

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 384 084**

51 Int. Cl.:
G10L 21/02 (2006.01)
G10L 11/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09708900 .7**
- 96 Fecha de presentación: **28.01.2009**
- 97 Número de publicación de la solicitud: **2238594**
- 97 Fecha de publicación de la solicitud: **13.10.2010**

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

30 Prioridad:
01.02.2008 US 24620

45 Fecha de publicación de la mención BOPI:
29.06.2012

45 Fecha de la publicación del folleto de la patente:
29.06.2012

73 Titular/es:
**Motorola Mobility, Inc.
600 North US Highway 45
Libertyville, IL 60048 , US**

72 Inventor/es:
**RAMABADRAN, Tenkasi V. y
JASIUK, Mark A.**

74 Agente/Representante:
de Elzaburu Márquez, Alberto

ES 2 384 084 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 de ancho de banda.

Solicitud Relacionada

5 Esta solicitud está relacionada con el número de solicitud de patente U.S. en tramitación y de co-propiedad, titulada 11/946.978 y clasificada el 29 de noviembre de 2007.

Campo técnico

Esta invención se refiere generalmente a reproducir contenido audible y más particularmente a técnicas de extensión de ancho de banda.

Antecedentes

10 La reproducción audible de contenido audio de una representación digital comprende un área conocida de esfuerzo. En algunos ajustes de la solicitud la representación digital comprende un ancho de banda correspondiente completo que pertenece a una muestra de audio original. En tal caso, la reproducción audible puede comprender una salida de sonido altamente precisa y natural. Tal planteamiento, no obstante, requiere considerables recursos de sobredimensionamiento para acomodar la cantidad de datos correspondiente. En muchos ajustes de la aplicación, 15 tales como, por ejemplo, ajustes de comunicación inalámbrica, tal cantidad de información no siempre se puede soportar adecuadamente.

Para acomodar tal limitación, las denominadas técnicas de habla de banda estrecha pueden servir para limitar la cantidad de información limitando, a su vez, la representación a menos que el ancho de banda completo correspondiente que pertenece a una muestra de audio original. Pero como un ejemplo en este sentido, mientras el 20 habla natural incluye componentes significativos hasta 8kHz (o mayores), una representación de banda estrecha puede proporcionar solamente información con respecto, digamos, a la gama de 300 – 3.400 Hz. El contenido resultante, cuando se reproduce audible, típicamente es suficientemente inteligible para soportar las necesidades funcionales de la comunicación basada en habla. Desafortunadamente, no obstante, el procesamiento de habla de banda estrecha también tiende a producir habla que suena apagada y puede incluso haber reducido la inteligibilidad comparado con el habla de banda completa. 25

Para cumplir esta necesidad, las técnicas de extensión de ancho de banda son a veces empleadas. Una artificialmente genera la información que se omite en las bandas más altas y/o más bajas en base a la información de banda estrecha disponible así como otra información para seleccionar la información que puede ser añadida al contenido de banda estrecha para sintetizar por ello una señal de banda pseudo ancha (o completa). Usando tales 30 técnicas, por ejemplo, uno puede transformar el habla de banda estrecha en la gama de 300 – 3400 Hz para habla de banda ancha, digamos, en la gama de 100 – 8000 Hz. Con este fin, una pieza crítica de información que es requerida es la envoltura espectral en la banda alta (3400 – 8000 Hz). Si se estima la envoltura espectral de banda ancha, la envoltura espectral de banda alta entonces puede normalmente ser fácilmente extraída de ella. Uno puede pensar de la envoltura espectral de banda alta como comprendida de una forma y una ganancia (o 35 equivalentemente, energía).

Mediante un planteamiento, por ejemplo, la forma de envoltura espectral de banda alta es estimada estimando la envoltura espectral de banda ancha a partir de la envoltura espectral de banda estrecha a través de la asignación del libro de códigos. La energía de banda alta entonces se estima ajustando la energía dentro de la sección de banda estrecha de la envoltura espectral de banda ancha para hacer coincidir la energía de la envoltura 40 espectral de banda estrecha. En este planteamiento, la forma de la envoltura espectral de banda alta determina la energía de banda alta y cualquier error en la estimación de la forma también afectará correspondientemente las estimaciones de la energía de banda alta.

En otro planteamiento, la forma de la envoltura espectral de banda alta y la energía de banda alta se estiman separadamente, y la envoltura espectral de banda alta que finalmente se usa es ajustada para hacer coincidir la energía de banda alta estimada. Mediante un planteamiento relacionado la energía de banda alta estimada es usada, junto con otros parámetros, para determinar la forma de la envoltura espectral de banda alta. No obstante, la envoltura espectral de banda alta resultante no se asegura necesariamente que tenga la energía de banda alta apropiada. Un paso adicional se requiere por lo tanto para ajustar la energía de la envoltura espectral de banda alta al valor estimado. A menos que se tenga especial cuidado, este planteamiento provocará una discontinuidad en 50 la envoltura espectral de banda ancha en el límite entre la banda estrecha y la banda alta. Aunque los planteamientos existentes para extensión de ancho de banda, y, en particular, para estimación de envoltura de banda alta son razonablemente exitosos, no necesariamente producen habla resultante de calidad adecuada en al menos algunos ajustes de aplicaciones.

Para generar habla de ancho de banda extendido de calidad aceptable, el número de artefactos en tal habla debería ser minimizado. Se conoce que sobre estimación de energía de banda alta provoca artefactos molestos. La estimación incorrecta de la forma de envoltura espectral de banda alta también puede conducir a artefactos pero 55

estos artefactos son normalmente más leves y son fácilmente enmascarados por el habla de banda estrecha.

De acuerdo con el documento de la técnica anterior US 7.181.402 B2, se conoce un método de estimación de energía de banda alta basado en la energía de la banda estrecha entera la cual podría causar una sobreestimación potencial de la energía.

5 Breve descripción de los dibujos

Las necesidades anteriores se cumplen al menos parcialmente a través del suministro del método y aparato para estimar energía de banda alta en un sistema de extensión de ancho de banda descrito en la siguiente descripción detallada. Las figuras anexas donde números referencia similares se refieren a elementos idénticos o funcionalmente similares en todas las vistas separadas y las cuales junto con la descripción detallada de más adelante son incorporadas en y forman parte de la especificación, sirven para ilustrar además diversas realizaciones y para explicar diversos principios y ventajas todas de acuerdo con la presente invención.

10

La FIG. 1 comprende un diagrama de flujo como se configura de acuerdo con diversas realizaciones de la invención;

La FIG. 2 comprende un gráfico como se configura de acuerdo con diversas realizaciones de la invención;

15 La FIG. 3 comprende un diagrama de bloques como se configura de acuerdo con diversas realizaciones de la invención;

La FIG. 4 comprende un diagrama de bloques como se configura de acuerdo con diversas realizaciones de la invención;

La FIG. 5 comprende un diagrama de bloques como se configura de acuerdo con diversas realizaciones de la invención; y

20 La FIG. 6 comprende un gráfico como se configura de acuerdo con diversas realizaciones de la invención.

Los técnicos expertos apreciarán que los elementos en las figuras son ilustrados por simplicidad y claridad y no han sido necesariamente dibujados a escala. Por ejemplo, las dimensiones y/o posicionamiento relativo de algunos de los elementos en las figuras pueden ser exageradas en relación con otros elementos para ayudar a mejorar la comprensión de diversas realizaciones de la presente invención. También, elementos comunes pero bien entendidos que son útiles o necesarios en una realización factible comercialmente a menudo no se representan para facilitar una vista menos obstruida de estas diversas realizaciones de la presente invención. Se apreciará además que ciertas acciones y/o pasos pueden ser descritos o representados en un orden particular de ocurrencia mientras que aquellos expertos en la técnica entenderán que tal especificidad con respecto a la secuencia no es realmente requerida. También se entenderá que los términos y expresiones usados aquí dentro tienen el significado técnico habitual como es acordado para tales términos y expresiones por las personas expertas en el campo técnico como se establece en adelante anteriormente excepto donde se han establecido de otra manera aquí dentro diferentes significados específicos.

25

30

Resumen de la invención

La invención se define por el método según la reivindicación 1 y el aparato según la reivindicación 8.

35 Descripción detallada

Las enseñanzas tratadas aquí dentro están dirigidas a un método y sistema rentables para extensión de ancho de banda artificial. De acuerdo con tales enseñanzas, se recibe una señal audio digital de banda estrecha. La señal de audio digital de banda estrecha puede ser una señal recibida a través de una estación móvil en una red celular, por ejemplo, y la señal de audio digital de banda estrecha puede incluir habla en la gama de frecuencias de 300 – 3400 Hz. Las técnicas de extensión de ancho de banda artificial se implementan para esparcir 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 la extensión de ancho de banda artificial para esparcir el espectro para incluir frecuencias de banda baja y banda alta, se crea una señal de audio digital de sonido más natural que es más placentera para un usuario de una estación móvil que implementa la técnica.

40

45 En las técnicas de extensión de ancho de banda artificial, la información omitida en las bandas más altas (3400 – 8000 Hz) y más bajas (100 – 300 Hz) son generadas artificialmente en base a la información de banda estrecha disponible así como la información *a priori* derivada y almacenada desde una base de datos de habla y añadida a la señal de banda estrecha para sintetizar una señal de banda pseudo ancha. Tal solución es bastante atractiva porque requiere cambios mínimos a un sistema de transmisión existente. Por ejemplo, no se necesita tasa de bit adicional.

50 La extensión de ancho de banda artificial 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 habla usada en el sistema de comunicación o la naturaleza del sistema de comunicación en sí mismo, por ejemplo, analógica, digital, de línea terrestre, o celular. Por ejemplo, las técnicas de extensión de ancho de banda artificial pueden ser implementadas

por una estación móvil que recibe una señal de audio digital de banda estrecha, y la señal de banda amplia resultante es utilizada para generar audio reproducido a un usuario de la estación móvil.

En la determinación de la información de banda alta, se estima primero la energía en la banda alta. 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 próximo a las frecuencias de banda alta generalmente tiene la correlación más alta con la señal de banda alta. Por consiguiente, solamente se utiliza un subconjunto de la banda estrecha, opuesto a la banda estrecha entera, para estimar la energía de banda alta. El subconjunto que es usado se conoce como la "banda de transición" y puede incluir frecuencias tales como 2500 – 3400 Hz. Más específicamente, la banda de transición es definida aquí dentro como una banda de frecuencias que está contenida dentro de la banda estrecha y es cercana a la banda alta, es decir, sirve como una transición para la banda alta. Este planteamiento está en contraste con los sistemas de extensión de ancho de banda de la técnica anterior los cuales estiman la energía de banda alta en términos de la energía en la banda estrecha entera, típicamente como una relación.

Para estimar la energía de banda alta, se estima primero la energía de la banda de transición mediante técnicas tratadas más adelante con respecto a FIG. 4 y 5. Por ejemplo, la energía de la banda de transición de la banda de transición puede ser calculada primero sobre-muestreando una señal de banda estrecha de entrada, calculando el espectro de frecuencia de la señal de banda estrecha sobre-muestreada, y luego sumando las energías de las componentes espectrales en la banda de transición. La energía de la banda de transición estimada se inserta posteriormente en una ecuación polinómica como una variable independiente para estimar la energía de banda alta. Los coeficientes o pesos de las diferentes potencias de la variable independiente en la ecuación polinómica incluyendo aquel de potencia cero, es decir, el término constante, se seleccionan para minimizar el error cuadrático medio entre valores verdaderos y estimados de la energía de banda alta sobre un gran número de tramas de una base de datos de entrenamiento de habla. La precisión de la estimación se puede mejorar además acondicionando la estimación en parámetros derivados de la señal de banda estrecha así como parámetros derivados de la señal de banda de transición como se trata en más detalle más adelante. Después de que se ha estimado la energía de banda alta, el espectro de banda alta se estima en base a la estimación de energía de banda alta.

Utilizando la banda de transición de esta manera, se proporciona una técnica de extensión de ancho de banda robusta que produce una señal de audio correspondiente de mayor calidad que sería posible si la energía en la banda estrecha entera fuera usada para estimar la energía de banda alta. Además, esta técnica puede ser utilizada sin afectar demasiado adversamente los sistemas de comunicaciones existentes porque las técnicas de extensión de ancho de banda se aplican a una señal de banda estrecha recibida a través del sistema de comunicación, es decir, los sistemas de comunicación existentes pueden ser utilizados 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. Primero, en la operación 101, se recibe una señal de audio digital de banda estrecha. En un ajuste de aplicación típico, esto comprenderá proporcionar una pluralidad de tramas de tal contenido. Estas enseñanzas acogerán en seguida procesar cada tal trama por los pasos descritos. Mediante un planteamiento, por ejemplo, cada tal trama puede corresponder a 10 – 40 milisegundos de contenido audio original.

Esto puede comprender, por ejemplo, proporcionando una señal audio digital que comprende contenido vocal sintetizado. Tal es el caso, por ejemplo, cuando se emplean estas enseñanzas en conjunto con el contenido de habla codificado por voz recibido en un dispositivo de comunicaciones inalámbricas portátil. Otras posibilidades existen también, no obstante, como será bien entendido por aquellos expertos en la técnica. Por ejemplo, la señal de audio digital puede comprender en su lugar una señal de habla original o una versión vuelta a muestrear de o bien una señal de habla original o bien un contenido de habla sintetizado.

Con referencia de manera momentánea a la FIG. 2, se entenderá que esta señal de audio digital pertenece a alguna señal de audio original 201 que tiene un ancho de banda de señal correspondiente original 202. Este ancho de banda de señal correspondiente original 202 típicamente será más grande que el ancho de banda de la señal antes mencionado como corresponde a la señal de audio digital. Esto puede ocurrir, por ejemplo, cuando la señal de audio digital representa sólo una parte 203 de la señal de audio original 201 con otras partes que se dejan fuera de banda. En el ejemplo ilustrativo mostrado, esto incluye una parte de banda baja 204 y una parte de banda alta 205. Aquellos expertos en la técnica reconocerán que este ejemplo sirve solamente para un propósito ilustrativo y que la parte no representada solamente puede comprender una parte de banda baja o una parte de banda alta. Estas enseñanzas también serían aplicables para uso en un ajuste de aplicación en que la parte no representada cae en medio de la banda en dos o más partes representadas (no mostradas).

Se entenderá por lo tanto fácilmente que la(s) parte(s) no representada(s) de la señal de audio original 201 comprende(n) contenido que estas presentes enseñanzas pueden buscar razonablemente para reemplazar o de otra manera representar de alguna manera razonable y aceptable. También se comprenderá que este ancho de banda de señal ocupa solamente una parte del ancho de banda de Nyquist determinado por frecuencia de muestreo relevante. Esto, a su vez, se entenderá para proporcionar además una región de frecuencia en la cual efectuar la extensión de ancho de banda deseada.

Con referencia de nuevo a la FIG. 1, la señal de audio digital de entrada se procesa 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 sobre-muestreo. Mediante otro planteamiento, puede ser un sistema de ganancia 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 se estima en base a una banda de transición de la señal de audio digital procesada dentro de una gama de frecuencia superior predeterminada de un ancho de banda de banda estrecha.

Usando las componentes de banda de transición como la base para la estimación, se obtiene una estimación más precisa que sería posible generalmente si todas las componentes de banda estrecha fueran usadas colectivamente para estimar el valor de energía de las componentes de banda alta. Mediante un planteamiento, el valor de la energía de banda alta se usa para acceder 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 la envolvente espectral de banda alta apropiada al nivel de energía correcto.

Este proceso 100 entonces se acomodará opcionalmente combinando 104 la señal de audio digital con el contenido de banda alta correspondiente con el valor de energía estimado y al espectro de las componentes de banda alta para proporcionar una versión extendida de ancho de banda de la señal de audio digital de banda estrecha a ser reproducida. Aunque el proceso mostrado en la FIG. 1 solamente ilustra añadir las componentes de banda alta estimadas, se debería apreciar que las componentes de banda baja también se pueden estimar y combinar con la señal de audio digital de banda estrecha para generar una señal de banda ancha de ancho de banda extendido.

La señal de audio de ancho de banda extendido resultante (obtenida combinando la señal de audio digital de entrada con el contenido de ancho de banda fuera de la señal generado artificialmente) tiene una calidad de audio mejorada frente a la señal de audio digital de banda estrecha original cuando se reproduce de forma audible. Mediante un planteamiento, esto puede comprender combinar dos elementos que son mutuamente exclusivos con respecto a su contenido espectral. En tal caso, tal combinación puede tomar la forma, por ejemplo, de simplemente concatenar o de otra manera unir los dos (o más) segmentos juntos. Mediante otro planteamiento, si se desea, el contenido de ancho de banda de banda alta y/o banda baja puede tener una parte que está dentro del correspondiente ancho de banda de señal de la señal de audio digital. Tal solapamiento puede ser útil en al menos algunos ajustes de aplicaciones para suavizar y/o poner horizontal la transición desde una parte a la otra combinando la parte de solapamiento del contenido de ancho de banda de banda alta y/o banda baja con la correspondiente parte en banda de la señal audio digital.

Aquellos expertos en la técnica apreciarán que los procesos anteriormente descritos rápidamente se habilitan usando cualquiera de una variedad amplia de plataformas disponibles y/o configuradas fácilmente, que incluyen parcialmente o completamente plataformas programables según se conocen en la técnica o plataformas de propósito dedicado como se puede desear para algunas aplicaciones. Con referencia ahora a la FIG. 3, se proporcionará ahora un planteamiento ilustrativo para tal plataforma.

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

El procesador 301, a su vez, puede ser configurado y adaptado (a través de, por ejemplo, la programación correspondiente cuando el procesador 301 comprende una plataforma parcialmente o completamente programable como se conoce en la técnica) para llevar a cabo uno o más de los pasos u otra funcionalidad fijada en adelante aquí dentro. Esto puede comprender, por ejemplo, estimar el valor de energía de banda alta de la energía a partir de la energía de la banda de transición y entonces usar el valor de energía de la banda alta y un conjunto de formas de índice de energía para determinar la envolvente espectral de banda alta.

Como se describió anteriormente, mediante un planteamiento, el valor de energía de banda alta anteriormente mencionado puede servir para facilitar acceso a una tabla de búsqueda que contiene una pluralidad de formas de envolventes espectrales candidatas correspondientes. Para soportar tal planteamiento, este aparato también puede comprender, si se desea, una o más tablas de búsqueda 304 que están acopladas operativamente al procesador 301. Así configurado, el procesador 301 puede acceder fácilmente a la tabla de búsqueda 304 como es adecuado.

Aquellos expertos en la técnica reconocerán y entenderán que tal aparato 300 puede ser compuesto de una pluralidad de elementos físicamente distintos como se sugiere por la ilustración mostrada en la FIG. 3. También es posible, no obstante, ver esta ilustración como que comprende una vista lógica, en cuyo caso uno o más de estos elementos pueden ser habilitados y realizados a través de una plataforma compartida. También se entenderá que tal plataforma compartida puede comprender una plataforma completamente o al menos parcialmente programable como se conocen en la técnica.

Se debería apreciar que el procesamiento tratado anteriormente se puede realizar por una estación móvil en comunicación inalámbrica con una estación base. Por ejemplo, la estación base puede transmitir la señal de audio

digital de banda estrecha a través de medios convencionales a la estación móvil. Una vez recibida, el(los) procesador(es) dentro de la estación móvil realiza(n) las operaciones precisas para generar una versión extendida de ancho de banda de la señal de audio digital que es más clara y más audiblemente placentera para un usuario de la estación móvil.

5 Con referencia ahora a la FIG. 4, el habla de banda estrecha de entrada s_{nb} muestreada a 8 kHz primero se sobre muestrea por 2 usando un sobre muestreador 401 correspondiente para obtener habla de banda estrecha sobre muestreada \dot{s}_{nb} muestreada a 16 kHz. Esto puede comprenderse realizando una interpolación 1:2 (por ejemplo, insertando una muestra de valor cero entre cada par de muestras de habla original) seguida por un filtrado paso bajo usando, por ejemplo, un filtro paso bajo (LPF) que tiene una banda de paso entre 0 y 3400 Hz.

10 A partir de s_{nb} , los parámetros predictivos lineales de banda estrecha (LP), $A_{nb} = \{1, a_1, a_2, \dots, a_P\}$ donde P es el orden del modelo, también se calculan usando un analizador LP 402 que emplea técnicas de análisis LP bien conocidas. (Existen otras posibilidades, por supuesto; por ejemplo, los parámetros LP pueden ser calculados a partir de una versión diezmada 2:1 de \dot{s}_{nb}). Estos parámetros LP modelan la envolvente espectral del habla de entrada de banda estrecha como:

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

En la ecuación anterior, la frecuencia angular ω en radianes/muestra es 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 modelo conveniente P, por ejemplo, es 10.

20 Los parámetros LP A_{nb} entonces se interpolan por 2 usando un módulo de interpolación 403 para obtener $\dot{A}_{nb} = \{1, 0, a_1, 0, a_2, 0, \dots, 0, a_P\}$. Usando \dot{A}_{nb} , el habla de banda estrecha sobre muestreada \dot{s}_{nb} se filtra inversamente usando un filtro de análisis 404 para obtener la señal residual LP \dot{r}_{nb} (la cual también se muestrea a 16 kHz). Mediante un planteamiento, esta operación de filtrado (o análisis) inverso puede ser descrita por la ecuación

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

25 donde n es el índice de muestra.

En un ajuste típico de aplicación, el filtrado inverso de \dot{s}_{nb} para obtener \dot{r}_{nb} puede ser hecho de una forma trama por trama donde se define una trama como una secuencia de N muestras consecutivas sobre una duración de T segundos. Para muchas aplicaciones de señal de habla, una buena elección para T es alrededor de 20 ms con valores correspondientes para N de alrededor de 160 en 8 kHz y alrededor de 320 en 16 kHz de frecuencia de muestreo. Las tramas sucesivas pueden solaparse una con otra, por ejemplo, hasta o alrededor del 50%, en cuyo caso, la segunda mitad de las muestras en la trama actual y la primera mitad de las muestras en la trama siguiente son la misma, y una nueva trama es procesada cada T/2 segundos. Para una elección de T como 20 ms y 50% de solapamiento, por ejemplo, los parámetros LP A_{nb} se calculan a partir de 160 muestras s_{nb} consecutivas cada 10 ms, y son usadas para filtrar a la inversa las 160 muestras del medio de la correspondiente trama \dot{s}_{nb} de 320 muestras para producir 160 muestras de \dot{r}_{nb} .

Uno también puede calcular los parámetros LP de orden 2P para la operación de filtrado inverso directamente desde el habla de banda estrecha sobre muestreada. Este planteamiento, no obstante, puede aumentar la complejidad tanto de calcular los parámetros LP como la operación de filtrado inverso, sin aumentar necesariamente el rendimiento bajo al menos algunas condiciones de operación.

La señal LP residual \dot{r}_{nb} es a continuación rectificadora a onda completa usando un rectificador de onda completa 405 y filtrando paso alto el resultado (usando, por ejemplo, un filtro paso alto (HPF) 406 con un paso banda 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 paso alto 408 para obtener la señal de ruido de banda alta r_{hb} .

Alternativamente, una secuencia de ruido filtrado paso alto puede ser pre almacenada en un almacenador temporal (tal como, por ejemplo, un almacenador temporal circular) y accedida como se requiera para generar n_{hb} . El uso de tal almacenador temporal elimina los cálculos asociados con el filtrado paso alto de las muestras de ruido pseudo aleatorio en tiempo real. Estas dos señales, viz., rr_{hb} y n_{hb} , entonces se mezclan en un mezclador 409 de acuerdo con el nivel de voz v proporcionado por un Módulo de Estimación y Control (ECM) 410 (cuyo módulo será descrito en más detalle más adelante). En este ejemplo ilustrativo, este nivel de voz v oscila de 0 a 1, con 0 que indica un nivel sin voz y 1 que indica un nivel de voz completo. El mezclador 409 esencialmente forma una suma ponderada de las dos señales de entrada en su salida después de asegurar que las dos señales de entrada están ajustadas para tener el mismo nivel de energía. La señal de salida del mezclador m_{hb} es dada por:

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

Aquellos expertos en la técnica apreciarán que también son posibles otras reglas de mezcla. También es posible mezclar primero las dos señales, viz., la señal residual LP rectificadas de onda completa y la señal de ruido pseudo aleatoria, y después filtrar paso alto la señal mezclada. En este caso, los dos filtros paso alto 406 y 408 se sustituyen por un único filtro de paso alto situado en la salida del mezclador 409.

La señal resultante m_{hb} entonces se pre procesa usando un preprocesador de excitación de banda alta (HB) 411 para formar la señal de excitación de banda alta ex_{hb} . Los pasos de pre procesamiento pueden comprender: (i) escalar la señal de salida del mezclador m_{hb} para hacer coincidir el nivel de energía de banda alta E_{hb} , y (ii) opcionalmente dar forma a la señal de salida del mezclador m_{hb} para hacer coincidir la envolvente espectral de banda alta SE_{hb} . Tanto E_{hb} como SE_{hb} se proporcionan al preprocesador de excitación HB 411 por el ECM 410. Cuando se emplea este planteamiento, puede ser útil en muchos ajustes de aplicación asegurar que tal formación no afecta al espectro de fase de la señal de salida del mezclador m_{hb} ; es decir, la formación se puede realizar preferentemente mediante un filtro de repuesta de fase cero.

La señal de habla de banda estrecha sobre muestreada \hat{s}_{nb} y la señal de excitación de banda alta ex_{hb} son añadidas juntas usando un sumador 412 para formar la señal de banda mezclada \hat{s}_{mb} . Esta señal de banda mezclada resultante \hat{s}_{mb} es introducida en un filtro ecualizador 413 que filtra esa entrada usando información de envolvente espectral de banda amplia SE_{wb} proporcionada por el ECM 410 para formar la señal de banda amplia estimada \hat{s}_{wb} . El filtro ecualizador 413 esencialmente impone la envolvente espectral de banda amplia SE_{wb} en la señal de entrada \hat{s}_{mb} para formar \hat{s}_{wb} . (La discusión adicional a este respecto aparece más adelante). La señal de banda ancha estimada resultante \hat{s}_{wb} es filtrada paso alto, por ejemplo, usando un filtro paso alto 414 que tiene una banda de paso desde 3400 a 8000 Hz, y filtrada paso bajo, por ejemplo, usando un filtro paso bajo 415 que tiene una banda de paso desde 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 sobre muestreada \hat{s}_{nb} son añadidas juntas en otro sumador 416 para formar la señal extendida de ancho de banda s_{bwe} .

Aquellos expertos en la técnica apreciarán que hay varias otras configuraciones de filtro posibles para obtener la señal extendida de ancho de banda s_{bwe} . Si el filtro ecualizador 413 retiene con precisión el contenido espectral de la señal de habla de banda estrecha sobre muestreada \hat{s}_{nb} la cual es parte de su señal de entrada \hat{s}_{mb} , entonces la señal de banda ancha estimada \hat{s}_{wb} se puede sacar directamente como la señal extendida de ancho de banda s_{bwe} eliminando por ello el filtro paso alto 414, el filtro paso bajo 415, y el sumador 416. Alternativamente, se pueden usar dos filtros ecualizadores, uno para recuperar la parte de baja frecuencia y otro para recuperar la parte de alta frecuencia, y la salida de la primera se puede añadir a la salida filtrada paso alto de la última para obtener la señal extendida de ancho de banda s_{bwe} .

Aquellos expertos en la técnica entenderán y apreciarán que, con este ejemplo ilustrativo particular, la excitación residual rectificadas de banda alta y la excitación de ruido de banda alta se mezclan juntas de acuerdo con el nivel de voz. Cuando el nivel de voz es 0 que indica habla sin voz, es usada exclusivamente la excitación de ruido. De manera similar, cuando el nivel de voz es 1 que indica habla con voz, se usa exclusivamente la excitación residual rectificadas de banda alta. Cuando el nivel de voz está entre 0 y 1 que indica habla con voz mezclada, las dos excitaciones se mezclan en proporción adecuada según se determina por el nivel de voz y se usan. La excitación de banda alta mezclada es adecuada de esta manera para sonidos con voz, sin voz y con voz mezclada.

Además se entenderá y apreciará que, en este ejemplo ilustrativo, se usa un filtro ecualizador para sintetizar \hat{s}_{wb} . El filtro 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 hacer coincidir con la ideal. Dado que solamente están implicadas las magnitudes en la ecualización de la envolvente espectral, la respuesta en fase del filtro ecualizador se elige que sea cero. La respuesta en magnitud del filtro ecualizador se especifica por $SE_{wb}(\omega)/SE_{mb}(\omega)$. El diseño e implementación de tal filtro ecualizador para una aplicación de codificación de habla comprende un área bien entendida de esfuerzo. Brevemente, no obstante, el filtro ecualizador funciona como sigue usando análisis de superposición-adición (OLA).

La señal de entrada \hat{s}_{mb} primero se divide en tramas de solapamiento, por ejemplo, tramas de 20 ms (320 muestras a 16 kHz) con solapamiento del 50%. Cada trama de muestras entonces se multiplica (punto a punto) por una ventana adecuada, por ejemplo, una ventana de coseno alzado con propiedad de reconstrucción perfecta. La trama de habla de ventana a continuación se analiza para estimar los parámetros LP que modelan su envolvente espectral. La envolvente espectral de banda ancha ideal para la trama se proporciona por el ECM. A partir de las dos envolventes espectrales, el ecualizador calcula la respuesta en magnitud del filtro como $SE_{wb}(\omega)/SE_{mb}(\omega)$ y fija la respuesta de fase a cero. La trama de entrada entonces se ecualiza para obtener la trama de salida correspondiente. Las tramas de salida ecualizadas finalmente se añaden solapadas para sintetizar el habla de banda ancha estimada \hat{s}_{wb} .

Aquellos expertos en la técnica apreciarán que además del análisis LP, hay otros métodos para obtener la envolvente espectral de una trama de habla dada, por ejemplo, análisis cepstral, adecuación de curva de orden más alto o lineal por tramos de picos de magnitud, etc.

Aquellos expertos en la técnica también apreciarán que en lugar de hacer ventanas con la señal de entrada \hat{s}_{mb} directamente, uno podría haber comenzado con las versiones de ventanas de \hat{s}_{nb} , r_{nb} , y n_{nb} , para lograr el mismo resultado. También puede ser conveniente mantener el tamaño de la trama y el porcentaje de solapamiento para el filtro ecualizador el mismo que aquel usado en el bloque de filtro de análisis usado para obtener \hat{r}_{nb} a partir de \hat{s}_{nb} .

El planteamiento de filtro ecualizador descrito para sintetizar \hat{s}_{wb} ofrece una serie de ventajas: i) Dado que la respuesta en fase del filtro ecualizador 413 es cero, las componentes de frecuencia diferentes de la salida del ecualizador están alineadas en el tiempo con las componentes correspondientes de la entrada. Esto puede ser útil para habla con voz porque los segmentos de energía alta (tales como los segmentos de pulso glotales) de la excitación de banda alta residual rectificadas $e_{x_{hb}}$ están alineados en el tiempo con los segmentos de alta energía correspondientes del habla de banda estrecha sobre muestreada \hat{s}_{nb} en la entrada del ecualizador, y la conservación de esta alineación en el tiempo en la salida del ecualizador a menudo actuará para asegurar buena calidad de habla; ii) la entrada al filtro ecualizador 413 no necesita tener un espectro plano como en el caso del filtro de síntesis LP; iii) el filtro ecualizador 413 se especifica en el dominio de la frecuencia, y por lo tanto es factible un mejor control y más fino sobre diferentes partes del espectro; y iv) son posibles iteraciones para mejorar la efectividad de filtrado a costa de complejidad y retardo adicionales (por ejemplo, la salida del ecualizador se puede realimentar a la entrada para ser de nuevo ecualizada y de nuevo mejorar el rendimiento).

Algunos detalles adicionales con respecto a la configuración descrita se presentarán ahora.

Pre-procesamiento de excitación de banda alta: La respuesta en magnitud del filtro ecualizador 413 se da por $SE_{wb}(\omega)/SE_{mb}(\omega)$ y su respuesta en fase se puede fijar a cero. Cuanto más próxima esté la envolvente espectral de entrada $SE_{mb}(\omega)$ a la envolvente espectral ideal $SE_{wb}(\omega)$, más fácil es para el ecualizador corregir la envolvente espectral de entrada para hacer coincidir con la ideal. Al menos una función del preprocesador de excitación de banda alta 411 va a mover $SE_{mb}(\omega)$ más cerca a $SE_{wb}(\omega)$ y de esta manera hacer más fácil el trabajo del filtro ecualizador 413. Primero, esto se hace subiendo la señal de salida del mezclador m_{hb} al nivel de energía de banda alta correcto E_{hb} proporcionado por el ECM 410. Segundo, la señal de salida del mezclador m_{hb} se da forma opcionalmente de manera que su envolvente espectral se hace coincidir con la envolvente espectral de banda alta SE_{hb} proporcionada por el ECM 410 sin afectar su espectro de fase. Un segundo paso puede comprender esencialmente un paso de pre ecualización.

Excitación de banda baja: A diferencia de la pérdida de información en la banda alta causada 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 de limitación de banda de la función de transferencia de canal que consta de, por ejemplo, un micrófono, amplificador, codificador de habla, canal de transmisión, o similares. Consecuentemente, en una señal de banda estrecha limpia, la información de banda baja aún está presente a un nivel muy bajo. Esta información de bajo nivel se puede amplificar de una manera hacia delante para restaurar la señal original. Pero se debería tener cuidado en este proceso dado que las señales de nivel bajo se corrompen fácilmente 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. Es decir, la señal de excitación de banda baja se puede formar mezclando la señal residual rectificadas de banda baja r_{nb} y la señal de ruido de banda baja n_{nb} de una manera similar a la formación de la señal de salida del mezclador de banda alta m_{hb} .

Con referencia ahora a la FIG. 5, el Módulo de Estimación y Control (ECM) 410 toma como entrada el habla de banda estrecha S_{nb} , el habla de banda estrecha sobre muestreada \hat{S}_{nb} , y los parámetros 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} .

- 5 Estimación del nivel de voz: Para estimar el nivel de voz, un calculador de cruce con cero 501 calcula el número de cruces con cero z_c en cada trama del habla 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

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

- 10 n es el índice de muestras, y N es el tamaño de trama en muestras. Es conveniente mantener el tamaño de trama y el porcentaje de solapamiento usado en el ECM 410 el mismo que aquel usado en el filtro ecualizador 413 y los bloques de filtro de análisis, por ejemplo, $T = 20$ ms, $N = 160$ para muestreo de 8 kHz, $N = 320$ para muestreo de 16 kHz, y 50% de solapamiento con referencia a los valores ilustrativos presentados anteriormente. El valor del parámetro z_c calculado como anteriormente oscila de 0 a 1. A partir del parámetro z_c , un estimador del nivel de voz
15 502 puede estimar el nivel de voz v como sigue.

$$v = \begin{cases} 1 & \text{si } z_c < ZC_{bajo} \\ 0 & \text{si } z_c > ZC_{alto} \\ 1 - \left[\frac{z_c - ZC_{bajo}}{ZC_{alto} - ZC_{bajo}} \right] & \text{de otro modo} \end{cases}$$

- donde, ZC_{bajo} y ZC_{alto} representan respectivamente los umbrales bajo y alto elegidos adecuadamente, por ejemplo, $ZC_{bajo} = 0,40$ y $ZC_{alto} = 0,45$. La salida d de un detector de aparición/oclusiva 503 también se puede alimentar en el detector de nivel de voz 502. Si una trama se marca como que contiene una aparición o una oclusiva con $d = 1$, el nivel de voz de esa trama así como la trama siguiente se puede fijar a 1. Recordar que, mediante un planteamiento, cuando el nivel de voz es 1, se usa exclusivamente la excitación residual rectificada de banda alta. Esto es ventajoso en una aparición/oclusiva, comparado con la excitación de banda alta mezclada o solamente ruido, porque la excitación residual rectificada sigue de manera cercana la energía en función del contorno de tiempo del habla de banda estrecha sobre muestreada reduciendo de esta manera la posibilidad de artefactos de tipo eco previo debidos a la dispersión en el tiempo en la señal extendida de ancho de banda
25

- Para estimar la energía de banda alta, el estimador de energía de banda de transición 504 estima la energía de banda de transmisión a partir de la señal de habla de banda estrecha sobre muestreada \hat{S}_{nb} . La banda de transición se define aquí como una banda de frecuencia que está contenida dentro de la banda estrecha y cerca de la banda alta, es decir, sirve como una transición a la banda alta, (la cual, en este ejemplo ilustrativo, es de alrededor de 2500 – 3400 Hz). Intuitivamente, uno esperaría que la energía de la banda alta esté bien correlacionada con la energía de la banda de transición, lo cual se confirma en los experimentos. Una forma simple de calcular la energía de la banda de transición E_{tb} es calcular el espectro de frecuencia de \hat{S}_{nb} (por ejemplo, a través de una Transformada Rápida de Fourier (FFT)) y sumar las energías de las componentes espectrales dentro de la banda de transición.
30

- A partir de la energía de la banda de transición E_{tb} en dB (decibelios), se estima la energía de la banda alta E_{hb0} en dB como
35

$$E_{hb0} = \alpha E_{tb} + \beta,$$

donde los coeficientes α y β se seleccionan para minimizar el error cuadrático medio entre los valores verdadero y estimado de la energía de banda alta sobre un gran número de tramas de una base de datos de entrenamiento de habla.

- 40 La precisión de la estimación se puede mejorar además explotando la información contextual a partir de los

parámetros de habla adicionales tales como el parámetro de cruce con cero zc y el parámetro de pendiente espectral de banda de transición $s/$ que puede ser proporcionado por un estimador de pendiente de banda de transición 505. El parámetro de cruce con cero, como se trató anteriormente, es indicativo del nivel de voz de habla. El parámetro de pendiente indica la tasa de cambio de energía espectral dentro de la banda de transición. Se puede estimar a partir de los parámetros LP de banda estrecha A_{nb} aproximando la envolvente espectral (en dB) dentro de la banda de transición como una línea recta, por ejemplo, a través de regresión lineal, y calcular su pendiente. El plano de parámetros $zc-s/$ se parte entonces en una serie de regiones, y los coeficientes α y β se seleccionan separadamente para cada región. Por ejemplo, si las gamas de los parámetros zc y $s/$ están divididas cada una en 8 intervalos iguales, el plano de parámetros $zc-s/$ entonces se parte en 64 regiones, y se seleccionan 64 conjuntos de coeficientes α y β , uno para cada región.

Mediante otro planteamiento (no se muestra en la FIG. 5), la mejora adicional en la precisión de la estimación se logra como sigue. Señalar que en lugar del parámetro de pendiente $s/$ (el cual solamente es una representación de primer orden de la envolvente espectral dentro de la banda de transición), se puede emplear una representación de mayor resolución para mejorar el rendimiento del estimador de energía de banda alta. Por ejemplo, se puede usar una representación cuantificada del vector de las formas de envolvente espectral de banda de transición (en dB). Como un ejemplo ilustrativo, el libro de códigos del cuantificador de vectores (VQ) consta de 64 formas conocidas como parámetros de forma de envolvente espectral de banda de transición tbs que se calculan a partir de una base de datos de entrenamiento grande. Uno podría sustituir el parámetro $s/$ en el plano de parámetros $zc-s/$ con el parámetro tbs para lograr mejorar el rendimiento. Mediante otro planteamiento, no obstante, se introduce un tercer parámetro conocido como la medida de uniformidad espectral sfm . La medida de uniformidad espectral se define como la relación de la media geométrica a la media aritmética de la envolvente espectral de banda estrecha (en dB) dentro de una gama de frecuencia adecuada (tal como, por ejemplo, 300 – 3400 Hz). El parámetro sfm indica cómo es de plana la envolvente espectral – oscilando en este ejemplo de alrededor de 0 para una envolvente picuda a 1 para una envolvente completamente plana. El parámetro sfm también se relaciona con el nivel de voz del habla pero de una forma diferente que zc . Mediante un planteamiento, el espacio de parámetros tridimensional $zc-sfm-tbs$ se divide en una serie de regiones como sigue. El plano $zc-sfm$ se divide en 12 regiones dando lugar por ello a $12 \times 64 = 768$ regiones posibles en el espacio tridimensional. No todas de estas regiones, no obstante, tienen suficientes puntos de datos de la base de datos de entrenamiento. Así, para muchos ajustes de aplicaciones, el número de regiones útiles está limitado a alrededor de 500, con un conjunto separado de coeficientes α y β que se seleccionan para cada una de estas regiones.

Un estimador de energía de banda alta 506 puede proporcionar mejora adicional en la precisión de la estimación usando potencias más altas de E_{tb} en la estimación de E_{hb0} , por ejemplo,

$$E_{hb0} = \alpha_4 E_{tb}^4 + \alpha_3 E_{tb}^3 + \alpha_2 E_{tb}^2 + \alpha_1 E_{tb} + \beta.$$

En este caso, se seleccionan cinco coeficientes diferentes, viz., α_4 , α_3 , α_2 , α_1 , y β , para cada partición del plano de parámetros $zc-s/$ (o alternativamente, para cada partición del espacio de parámetros $zc-sfm-tbs$). Dado que las ecuaciones anteriores (referidas a los párrafos 69 a 74) para estimar E_{hb0} son no lineales, se debe tener especial cuidado para ajustar la energía de banda alta estimada según cambia el nivel de señal de entrada, es decir, la energía. Una forma de lograr esto es estimar el nivel de señal de entrada en dB, ajustar E_{tb} arriba o abajo para corresponder con el nivel de señal nominal, estimar E_{hb0} , y ajustar E_{hb0} abajo o arriba para corresponder con el nivel de señal real.

Aunque el método de estimación de energía de banda alta descrito anteriormente funciona bastante bien para la mayoría de las tramas, ocasionalmente hay tramas para las cuales la energía de banda alta está gravemente infra o sobre estimada. Tales errores de estimación se pueden corregir al menos parcialmente por medio de un suavizador de paso de energía 507 que comprende un filtro de suavización. El filtro de suavización se puede diseñar de manera que permita transiciones reales en el paso de energía para pasar a través de transiciones no afectadas, por ejemplo, entre segmentos con voz y sin voz, pero corrige los errores graves ocasionales en un paso de energía suave de otro modo, por ejemplo, dentro de un segmento con voz o sin voz. Un filtro adecuado para este propósito es un filtro de mediana, por ejemplo, un filtro de mediana de 3 puntos descrito por la ecuación

$$E_{hb1}(k) = \text{mediana}(E_{hb0}(k-1), E_{hb0}(k), E_{hb0}(k+1))$$

donde k es el índice de trama, y el operador de mediana (\cdot) selecciona la mediana de sus tres argumentos. El filtro de mediana de 3 puntos introduce un retardo de una trama. Otros tipos de filtros con o sin retardo se pueden diseñar para suavizar el paso de energía.

El valor de energía suavizado E_{hbl} además se puede adaptar por un adaptador de energía 508 para obtener la estimación de energía de banda alta adaptada final E_{hb} . Esta adaptación puede implicar o bien disminuir o bien aumentar el valor de energía suavizada en base al parámetro de nivel de voz v y/o la salida del parámetro d mediante el detector de aparición/oclusivo 503. Mediante un planteamiento, la adaptación del valor de energía de banda alta cambia no solamente el nivel de energía sino también la forma de la envolvente espectral dado que la selección del espectro de banda alta se puede vincular a la energía estimada.

En base al parámetro de nivel de voz v , se puede lograr la adaptación de energía como sigue. Para $v = 0$ que corresponde a una trama sin voz, el valor de energía suavizado E_{hbl} se incrementa ligeramente, por ejemplo, en 3 dB, para obtener el valor de energía adaptado E_{hb} . El nivel de energía aumentado enfatiza el habla sin voz en la salida extendida de banda ancha comparada con la entrada de banda estrecha y también ayuda a seleccionar una forma de envolvente espectral adecuada para segmentos sin voz. Para $v = 1$ que corresponde a una trama con voz, el valor de energía suavizado E_{hbl} se disminuye ligeramente, por ejemplo, en 6 dB, para obtener el valor de energía adaptado E_{hb} . El nivel de energía disminuido ligeramente ayuda a enmascarar cualquier error en la selección de la forma de envolvente espectral adecuada para los segmentos con voz y los artefactos ruidosos consiguientes.

Cuando el nivel de voz v está entre medias de 0 y 1 que corresponde a una trama con voz mezclada, no se hace adaptación del valor de energía. Tales tramas con voz mezcladas representan solamente una fracción pequeña del número total de tramas y los valores de energía sin adaptar funcionan bien para tales tramas. En base a la salida del detector de aparición/oclusivo d , la adaptación de energía se hace como sigue. Cuando $d = 1$, indica que la trama correspondiente contiene una aparición, por ejemplo, transición de silencio a sonido sin voz o con voz, o un sonido oclusivo, por ejemplo, /t/. En este caso, la energía de banda alta de la trama particular así como la trama siguiente se adapta a un valor muy bajo de manera que su contenido de energía de banda alta es bajo en el habla extendida de ancho de banda. Esto ayuda a impedir los artefactos ocasionales asociados con tales tramas. Para $d = 0$, no se hace adaptación adicional de la energía; es decir, se mantiene la adaptación de energía basada en el nivel de voz v , como se describió anteriormente.

La estimación de la envolvente espectral de banda ancha SE_{wb} se describe a continuación. Para estimar SE_{wb} , uno 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 juntas.

Un estimador de espectro de banda estrecha 509 puede estimar la envolvente espectral de banda estrecha SE_{nb} a partir del habla de banda estrecha sobre muestreada \hat{s}_{nb} . A partir de S_{nb} , los parámetros LP, $\mathbf{B}_{nb} = \{1, b_1, b_2, \dots, b_Q\}$ donde Q es el orden de modelo, se calculan primero usando técnicas de análisis LP bien conocidas. Para una frecuencia sobre muestreada de 16 kHz, un orden de modelo adecuado Q , por ejemplo, es 20. Los parámetros LP \mathbf{B}_{nb} modelan la envolvente espectral del habla de banda estrecha sobre muestreada 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 se da 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. Señalar que las envolventes espectrales SE_{nb} y SE_{usnb} son diferentes dado que la primera se deriva del habla de entrada de banda estrecha y la última del habla de banda estrecha sobre muestreada. No obstante, dentro de la banda de paso de 300 a 3400 Hz, están relacionadas aproximadamente por $SE_{usnb}(\omega) \approx SE_{nb}(2\omega)$ dentro de una constante. Aunque la envolvente espectral SE_{usnb} está definida sobre la gama 0 – 8000 (F_s) Hz, la parte 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 hace usando la FFT como sigue. Primero, la respuesta al impulso del filtro inverso $\mathbf{B}_{nb}(z)$ se calcula para una longitud adecuada, por ejemplo, 1024, como $\{1, b_1, b_2, \dots, b_Q, 0, 0, \dots, 0\}$. Entonces una FFT de la respuesta al impulso se toma, y se obtiene una envolvente espectral de magnitud SE_{usnb} calculando la magnitud inversa en cada índice de FFT. Para una longitud de FFT de 1024, la resolución de frecuencia de SE_{usnb} calculada como anteriormente es $16000/1024=15,625$ Hz. A partir de SE_{usnb} , la envolvente espectral de banda estrecha SE_{nb} se estima simplemente extrayendo las magnitudes espectrales dentro de la gama aproximada, de 300 a 3400 Hz.

Aquellos expertos en la técnica apreciarán que además del análisis LP, hay otros métodos para obtener la envolvente espectral de una trama de habla dada, por ejemplo, análisis cepstral, adecuación de curva de orden más alto o lineal por tramos de picos de 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 es consistente con la energía de banda alta estimada. Una técnica para aparecer con diferentes formas de envolvente espectral de banda alta que corresponden a diferentes energías de banda alta se describe a continuación.

Comenzando con una base de datos de entrenamiento grande de habla de banda ancha muestreada a 16 kHz, la envolvente de magnitud espectral de banda ancha se calcula para cada trama de habla usando análisis LP estándar u otras técnicas. A partir de la envolvente espectral de banda ancha para cada trama, la parte de banda alta correspondiente a 3400 – 8000 Hz se extrae y normaliza dividiendo a través de la magnitud espectral a 3400 Hz. Las envolventes espectrales de banda alta resultantes tienen de esta manera una magnitud de 0 dB a 3400 Hz. La energía de banda alta correspondiente a cada envolvente de banda alta normalizada se calcula a continuación. La colección de envolventes espectrales de banda alta entonces se parte en base a la energía de banda alta, por

ejemplo, se selecciona una secuencia de valores de energía nominales que difieren en 1 dB para cubrir la gama entera y todas las envolventes con energía dentro de 0,5 dB de un valor nominal son agrupadas juntas.

Para cada grupo formado de esta manera, se calcula la forma de envolvente espectral de banda alta y posteriormente la energía de banda alta correspondiente. En la FIG. 6, se muestra un conjunto de 60 formas de envolvente espectral de banda alta 600 (con magnitud en dB 3n función de la frecuencia en Hz) en niveles de energía diferentes. Contando desde la parte de abajo de la figura, las 1ª, 10ª, 20ª, 30ª, 40ª, 50ª, y 60ª formas (conocidas aquí dentro como formas pre calculadas) se obtuvieron usando una técnica similar a la descrita anteriormente. Las formas restantes 53 fueron obtenidas por simple interpolación lineal (en el dominio de dB) entre las formas pre calculadas más próximas.

Las energías de estas formas oscilan desde alrededor de 4,5 dB para la forma 1ª a alrededor de 43,5 dB para la forma 60ª. Dado que la energía de banda alta para una trama, es un asunto simple seleccionar la forma de envolvente espectral de banda alta de mayor grado de coincidencia como se describirá más tarde en el documento. La forma seleccionada representa la envolvente espectral de banda alta estimada SE_{hb} dentro de una constante. En la FIG. 6, la resolución de energía media es aproximadamente de 0,65 dB. Claramente, es posible mejor resolución aumentando el número de formas. Dadas las formas en la FIG. 6, la selección de una forma para una energía particular es única. Uno puede pensar también en una situación donde 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. Adicionalmente, uno puede tener múltiples conjuntos de formas cada conjunto indexado por la energía de banda alta, por ejemplo, dos conjuntos de formas seleccionables por el parámetro de voz v , uno para tramas con voz y el otro para tramas sin voz. Para una trama con voz mezclada, las dos formas seleccionadas de los dos conjuntos se pueden combinar adecuadamente.

El método de estimación de espectro de banda alta descrito anteriormente ofrece algunas claras ventajas. Por ejemplo, este planteamiento ofrece control explícito sobre la evolución en el tiempo de las estimaciones de espectro de banda alta. Una evolución suave de las estimaciones de espectro de banda alta dentro de distintos segmentos de habla, por ejemplo, habla con voz, habla sin voz, y así sucesivamente es a menudo importante para habla extendida de banda ancha libre de artefactos. Para el método de estimación de banda alta descrito anteriormente, es evidente a partir de la FIG. 6 que pequeños cambios en la energía de banda alta provocan pequeños cambios en las formas de envolvente espectral de banda alta. De esta manera, la evolución suave del espectro de banda alta se puede asegurar esencialmente asegurando que la evolución en el tiempo de la energía de banda alta dentro de distintos segmentos de habla también es suave. Esto se consuma explícitamente suavizando el paso de energía como se describió anteriormente.

Señalar que distintos segmentos de habla, dentro de los cuales se hace la suavización de energía, se pueden identificar con resolución incluso más fina, por ejemplo, haciendo el seguimiento del cambio en el espectro de banda estrecha o el espectro de habla de banda estrecha sobre muestreada de trama a trama usando cualquiera de las medidas de distancia espectral bien conocidas tales como la distorsión espectral logarítmica o la distorsión Itakura basada en LP. Usando este planteamiento, se puede definir un segmento de habla distinto como una secuencia de tramas dentro de la cual el espectro está evolucionando lentamente y que se encuentra rodeado en cada lado por una trama en la cual el cambio espectral calculado excede un umbral fijo o uno adaptativo indicando por ello la presencia de una transición espectral en cada lado del segmento de habla distinto. La suavización del paso de energía se puede hacer entonces dentro del segmento de habla distinto, pero no a través de límites de los segmentos.

Aquí, la evolución suave del paso de energía de banda alta se traduce en una evolución suave de la envolvente espectral de banda alta estimada, la cual es una característica deseable dentro de un segmento de habla distinto. Señalar también que este planteamiento para asegurar una evolución suave de la envolvente espectral de banda alta dentro de un segmento de habla distinto también se puede aplicar como un paso de procesamiento posterior a una secuencia de envolventes espectrales de banda alta estimadas obtenidas por métodos de la técnica anterior. En ese caso, no obstante, las envolventes espectrales de banda alta pueden necesitar ser suavizadas explícitamente dentro de un segmento de habla distinto, a diferencia de la suavización del paso de energía hacia delante de las enseñanzas actuales las cuales provocan automáticamente la evolución suave de la envolvente espectral de banda alta.

La pérdida de información de la señal de habla de banda estrecha en la banda baja (la cual, en este ejemplo ilustrativo, puede ser de 0 – 300 Hz) no es debida a la restricción de ancho de banda impuesta por la frecuencia de muestreo como en el caso de la banda alta sino debido al efecto de limitación de banda de la función de transferencia de canal que consta de, por ejemplo, el micrófono, amplificador, codificador de habla, canal de transmisión, y así sucesivamente.

Una planteamiento hacia delante para restaurar la señal de banda baja es entonces contrarrestar el efecto de esta función de transferencia de canal dentro de la gama de 0 a 300 Hz. Una forma simple para hacer esto es usar un estimador de espectro de banda baja 511 para estimar la función de transferencia de canal en la gama de frecuencia de 0 a 300 Hz a partir de los datos disponibles, obtener su inversa, y usar la inversa para aumentar la envolvente espectral del habla de banda estrecha sobre muestreada. Es decir, la envolvente espectral de banda baja SE_{lb} se

estima como la suma de SE_{usnb} y una característica de aumento de envolvente espectral $SE_{aumentada}$ diseñada a partir de la inversa de la función de transferencia de canal (suponiendo que las magnitudes de envolvente espectral se expresan en el dominio logarítmico, por ejemplo, en dB). Para muchos ajuste de aplicaciones, se debería tener cuidado en el diseño de $SE_{aumentada}$. Dado que la restauración de la señal de banda baja se basa esencialmente en la

- 5 amplificación de una señal de bajo nivel, implica el peligro de amplificar errores, ruido, y distorsiones típicamente asociados con señales de bajo nivel. Dependiendo de la calidad de la señal de bajo nivel, el valor de aumento máximo se debería restringir adecuadamente. También, dentro de la gama de frecuencia de 0 a alrededor de 60 Hz, es deseable diseñar $SE_{aumentada}$ para tener valores bajos (o incluso negativos, es decir, que atenúan) para impedir amplificar el zumbido eléctrico y el ruido de fondo.
- 10 Un estimador de espectro de banda ancha 512 entonces puede estimar la envolvente espectral de banda ancha combinando las envolventes espectrales estimadas en la banda estrecha, banda alta, y banda baja. Una forma de combinar las tres envolventes para estimar la envolvente espectral de banda ancha es como sigue.

La envolvente espectral de banda estrecha SE_{nb} se estima a partir de \hat{s}_{nb} como se describió anteriormente y sus valores dentro de la gama de 400 a 3200 Hz se usan sin ningún cambio en la estimación de la envolvente espectral de banda ancha SE_{wb} . Para seleccionar la forma de banda alta adecuada, son necesarios la energía de banda alta y el valor de la magnitud de inicio a 3400 Hz. La energía de banda alta E_{hb} en dB se estima como se describió anteriormente. El valor de la magnitud de inicio a 3400 Hz se estima modelando el espectro de magnitud FFT de \hat{s}_{nb} en dB dentro de la banda de transición, viz., de 2500 – 3400 Hz, por medio de una línea recta a través de regresión lineal y encontrando el valor de la línea recta a 3400 Hz. Permitamos que este valor de la magnitud sea indicado por M_{3400} en dB. La envolvente espectral de banda alta entonces se selecciona como una entre muchos valores, por ejemplo, como se muestra en la FIG. 6, que tiene un valor de energía más próximo a $E_{hb} - M_{3400}$. Permitamos que esta forma se indique por $SE_{más\ cercano}$. Entonces la estimación de la envolvente espectral de banda alta SE_{hb} y por lo tanto la envolvente espectral de banda ancha SE_{wb} dentro de la gama de 3400 a 8000 Hz se estima como $SE_{más\ cercano} + M_{3400}$.

- 15 Entre 3200 y 3400 Hz, se estima SE_{wb} como el valor interpolado linealmente en dB entre SE_{nb} y una línea recta que une la SE_{nb} a 3200 Hz y M_{3400} a 3400 Hz. El valor de interpolación por sí mismo se cambia linealmente de manera que la SE_{wb} estimada se mueve gradualmente desde SE_{nb} a 3200 Hz a M_{3400} 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} se estiman como $SE_{nb} + SE_{aumentada}$, donde $SE_{aumentada}$ representa una característica de aumento diseñada adecuadamente a partir de la inversa de la función de transferencia de canal como se describió anteriormente.

Como se aludió anteriormente, las tramas que contienen apariciones y/o oclusivas pueden beneficiarse del manejo especial para evitar artefactos ocasionales en el habla extendida de banda ancha. Tales tramas se pueden identificar por el aumento repentino en su energía en relación con las tramas precedentes. La salida del detector de aparición/oclusivo 503 d para una trama se fija a 1 siempre que la energía de la trama precedente es baja, es decir, por debajo de un cierto umbral, por ejemplo, -50 dB, y el aumento en la energía de la trama actual en relación con la trama precedente excede otro umbral, por ejemplo, 15 dB. De otro modo, la salida del detector d se fija a 0. La energía de la trama por sí misma se calcula a partir de la energía del espectro de magnitud FFT del habla de banda estrecha sobre muestreada \hat{s}_{nb} dentro de la banda estrecha, es decir, 300 – 3400 Hz. Como se señaló anteriormente, la salida del detector de aparición/oclusivo 503 d se alimenta en el estimador de nivel de voz 502 y el adaptador de energía 508. Como se describió anteriormente, siempre que se marca una trama como que contiene una aparición o un oclusivo con $d = 1$, el nivel de voz v de esa trama así como la trama siguiente se fija a 1. También, el valor de energía de banda alta adaptada E_{hb} de esa trama así como la trama siguiente se fija a un valor bajo. Alternativamente, la extensión de banda ancha se puede desviar en conjunto para esas tramas.

- Aquellos expertos en la técnica apreciarán que las técnicas de estimación de energía de banda alta descritas se pueden usar en conjunto con otros sistemas de extensión de ancho de banda de la técnica anterior para subir el contenido de señal de banda alta generado artificialmente para tales sistemas a un nivel de energía adecuado. Adicionalmente, señalar que aunque la técnica de estimación de energía se ha descrito con referencia a la banda de frecuencia alta, (por ejemplo, 3400 – 8000 Hz), también se puede aplicar para estimar la energía en cualquier otra banda redefiniendo adecuadamente la banda de transición. Por ejemplo, para estimar la energía en un contexto de banda baja, tal como 0 – 300 Hz, la banda de transición se puede redefinir como la banda de 300 – 600 Hz. Aquellos expertos en la técnica reconocerán también que las técnicas de estimación de energía de banda alta descritas aquí dentro se pueden emplear para propósitos de codificación de habla/audio. De igual manera, las técnicas descritas aquí dentro para estimar la envolvente espectral de banda alta y la excitación de banda alta también se pueden usar en el contexto de codificación de habla/audio.

- Señalar que aunque la estimación de parámetros tales como la envolvente espectral, cruces por cero, coeficientes LP, energías de banda, y así sucesivamente se ha descrito en los ejemplos específicos dados previamente como que se hace a partir del habla de banda estrecha en algunos casos y el habla de banda estrecha sobre muestreada en otros casos, se apreciará por aquellos expertos en la técnica que la estimación de los parámetros respectivos y su uso y aplicación posterior, se puede modificar para ser hecho de cualquiera de las dos de esas dos señales (habla de banda estrecha o el habla de banda estrecha sobre muestreada).

Aquellos expertos en la técnica reconocerán que se puede hacer una amplia variedad de modificaciones, alteraciones, y combinaciones con respecto a las realizaciones descritas anteriormente sin salirse del alcance de la invención definido por las reivindicaciones anexas.

REIVINDICACIONES

1. Un método que comprende:
 - recibir una señal de audio digital que comprende una señal dentro de una primera banda de frecuencia;
 - 5 procesar la señal de audio digital de entrada a través de una operación de sobre muestreo, o mediante un sistema de ganancia unitaria simple para el cual la salida es igual a la entrada, para generar una señal de audio digital procesada;
 - 10 estimar un nivel de energía en una segunda banda de frecuencia para la señal de audio digital de entrada, en base a una energía de una banda de transición de la señal de audio digital procesada dentro de una gama de frecuencia superior predeterminada de la primera banda de frecuencia, la segunda banda de frecuencia que está por encima de la primera banda de frecuencia en el espectro de frecuencia;
 - generar una segunda señal de audio digital en base a al menos el nivel de energía y una envolvente espectral estimada correspondiente al nivel de energía; y
 - combinar la señal de audio digital de entrada y la segunda señal de audio digital para generar una señal de audio digital resultante que tiene un ancho de banda de señal extendido.
- 15 2. El método de la reivindicación 1, en el que el procesamiento comprende sobre muestrear la señal de audio digital de entrada para generar la señal de audio digital procesada.
3. El método de la reivindicación 1, en el que la estimación comprende calcular un nivel de energía de la señal de audio digital procesada calculando un espectro de frecuencia de la señal de audio digital procesada y sumando las energías de las componentes espectrales dentro de la banda de transición.
- 20 4. El método de la reivindicación 1, en el que la estimación comprende además utilizar al menos un parámetro de habla predeterminado, basado en la señal de audio digital de entrada, para generar un espacio de parámetros.
5. El método de la reivindicación 4, en el que el parámetro de habla predeterminado es al menos uno de un parámetro de cruce por cero, un parámetro de medida de uniformidad espectral, un parámetro de pendiente espectral de banda de transmisión, y un parámetro de forma de envolvente espectral de banda de transición.
- 25 6. El método de la reivindicación 4, en el que la estimación además comprende partir el espacio de parámetros en regiones y asignar coeficientes para cada región para estimar el nivel de energía.
7. El método de la reivindicación 1, en el que la señal dentro de la primera banda de frecuencia tiene un ancho de banda de alrededor de 300 – 3400 Hz.
8. Un aparato, que comprende:
 - 30 una entrada configurada y dispuesta para recibir una señal de audio digital de entrada que comprende una señal dentro de una primera banda de frecuencia;
 - un procesador operable acoplado con la entrada y que se configura y dispone para:
 - 35 procesar la señal de audio digital de entrada a través de una operación de sobre muestreo, o mediante un sistema de ganancia unitaria simple para el cual la salida es igual a la entrada, para generar una señal de audio digital procesada;
 - estimar un nivel de energía en una segunda banda de frecuencia para la señal de audio digital de entrada, en base a una energía de una banda de transición de la señal de audio digital procesada dentro de una gama de frecuencia superior predeterminada de la primera banda de frecuencia, la segunda banda de frecuencia que está por encima de la primera banda de frecuencia en el espectro de frecuencia;
 - 40 generar una segunda señal de audio digital en base a al menos al nivel de energía y una envolvente espectral estimada correspondiente al nivel de energía; y
 - combinar la señal de audio digital de entrada y la segunda señal de audio digital para generar una señal de audio digital resultante que tiene un ancho de banda de señal extendido.

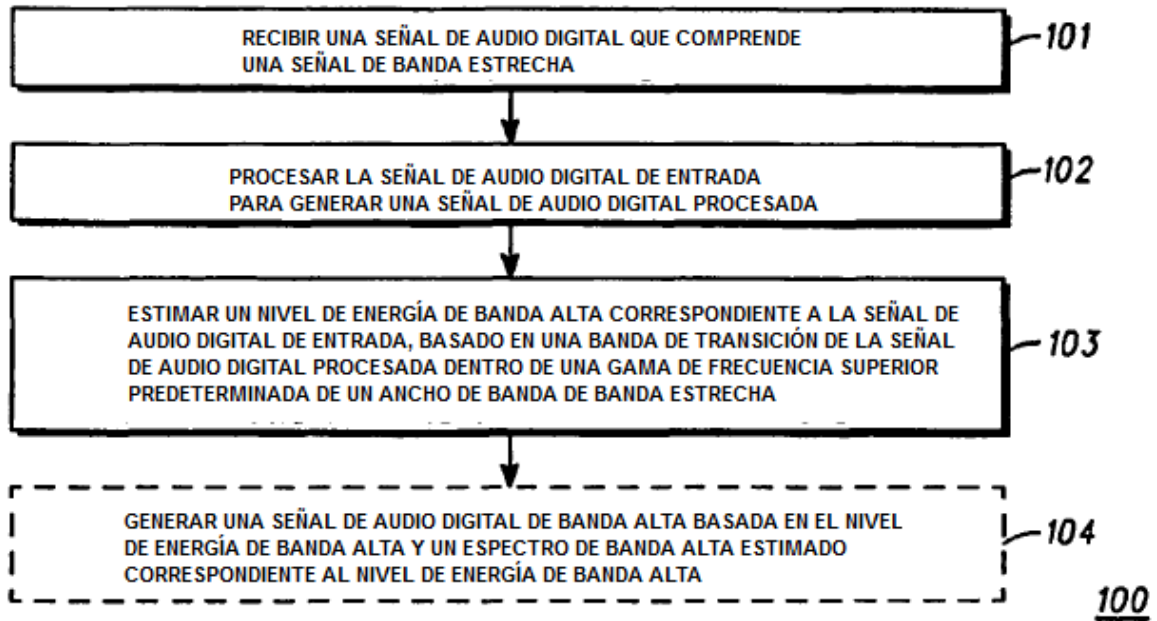


FIG. 1

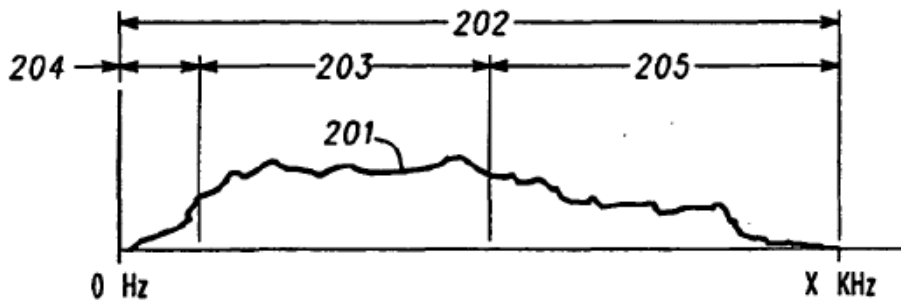


FIG. 2

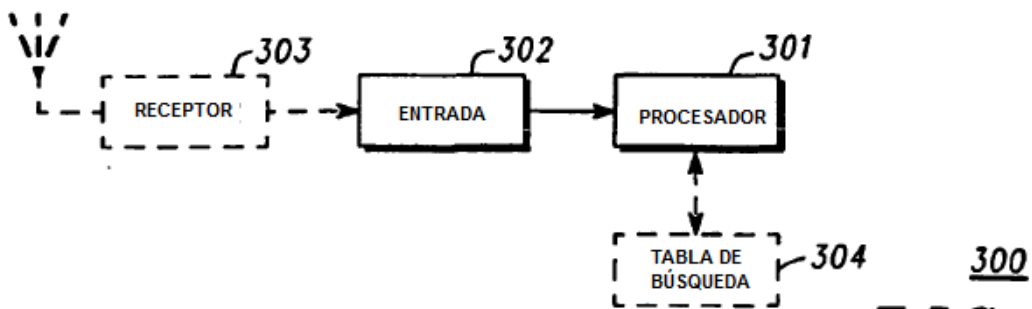


FIG. 3

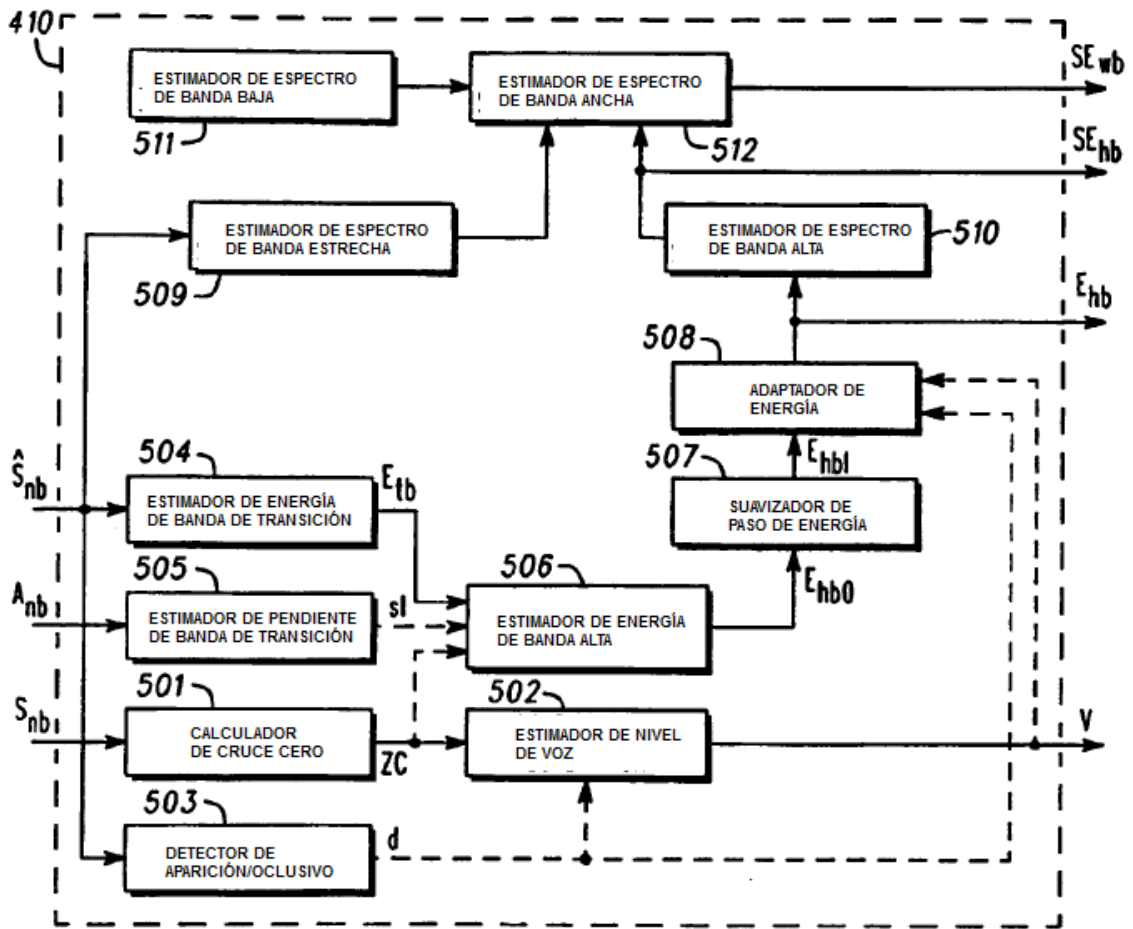


FIG. 5

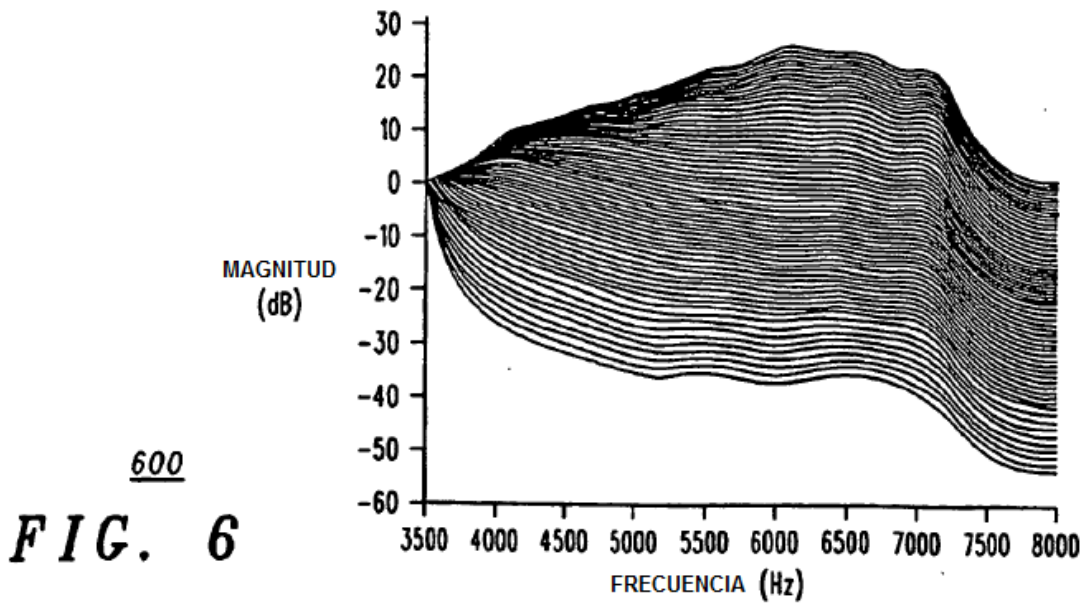


FIG. 6