

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 819 032**

51 Int. Cl.:

G10L 25/78 (2013.01)

G10L 25/06 (2013.01)

G10L 25/21 (2013.01)

G10L 21/0216 (2013.01)

G10L 19/24 (2013.01)

G10L 25/84 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.12.2014** **E 18195924 (8)**

97 Fecha y número de publicación de la concesión europea: **24.06.2020** **EP 3438979**

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

30 Prioridad:

19.12.2013 US 201361918258 P

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

14.04.2021

73 Titular/es:

TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)
(100.0%)

164 83 Stockholm, SE

72 Inventor/es:

SEHLSTEDT, MARTIN

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 819 032 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

Campo técnico

Las realizaciones de la presente invención se refieren a codificación de audio y, en particular, a estimación de ruido de fondo para soportar una decisión de actividad de sonido.

Antecedentes

En sistemas de comunicación que utilizan transmisión discontinua (DTX) es importante encontrar un equilibrio entre eficiencia y no reducción de la calidad. En tales sistemas, se usa un detector de actividad para indicar señales activas, por ejemplo, habla o música, que han de ser codificadas activamente, y segmentos con señales de fondo que se pueden sustituir con ruido de confort generado en el lado receptor. Si el detector de actividad es demasiado eficiente en la detección de no actividad, introducirá recorte en la señal activa, lo que luego se percibe como una degradación de la calidad subjetiva cuando el segmento activo recortado se sustituye por ruido de confort. Al mismo tiempo, la eficiencia de la DTX se reduce si el detector de actividad no es lo bastante eficiente y clasifica segmentos de ruido de fondo como activos y luego codifica activamente el ruido de fondo en lugar de entrar en 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 general de un detector de actividad de sonido, SAD o detector de actividad de voz, VAD, generalizado 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, por ejemplo, de 5-30 ms, dependiendo de la implementación, y se produce como salida una decisión de actividad por trama.

Una decisión primaria, "prim", se toma por el detector primario ilustrado en la Figura 1. La decisión primaria es básicamente sólo una comparación de los rasgos de una trama actual con los rasgos de fondo, que se estiman a partir de tramas de entrada previas. Una diferencia entre los rasgos de la trama actual y los rasgos de fondo que es más grande que un umbral causa una decisión primaria activa. El bloque de adición de retención se usa para extender la decisión primaria basada en decisiones primarias pasadas para formar la decisión final, "marca". La razón para usar la retención es principalmente reducir/eliminar el riesgo de recorte de actividad de parte media y trasera de ráfaga de actividad. Como se indica en la figura, un controlador de operación puede ajustar el umbral o los umbrales para el detector primario y la duración de la adición de retención según las características de la señal de entrada.

El bloque del estimador de fondo se usa para estimar el ruido de fondo en la señal de entrada. También se puede hacer referencia al ruido de fondo como "el fondo" o "el rasgo de fondo" en la presente memoria.

La estimación del rasgo de fondo se puede hacer según dos principios básicamente diferentes, o bien usando la decisión primaria, es decir, con la decisión o la realimentación de la métrica de decisión, que se indica mediante una línea de puntos y rayas en la Figura 1, o bien usando algunas otras características de la señal de entrada, es decir, sin realimentación de la decisión. También es posible usar combinaciones de las dos estrategias.

Un ejemplo de un códec que usa realimentación de decisión para la estimación de fondo es AMR-NB (Banda Estrecha de Múltiples Tasas Adaptativa) y ejemplos de códec donde no se usa realimentación de decisión son EVRC (CODEC de Tasa Variable Mejorada) y G.718.

Hay una serie de diferentes rasgos o características de señal que se pueden usar, pero un rasgo común utilizado en los VAD son las características de frecuencia de la señal de entrada. Un tipo de características de frecuencia comúnmente usado es la energía de trama de subbanda, debido a su baja complejidad y operación fiable en SNR baja. 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, uno de los rasgos del 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 hacer el seguimiento del ruido de fondo, la actualización de la estimación del ruido de fondo real se puede hacer de al menos tres formas diferentes. Una forma es usar un proceso de Regresión Automática, AR, por compartimento de frecuencia para manejar la actualización. Ejemplos de tales códec 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 un códec que usa esta técnica para actualizar la estimación de fondo para la función de VAD. Obsérvese que EVRC usa diferentes estimaciones de fondo para VAD y supresión de ruido. Se debería observar que un VAD se puede usar en otros contextos distintos de DTX. Por ejemplo, en códec de tasa variable, tales como EVRC, el VAD se puede usar como parte de una función de determinación de tasa.

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

- 5 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. No obstante, es muy difícil separar las señales activas y no activas en casos de SNR baja, y en particular cuando el fondo no es estacionario o incluso similar a la señal activa en sus características.

Compendio

- 10 Sería deseable tomar decisiones más adecuadas de si una señal de audio comprende o no habla activa o música. En la presente memoria se proporciona un método mejorado para generar una estimación de ruido de fondo, que permite que un detector de actividad de sonido tome decisiones más adecuadas.

- 15 Según un primer aspecto, se proporciona un método para la estimación de ruido de fondo en un segmento de señal de audio según la reivindicación 1. El método comprende reducir una estimación de ruido de subbanda actual cuando se determina que el segmento de señal de audio comprende música y la estimación de subbanda actual excede un valor mínimo. Esto se ha de realizar cuando un nivel de energía de un segmento de señal de audio es menor que un umbral más alto que un nivel de energía mínimo a largo plazo, I_{t_min} , pero no se detecta ninguna pausa en el segmento de señal de audio.

Según un segundo aspecto, se proporciona un aparato para la estimación de ruido de fondo en un segmento de señal de audio según la reivindicación 4.

- 20 Según un tercer aspecto, se proporciona un detector de actividad de sonido, que comprende un aparato según el segundo aspecto.

Según un cuarto aspecto, se proporciona un códec, que comprende un aparato según el segundo aspecto.

Según un quinto aspecto, se proporciona un dispositivo de comunicación, que comprende un aparato según el segundo aspecto.

- 25 Según un sexto aspecto, se proporciona un nodo de red, que comprende un aparato según el segundo aspecto.

Según un séptimo aspecto, se proporciona un programa de ordenador que comprende instrucciones que, cuando se ejecutan en al menos un procesador, hacen que al menos un procesador lleve a cabo el método según el primer aspecto.

- 30 Según un octavo aspecto, se proporciona un portador, que contiene un programa de ordenador según el séptimo aspecto.

Breve descripción de los dibujos

- 35 Los precedentes y otros objetos, rasgos y ventajas de la tecnología descrita en la presente memoria serán evidentes a partir de la siguiente descripción más particular de las realizaciones, como se ilustra en los dibujos que se acompañan. Los dibujos no están necesariamente a escala, siendo el énfasis colocado, en su lugar, en ilustrar los principios de la tecnología descrita en la presente memoria.

La Figura 1 es un diagrama de bloques que ilustra un detector de actividad y una lógica de determinación de retención.

La Figura 2 es un diagrama de flujo que ilustra una lógica de decisión de actualización de fondo, según una realización ejemplificativa.

- 40 Las Figuras 4 y 5 muestran un estimador de fondo según diferentes realizaciones ejemplificativas.

La Figura 5 es un diagrama de bloques que muestra un estimador de fondo de energía de subbanda.

Las Figuras 6-9 son diagramas que muestran cómo las realizaciones permiten un mejor seguimiento del ruido de fondo en señales de audio.

Descripción detallada

- 45 La solución descrita en la presente memoria se relaciona con estimación de ruido de fondo en señales de audio. En el detector de actividad generalizada ilustrado en la Figura 1, la función de estimación de ruido de fondo se realiza por el bloque denotado "estimador de fondo". Algunas realizaciones de la solución descrita en la presente memoria se pueden ver en relación con las soluciones descritas previamente en los documentos WO2011/049514 y WO2011/049515.

La solución descrita en la presente memoria se comparará con implementaciones de estas aplicaciones descritas previamente. Aún cuando las soluciones descritas en los documentos WO2011/049514 y WO2011/049515, así como la solución presentada en el documento US 6 157 670, son buenas soluciones, la solución presentada en la presente memoria todavía tiene ventajas en relación con estas soluciones. Por ejemplo, la solución presentada en la presente memoria tiene una implementación incluso menos compleja y 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 hacer el seguimiento de las características del fondo – en particular cuando se trata de fondos no estacionarios. Con un mejor seguimiento, es posible hacer que el VAD sea más eficiente sin aumentar el riesgo de recorte de habla.

Un problema con los métodos actuales de estimación de ruido es que para lograr un buen seguimiento del ruido de fondo en SNR baja, se necesita un detector de pausa fiable. Para la entrada solamente de habla, es posible utilizar la tasa silábica o el hecho de que una persona no puede hablar todo el tiempo para encontrar pausas en el habla. Tales soluciones podrían implicar que después de un tiempo suficiente de no hacer actualizaciones de fondo, los requisitos para la detección de pausa se “relajan”, de manera que sea más probable detectar una pausa en el habla. Esto permite responder a cambios bruscos en las características o el nivel de ruido. Algunos ejemplos de tales lógicas de recuperación de ruido son: 1) Como las expresiones del habla contienen segmentos con alta correlación después de un número de tramas suficiente sin correlación, es normalmente seguro suponer que hay una pausa en el habla. 2) Cuando la Relación Señal a Ruido, $SNR > 0$, la energía del habla es más alta que el ruido de fondo, así que, si la energía de la trama está cerca de la energía mínima durante un tiempo más largo, por ejemplo, 1-5 segundos, también es seguro suponer que uno está en una pausa de habla. Si bien las técnicas previas funcionan bien con entrada solamente de habla, no son suficientes cuando la música se considera una entrada activa. En música puede haber largos segmentos con baja correlación que aún son música. Además, la dinámica de la energía en la música también puede desencadenar la detección de una pausa falsa, lo que puede dar como resultado actualizaciones no deseadas y erróneas de la estimación de ruido de fondo.

Idealmente, una función inversa de un detector de actividad, o lo que se llamaría un “detector de aparición de pausa”, se necesitaría para controlar la estimación de ruido. Esto aseguraría que la actualización de las características de ruido de fondo se haga solamente cuando no haya ninguna señal activa en la trama actual. No obstante, como se ha indicado anteriormente, no es una tarea fácil determinar si un segmento de señal de audio comprende o no una señal activa.

Tradicionalmente, cuando se conocía que la señal activa era una señal de habla, 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. No obstante, en los códec modernos, también es común referirse al detector de actividad como un Detector de Actividad de Sonido (SAD) cuando también se ha de detectar música como señal activa.

El estimador de fondo ilustrado en la Figura 1 utiliza realimentación del detector primario y/o el bloque de retención para localizar segmentos de señal de audio inactivos. Cuando se desarrolla la tecnología descrita en la presente memoria, ha sido un deseo eliminar, o al menos reducir la dependencia de tal realimentación. Para la estimación de fondo descrita en la presente memoria, se ha identificado, por lo tanto, por los inventores como importante ser capaz de encontrar rasgos fiables para identificar las características de las señales de fondo cuando solamente está disponible una señal de entrada con una mezcla desconocida de señal activa y de fondo. Los inventores se han dado cuenta además de que no se puede suponer que la señal de entrada comience con un segmento de ruido, o incluso que la señal de entrada tenga habla mezclada con ruido, ya que puede ser que la señal activa sea música.

Una contribución de las realizaciones en la presente memoria a la técnica anterior es la selección de rasgos para usar, e incluso más, cómo combinar los rasgos seleccionados para lograr una lógica de estimación de ruido que funcione de manera fiable para diferentes tipos de entrada.

Como hemos visto anteriormente, hay varios rasgos que funcionan bien para condiciones particulares. La dificultad es combinarlos de una forma que beneficie la estimación de ruido y el seguimiento de fondo. En particular, si uno es evitar suposiciones acerca de las condiciones iniciales, pero se basa solamente en las características de la señal hasta el momento y ser capaces de manejar condiciones donde tanto el habla como la música han de ser considerados entradas activas.

La Figura 2 es un diagrama de flujo que ilustra una realización ejemplificativa de un método para estimación de ruido de fondo según la tecnología propuesta en la presente memoria. 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 estimación de nuevo ruido de subbanda, independientemente de si la trama es de contenido de fondo o activo, si el nuevo valor es menor que el actual, se usa directamente como que lo más probable es que sea de una trama de fondo. La siguiente lógica de estimación de ruido es un segundo paso donde se decide si se puede aumentar la estimación de ruido de subbanda y, en caso afirmativo, cuánto, el aumento se basa en la estimación de nuevo ruido de subbanda posible calculada previamente.

Básicamente, esta lógica forma la decisión de que la trama actual es una trama de fondo y, si no es seguro, puede permitir un aumento menor en comparación con lo que se estimó originalmente.

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

- reducir 206 una estimación de ruido de fondo actual cuando se determina que el segmento de señal de audio 203:2 comprende música y la estimación de ruido de fondo actual excede un valor mínimo 205:1, denotado "T" en la Figura 2, y se ejemplifica, además, por ejemplo, como $2 \cdot E_MIN$ en el código a continuación.

Realizando lo anterior, y proporcionando la estimación de ruido de fondo a un SAD, se permite que el SAD realice una detección de actividad de sonido más adecuada. Además, se permite la recuperación de las actualizaciones erróneas de estimación de ruido de fondo.

Se puede hacer referencia alternativamente al nivel de energía del segmento de señal de audio usado en el método descrito anteriormente como, por ejemplo, la energía de la trama actual, $Etot$, o como la energía del segmento de señal, o trama, que se puede calcular sumando las energías de subbandas para el segmento de señal actual.

El otro rasgo de energía usada en el método anterior, es decir, el nivel de energía mínimo 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 se podría denotar alternativamente, por ejemplo, $Etot_lp$. Una forma básica de derivar It_min sería usar el valor mínimo del historial de energía de la trama actual sobre algún número de tramas pasadas. Si el valor calculado como: "energía de la trama actual - estimación mínima a largo plazo" está por debajo de un valor de umbral, denotado, por ejemplo, $THR1$, se dice en la presente memoria que la energía de la trama actual está cercana a la energía mínima a largo plazo, o que está cerca de la energía mínima a largo plazo. Es decir, cuando $(Etot - It_min) < THR1$, se puede determinar 202 que la energía de la trama actual, $Etot$, está cerca de la energía mínima a largo plazo It_min . El caso cuando $(Etot - It_min) = THR1$ se puede referir 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 la trama actual no está cerca de It_min , mientras que 202:2 indica la decisión de que la energía de la trama actual está cerca de It_min . Otra numeración en la Figura 2 en forma XXX:Y indica las decisiones correspondientes. El rasgo It_min se describirá aún más a continuación.

El valor mínimo, que ha de exceder la estimación de ruido de fondo actual, con el fin de ser reducido, se puede suponer que es cero o un valor positivo pequeño. Por ejemplo, como se ejemplificará en el código a continuación, una energía total actual de la estimación de fondo, que se puede denotar "totalNoise" y se puede determinar, por ejemplo, como $10 \cdot \log_{10} \Sigma backr[i]$, se puede requerir que exceda un valor mínimo de cero con el fin de que la reducción sea cuestionable. 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 , con el fin de que sea realizada la reducción. En el ejemplo de código a continuación, E_MIN es un valor positivo pequeño.

Se debería señalar que, según una realización preferida de la solución sugerida en la presente memoria, la decisión de si el nivel de energía del segmento de señal de audio es mayor que un umbral más alto que It_min se basa solamente en información derivada de la señal de audio de entrada, es decir, no se basa en la realimentación de una decisión del detector de actividad de sonido.

La determinación 204 de si una trama actual comprende o no una pausa se puede realizar de diferentes formas en base a uno o más criterios. También se puede hacer referencia a un criterio de pausa como detector de pausa. Se podría aplicar un detector de pausa único, o una combinación de detectores de pausa diferentes. Con una combinación de detectores de pausa, cada uno se puede usar para detectar pausas en diferentes condiciones. Un indicador de que una trama actual puede comprender una pausa, o inactividad, es que un rasgo de correlación para la trama es bajo, y que un número de tramas precedentes también han tenido rasgos de correlación baja. 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 según la entrada actual, como se ilustra en la Figura 2. Se puede considerar que una pausa se detecte cuando, además de que el nivel de energía del segmento de señal de audio sea menor que un umbral más alto que It_min : se haya determinado que un número predefinido de segmentos de señal de audio precedentes consecutivos no comprendan una señal activa y/o una dinámica de la señal de audio exceda un umbral. Esto también se ilustra en el ejemplo de código más adelante.

La reducción 206 de la estimación de ruido de fondo permite el manejo de situaciones donde la estimación de ruido de fondo ha llegado a ser "demasiado alta", es decir, en relación con un ruido de fondo verdadero. Esto también se podría expresar, 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 parte del SAD, donde se determina que el segmento de la señal actual está inactivo, aún cuando comprenda habla activa o música. Una razón para que la estimación de ruido de fondo llegue a ser 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 que la estimación de ruido se aumente. El método descrito permite que tal estimación de ruido de fondo actualizada erróneamente se ajuste, por ejemplo, cuando una trama siguiente de la señal de entrada se determina que comprende música. Este ajuste se hace 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 más alta que la estimación de ruido de fondo actual, por ejemplo, en una subbanda. Se debería observar que la lógica descrita anteriormente para la estimación de ruido de fondo se usa para controlar el aumento de la energía de subbanda de fondo. Siempre se permite bajar la energía de subbanda cuando la energía de subbanda de la trama actual es menor que la estimación de ruido de fondo. Esta función no se muestra explícitamente en la Figura 2. Tal disminución normalmente tiene una configuración fija para el tamaño del paso. No obstante, la estimación de ruido de fondo solamente se debería permitir que sea aumentada en asociación con la lógica de decisión según el método descrito anteriormente. Cuando se detecta una pausa, los rasgos de energía y correlación también se pueden usar para decidir lo grande que debería ser el tamaño del paso de ajuste para el aumento de la estimación de fondo antes de que se haga la actualización del ruido de fondo real.

Como se mencionó anteriormente, algunos segmentos de música pueden ser difíciles de separar del ruido de fondo, debido a que son muy ruidosos. De este modo, la lógica de actualización de ruido puede permitir accidentalmente un aumento de las estimaciones de energía de subbanda, aún cuando la señal de entrada fuera una señal activa. Esto puede causar problemas en la medida que la estimación de ruido puede llegar a 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 se podrían reducir solamente cuando una energía de subbanda de entrada fuese por debajo de una estimación de ruido actual. No obstante, 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 la presente memoria, tal recuperación se puede hacer mediante una reducción forzada de estimación de ruido cuando la señal de entrada vuelve a características de tipo música. Es decir, cuando la energía y la lógica de pausa descritas anteriormente evita, 202:1, 204:1, que la estimación de ruido sea aumentada, se prueba 203 si se sospecha que la entrada sea música y, si es así, 203:2, las energías de subbanda se reducen en una pequeña cantidad cada trama hasta que las estimaciones de ruido alcancen el nivel más bajo 205:2.

La solución descrita en la presente memoria también se refiere a un estimador de fondo implementado en hardware y/o software. Un estimador de fondo según una realización ejemplificativa se ilustra esquemáticamente en la Figura 3. Se puede suponer que el estimador de fondo 500 comprende una unidad de entrada 502 para recibir medidas de energía y posiblemente medidas de correlación; y una unidad de salida 505 para proporcionar una estimación de ruido de fondo actualizada. El estimador de fondo 500 comprende además un procesador 503 y una memoria 504, dicha memoria que contiene instrucciones 507 ejecutables por dicho procesador 504. La ejecución de las instrucciones 507 hace que dicho estimador de fondo 500 operativo realice al menos una realización del método para estimación de ruido de fondo descrito anteriormente. En otras palabras, la ejecución de las instrucciones 507 por los medios de procesamiento 503 hace el estimador de fondo 500 operativo para: cuando un nivel de energía de un segmento de señal de audio es mayor que un umbral más alto que un nivel de energía mínimo a largo plazo, It_min , que se determina sobre una pluralidad de segmentos de señal de audio precedentes, o, cuando el nivel de energía del segmento de señal de audio es menor que un umbral más alto que It_min , pero no se detecta ninguna pausa en el segmento de señal de audio:

- reducir una estimación de ruido de fondo actual cuando se determina que el segmento de señal de audio comprende música y la estimación de ruido de fondo actual excede un valor mínimo; Las instrucciones 507 se pueden almacenar en forma de código de ordenador, por ejemplo, como el que se presenta más adelante en esta descripción. Las instrucciones o el programa de ordenador pueden ser transportados por un portador antes de ser almacenados y/o ejecutados por el estimador de fondo. Tal portador puede ser, por ejemplo, una señal electrónica, una señal óptica, una señal de radio o un medio de almacenamiento legible por ordenador.

La Figura 4 muestra una implementación alternativa de un estimador de fondo. El estimador de fondo 400 comprende una unidad de entrada/salida 401, medios de determinación de energía 402 configurados para determinar si la energía de la trama actual está cerca o no de una estimación de energía mínima a largo plazo, un detector de pausa 403, configurado para determinar si una trama actual comprende una pausa, un detector de música, configurado para determinar si una trama actual comprende o no música. El estimador de fondo 400 comprende además un ajustador de estimador de fondo 405, configurado para: cuando un nivel de energía de un segmento de señal de audio es mayor que un umbral más alto que un nivel de energía mínimo a largo plazo, It_min , o cuando el nivel de energía del segmento de señal de audio es menor que un umbral más alto que It_min , pero no se detecta ninguna pausa en el segmento de señal de audio: - reducir una estimación de ruido de fondo actual cuando se determina que el segmento de señal de audio comprende música y la estimación de ruido de fondo actual excede un valor mínimo. El ajustador de estimador de fondo 405 también se puede configurar para realizar, por ejemplo, un ajuste regular, como el ilustrado como 208 en la Figura 3.

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 un decodificador, en donde el codificador y/o el decodificador se pueden implementar en un dispositivo de usuario, tal 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 según una realización ejemplificativa. Un bloque de tramas de entrada 51 divide primero la señal de entrada en tramas de longitud adecuada, por ejemplo, 5-30 ms. Para cada trama, un extractor de rasgos 52 calcula al menos los siguientes rasgos de la entrada: 1) El extractor de rasgos analiza la trama en el dominio de frecuencia y se calcula la energía para un conjunto de subbandas. Las subbandas son las mismas subbandas que han de ser usadas para la estimación de fondo. 2) El extractor de rasgos analiza además la trama en el dominio de tiempo y calcula una correlación denotada, por ejemplo, cor_est y/o lt_cor_est , que se usa en la determinación de si la trama comprende o no contenido activo. 3) El extractor de rasgos utiliza además la energía total de la trama actual, por ejemplo, denotada $Etot$, para actualizar los rasgos del historial de energía de las tramas de entrada actual y anteriores, tal como la energía mínima a largo plazo, lt_min . Los rasgos de correlación y energía se alimentan entonces al bloque de Lógica de Decisión de Actualización 53.

Aquí, una lógica de decisión según la solución descrita en la presente memoria se implementa en el bloque de Lógica de Decisión de Actualización 53, donde los rasgos de correlación y energía se usan para formar decisiones sobre si la energía de la trama actual está cerca o no de una energía mínima a largo plazo; sobre si la trama actual es o no parte de una pausa (no una señal activa); y si la trama actual es o no parte de una música. La solución según las realizaciones descritas en la presente memoria implica cómo estos rasgos y decisiones se usan para actualizar la estimación de ruido de fondo de una forma robusta.

A continuación, se describirán algunos detalles de implementación de las realizaciones de la solución descrita en la presente memoria. Los detalles de implementación a continuación se toman a partir de una realización en un codificador basado en G.718. Esta realización usa algunos de los rasgos descritos en los documentos WO2011/049514 y WO2011/049515, de los cuales se adjuntan partes a esta descripción.

Los siguientes rasgos se definen en el documento G.718 modificado descrito en el documento WO2011/09514:

Etot;	La energía total para la trama de entrada actual
Etot_l	Hace el seguimiento de 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 estimación de nuevo fondo potencial calculada previamente
aEn;	Un detector de fondo que usa múltiples rasgos (un contador)
harm_cor_cnt	Cuenta las tramas desde la última trama con evento de correlación o armónico
act_pred	Una predicción de actividad de los rasgos de trama de entrada solamente
cor[i]	Vector con estimaciones de correlación para, $i = 0$ fin de la trama actual, $i = 1$ inicio de la
	trama actual, $i = 2$ fin de la trama previa

Los siguientes rasgos se definen en el documento G.718 modificado descrito en el documento WO2011/09515

Etot_h	Hace el seguimiento de la envoltura de energía máxima
sign_dyn_lp;	Una dinámica de señal de entrada suavizada.

También se definió el rasgo Etot_v_h en el documento WO2011/049514, pero en esta realización se ha modificado y ahora se implementa de la siguiente manera:

```

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

```

Etot_v mide la variación de energía absoluta entre tramas, es decir, el valor absoluto de la variación de energía instantánea entre tramas. En el ejemplo anterior, se determina que la variación de energía entre dos tramas es “baja” cuando la diferencia entre la energía de la última trama y la actual es menor que 7 unidades. Esto se utiliza como indicador de que la trama actual (y la trama previa) puede ser parte de una pausa, es decir, que comprende solamente ruido de fondo. No obstante, tal varianza baja se podría encontrar alternativamente, por ejemplo, en medio de una ráfaga de habla. La variable Etot_last es el nivel de energía de la trama previa.

Los pasos anteriores descritos en el código se pueden realizar como parte de los pasos “calcular/actualizar correlación y energía” en el diagrama de flujo de la Figura 2, es decir, como parte de las acciones 201. En la implementación del documento WO2011/049514, se usó una marca VAD para determinar si el segmento de la señal de audio actual comprendía o no ruido de fondo. Los inventores se han dado cuenta de que la dependencia de la información de realimentación puede ser problemática. En la solución descrita en la presente memoria, la decisión de si actualizar o no la estimación de ruido de fondo no es dependiente de una decisión de VAD (o SAD).

Además, en la solución descrita en la presente memoria, los siguientes rasgos, que no son parte de la implementación del documento WO2011/049514 se pueden calcular/actualizar como parte de los mismos pasos, es decir, los pasos de calcular/actualizar correlación y energía ilustrados en la Figura 2. Estos rasgos se usan también en la lógica de decisión de si actualizar o no la estimación de fondo.

Con el fin de lograr una estimación de ruido de fondo más adecuada, se definen a continuación una serie de rasgos. Por ejemplo, se definen los nuevos rasgos relacionados con la correlación cor_est e lt_cor_est. El rasgo 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 a largo plazo suavizada de la correlación.

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

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

Como se ha definido 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 previa.

Además, se calcula un nuevo rasgo, lt_tn_track, que da una estimación a largo plazo de lo a menudo que las estimaciones de fondo están cerca de la energía de la trama actual. Cuando la energía de la trama actual está lo suficientemente cerca de la estimación de fondo actual esto se registra mediante 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->lt_tn_track = 0,03f* (Etot - st->totalNoise < 10) + 0.97f*st->lt_tn_track

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

Además, la distancia entre la estimación de fondo actual, Etot, y la energía de la trama actual, totalNoise, se usa para determinar un rasgo, lt_tn_dist, que da una estimación a largo plazo de esta distancia. Un rasgo similar, lt_Ellp_dist, se crea para la distancia entre la energía mínima a largo plazo Etot_l_lp y la energía de la trama actual, Etot.

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

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

El rasgo `harm_cor_cnt`, introducido anteriormente, se usa para contar el número de tramas desde que la última trama que tiene un evento de correlación o 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 la trama actual es más probablemente una trama activa, ya que muestra un evento de correlación o armónico. Esto se usa para formar una estimación suavizada a largo plazo, `lt_haco_ev`, de lo a menudo que ocurren tales eventos. En este caso, la actualización no es simétrica, es decir, se usan diferentes constantes de tiempo si la estimación se aumenta o disminuye como se puede ver a continuación.

```

if (st->harm_cor_cnt == 0)                /*cuando está probablemente activo*/
{
    st->lt_haco_ev = 0,03f + 0,97f*st->lt_haco_ev; /*aumentar estimación a largo plazo*/
}
else
{
    st->lt_haco_ev = 0,99f*st->lt_haco_ev;      /*disminuir estimación a largo plazo*/
}

```

Un valor bajo del rasgo `lt_tn_track`, introducido anteriormente, indica que la energía de la trama de entrada no ha estado cerca de la energía de fondo para algunas tramas. Esto es debido a que `lt_tn_track` se disminuye para cada trama donde la energía de la trama actual no está cerca de la estimación de energía de fondo. `lt_tn_track` se aumenta solamente cuando la energía de la trama actual está cerca de la estimación de energía de fondo como se ha mostrado anteriormente. Para obtener una mejor estimación de cuánto de este “no seguimiento”, es decir, la energía de la trama que 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:

```

if (st->lt_tn_track < 0,05f)              /*cuando lt_tn_track es bajo*/
{
    st->low_tn_track_cnt++; /*añadir 1 al contador*/
}
else
{
    st->low_tn_track_cnt = 0; /*reiniciar contador*/
}

```

En el ejemplo anterior, “bajo” se define como por debajo del valor 0,05. Esto se debería ver como un valor ejemplificativo, que se podría 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 denotada detección de fondo. En otras realizaciones e implementaciones, también se pueden añadir otros criterios para la detección de pausa. La decisión de música real se forma en el código usando rasgos de correlación y energía.

1: `bg_bgd = Etot < Etot_l_lp + 0,6f * st-> Etot_v_h;`

`bg_bgd` llegará a ser “1” o “verdadero” cuando `Etot` esté cerca de la estimación de ruido de fondo.

`bg_bgd` sirve como máscara para otros detectores de fondo. Es decir, si `bg_bgd` no es “verdadero”, no necesitan ser evaluados los detectores de fondo 2 y 3 a continuación.

`Etot_v_h` es una estimación de la varianza de ruido, que se podría denotar alternativamente N_{var} .

`Etot_v_h` se deriva de la energía total de entrada (en el dominio log) usando `Etot_v` que mide la variación de energía absoluta entre tramas. Obsérvese que el rasgo

`Etot_v_h` está limitado a solamente 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` llega a ser “1” o “verdadero”. `aEn` es un contador que se aumenta cuando se determina que una señal activa esté presente en una trama actual, y se disminuye cuando se determina que la trama actual no comprende una señal activa. `aEn` no se puede aumentar más de un cierto número, por ejemplo, 6, y

no se puede reducir a menos que cero. Después de un número 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,

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

10 La función de la bg_bgd ha de ser una marca para detectar que la energía de la 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 habla en SNR alta.

15 Una nueva lógica de decisión según una realización de la tecnología descrita en la presente memoria se construye como sigue 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 lo bien que el totalNoise hace el seguimiento de la estimación de energía mínima. Las condiciones evaluadas si la primera línea es verdadera es la lógica de decisión sobre lo grande que debería ser el tamaño del paso, updt_step y la actualización de la estimación del ruido real es la asignación de valor a "st-> bckr[i] =-". Obsérvese que el tmpN[i] es un nivel de nuevo ruido potencialmente calculado previamente según la solución descrita en el documento WO2011/049514. La lógica de decisión a continuación sigue la parte 209 de la Figura 2, que está parcialmente indicada en conexión con el código a continuación

```

if (bg_bgd && ( aE_bgd || sd1_bgd || st->lt_tn_track > 0.90f ) )           /*if 202:2 and 204:2)*/
{
    if( (st->act_pred < 0.85f || ( aE_bgd && st->lt_haco_ev < 0.05f ) ) &&
        (st->lt_Ellp_dist < 10 || sd1_bgd ) && st->lt_tn_dist < 40 &&
        ( (Etot - st->totalNoise ) < 15.0f || st->lt_haco_ev < 0.10f ) )    /*207*/
    {
        st->first_noise_updt = 1;
        for( i=0; i< NB_BANDS; i++ )
        {
            st->bckr[i] = tmpN[i]                                           /*208*/
        }
    }
    else if (aE_bgd && st->lt_haco_ev < 0.15f)
    {
        updt_step=0.1f;
        if (st->act_pred > 0.85f )
        {
            updt_step=0.01f                                              /*207*/
        }
        if (updt_step > 0.0f)
        {
            st->first_noise_updt = 1;
            for( i=0; i< NB_BANDS; i++ )
            {
                st->bckr[i] = st->bckr[i] + updt_step * (tmpN[i]-st->bckr[i]); /*208*/
            }
        }
    }
}

```

20

```

    }}}
    else
    {
        (st->first_noise_updt) +=1;
    }
}
else
{
    /*si en música disminuir bckr para caer más*/ /*si 203:2 y 205:1*/
    If ( st->low_tn_track_cnt > 300 && st->lt_haco_ev > 0.9f && st-> totalNoise > 0.0f)
    {
        For ( i=0; i< NB_BANDS; i++)
        {
            If (st->bckr[i] > 2 * E_MIN
            {
                St->bckr[i] = 0.98f * st->bckr[i];          /*206*/
            }
        }
    }
    Else
    {
        (st->first_noise_updt) += 1;
    }
}
}

```

El segmento de código en el último bloque de código que comienza con “/ * Si en música ... */” contiene el escalado forzado hacia abajo de la estimación de fondo que se usa si se sospecha que la entrada actual es música. Esto se decide como función de:

- 5 un período largo de seguimiento deficiente del ruido de fondo en comparación con la estimación de energía mínima, Y, apariciones frecuentes de eventos armónicos o de correlación, Y, la última condición “totalNoise > 0” es una comprobación 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 valor positivo pequeño. Esta es una comprobación de cada entrada en un vector que comprende las
- 10 estimaciones de fondo de subbanda, de manera que una entrada necesita exceder E_MIN con el fin de ser reducido (en el ejemplo siendo multiplicado por 0,98). Estas comprobaciones se hacen con el fin de evitar reducir las estimaciones de fondo a valores demasiado pequeños.

15 Las realizaciones mejoran la estimación del ruido de fondo que permite un rendimiento mejorado del SAD/VAD para lograr una solución DTX de alta eficiencia y evitar la degradación en la calidad del habla o la música causada recortando.

Con la eliminación de la realimentación de decisión descrita en el documento WO2011/09514 del Etot_v_h, hay una mejor separación entre la estimación del ruido y el SAD. Esto tiene beneficios ya que la estimación de ruido no se cambia si/cuando se cambia la función/sintonización de SAD. Es decir, la determinación de una estimación de ruido de fondo llega a ser independiente de la función del SAD. También, la sintonización de la lógica de estimación de

20 ruido llega a ser más fácil ya que uno no se ve afectado por los efectos secundarios del SAD cuando se cambian las estimaciones de fondo.

A continuación, sigue una descripción de las figuras que ilustran los problemas resueltos con las realizaciones descritas en la presente memoria.

25 La Figura 6 es un diagrama que muestra la energía, Etot (puntos) de una serie de tramas de una señal de audio. El diagrama muestra el fondo estimado con la solución de la técnica anterior (curva inferior, más delgada, “x”) y estimado según una realización de la solución sugerida (curva superior, más gruesa, “+”). Este diagrama muestra cómo las realizaciones permiten un mejor seguimiento del ruido de fondo manteniendo la estimación de energía total en un nivel más alto y reaccionando más rápido, por ejemplo, alrededor de la trama 2510 en comparación con 2610 para la solución original.

La Figura 7 es también un diagrama que muestra la energía, Etot (puntos) de un número de tramas de una señal de audio. El diagrama muestra el fondo estimado con la solución de la técnica anterior (curva inferior, más delgada, "x") y estimado según una realización de la solución sugerida (curva superior, más gruesa, "+"). Se puede ver que la estimación según la solución sugerida en la presente memoria hace el seguimiento del ruido de fondo de manera más eficiente, por ejemplo, el ruido de fondo entre las expresiones, en particular en el intervalo de número de tramas 1600-1700.

La Figura 8 es también un diagrama que muestra la energía, Etot (puntos) de un número de tramas de una señal de audio. El diagrama muestra el fondo estimado con la solución de la técnica anterior (curva más superior, más delgada, "x"), y estimado según una realización de la solución sugerida (curva más inferior, más gruesa, "+"). El diagrama muestra el beneficio de la solución sugerida en comparación con cuando el seguimiento (de la técnica anterior) del fondo es demasiado eficiente. Mientras que hay una ráfaga de energía en el fondo entre las tramas 2300 y 2400, hay un aumento del riesgo de recorte de la parte delantera de la expresión que comienza en la trama 2400.

La Figura 9 es también un diagrama que muestra la energía, Etot (puntos) de un número de tramas de una señal de audio, en este caso una señal musical. El archivo de música para esta ilustración tiene mucho ruido como inicio y esto hace que la estimación del ruido tome una decisión equivocada y permita una actualización un poco en el archivo (alrededor de la trama 200). No obstante, con la reducción de fondo forzada, comienza a recuperarse en la trama 1700 y mediante la trama 2100 la estimación de ruido se baja al nivel más bajo para la reducción forzada. Como se ve a partir de la figura, no sería posible tener la misma reducción de nivel de fondo con la lógica de actualización normal ya que la entrada es más alta que la estimación de fondo para la mayoría de las tramas.

Observaciones finales

El estimador de fondo descrito anteriormente puede estar comprendido en un SAD, 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 sobremesa, ordenador portátil, decodificador de TV o servidor doméstico/pasarela doméstica/punto de acceso doméstico/encaminador doméstico. El dispositivo de comunicación puede ser, en algunas realizaciones, un dispositivo de red de comunicaciones adaptado para codificación y/o transcodificación. Ejemplos de tales dispositivos de red de comunicaciones son servidores, tales como servidores de medios, servidores de aplicaciones, encaminadores, pasarelas y estaciones base de radio. El dispositivo de comunicación también se puede adaptar para ser colocado en, es decir, estar integrado en, un buque, tal como un barco, un vehículo aéreo no tripulado, un avión y un vehículo de carretera, tal como un automóvil, un autobús o un camión. Tal dispositivo integrado pertenecería típicamente a una unidad telemática del vehículo o un sistema de información y entretenimiento del vehículo.

Los pasos, funciones, procedimientos, módulos, unidades y/o bloques descritos en la presente memoria se pueden implementar en hardware usando cualquier tecnología convencional, tal como tecnología de circuitos discretos o de circuitos integrados, incluyendo tanto circuitería electrónica de propósito general como circuitería específica de la aplicación.

Los ejemplos particulares incluyen uno o más procesadores de señal digital configurados adecuadamente y otros circuitos electrónicos conocidos, por ejemplo, puertas lógicas discretas interconectadas para realizar una función especializada, o Circuitos Integrados de Aplicaciones Específicas (ASIC).

Alternativamente, al menos algunos de los pasos, funciones, procedimientos, módulos, unidades y/o bloques descritos anteriormente se pueden implementar en un software tal como un programa de ordenador para su ejecución mediante circuitería de procesamiento adecuada que incluya una o más unidades de procesamiento. El software podría ser transportado por un portador, 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 de ordenador en los nodos de red.

El diagrama o diagramas de flujo presentados en la presente memoria se pueden considerar como un diagrama o diagramas de flujo de ordenador, 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 de ordenador que se ejecuta en el procesador.

Ejemplos de circuitería de procesamiento incluyen, pero no se limitan a, uno o más microprocesadores, uno o más Procesadores de Señales Digitales, DSP, una o más Unidades de Procesamiento Central, CPU y/o cualquier circuitería lógica programable adecuada, tal como una o más Agrupaciones de Puertas Programables en Campo, FPGA, o uno o más Controladores de Lógica Programable, PLC. Es decir, las unidades o módulos en las disposiciones en los diferentes nodos descritos anteriormente se podrían implementar mediante una combinación de circuitos analógicos y digitales, y/o uno o más procesadores configurados con software y/o microprograma, por ejemplo, almacenados en una memoria. Uno o más de estos procesadores, así como el otro hardware digital, se pueden incluir en un único Circuito Integrado de Aplicaciones Específicas, ASIC, o varios procesadores y diverso

hardware digital se puede distribuir entre diversos componentes separados, ya sea empaquetado individualmente o ensamblado en un sistema en un chip, SoC.

5 También se debería entender que puede ser posible reutilizar las capacidades de procesamiento generales de cualquier dispositivo o unidad convencional en el que se implementa la tecnología propuesta. También puede ser posible reutilizar el software existente, por ejemplo, mediante la reprogramación del software existente o añadiendo nuevos componentes de software.

Abreviaturas

	AMR	Tasa Múltiple Adaptativa
	DTX	Transmisión discontinua
10	VAD	Detector de Actividad de Voz
	3GPP	Proyecto de Cooperación de Tercera Generación
	SID	Descriptor de Inserción de Silencio.
	SAD	Detector de Actividad de Voz
	SNR	Relación Señal a Ruido
15	WB	Banda Ancha

REIVINDICACIONES

1. Un método para la estimación de ruido de fondo en un segmento de señal de audio que comprende una pluralidad de subbandas, comprendiendo el método:
5 calcular una posible estimación de nuevo ruido de subbanda y actualizar una estimación de ruido de subbanda actual con la estimación de nuevo ruido de subbanda si el nuevo valor es menor que el valor actual; y
cuando el nivel de energía del segmento de señal de audio es menor que un umbral más alto (202:2) que un nivel de energía mínimo a largo plazo I_{t_min} , pero no se detecta ninguna pausa (204:1) en el segmento de señal de audio:
- determinar (203) si el segmento de señal de audio comprende música; y
10 - reducir (206) la estimación de ruido de subbanda actual si se determina que el segmento de señal de audio (203:2) comprende música y la estimación de ruido de subbanda actual excede un valor mínimo (205:1).
2. El método según la reivindicación 1, en donde la decisión de si el nivel de energía del segmento de señal de audio es más que un umbral más alto que I_{t_min} se basa solamente en información derivada de una señal de audio de entrada.
3. El método según la reivindicación 1 o 2, en donde se considera que se detecta una pausa cuando uno o
15 ambos de las siguientes se cumple además de que el nivel de energía del segmento de señal de audio sea menor que un umbral más alto que I_{t_min} :
- se ha determinado que un número predefinido de segmentos de señal de audio precedentes consecutivos no comprende una señal activa;
- una dinámica de la señal de audio excede un umbral.
4. Un aparato para la estimación de ruido de fondo en un segmento de señal de audio que comprende una pluralidad de subbandas, estando configurado el aparato para:
calcular una posible estimación de nuevo ruido de subbanda y actualizar una estimación de ruido de subbanda
actual con la estimación de nuevo ruido de subbanda si el nuevo valor es menor que el valor actual; y
20 cuando el nivel de energía del segmento de señal de audio es menor que un umbral más alto que un nivel de energía mínimo a largo plazo I_{t_min} , pero no se detecta ninguna pausa en el segmento de señal de audio:
25 - determinar si el segmento de señal de audio comprende música; y
- reducir la estimación de ruido de subbanda actual si se determina que el segmento de señal de audio comprende música y la estimación de ruido de subbanda actual excede un valor mínimo.
5. El aparato según la reivindicación 4, que está configurado para decidir si el nivel de energía del segmento de
30 señal de audio es mayor que un umbral más alto que I_{t_min} , en base solamente a la información derivada de una señal de audio de entrada.
6. El aparato según la reivindicación 4 o 5, que está configurado para detectar una pausa cuando se cumple uno o ambos de los siguientes, además de que el nivel de energía del segmento de señal de audio sea menor que un umbral más alto que I_{t_min} :
35 - se ha determinado que un número predefinido de segmentos de señal de audio precedentes consecutivos no comprende una señal activa;
- una dinámica de la señal de audio excede un umbral.
7. Un detector de actividad de sonido que comprende un aparato según cualquiera de las reivindicaciones 4-6.
8. Un códec que comprende un aparato según cualquiera de las reivindicaciones 4-6.
9. Un dispositivo inalámbrico que comprende un aparato según cualquiera de las reivindicaciones 4-6.
10. Un nodo de red que comprende un aparato según cualquiera de las reivindicaciones 4-6.
11. Un programa de ordenador, que comprende instrucciones que, cuando se ejecutan en al menos un
procesador, hacen que al menos un procesador lleve a cabo el método según cualquiera de las reivindicaciones 1-3.
12. Un portador que contiene el programa de ordenador de la reivindicación 11, en donde el portador es uno de
45 una señal electrónica, señal óptica, señal de radio o medio de almacenamiento legible por ordenador.

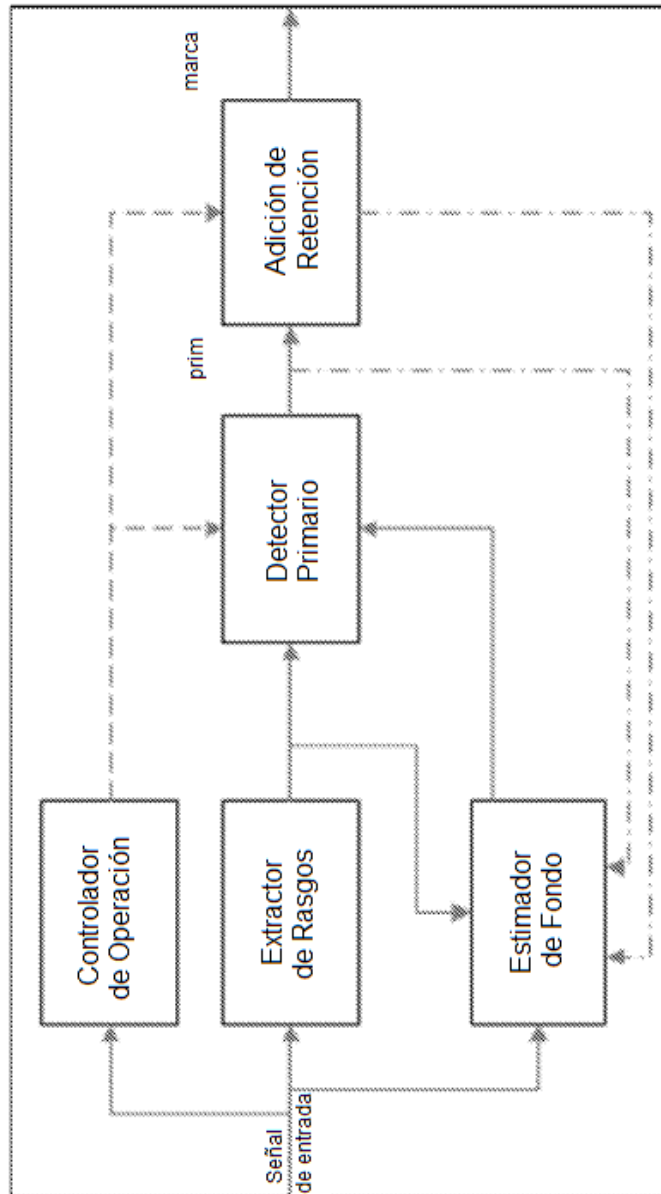


Figura 1

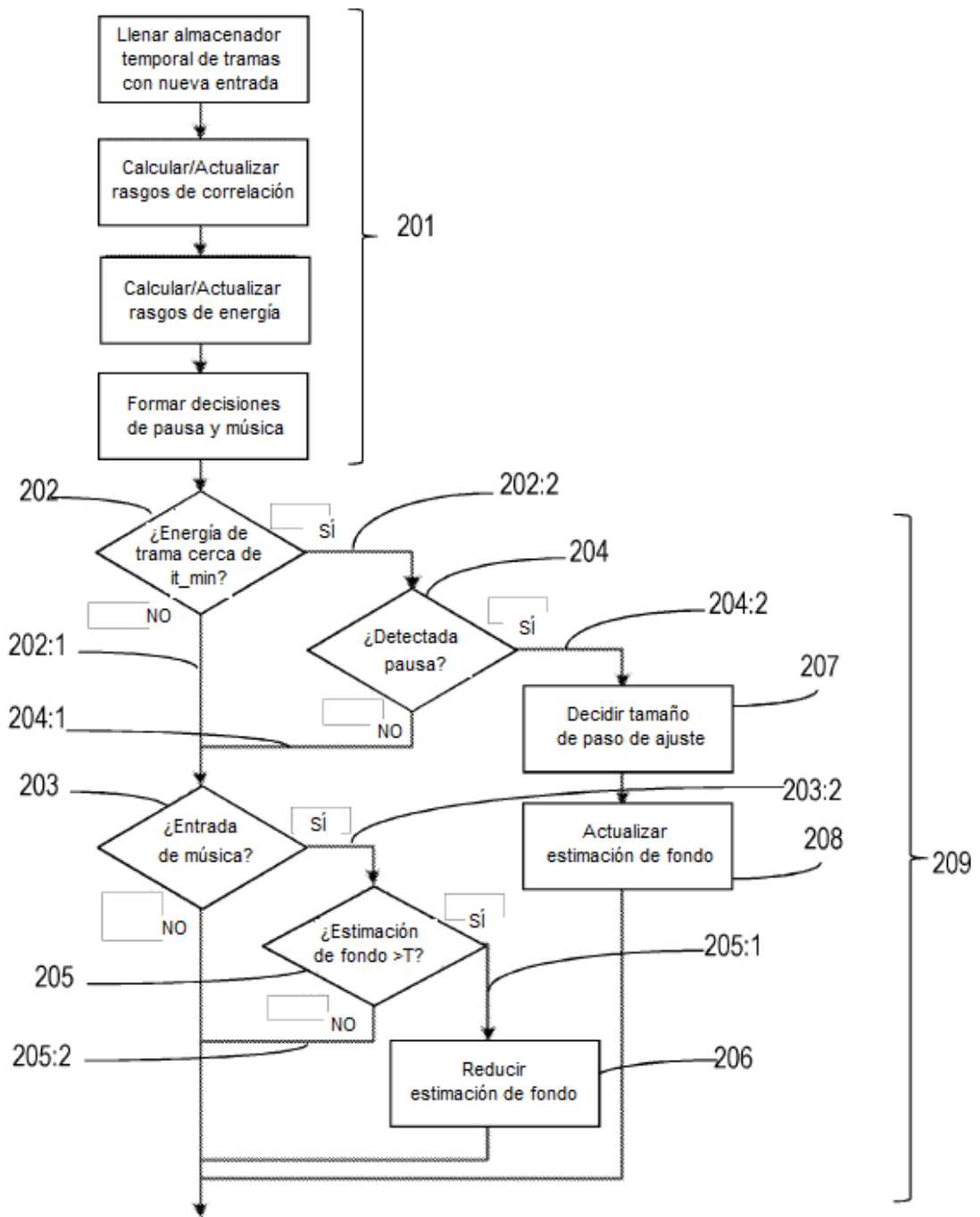


Figura 2

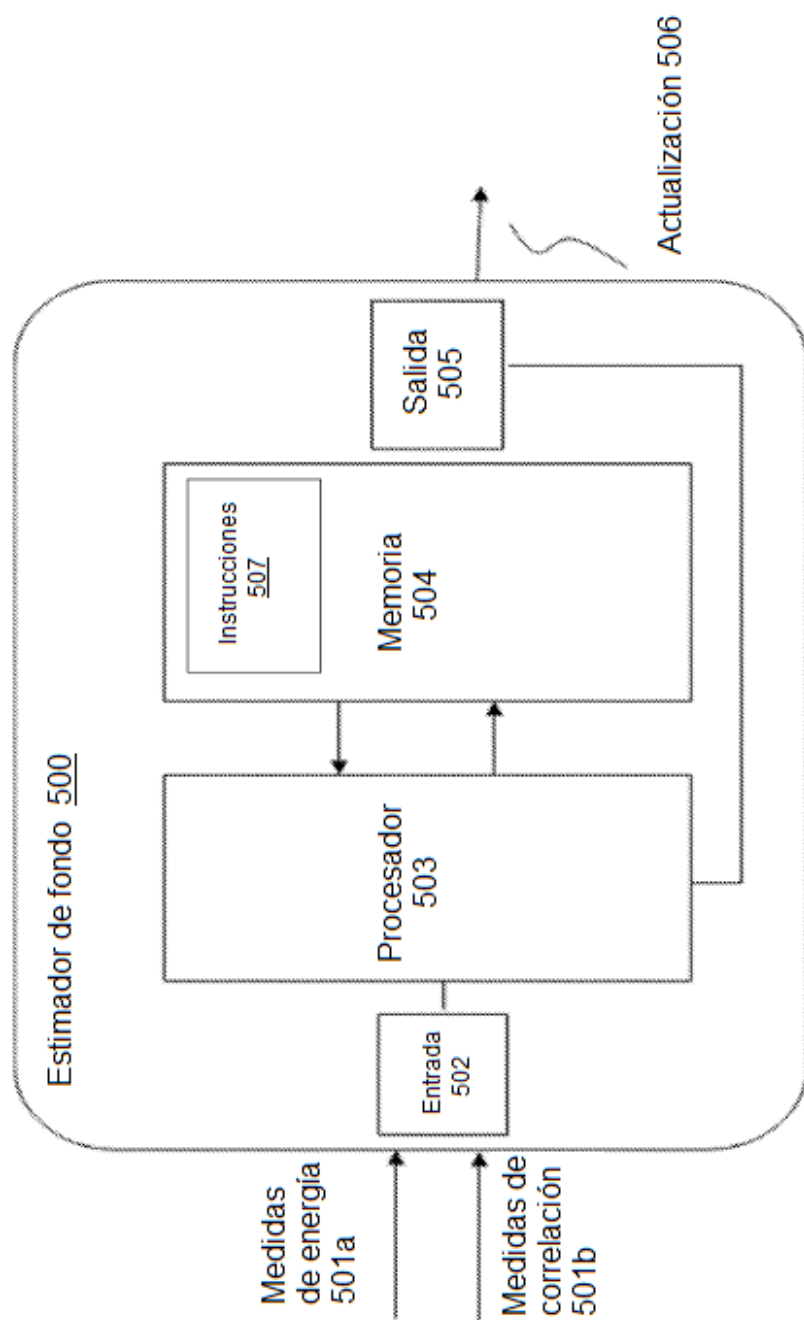


Figura 3

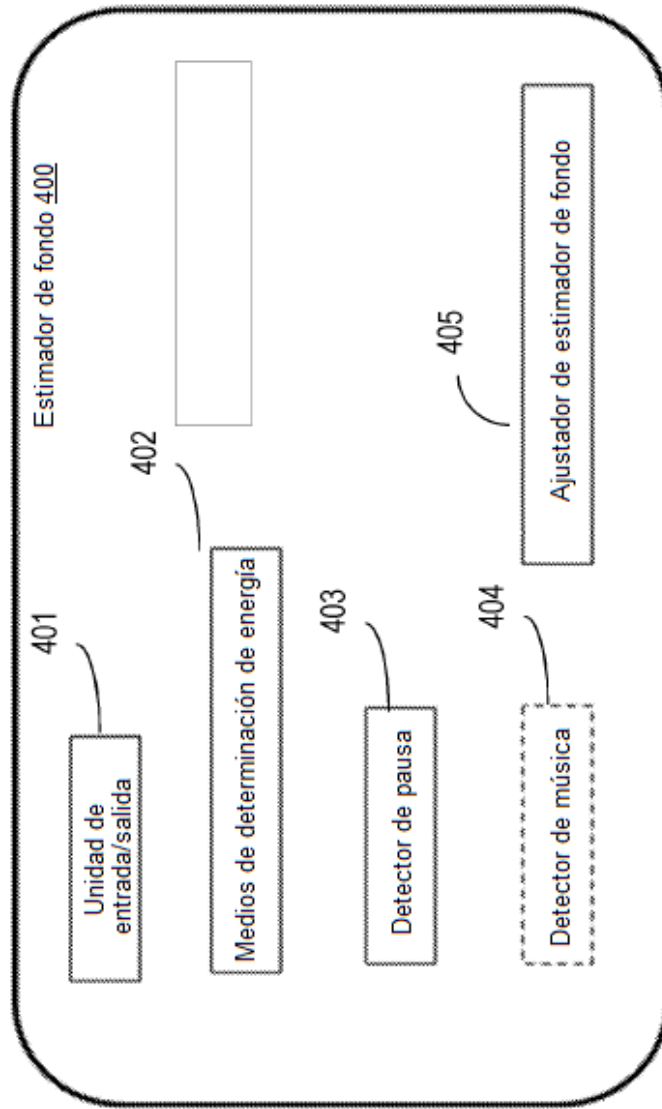


Figura 4

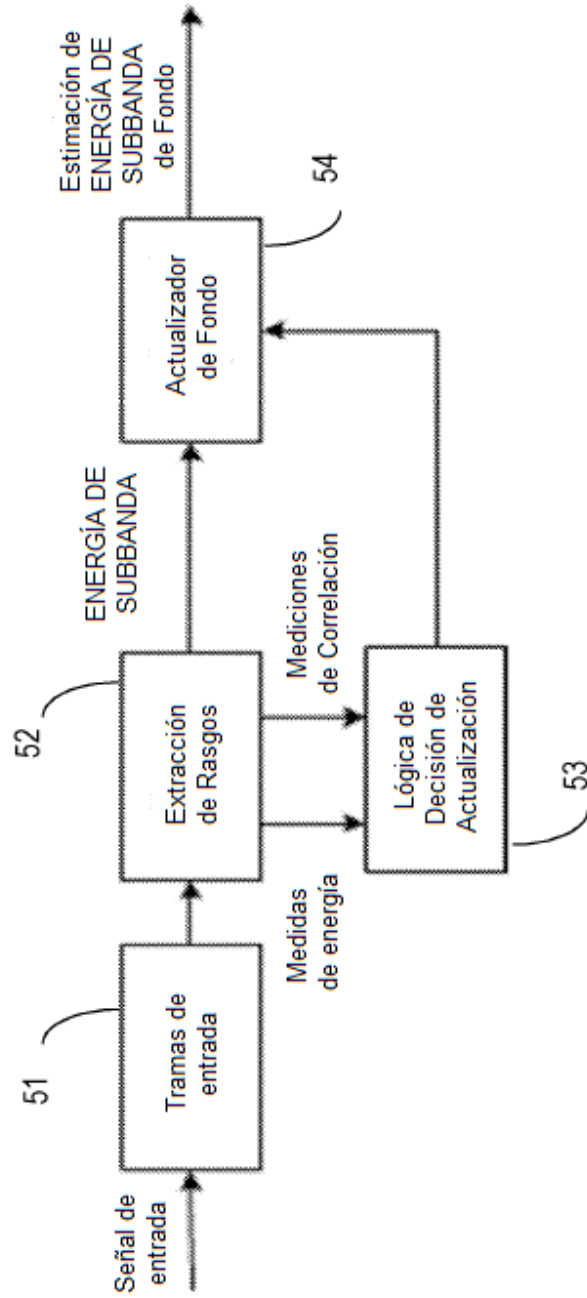


Figura 5

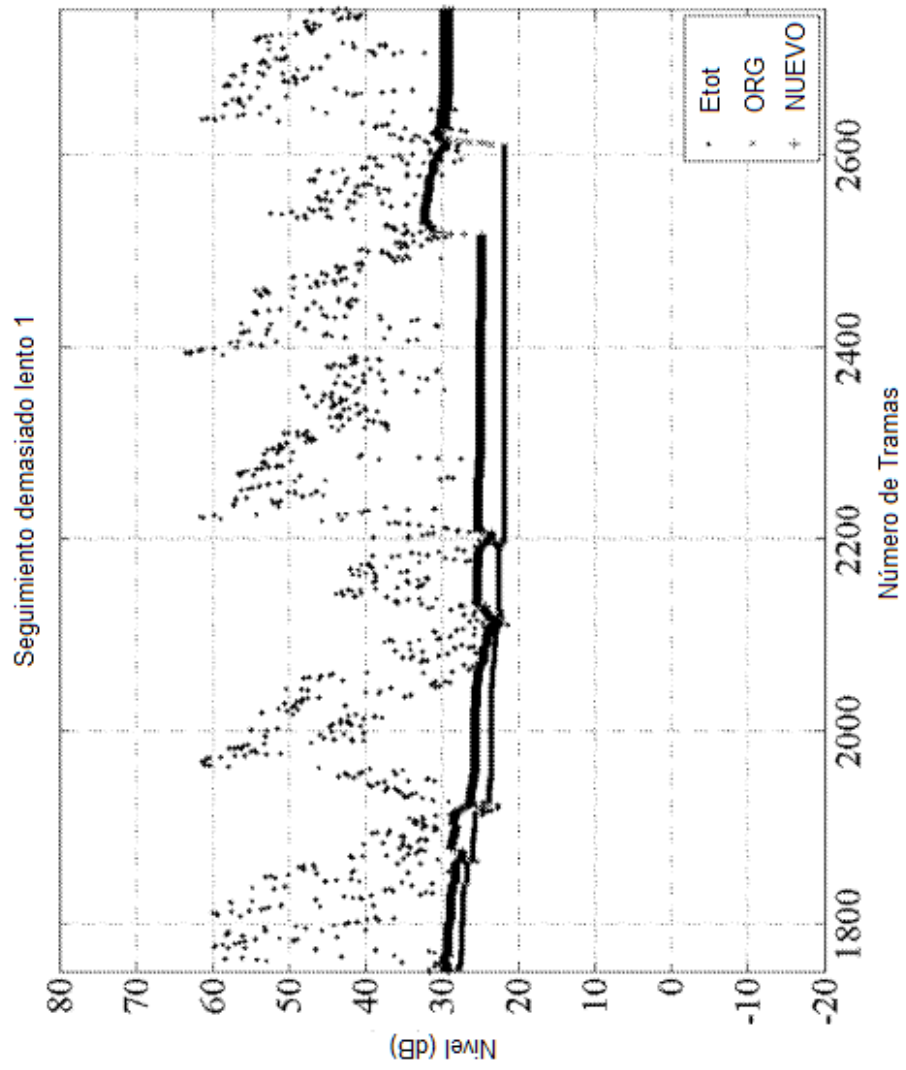


Figura 6

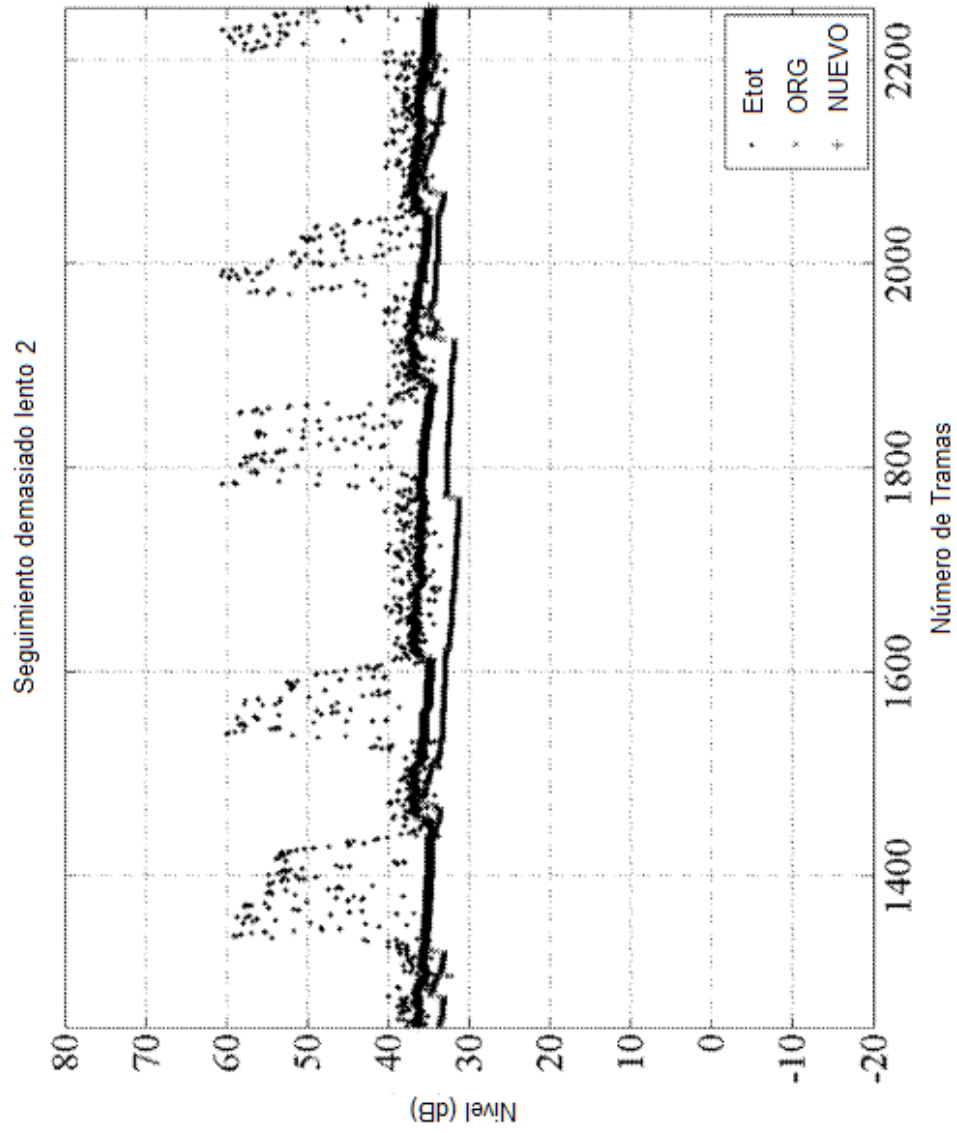


Figura 7

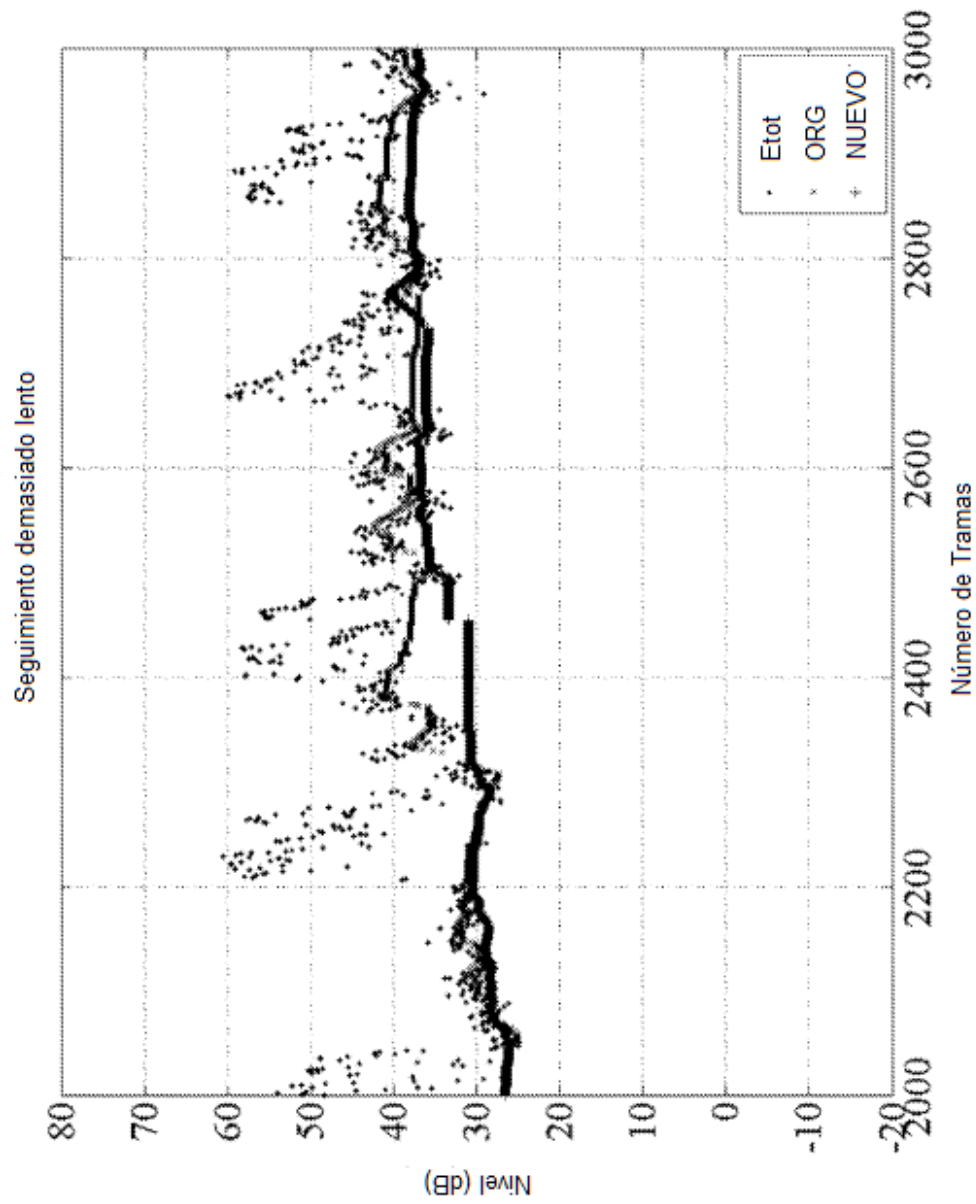


Figura 8

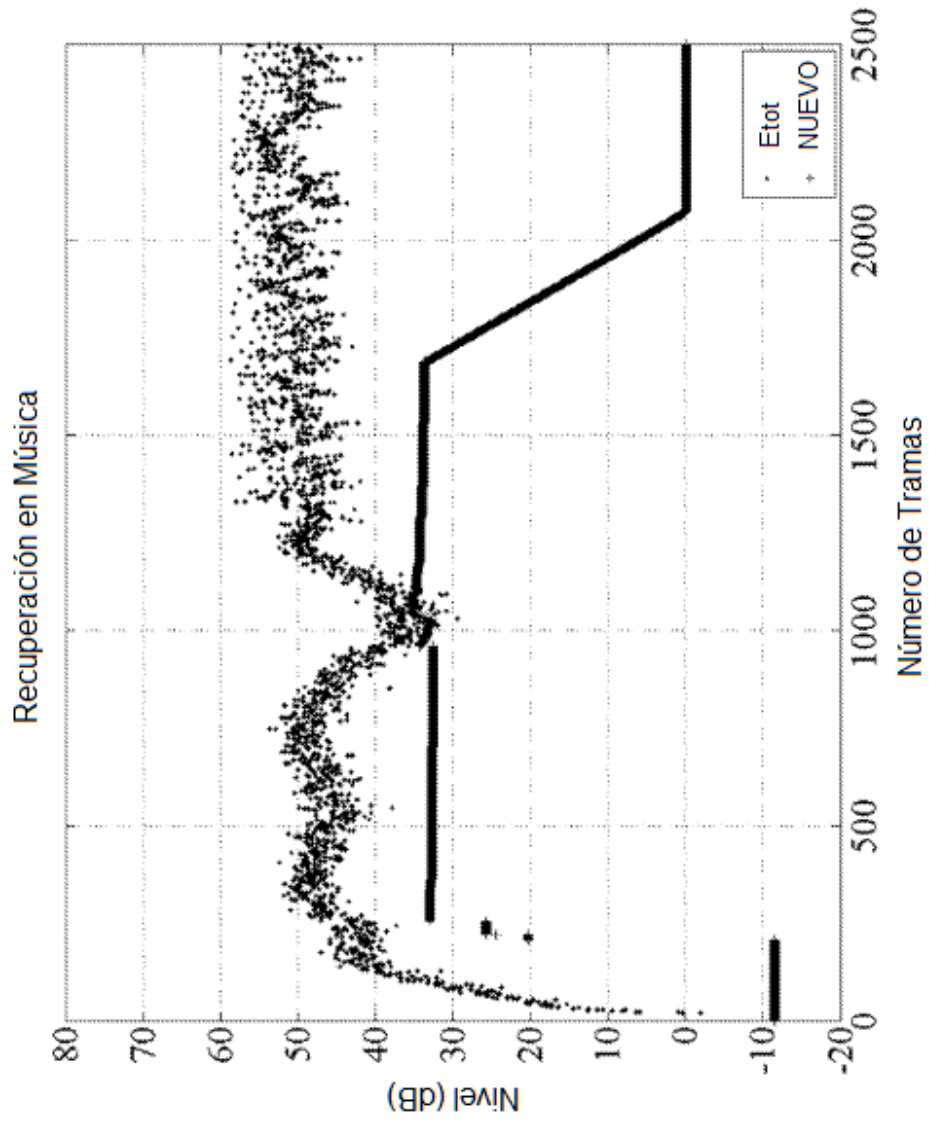


Figura 9