segunda, por ejemplo mediante el filtrado de paso bajo. Al menos tal estimación a largo plazo se basaría entonces en las ganancias de predicción lineal correspondientes asociadas con al menos un segmento de señal de audio precedente. Se podría crear más de una estimación a largo plazo, donde, por ejemplo, una primera y una segunda estimación a largo plazo 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. Tal primera estimación a largo plazo puede alternativamente indicarse como una estimación a corto plazo.

La obtención de al menos un parámetro puede comprender además determinar una diferencia, tal como la diferencia absoluta Gd_0_2 (Ec.3) descrita a continuación, 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. Alternativamente o además, se podría determinar una diferencia entre dos estimaciones a largo plazo, como en la ecuación 9 a continuación. El término determinante podría alternativamente intercambiarse para calcular, crear o derivar.

La obtención de al menos un parámetro puede, como se indica anteriormente, comprender el filtrado de paso bajo de las ganancias de predicción lineal, derivando así estimaciones a largo plazo, algunas de las cuales pueden denominarse estimaciones a corto plazo, dependiendo de cuántos segmentos se tengan en cuenta 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 solo, con el segmento de señal de audio actual y una media, denominada por ejemplo media a largo plazo, o estimación a largo plazo, de una ganancia de predicción correspondiente obtenida basándose en una pluralidad de segmentos de señal de audio precedentes. Esto se puede realizar para crear, por ejemplo además, las estimaciones a largo plazo de las ganancias de predicción. El filtrado de paso bajo puede realizarse en dos o más pasos, donde cada paso puede dar como resultado un parámetro, o estimación, que se usa para tomar una decisión con respecto a la presencia de una pausa en el segmento de señal de audio. Por ejemplo, diferentes estimaciones a largo plazo (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 a continuación) que reflejan cambios en la señal de audio de diferentes maneras, pueden analizarse o compararse para detectar una pausa en un segmento de señal de audio actual.

La determinación 202 de si el segmento de señal de audio comprende una pausa o no puede además basarse en una medida de cercanía espectral asociada con el segmento de señal de audio. La medida de cercanía espectral indicará qué tan cerca está el nivel de energía "por banda de frecuencia" del segmento de señal de audio actualmente procesado al nivel de energía "por banda de frecuencia" de la estimación de ruido de fondo actual, por ejemplo un valor inicial o una estimación que es el resultado de una actualización previa realizada antes del análisis del segmento de señal de audio actual. Un ejemplo de determinación o derivación de una medida de cercanía espectral se da a continuación en las ecuaciones Ec.12 y Ec.13. La medida de cercanía espectral se puede usar para evitar 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 media 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 está distribuida 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 trama, puede ser contenido activo de bajo nivel y una actualización de la estimación de ruido de

fondo basada en la trama podría por ejemplo prevenir la detección de futuros tramas con contenido similar. Como la SNR de subbanda es más sensible a los aumentos de energía que usan incluso contenido activo de bajo nivel, puede resultar en una gran actualización de la estimación de fondo si ese rango de frecuencia particular no existe en el ruido de fondo, como la parte de alta frecuencia de voz comparada con ruido de automóvil de baja frecuencia.
 5 Después de dicha actualización, será más difícil detectar la voz.

Como ya se sugirió anteriormente, la medida de cercanía espectral puede derivarse, obtenerse o calcularse basándose en las energías para un conjunto de bandas de frecuencias, subbandas denominadas alternativamente, del segmento de señal de audio analizado actualmente y las estimaciones de ruido de fondo actuales
 10 correspondientes al conjunto de bandas de frecuencia. Esto también se ejemplificará y describirá con más detalle más adelante, y se ilustra en la figura 5.

Como se indicó anteriormente, la medida de cercanía espectral puede derivarse, obtenerse o calcularse comparando una corriente por nivel de energía de banda de frecuencia del segmento de señal de audio procesado
 15 actualmente con un nivel de energía de banda por frecuencia de una estimación de ruido de fondo actual. Sin embargo, para comenzar, 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 ya que todavía no se habrá realizado una actualización fiable de una estimación de 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
 20 niveles de energía por banda de frecuencia del segmento de señal de audio actual se compararán en cambio con una estimación de fondo inicial, que puede ser, por ejemplo un valor constante configurable. En los ejemplos que se muestran a continuación, esta estimación inicial de ruido de fondo se establece en el valor ejemplar $E_{\min} = 0,0035$. Después del período de inicialización, el procedimiento puede cambiar a funcionamiento normal y comparar el nivel de energía actual por banda de frecuencia del segmento de señal de audio actualmente procesado con un nivel de
 25 energía de banda por frecuencia de una estimación de ruido de fondo actual. La duración del período de inicialización puede configurarse, por ejemplo basándose en simulaciones o pruebas que indiquen el tiempo que transcurre antes de que, sea provista una estimación, por ejemplo fiable y/o satisfactoria, del ruido de fondo. Un ejemplo usado a continuación, la comparación con una estimación de ruido de fondo inicial (en lugar de con una estimación "real" derivada basada en la señal de audio actual) se realiza durante las primeras 150 tramas.

Al menos dicho parámetro puede ser el parámetro ejemplificado en el código más abajo, denominado NEW_POS_BG, y/o uno o más de la pluralidad de parámetros descritos más adelante, que conducen a la formación de un criterio de decisión o un componente en un criterio de decisión para la detección de pausa. En otras palabras,
 30 al menos dicho parámetro, o característica, obtenido 201 basándose en 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 basarse en uno o más de los parámetros descritos a continuación.

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

40 La figura 3 muestra un diagrama de bloques general de la derivación de características o parámetros relacionados con E(0) y E(2), de acuerdo con una realización ejemplar. 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

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

45 donde E(0) representa la energía de la señal de entrada y E(2) es la energía residual después de una predicción lineal de 2º orden. La expresión en la ecuación 1 limita la ganancia de predicción a un intervalo entre 0 y 8. La ganancia de predicción para casos normales debería ser mayor que cero, pero pueden ocurrir anomalías, por ejemplo para valores cercanos a cero, y por lo tanto una limitación "mayor que cero" ($0 <$) puede ser útil. La razón para limitar la ganancia de predicción a un máximo de 8 es que, para el propósito de la solución descrita en el
 50 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 hay 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 más alto no tiene más éxito en el modelado de la señal de audio que el filtro de un orden de modelo inferior. Además, si la ganancia de predicción G_{0_2} tomara valores demasiado grandes en las
 55 siguientes expresiones, puede poner en riesgo la estabilidad de los parámetros derivados. Debe observarse que 8 es solo un valor ejemplar, que se ha seleccionado para una realización específica. El parámetro G_{0_2} podría alternativamente denominarse, por ejemplo epsP_{0_2} , o $g_{LP_{0_2}}$.

La ganancia de predicción limitada se filtra en dos pasos para crear estimaciones a largo plazo de esta ganancia. El
 60 primer filtro de paso bajo y, por lo tanto, la derivación de una primera característica o parámetro a largo plazo se realiza como:

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

Donde el segundo "G1_0_2" en la expresión debería leerse como el valor de un segmento de señal de audio precedente. Este parámetro generalmente será 0 u 8, dependiendo del tipo de ruido de fondo en la entrada una vez que haya un segmento de entrada de solo fondo. El parámetro G1_0_2 podría alternativamente denominarse, por ejemplo epsP_0_2_lp o g_{LP_0_2}. Se puede crear o calcular otra característica o parámetro usando la diferencia entre la primera característica de largo plazo G1_0_2 y la ganancia de predicción limitada de trama por trama G_0_2, de acuerdo con:

$$Gd_0_2 = \text{abs}(G1_0_2 - G_0_2) \quad (\text{Ec. 3})$$

Esto 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 podría alternativamente denominarse, por ejemplo epsP_0_2_ad o g_{ad_0_2}. En la figura 4, esta diferencia se usa para crear una segunda estimación o característica a largo plazo 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 media estimada actualmente de acuerdo con:

$$Gad_0_2 = (1-a) \overline{Gad_0_2} + a Gd_0_2 \quad (\text{Ec. 4})$$

donde, si Gd_0_2 < $\overline{Gad_0_2}$ entonces a = 0,1 si no a = 0,2

Donde el segundo "Gad_0_2" en la expresión debería leerse como el valor de un segmento de señal de audio precedente.

El parámetro Gad_0_2 podría alternativamente denominarse, por ejemplo Glp_0_2, epsP_0_2_ad_lp o $\overline{Gad_0_2}$. Para evitar que el filtrado enmascare las diferencias ocasionales de trama alta, se puede derivar otro parámetro, que no se muestra en la figura. Es decir, la segunda característica a largo plazo Gad_0_2 se puede combinar con la diferencia de trama para evitar dicha máscara. Este parámetro puede derivarse tomando el máximo de la versión de trama Gd_0_2 y la versión a largo plazo Gad_0_2 de la característica de ganancia de predicción, como:

$$Gmax_0_2 = \text{max}(Gad_0_2, Gd_0_2) \quad (\text{Ec. 5})$$

El parámetro Gmax_0_2 podría alternativamente denominarse, por ejemplo epsP_0_2_ad_lp_max o g_{max_0_2}.

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

La figura 4 muestra un diagrama de bloques de general de la derivación de características o parámetros relacionados con E(2) y E(16), de acuerdo con una realización ejemplar. 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 usando la diferencia o relación entre la energía residual de 2º orden y la energía residual de 16º orden se derivan ligeramente de forma diferente de los descritos anteriormente relacionados con la relación entre las energías residuales de orden 0 y 2º.

Aquí, también, una ganancia de predicción limitada se calcula como

$$G_2_16 = \text{max}(0, \text{min}(8, E(2)/E(16))) \quad (\text{Ec. 6})$$

donde E(2) representa la energía residual después de una predicción lineal de 2º orden y E(16) representa la energía residual después de una predicción lineal de 16º orden. El parámetro G_2_16 podría alternativamente denominarse por ejemplo epsP_2_16 o g_{LP_2_16}. Esta ganancia de predicción limitada se usa para crear dos estimaciones a largo plazo de esta ganancia: una donde el coeficiente del filtro difiere de si la estimación a largo plazo se aumenta o no como se muestra en:

$$G1_2_16 = (1-a) \overline{G1_2_16} + a G_2_16 \quad (\text{Ec. 7})$$

donde si G_2_16 > $\overline{G1_2_16}$ entonces a = 0,2 si no a = 0,03

El parámetro G1_2_16 podría alternativamente denominarse por ejemplo epsP_2_16_lp o $\overline{G1_2_16}$.

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

$$G2_2_16 = (1-b) \overline{G2_2_16} + b G_2_16, \text{ donde } b = 0,02 \quad (\text{Ec. 8})$$

El parámetro G2_2_16 podría alternativamente denominarse por ejemplo epsP_2_16_lp2 o $\overline{G2_2_16}$.

Para la mayoría de los tipos de señales de fondo, tanto G1_2_16 como G2_2_16 estarán cerca de 0, pero tendrán diferentes respuestas al contenido donde se necesita la predicción lineal de 16º orden, que típicamente es para voz y otro contenido activo. La primera estimación a largo plazo, G1_2_16 generalmente será 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 denominarse como epsP_2_16_dlp o g_{ad_2_16}.

10 Gd_2_16 se puede usar 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 si no c = 0,05

15 Este filtro aplica diferentes coeficientes de filtro dependiendo de si la tercera señal de largo plazo debe ser aumentada o no. El parámetro Gad_2_16 puede alternativamente denominarse por ejemplo epsP_2_16_dlp_lp2 o g_{ad_2_16}. Además, la señal a largo plazo Gad_2_16 se puede combinar con la señal de entrada del filtro Gd_2_16 para evitar que el filtrado enmascare las altas entradas 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

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

El parámetro Gmax_2_16 podría alternativamente denominarse, por ejemplo epsP_2_16_dlp_max o g_{max_0_2}.

25 Cercanía espectral/medida de diferencia

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

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

40 Entonces, durante la inicialización, nonstaB se calcula usando un Emin, que aquí se establece en Emin = 0,0035 como:

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

donde la suma se hace sobre i = 2 ... 16.

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

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

50 donde la suma se hace sobre i = 2 ... 16.

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

55 En la figura 6 se muestra un diagrama de bloques que ilustra una realización ejemplar de un estimador de fondo. La realización en la figura 6 comprende un bloque para el entramado 601 de entrada, que divide la señal de audio de

entrada en tramas o segmentos de longitud adecuada, por ejemplo 5-30 ms. La realización comprende además un bloque para la extracción 602 de características que calcula las características, también denominadas parámetros en el presente documento, para cada trama o segmento de la señal de entrada. La realización comprende además un bloque para la lógica 603 de decisión de actualización, para determinar si una estimación de fondo puede o no actualizarse basándose en 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 además un actualizador 604 de fondo, para actualizar la estimación de ruido de fondo cuando la lógica de decisión de actualización indica que es adecuado hacerlo. En la realización ilustrada, puede derivarse una estimación de ruido de fondo por subbanda, es decir, para un número de bandas de frecuencia.

La solución descrita en el presente documento se puede usar para mejorar una solución previa para la estimación de ruido de fondo, descrita en el Anexo A en el presente documento, y también en el documento WO 2011/049514. A continuación, la solución descrita en el presente documento se describirá en el contexto de esta solución descrita anteriormente. Se darán ejemplos de código desde una implementación de códigos de una realización de un estimador de ruido de fondo.

A continuación, se describen detalles de implementación reales para una realización de la invención en un codificador basado en G.718. Esta implementación usa muchas de las características de energía descritas en la solución en el Anexo A y el documento WO 2011/049514 incorporado en el presente documento como referencia. Para más detalles que los presentados a continuación, nos referimos al Anexo A y al documento WO 2011/049514.

Las siguientes características de energía se definen en el documento WO 2011/049514:

```

Etot;
Etot_l_lp;
Etot_v_h;
totalNoise;
sign_dyn_lp;

```

Las siguientes características de correlación se definen en el documento WO 2011/049514:

```

aEn;
harm_cor_cnt
act_pred
cor_est

```

Las siguientes características se definieron en la solución que figura en el Anexo A:

```

Etot_v_h;
It_cor_est = 0.01f* cor_est + 0.99f* lt_cor_est;
It_tn_track = 0.03f* (Etot - totalNoise <10) + 0.97f*lt_tn_track;
It_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

```

La lógica de actualización de ruido de la solución que figura en el Anexo A se muestra en la figura 7. Las mejoras, relacionadas 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 las características; la parte 702, donde las decisiones de pausa se basan en diferentes parámetros; y además a la parte 703, donde se toman diferentes acciones basadas en si se detecta o no una pausa. Además, las mejoras pueden tener un efecto sobre la actualización 704 de la estimación de ruido de fondo, que podrían por ejemplo actualizarse cuando se detecte una pausa basada en las nuevas características, que no se habrían detectado antes de introducir la solución descrita en este documento. En la implementación ejemplar descrita aquí, las nuevas características introducidas en el presente documento se calculan de la siguiente manera, comenzando con non_staB, que se determina usando las energías de subbanda de la trama actual enr [i], que corresponde a Ecb (i) anterior y en la figura 6, y la estimación de ruido de fondo actual bckr [i], que corresponde a Ncb (i) anterior y en la figura 6. La primera parte de la primera sección del código siguiente está relacionada con un procedimiento inicial especial para las primeras 150 tramas de una señal de audio, antes de que se haya derivado una estimación de fondo adecuada.

```

/*calcular el fondo relativo de la característica de no estacionariedad (característica de cercanía espectral non_staB*/
if (ini_frame <150)
{
/* Durante init no incluye actualizaciones*/
if (i >= 2 && i <= 16)
{

```

```

non_staB += (float) fabs (log (en[i]) + 1.0f) -
log (E_MIN + 1.0f));
}
}
5 else
{
/*Después de la comparación de init con la estimación de fondo*/
if (i>= 2 && i <= 16)
{
10 non_staB += (float) fabs (log (enr[i]) + 1.0f) -
log (bckr [i] + 1.0f));
}
}
if (non_staB>= 128)
15 {
non_staB = 32767.0/256.0f;
}

```

20 Las siguientes secciones de códigos muestran cómo se calculan las nuevas características para las energías residuales de predicción lineal, es decir, para la ganancia de predicción lineal. Aquí las energías residuales se llaman epsP [m] (cf. E(m) usado previamente).

```

/*-----*
* Eficacia de predicción lineal de orden de 0 a 2
25 * (ganancia de predicción lineal que va del modelo de orden 0 a 2° del filtro de
predicción lineal)
*-----*/
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;
30 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
35 {
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);
/*-----*
* Eficacia de predicción lineal de orden de 2 a 16
40 * (ganancia de predicción lineal que va del modelo de orden 2° a 16° del filtro de
predicción lineal)
*-----*/
epsP_2_16 = max (0, min (8, epsP[2]/epsP[16]));
if (epsP_2_16 > epsP_2_16_lp)
45 {
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;
50 }
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)
{
55 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);}
60
El siguiente código ilustra la creación de métricas, umbrales e indicadores combinados utilizados para la decisión de
actualización real, es decir, la determinación de si se debe actualizar o no la estimación de 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.
65 comb_ahc_epsP = max (max (act_pred, lt_haco_ev), epsP_2_16_dlp);

```

ES 2 664 348 T3

```

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;
5
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;
10 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));

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

PD_1 = (epsP_2_16_dlp_max < 0.1 0f);
PD_2 = (epsP_0_2_ad_lp_max < 0.1 0f);
PD_3 = (comb_ahc_epsP < 0.85f);
20 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);

25 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;

/* El detector de silencio original funciona en la mayoría de los casos*/
30 aE_bgd = aEn == 0;

/* Cuando la dinámica de la señal es alta y la energía está cerca de la estimación de fondo*/
sd1_bgd = (st-> sign_dyn_lp > 15) && (Etot - st-> Etot_l_lp) < 2*st-> Etot_v_h2 && st-> harm_cor_cnt > 20;

35 /* Las condiciones de init disminuyen continuamente act_pred y/o 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 ||
40 st-> lt_haco_ev < 0.15f ||
non_staB < 50.0f ||
aE_bgd);
/* La energía cercana a la estimación de fondo sirve como máscara para otros detectores de fondo*/
bg_bgd2 = Etot < Etot_l_lp_thr || tn_ini;

45 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 para decidir si se debe hacer una actualización. El
principal paso de decisión en la lógica de actualización de ruido es si se debe realizar o no una actualización, y esto
se forma mediante la evaluación de una expresión lógica, que se subraya a continuación. El nuevo parámetro
50 NEW_POS_BG (nuevo en relación con la solución en el Anexo A y el documento WO 2011/049514) es un detector
de pausa, y se obtiene basándose en las ganancias de predicción lineal que van desde modelo de orden de 0 a 2º, y
de 2º a 16º de un filtro de predicción lineal, y tn_ini se obtiene basándose en las características relacionadas con la
cercanía espectral. Aquí sigue una lógica de decisión que usa las nuevas características, de acuerdo con la
realización ejemplar.

55 updt_step = 0.0f;
if ((bg_bgd2 && (aE_bgd || sd1_bgd || lt_tn_track > 0.90f || NEW_POS_BG))
tn_ini)
{
60 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))
65 )
{

```

```

updt_step = 1.0f;
st-> first_noise_updt = 1;
for (i = 0; i <NB_BANDS; i ++)
{
5  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) ||
10 (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
)
{
15  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)))
20  {
updt_step = 0.01f;
}
if (updt_step> 0.0f)
{
25  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]);
}
30  }
}
else if (aE_bgd || st-> harm_cor_cnt> 100)
{
(st-> first_noise_updt) + = 1;
35  }
}
else
{
/* Si en la música baja bckr para caer más*/
40  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 ++)
{
45  if (st-> bckr [i]> 2*E_MIN)
{
st-> bckr [i] = 0.98f* st-> bckr [i];
}
}
50  }
}
st-> lt_aEn_zero = 0.2f * (st-> aEn == 0) + (1 -0.2f) * st-> lt_aEn_zero;

```

55 Como se indicó anteriormente, las características de la predicción lineal proporcionan un análisis independiente del nivel de la señal de entrada que mejora la decisión para la actualización de ruido de fondo, que 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 señales de voz

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

65

Las figuras 8-10 muestran cómo los parámetros o métricas respectivos se comportan para la voz en el fondo con un ruido de automóvil de SNR de 10dB. 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 por 10 para que sea 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 emisiones, donde la posición aproximada para la primera emisión está en las tramas 1310-1420 y para la segunda emisión, en las tramas 1500-1610.

La figura 8 muestra la energía de la trama (/10) (punto, ".") y las características G_{0_2} (círculo, "o") y $G_{max_0_2}$ (más, "+"), para voz de SNR de 10 dB con ruido de automóvil. Téngase en cuenta que el G_{0_2} es 8 durante el ruido de automóvil ya que hay cierta correlación en la señal que puede modelarse utilizando la predicción lineal con el orden de modelo 2. Durante los enunciados, la característica $G_{max_0_2}$ pasa a ser más de 1,5 (en este caso) y después de la ráfaga de voz cae a 0. En una implementación específica de una lógica de decisión, el $G_{max_0_2}$ necesita estar por debajo de 0,1 para permitir que las actualizaciones de ruido usen esta característica.

La figura 9a muestra la energía de trama (/10) (punto, ".") y las características G_{2_16} (círculo, "o"), $G_{1_2_16}$ (cruz, "x"), $G_{2_2_16}$ (más, "+"). La figura 9b muestra la energía de trama (/10) (punto, "."), y las características G_{2_16} (círculo, "o") $G_{d_2_16}$ (cruz, "x"), y $G_{ad_2_16}$ (más, "+"). La figura 9c muestra la energía de trama (/10) (punto, ".") y las características G_{2_16} (círculo, "o") y $G_{max_2_16}$ (más, "+"). Los diagramas que se muestran en las figuras 9a-c también se refieren a voz SNR de 10 dB de con ruido de automóvil. Las características se muestran en tres diagramas para facilitar la visualización de cada parámetro. Téngase en cuenta que el G_{2_16} (círculo, "o") está justo por encima de 1 durante el ruido de automóvil (es decir, emisiones exteriores) que indica que la ganancia del orden superior del modelo es baja para este tipo de ruido. Durante los enunciados, la característica $G_{max_2_16}$ (más, "+" en la figura 9c) aumenta, y luego comienza a caer de nuevo a 0. En una implementación específica de una lógica de decisión, la característica $G_{max_2_16}$, también tiene que ser inferior a 0,1 para permitir actualizaciones de ruido. En esta muestra de señal de audio particular, esto no ocurre.

La figura 10 muestra la energía de trama (punto, ".") (no dividido por 10 esta vez) y la función nonstaB (más, "+") para voz SNR de 10 dB con ruido de automóvil. La característica nonstaB está en el rango 0-10 durante los segmentos de solo ruido, y para los enunciados, 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 emisiones, hay tramas en las que la característica nonstaB entra en el rango de 0-10. Para estas tramas podría existir la posibilidad de realizar actualizaciones de ruido de fondo y, por lo tanto, realizar un mejor seguimiento del ruido de fondo.

La solución divulgada en el presente documento también se refiere a un estimador de ruido de fondo implementado en hardware y/o software.

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

Una realización ejemplar de un estimador de ruido de fondo se ilustra de forma general en la figura 11a. Por estimador de ruido de fondo se hace referencia a un módulo o entidad configurada 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 del método descrito anteriormente. El estimador de ruido de fondo se describirá brevemente para evitar repeticiones innecesarias.

El estimador de ruido de fondo puede implementarse y/o describirse de la siguiente manera:

El estimador 1100 de ruido de fondo está configurado para estimar un ruido de fondo de una señal de audio. El estimador 1100 de ruido de fondo comprende circuitería de procesamiento, o medios 1101 de procesamiento 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, basándose en una primera ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de orden 0 y una señal residual de una predicción lineal de 2º orden para el segmento de señal de audio; y una segunda ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de 2º orden y una señal residual de una predicción lineal de 16º orden para el segmento de señal de audio.

La circuitería 1101 de procesamiento está configurada además para hacer 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, basándose en al menos un parámetro. La circuitería 1101 de procesamiento está configurada además para hacer que el estimador de ruido de fondo actualice una estimación de ruido de fondo basándose en el segmento de señal de audio cuando el segmento de señal de audio comprende una pausa.

La interfaz 1102 de comunicación, que también se puede indicar, por ejemplo interfaz de entrada/salida (E/S), incluye una interfaz para enviar datos y recibir datos de otras entidades o módulos. Por ejemplo, las señales residuales relacionadas con los órdenes 0, 2 y 16 de modelo de predicción lineal pueden obtenerse, por ejemplo

recibirse, a través de la interfaz de E/S desde un codificador de señal de audio que realiza una codificación predictiva lineal.

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

10 Una implementación alternativa de la circuitería 1101 de procesamiento se muestra en la figura 11c. La circuitería de procesamiento aquí comprende una unidad o módulo 1106 de obtención o determinación, configurada 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, basándose en una primera ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de orden 0 y una señal residual de una predicción lineal de 2º orden para el segmento de señal de audio; y una segunda ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de 2º orden y una señal residual de una predicción lineal de 16º orden para el segmento de señal de audio. El circuito de procesamiento comprende además una unidad o módulo 1107 de determinación, configurada 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 como voz y música, basándose al menos en al menos dicho parámetro. La circuitería 1101 de procesamiento comprende además una unidad o módulo 1110 de actualización o estimación, configurada para hacer que el estimador de ruido de fondo actualice una estimación de ruido de fondo basándose en el segmento de señal de audio cuando el segmento de señal de audio comprende una pausa.

25 La circuitería 1101 de procesamiento podría comprender más unidades, tales como una unidad o módulo de filtro configurada para hacer que el estimador de ruido de fondo filtre de paso bajo a las ganancias de predicción lineal, creando así una o más estimaciones a largo plazo de las ganancias de predicción lineal. Acciones como el filtrado de paso bajo pueden realizarse de otro modo, por ejemplo por la unidad o módulo 1107 de determinación.

30 Las realizaciones de un estimador de ruido de fondo descrito anteriormente podrían configurarse para las diferentes realizaciones de método descritas en el presente documento, tales como limitar y filtrar de paso bajo las ganancias de predicción lineal; determinar una diferencia entre las ganancias de predicción lineal y las estimaciones a largo plazo y entre las estimaciones a largo plazo; y/o obtener y usar una medida de cercanía espectral, etc.

35 Puede suponerse que el estimador 1100 de ruido de fondo comprende una funcionalidad adicional, para llevar a cabo una 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 ejemplar. El estimador 1200 de ruido de fondo comprende una unidad de entrada, por ejemplo para recibir energías residuales para los órdenes 0, 2 y 16 de modelo. El estimador de ruido de fondo comprende además un procesador y una memoria, conteniendo dicha memoria instrucciones ejecutables por dicho procesador, por lo que dicho estimador de ruido de fondo es operativo para: realizar un método de acuerdo con una realización descrita en el presente documento.

45 Por consiguiente, el estimador de ruido de fondo puede comprender, como se ilustra en la figura 13, una unidad 1301 de entrada/salida, un calculador 1302 para calcular los primeros dos conjuntos de características de las energías residuales para los órdenes 0, 2 y 16 de modelo 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 video, grabadora de sonido, tableta, ordenador de escritorio, ordenador portátil, decodificador de televisión o servidor doméstico/puerta de enlace doméstica/punto de acceso doméstico/enrutador doméstico. El dispositivo de comunicación puede ser en algunas realizaciones un dispositivo de red de comunicaciones adaptado para codificar y/o transcodificar señales de audio. Ejemplos de tales dispositivos de red de comunicaciones son servidores, tales como servidores de medios, servidores de aplicaciones, enrutadores, pasarelas y estaciones base de radio. El dispositivo de comunicación también puede estar adaptado para colocarse, es decir, estar integrado en, un buque, tal como un barco, dron volador, avión y un vehículo de carretera, tal como un automóvil, autobús o camión. Tal dispositivo integrado típicamente pertenecería a una unidad telemática de vehículo o sistema de infoentretenimiento de vehículo.

65 Los pasos, funciones, procedimientos, módulos, unidades y/o bloques descritos en el presente documento pueden implementarse en hardware que use cualquier tecnología convencional, tal como tecnología de circuito discreto o de circuito integrado, que incluye circuitería electrónica de propósito general y circuitería específica de aplicación.

Los ejemplos particulares incluyen uno o más procesadores de señal digital configurados adecuadamente y otros

circuitos electrónicos conocidos, por ejemplo pasarelas lógicas discretas interconectadas para realizar una función especializada, o circuitos integrados de aplicación específica (los ASIC).

Alternativamente, al menos algunos de los pasos, funciones, procedimientos, módulos, unidades y/o bloques descritos anteriormente pueden implementarse en un software tal como un programa informático para su ejecución mediante una circuitería de procesamiento adecuada que incluye una o más unidades de procesamiento. El software podría ser transportado por una portadora, tal 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.

El diagrama o diagramas de flujo presentados en este documento pueden considerarse como un diagrama o diagramas de flujo informático, cuando se realizan por uno o más procesadores. Un aparato correspondiente se puede definir 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 implementan como un programa informático que se ejecuta en el procesador.

Ejemplos de circuitería de procesamiento incluyen, pero no están limitados a, uno o más microprocesadores, uno o más procesadores de señal digital, DSP, una o más unidades de procesamiento central, CPU y/o cualquier circuitería lógica programable adecuada tal como una o más matrices de pasarela programables de campo, las FPGA, o uno o más controladores lógicos programables, PLC. Es decir, las unidades o módulos en las disposiciones en los diferentes nodos descritos anteriormente podrían implementarse mediante una combinación de circuitos analógicos y digitales, y/o uno o más procesadores configurados con software y/o firmware, por ejemplo almacenados en una memoria. Uno o más de estos procesadores, así como el otro hardware digital, se pueden incluir en una única circuitería integrada de aplicación específica, ASIC, o varios procesadores y diversos hardware digitales se pueden distribuir entre varios componentes separados, ya sea empaquetados individualmente o ensamblados en un sistema en un chip, SoC.

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

Las realizaciones descritas anteriormente se dan meramente como ejemplos, y debe entenderse que la tecnología propuesta no está limitada a ellas. Los expertos en la técnica entenderán que pueden realizarse diversas modificaciones, combinaciones y cambios en las realizaciones sin apartarse del alcance actual. En particular, diferentes soluciones de piezas en las diferentes realizaciones se pueden combinar en otras configuraciones, cuando sea técnicamente posible.

Cuando se usa la palabra "comprender" o "que comprende", debe interpretarse como no limitativa, es decir, que significa "consiste al menos en".

También se debe tener en cuenta que en algunas implementaciones alternativas, las funciones/actos observados en los bloques pueden ocurrir fuera del orden indicado en los diagramas de flujo. Por ejemplo, dos bloques mostrados en sucesión pueden, de hecho, ejecutarse de manera sustancialmente simultánea o los bloques a veces pueden ejecutarse en el orden inverso, dependiendo de la funcionalidad/actos involucrados. Además, la funcionalidad de un bloque dado de los diagramas de flujo y/o diagramas de bloques se puede separar en múltiples bloques y/o la funcionalidad de dos o más bloques de los diagramas de flujo y/o diagramas de bloques se puede integrar al menos parcialmente. Finalmente, se pueden añadir/insertar otros bloques entre los bloques que se ilustran, y/o se pueden omitir bloques/operaciones sin apartarse del alcance de los conceptos de la invención.

Debe entenderse que la elección de las unidades que interactúan, así como la denominación de las unidades dentro de esta divulgación son solo para fines ejemplares, y los nodos adecuados para ejecutar cualquiera de los métodos descritos anteriormente pueden configurarse de varias maneras alternativas para poder ejecutar las acciones de procedimiento sugeridas.

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

La referencia a un elemento en singular no pretende significar "uno y solo uno" a menos que así se indique explícitamente, sino más bien "uno o más". Además, no es necesario que un dispositivo o método aborde todos y cada uno de los problemas que se pretende resolver mediante la tecnología divulgada en el presente documento, para que sean abarcados por este documento.

En algunos ejemplos del presente documento, se omiten descripciones detalladas de dispositivos, circuitos y métodos bien conocidos para no ocultar la descripción de la tecnología divulgada con detalles innecesarios. Todas las afirmaciones en el presente documento que enumeran principios, aspectos y realizaciones de la tecnología

divulgada, así como sus ejemplos específicos, pretenden abarcar sus equivalentes estructurales y funcionales. Adicionalmente, se pretende que tales equivalentes incluyan tanto equivalentes actualmente conocidos como equivalentes desarrollados en el futuro, por ejemplo cualquier elemento desarrollado que realice la misma función, independientemente de la estructura.

5

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" a continuación corresponde a la figura A2 en los dibujos.

10

La figura 2 es un diagrama de flujo que ilustra una realización ejemplar de un método para la estimación de 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 además 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, el ajuste de 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 inferior a la actual se usa directamente, como muy probablemente sería desde una trama de fondo. La siguiente lógica de estimación de ruido es un segundo paso en el que se decide si puede aumentarse la estimación de ruido de subbanda y, en caso afirmativo, cuánto, el aumento se basa en la posible nueva estimación de ruido de subbanda calculada previamente. Básicamente, esta lógica forma la decisión de la trama actual en una trama de fondo y si no está seguro puede permitir un aumento menor en comparación con lo que se estimó originalmente.

15

20

El método ilustrado en la figura 2 comprende: cuando un nivel de energía de un segmento de señal de audio es más de un umbral superior 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 menos que un umbral superior 202:2 que It_{min} , pero no se detecta pausa 204:1 en el segmento de señal de audio:

25

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

30

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

35

El nivel de energía del segmento de señal de audio usado en el método descrito anteriormente puede referirse 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.

40

La otra característica de energía usada en el método anterior, es decir, el nivel mínimo de energía a largo plazo, It_{min} , es una estimación, que se determina sobre una pluralidad de segmentos o tramas de señal de audio precedentes. It_{min} podría alternativamente denominarse, por ejemplo $Etot_l_lp$. Una forma básica de derivar It_{min} sería usar el valor mínimo de la historia de la energía de trama actual sobre un cierto número de tramas anteriores. Si el valor calculado como: "energía de trama actual - estimación mínima a largo plazo" está por debajo de un valor umbral, denominado por ejemplo THR1, se dice que la energía de trama actual es cercana a la energía mínima a largo plazo, o está cerca de la energía mínima a largo plazo. Es decir, cuando $(Etot - It_{min}) < THR1$, la energía de trama actual, $Etot$, puede determinarse 202 para estar cerca de la energía mínima a largo plazo It_{min} . El caso en que $(Etot - It_{min}) = THR1$ puede referirse a cualquiera de las decisiones, 202:1 o 202:2, dependiendo de la implementación. La numeración 202:1 en la figura 2 indica la decisión de que la energía de trama actual no está cerca de It_{min} , mientras que 202:2 indica la decisión de que la energía de trama actual está cerca de It_{min} . Otra numeración en la figura 2 en la forma XXX:Y indica las decisiones correspondientes. La característica It_{min} se describirá con más detalle a continuación.

45

50

Se puede suponer que el valor mínimo, que la estimación de ruido de fondo actual debe exceder, para ser reducida, es cero o un pequeño valor positivo. Por ejemplo, como se ejemplificará en el código a continuación, una energía total actual de la estimación de fondo, que puede denominarse "totalNoise" y se puede determinar, por ejemplo como $10 * \log_{10}(\sum backr[i])$, se puede requerir que exceda un valor mínimo de cero para que la reducción sea cuestionada. Alternativamente, o además, cada entrada en un vector $backr[i]$ que comprende las estimaciones de fondo de subbanda se puede comparar con un valor mínimo, E_{MIN} , para que se realice la reducción. En el siguiente ejemplo de código, E_{MIN} es un pequeño valor positivo.

60

Debe observarse 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 que un umbral superior a It_{min} se basa únicamente en información derivada de la señal de audio de entrada, es decir, no se basa en la retroalimentación de una decisión de detector de actividad de sonido.

65

La determinación 204 de si una trama actual comprende una pausa o no se puede realizar de diferentes maneras basándose en uno o más criterios. Un criterio de pausa también puede denominarse detector de pausa. Se podría aplicar un solo detector de pausa, o una combinación de diferentes detectores de pausa. Con una combinación de detectores de pausa, cada uno puede usarse 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 es baja, y que una serie de tramas precedentes también han 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 I_{t_min} : se ha determinado que un número predefinido de segmentos de señal de audio precedentes consecutivos no comprende una señal activa y/o una dinámica de la señal de audio excede un umbral. Esto también se ilustra en el siguiente ejemplo de código.

La reducción 206 de la estimación de ruido de fondo permite el manejo de situaciones en las que la estimación de ruido de fondo se ha vuelto "demasiado alta", es decir, en relación con un verdadero ruido de fondo. Esto también podría expresarse por ejemplo como que la estimación de ruido de fondo se desvía del ruido de fondo real. Una estimación de ruido de fondo demasiado alta puede conducir a decisiones inadecuadas por el SAD, donde se determina que el segmento de señal actual está inactivo a pesar de que comprende voz o música activas. Una razón para que la estimación de ruido de fondo sea demasiado alta es, por ejemplo 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 erróneamente actualizada sea ajustada por ejemplo cuando se determina que una siguiente trama de la señal de entrada comprende música. Este ajuste se realiza mediante una reducción forzada de la estimación de ruido de fondo, donde la estimación de ruido se reduce, incluso si la energía del segmento de señal de entrada actual es mayor que la estimación de ruido de fondo actual, por ejemplo en una subbanda. Debe observarse que la lógica descrita anteriormente para la estimación de ruido de fondo se usa para controlar el aumento de la energía de la subbanda de fondo. Siempre se permite bajar la energía de la subbanda cuando la energía de la subbanda de trama actual es menor que la estimación de 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 de ruido de fondo solo debe aumentarse en asociación con la lógica de decisión de acuerdo con el método descrito anteriormente. Cuando se detecta una pausa, las características de energía y correlación también se pueden usar para decidir 207 qué tan grande debe ser el tamaño del paso de ajuste para el aumento de la estimación de fondo antes de que se realice la actualización real del ruido de fondo.

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

En los estimadores de ruido de fondo de la técnica anterior, las estimaciones de energía de subbanda solo se podrían reducir cuando una energía de subbanda de entrada fuera 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, dicha recuperación puede realizarse mediante la reducción de la estimación de ruido forzado cuando la señal de entrada vuelve a características similares a la música. Es decir, cuando la lógica de energía y pausa descrita anteriormente impiden, 202:1, 204:1, que la estimación de ruido aumente, se prueba 203 si se sospecha que la entrada es música, y si es así 203:2, las energías de subbanda se reducen 206 en una pequeña cantidad en cada trama hasta que las estimaciones de ruido alcanzan un nivel más bajo 205:2.

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

La figura 5 es un diagrama de bloques que ilustra esquemáticamente una implementación de un estimador de fondo de acuerdo con una realización ejemplar. Un bloque 51 de entramado de entrada primero divide 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 desde la entrada: 1) El extractor de características analiza la trama en el dominio de la frecuencia y se calcula la energía para un conjunto de subbandas. Las subbandas son las mismas subbandas que se usarán 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 $I_{t_cor_est}$, que se usa para determinar si la trama comprende contenido activo o no. 3) El extractor de características utiliza además 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, como la energía mínima a largo plazo, I_{t_min} . Las características de correlación y energía se suministran luego al bloque 53 de lógica de decisión de actualización.

5 Aquí, se implementa una lógica de decisión de acuerdo con la solución divulgada 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 de una energía mínima a largo plazo o no; sobre si la trama actual es parte de una pausa (señal no activa) o no; 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 estas características y decisiones se usan para actualizar la estimación de ruido de fondo de una manera robusta.

10 A continuación, se describirán algunos detalles de implementación de las realizaciones de la solución divulgada en el presente documento. Los detalles de implementación a continuación se toman de una realización en un codificador basado en G.718. Esta realización usa algunas de las características descritas en los documentos WO 2011/049514 y WO 2011/049515.

15 Las siguientes características se definen en el G.718 modificado descrito en el documento WO 2011/09514

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 de fondo potencial precalculada
aEn	Un detector de fondo que usa múltiples características (un contador)
harm_cor_cnt	Cuenta las tramas desde la última trama con correlación o evento armónico
act_pred	Una predicción de actividad desde las características de la trama de entrada solo
cor [i]	Vector con estimaciones de correlación para, $i = 0$ fin de la trama actual, $i = 1$ inicio de la trama actual, $i = 2$ final de la trama anterior

20 Las siguientes características están definidas en el G.718 modificado descrito en el documento WO 2011/09515

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

También la característica Etot_v_h se definió en el documento WO 2011/049514, pero en esta realización se ha modificado y ahora se implementa de la siguiente manera:

```

25 Etot_v = (float) fabs (*Etot_last - Etot);
   if (Etot_v < 7.0f) /*obsérvese que no se usa aquí ningún indicador de VAD o similar*/
   {
   *Etot_v_h -= 0.01f;
   if (Etot_v > * Etot_v_h)
30 {if ((*Etot_v - *Etot_v_h) > 0.2f)
   {
   *Etot_v_h = *Etot_v_h + 0.2f;
   }
   }
   else
35 {
   *Etot_v_h = Etot_v; }}}

```

40 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, la variación de energía entre dos tramas se determina como "baja" cuando la diferencia entre la última y la actual energía de trama es menor que 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, que comprende solo ruido de fondo. Sin embargo, tal baja varianza podría encontrarse alternativamente, por ejemplo en medio de una ráfaga de voz. La variable Etot_last es el nivel de energía de la trama anterior.

45 Los pasos anteriores descritos en el código se pueden realizar como parte de los pasos "calcular/actualizar la correlación y la energía" en el diagrama de flujo de la figura 2, es decir, como parte de las acciones 201. En la

implementación del documento WO 2011/049514, se usó 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 de ruido de fondo no depende de una decisión de VAD (o SAD).

5 Además, en la solución divulgada en el presente documento, las siguientes características, que no son parte de la implementación del documento WO 2011/049514, pueden calcularse/actualizarse como parte de los mismos pasos, es decir, los pasos de cálculo/actualización de correlación y energía ilustrados en la figura 2. Estas características también se usan en la lógica de decisión de si actualizar o no la estimación de fondo.

10 Para lograr una estimación de ruido de fondo más adecuada, a continuación se definen varias características. Por ejemplo, se definen las nuevas características 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` también se usa para producir `lt_cor_est`, que es una estimación suavizada a largo plazo de la correlación.

```
15 cor_est = (cor[0] + cor[1] + cor[2])/3.0f;
st-> It_cor_est = 0.01f*cor_est + 0.99f * st-> It_cor_est;
```

20 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 inicio de la trama actual, y `cor[2]` representa el final de una trama anterior.

25 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 usa para formar la medida a largo plazo `lt_tn_track`.

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

30 En este ejemplo, se añade 0,03 cuando la energía de trama actual está cerca de la estimación de ruido de fondo, y de lo contrario 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 de ruido de fondo, `totalNoise`, es inferior a 10 unidades. Otras definiciones de "cerca" también son posibles.

35 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. Una característica similar, `lt_Ellp_dist`, se crea para la distancia entre la energía mínima a largo plazo `Etot_l_lp` y la energía de trama actual, `Etot`.

```
40 st-> It_tn_dist = 0.03f* (Etot - st-> totalNoise) + 0.97f* st-> It_tn_dist;
st-> It_Ellp_dist = 0.03f* (Etot - st-> Etot_l_lp) + 0.97f* st-> It_Ellp_dist;
```

45 La característica `harm_cor_cnt`, presentada 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 que cumple ciertos criterios relacionados con la actividad. Es decir, cuando la condición `harm_cor_cnt == 0`, esto implica que es muy probable que la trama actual sea una trama activa, ya que muestra una correlación o un evento armónico. Esto se usa 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, se utilizan constantes de tiempo diferentes si la estimación se aumenta o disminuye, como se puede ver a continuación.

```
50 if (st->harm_cor_cnt == 0) /*cuando probablemente está activo*/
{
st->It_haco_ev = 0,03f + 0.97f*st->It_haco_ev; /*aumentar la estimación a largo
plazo*/ }
55 else
{
st->It_haco_ev = 0.99f*st->It_haco_ev; /* disminuir la estimación a largo plazo*/ }
```

60 Un valor bajo de la característica `lt_tn_track`, presentada 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 `lt_tn_track` se reduce para cada trama donde la energía de trama actual no está cerca de la estimación de energía de fondo. `lt_tn_track` aumenta solo cuando la energía de trama actual está cerca de la estimación de energía de fondo como se muestra arriba. Para obtener una mejor estimación de cuánto tiempo este "no seguimiento", es decir, la energía de trama está lejos de la estimación de fondo, ha durado, un contador, `low_tn_track_cnt`, para el número de tramas con esta ausencia de seguimiento se forma como:

65

```

if (st-> lt_tn_track <0.05f) /*cuando It_tn_track es bajo*/
{
st->low_tn_track_cnt ++; /*añadir 1 en el contador*/ }
else
5 {
st->low_tn_track_cnt = 0; /*resetear contador*/ }

```

En el ejemplo anterior, "bajo" se define como debajo del valor 0,05. Esto debe verse como un valor ejemplar, que se puede seleccionar de manera diferente.

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

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

bg_bgd se convertirá en "1" o "verdadero" cuando Etot esté cerca de la estimación de ruido de fondo. bg_bgd sirve como una máscara para otros detectores de fondo. Es decir, si bg_bgd no es "verdadero", los detectores 2 y 3 de fondo a continuación no necesitan ser evaluados. Etot_v_h es una estimación de varianza de ruido, que alternativamente podría denominarse Nvar. Etot_v_h se deriva de la energía total de entrada (en el dominio de registro) usando Etot_v que mide la variación de energía absoluta entre tramas. Téngase en cuenta que la función Etot_v_h está limitada solo a aumentar un máximo de un valor constante pequeño, por ejemplo 0,2 para cada trama. Etot_l_lp es una versión suavizada de la envoltura de energía mínima Etot_l.

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

cuando aEn es cero, aE_bgd se convierte en "1" o "verdadero". aEn es un contador que se 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. aEn no puede aumentarse más que hasta un cierto número, por ejemplo 6, y no se reduce a menos de cero. Después de una cantidad de tramas consecutivas, por ejemplo 6, sin una señal activa, aEn será igual a cero.

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

aquí, sd1_bgd será "1" o "verdadero" 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: un cierto número de tramas han pasado sin correlación o eventos armónicos, en este ejemplo 20 tramas.

La función de 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, aE_bgd y sd1_bgd representan la detección de pausa o de fondo en diferentes condiciones. aE_bgd es el detector más general de los dos, mientras que sd1_bgd detecta principalmente las pausas de voz en SNR alta.

Una nueva lógica de decisión de acuerdo con una realización de la tecnología divulgada en el presente documento, se construye de la siguiente manera en el código a continuación. La lógica de decisión comprende la condición de enmascaramiento bg_bgd y los dos detectores de pausa aE_bgd y sd1_bgd. También podría haber un tercer detector de pausa, que evalúa las estadísticas a largo plazo de qué tan bien TotalNoise rastrea la estimación de energía mínima. Las condiciones evaluadas si la primera línea es verdadera es la lógica de decisión sobre qué tan grande debe ser el tamaño del paso, updt_step y la actualización de la estimación de ruido real es la asignación de valor a "st-> bckr [i] = -". Téngase en cuenta que tmpN [i] es un nivel de ruido potencialmente calculado previamente calculado de acuerdo con la solución descrita en el documento WO 2011/049514. La lógica de decisión siguiente sigue la parte 209 de la figura 2, que se indica en parte en conexión con el código siguiente

```

55 if (bg_bgd && (aE_bgd II sd1_bgd II st-> It_tn_track > 0.90f)) /*si 202:2 y 204:2)*/
{
if ((st->act_pred <0.85f | (aE_bgd && st->It_haco_ev <0.05f)) &&
(st-> It_Ellp_dist <10 || sd1_bgd) && st->It_tn_dist <40 &&
((Etot - st-> totalNoise) <15.0f || st-> It_haco_ev <0.10f)) /*207*/
60 st-> first_noise_updt = 1;
for (i = 0; i <NB_BANDS; i ++ )
{
st-> bckr[i] = tmpN[i] /*208*/ }
}
65 else if (aE_bgd && st->It_haco_ev <0.15f)
{

```

```

updt_step = 0.1f;
if (st->act_pred > 0.85f)
{
updt_step = 0.01f /*207*/ }
5 if (updt_step > 0.0f)
{
st-> first_noise_updt = 1;
para [i = 0; i < NB_BANDS; i ++ )
{
10 st->bckr[i] = st->bckr[i] + updt_step * (tmpN [i] -st->bckr[i]); /*208*/
else
(st-> first_noise_updt) + = 1;
}
else
15 /*Si en la música baja bckr para caer más*/ /*if 203:2 y 205:1*/
If (st-> low_tn_track_cnt > 300 && st-> It_haco_ev > 0.9f && st-> totalNoise > 0.0f)
For (i = 0; i < NB_BANDS; i ++ )
If (st-> bckr[i] > 2 * E_MIN )
{
20 St-> bckr[i] = 0.98f * st-> bckr [i]; /*206*/
}}
Else
{
} (st-> first_noise_updt) + = 1;
25 }

```

El segmento de código en el último bloque de códigos que comienza con “/*Si en la música ... */” contiene la reducción progresiva forzada de la estimación de fondo que se utiliza si se sospecha que la entrada actual es música. Esto se decide como una función: un período prolongado de ruido de fondo de seguimiento pobre en comparación con la estimación de energía mínima, y, ocurrencias frecuentes de eventos armónicos o de correlación, y, la última condición “totalNoise > 0” es una verificación de que la energía total actual de la estimación de fondo es mayor que cero, lo 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 pequeño valor positivo. Esta es una comprobación de cada entrada en un vector que comprende las estimaciones de fondo de subbanda, de modo que una entrada necesita exceder E_MIN para poder ser reducida (en el ejemplo, multiplicándola por 0,98). Estas comprobaciones se realizan para evitar reducir las estimaciones de fondo en valores demasiado pequeños.

Las realizaciones mejoran la estimación de ruido de fondo que permite un rendimiento mejorado del SAD/VAD para lograr una solución DTX de alta eficacia y evitar la degradación en la calidad de la voz o la música causada por el recorte.

Con la eliminación de la retroalimentación de decisión descrita en el documento WO 2011/09514 del Etot_v_h, hay una mejor separación entre la estimación de ruido y el SAD. Esto tiene la ventaja de que la estimación de ruido no cambia si/cuando se cambia la función/afinación SAD. Es decir, la determinación de una estimación de ruido de fondo se vuelve independiente de la función del SAD. Además, la sintonización de la lógica de estimación de ruido se vuelve más fácil ya que uno no se ve afectado por los efectos secundarios del SAD cuando se modifican las estimaciones de fondo.

REIVINDICACIONES

1.- 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:

- 5 • obtener (201) al menos un parámetro asociado con un segmento de señal de audio, basándose en:
 - una primera ganancia de predicción lineal calculada como un 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 2º orden para el segmento de señal de audio; y
 - 10 - una segunda ganancia de predicción lineal calculada como un cociente entre una señal residual (E(2)) de una predicción lineal de 2º orden y una señal residual (E(16)) de una predicción lineal de 16º orden para el segmento de señal de audio;
 - 15 • determinar (202) si el segmento de señal de audio comprende una pausa, es decir, está libre de contenido activo tal como voz y música, basándose al menos en al menos dicho parámetro obtenido; y
 - cuando el segmento de señal de audio comprende una pausa, actualizar (203) una estimación de ruido de fondo basándose en el segmento de señal de audio.

2.- El método de acuerdo con la reivindicación 1, en el que la obtención de al menos dicho parámetro comprende:

- 25 - limitar las ganancias de predicción lineal primera y segunda, para tomar valores en un intervalo predefinido.

3.- El método de acuerdo con cualquiera de las reivindicaciones 1-2, en el que la obtención de al menos dicho parámetro comprende:

- 30 - 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 un filtrado de paso bajo, en el que la estimación a largo plazo se basa además en las correspondientes ganancias de predicción lineal asociadas con al menos un segmento de señal de audio precedente.

4.- El método de acuerdo con cualquiera de las reivindicaciones 1-3, en el que la obtención de al menos dicho parámetro comprende:

- 40 - 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.

5.- El método de acuerdo con cualquiera de las reivindicaciones precedentes, en el que la obtención de al menos dicho parámetro comprende el filtrado de paso bajo de la ganancia de predicción lineal primera y segunda.

6.- El método de acuerdo con la reivindicación 5, 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 una media de una ganancia de predicción lineal correspondiente obtenida basándose en una pluralidad de segmentos de señal de audio precedentes.

7.- 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:

- obtener al menos un parámetro basándose en:
 - 55 - una primera ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de orden 0 y una señal residual de una predicción lineal de 2º orden para el segmento de señal de audio; y
 - una segunda ganancia de predicción lineal calculada como un cociente entre una señal residual de una predicción lineal de 2º orden y una señal residual de una predicción lineal de 16º orden para el segmento de señal de audio;
 - 60 • determinar si el segmento de señal de audio comprende una pausa, es decir, está libre de contenido activo tal como voz y música, basándose al menos en al menos un parámetro; y
 - cuando el segmento de señal de audio comprende una pausa, actualizar una estimación de ruido de fondo basándose en el segmento de señal de audio.

- 8.- El estimador de ruido de fondo de acuerdo con la reivindicación 7, en el que el estimador de ruido de fondo está configurado para obtener al menos dicho parámetro limitando la ganancia de predicción lineal primera y segunda para tomar valores en un intervalo predefinido.
- 5 9.- El estimador de ruido de fondo de acuerdo con cualquiera de las reivindicaciones 7-8, en el que el estimador de ruido de fondo está configurado para obtener al menos dicho parámetro:
- 10 - creando al menos una estimación a largo plazo de cada una de las ganancias de predicción lineal primera y segunda, por ejemplo por medio de un filtrado de paso bajo, en el que la estimación a largo plazo se basa además en las correspondientes ganancias de predicción lineal asociadas con al menos un segmento de señal de audio precedente.
- 15 10.- El estimador de ruido de fondo de acuerdo con cualquiera de las reivindicaciones 7-9, en el que el estimador de ruido de fondo está configurado para obtener al menos dicho parámetro:
- determinando 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.
- 20 11.- El estimador de ruido de fondo de acuerdo con cualquiera de las reivindicaciones 7-10, en el que el estimador de ruido de fondo está configurado para obtener al menos dicho parámetro filtrando de paso bajo la ganancia de predicción lineal primera y segunda.
- 25 12.- El estimador de ruido de fondo de acuerdo con la reivindicación 11, 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 una media de una ganancia de predicción lineal correspondiente obtenida basándose en una pluralidad de segmentos de señal de audio precedentes.
- 30 13.- Un dispositivo inalámbrico que comprende un estimador de ruido de fondo de acuerdo con cualquiera de las reivindicaciones 7-12.
- 14.- Un nodo de red que comprende un estimador de ruido de fondo de acuerdo con cualquiera de las reivindicaciones 7-12.
- 35 15.- Programa informático, que comprende instrucciones que, cuando se ejecutan en al menos un procesador, hacen que al menos dicho procesador lleve a cabo el método de acuerdo con cualquiera de las reivindicaciones 1-6.
- 16.- Una portadora que contiene el programa informático de la reivindicación precedente, en el que la portadora es una de una señal electrónica, señal óptica, señal de radio o medio de almacenamiento legible por ordenador.

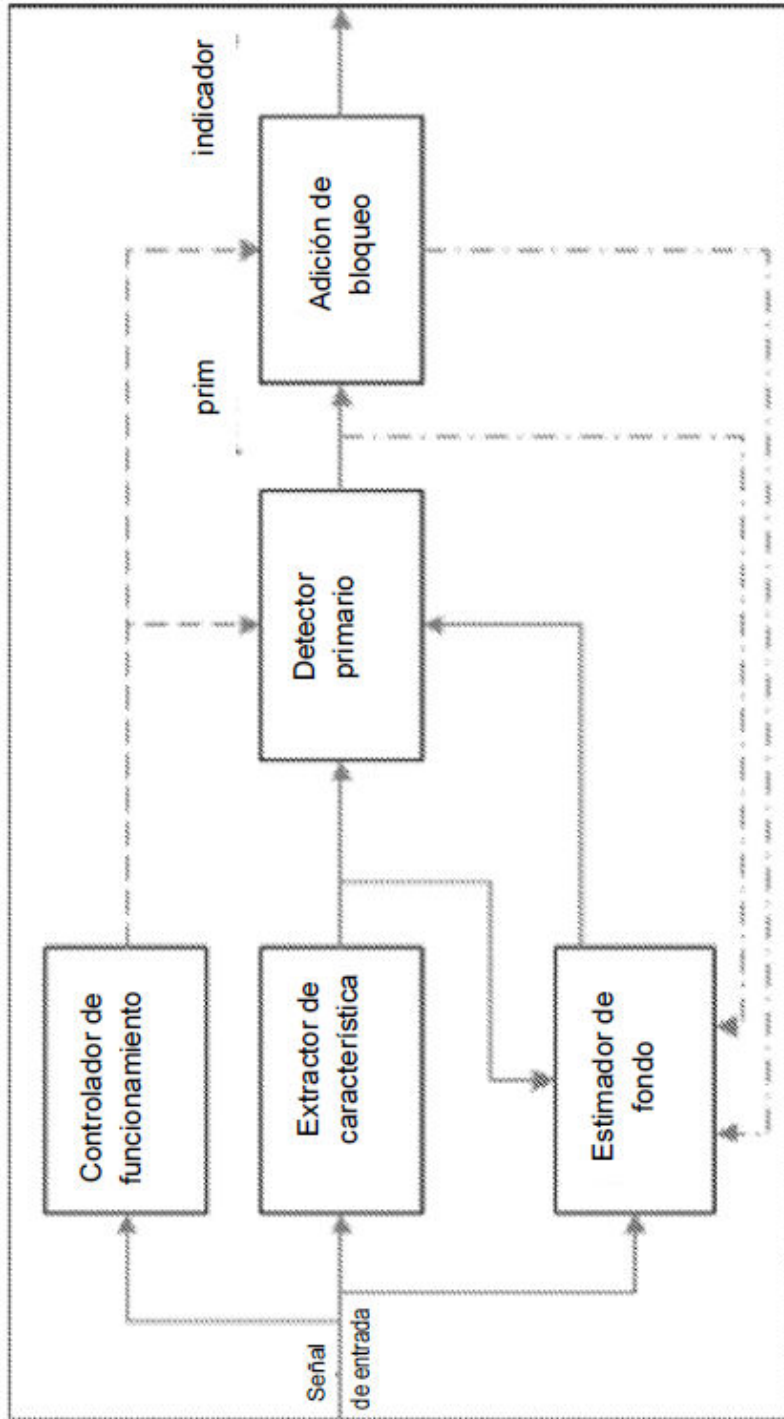


Figura 1

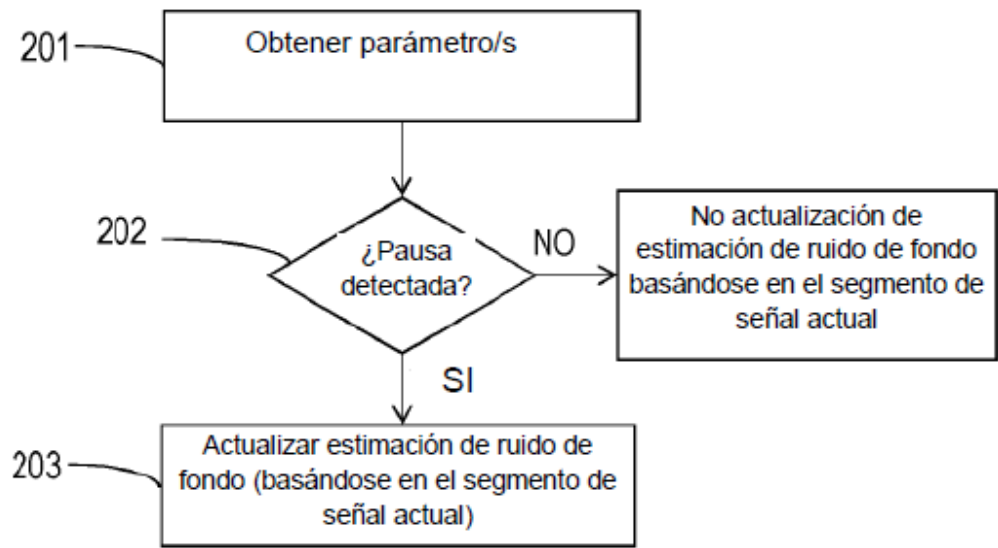


Figura 2

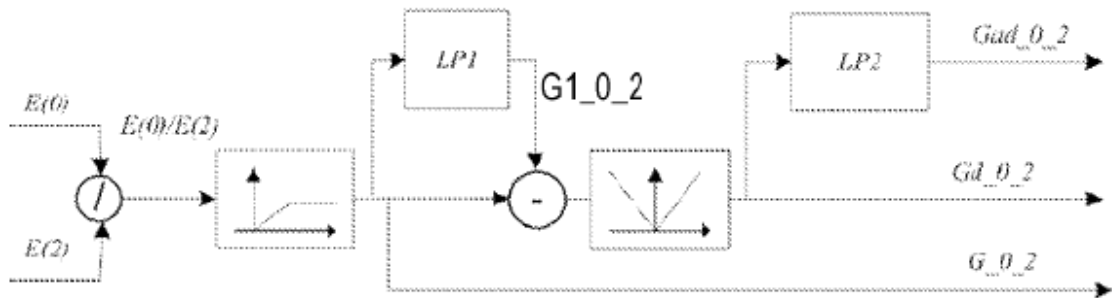


Figura 3

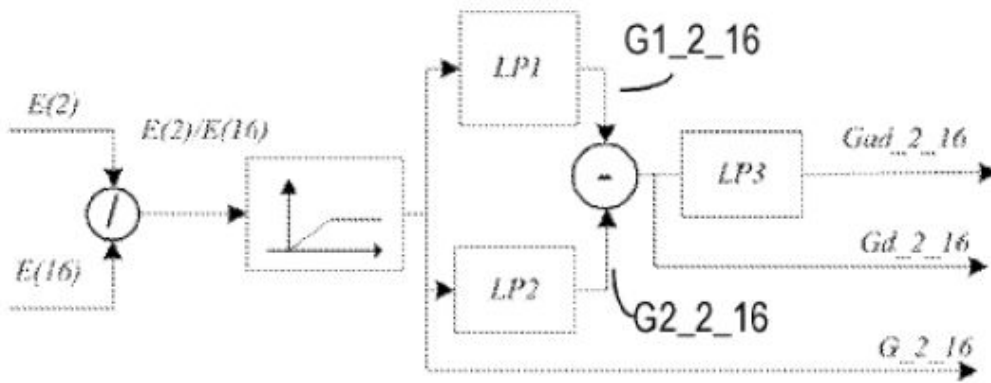


Figura 4

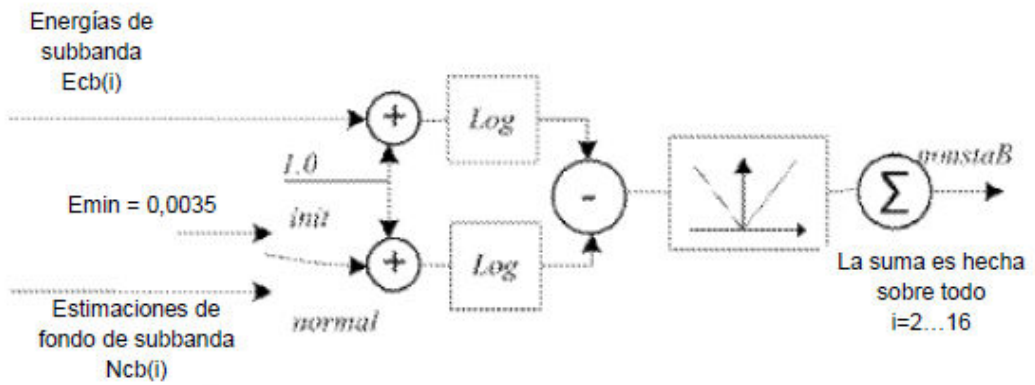


Figura 5

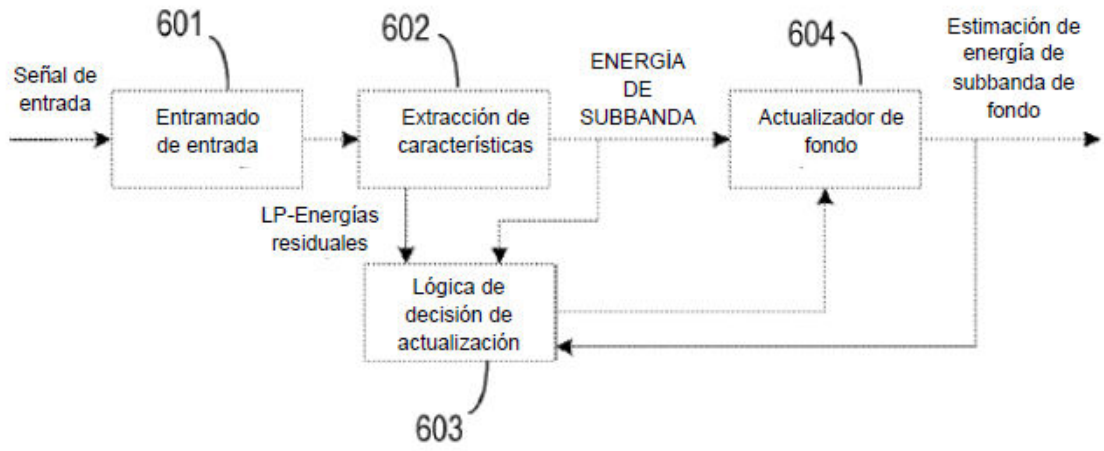


Figura 6

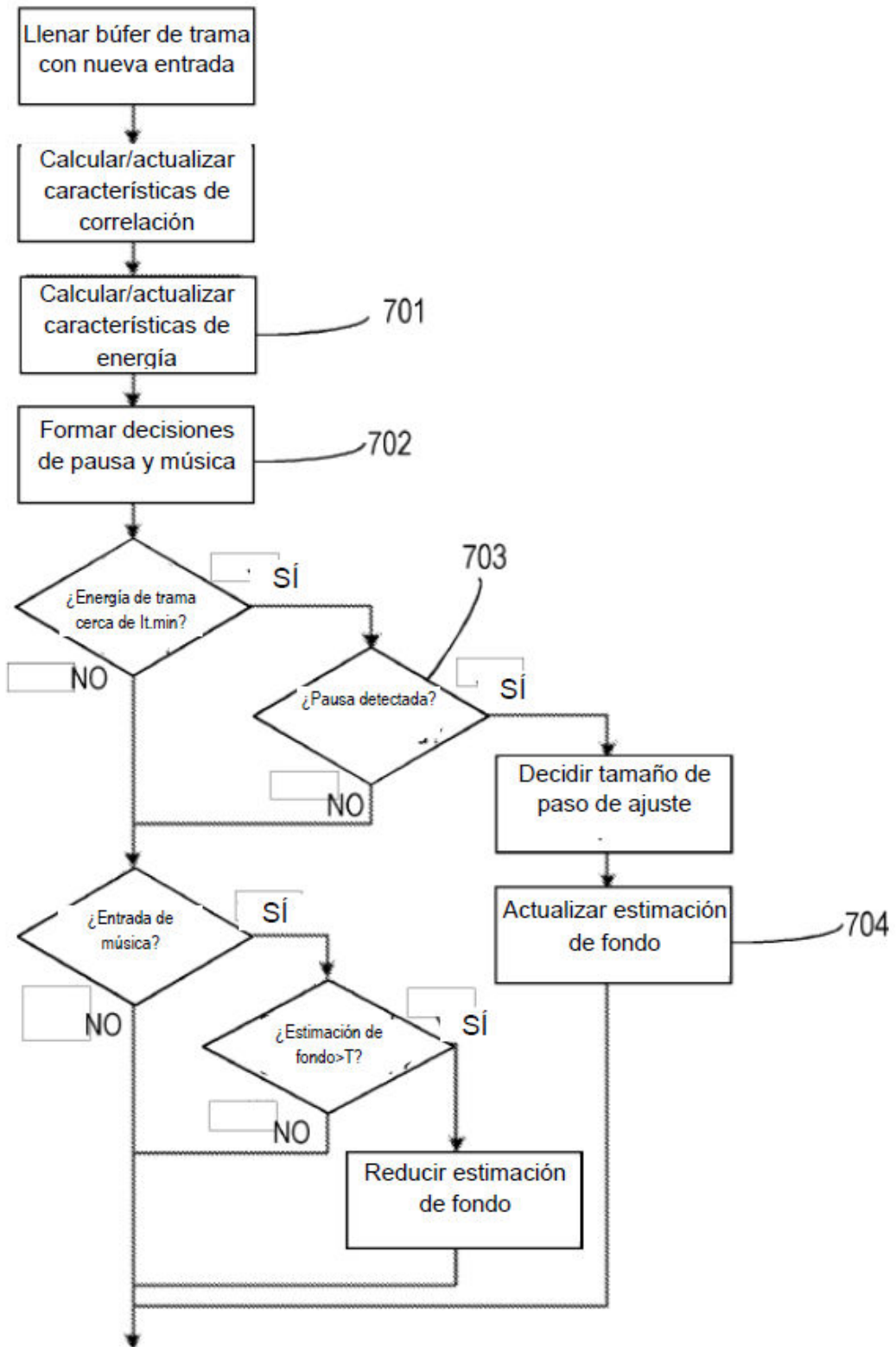


Figura 7

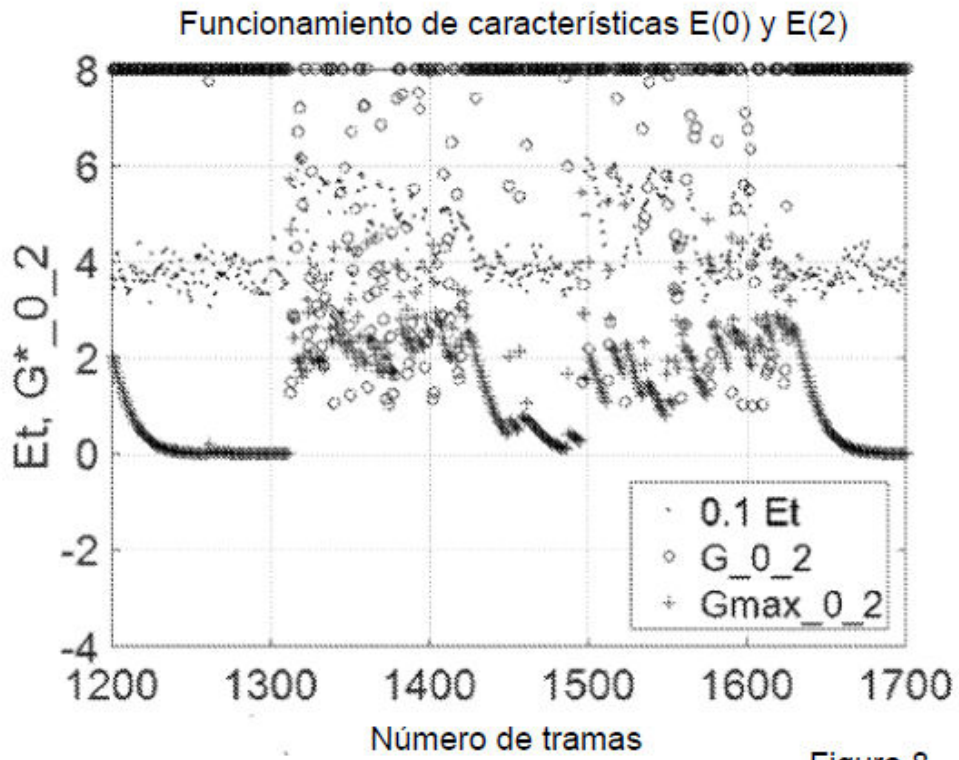


Figura 8

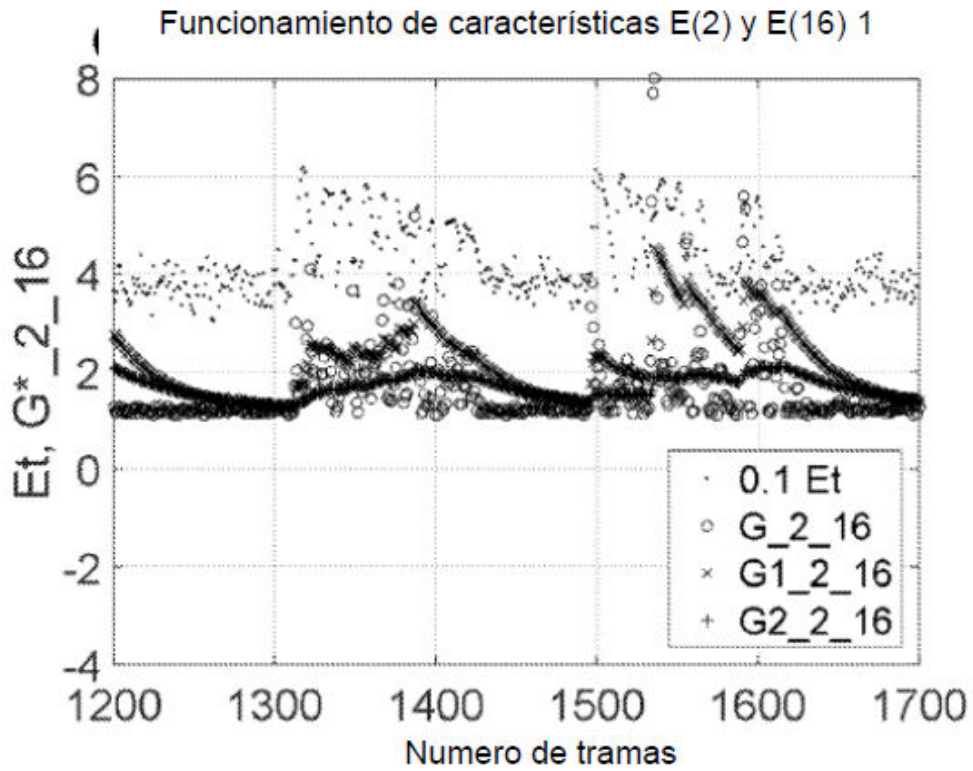
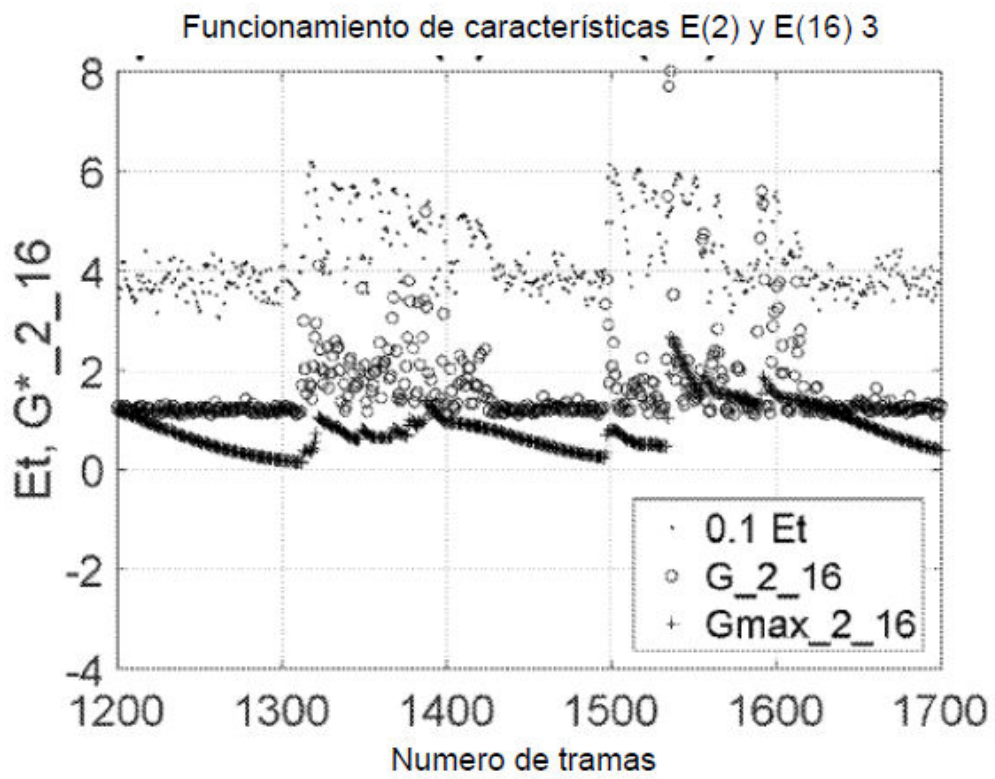
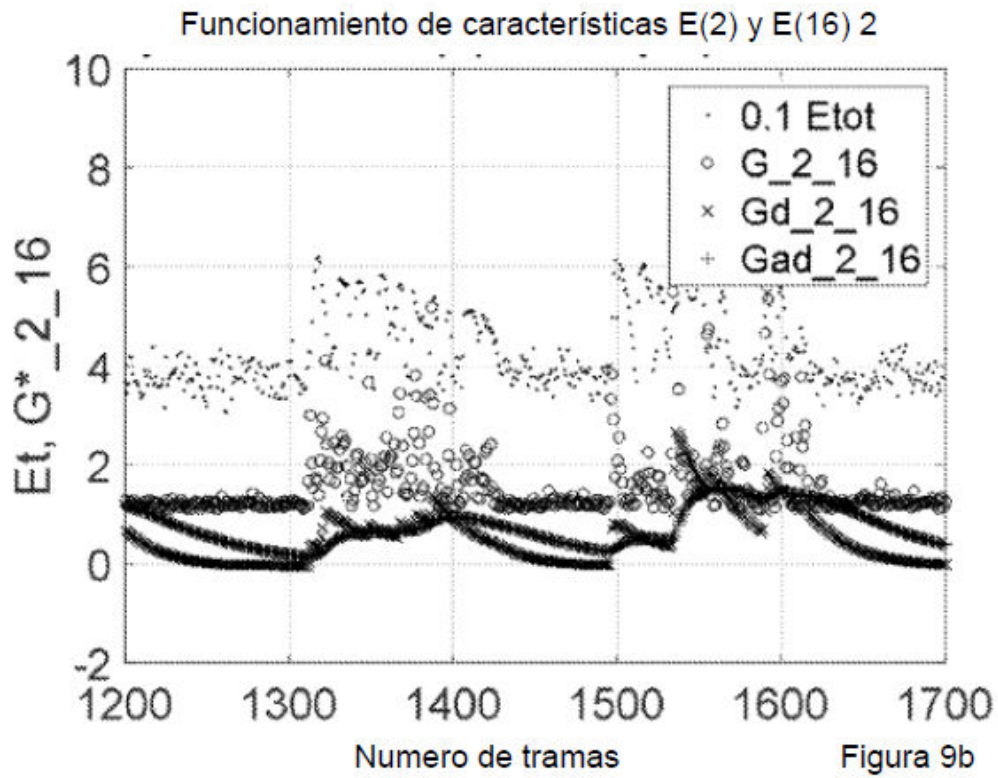


Figura 9a



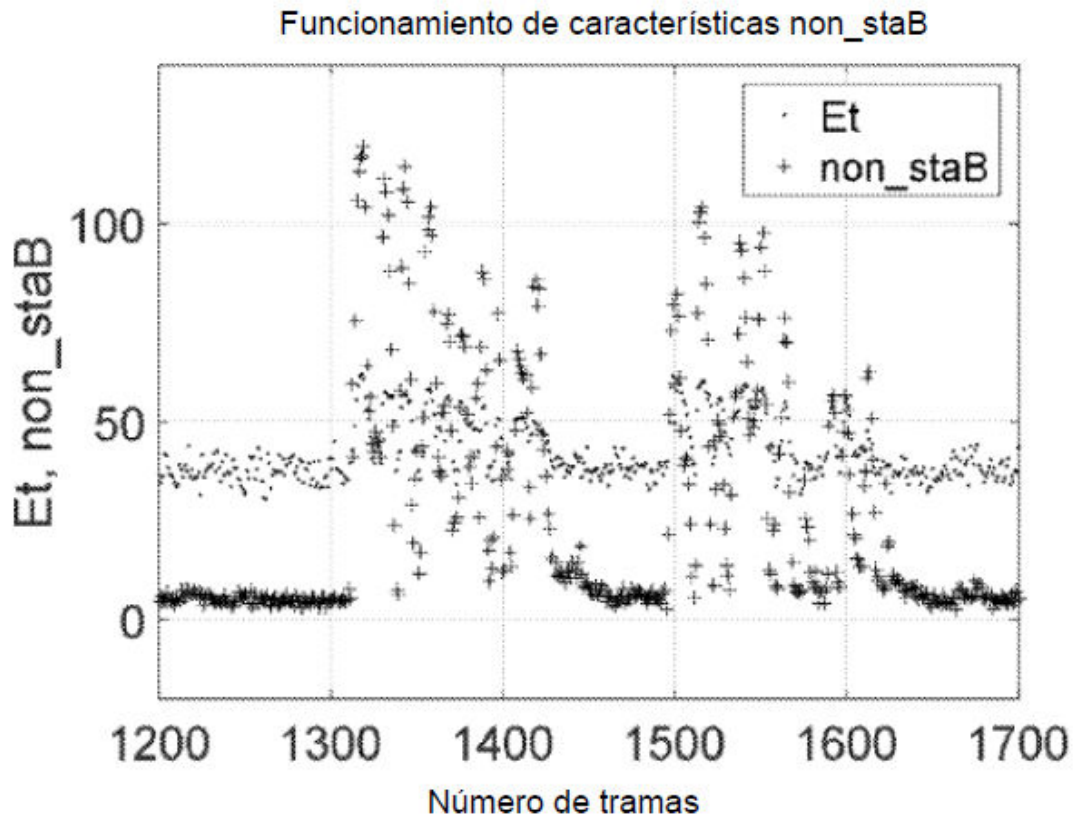


Figura 10

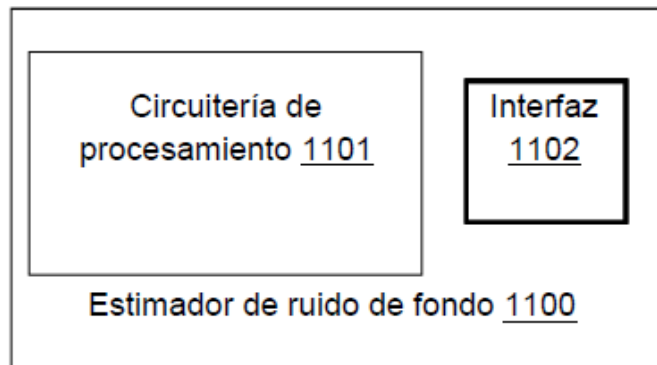


Figura 11a

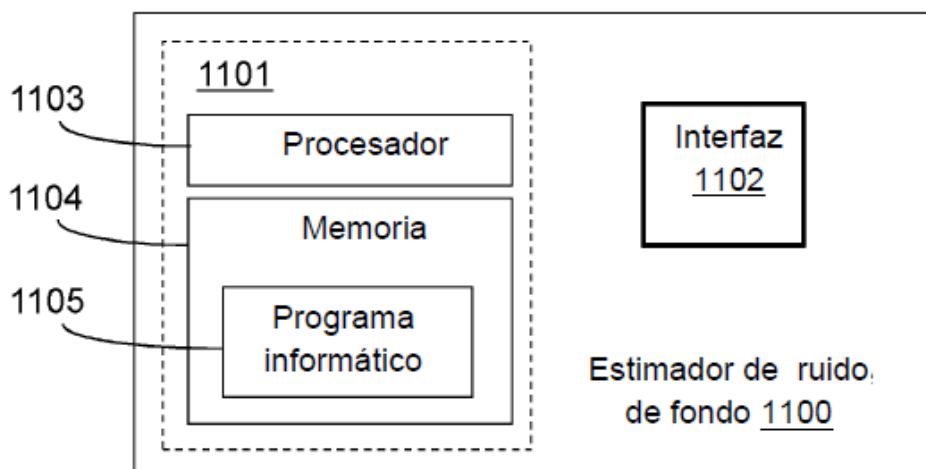


Figura 11b

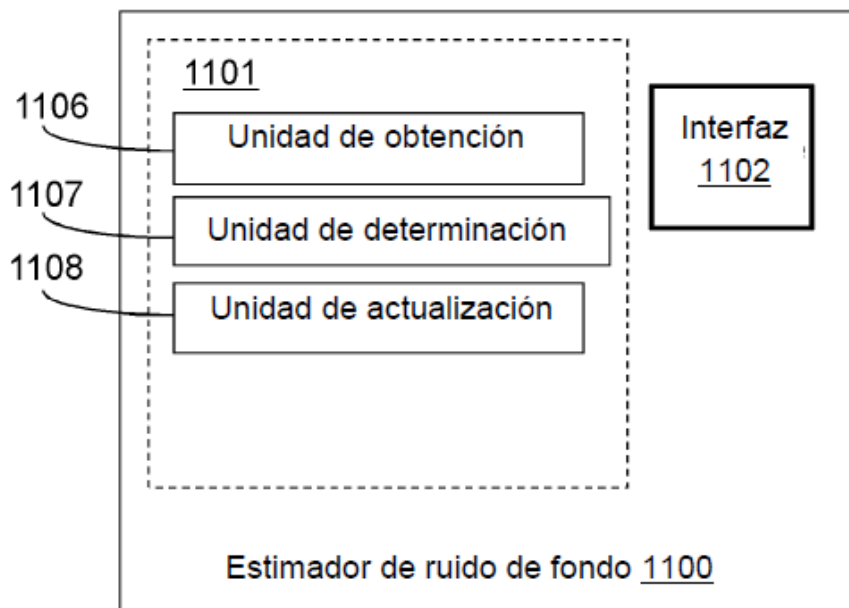


Figura 11c

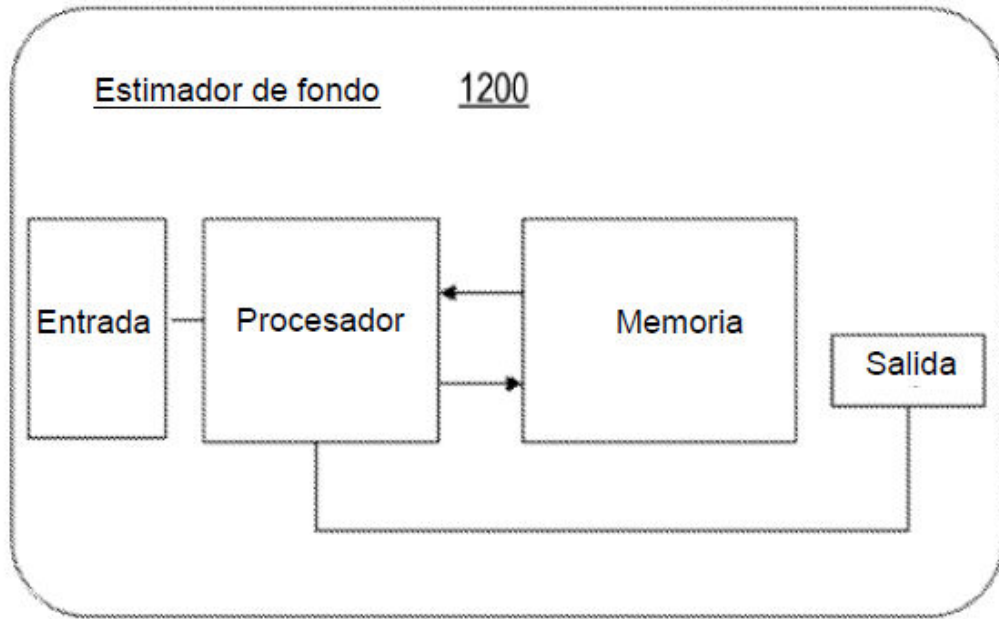


Figura 12

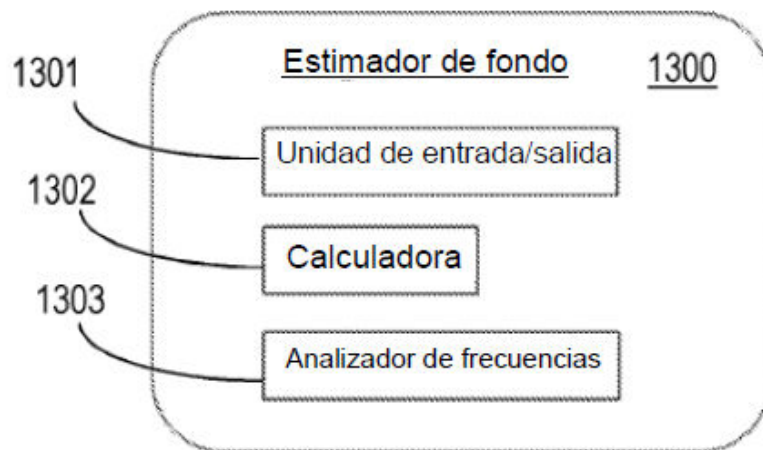


Figura 13

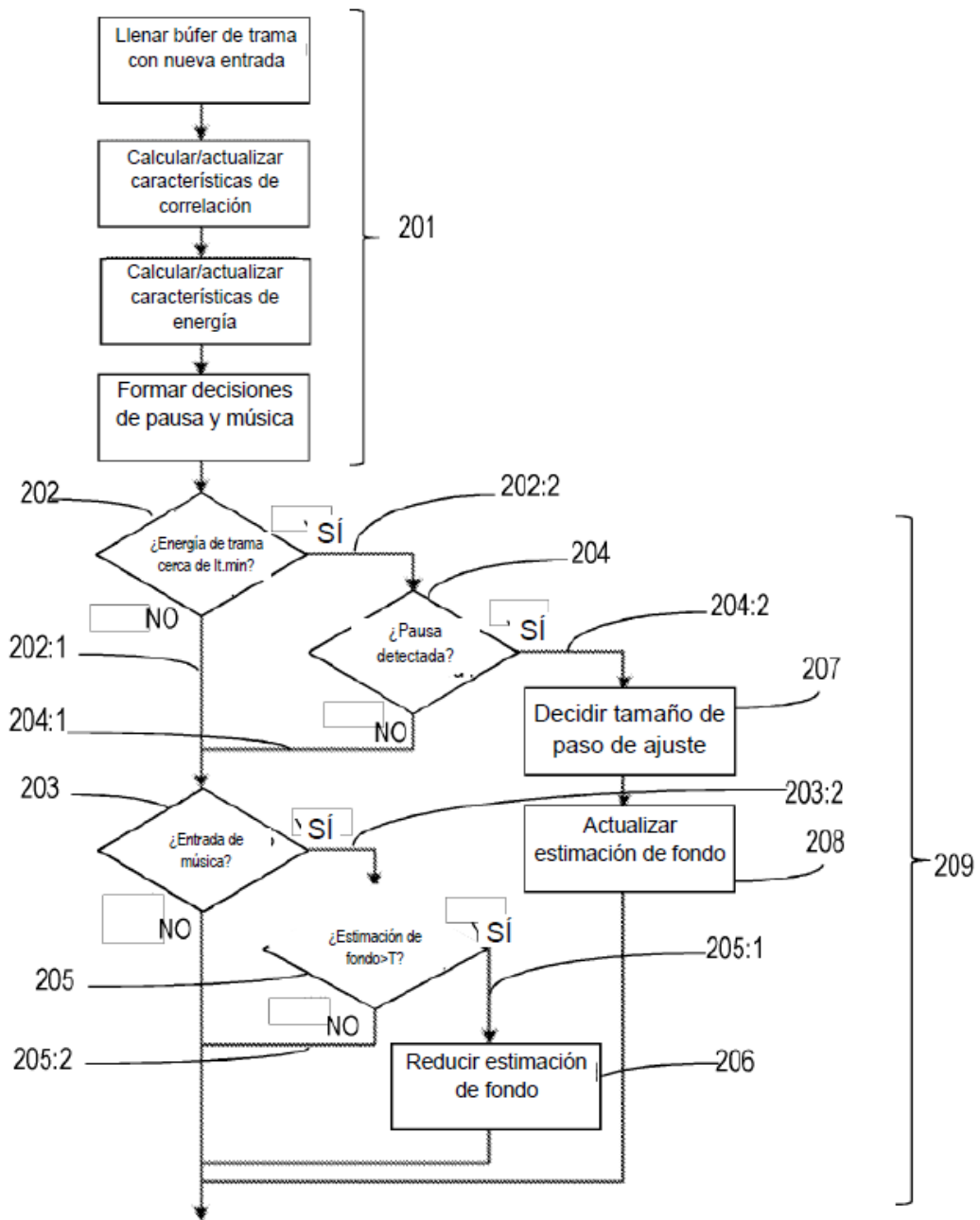


Figura A2

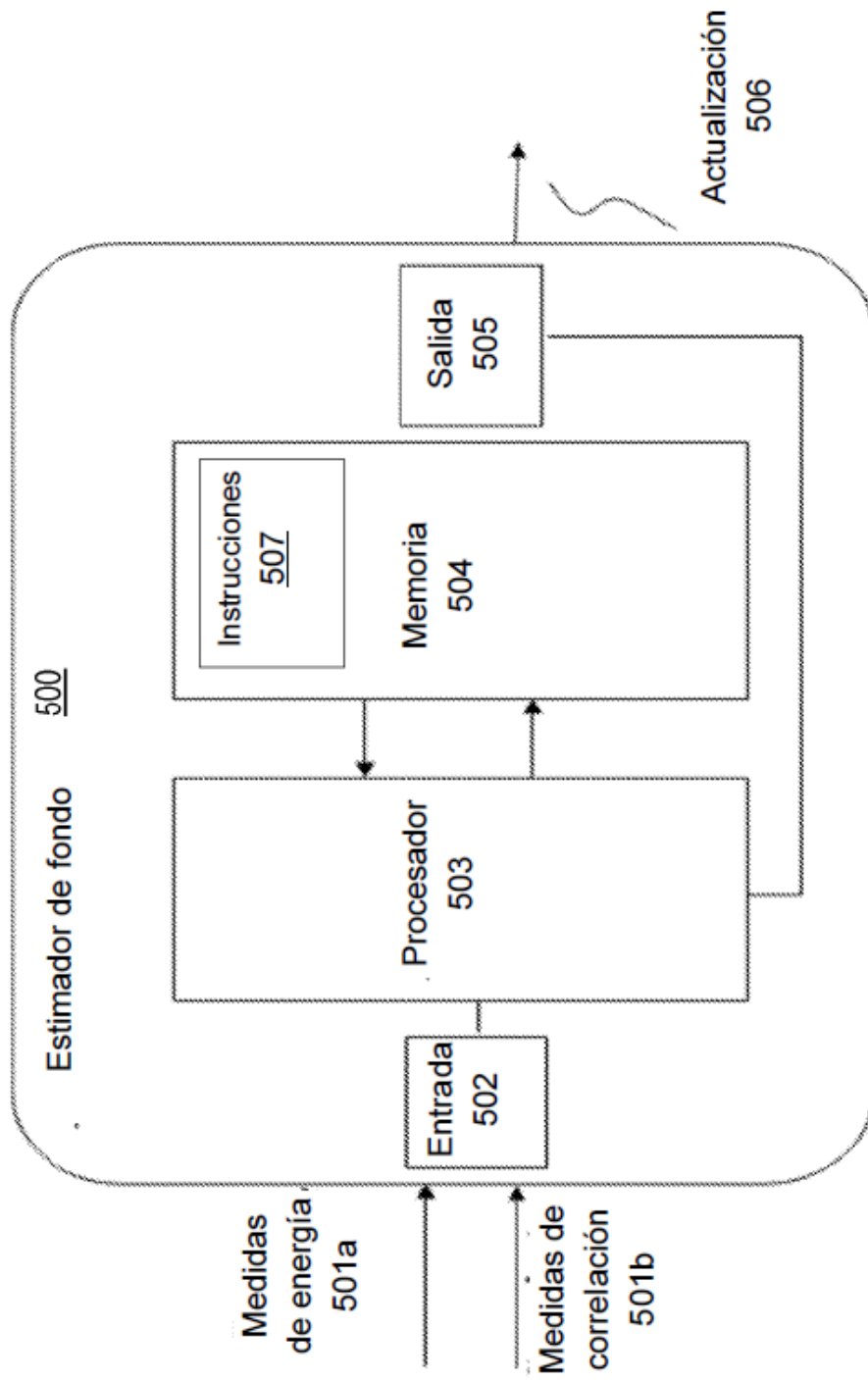


Figura A3

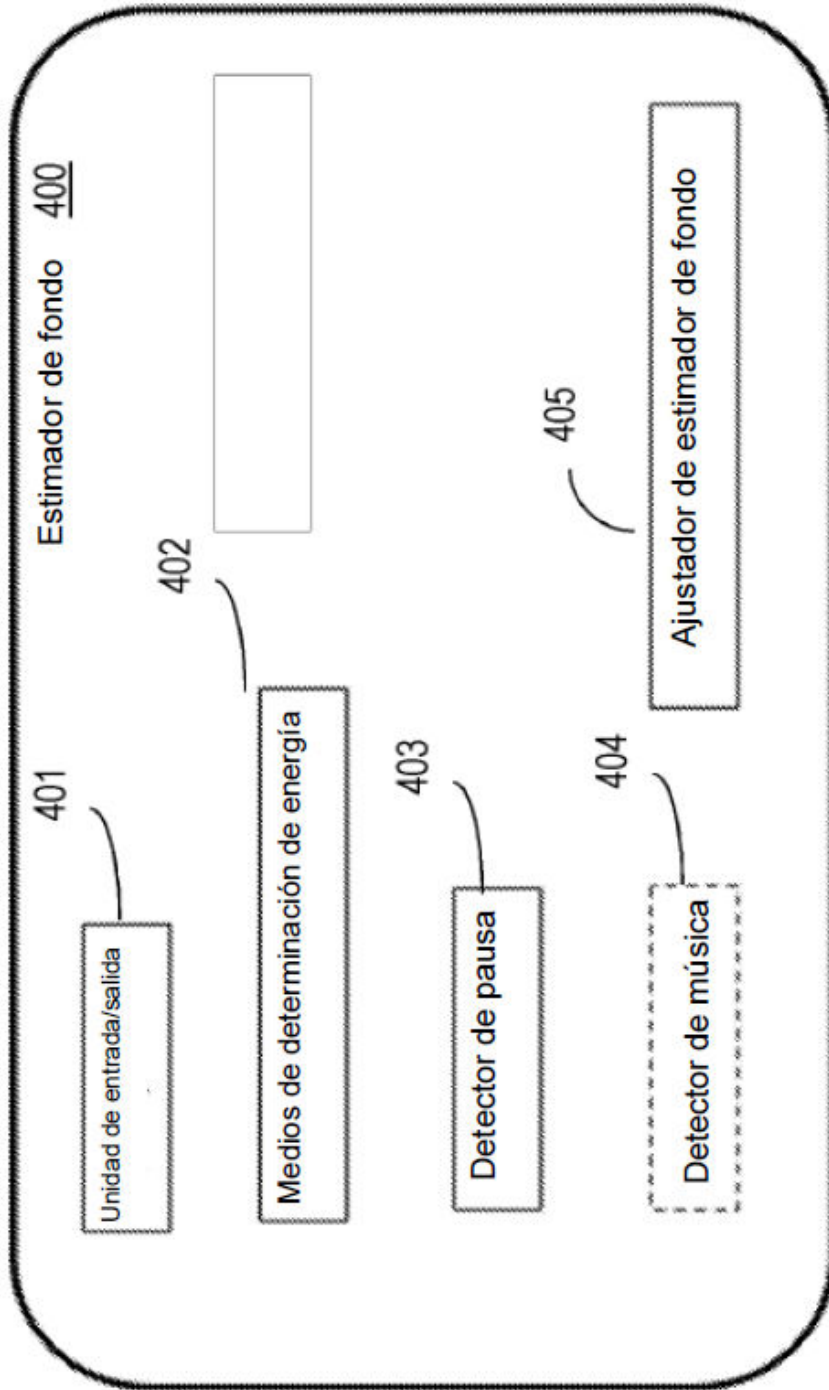


Figura A4

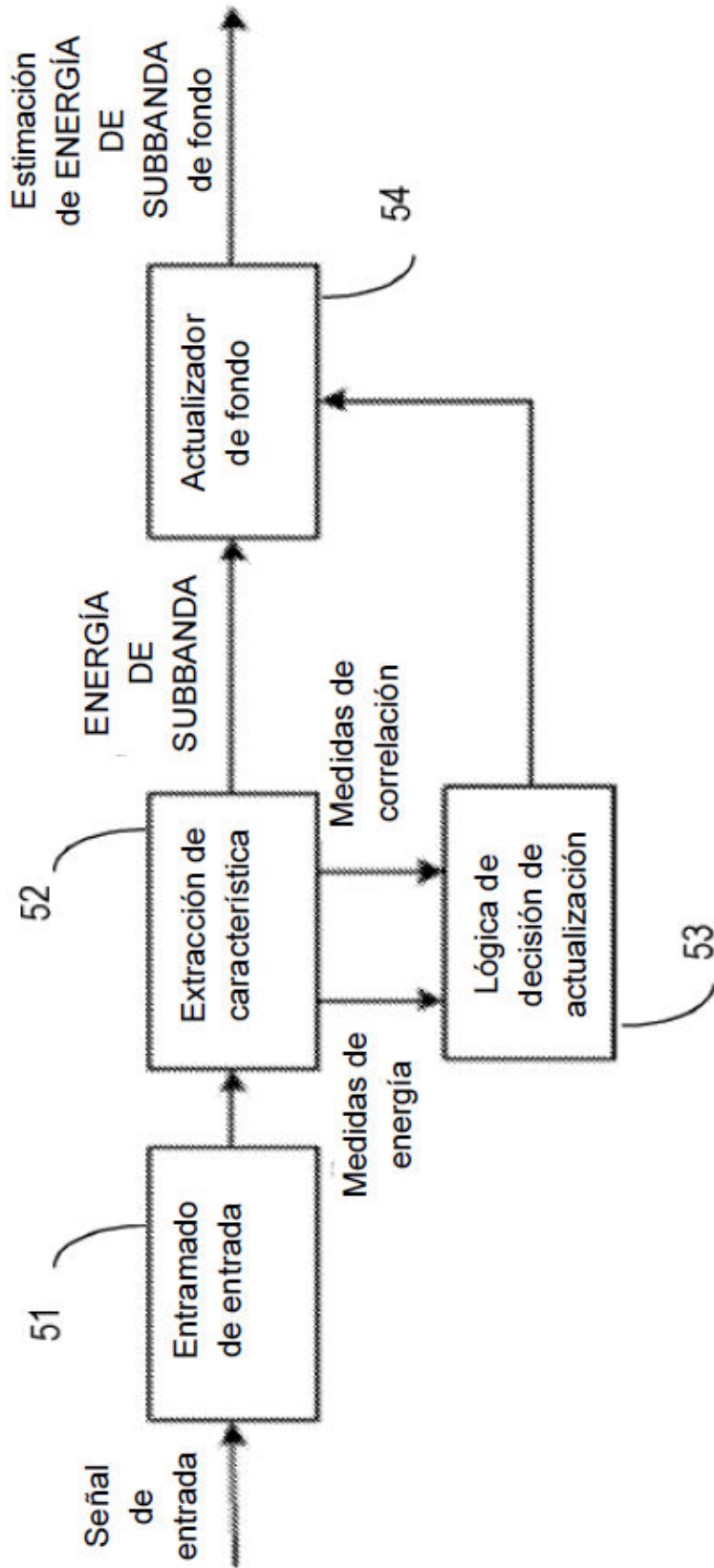


Figura A5

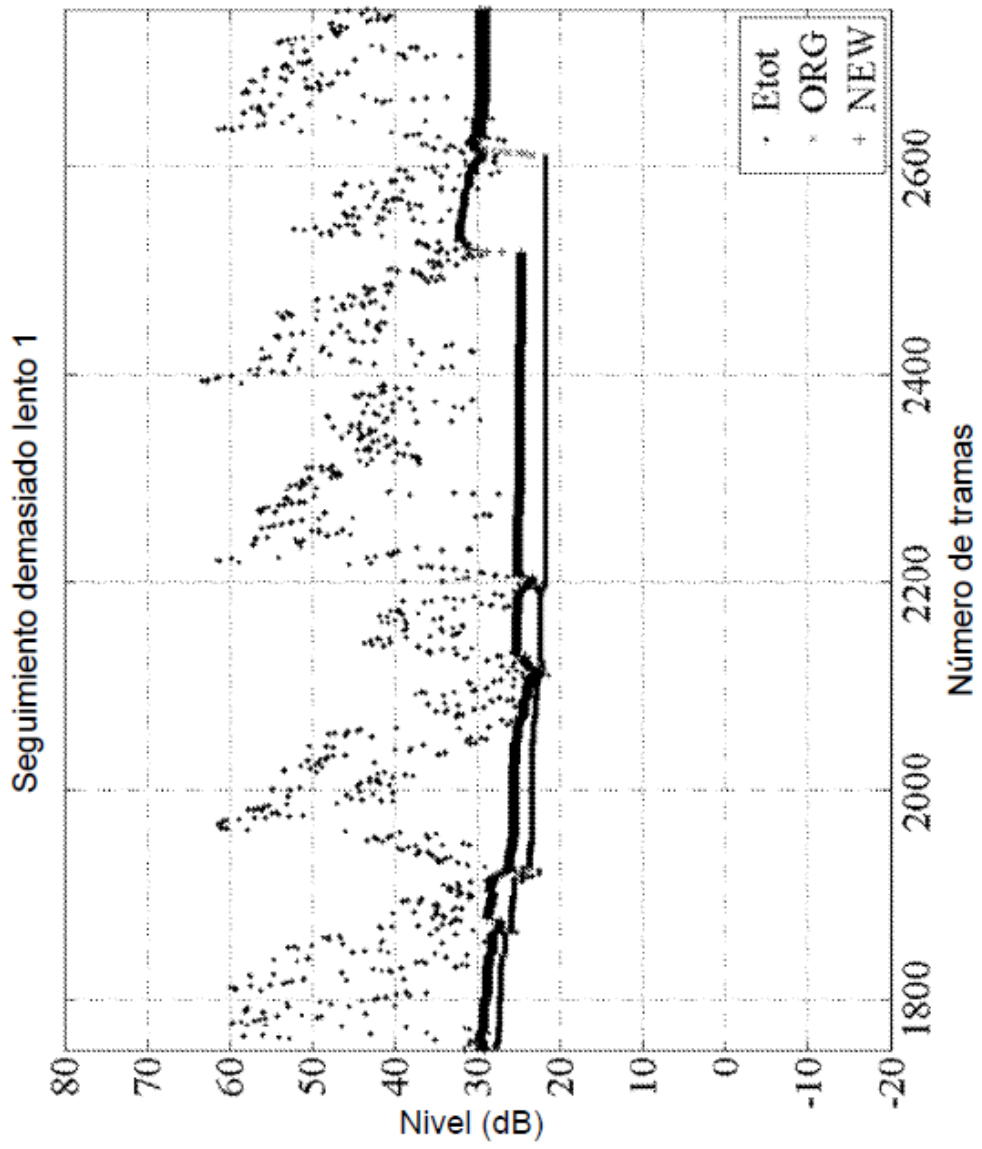


Figura A6

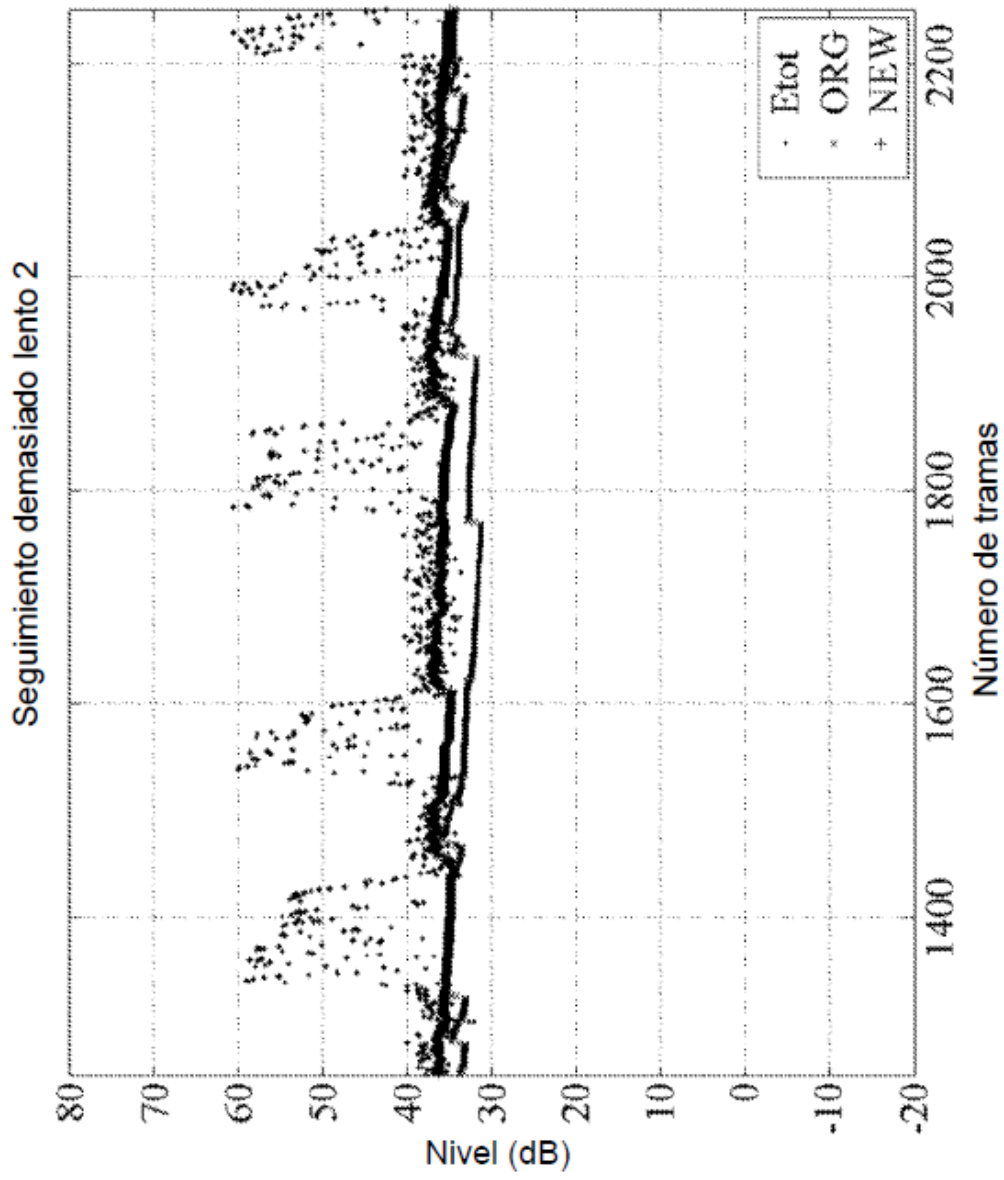


Figura A7

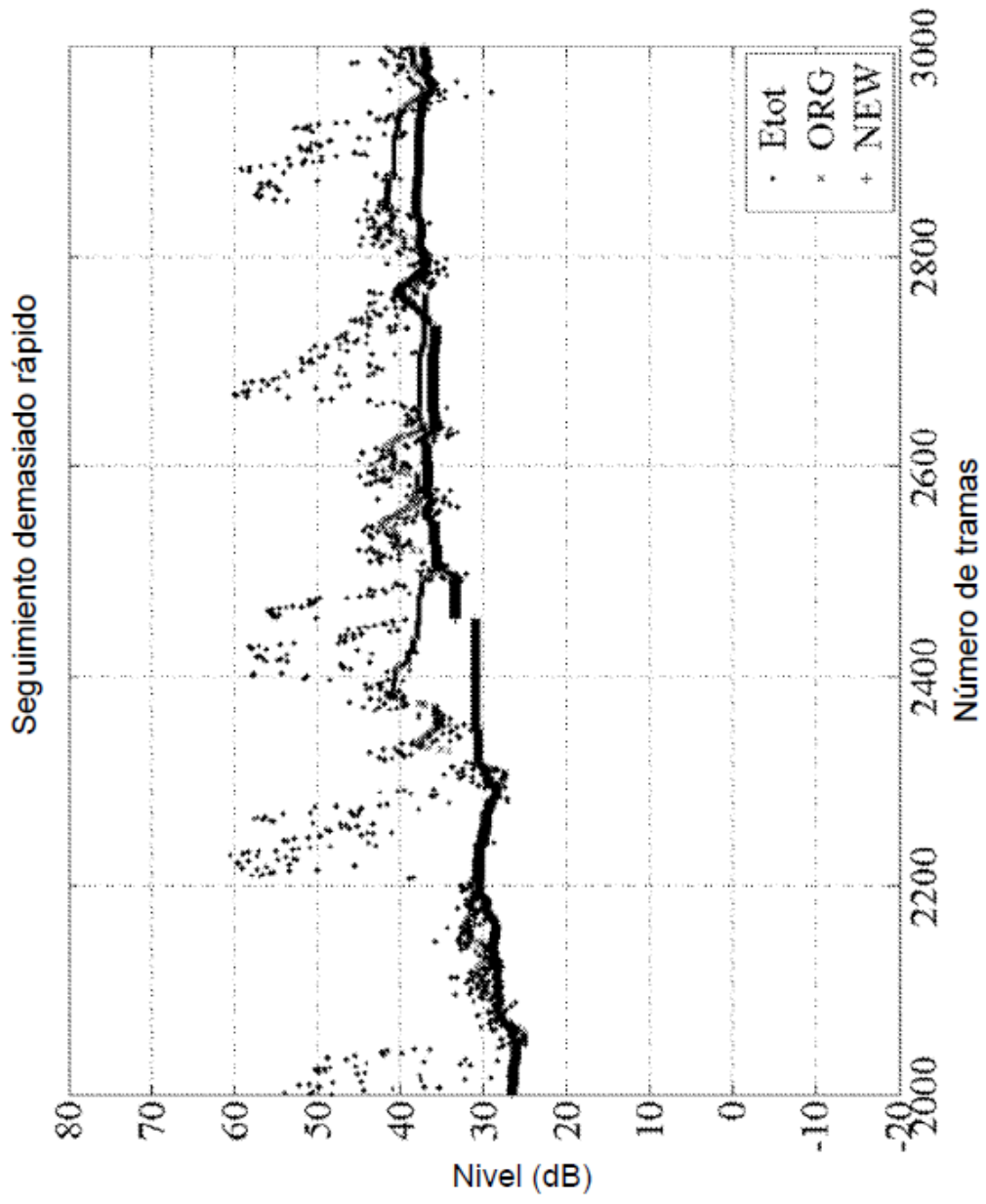


Figura A8

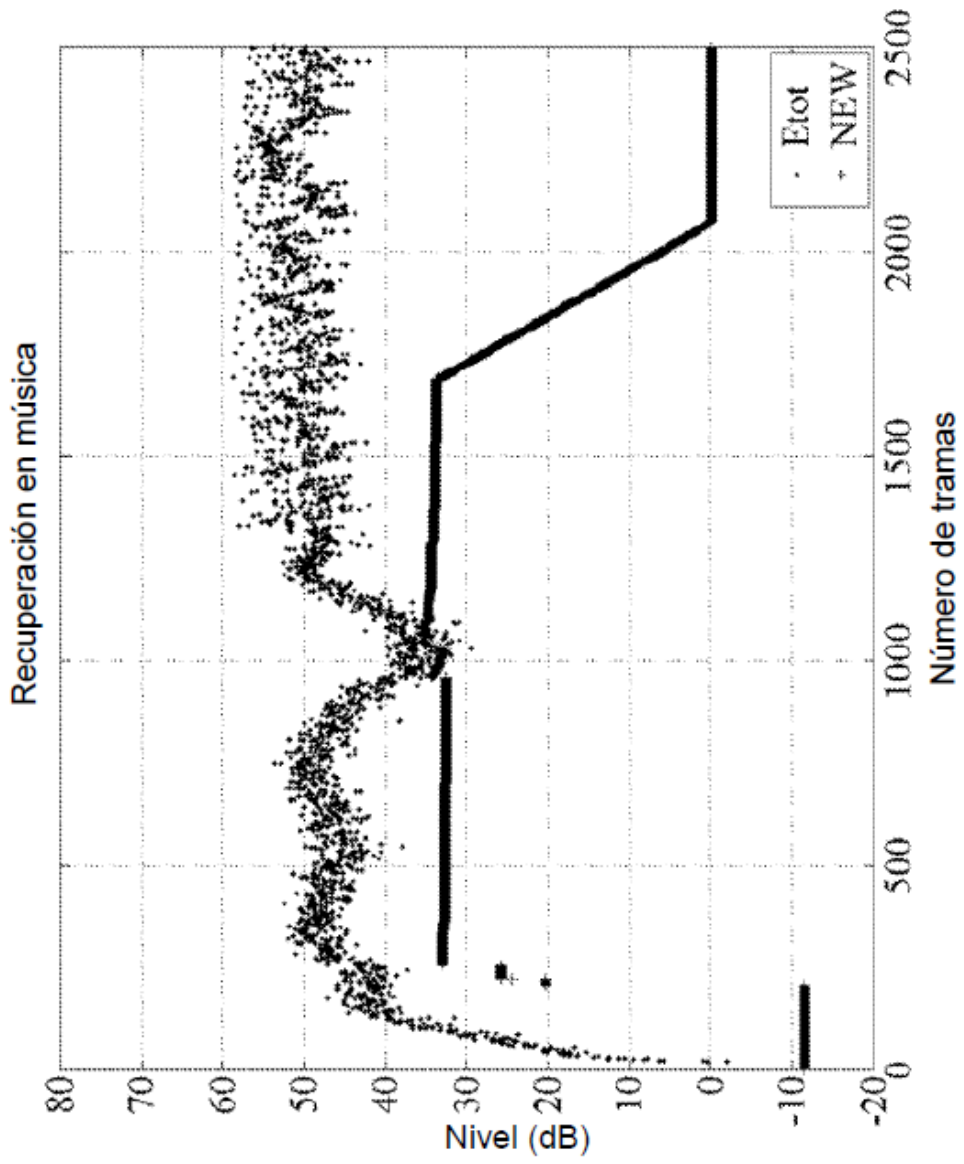


Figura A9