

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 467 966**

51 Int. Cl.:

G10L 21/038 (2013.01)

G10L 25/21 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.02.2009 E 09707285 (4)**

97 Fecha y número de publicación de la concesión europea: **14.05.2014 EP 2238593**

54 Título: **Método y aparato para estimar la energía de banda alta en un sistema de extensión del ancho de banda para señales de audio**

30 Prioridad:

07.02.2008 US 27571

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

13.06.2014

73 Titular/es:

**MOTOROLA MOBILITY LLC (100.0%)
600 North US Highway 45
Libertyville, IL 60048 , US**

72 Inventor/es:

**JASIUK, MARK A. y
RAMABADRAN, TENKASI V.**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 467 966 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Método y aparato para estimar la energía de banda alta en un sistema de extensión del ancho de banda para señales de audio

Aplicaciones relacionadas

- 5 Esta solicitud está relacionada con la solicitud de patente de U.S. co-dependiente y de propiedad conjunta de número 11/946,978 presentada el 29 de Noviembre de 2007. Esta solicitud está adicionalmente relacionada con la solicitud de patente de U.S. co-dependiente y de propiedad conjunta de número 12/024,620 presentada el 1 de Febrero de 2008.

Campo técnico

- 10 Esta invención se refiere en general a hacer audible un contenido y más particularmente a las técnicas de extensión del ancho de banda.

Antecedentes

- 15 El hacer audible un contenido de audio a partir de una representación digital comprende un área de trabajo conocida. En algunos de los ajustes de la aplicación la representación digital comprende un ancho de banda correspondiente que pertenece a una muestra de audio original. En tal caso, el hacer audible puede comprender una salida altamente precisa y que suene natural. Tal planteamiento, no obstante, requiere un considerable incremento de recursos para albergar la correspondiente cantidad de datos. En muchos ajustes de la aplicación, tales como, por ejemplo, ajustes de comunicación inalámbrica, tal cantidad de información no siempre puede ser adecuadamente soportada.

- 20 Para acomodar tal limitación, las llamadas técnicas de conversación de banda estrecha pueden servir para limitar la cantidad de información limitando, a su vez, la representación a menos de un ancho de banda correspondiente completo que pertenece a una muestra de audio original. Como un único ejemplo a este respecto, aunque la conversación natural incluye componentes significativos de hasta 8 kHz (o más), una representación de banda estrecha sólo puede proporcionar información relativa, digamos, al intervalo de 300 – 3400 Hz. El contenido resultante, cuando se hace audible, es típicamente suficientemente inteligible para soportar las necesidades funcionales de la comunicación basada en conversación. Desgraciadamente, no obstante, el procesamiento de la conversación de banda estrecha también tiende a conseguir una conversación que suena amortiguada y puede incluso tener una inteligibilidad reducida en comparación con la conversación de banda completa.

- 30 Para cubrir esta necesidad, se emplean en ocasiones técnicas de extensión de ancho de banda. Se genera artificialmente la información faltante en las bandas superior y/o inferior sobre la base de la información de banda estrecha disponible, así como de otra información, para seleccionar información que puede ser añadida al contenido de banda estrecha, para sintetizar con ello una señal de banda pseudo ancha (o completa). Utilizando tales técnicas, por ejemplo, se puede transformar conversación de banda estrecha en el intervalo de 300 - 3400 Hz en conversación de banda ancha, es decir, en el intervalo de 100 - 8000 Hz. Con este fin, una parte crítica de la información que se requiere es la envolvente espectral en la banda alta (3400 - 8000 Hz). Si se estima la envolvente espectral de banda ancha, la envolvente espectral de banda alta puede entonces normalmente ser obtenida fácilmente a partir de ella. Se puede considerar la envolvente espectral de banda alta comprendida por una forma y una ganancia (o de manera equivalente, energía).

- 40 Mediante un planteamiento, por ejemplo, la forma de envolvente espectral de banda alta es estimada estimando la envolvente espectral de banda ancha a partir de la envolvente espectral de banda estrecha mediante el mapeo del libro de códigos. La energía de banda alta es entonces estimada ajustando la energía dentro de la sección de banda estrecha de la envolvente espectral de banda ancha para que coincida con la energía de la envolvente espectral de banda estrecha. En este planteamiento, la forma de la envolvente espectral de banda alta determina la energía de banda alta y cualquier error en la estimación de la forma afectará de manera correspondiente a las estimaciones de la energía de banda alta.

- 45 En otro planteamiento, la forma de la envolvente espectral de banda alta y la energía de banda alta son estimadas de manera separada, y la envolvente espectral de banda alta que es finalmente utilizada es ajustada para que coincida con la energía de banda alta estimada. Mediante otro planteamiento relacionado se utiliza la energía de banda alta estimada, además de otros parámetros, para determinar la forma de la envolvente espectral de banda alta. No obstante, no está necesariamente asegurado el que la envolvente espectral de banda alta resultante tenga la energía de banda alta apropiada. En una etapa adicional se requiere por lo tanto ajustar la energía de la envolvente espectral de banda alta al valor estimado. A menos que se tenga especial cuidado, este planteamiento resultará en una discontinuidad en la envolvente espectral de banda ancha en la frontera entre la banda estrecha y la banda alta. Aunque los planteamientos existentes para la extensión de la banda ancha y, en particular, para la estimación de la envolvente de banda alta tienen un razonable éxito, no necesariamente conducen a una conversación resultante de calidad adecuada en al menos algunos ajustes de la aplicación.

5 Con el fin de generar conversación de ancho de banda extendido de una calidad aceptable, el número de aberraciones en tal conversación debe ser minimizado. Se conoce que la sobre-estimación de la energía de banda alta resulta en aberraciones molestas. Una incorrecta estimación de la envolvente espectral de banda alta puede también conducir a aberraciones pero estas aberraciones son normalmente más suaves y son fácilmente enmascaradas por la conversación de banda estrecha.

La publicación de M. Nilsson, W.B. Kleijn "Avoiding over-estimation in bandwidth extension of telephony speech", Procedimientos del IEEE ICASSP 2001, 7 de Mayo de 2001, vol. 2, páginas 869-872, describe un método de compensar la sobre-estimación de la energía de banda alta en la extensión del ancho de banda utilizando un función de coste asimétrica en el proceso de estimación.

10 La solicitud de patente internacional WO2009/070387 A1, describe que las tramas que contienen ataques y/o sonidos oclusivos pueden beneficiarse de un manejo especial cuando adaptan un valor de energía de banda alta estimado.

Compendio de la invención

15 La presente invención define un método de extensión de ancho de banda de acuerdo con la reivindicación 1 y un aparato para la extensión de ancho de banda de acuerdo con la reivindicación 3.

Breve descripción de los dibujos

20 Las necesidades anteriores se cubren al menos parcialmente mediante la provisión del método y aparato para estimar la energía de banda alta en un sistema de extensión de ancho de banda descrito en la descripción detallada que sigue. Las figuras que se acompañan, en las que números de referencia iguales se refieren a elementos idénticos o funcionalmente similares en las vistas separadas y, las cuales, junto con la descripción detallada que sigue, están incorporadas en y forman parte de la memoria, sirven para ilustrar con más detalle varias realizaciones y para explicar varios principios y ventajas de acuerdo por completo con la presente invención.

La FIG. 1 comprende un diagrama de flujo configurado de acuerdo con varias realizaciones de la invención;

la FIG. 2 comprende un gráfico configurado de acuerdo con varias realizaciones de la invención;

25 la FIG. 3 comprende un diagrama de bloques configurado de acuerdo con varias realizaciones de la invención;

la FIG. 4 comprende un diagrama de bloques configurado de acuerdo con varias realizaciones de la invención;

la FIG. 5 comprende un diagrama de bloques configurado de acuerdo con varias realizaciones de la invención; y

la FIG. 6 comprende un gráfico configurado de acuerdo con varias realizaciones de la invención;

30 Resultará evidente para los expertos en la materia que los elementos de las figuras se ilustran por sencillez y claridad y no necesariamente están dibujados a escala. Por ejemplo, las dimensiones y/o posicionamiento relativo de algunos de los elementos de las figuras pueden estar exageradas con respecto a otros elementos, para ayudar a mejorar la comprensión de varias realizaciones de la presente invención. También, elementos comunes pero bien comprendidos que resultan útiles o necesarios en una realización comercialmente factible típicamente no están representados con el fin de facilitar una vista menos obstruida de estas varias realizaciones de la presente invención. Resultará también evidente que ciertas acciones y/o etapas pueden ser descritas o representadas en un orden particular de ocurrencia, aunque resultará evidente para los expertos en la materia que tal especificidad con respecto a la secuencia no se requiere realmente. Debe comprenderse también que los términos y expresiones utilizados en esta memoria tienen el significado técnico ordinario que está acordado para tales términos y expresiones por los expertos en el campo técnico presentado anteriormente, excepto donde significados y expresiones diferentes hayan sido indicados de otro modo en esta memoria.

Descripción detallada

45 Las enseñanzas explicadas en esta memoria están dirigidas a un método y sistema rentable para una extensión artificial del ancho de banda. De acuerdo con tales enseñanzas, se recibe una señal de audio digital de banda estrecha. La señal de audio digital de banda estrecha puede ser una señal recibida a través de una estación de telefonía móvil en una red de telefonía móvil, por ejemplo, y la señal de audio digital de banda estrecha puede incluir conversación en el intervalo de frecuencias de 300 - 3400 Hz. Las técnicas de extensión artificial del ancho de banda son implementadas para extender el espectro de la señal de audio digital para incluir frecuencias de banda baja tales como 100 - 300 Hz y frecuencias de banda alta tales como 3400 - 8000 Hz. Utilizando extensión de ancho de banda artificial para extender el espectro para que incluya frecuencias de banda baja y de banda alta, se crea una
50 señal de audio digital que suena más natural que es más agradable para un usuario de una estación de telefonía móvil que implementa la técnica.

En las técnicas de extensión artificial del ancho de banda, la información faltante en las bandas superior (3400 - 8000 Hz) e inferior (100 - 300 Hz) es artificialmente generada sobre la base de la información de banda estrecha

disponible así como de información a priori derivada y almacenada de una base de datos de conversación y añadida a la señal de banda estrecha para sintetizar una señal de banda pseudo-ancha. Tal solución es bastante atractiva porque requiere mínimos cambios a un sistema de transmisión existente. Por ejemplo, no es necesaria ninguna velocidad de bits adicional. La extensión artificial del ancho de banda puede ser incorporada en un elemento de post-procesamiento en el extremo receptor, y es por lo tanto independiente de la tecnología de codificación de conversación utilizada en el sistema de comunicación o de la naturaleza del propio sistema de comunicación, por ejemplo, analógico, digital, terrestre o celular. Por ejemplo, las técnicas de extensión artificial del ancho de banda pueden ser implementadas por una estación de telefonía móvil que recibe una señal de audio digital de banda estrecha, y la señal de banda ancha resultante es utilizada para generar audio reproducido para un usuario de la estación de telefonía móvil.

Determinando la información de banda alta, la energía en la banda alta es estimada en primer lugar. Un subconjunto de la señal de banda estrecha es utilizado para estimar la energía de banda alta. El subconjunto de la señal de banda estrecha que está más cerca de las frecuencias de banda alta generalmente tiene la mayor correlación con la señal de banda alta. De acuerdo con esto, sólo un subconjunto de la banda estrecha, en lugar de toda la banda estrecha, es utilizado para estimar la energía de banda alta. El subconjunto que es utilizado se denomina la "banda de transición" y puede incluir frecuencias tales como 2500 - 3400 Hz. Más específicamente, la banda de transición se define en esta memoria como una banda de frecuencias que está contenida en la banda estrecha y está cerca de la banda alta, es decir, sirve como transición a la banda alta. Este planteamiento contrasta con los sistemas de extensión del ancho de banda de la técnica anterior, que estiman la energía de banda alta en términos de la energía en toda la banda estrecha, típicamente como una relación.

Con el fin de estimar la energía de banda alta, se estima primero la energía de banda de transición mediante técnicas que se explican en lo que sigue con respecto a las FIGS. 4 y 5. Por ejemplo, la energía de banda de transición de la banda de transición puede ser calculada aumentando en primer lugar la frecuencia de muestreo a una señal de banda estrecha de entrada, calculando el espectro de frecuencias de la señal de banda estrecha con frecuencia de muestreo aumentada (up sampled, en inglés), y sumando a continuación las energías de los componentes espectrales dentro de la banda de transición. La energía de banda de transición estimada es insertada a continuación en una ecuación polinómica como variable independiente para estimar la energía de banda alta. Los coeficientes o ponderaciones de las diferentes potencias de la variable independiente en la ecuación polinómica, incluyendo el de la potencia cero-ésima, esto es, el término constante, son seleccionados para minimizar el error medio cuadrático entre los valores verdaderos y estimados de la energía de banda alta sobre un gran número de tramas de una base de datos de conversación de entrenamiento. La precisión de la estimación puede ser también mejorada condicionando la estimación a parámetros derivados de la señal de banda estrecha así como a parámetros derivados de la señal de banda de transición tal como se explica con más detalle en lo que sigue. Después de que la energía de banda alta ha sido estimada, el espectro de banda alta es estimado sobre la base de la estimación de la energía de banda alta.

Utilizando la banda de transición de esta manera, se proporciona una técnica de extensión del ancho de banda robusta que produce una señal de audio correspondiente de mayor calidad de lo que sería posible si la energía de toda la banda estrecha fuese utilizada para estimar la energía de banda alta. además, esta técnica puede ser utilizada sin afectar adversamente de manera indebida a los sistemas de comunicación existentes, debido a que las técnicas de extensión de ancho de banda son aplicadas a una señal de banda estrecha recibida a través del sistema de comunicación, es decir, pueden utilizarse los sistemas de comunicación existentes para enviar las señales de banda estrecha.

La FIG. 1 ilustra un proceso 100 para generar una señal de audio digital de ancho de banda extendido de acuerdo con varias realizaciones de la invención. En primer lugar, en la operación 101, se recibe una señal de audio digital de banda estrecha. En un ajuste de la aplicación típico, éste comprenderá proporcionar una pluralidad de tramas de tal contenido. Estas enseñanzas acomodarán fácilmente el procesamiento de cada una de tales tramas de acuerdo con las etapas descritas. Mediante un planteamiento, por ejemplo, cada una de tales tramas puede corresponder a 10 - 40 milisegundos de contenido de audio original.

Esto puede comprender, por ejemplo, proporcionar una señal de audio digital que comprenda contenido vocal sintetizado. Tal es el caso, por ejemplo cuando se emplean estas enseñanzas junto con el contenido de la conversación codificado en voz recibido en un dispositivo de comunicaciones inalámbrico portátil. Existen también otras posibilidades, no obstante, como resultará evidente para los expertos en la materia. Por ejemplo, la señal de audio digital podría por el contrario comprender una señal de conversación original o una versión re-muestreada de cualquiera de las señales de conversación originales o del contenido de conversación sintetizado.

En referencia momentáneamente a la FIG. 2, debe entenderse que esta señal de audio digital pertenece a alguna señal de audio original 201 que tiene un ancho de banda 202 de señal correspondiente original. Este ancho de banda 202 de señal correspondiente original será típicamente mayor que el ancho de banda de señal mencionado anteriormente que corresponde a la señal de audio digital. Esto puede ocurrir, por ejemplo, cuando la señal de audio digital representa sólo una porción 203 de la señal de audio original 201, quedándose otras porciones fuera de banda. En el ejemplo ilustrativo mostrado, esto incluye una porción de banda baja 204 y una porción de banda alta 205. Resultará evidente para los expertos en la materia que este ejemplo sirve sólo como propósito ilustrativo y que

la porción no representada sólo puede comprender una porción de banda baja o una porción de banda alta. Estas enseñanzas serían también aplicables para su uso en un ajuste de aplicación en el que la porción no representada se encuentra en la banda media para dos o más porciones representadas (no mostradas).

5 Por lo tanto se comprenderá fácilmente que la porción o porciones no representada o no representadas de la señal de audio original 201 comprende o comprenden un contenido que estas enseñanzas actuales pueden razonablemente buscar reemplazar o representar de otro modo de alguna manera razonable y aceptable. Debe entenderse también que este ancho de banda de señal ocupa sólo una porción del ancho de banda de Nyquist determinado por la frecuencia de muestreo relevante. Esto, a su vez, se comprenderá que proporciona además una región de frecuencias en la cual efectuar la deseada extensión de ancho de banda.

10 En referencia de nuevo a la FIG. 1, la señal de audio digital de entrada es procesada para generar una señal de audio digital procesada en la operación 102. Mediante un planteamiento, el procesamiento en la operación 102 es una operación de aumento de frecuencia (up-sampling, en inglés). Mediante otro planteamiento, puede ser un sistema de ganancia de unidad simple para el cual la salida es igual a la entrada. En la operación 103, un nivel de energía de banda alta correspondiente a la señal de audio digital de entrada es estimado en una banda de transición de la señal de audio digital procesada dentro de un intervalo de frecuencias superior predeterminado de un ancho de banda de banda estrecha.

15 Utilizando los componentes de la banda de transición como base para la estimación, se obtiene una estimación más precisa de lo que sería generalmente posible si todos los componentes de banda estrecha fuesen utilizados colectivamente para estimar el valor de energía de los componentes de banda alta. Mediante un planteamiento, el valor de energía de banda alta es utilizado para acceder a una tabla de búsqueda que contiene una pluralidad de formas de envolvente espectral de banda alta candidatas correspondientes para determinar la envolvente espectral de banda alta, es decir, la forma de envolvente espectral de banda alta apropiada en el nivel de energía correcto.

20 En 104, el nivel de energía de banda alta estimado es modificado sobre la base de una precisión de estimación y/o de las características de la señal de banda estrecha para reducir las aberraciones y mejorar con ello la calidad de la señal de audio extendida en el ancho de banda. Esto se describirá con detalle en lo que sigue. Finalmente, en 105, una señal de audio digital de banda alta es opcionalmente generada sobre la base de la estimación modificada del nivel de energía de banda alta y en una estimación del espectro de banda alta correspondiente a la estimación modificada del nivel de energía de banda alta.

25 Este proceso 100 acomodará entonces opcionalmente el combinar la señal de audio digital con el contenido de banda alta correspondiente al valor de energía estimado y al espectro de los componentes de banda alta para proporcionar una versión de ancho de banda extendido de la señal de audio digital de banda estrecha para ser entregada. Aunque el proceso mostrado en la FIG. 1 sólo ilustra la adición de los componentes de banda alta estimados, resultará evidente que los componentes de banda baja pueden también ser estimados y combinados con la señal de audio digital de banda estrecha para generar una señal de banda ancha de ancho de banda extendido.

30 La señal de audio de ancho de banda extendido resultante (obtenida mediante la combinación de la señal de audio digital de entrada con el contenido de ancho de banda de fuera de señal generado artificialmente) tiene una mayor calidad de audio frente a la señal de audio digital de banda estrecha original cuando es proporcionada en forma audible. Mediante un planteamiento, esto puede comprender combinar dos elementos que son excluyentes entre sí con respecto a su contenido espectral. En tal caso, tal combinación puede tomar la forma, por ejemplo, de simplemente concatenar o unir los dos (o más) segmentos entre sí. Mediante otro planteamiento, si se desea, el contenido de ancho de banda de banda alta y/o de banda baja puede tener una porción que está dentro del correspondiente ancho de banda de señal de la señal de audio digital. Tal superposición puede resultar útil en al menos algunos ajustes de la aplicación para suavizar y/o aligerar la transición desde una porción a la otra combinando la porción que se superpone del contenido de ancho de banda de banda alta y/o de banda baja con la porción en banda correspondiente de la señal de audio digital.

35 Resultará evidente para los expertos en la materia que los procesos descritos anteriormente son fácilmente habilitados utilizando cualquiera de una amplia variedad de plataformas disponibles y/o fácilmente configuradas, que incluyen parcial o totalmente programables plataformas que son conocidas en el sector o plataformas de propósito dedicado como puede resultar deseable para algunas aplicaciones. En referencia ahora a la FIG. 3, se proporcionará ahora un planteamiento ilustrativo para tal plataforma.

40 En este ejemplo ilustrativo, en un aparato 300 un procesador 301 de elección se acopla de manera operable a una entrada 302 que está configurada y dispuesta para recibir una señal de audio digital que tiene un correspondiente ancho de banda de señal. Cuando el aparato 300 comprende un dispositivo de comunicaciones bidireccional inalámbrico, tal señal de audio digital puede ser proporcionada por un receptor 303 correspondiente como es bien conocido en la técnica. En tal caso, por ejemplo, la señal de audio digital puede comprender contenido vocal sintetizado formado como función de un contenido de conversación codificado en voz recibido.

45 El procesador 301, a su vez, puede ser configurado y dispuesto (mediante, por ejemplo, la programación correspondiente cuando el procesador 301 comprende una plataforma parcial o completamente programable que

son conocidas en la técnica) para realizar una o más de las etapas u otra funcionalidad presentada en esta memoria. Esto puede comprender, por ejemplo, estimar el valor de la energía de banda alta a partir de la energía de banda de transición y a continuación utilizar el valor de la energía de banda alta y un conjunto de formas clasificadas en función de la energía para determinar la envolvente espectral de banda alta.

5 Como se ha descrito anteriormente, mediante un planteamiento, el anteriormente mencionado valor de la energía de banda alta puede servir para facilitar el acceso a una tabla de búsqueda que contiene una pluralidad de formas de envolvente espectral candidatas correspondientes. Para soportar tal planteamiento, este aparato puede también comprender, si se desea, una o más tablas de búsqueda 304 que están acopladas en operación al procesador 301. Así configurado, el procesador 301 puede acceder fácilmente a la tabla de búsqueda 304 según sea apropiado.

10 Resultará evidente para los expertos en la materia y comprenderán que tal aparato 300 puede estar comprendido por una pluralidad de elementos distintos físicamente tal como se sugiere mediante la ilustración mostrada en la FIG. 3. Es también posible, no obstante, considerar esta ilustración con una vista lógica, en cuyo caso uno o más de estos elementos puede ser habilitado y realizado por medio de la plataforma compartida. Debe comprenderse también que tal plataforma compartida puede comprender una plataforma completamente o al menos parcialmente programable de las que se conocen en la técnica.

15 Resultará evidente que el procesamiento explicado anteriormente puede ser realizado por una estación de telefonía móvil en comunicación inalámbrica con una estación de base. Por ejemplo, la estación de base puede transmitir la señal de audio digital de banda estrecha a través de un medio convencional a la estación de telefonía móvil. Una vez recibida, el procesador o procesadores que se encuentran en la estación de telefonía móvil lleva o llevan a cabo las operaciones de requisito para generar una versión de ancho de banda extendido de la señal de audio digital que es más clara y más agradable audiblemente para un usuario de la estación de telefonía móvil.

20 En referencia ahora a la FIG. 4, la conversación de banda estrecha s_{nb} de entrada muestreada a 8 kHz se le aplica primero un aumento de la frecuencia de muestreo (up-sampled, en inglés) al doble utilizando un elevador de frecuencia de muestreo 401 correspondiente para obtener una conversación de banda estrecha de frecuencia de muestreo aumentada \acute{s}_{nb} muestreada a 16 kHz. Esto puede comprender realizar una interpolación de 1:2 (por ejemplo, insertando una muestra de valor cero entre cada par de muestras de conversación originales) seguida de un filtrado de paso bajo utilizando, por ejemplo, un filtro de paso bajo (LPF - Low Pass Filter, en inglés) que tiene una banda de paso entre 0 y 3400 Hz.

25 A partir de la s_{nb} , los parámetros lineales (LP - Linear Predictive, en inglés) de banda estrecha, $A_{nb} = \{1, a_1, a_2, \dots, a_P\}$ donde P es el orden del modelo, son también calculados utilizando un analizador de LP 402 que emplea técnicas de análisis de LP bien conocidas. (Existen otras posibilidades, por supuesto; por ejemplo, los parámetros de LP pueden ser calculados a partir de una versión decimada en 2:1 de la \acute{s}_{nb} .) Estos parámetros de LP modelizan la envolvente espectral de la conversación de entrada de banda estrecha como

$$SE_{nb}(j\omega) = \frac{1}{1 + a_1 e^{-j\omega} + a_2 e^{-j2\omega} + \dots + a_P e^{-jP\omega}}$$

35 En la ecuación anterior, la frecuencia angular ω en radianes/muestra viene dada por $\omega = 2\pi f/F_s$, donde f es la frecuencia de la señal en Hz y F_s es la frecuencia de muestreo en Hz. Para una frecuencia de muestreo F_s de 8 kHz, un orden de modelo P adecuado, por ejemplo, es 10.

40 Los parámetros de LP A_{nb} son a continuación interpolados por 2 utilizando un módulo de interpolación 403 para obtener los $\acute{A}_{nb} = \{1, 0, a_1, 0, a_2, 0, \dots, 0, a_P\}$. Utilizando los \acute{A}_{nb} , la conversación de banda estrecha de frecuencia de muestreo aumentada \acute{s}_{nb} es filtrada inversamente utilizando un filtro de análisis 404 para obtener la señal residual de LP \acute{r}_{nb} (que es también muestreada a 16 kHz). Mediante un planteamiento, esta operación de filtrado inverso (o análisis) puede ser descrita mediante la ecuación

$$\acute{r}_{nb}(n) = \acute{s}_{nb}(n) + a_1 \acute{s}_{nb}(n-2) + a_2 \acute{s}_{nb}(n-4) + \dots + a_P \acute{s}_{nb}(n-2P)$$

donde n es el índice de la muestra.

45 En un ajuste de la aplicación típico, el filtrado inverso de la \acute{s}_{nb} para obtener la \acute{r}_{nb} puede ser realizado de trama en trama, donde una trama se define como una secuencia de N muestras consecutivas sobre una duración de T segundos. Para muchas aplicaciones de la señal de conversación, una buena opción para T es aproximadamente 20 ms con valores correspondientes de frecuencia de muestreo para N de aproximadamente 160 a 8 kHz y aproximadamente 320. Sucesivas tramas pueden superponerse entre sí, por ejemplo, en aproximadamente 50%, en cuyo caso, la segunda mitad de las muestras en la trama actual y la primera mitad de las muestras en la siguiente trama son las mismas, y una nueva trama es procesada cada T/2 segundos. Para una elección de T de 20 ms y 50% de superposición, por ejemplo, los parámetros de LP A_{nb} son calculados a partir de 160 muestras de s_{nb}

consecutivas cada 10 ms, y son utilizadas para filtrar en inverso las 160 muestras del medio de la correspondiente trama \hat{s}_{nb} de 320 muestras para conseguir 160 muestras de \hat{r}_{nb} .

5 Se pueden calcular también los parámetros de LP de orden $2P$ para la operación de filtrado inverso directamente a partir de la conversación de banda estrecha de frecuencia de muestreo aumentada. Este planteamiento, no obstante, puede aumentar la complejidad tanto del cálculo de los parámetros de LP como de la operación de filtrado inverso, sin aumentar necesariamente el rendimiento bajo al menos algunas condiciones de operación.

10 La señal residual de LP \hat{r}_{nb} es a continuación rectificadora en onda completa utilizando un rectificador de onda completa 405 y filtrando en paso alto el resultado (utilizando, por ejemplo, un filtro de paso alto (HPF - High Pass Filter, en inglés) 406 con una banda de paso entre 3400 y 8000 Hz) para obtener la señal residual rectificadora de banda alta r_{hb} . En paralelo, la salida de una fuente de ruido pseudo-aleatorio 407 es también filtrada en paso alto 408 para obtener la señal de ruido de banda alta n_{hb} . Alternativamente, una secuencia de ruido filtrado en paso alto puede ser pre-almacenada en una memoria temporal (tal como, por ejemplo, una memoria temporal circular) y se puede acceder a ella según se requiera para generar n_{hb} . El uso de tal memoria temporal elimina los cálculos asociados con el filtrado en paso alto de las muestras de ruido pseudo-aleatorio en tiempo real. Estas dos señales, a saber, r_{hb} y n_{hb} , son a continuación mezcladas en un mezclador 409 de acuerdo con el nivel de voz v proporcionado por un Módulo de Estimación y Control (ECM - Estimation & Control Module, en inglés) 410 (cuyo módulo será descrito con más detalle en lo que sigue). En este ejemplo ilustrativo, este nivel de voz v va de 0 a 1, indicando 0 un nivel sin voz e indicando 1 un nivel de voz completa. El mezclador 409 forma esencialmente una suma ponderada de las dos señales de entrada en su salida tras asegurar que las dos señales de entrada están ajustadas para tener el mismo nivel de energía. La señal de salida del mezclador m_{hb} viene dada por

$$m_{hb} = (v) r_{hb} + (1-v) n_{hb}.$$

25 Resultará evidente para los expertos en la materia que también son posibles otras reglas de mezclado. Es también posible mezclar primero las dos señales, a saber, la señal residual de LP rectificadora en onda completa y la señal de ruido pseudo-aleatorio, y a continuación filtrar en paso alto la señal mezclada. En este caso, los dos filtros de paso alto 406 y 408 son reemplazados por un único filtro de paso alto situado a la salida del mezclador 409.

30 La señal resultante m_{hb} es a continuación pre-procesada utilizando un preprocesador de excitación 411 de banda alta (HB - High Band, en inglés) para formar la señal de excitación de banda alta e_{hb} . Las etapas del pre-procesamiento pueden comprender: (i) escalar la señal de salida del mezclador m_{hb} para que coincida con el nivel de energía de banda alta E_{hb} , y (ii) opcionalmente conformar la señal de salida del mezclador m_{hb} para que coincida con la envolvente espectral de banda alta SE_{hb} . Tanto la E_{hb} como la SE_{hb} son proporcionadas al pre-procesador de excitación 411 HB por el ECM 410. Cuando se emplea este planteamiento, puede resultar útil en muchos ajustes de la aplicación asegurar que tal conformado no afecta al espectro de fase de la señal de salida del mezclador m_{hb} ; esto es, el conformado puede ser realizado preferiblemente mediante un filtro de respuesta de fase cero.

35 La señal de conversación de banda estrecha de frecuencia de muestreo aumentada \hat{s}_{nb} y la señal de excitación de banda alta e_{hb} son sumadas utilizando un sumador 412 para formar la señal de banda mixta \hat{s}_{mb} . Esta señal de banda mixta \hat{s}_{mb} resultante es introducida en un filtro de ecualizador 413 que filtra esa entrada utilizando información de envolvente espectral de banda ancha SE_{wb} proporcionada por el ECM 410 para formar la señal de banda ancha estimada \hat{s}_{wb} . El filtro de ecualizador 413 esencialmente impone la envolvente espectral de banda ancha SE_{wb} en la señal de salida \hat{s}_{mb} para formar la \hat{s}_{wb} (una explicación más detallada a este respecto aparece en lo que sigue). La señal de banda ancha estimada \hat{s}_{wb} resultante es filtrada en paso alto, por ejemplo utilizando un filtro de paso alto 414 que tiene una banda de paso de 3400 a 8000 Hz, y filtrada en paso bajo, por ejemplo, utilizando un filtro de paso bajo 415 que tiene una banda de paso de 0 a 300 Hz, para obtener respectivamente la señal de banda alta \hat{s}_{hb} y la señal de banda baja \hat{s}_{lb} . Estas señales \hat{s}_{hb} , \hat{s}_{lb} , y la señal de banda estrecha de frecuencia de muestreo aumentada \hat{s}_{nb} son sumadas en otro sumador 416 para formar la señal de ancho de banda extendido S_{bwe} .

45 Resultará evidente para los expertos en la materia que existen varias configuraciones posibles diferentes para obtener la señal de ancho de banda extendido s_{bwe} . Si el filtro de ecualizador 413 guarda de manera precisa el contenido espectral de la señal de conversación de banda estrecha de frecuencia de muestreo aumentada \hat{s}_{nb} que forma parte de su señal de entrada \hat{s}_{mb} , entonces la señal de banda ancha \hat{s}_{wb} estimada puede salir directamente como la señal de ancho de banda extendido s_{bwe} , eliminando con ello el filtro de paso alto 414, el filtro de paso bajo 415 y el sumador 416. Alternativamente, pueden utilizarse dos filtros de ecualizador, uno para recuperar la porción de baja frecuencia y otro para recuperar la porción de alta frecuencia, y la salida del primero puede ser sumada a la salida filtrada en paso alto de la última para obtener la señal de ancho de banda extendido s_{bwe} .

50 Los expertos en la materia comprenderán y podrán apreciar que, con este ejemplo ilustrativo particular, la excitación residual rectificadora de banda alta y la excitación de ruido de banda alta son mezcladas entre sí de acuerdo con el nivel de voz. Cuando el nivel de voz es 0, indicando conversación sin voz, se utiliza exclusivamente la excitación de ruido. De manera similar, cuando el nivel de voz es 1, indicando conversación con voz, se utiliza exclusivamente la excitación residual rectificadora de banda alta es. Cuando el nivel de voz está entre 0 y 1, indicando conversación de voz mixta, las dos excitaciones son mezcladas en proporción adecuada tal como se determina y utiliza mediante el

nivel de voz. La excitación de banda alta mixta es así adecuada para sonidos con voz, sin voz y mixtos con y sin voz.

Debe comprenderse y apreciarse también que, en este ejemplo ilustrativo, se utiliza un filtro de ecualizador para sintetizar la \hat{s}_{wb} . El filtro de ecualizador considera la envolvente espectral de banda ancha SE_{wb} proporcionada por el ECM como la envolvente ideal y corrige (o ecualiza) la envolvente espectral de su señal de entrada \hat{s}_{mb} para que coincida con la ideal. Puesto que sólo las magnitudes están implicadas en la ecualización de la envolvente espectral, la respuesta de fase del filtro de ecualizador se elige que sea cero. La respuesta de magnitud del filtro de ecualizador es especificada por $SE_{wb}(\omega)/SE_{mb}(\omega)$. El diseño e implementación de tal filtro de ecualizador para una aplicación de codificación de conversación comprende un área de trabajo. Brevemente, no obstante, el filtro de ecualizador opera como sigue utilizando análisis de superponer - sumar (OLA - OverLap ADD, en inglés).

La señal de entrada \hat{s}_{mb} es primero dividida en tramas superpuestas, por ejemplo, tramas de 20 ms (320 muestras a 16 kHz) con 50% de superposición. Cada trama de muestras es a continuación multiplicada (por puntos) mediante una ventana adecuada, por ejemplo, una ventana de coseno alzado con propiedad de reconstrucción perfecta. La trama de conversación de ventana es a continuación analizada para estimar los parámetros de LP que modelizan su envolvente espectral. La envolvente espectral de banda ancha ideal para la trama es proporcionada por el ECM. A partir de las dos envolventes espectrales, el ecualizador calcula la respuesta de magnitud del filtro como $SE_{wb}(\omega)/SE_{mb}(\omega)$ y ajusta la respuesta de fase a cero. La trama de entrada es a continuación ecualizada para obtener la correspondiente trama de salida. Las tramas de salida ecualizadas son finalmente superpuestas - sumadas para sintetizar la conversación de banda ancha estimada \hat{s}_{wb} .

Resultará evidente para los expertos en la materia que además del análisis de LP, existen otros métodos para obtener la envolvente espectral de una trama de conversación dada, por ejemplo, análisis cepstral, ajuste de curva lineal por trozos o de orden superior de los picos de la magnitud espectral, etc.

Resultará también evidente para los expertos en la materia que en lugar de poner en una ventana la señal de entrada \hat{s}_{mb} directamente, se podría haber empezado con versiones de ventana de \hat{s}_{nb} , r_{hb} , y n_{hb} para conseguir el mismo resultado. Puede resultar también conveniente mantener el tamaño de trama y el porcentaje de superposición para el filtro del ecualizador iguales a los utilizados en el bloque de filtro de análisis utilizado para obtener \hat{f}_{nb} a partir de \hat{s}_{nb} .

El planteamiento del filtro de ecualizador descrito a la sintetización de la \hat{s}_{wb} ofrece varias ventajas: i) Puesto que la respuesta de fase del filtro de ecualizador 413 es cero, los componentes de diferente frecuencia de la salida del ecualizador están alineados en el tiempo con los correspondientes componentes de la entrada. Esto puede resultar útil para conversación de voz porque los segmentos de alta energía (tales como los segmentos de impulso glotal) de la excitación de banda alta $e_{x_{hb}}$ residual rectificadas están alineados en el tiempo con los correspondientes segmentos de alta energía de la conversación de banda estrecha de frecuencia de muestreo aumentada \hat{s}_{nb} en la entrada del ecualizador, y la preservación de este alineamiento en el tiempo en la salida del ecualizador a menudo actuará para asegurar una buena calidad de conversación; ii) la salida al filtro de ecualizador 413 no necesita tener un espectro plano como en el caso del filtro de síntesis de LP; iii) el filtro de ecualizador 413 está especificado en el dominio de la frecuencia, y por lo tanto es factible un mejor y más fino control sobre diferentes partes del espectro; y iv) son posibles iteraciones para mejorar la efectividad del filtrado al coste de una complejidad y retardo adicionales (por ejemplo, la salida del ecualizador puede ser alimentada de nuevo a la entrada para ser ecualizada una y otra vez para mejorar el rendimiento).

Se presentarán ahora algunos detalles adicionales relativos a la configuración descrita.

Pre-procesamiento de excitación de banda alta: La respuesta de magnitud del filtro de ecualizador 413 viene dada por $SE_{wb}(\omega)/SE_{mb}(\omega)$ y su respuesta en fase puede ser ajustada a cero. Cuanto más cerca esté la envolvente espectral de entrada $SE_{mb}(\omega)$ de la envolvente espectral ideal $SE_{wb}(\omega)$, más fácil es que el ecualizador corrija la envolvente espectral de entrada para que coincida con el ideal. Al menos una función del preprocesador de excitación de banda alta 411 es acercar la $SE_{mb}(\omega)$ a la $SE_{wb}(\omega)$ y así hacer el trabajo del filtro de ecualizador 413 más fácil. Primero, esto se lleva a cabo escalando la señal de salida del mezclador m_{hb} al nivel de energía de banda alta E_{hb} correcto proporcionado por el ECM 410. Segundo, la señal de salida del mezclador m_{hb} es opcionalmente conformada para que la envolvente espectral coincida con la envolvente espectral de banda alta SE_{hb} proporcionada por el ECM 410 sin afectar a su espectro de fase. Una segunda etapa puede comprender esencialmente una etapa de pre-ecualización.

Excitación de banda baja: A diferencia de la pérdida de información en la banda alta provocada por la restricción de ancho de banda impuesta, al menos en parte, por la frecuencia de muestreo, la pérdida de información en la banda baja (0 - 300 Hz) de la señal de banda estrecha es debida, al menos en gran medida, al efecto limitador de banda de la función de transferencia de canal que consiste en, por ejemplo, un micrófono, un amplificador, un codificador de conversación, un canal de transmisión, etc. En consecuencia, en una señal de banda estrecha limpia, la información de banda estrecha está aún presente aunque a muy bajo nivel. Esta información de nivel bajo puede ser amplificada de una manera directa para restaurar la señal original. Pero debe tenerse cuidado en este proceso puesto que las señales de nivel bajo son fácilmente corrompidas por errores, ruido y distorsiones. Una alternativa es sintetizar una

señal de excitación de banda baja similar a la señal de excitación de banda alta descrita anteriormente. Esto es, la señal de excitación de banda baja puede formarse mezclando la señal residual rectificada de banda baja r_{lb} y la señal de ruido de banda baja n_{lb} de una manera similar a la formación de la señal de salida del mezclador de banda alta m_{hb} .

5 En referencia ahora a la FIG. 5, se muestra el módulo de Estimación y Control (ECM - Estimation and Control Module, en inglés) 410 que comprende un detector de ataques / sonidos oclusivos 503, un calculador de cruces por
 10 cero 501, un estimador de pendiente de banda de transición 504, un estimador de espectro de banda estrecha 509, un estimador de espectro de banda baja 511, un estimador de espectro de banda ancha 512, un estimador de espectro de banda alta 510, un detector de SS / Transición 513, un estimador de energía de banda alta 506, un
 15 estimador de nivel de voz 502, un adaptador de energía 514, un suavizador de trayectoria energética 507 y un adaptador de energía 508.

El ECM 410 toma como entrada la conversación de banda estrecha s_{nb} , la conversación de banda estrecha de
 15 frecuencia de muestreo aumentada \hat{s}_{nb} , y los parámetros de LP de banda estrecha A_{nb} y proporciona como salida el nivel de voz v , la energía de banda alta E_{hb} , la envolvente espectral de banda alta SE_{hb} y la envolvente espectral de banda ancha SE_{wb} .

Estimación de nivel de voz: Para estimar el nivel de voz, un calculador de cruces por cero 501 calcula el número de cruces por cero z_c en cada trama de la conversación de banda estrecha s_{nb} como sigue:

$$z_c = \frac{1}{2(N-1)} \sum_{n=0}^{N-2} |Sgn(s_{nb}(n)) - Sgn(s_{nb}(n+1))|$$

donde

$$Sgn(s_{nb}(n)) = \begin{cases} 1 & \text{si } s_{nb}(n) \geq 0 \\ -1 & \text{si } s_{nb}(n) < 0 \end{cases}$$

20 n es el índice de muestra, y N es el tamaño de trama en las muestras. Resulta conveniente hacer que el tamaño de trama y el porcentaje de superposición utilizado en el ECM 410 sea el mismo que se utilizó en el filtro de ecualizador 413 y los bloques de filtro de análisis, por ejemplo, $T = 20$ ms, $N = 160$ para muestreo de a kHz, $N = 320$ para muestreo a 16 kHz, y 50% de superposición con referencia a los valores ilustrativos presentados anteriormente. El
 25 valor del parámetro z_c calculado como se ha indicado anteriormente va de 0 a 1. A partir del parámetro z_c , un estimador de nivel de voz 502 puede estimar el nivel de voz v como sigue.

$$v = \begin{cases} 1 & \text{si } z_c < ZC_{bajo} \\ 0 & \text{si } z_c > ZC_{alto} \\ 1 - \left[\frac{z_c - ZC_{bajo}}{ZC_{alto} - ZC_{bajo}} \right] & \text{Si no} \end{cases}$$

30 donde, ZC_{bajo} y ZC_{alto} representan umbrales alto y bajo apropiadamente elegidos respectivamente, por ejemplo, $ZC_{bajo} = 0,40$ y $ZC_{alto} = 0,45$. La salida d de un detector de ataques / sonidos oclusivos 503 puede también ser alimentada en el detector de nivel de voz 502. Si la trama está marcada como conteniendo un ataque o un sonido oclusivo con $d = 1$, el nivel de voz de esa trama así como de la siguiente trama puede ser ajustado a 1. Debe recordarse que, mediante un planteamiento, cuando el nivel de voz es 1, indicando conversación con voz, se utiliza exclusivamente la excitación residual rectificada de banda alta. Esto resulta ventajoso en un ataque / sonido oclusivo, en comparación con la excitación de banda alta sólo de ruido o mezclada, porque la excitación residual
 35 rectificada sigue de cerca el contorno de energía frente a tiempo de la conversación de banda estrecha de frecuencia de muestreo aumentada, reduciendo así la posibilidad de aberraciones de tipo de pre-eco debido a dispersión en el tiempo en la señal de ancho de banda extendido.

Con el fin de estimar la energía de banda alta, un estimador de energía de banda de transición 504 estima la energía de banda de transición a partir de la señal de conversación de banda estrecha de frecuencia de muestreo
 40 aumentada \hat{s}_{nb} . La banda de transición se define en esta memoria como una banda de frecuencias que está contenida en la banda estrecha y está cerca de la banda alta, es decir, sirve como transición a la banda alta, (que, en este ejemplo ilustrativo, es aproximadamente 2500 - 3400 Hz). De manera intuitiva, se esperaría que la energía de banda alta estuviese bien correlacionada con la energía de banda de transición, que se obtiene en experimentos. Una manera simple de calcular la energía de la banda de transición E_{tb} es calcular el espectro de frecuencias de \hat{s}_{nb}

(por ejemplo, mediante una Transformada de Fourier Rápida (FFT - Fast Fourier Transform, en inglés) y sumar las energías de los componentes espectrales dentro de la banda de transición.

A partir de la energía de la banda de transición E_{tb} en dB (decibelios), la energía de banda alta E_{hbo} en dB es estimada como

$$E_{hbo} = \alpha E_{tb} + \beta$$

5 donde, los coeficientes α y β son seleccionados para minimizar el error medio cuadrático entre los valores verdaderos y estimados de la energía de banda alta sobre un gran número de tramas de una base de datos de conversación de entrenamiento.

10 La precisión de estimación puede ser también mejorada aprovechando la información contextual de parámetros de conversación adicionales tales como el parámetro cruces por cero zc y el parámetro pendiente espectral de la banda de transición sl , que puede ser proporcionado por un estimador de pendiente de banda de transición 505. El parámetro cruces por cero, como se ha explicado anteriormente, es indicativo del nivel de voz de conversación. El parámetro pendiente indica la velocidad de cambio de la energía espectral dentro de la banda de transición. Puede estimarse a partir de los parámetros LP de banda estrecha A_{nb} mediante la aproximación de la envolvente espectral (en dB) dentro de la banda de transición como una línea recta, por ejemplo, mediante regresión lineal, y calculando su pendiente. El plano de parámetros zc - sl es a continuación dividido en varias regiones, y los coeficientes α y β son separadamente seleccionados para cada región. Por ejemplo, si los intervalos de los parámetros zc y sl están divididos cada uno en 8 intervalos iguales, el plano de parámetros zc - sl en entonces dividido en 64 regiones, y se seleccionan 64 conjuntos de α y β coeficientes, uno para cada región.

20 Mediante otro planteamiento (no mostrado en la FIG. 5) se consigue otra mejora en la precisión de la estimación como sigue. Debe observarse que en lugar del parámetro de pendiente sl (que es sólo una representación de primer orden de la envolvente espectral dentro de la banda de transición), puede emplearse una representación de mayor resolución para mejorar el rendimiento del estimador de la energía de banda alta. Por ejemplo, puede utilizarse una representación cuantificada de vector de formas de la envolvente espectral de la banda de transición (en dB). Como un ejemplo ilustrativo, el libro de códigos del cuantificador de vector (VQ - Vector Quantizer, en inglés) consiste en 64 formas denominadas parámetros de forma de envolvente espectral de banda de transición tbs que son calculadas a partir de una extensa base de datos de entrenamiento. Se podría reemplazar el parámetro sl en el plano de parámetros zc - sl con el parámetro tbs para conseguir un mayor rendimiento. Mediante otro planteamiento, no obstante, se introduce un tercer parámetro denominado medida de planicidad espectral $sfinis$. La medida de planicidad espectral se define como la relación de la media geométrica con respecto a la media aritmética de la envolvente espectral de banda estrecha (en dB) dentro del intervalo de frecuencias apropiado (tal como, por ejemplo, 300 - 3400 Hz). El parámetro sfm indica cómo es de plana la envolvente espectral - que va en este ejemplo de aproximadamente 0 para una envolvente con picos a 1 para una envolvente completamente plana. El parámetro sfm está también relacionado con el nivel de voz de conversación pero de una manera muy diferente que zc .
30 Mediante un planteamiento, el espacio de parámetros zc - sfm - tbs tridimensional está dividido en un número de regiones como sigue: El plano zc - sfm está dividido en 12 regiones dando lugar con ello a $12 \times 64 = 768$ posibles regiones en el espacio tridimensional. No todas estas regiones, no obstante, tienen suficientes puntos de datos de la base de datos de entrenamiento. Así, para muchos ajustes de la aplicación, el número de regiones útiles está limitado aproximadamente a 500, siendo un conjunto separado de coeficientes α y β seleccionado para cada una de estas regiones.

Un estimador de energía de banda alta 506 puede proporcionar una mejora adicional en la precisión de estimación utilizando potencias altas de E_{tb} en la estimación de E_{hbo} , por ejemplo,

$$E_{hbo} = \alpha_4 E_{tb}^4 + \alpha_3 E_{tb}^3 + \alpha_2 E_{tb}^2 + \alpha_1 E_{tb} + \beta.$$

45 En este caso, cinco coeficientes diferentes, a saber, α_4 , α_3 , α_2 , α_1 , y β , son seleccionados para cada partición del plano de parámetros zc - sl (o alternativamente, para cada partición del espacio paramétrico zc - sfm - tbs). Puesto que las ecuaciones anteriores (véanse los párrafos 70 y 75) para estimar E_{hbo} no son lineales, debe tenerse especial cuidado en ajustar la energía de banda alta estimada a medida que el nivel de señal de entrada, es decir, la energía, cambia. Una manera de conseguir esto es estimar el nivel de la señal de entrada en dB, ajustar E_{tb} arriba o abajo para que corresponda con el nivel nominal de señal, estimar E_{hbo} , y ajustar E_{hbo} abajo o arriba para que corresponda con el nivel de señal real.

50 La estimación de la energía de banda alta es proclive a los errores. Puesto que la sobre-estimación conduce a aberraciones, la energía de banda alta estimada es desviada para que sea menor que una cantidad proporcional a la desviación estándar de la estimación de E_{hbo} . Esto es, la energía de banda alta es adaptada en el adaptador de energía 1 (514) como:

$$E_{hb1} = E_{hb0} - \lambda \cdot \sigma$$

donde, E_{hb1} es la energía de banda alta adaptada en dB, E_{hb0} es la energía de banda alta estimada en dB, $\lambda \geq 0$ es un factor de proporcionalidad, y σ es la desviación estándar del error estimado en dB. De este modo, tras recibir la señal de audio digital de entrada que comprende la señal de banda estrecha, y determinar el nivel de energía de banda alta a partir de la correspondiente señal de audio digital, el nivel de energía de banda alta estimado es modificado sobre la base de una precisión de estimación de la energía de banda alta estimada. Con referencia a la FIG. 5, el estimador de energía de banda alta 506 determina adicionalmente una medida de no fiabilidad en la estimación del nivel de energía de banda alta y el adaptador de energía 514 desvía el nivel de energía de banda alta estimado para que sea menor en una cantidad proporcional a la medida de no fiabilidad. En una realización de la presente invención la medida de no fiabilidad comprende una desviación estándar del error en el nivel de energía de banda alta estimado. Debe observarse que pueden emplearse también otras medidas de no fiabilidad sin separarse del alcance de esta invención.

Mediante "desvío de reducción" de la energía de banda alta estimada, la probabilidad (o número de ocurrencias) de sobre-estimación de energía se reduce, reduciendo con ello el número de aberraciones. También, la cantidad mediante la cual se reduce la energía de banda alta estimada es proporcional a lo buena que sea la estimación - una estimación más fiable (es decir menor valor de σ) se reduce en una menor cantidad que una estimación menos fiable. Aunque se diseña el estimador de energía de banda alta, el valor σ correspondiente a cada partición del plano de parámetros zc-sl (o alternativamente, cada partición del espacio paramétrico zc-sfm-tbs) es calculado a partir de la base de datos de conversación de entrenamiento y almacenado para un uso posterior en el "desvío de reducción" de la energía de banda alta estimada. El valor σ de las aproximadamente 500 particiones del espacio paramétrico zc-sfm-tbs, por ejemplo, va de aproximadamente 3 dB aproximadamente a 10 dB, con un valor medio de aproximadamente 5,8 dB. Un valor adecuado de λ para este elemento de predicción de la energía de banda alta, por ejemplo, es 1,5.

En un planteamiento de la técnica anterior, la sobre-estimación de la energía de banda alta es manejada utilizando una función de coste asimétrica que penaliza los errores sobre-estimados más que los errores subestimados en el diseño del estimador de energía de banda alta. En comparación con este planteamiento de la técnica anterior, el planteamiento de "desviación de reducción" descrito en esta invención tiene las siguientes ventajas: (A) El diseño del estimador de la energía de banda alta es más simple porque está basado en la función de coste de "error cuadrático" simétrica estándar; (B) la "desviación de reducción" se da explícitamente durante la fase de operación (y no implícitamente durante la fase de diseño) y por lo tanto la cantidad de "desviación de reducción" puede ser fácilmente controlada a voluntad; y (C) la dependencia de la cantidad de "desviación de reducción" de la fiabilidad de la estimación es explícita y directa (en lugar de depender implícitamente de la función de coste específica utilizada durante la fase de diseño).

Además de reducir las aberraciones debido a la sobre-estimación de la energía, el planteamiento de la "desviación de reducción" descrito anteriormente tiene un beneficio añadido para las tramas de voz - a saber, el de enmascaramiento de cualquier error en la estimación de la forma de la envolvente espectral de banda alta y la reducción por ello de las aberraciones "ruidosas" resultantes. No obstante, para tramas sin voz, si la reducción en la energía de banda alta estimada es demasiado grande, la conversación de salida de ancho de banda extendido ya no se escucha como una conversación de banda ancha. Para contrarrestar esto, la energía de banda alta estimada es adaptada de nuevo en el adaptador de energía 1 (514) dependiendo de su nivel de voz como

$$E_{hb2} = E_{hb1} + (1-v) \cdot \delta_1 + v \cdot \delta_2$$

donde, E_{hb2} es la energía de banda alta adaptada al nivel de voz en dB, v es el nivel de voz que va de 0 para una conversación sin voz a 1 para una conversación con voz, y δ_1 y δ_2 ($\delta_1 > \delta_2$) son constantes en dB. La elección de δ_1 y δ_2 depende del valor de λ utilizado para la "desviación de reducción" y es determinado empíricamente para conseguir la conversación de salida de mejor suene. Por ejemplo, cuando se elige λ como 1,5, δ_1 y δ_2 pueden ser elegidas como 7,6 y -0,3 respectivamente. Debe observarse que otras elecciones para el valor de λ pueden resultar en diferentes elecciones para δ_1 y δ_2 - los valores de δ_1 y δ_2 pueden ser ambos positivos o negativos o de signos opuestos. El mayor nivel de energía para la conversación sin voz enfatiza tal conversación en la salida de ancho de banda extendido en comparación con la salida de banda estrecha y también ayuda a seleccionar una forma de la envolvente espectral más apropiada para tales segmentos sin voz.

Con referencia a la FIG. 5, el estimador de nivel de voz proporciona un nivel de voz al adaptador de energía 1 que modifica también el nivel de energía de banda alta estimado sobre la base de las características de la señal de banda estrecha modificando también el nivel de energía de banda alta estimado sobre la base del nivel de voz. La nueva modificación puede comprender reducir el nivel de energía de banda alta para una conversación substancialmente de voz y/o incrementar el nivel de energía de banda alta para una conversación substancialmente sin voz.

Aunque el estimador de energía de banda alta 506 seguido del adaptador de energía 1 (514) funciona bastante bien para la mayoría de las tramas, ocasionalmente existen tramas para las cuales la energía de banda alta está extremadamente sub o sobre-estimada. Tales errores de estimación pueden ser al menos parcialmente corregidos por medio de un suavizador de trayectoria energética 507 que comprende un filtro de suavizado. Así, la etapa de modificar el nivel de energía de banda alta estimado sobre la base de las características de la señal de banda estrecha puede comprender el suavizado del nivel de energía de banda alta estimado (que ha sido previamente modificado tal como se ha descrito anteriormente sobre la base de la desviación estándar de la estimación σ y del nivel de voz v), esencialmente reduciendo una diferencia de energía entre tramas consecutivas.

Por ejemplo, la energía de banda alta adaptada al nivel de voz E_{hb2} puede ser suavizada utilizando un filtro de promediación de 3 puntos como

$$E_{hb3} = [E_{hb2}(k-1) + E_{hb2}(k) + E_{hb2}(k+1)] / 3$$

donde, E_{hb3} es la estimación suavizada y k es el índice de trama. El suavizado reduce la diferencia de energía entre tramas consecutivas, especialmente cuando una estimación es una "excepción", es decir, la estimación de energía de banda alta de una trama es demasiado alta o demasiado baja en comparación con las estimaciones de las tramas vecinas. Así, el suavizado ayuda a reducir el número de aberraciones en la conversación de ancho de banda extendido de salida. El filtro de promediación de 3 puntos introduce un retardo de una trama. Otros tipos de filtros con o sin retardo pueden ser también diseñados para suavizar la trayectoria energética.

El valor de energía suavizado E_{hb3} puede ser además adaptado por el adaptador de energía 2 (508) para obtener la estimación de energía de banda alta E_{hb} adaptada final. Esta adaptación puede implicar disminuir o aumentar el valor de energía suavizado sobre la base del parámetro ss proporcionado por el detector de estado estacionario / transición 513 y/o el parámetro d proporcionado por el detector de ataques / sonidos oclusivos 503. De este modo, la etapa de modificar el nivel de energía de banda alta estimado sobre la base de las características de la señal de banda estrecha puede comprender la etapa de modificar el nivel de energía de banda alta estimado (o el nivel de energía de banda alta estimado previamente modificado) sobre la base de si una trama es o no de estado estacionario o transitorio. Esto puede comprender reducir el nivel de energía de banda alta para tramas transitorias y/o incrementar el nivel de energía de banda alta para tramas de estado estacionario, y puede también comprender modificar el nivel de energía de banda alta estimado sobre la base de una ocurrencia de un ataque / sonido oclusivo. Mediante un planteamiento, la adaptación del valor de la energía de banda alta cambia no sólo el nivel de energía sino también la forma de envolvente espectral puesto que la selección del espectro de banda alta puede estar ligada a la energía estimada.

Una trama se define como una trama de estado estacionario si tiene suficiente energía (esto es, es una trama de conversación y no una trama de silencio) y está cerca de cada una de las tramas vecinas tanto en un sentido espectral como en términos de energía. Dos tramas pueden ser consideradas espectralmente cerca si la distancia de Itakura entre las dos tramas está por debajo de un umbral especificado. Pueden utilizarse también otros tipos de medidas de distancia espectral. Dos tramas se consideran cercanas en términos de energía si la diferencia en las energías de banda estrecha de las dos tramas está por debajo de un umbral especificado. Cualquier trama que no esté en estado estacionario se considera una trama de transición. Una trama de estado estacionario es capaz de enmascarar errores en la estimación de la energía de banda alta mucho mejor que las tramas de estado transitorio. De acuerdo con esto, la energía de banda alta estimada de una trama es adaptada sobre la base del parámetro ss , es decir, dependiendo de si es una trama de estado estacionario ($ss = 1$) o una trama de transición ($ss = 0$) como

$$E_{hb4} = \begin{cases} E_{hb3} + \mu_1 & \text{para tramas de estado estacionario} \\ \min(E_{hb3} - \mu_2, E_{hb2}) & \text{para tramas de transición} \end{cases}$$

donde, $\mu_2 > \mu_1 \geq 0$, son constantes empíricamente elegidas en dB para conseguir una buena calidad de la conversación de salida. Los valores de μ_1 y μ_2 dependen de la elección de la constante de proporcionalidad λ utilizada para la "desviación de reducción". Por ejemplo, cuando se elige λ como 1,5, δ_1 como 7,66, y δ_2 como -0,3, μ_1 y μ_2 pueden ser elegidos como 1,5 y 6,0 respectivamente. Nótese que en este ejemplo estamos aumentando ligeramente la energía de banda alta estimada para las tramas de estado estacionario y disminuyéndola significativamente más para las tramas de transición. Debe observarse que otras elecciones para los valores de λ , δ_1 , y δ_2 pueden resultar en diferentes elecciones para μ_1 y μ_2 - los valores de μ_1 y μ_2 pueden ser ambos positivos o negativos o de signos opuestos. Además, debe observarse que pueden utilizarse también otros criterios para identificar las tramas de estado estacionario / transición.

Sobre la base de la salida del detector de ataques / sonidos oclusivos d , el nivel de energía de banda alta estimado puede ser ajustado como sigue: Cuando $d = 1$, indica que la correspondiente trama contiene un ataque, por ejemplo, transición de silencio a sonido sin voz o con voz, o a un sonido oclusivo. Un ataque / sonido oclusivo es detectado en la trama actual si la energía de banda estrecha de la trama precedente está por debajo de un cierto umbral y la

diferencia de energía entre las tramas actuales y precedentes excede otro umbral. Pueden emplearse también otros métodos para detectar un ataque / sonido oclusivo. Un ataque / sonido oclusivo presenta un especial problema debido a las siguientes razones: A) La estimación la energía de banda alta cerca de ataques / sonidos oclusivos resulta difícil; B) pueden aparecer aberraciones de tipo de pre-eco en la conversación de salida debido al procesamiento de bloques típico empleado; y C) los sonidos oclusivos (por ejemplo, [p], [t], y [k]), tras su ráfaga de energía inicial, tienen características similares a ciertos sonidos sibilantes (por ejemplo, [s], [j], y [ʒ]) en la banda estrecha pero bastante diferentes en la banda alta, lo que lleva a una sobre-estimación de la energía y a las aberraciones consecuentes. La adaptación de la energía de banda alta para un ataque / sonido oclusivo ($d = 1$) se realiza como sigue:

$$E_{hb}(k) = \begin{cases} E_{\min} & \text{para } k = 1, \dots, K_{\min} \\ E_{hb4}(k) - \Delta & \text{para } k = K_{\min} + 1, \dots, K_T \text{ si } v(k) > V_1 \\ E_{hb4}(k) - \Delta + \Delta_T(k - K_T) & \text{para } k = K_T + 1, \dots, K_{\max} \text{ si } v(k) > V_1 \end{cases}$$

donde k es el índice de trama. Para las primeras K_{\min} tramas que empiezan con la trama ($k = 1$) en la cual se detecta el ataque / sonido oclusivo, la energía de banda alta es ajustada al valor más bajo posible E_{\min} . Por ejemplo, E_{\min} puede ser ajustada a $-\infty$ dB o a la energía de la forma de la envolvente espectral de banda alta con menor energía. Para las tramas subsiguientes (es decir, para el intervalo dado por $k = K_{\min} + 1$ a $k = K_{\max}$), la adaptación de la energía se realiza sólo siempre que el nivel de voz $v(k)$ de la trama excede el umbral V_1 . Siempre que el nivel de voz de una trama dentro de este intervalo resulta ser menor o igual que V_1 , la adaptación de la energía de un ataque se detiene inmediatamente, es decir, $E_{hb}(k)$ se hace igual a $E_{hb4}(k)$ hasta que se detecta el siguiente ataque. Si el nivel de voz $v(k)$ es mayor que V_1 , entonces para $k = K_{\min} + 1$ a $k = K_T$, la energía de banda alta se disminuye en una cantidad Δ fija. Para $k = K_T + 1$ a $k = K_{\max}$, la energía de banda alta es gradualmente incrementada de $E_{hb4}(k) - \Delta$ hacia $E_{hb4}(k)$ por medio de la secuencia $\Delta T(k - K_T)$ previamente especificada y en $k = K_{\max} + 1$, $E_{hb}(k)$ se iguala a $E_{hb4}(k)$, y esto continúa hasta que se detecta el siguiente ataque. Valores típicos de los parámetros utilizados para la adaptación de la energía basada en ataques / sonidos oclusivos, por ejemplo, son $K_{\min} = 2$, $K_T = 5$, $K_{\max} = 7$, $V_1 = 0,4$, $\Delta = -12$ dB, $\Delta T(1) = 6$ dB, y $\Delta T(2) = 9,5$ dB. Para $d = 0$, no se realiza ninguna adaptación de la energía, esto es, E_{hb} se iguala a E_{hb4} . De este modo, la etapa de modificar el nivel de energía de banda alta estimado sobre la base de las características de la señal de banda estrecha puede comprender la etapa de modificar el nivel de energía de banda alta estimado (o el nivel de energía de banda alta estimado previamente modificado) sobre la base de una ocurrencia de un ataque / sonido oclusivo.

La adaptación de la energía de banda alta tal como se presenta en los párrafos 77 al párrafo 95 ayuda a minimizar el número de aberraciones en la conversación de salida de ancho de banda extendido y mejorar con ello su calidad. Aunque la secuencia de operaciones utilizada para adaptar la energía de banda alta estimada ha sido presentada de una manera particular, resultará evidente para los expertos en la materia que tal especificidad con respecto a la secuencia no se requiere realmente. También, las operaciones descritas para modificar el nivel de energía de banda alta pueden ser selectivamente aplicadas.

La estimación de la envolvente espectral de banda ancha SE_{wb} se describe a continuación. Para estimar SE_{wb} , se puede estimar separadamente la envolvente espectral de banda estrecha SE_{nb} , la envolvente espectral de banda alta SE_{hb} , y la envolvente espectral de banda baja SE_{lb} , y combinar las tres envolventes entre sí.

Un estimador de espectro de banda estrecha 509 puede estimar la envolvente espectral de banda estrecha SE_{nb} a partir de la conversación de frecuencia de muestreo aumentada \hat{s}_{nb} . A partir de la \hat{s}_{nb} , los parámetros de LP, $B_{nb} = \{1, b_1, b_2, \dots, b_Q\}$ en los que Q es el orden del modelo, son primero calculados utilizando técnicas de análisis de LP bien conocidas. Para una frecuencia de frecuencia de muestreo aumentada de 16 kHz, un orden P de modelo adecuado, por ejemplo, es 20. Los parámetros de LP B_{nb} modelizan la envolvente espectral de la conversación de banda estrecha de frecuencia de muestreo aumentada como

$$SE_{usnb}(\omega) = \frac{1}{1 + b_1 e^{-j\omega} + b_2 e^{-j2\omega} + \dots + b_Q e^{-jQ\omega}}$$

En la ecuación anterior, la frecuencia angular ω en radianes/muestra viene dada por $\omega = 2\pi f / 2F_s$, donde f es la frecuencia de la señal en Hz y F_s es la frecuencia de muestreo en Hz. Debe observarse que las envolventes espectrales SE_{nb} y SE_{usnb} son diferentes puesto que la primera es derivada de la conversación de entrada de banda estrecha y la última de la conversación de banda estrecha de frecuencia de muestreo aumentada. No obstante, dentro de la banda de paso de 300 a 3400 Hz, están aproximadamente relacionadas por la $SE_{usnb}(\omega) \approx SE_{nb}(2\omega)$ a dentro de una constante. Aunque la envolvente espectral SE_{usnb} está definida sobre el intervalo 0 - 8000 (F_s) Hz, la porción útil se encuentra dentro de la banda de paso (en este ejemplo ilustrativo, 300 - 3400 Hz).

Como un ejemplo ilustrativo a este respecto, el cálculo de SE_{usnb} se realiza utilizando FFT como sigue. Primero, la respuesta del impulso del filtro inverso $B_{nb}(z)$ se calcula a una longitud adecuada, por ejemplo, 1024, como $\{1, b_1, b_2,$

... , b_Q), 0, 0, ... , 0}. A continuación se toma una FFT de la respuesta del impulso, se obtiene una envolvente espectral de magnitud SE_{usnb} calculando la magnitud inversa de cada índice en la FFT. Para una longitud de la FFT de 1024, la resolución de la frecuencia de SE_{usnb} calculada como se indicó anteriormente está por encima de $16000/1024 = 15,625$ Hz. A partir de la SE_{usnb} , la envolvente espectral de banda estrecha SE_{nb} es estimada simplemente extrayendo las magnitudes espectrales de dentro del intervalo aproximado, 300 - 3400 Hz.

Resultará evidente para los expertos en la materia que además del análisis de LP, existen otros métodos para obtener la envolvente espectral de una trama de conversación dada, por ejemplo, análisis cepstral, ajuste de curva lineal por trozos o de orden superior de los picos de la magnitud espectral, etc.

Un estimador de espectro de banda alta 510 toma una estimación de la energía de banda alta como entrada y selecciona una forma de envolvente espectral de banda alta que sea consistente con la energía de banda alta estimada. Una técnica para aportar diferentes formas de envolvente espectral de banda alta correspondientes a diferentes energías de banda alta se describe a continuación.

Empezando con una gran base de datos de entrenamiento de conversación de banda ancha muestreada a 16 kHz, la envolvente de magnitud espectral de banda ancha es calculada para cada trama de conversación utilizando análisis de LP u otras técnicas. A partir de la envolvente espectral de banda ancha de cada trama, la porción de banda alta correspondiente a 3400 - 8000 Hz es extraída y normalizada dividiendo por la magnitud espectral a 3400 Hz. Las envolventes espectrales de banda alta resultantes tienen así una magnitud de 0 dB a 3400 Hz. La energía de banda alta correspondiente a cada envolvente de banda alta normalizada es calculada a continuación. La colección de envolventes espectrales de banda alta es a continuación dividida en la energía de banda alta, por ejemplo, una secuencia de valores de energía nominales que difieren en 1 dB es seleccionada para cubrir todo el intervalo y todas las envolventes con energía dentro de 0,5 dB de un valor nominal están agrupadas juntas.

Para cada grupo así formado, la forma de la envolvente espectral de banda alta media es calculada y subsiguientemente la correspondiente energía de banda alta. En la FIG. 6, se muestra un conjunto de 60 formas 600 de envolvente espectral de banda alta (con magnitud en dB en lugar de en Hz) a diferentes niveles de energía. Contando desde la parte inferior de la figura, las formas 1^a, 10^a, 20^a, 30^a, 40^a, 50^a, y 60^a (denominadas en esta memoria formas pre-calculadas) fueron obtenidas utilizando una técnica similar a la descrita anteriormente. Las restantes 53 formas fueron obtenidas mediante simple interpolación lineal (en el dominio de dB) entre las forma pre-calculadas más cercanas.

Las energías de estas formas van de aproximadamente 4,5 dB para la 1^a forma aproximadamente a 43,5 dB para la forma 60^a. Dada la energía de banda alta para una trama, resulta sencillo seleccionar la coincidencia más cercana de la forma de la envolvente espectral de banda alta como se describirá más tarde en este documento. La forma seleccionada representa la envolvente espectral de banda alta SE_{nb} estimada dentro de una constante. En la FIG. 6, la resolución de energía media es aproximadamente 0,65 dB. Claramente, una mejor resolución es posible aumentando el número de formas. Dadas las formas de la FIG. 6, la selección de una forma para una energía particular es única. Puede pensarse también en una situación en la cual hay más de una forma para una energía dada, por ejemplo, 4 formas por nivel de energía, y en este caso, se necesita información adicional para seleccionar una de las 4 formas para cada nivel de energía dado. Además, se pueden tener múltiples conjuntos de formas, estando cada uno clasificado en función de la energía de banda alta, por ejemplo, dos conjuntos de formas seleccionables mediante el parámetro de voz v , uno para tramas de voz y el otro para tramas sin voz. Para una trama de voz mixta, las dos formas seleccionadas de los dos conjuntos pueden ser apropiadamente combinadas.

El método de estimación del espectro de banda alta descrito anteriormente ofrece algunas ventajas claras. Por ejemplo, este planteamiento ofrece un control explícito sobre la evolución en el tiempo de las estimaciones del espectro de banda ancha. Una evolución suave de las estimaciones del espectro de banda alta dentro de segmentos de conversación distintos, por ejemplo, conversación de voz, conversación sin voz, etc, es a menudo importante para una conversación de ancho de banda extendido sin aberraciones. Para el método de estimación del espectro de banda alta descrito anteriormente, es evidente a partir de la FIG. 6 que pequeños cambios en la energía de banda alta resultan en pequeños cambios en las formas de la envolvente espectral de banda alta. Así, una evolución suave del espectro de banda alta puede ser esencialmente asegurada asegurando que la evolución en el tiempo de la energía de banda alta dentro de segmentos de conversación distintos es también suave. Esto se consigue explícitamente mediante un suavizado de la trayectoria energética tal como se ha descrito anteriormente.

Debe observarse que segmentos de conversación distintos, dentro de los cuales se realiza el suavizado de energía, pueden ser identificados con una resolución incluso más fina, por ejemplo, rastreando el cambio en el espectro de conversación de banda estrecha o en el espectro de conversación de banda estrecha de frecuencia de muestreo aumentada de trama en trama utilizando cualquiera de las medidas de distancia espectral bien conocidas tales como la distorsión espectral logarítmica o la distorsión de Itakura basada en LP. Utilizando este planteamiento, puede definirse un segmento de conversación distinto como una secuencia de tramas dentro de la cual el espectro está evolucionando lentamente y que está flanqueado en cada lado por una trama, donde el cambio espectral calculado excede un umbral fijo o adaptativo, indicando con ello la presencia de una transición espectral a cada lado del segmento de conversación distinto. El suavizado de la trayectoria energética puede entonces ser realizado dentro del segmento de conversación distinto, pero no a través de las fronteras del segmento.

- En esta memoria, la evolución suave de la trayectoria energética de banda alta se traduce en una evolución suave de la envolvente espectral de banda alta estimada, lo que es una característica deseable dentro de un segmento de conversación distinto. Debe observarse también que este planteamiento para asegurar una evolución suave de la envolvente espectral de banda alta dentro de un segmento de conversación distinto puede ser también aplicado
- 5 como una etapa de post-procesamiento a una secuencia de envolventes espectrales de banda alta estimadas obtenida mediante los métodos de la técnica anterior. En ese caso, no obstante, las envolventes espectrales de banda alta necesitan ser explícitamente suavizadas dentro de un segmento de conversación distinto, a diferencia del suavizado directo de la trayectoria energética de las enseñanzas actuales, lo que automáticamente resulta en la evolución suave de la envolvente espectral de banda alta.
- 10 La pérdida de información de la señal de conversación de banda estrecha en la banda baja (la cual, en este ejemplo ilustrativo, puede ser de 0 - 300 Hz) no se debe a la restricción del ancho de banda impuesta por la frecuencia de muestreo, como en el caso de la banda alta, sino que se debe al efecto limitativo de la función de transferencia de canal que consiste, por ejemplo, en el micrófono, amplificador, codificador de conversación, canal de transmisión, etc.
- 15 Un planteamiento directo para restaurar la señal de banda baja es entonces contrarrestar el efecto de esta función de transferencia de canal dentro del intervalo de 0 a 300 Hz. Una manera sencilla de hacer esto es utilizar un estimador de espectro de banda baja 511 para estimar la función de transferencia de canal en el intervalo de frecuencias de 0 a 300 Hz a partir de los datos disponibles, obtener su inverso, y utilizar el inverso para potenciar la envolvente espectral de la conversación de banda estrecha de frecuencia de muestreo aumentada. Esto es, la
- 20 envolvente espectral de banda baja SE_{nb} es estimada como la suma de SE_{usnb} y una característica de potenciación de envolvente espectral $SE_{potenciación}$ diseñada a partir de la inversa de la función de transferencia de canal (asumiendo que las magnitudes de la envolvente espectral están expresadas en el dominio logarítmico, por ejemplo, dB). Para muchos ajustes de la aplicación, debe tenerse cuidado en el diseño del $SE_{potenciación}$. Puesto que la restauración de la señal de banda baja se basa esencialmente en la amplificación de una señal de nivel bajo, implica
- 25 el peligro de amplificar los errores, el ruido y las distorsiones típicamente asociadas con las señales de nivel bajo. Dependiendo de la calidad de la señal de nivel bajo, el máximo valor de potenciación debe ser restringido apropiadamente. También, dentro del intervalo de frecuencias de 0 aproximadamente a 60 Hz, resulta deseable un diseño de $SE_{potenciación}$ para que tenga valores bajos (o incluso negativos, es decir, que atenúen) para evitar amplificar el zumbido eléctrico y el ruido de fondo.
- 30 Un estimador de espectro de banda ancha 512 puede entonces estimar la envolvente espectral de banda ancha combinando las envolventes espectrales estimadas en la banda estrecha, la banda alta y la banda baja. Una manera de combinar las tres envolventes para estimar la envolvente espectral de banda ancha es como sigue.
- La envolvente espectral de banda estrecha SE_{nb} es estimada a partir de la \hat{s}_{nb} tal como se ha descrito anteriormente y sus valores dentro del intervalo de 400 a 3200 Hz son utilizados sin ningún cambio en la estimación de la
- 35 envolvente espectral de banda ancha SE_{wb} . Para seleccionar la forma de banda alta apropiada, son necesarios la energía de banda alta y el valor de la magnitud de inicio a 3400 Hz. La energía de banda alta E_{nb} en dB se estima como se ha descrito anteriormente. El valor de la magnitud de inicio a 3400 Hz se estima modelizando el espectro de magnitud de la FFT de la \hat{s}_{nb} en dB dentro de la banda de transición, a saber, 2500 - 3400 Hz, por medio de una línea recta mediante una regresión lineal y encontrando el valor de la línea recta a 3400 Hz. Sea este valor de
- 40 magnitud denotado por $M3400$ in dB. La forma de la envolvente espectral de banda alta es a continuación seleccionada como la una entre muchos valores, por ejemplo, tal como se muestra en la FIG. 6, que tiene el valor de energía más cercano a $E_{nb} - M3400$. Sea esta forma denotada por $SE_{más\ cercana}$. A continuación la estimación de la envolvente espectral de banda alta SE_{hb} y por lo tanto la envolvente espectral de banda ancha SE_{wb} dentro del intervalo de 3400 a 8000 Hz son estimadas como $SE_{más\ cercana} + M3400$.
- 45 Entre 3200 y 3400 Hz, la SE_{wb} es estimada como el valor interpolado linealmente en dB entre la SE_{nb} y una línea recta que une la SE_{nb} a 3200 Hz y $M3400$ a 3400 Hz. El propio factor de interpolación es cambiado linealmente de manera que la SE_{wb} estimada se desplaza gradualmente de la SE_{nb} a 3200 Hz a la $M3400$ a 3400 Hz. Entre 0 a 400 Hz, la envolvente espectral de banda baja SE_{lb} y la envolvente espectral de banda ancha SE_{wb} son estimadas como
- 50 $SE_{nb} + SE_{potenciación}$, donde $SE_{potenciación}$ representa una característica de potenciación apropiadamente diseñada a partir del inverso de la función de transferencia de canal tal como se ha descrito anteriormente.
- Como se ha indicado anteriormente, las tramas que contienen ataques y/o sonidos oclusivos pueden aprovecharse de un manejo especial para evitar aberraciones ocasionales en la conversación de ancho de banda extendido. Tales tramas pueden ser identificadas por un repentino aumento en su energía con respecto a las tramas precedentes. La salida del detector de ataques / sonidos oclusivos 503 d para una trama es ajustada a 1 siempre que la energía de la
- 55 trama precedente sea baja, es decir, por debajo de un cierto umbral, por ejemplo, -50 dB, y el aumento en energía de la trama actual con respecto a la trama precedente excede otro umbral, por ejemplo, 15 dB. Si no, la salida d del detector es ajustada a 0. La propia energía de trama es calculada a partir de la energía del espectro de magnitud de la FFT de la conversación de banda estrecha de frecuencia de muestreo aumentada \hat{s}_{nb} dentro de la banda estrecha, es decir, 300 - 3400 Hz. Como se ha observado anteriormente, la salida del detector de ataques / sonidos oclusivos 503 d es alimentada en el estimador de nivel de voz 502 y en el adaptador de energía 508. Como se ha
- 60 descrito anteriormente, siempre que una trama está marcada como conteniendo un ataque o un sonido oclusivo con

$d = 1$, el nivel de voz de esa trama así como de la siguiente trama puede ser ajustado a 1. También, el valor de la energía de banda alta de esa trama así como de las siguientes tramas es modificado como se ha descrito anteriormente.

5 Resultará evidente para los expertos en la materia que las técnicas de estimación de la energía de banda alta pueden ser utilizadas junto con otros sistemas de extensión del ancho de banda de la técnica anterior para escalar el contenido de la señal de banda alta artificialmente generado para tales sistemas hasta un nivel de energía apropiado. Además, debe observarse que aunque la técnica de estimación de la energía ha sido descrita con referencia a la banda de frecuencias alta, (por ejemplo, 3400 - 8000 Hz), también puede ser aplicada una estimación de la energía en cualquier otra banda redefiniendo de manera apropiada la banda de transición. Por ejemplo, para 10 estimar la energía en un contexto de banda baja, tal como 0 - 300 Hz, la banda de transición puede ser redefinida como la banda de 300 - 600 Hz. Resultará también evidente para los expertos en la materia que las técnicas de estimación de la energía de banda alta descritas en esta memoria pueden ser empleadas con propósitos de codificación de conversación / audio. Asimismo, las técnicas descritas en esta memoria para estimar la envolvente espectral de banda alta y la excitación de banda alta pueden ser también utilizadas en el contexto de la codificación de conversación / audio. 15

Debe observarse que técnicas distintas de las descritas en esta invención pueden ser utilizadas para estimar el nivel de energía de banda alta. También es posible que el sistema de extensión del ancho de banda reciba una estimación del nivel de energía de banda alta transmitido desde algún otro lugar. El nivel de la energía de banda alta puede ser también implícitamente estimado, por ejemplo, se podría estimar por el contrario el nivel de energía de la 20 señal de banda ancha, y a partir de esta estimación y de otra información conocida, puede extraerse el nivel de energía de banda alta.

Debe observarse que aunque la estimación de parámetros tales como la envolvente espectral, los cruces por cero, los coeficientes de LP, las energías de banda, etc. han sido descritos en los ejemplos específicos proporcionados previamente como realizados desde la conversación de banda estrecha en algunos casos y la conversación de 25 banda estrecha de frecuencia de muestreo aumentada en otros casos, resultará evidente para los expertos en la materia que la estimación de los parámetros respectivos y su subsiguiente uso y aplicación, pueden ser modificados para ser realizados desde cualquiera de esas dos señales (conversación de banda estrecha o la conversación de banda estrecha de frecuencia de muestreo aumentada), sin separarse del espíritu y el alcance de las enseñanzas descritas.

30 Resultará evidente para los expertos en la materia que pueden realizarse una gran variedad de modificaciones, alteraciones y combinaciones con respecto a las reivindicaciones descritas anteriormente sin separarse del alcance de la invención tal como se define en las reivindicaciones dependientes, y que tales modificaciones, alteraciones y combinaciones deben ser consideradas como pertenecientes al alcance del concepto de la invención.

REIVINDICACIONES

1. Un método de extensión del ancho de banda que comprende:

recibir una señal de audio digital de entrada que comprende una señal de banda estrecha en un primer intervalo de frecuencias;

5 determinar un nivel de energía de banda alta estimado en un segundo intervalo de frecuencias, correspondientes a la señal de audio digital de entrada, donde el segundo intervalo de frecuencias es mayor en frecuencia que el primer intervalo de frecuencias y a la energía de banda alta estimada le falta información para ser estimada y utilizada en la extensión del ancho de banda; y

10 modificar el nivel de energía de banda alta estimado sobre la base de las características de la señal de banda estrecha;

donde la etapa de modificar el nivel de energía de banda alta estimado comprende la etapa de modificar el nivel de energía de banda alta estimado sobre la base de una ocurrencia de un ataque / sonido oclusivo;

donde los niveles de energía de banda alta estimados de una secuencia de K_{max} tramas que empieza en una trama en la cual se ha detectado el ataque / sonido oclusivo son modificados;

15 donde las primeras K_{min} tramas son ajustadas a un nivel de energía lo más bajo posible E_{min} ;

donde la modificación de los niveles de energía de banda alta estimados continúa hasta la trama K_{max} -ésima siempre que el nivel de voz de una trama dentro de la secuencia de K_{max} tramas excede un umbral; y

20 donde la modificación del nivel de energía de banda alta estimado viene dada por la disminución del nivel de energía de banda alta en una cantidad fija hasta una trama K_T en la que el nivel de voz de la trama excede un umbral y es aumentado de nuevo hacia la energía de banda alta estimada.

2. El método de la reivindicación 1, en el que la energía de banda alta es una adaptación para un ataque / sonido oclusivo que se realiza como:

$$E_{hb}(k) = \begin{cases} E_{min} & \text{para } k = 1, \dots, K_{min} \\ E_{hb4}(k) - \Delta & \text{para } k = K_{min} + 1, \dots, K_T \text{ si } v(k) > V_1 \\ E_{hb4}(k) - \Delta + \Delta_T(k - K_T) & \text{para } k = K_T + 1, \dots, K_{max} \text{ si } v(k) > V_1 \end{cases}$$

25 donde E_{hb} es el nivel de energía de banda alta, E_{min} es el nivel de energía de banda alta lo más bajo posible, k es el índice de trama y $v(k)$ es un nivel de voz.

3. Un aparato para la extensión del ancho de banda que comprende:

30 una estimación y módulo de control (ECM - Estimation and Control Module, en inglés) que recibe una señal de audio digital de entrada que comprende una señal de banda estrecha en un primer intervalo de frecuencias, generar un nivel de energía de banda alta estimado en un segundo intervalo de frecuencias, correspondiendo a la señal de audio digital de entrada y la energía de banda alta estimada es información faltante para ser estimada y utilizada en la extensión del ancho de banda, y modificar el nivel de energía de banda alta estimado sobre la base de las características de la señal de banda estrecha, donde el segundo intervalo de frecuencias es mayor en frecuencia que el primer intervalo de frecuencias, y donde la modificación del nivel de energía de banda alta estimado comprende modificar el nivel de energía de banda alta estimado sobre la base de una ocurrencia de un ataque / sonido oclusivo;

35 donde los niveles de energía de banda alta estimados de una secuencia de K_{max} tramas que empieza en una trama en la cual se ha detectado el ataque / sonido oclusivo son modificados;

donde las primeras K_{min} tramas son ajustadas a un nivel de energía lo más bajo posible E_{min} ;

40 donde la modificación de los niveles de energía de banda alta estimados continúa hasta la trama K_{max} -ésima siempre que el nivel de voz de una trama dentro de la secuencia de K_{max} tramas excede un umbral; y

donde la modificación del nivel de energía de banda alta estimado viene dada por la disminución del nivel de energía de banda alta en una cantidad fija hasta una trama K_T en la que el nivel de voz de la trama excede un umbral y es aumentado de nuevo hacia la energía de banda alta estimada.

45 4. El aparato de la reivindicación 3, en el que la energía de banda alta es una adaptación para un ataque / sonido oclusivo que viene dada como:

$$E_{hb}(k) = \begin{cases} E_{\min} & \text{para } k = 1, \dots, K_{\min} \\ E_{hb4}(k) - \Delta & \text{para } k = K_{\min} + 1, \dots, K_T \text{ si } v(k) > V_1 \\ E_{hb4}(k) - \Delta + \Delta_T(k - K_T) & \text{para } k = K_T + 1, \dots, K_{\max} \text{ si } v(k) > V_1 \end{cases}$$

donde E_{hb} es el nivel de energía de banda alta, E_{\min} es el nivel de energía de banda alta lo más bajo posible, k es el índice de trama y $v(k)$ es un nivel de voz.

100

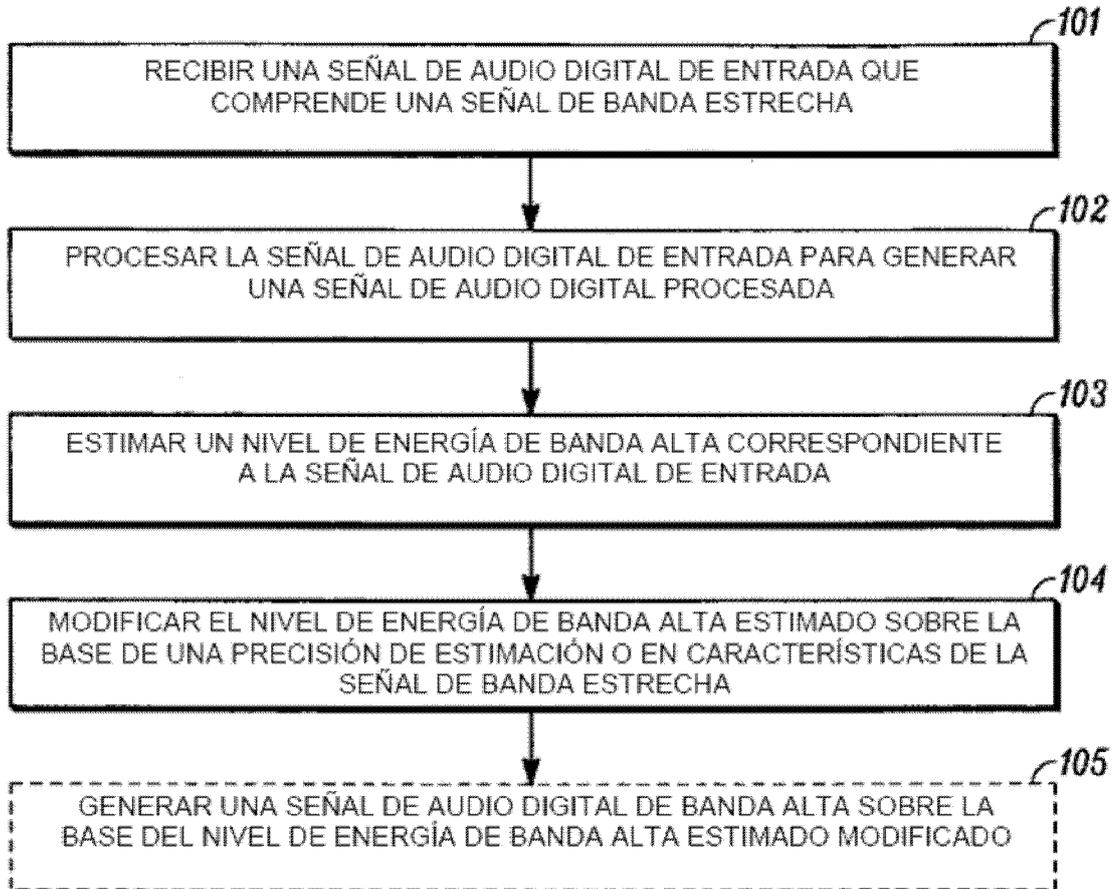


FIG. 1

200

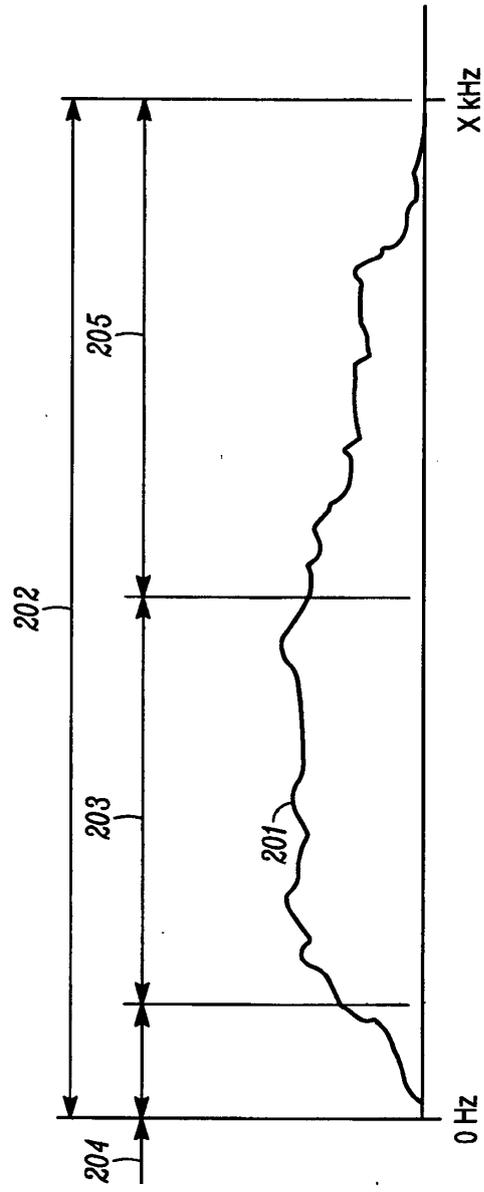


FIG. 2

300

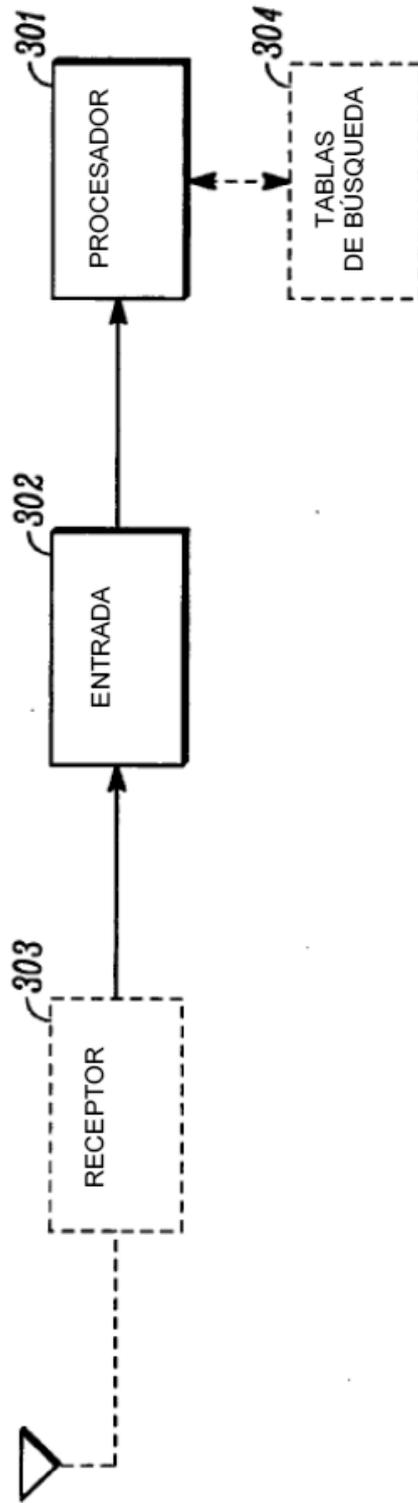


FIG. 3

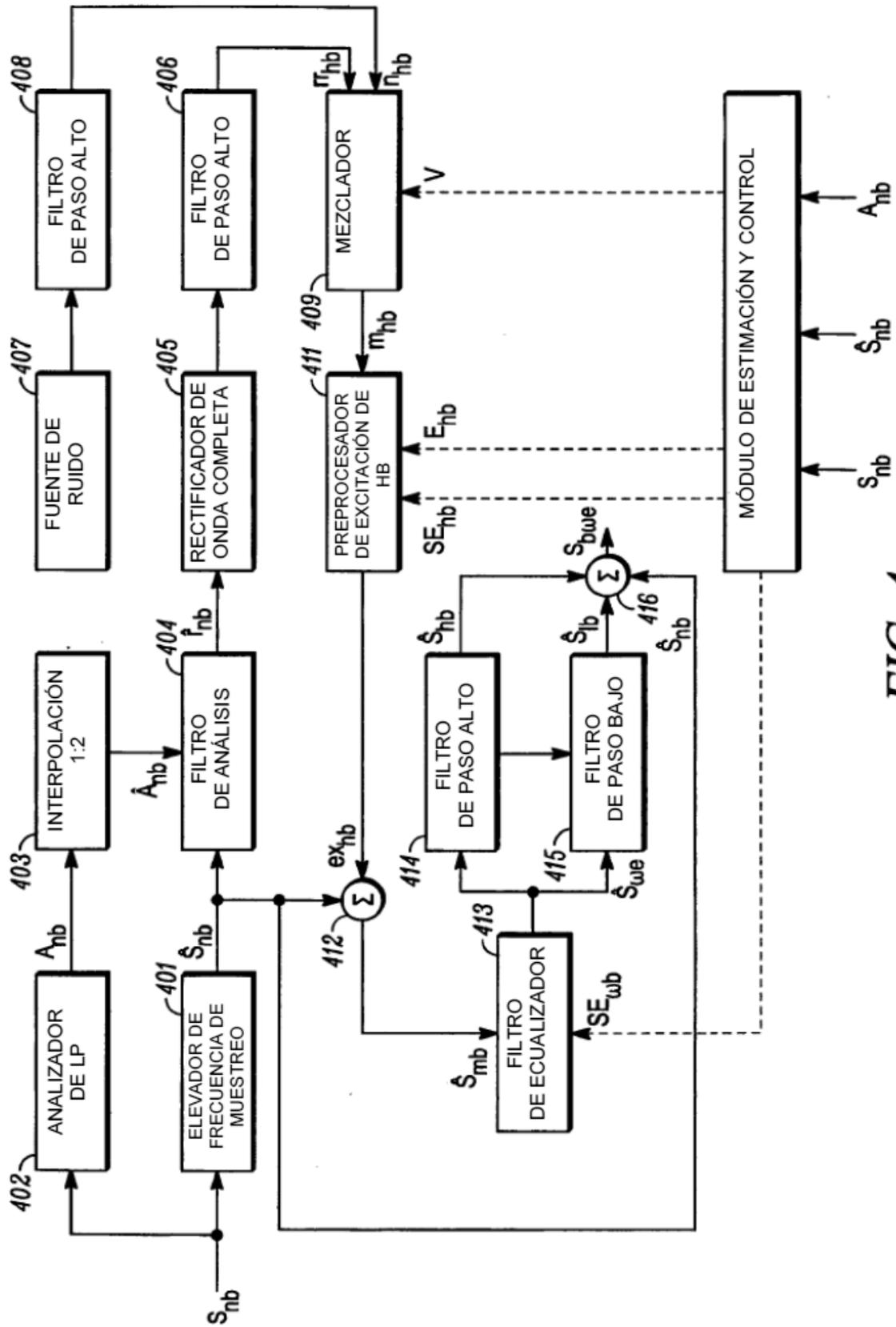


FIG. 4

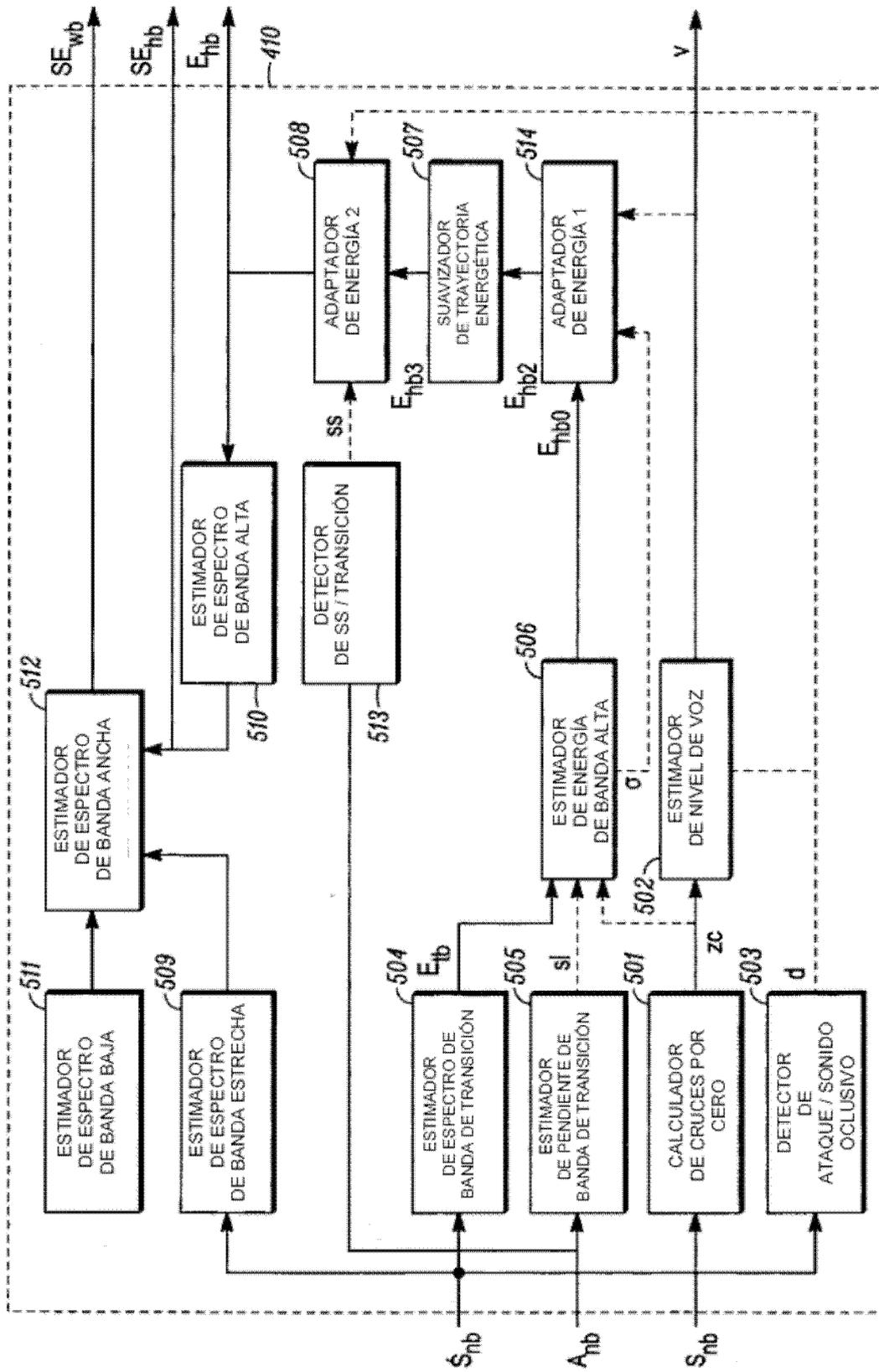


FIG. 5

600

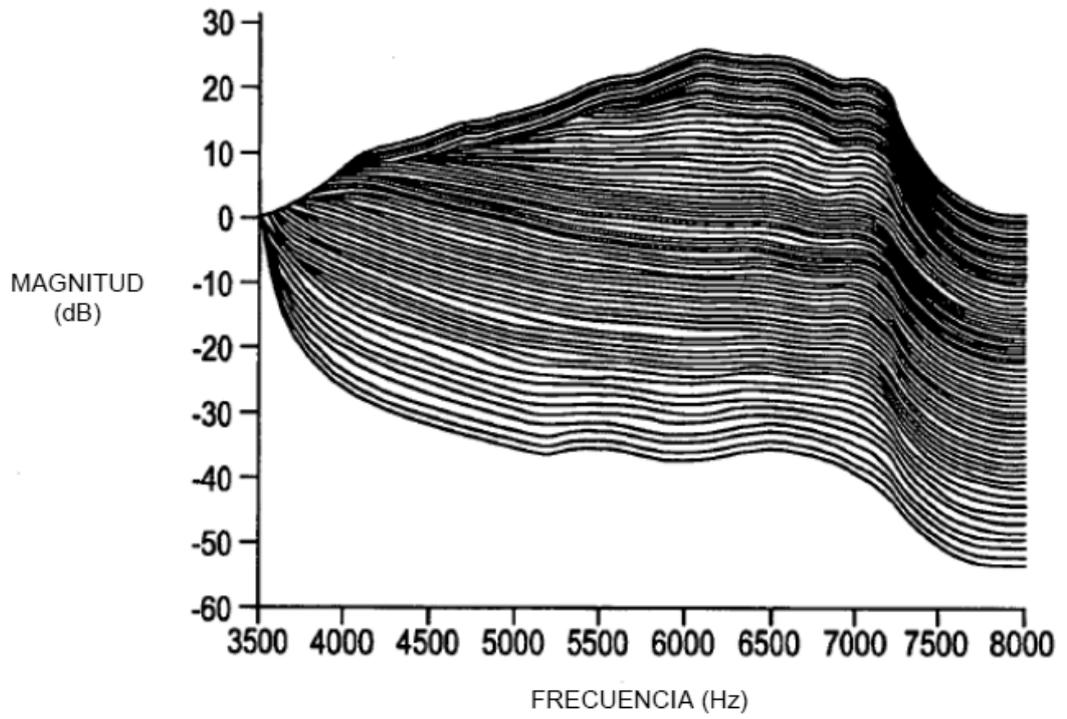


FIG. 6