

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 725 358**

51 Int. Cl.:

G10L 21/0388 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.01.2014 PCT/EP2014/051591**

87 Fecha y número de publicación internacional: **07.08.2014 WO14118155**

96 Fecha de presentación y número de la solicitud europea: **28.01.2014 E 14701550 (7)**

97 Fecha y número de publicación de la concesión europea: **06.03.2019 EP 2951828**

54 Título: **Decodificador para generar una señal de audio mejorada en frecuencia, procedimiento de decodificación, codificador para generar una señal codificada y procedimiento de codificación que utiliza información lateral de selección compacta**

30 Prioridad:

29.01.2013 US 201361758092 P

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

23.09.2019

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**NAGEL, FREDERIK;
DISCH, SASCHA y
NIEDERMEIER, ANDREAS**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 725 358 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

Decodificador para generar una señal de audio mejorada en frecuencia, procedimiento de decodificación, codificador para generar una señal codificada y procedimiento de codificación que utiliza información lateral de selección compacta

Memoria descriptiva

[0001] La presente invención se refiere a la codificación de audio y, en particular, a la codificación de audio en el contexto de la mejora de la frecuencia, es decir que la señal de salida de un decodificador tenga un número mayor de bandas de frecuencia en comparación con una señal codificada. Tales procedimientos comprenden la extensión del ancho de banda, la replicación espectral o el relleno inteligente de espacios vacíos.

[0002] Los sistemas contemporáneos de codificación de voz tienen capacidad para codificar el contenido de audio digital de banda ancha (WB), es decir, señales con frecuencias de hasta 7 u 8 kHz, a tasas de bits de tan solo 6 kbit/s. Los ejemplos más ampliamente difundidos son las recomendaciones de ITU-T G.722.2 [1], así como las desarrolladas más recientemente G.718 [4, 10] y MPEG-D de Codificación Unificada de Voz y Audio (USAC) [8]. Tanto G.722.2, que también se conoce como AMR-WB, como G.718 emplean técnicas de extensión de ancho de banda (BWE) entre 6,4 y 7 kHz para permitir que el codificador de núcleo ACELP subyacente se “enfoque” en las frecuencias más bajas perceptualmente más relevantes (especialmente aquellas en las cuales el sistema auditivo humano es sensible a las fases) y de esa manera se obtenga una calidad perceptual suficiente especialmente a tasas de bits muy bajas. En el perfil de Codificación de Audio Avanzada de Alta Eficiencia Extendida (xHE-AAC) de USAC, se utiliza la replicación espectral mejorada de banda (de SBR) para extender el ancho de banda de audio más allá del ancho de banda del codificador de núcleo que por lo general es inferior a 6 kHz a 16 kbit/s. Los procesos BWE del estado actual de la técnica se pueden dividir en general en dos estrategias conceptuales:

- BWE *ciega o artificial*, en la cual se reconstruyen los componentes de alta frecuencia (HF) a partir solamente de la señal de baja frecuencia (LF) del codificador de núcleo decodificada, es decir sin necesidad de transmisión de información lateral desde el codificador. Este esquema es utilizado por AMR-WB y G.718 a 16 kbit/s y menos, así como por algunos post-procesadores de BWE compatibles con los anteriores que operan en la voz telefónica de banda estrecha tradicional [5, 9, 12] (Ejemplo: figura 15).
- BWE *Guiada*, que difiere de la BWE ciega en que algunos de los parámetros empleados para la reconstrucción del contenido de HF se transmiten al decodificador en forma de información lateral en lugar de ser deducidos de la señal de núcleo decodificada. AMR-WB, G.718, xHE-AAC, así como otros códecs [2, 7, 11] utilizan esta estrategia, aunque no a tasas de transmisión muy bajas (figura 16).

[0003] La Fig. 15 ilustra tal extensión de ancho de banda ciega o artificial descrita en la publicación Bernd Geiser, Peter Jax, and Peter Vary: "ROBUST WIDEBAND ENHANCEMENT OF SPEECH BY COMBINED CODING AND ARTIFICIAL EXTENSIÓN DE ANCHO DE BANDA", Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), 2005. El algoritmo de extensión de ancho de banda autónomo ilustrado en la Fig. 15 comprende un procedimiento de interpolación 1500, un filtro de análisis 1600, una extensión de excitación 1700, un filtro de síntesis 1800, un procedimiento de extracción de características 1510, un procedimiento de estimación de la envolvente 1520 y un modelo estadístico 1530. Después de una interpolación de la señal de banda estrecha a una tasa de muestreo de banda ancha, se calcula un vector de característica. A continuación, mediante un modelo oculto de Markov previamente entrenado (HMM), se determina un cálculo estimativo de la envolvente espectral de banda ancha en términos de coeficientes de predicción lineal (LP). Estos coeficientes de banda ancha son utilizados para el filtrado de análisis de la señal de banda estrecha interpolada. Después de la extensión de la excitación así producida se aplica un filtro de síntesis inverso. La elección de una extensión de excitación que no altera la banda estrecha es transparente con respecto a los componentes de banda estrecha.

[0004] La Fig. 16 ilustra una extensión de ancho de banda con información lateral de acuerdo con lo descrito en la publicación antes citada, donde la extensión de ancho de banda comprende un paso banda telefónico 1620, un bloque de extracción de información lateral 1610, un codificador (conjunto) 1630, un decodificador 1640 y un bloque de extensión de ancho de banda 1650. En la Fig. 16 se ilustra este sistema para mejorar el ancho de banda de una señal de voz de banda con errores mediante la combinación de codificación y extensión de ancho de banda. En la terminal de transmisión, se analiza la envolvente espectral de banda alta de la señal de entrada de banda ancha y se determina la información lateral. El mensaje así obtenido m se codifica por separado o conjuntamente con la señal de voz de banda estrecha. En el receptor, se utiliza la información lateral del decodificador para sustentar el cálculo de la envolvente de banda ancha dentro del algoritmo de extensión de ancho de banda. El mensaje m se obtiene por varios procedimientos. Se extrae una representación espectral de frecuencias de 3,4 kHz a 7 kHz de la señal de banda ancha disponible solo del lado del envío.

[0005] Esta envolvente de subbanda se computa por predicción lineal selectiva, es decir, el cómputo del espectro de potencia de la banda ancha seguida por una IOFT de sus componentes de banda superior y la posterior recursión

de Levinson-Durbin de orden 8. Los coeficientes de LPC de subbanda así obtenidos se convierten al dominio cepstral y por último son cuantificados por un cuantificador de vectores con un tamaño de libreo de códigos de $M = 2^N$. En el caso de una longitud de trama de 20 ms, esto da lugar a una tasa de datos de información lateral de 300 bit/s. Una estrategia de cálculo estimativo combinado extiende el cálculo de las probabilidades a posteriori y reintroduce las dependencias de la característica de banda estrecha. De esa manera se obtiene una forma mejorada de ocultamiento de errores que utiliza más de una fuente de información para su estimación de parámetros.

[0006] Se puede observar un cierto dilema de calidad en los códecs de WB a bajas tasas de bits, típicamente por debajo de 10 kbit/s. Por un lado, esas tasas ya son demasiado bajas para justificar la transmisión incluso de cantidades moderadas de datos de BWE, excluyendo los sistemas típicos de BWE guiada con 1 kbit/s o más de información lateral. Por otro lado, se encuentra que una BWE ciega factible suena considerablemente peor en algunos tipos de materiales de voz o música debido a la incapacidad de predicción correcta de los parámetros desde la señal de núcleo. Esto es así especialmente con respecto a cierto sonido vocal con fricativos con baja correlación entre HF y LF. Por lo tanto, es conveniente reducir la tasa de información lateral de un esquema de BWE guiada a un nivel muy inferior a 1 kbit/s, lo que permitiría la adopción de una codificación a tasas de bits muy bajas.

[0007] En los últimos años se han documentados múltiples estrategias de BWE [1–10]. En general, todas estas son completamente ciegas o completamente guiadas en un punto operativo dado, independientemente de las características instantáneas de la señal de entrada. Además, muchos sistemas de BWE ciega [1, 3, 4, 5, 9, 10] se optimizan especialmente para señales de voz en lugar de música y, por lo tanto, pueden producir resultados no satisfactorios para la música. Por último, la mayoría de las realizaciones de BWE son relativamente complejas desde el punto de vista informático, y emplean transformadas de Fourier, cómputos de filtros LPC o cuantificación de vectores de la información lateral (Codificación por Vectores Predictivos en MPEG–D USAC [8]). Esto puede ser una desventaja en la adopción de nueva tecnología de codificación en los mercados de telecomunicaciones móviles, dado que la mayoría de los dispositivos móviles ofrecen una limitadísima potencia informática y capacidad de la batería.

[0008] Una estrategia que extiende la BWE ciega mediante una pequeña formación lateral es la presentada en [12] e ilustrada en la Fig. 16. Sin embargo, la información lateral "m" se limita a la transmisión de una envolvente espectral del intervalo de frecuencias del ancho de banda extendido.

[0009] Un problema adicional del procedimiento ilustrado en la Fig. 16 es la manera muy complicada de estimar la envolvente mediante el uso de la característica de banda baja por un lado y la información lateral adicional de la envolvente por el otro. Ambas entradas, es decir, la característica de banda baja y la envolvente adicional de banda alta influyen sobre el modelo estadístico. Esto da lugar a una implementación complicada del lado del decodificador que es particularmente problemática para los dispositivos móviles debido al mayor consumo de energía. Por añadidura, el modelo estadístico es aún más difícil de actualizar debido al hecho de que no es influenciado por los datos adicionales de envolvente de banda alta.

[00010] Es un objetivo de la presente invención proporcionar un concepto mejorado de codificación/ decodificación de audio.

[00011] Este objetivo se obtiene por medio de un decodificador según la reivindicación 1, un codificador según la reivindicación 12, un procedimiento para generar una señal de audio mejorada en frecuencia según la reivindicación 14, un procedimiento de generación de una señal codificada según la reivindicación 15, un programa informático según la reivindicación 16.

[00012] La presente invención se basa en el hallazgo de que para reducir aún más la cantidad de información lateral y, además, para poder lograr que un codificador/decodificador en su conjunto no sea demasiado complejo, se debe reemplazar, o al menos mejorar, la codificación paramétrica de una porción de banda alta de la técnica anterior por información lateral de selección que se relaciona, en realidad, con el modelo estadístico utilizado junto con un extractor de características en un decodificador de mejora de la frecuencia. Debido a que la extracción de características en combinación con un modelo estadístico proporciona alternativas de representación paramétrica que presentan ambigüedades específicamente en el caso de ciertas porciones de voz, se ha encontrado que el control verdadero del modelo estadístico dentro de un generador de parámetros del lado del decodificador, cuál de las alternativas propuestas sería la mejor, es superior a la codificación paramétrica real de una determinada característica de la señal específicamente en aplicaciones de muy baja tasa de bits, donde la información lateral correspondiente a la extensión de ancho de banda es limitada.

[00013] De esa manera se mejora una BWE ciega, que aprovecha un modelo de origen de la señal codificada, mediante extensión con poca información lateral adicional, en particular si la señal en sí no da lugar a la reconstrucción del contenido de HF en un nivel aceptable de calidad perceptual. Por lo tanto, el procedimiento combina los parámetros del modelo de origen, que se generan en el contenido codificado del codificador de núcleo, mediante información extra. Esto es ventajoso, especialmente para mejorar la calidad perceptual de los sonidos difíciles de codificar dentro de un modelo de origen. Tales sonidos exhiben, por lo general, una baja correlación entre

el contenido de HF y LF.

5 **[00014]** La presente invención aborda los problemas de la BWE convencional en la codificación de audio de muy baja tasa de bits y las falencias de las técnicas de BWE existentes en el estado actual de la técnica. Se presenta una solución al problema antes planteado al proponer una BWE mínimamente guiada como combinación adaptativa a la señal de una BWE ciega y una guiada. La BWE de la invención agrega cierta información lateral a la señal, lo que da lugar a una mayor discriminación de los sonidos codificados de otro modo problemáticos. En la codificación de voz, esto se aplica especialmente a los sibilantes o fricativos.

10 **[00015]** Se ha encontrado que, en los códecs de WB, la envolvente espectral de la región de HF por encima de la región del codificador de núcleo representa los datos más críticos necesarios para ejecutar la BWE con calidad perceptual aceptable. Todos los demás parámetros como, por ejemplo, la estructura fina espectral y la envolvente temporal, pueden ser derivados, con frecuencia, de la señal de núcleo decodificada de forma bastante precisa o son de poca importancia perceptual. Los fricativos, sin embargo, a menudo carecen de una reproducción correcta en la
15 señal de BWE. Por lo tanto, la información lateral puede incluir información adicional que distingue entre sibilantes o fricativos tales como "f", "s", "ch" y "sh".

[00016] Otra información acústica problemática para la extensión de ancho de banda, cuando se producen plosivos o africadas tales como "t" o "tsch".

20 **[00017]** La presente invención permite utilizar solo esta información lateral y transmitir, en realidad, esta información lateral cuando es necesario y no transmitir esta información lateral, cuando no hay una ambigüedad estimada en el modelo estadístico.

25 **[00018]** Además, las formas de realización preferidas de la presente invención solo hacen uso de una cantidad muy pequeña de información lateral como, por ejemplo, tres o menos bits por trama, una detección de actividad de voz/detección de voz o no voz combinadas para controlar un estimador de señales, diferentes modelos estadísticos determinados por un clasificador de señales o alternativas de representación paramétrica no solo en referencia a una estimación de la envolvente sino también en referencia a otras herramientas de extensión de ancho de banda o
30 la mejora de los parámetros de extensión de ancho de banda o la adición de nuevos parámetros a los ya existentes y los parámetros de extensión de ancho de banda transmitidos en realidad.

[00019] Seguidamente se describen las formas de realización preferidas de la presente invención en el contexto de los dibujos que la acompañan y también se exponen en las reivindicaciones dependientes.

- 35 Fig. 1 ilustra un decodificador para generar una señal de audio mejorada en frecuencia;
Fig. 2 ilustra una implementación preferida en el contexto del extractor de información lateral de la Fig. 1;
Fig. 3 ilustra una tabla que relaciona con un número de bits de la información lateral de
40 selección con el número de alternativas de representación paramétrica;
Fig. 4 ilustra un procedimiento preferido ejecutado en el generador de parámetros;
Fig. 5 ilustra una implementación preferida del estimador de señales controlado por un detector de actividad de voz o un detector de voz/no voz;
Fig. 6 ilustra una implementación preferida del generador de parámetros controlado por un
45 clasificador de señales;
Fig. 7 ilustra un ejemplo correspondiente a un resultado de un modelo estadístico y la información lateral de selección asociada;
Fig. 8 ilustra una señal codificada a modo de ejemplo que comprende una señal de núcleo codificada e información lateral asociada;
50 Fig. 9 ilustra un esquema de procesamiento de señales de extensión de ancho de banda correspondiente a una mejora de la estimación de envolvente;
Fig. 10 ilustra una implementación adicional de un decodificador en el contexto de procedimientos de replicación de la banda espectral;
Fig. 11 ilustra otra forma de realización de un decodificador en el contexto de información lateral transmitida
55 adicionalmente;
Fig. 12 ilustra una forma de realización de un codificador para generar una señal codificada;
Fig. 13 ilustra una implementación del generador de información lateral de selección de la Fig. 12;
Fig. 14 ilustra una implementación adicional del generador de información lateral de selección de la Fig. 12;
Fig. 15 ilustra un algoritmo de extensión de ancho de banda autónomo de la técnica anterior; y
60 Fig. 16 ilustra un panorama general de un sistema de transmisión con un mensaje de adición.

[00020] La Fig. 1 ilustra un decodificador para generar una señal de audio mejorada en frecuencia 120. El decodificador comprende un extractor de características 104 para extraer (al menos) una característica de una señal de núcleo 100. En general, el extractor de características puede extraer una característica única o una pluralidad de
65 características, es decir, dos o más características, y es aún más preferible que el extractor de características

extraiga una pluralidad de características. Esto se aplica no solo al extractor de características incluido en el decodificador sino también al extractor de características del codificador.

[00021] Además, se incluye un extractor de información lateral 110 para extraer una información lateral de selección 114 asociada a la señal de núcleo 100. Por añadidura, hay un generador de parámetros 108 conectado al extractor de características 104 a través de una línea de transmisión de características 112 y al extractor de información lateral 110 a través de la información lateral de selección 114. El generador de parámetros 108 está configurado para generar una representación paramétrica para estimar un alcance espectral de la señal de audio mejorada en la frecuencia no definida por la señal de núcleo. El generador de parámetros 108 está configurado para presentar un número de alternativas de representación paramétrica en respuesta a las características 112 y para elegir una de las alternativas de representación paramétrica como representación paramétrica en respuesta a la información lateral de selección 114. El decodificador comprende además un estimador de señales 118 para estimar una señal de audio mejorada en frecuencia mediante el uso de la representación paramétrica seleccionada por el selector, es decir, la representación paramétrica 116.

[00022] En particular, el extractor de características 104 puede ser implementado para extraer de la señal de núcleo decodificada de la manera ilustrada en la Fig. 2. A continuación, una interfaz de entrada 110 está configurada para recibir una señal de entrada codificada 200. Esta señal de entrada codificada 200 es introducida en la interfaz 110 y, a continuación, la interfaz de entrada 110 separa la información lateral de selección de la señal de núcleo codificada. De esa manera, la interfaz de entrada 110 opera como extractor de información lateral 110 de la Fig. 1. A continuación se introduce la señal de núcleo codificada 201 emitida por la interfaz de entrada 110 en un decodificador de núcleo 124 para dar origen a una señal de núcleo decodificada que puede ser la señal de núcleo 100.

[00023] Alternativamente, sin embargo, el extractor de características también puede cumplir la función de extraer una característica de la señal de núcleo codificada. Por lo general, la señal de núcleo codificada comprende una representación de factores de escala correspondientes a bandas de frecuencia o cualquier otra representación de información de audio. Dependiendo del tipo de extracción de características, la representación codificada de la señal de audio es representativa de la señal de núcleo decodificada y, por lo tanto, se pueden extraer las características. Alternativa o adicionalmente, se puede extraer una característica no solo de una señal de núcleo totalmente decodificada sino de una señal de núcleo parcialmente decodificada. En la codificación en el dominio de la frecuencia, la señal codificada está representando una representación en el dominio de la frecuencia que comprende una secuencia de tramas espectrales. Por lo tanto, la señal de núcleo codificada puede ser decodificada solo en parte para obtener una representación decodificada de una secuencia de tramas espectrales, antes de efectuar en realidad una conversión de espectro-tiempo. De este modo, el extractor de características 104 puede extraer características de la señal de núcleo codificada o de una señal de núcleo parcialmente decodificada o bien de una señal de núcleo completamente decodificada. El extractor de características 104 puede ser implementado, en lo que respecta a sus características extraídas, de la manera conocida en la técnica y el extractor de características puede ser implementado, por ejemplo, como en las tecnologías de impresiones digitales de audio o ID de audio.

[00024] De preferencia, la información lateral de selección 114 comprende un número N de bits por trama de la señal de núcleo. La Fig. 3. Ilustra una tabla que corresponde a diferentes alternativas. El número de bits correspondiente a la información lateral de selección es fijo o es seleccionado dependiendo del número de alternativas de representación paramétrica proporcionadas por un modelo estadístico en respuesta a una característica extraída. Un bit de información lateral de selección es suficiente cuando el modelo estadístico solo ofrece dos alternativas de representación paramétrica en respuesta a una característica. Cuando el modelo estadístico presenta un número máximo de cuatro alternativas de representación, en ese caso se necesitan dos bits para la información lateral de selección. Tres bits de información lateral de selección permiten un máximo de ocho alternativas de representación paramétrica concurrentes. Cuatro bits de información lateral de selección dan origen en realidad a 16 alternativas de representación paramétrica y cinco bits de información lateral de selección dan origen a 32 alternativas de representación paramétrica concurrentes. Es preferible utilizar solo tres o menos de tres bits de información lateral de selección por trama para dar lugar a una tasa de información lateral de 150 bits por segundo cuando un segundo se divide en 50 tramas. Esta tasa de información lateral se puede reducir incluso debido a que la información lateral de selección solo es necesaria cuando el modelo estadístico ofrece en realidad alternativas de representación. De este modo, cuando el modelo estadístico solo ofrece una única alternativa para una característica, no es necesario en absoluto un bit de información lateral de selección. Por otra parte, cuando el modelo estadístico solo presenta cuatro alternativas de representación paramétrica, en ese caso solo se necesitan dos bits, en lugar de tres bits de información lateral de selección. Por lo tanto, en los casos típicos, la tasa de información lateral adicional se puede reducir incluso a menos de 150 bits por segundo.

[00025] Además, el generador de parámetros está configurado para presentar, a lo sumo, una cantidad de alternativas de representación paramétrica que es igual a 2^N . Por otro lado, cuando el generador de parámetros 108 presenta, por ejemplo, solo cinco alternativas de representación paramétrica, de todas maneras se necesitan tres bits de información lateral de selección.

[00026] La Fig. 4 ilustra una implementación preferida del generador de parámetros 108. En particular, el generador de parámetros 108 está configurado de manera que la característica 112 de la Fig. 1 sea introducida en un modelo estadístico de la manera indicada en la etapa 400. A continuación, como se indica en la etapa 402, el modelo presenta una pluralidad de alternativas de representación paramétrica.

[00027] Además, el generador de parámetros 108 está configurado para recobrar la información lateral de selección 114 del extractor de información lateral como se indica en la etapa 404. Seguidamente, en la etapa 406, se selecciona una alternativa específica de representación paramétrica utilizando la información lateral de selección 114. Por último, en la etapa 408, se envía la alternativa de representación paramétrica seleccionada al estimador de señales 118.

[00028] De preferencia, el generador de parámetros 108 está configurado para usar, al seleccionar una de las alternativas de representación paramétrica, un orden predefinido de alternativas de representación paramétrica o, por otro lado, un orden de señales del codificador de las alternativas de representación. Para este fin, se hace referencia a la Fig. 7. La Fig. 7 ilustra un resultado del modelo estadístico que presenta cuatro alternativas de representación paramétrica 702, 704, 706, 708. También se ilustra el código de información lateral de selección correspondiente. La alternativa 702 corresponde al patrón de bits 712. La alternativa 704 corresponde al patrón de bits 714. La alternativa 706 corresponde al patrón de bits 716 y la alternativa 708 corresponde al patrón de bits 718. De esa manera, cuando el generador de parámetros 108 o, por ejemplo, la etapa 402 obtiene las cuatro alternativas 702 a 708 en el orden ilustrado en la Fig. 7, entonces una información lateral de selección que tiene el patrón de bits 716 ha de identificar de manera unívoca la alternativa de representación paramétrica 3 (número de referencia 706) y el generador de parámetros 108 selecciona entonces esta tercera alternativa. Sin embargo, cuando el patrón de bits de la información lateral de selección es el patrón de bits 712, se seleccionaría entonces la primera alternativa 702.

[00029] El orden predefinido de las alternativas de representación paramétrica puede ser, por lo tanto, el orden en el cual el modelo estadístico entrega realmente las alternativas en respuesta a una característica extraída. Por otro lado, si la alternativa individual está asociada a diferentes probabilidades que, de todas maneras, se aproximan bastante entre sí, entonces el orden predefinido podría ser que la representación paramétrica con mayores probabilidades sea la primera y así sucesivamente.

[00030] De manera alternativa, el orden podría ser señalado, por ejemplo, por un bit único, aunque, para ahorrar incluso este bit, es preferible un orden predefinido.

[00031] A continuación se hace referencia a las Figs. 9 a 11.

[00032] En una forma de realización según la Fig. 9, la invención se adecua particularmente a las señales de voz, ya que se aprovecha un modelo fuente de voz dedicado para la extracción de parámetros. Sin embargo, la invención no se limita a la codificación de voz. Diferentes formas de realización podrían emplear también otros modelos de fuente.

[00033] En particular, la información lateral de selección 114 también se denomina "información de fricativos", ya que esta información lateral de selección distingue entre sibilantes o fricativos problemáticos tales como "f", "s" o "sh". De este modo, la información lateral de selección ofrece una definición clara de una de tres alternativas problemáticas que son proporcionadas, por ejemplo, por el modelo estadístico 904 en el proceso de la estimación de envolvente 902 que se realiza en el generador de parámetros 108. La estimación de envolvente da lugar a una representación paramétrica de la envolvente espectral de las porciones espectrales no incluidas en la señal de núcleo.

[00034] Por lo tanto, el bloque 104 puede corresponder al bloque 1510 de la Fig. 15. Además, el bloque 1530 de la Fig. 15 puede corresponder al modelo estadístico 904 de la Fig. 9.

[00035] Además, es preferible que el estimador de señales 118 comprenda un filtro de análisis 910, un bloque de extensión de excitación 112 y un filtro de síntesis 940. De este modo, los bloques 910, 912, 914 pueden corresponder a los bloques 1600, 1700 y 1800 de la Fig. 15. En particular, el filtro de análisis 910 es un filtro de análisis de LPC. El bloque de estimación de envolvente 902 controla los coeficientes de filtro del filtro de análisis 910 de manera que el resultado del bloque 910 sea la señal de excitación de filtro. Esta señal de excitación de filtro se extiende con respecto a la frecuencia para obtener una señal de excitación a la salida del bloque 912, que no solo tiene el intervalo de frecuencias del decodificador 120 correspondiente a una señal de salida sino que también tiene la frecuencia o intervalo espectral no definida por el codificador de núcleo y/o que excede el intervalo espectral de la señal de núcleo. De este modo, se sobremuestra la señal de audio 909 a la salida y es interpolada por un interpolador 900 y, a continuación, se somete a la señal interpolada al proceso ejecutado en el estimador de señales 118. De este modo, el interpolador 900 de la Fig. 9 puede corresponder a interpolador 1500 de la Fig. 15. Sin embargo, es preferible, a diferencia de la Fig. 15, que la extracción de características 104 se lleve a cabo utilizando la señal no interpolada en lugar de la señal interpolada, como se ilustra en la Fig. 15. Esto es ventajoso por el

hecho de que el extractor de características 104 opera de manera más eficiente debido a que la señal de audio no interpolada 909 tiene un número menor de muestras en comparación con una determinada porción de tiempo de la señal de audio en comparación con la señal sobremuestreada e interpolada a la salida del bloque 900.

5 **[00036]** La Fig. 10 ilustra otra forma de realización de la presente invención. A diferencia de la Fig. 9, la Fig. 10 tiene un modelo estadístico 904 que no solo ofrece una estimación de la envolvente como en la Fig. 9, sino que también presenta representaciones paramétricas adicionales que comprenden información correspondiente a la generación de tonos que faltan 1080 o la información para el filtrado inverso 1040 o la información sobre un ruido de fondo 1020 que se va a añadir. Los bloques 1020, 1040, los procedimientos de generación de envolvente espectral
10 1060 y tonos que faltan 1080 han sido descritos por la norma MPEG-4 en el contexto de HE-AAC (del inglés, Codificación de Audio Avanzada de Alta Eficiencia).

[00037] De este modo, también se pueden codificar otras señales diferentes de la voz, como se ilustra en la Fig. 10. En ese caso, podría no ser suficiente para codificar la envolvente espectral 1060 solamente, sino también
15 más información lateral tal como la tonalidad (1040), el nivel de ruido (1020) o sinusoides que faltan (1080) como se efectúa en la tecnología de replicación de la banda espectral (SBR) lustrada en [6].

[00038] En la Fig. 11 se ilustra otra forma de realización, en la que se utiliza la información lateral 114, es decir, la información lateral de selección además de la información lateral de SBR ilustrada en 1100. De este modo,
20 la información lateral de selección que comprende, por ejemplo, información con respecto a los sonidos de voz detectados se suma a la información lateral de SBR 1100 existente anteriormente. Esto ayuda a regenerar con más precisión el contenido de alta frecuencia correspondiente a los sonidos de voz tales como sibilantes, incluyendo fricativos, plosivos o vocales. De este modo, el procedimiento ilustrado en la Fig. 11 tiene la ventaja de que la información lateral de selección 114 transmitida adicionalmente sustenta una clasificación del lado del decodificador
25 (fonema) a fin de producir una adaptación del lado del decodificador de los parámetros de SBR o BWE (extensión de ancho de banda). De este modo, a diferencia de la Fig. 10, la realización de la Fig. 11 proporciona, además de la información lateral de selección, la información lateral SBR que ya existía.

[00039] La Fig. 8 ilustra una representación ejemplar de la señal de entrada codificada. La señal de entrada
30 codificada consiste en las tramas subsiguientes 800, 806, 812. Cada trama tiene la señal de núcleo codificada. A título de ejemplo, la trama 800 tiene voz como señal de núcleo codificada. La trama 806 tiene música como señal de núcleo codificada y la trama 812 tiene, de nuevo, voz como señal de núcleo codificada. La trama 800 tiene como información lateral, por ejemplo, solo la información lateral de selección y no información lateral de SBR. De este modo, la trama 800 corresponde a la Fig. 9 o a la Fig. 10. A título de ejemplo, la trama 806 comprende información
35 SBR pero no contiene información lateral de selección. Además, la trama 812 comprende una señal de voz codificada y, a diferencia de la trama 800, la trama 812 no contiene ninguna información lateral de selección. Esto se debe a que la información lateral de selección no es necesaria, ya que no se han encontrado ambigüedades en el proceso de extracción de características/ modelo estadístico del lado del codificador.

40 **[00040]** A continuación se describe la Fig. 5. Se emplea un detector de actividad de voz o un detector de voz/no voz 500 que opera en la señal de núcleo a fin de decidir si se debe emplear la tecnología de mejora del ancho de banda o la frecuencia o una tecnología diferente de extensión de ancho de banda. De este modo, cuando el detector de actividad de voz o el detector de voz/no voz detecta voz o habla, en ese caso se utiliza una primera tecnología de extensión de ancho de banda BWEXT.1 ilustrada en 511 que funciona, por ejemplo, como se describe
45 en las Figs. 1, 9, 10, 11. De este modo, los conmutadores 502, 504 están configurados de tal manera que se toman los parámetros obtenidos del generador de parámetros de la entrada 512 y el conmutador 504 conecta estos parámetros con el bloque 511. Sin embargo, el detector 500 detecta una situación en que no se muestra ninguna señal de voz sino que muestra señales de música, en ese caso se introducen preferentemente los parámetros de extensión de ancho de banda 514 del flujo de bits en el procedimiento de la otra tecnología de extensión de ancho de banda 513. De este modo, el detector 500 detecta si se debe emplear o no la tecnología de extensión de ancho
50 de banda 511 de la invención. En el caso de las señales de no voz, el codificador puede conmutar a las otras técnicas de extensión de ancho de banda ilustradas por el bloque 513, tal como se ha mencionado en [6, 8]. De este modo, el estimador de señales 118 de la Fig. 5 está configurado para conmutar a un procedimiento diferente de extensión de ancho de banda y/o al uso de diferentes parámetros extraídos de una señal codificada, cuando el
55 detector 500 detecta una actividad que no es de voz o una señal que no es de voz humana. Para esta tecnología de extensión de ancho de banda 513 diferente, es preferible que la información lateral de selección no esté presente en el flujo de bits y tampoco se utiliza, lo que está simbolizado en la Fig. 5 por la activación del conmutador 502 ante la entrada 514.

60 **[00041]** La Fig. 6 ilustra una implementación adicional del generador de parámetros 108. El generador de parámetros 108 tiene preferentemente una pluralidad de modelos estadísticos tal como un primer modelo estadístico 600 y un segundo modelo estadístico 602. Además, se incluye un selector 604 que es controlado por la información lateral de selección para aportar la correcta alternativa de representación paramétrica. La opción de qué modelo estadístico se activa es controlada por un clasificador de señales adicional 606 que recibe, como entrada, la señal
65 de núcleo, es decir, la misma señal introducida en el extractor de características 104. De este modo, el modelo

estadístico de la Fig. 10 o de cualquier otra figura puede variar con el contenido codificado. En el caso de la voz, se emplea un modelo estadístico que representa un modelo fuente de producción de voz, mientras que para otras señales, tales como señales de música como, por ejemplo, las clasificadas por el clasificador de señales 606 se utiliza un modelo diferente que se entrena según una gran serie de datos musicales. Hay otros modelos estadísticos
5 útiles asimismo para diferentes idiomas, etc.

[00042] Como se ha tratado anteriormente, la Fig. 7 ilustra la pluralidad de alternativas obtenidas por un modelo estadístico tal como el modelo estadístico 600. Por lo tanto, la salida del bloque 600 es, por ejemplo, para diferentes alternativas, como se ilustra en la línea paralela 605. Del mismo modo, el segundo modelo estadístico 602
10 puede emitir también una pluralidad de alternativas tales como las alternativas ilustradas en la línea 606. Dependiendo del modelo estadístico específico, es preferible que solo se emitan como salida las alternativas con elevada probabilidad con respecto al extractor de características 104. De este modo, un modelo estadístico presenta, en respuesta a una característica, una pluralidad de representaciones paramétricas alternativas, donde cada representación paramétrica alternativa tiene una probabilidad que es idéntica a las probabilidades de otras
15 representaciones paramétricas alternativas o que es diferente de las probabilidades de las otras representaciones paramétricas alternativas en menos del 10 %. De este modo, en una forma de realización, solo la representación paramétrica con la probabilidad más elevada y un número de otras representaciones paramétricas alternativas, todas las cuales tienen una probabilidad que es solo un 10 % menor que la probabilidad de la mejor alternativa con mejor coincidencia.

[00043] La Fig. 12 ilustra un codificador para generar una señal codificada 1212. El codificador comprende un codificador de núcleo 1200 para codificar una señal original 1206 para obtener una señal de audio del núcleo codificada 1208 que tiene información sobre un número más pequeño de bandas de frecuencia en comparación con la señal original 1206. Además, se incluye un generador de información lateral de selección 1202 para generar
25 información lateral de selección 1210 (SSI - información lateral de selección). La información lateral de selección 1210 indica una representación paramétrica alternativa definida proporcionada por un modelo estadístico en respuesta a una característica extraída de la señal original 1206 o de la señal de audio codificada 1208 o de una versión decodificada de la señal de audio codificada. Además, el codificador comprende una interfaz de salida 1204 para emitir la señal codificada 1212. La señal codificada 1212 comprende la señal de audio codificada 1208 y la
30 información lateral de selección 1210. De preferencia, el generador de información lateral de selección 1202 está implementado de la manera ilustrada en la Fig. 13. Para este fin, el generador de información lateral de selección 1202 comprende un decodificador de núcleo 1300. Se incluye el extractor de características 1302 que opera en la señal de núcleo decodificada emitida por el bloque 1300. Se introduce la característica en un procesador de modelos estadísticos 1304 para generar un número de alternativas de representación paramétrica a fin de estimar un alcance
35 espectral de una señal mejorada en frecuencia no definida por la señal de núcleo decodificada emitida por el bloque 1300. Todas estas alternativas de representación paramétrica 1305 son introducidas en un estimador de señales 1306 para estimar una señal de audio mejorada en frecuencia 1307. A continuación se introducen todas estas señales de audio mejoradas en frecuencia 1307 en un comparador 1308 para comparar las señales de audio mejoradas en la frecuencia 1307 con la señal original 1206 de la Fig. 12. El generador de información lateral de
40 selección 1202 está configurado además para establecer la información lateral de selección 1210 de manera que la información lateral de selección defina de manera unívoca la representación paramétrica alternativa que da origen a una señal de audio mejorada en frecuencia que mejor coincide con la señal original según un criterio de optimización. El criterio de optimización puede ser un criterio basado en MMSE (mínimo error cuadrático medio), un criterio que minimiza la diferencia por muestra o preferentemente un criterio psico-acústico que minimiza la distorsión
45 percibida o cualquier otro criterio de optimización conocido por aquellos expertos en la técnica.

[00044] Aunque la Fig. 13 ilustra un procedimiento de bucle cerrado o de análisis por síntesis, la Fig. 14 ilustra una implementación alternativa de la información lateral de selección 1202 más similar a un procedimiento de bucle abierto. En la forma de realización de la Fig. 14, la señal original 1206 comprende meta información asociada
50 para el generador de información lateral de selección 1202 que describe una secuencia de información acústica (por ejemplo, anotaciones) correspondiente a una secuencia de muestras de la señal de audio original. El generador de información lateral de selección 1202 comprende, en esta forma de realización, un extractor de metadatos 1400 para extraer la secuencia de meta información y, además, un traductor de metadatos, que por lo general tiene conocimiento sobre el modelo estadístico empleado del lado del decodificador para traducir la secuencia de meta
55 información en una secuencia de información lateral de selección 1210 asociada a la señal de audio original. Los metadatos extraídos por el extractor de metadatos 1400 son descartados en el codificador y no se transmiten en la señal codificada 1212. Por el contrario, la información lateral de selección 1210 se transmite en la señal codificada junto con la señal de audio codificada 1208 generada por el codificador de núcleo que tiene un contenido de frecuencia diferente y, por lo general, un contenido de frecuencia menor en comparación con la señal decodificada
60 generada en último término o en comparación con la señal original 1206.

[00045] La información lateral de selección 1210 generada por el generador de información lateral de selección 1202 puede tener cualquiera de las características señaladas en el contexto de las figuras anteriores.

[00046] Aunque la presente invención ha sido descrita en el contexto de diagramas de bloques en los cuales

los bloques representan componentes de hardware reales o lógicos, la presente invención también puede ser implementada por un procedimiento implementado por informática. En este último caso, los bloques representan etapas correspondientes en el procedimiento, donde estas etapas representan las funcionalidades ejecutadas por los bloques de hardware lógicos o físicos correspondientes.

5

[00047] Aunque se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del procedimiento correspondiente, en el cual un bloque o dispositivo corresponde a un paso del procedimiento o a una característica de un paso del procedimiento. De manera análoga, los aspectos descritos en el contexto de un paso del procedimiento también representan una descripción de un

10

bloque o elemento correspondiente o de una característica de un aparato correspondiente. Algunas o todas las etapas del procedimiento pueden ser ejecutadas por medio de (o mediante el uso de) un aparato de hardware como, por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas formas de realización, una cualquiera o más de las etapas más importantes del procedimiento pueden ser ejecutadas por ese tipo de aparato.

15

[00048] La señal transmitida o codificada de la presente invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable tal como Internet.

20

[00049] Dependiendo de ciertos requisitos de implementación, las formas de realización de la invención pueden ser implementadas en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma, señales de control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema de computación programable de tal manera que se ejecute el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

25

[00050] Algunas formas de realización según la invención comprenden un soporte de datos que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema de computación programable de tal manera que se ejecute uno de los procedimientos descritos en esta invención.

30

[00051] En general, las formas de realización de la presente invención pueden ser implementadas en forma de producto de programa informático con un código de programa, donde el código de programa cumple la función de ejecutar uno de los procedimientos al ejecutarse el programa informático en un ordenador. El código de programa

35

puede ser almacenado, por ejemplo, en un soporte legible por máquina.

[00052] Otras formas de realización comprenden el programa informático para ejecutar uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por máquina.

40

[00053] En otras palabras, una forma de realización del procedimiento de la invención consiste, por lo tanto, en un programa informático que consta de un código de programa para realizar uno de los procedimientos descritos en esta invención al ejecutarse el programa informático en un ordenador.

45

[00054] Una forma de realización adicional del procedimiento de la invención consiste, por lo tanto, en un soporte de datos (o medio de almacenamiento digital, o medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para ejecutar uno de los procedimientos descritos en esta invención. El soporte de datos, el medio de almacenamiento digital o el medio grabado son por lo general tangibles y/o no transitorios.

50

[00055] Una forma de realización adicional del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para ejecutar uno de los procedimientos descritos en esta invención. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

55

[00056] Una forma de realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los procedimientos descritos en esta invención.

60

[00057] Una forma de realización adicional comprende un ordenador en el que se ha instalado el programa informático para ejecutar uno de los procedimientos descritos en esta invención.

65

[00058] Una forma de realización adicional según la invención comprende un aparato o sistema configurado para transferir (por ejemplo por vía electrónica u óptica) un programa informático para poner en práctica uno de los procedimientos descritos en esta invención en un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa informático al receptor.

[00059] En algunas formas de realización, se puede utilizar un dispositivo lógico programable (por ejemplo una matriz de puertas programables en el campo) para ejecutar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas formas de realización, una matriz de puertas programables
5 en el campo puede cooperar con un microprocesador para ejecutar uno de los procedimientos descritos en esta invención. Por lo general, los procedimientos son ejecutados preferentemente por cualquier aparato de hardware.

[00060] Las formas de realización anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles descritos en
10 esta invención han de ser evidentes para aquellos expertos en la técnica. Por lo tanto, es la intención limitarse únicamente al alcance de las siguientes reivindicaciones de patente y no a los detalles específicos presentados a manera de descripción y explicación de las formas de realización presentadas en esta invención.

Referencias:

15

[00061]

- [1] B. Bessette *et al.*, "The Adaptive Multi-rate Wideband Speech Codec (AMR-WB)," *IEEE Trans. on Speech y Audio Processing*, Vol. 10, No. 8, Nov. 2002.
- 20 [2] B. Geiser *eta/.*, "Wide band Extension for Hierarchical Speech and Audio Coding in ITU-T Rec. G.729.1," *IEEE Trans. on Audio, Speech, y Language Processing*, Vol. 15, No. 8, Nov. 2007.
- [3] B. Iser, W. Minker, y G. Schmidt, *Wide band Extension of Speech Signals*, Springer Lecture Notes in Electrical Engineering, Vol. 13, Nueva York, 2008.
- [4] M. Jelinek y R. Salami, "Speech Wideband Extension Coding Advances in VMR-WB Standard," *IEEE Trans. on*
25 *Audio, Speech, and Language Processing*, Vol. 15, No.4, mayo de 2007.
- [5] I. Katsir, I. Cohen, y D. Malah, "Speech Wideband Extension Coding Based on Speech Phonetic Content y Speaker Vocal Tract Shape Estimation," in *Proc. EUSIPCO 2011*, Barcelona, España, Sept. 2011.
- [6] E. Larsen y R. M. Aarts, *Audio Wideband Extension: Application of Psychoacoustics, Signal Processing y Loudspeaker Design*, Wiley, Nueva York, 2004.
- 30 [7] J. Makinen *eta/.*, "AMR-WB+: A New Audio Coding Standard for 3rd Generation Mobile Audio Services," in *Proc. /CASSP 2005*, Filadelfia, EE. UU., marzo de 2005.
- [8] M. Neuendorf *eta/.*, "MPEG Unified Speech y Audio Coding –The ISO/MPEG Standard for High-Efficiency Audio Coding of All Content Types," in *Proc. 132nd Convention of the AES*, Budapest, Hungría, abril de 2012. También aparece en the Journal of the AES, 2013.
- 35 [9] H. Pulakka y P. Alku, "Bandwidth extension of Telephone Speech Using a Neural Network y a Filter Bank Implementation for Highband Mel Spectrum," *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 19, No. 7, Sept. 2011.
- [10] T. Vaillancourt *et al.*, "ITU-T EV-VBR: A Robust 8–32 kbit/s Scalable Coder for Error Prone Telecommunications Channels," in *Proc. EUSIPCO 2008*, Lausanne, Suiza, agosto de 2008.
- 40 [11] L. Miao *et al.*, "G.711.1 Annex D y G.722 Annex B: New ITU-T Superwideband codecs," in *Proc. ICASSP 2011*, Praga, República Checa, mayo de 2011.
- [12] Bernd Geiser, Peter Jax, and Peter Vary: "ROBUST WIDEBAND ENHANCEMENT OF SPEECH BY COMBINED CODING AND ARTIFICIAL WIDEBAND EXTENSION ", *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2005
- 45 [13] P. Bauer and T. Fingscheidt: "A STATISTICAL FRAMEWORK FOR ARTIFICIAL BANDWIDTH EXTENSION EXPLOITING SPEECH WAVEFORM AND PHONETIC TRANSCRIPTION", 17th European Signal Processing Conference (EUSIPCO 2009) (17ª conferencia de procesamiento de señal europea) (EUSIPCO 2009)), agosto de 2009.

REIVINDICACIONES

1. Un decodificador para generar una señal de audio mejorada en frecuencia (120), que comprende: un extractor de características (104) para extraer una característica de una señal de núcleo (100);
- 5 un extractor de información lateral (110) para extraer una información lateral de selección asociada a la señal de núcleo; un generador de parámetros (108) para generar una representación paramétrica para estimar un alcance espectral de la señal de audio mejorada en la frecuencia (120) no definida por la señal de núcleo (100), donde el generador de parámetros (108) está configurado para presentar un número de alternativas de representación paramétrica (702, 704, 706, 708) en respuesta a la característica (112), y donde el generador de parámetros (108) está configurado para seleccionar una de las alternativas de representación paramétrica como representación paramétrica en respuesta a la información lateral de selección (712 a 718);
- 10 un estimador de señales (118) para estimar la señal de audio mejorada en la frecuencia (120) mediante el uso de la representación paramétrica seleccionada; y
- 15 un clasificador de señales (606) para clasificar una trama de la señal de núcleo (100), donde el generador de parámetros (108) está configurado para usar un primer modelo estadístico (600), cuando se clasifica una trama de la señal como perteneciente a una primera clase de señales y para usar un segundo modelo estadístico diferente (602), cuando la trama es clasificada como perteneciente a una segunda clase de señales diferentes.
- 20 2. Decodificador de la reivindicación 1, que comprende además: una interfaz de entrada (110) para recibir una señal de entrada codificada (200) que comprende una señal de núcleo codificada (201) y la información lateral de selección (114); y un decodificador de núcleo (124) para decodificar la señal de núcleo codificada para obtener la señal de núcleo
- 25 (100).
3. Decodificador de una de las reivindicaciones anteriores, en el que el generador de parámetros (108) está configurado para utilizar, cuando se selecciona una de las alternativas de representación paramétrica, un orden predefinido de alternativas de representación paramétrica o un orden de alternativas de representación paramétrica
- 30 señalizadas en el codificador.
4. Decodificador de una de las reivindicaciones anteriores, en el que el generador de parámetros (108) está configurado para proporcionar una representación de envolvente como representación paramétrica, en el que la información lateral de selección (114) indica uno de una pluralidad de sibilantes o fricativos diferentes, y
- 35 en el que el generador de parámetros (108) está configurado para proporcionar la representación de envolvente identificada por la información lateral de selección.
5. Decodificador de una de las reivindicaciones anteriores, en el que el estimador de señales (118) comprende un interpolador (900) para interpolar la señal de núcleo (100), y
- 40 en el que el extractor de características (104) está configurado para extraer la característica de la señal de núcleo (100) que no se está interpolando.
6. Decodificador de una de las reivindicaciones anteriores, en el que el estimador de señales (118) comprende:
- 45 un filtro de análisis (910) para analizar la señal de núcleo o una señal de núcleo interpolada para obtener una señal de excitación; un bloque de extensión de excitación (912) para generar una señal de excitación mejorada que tiene un intervalo espectral no incluido en la señal de núcleo (100); y un filtro de síntesis (914) para filtrar la señal de excitación extendida;
- 50 en el que el filtro de análisis (910) o el filtro de síntesis (914) está determinado por la representación paramétrica seleccionada.
7. Decodificador de una de las reivindicaciones anteriores, en el que el estimador de señales (118) comprende un procesador de extensión de ancho de banda espectral para
- 55 generar una banda espectral extendida que corresponde al intervalo espectral no incluido en la señal de núcleo mediante el uso de al menos una banda espectral de la señal de núcleo y la representación paramétrica, en el que la representación paramétrica comprende parámetros correspondientes al menos a uno entre el ajuste de la envolvente espectral (1060), una adición de ruido de fondo (1020), un filtro inverso (1040) y una adición de tonos que faltan (1080),
- 60 en el que el generador de parámetros está configurado para proporcionar, para una característica, una pluralidad de alternativas de representación paramétrica, donde cada alternativa de representación paramétrica tiene parámetros para al menos uno de un ajuste de la envolvente espectral (1060), una adición de ruido de fondo (1020), un filtrado inverso (1040) y adición de tonos que faltan (1080).
- 65 8. Decodificador de una de las reivindicaciones anteriores, que comprende además:

un detector de actividad de voz o un discriminador de voz/no voz (500), en el que el estimador de señales (118) está configurado para estimar la señal mejorada en frecuencia mediante el uso de la representación paramétrica solo cuando el detector de actividad de voz o el detector de voz/no voz (500) indica una actividad de voz o una actividad de habla.

5

9. Decodificador de la reivindicación 8, en el que el estimador de señales (118) está configurado para conmutar (502, 504) de un procedimiento de mejora en frecuencia (511) a otro procedimiento de mejora en frecuencia diferente (513) o para usar parámetros diferentes (514) extraídos de una señal codificada, cuando el detector de actividad de voz o el detector de voz/no voz (500) indica una señal de no voz o una señal que no tiene actividad de voz.

10

10. Decodificador de una de las reivindicaciones anteriores, en el que el modelo estadístico está configurado para proporcionar, en respuesta a una característica, una pluralidad de alternativas de representaciones paramétricas (702 a 708),

15

en el que cada representación paramétrica alternativa tiene una probabilidad que es idéntica a una probabilidad de una representación paramétrica alternativa diferente o que es diferente de la probabilidad de la representación paramétrica alternativa en menos de un 10 % de la probabilidad más elevada.

11. Decodificador de una de las reivindicaciones anteriores,

20

en el que la información lateral de selección solo está incluida en una trama (800) de la señal codificada, cuando el generador de parámetros (108) proporciona una pluralidad de alternativas de representación paramétrica y en el que la información lateral de selección no está incluida en una trama diferente (812) de la señal de audio codificada en la cual el generador de parámetros (108) proporciona solo una única alternativa de representación paramétrica en respuesta a la característica (112).

25

12. Decodificador para generar una señal codificada (1212), que comprende:

un codificador de núcleo (1200) para codificar una señal original (1206) para obtener una señal de audio codificada (1208) que tiene información sobre un número más pequeño de bandas de frecuencia en comparación con una señal original (1206);

30

un generador de información lateral de selección (1202) para generar información lateral de selección (1210) que indica una alternativa de representación paramétrica definida (702–708) proporcionada por un modelo estadístico en respuesta a una característica (112) extraída de la señal original (1206) o de la señal de audio codificada (1208) o de una versión decodificada de la señal de audio codificada (1208);

una interfaz de salida (1204) para emitir la señal codificada (1212), comprendiendo la señal codificada la señal de audio codificada (1208) y la información lateral de selección (1210);

35

un decodificador de núcleo (1300) para decodificar la señal de audio codificada (1208) para obtener una señal de núcleo decodificada,

en el que el generador de información lateral de selección (1202) comprende:

un extractor de características (1302) para extraer una característica de la señal de núcleo decodificada;

40

un procesador de modelos estadísticos (1304) para generar un número de alternativas de representación paramétrica (702 a 708) para estimar un intervalo espectral de una señal mejorada en frecuencia no definida por la señal de núcleo decodificada;

un estimador de señales (1306) para estimar señales de audio mejoradas en frecuencia para las alternativas de representación paramétrica (1305); y

45

un comparador (1308) para comparar las señales de audio mejoradas en frecuencia (1307) con la señal original (1206),

donde el generador de información lateral de selección (1202) está configurado para establecer la información lateral de selección (1210) de tal manera que la información lateral de selección defina de manera unívoca la alternativa de representación paramétrica que resulta en una señal de audio mejorada en frecuencia que mejor coincide con la señal original (1206) según un criterio de optimización.

50

13. Codificador de la reivindicación 12,

en el que la interfaz de salida (1204) está configurada para incluir solo la información lateral de selección (1210) en la señal codificada (1212), cuando el modelo estadístico proporciona una pluralidad de alternativas de representación paramétrica y para no incluir ninguna información lateral de selección en una trama para la señal de audio codificada (1208), en la cual el modelo estadístico es operativo para proporcionar solo una única representación paramétrica en respuesta a la característica.

55

14. Un procedimiento para generar una señal de audio mejorada en frecuencia (120), que comprende:

60

extracción (104) de una característica de una señal de núcleo (100);

extracción (110) de una información lateral de selección asociada a la señal de núcleo;

generación (108) de una representación paramétrica para estimar un intervalo espectral de la señal de audio mejorada en frecuencia (120) no definida por la señal de núcleo (100), donde se presenta un número de alternativas

- de representación paramétrica (702, 704, 706, 708) en respuesta a la característica (112), y donde se selecciona una de las alternativas de representación paramétrica como representación paramétrica en respuesta a la información lateral de selección (712 a 718);
- 5 estimación (118) de la señal de audio mejorada en frecuencia (120) mediante el uso de la representación paramétrica seleccionada; y
- 10 clasificación (606) de una trama de la señal de núcleo (100), en la que la generación (108) utiliza un primer modelo estadístico (600), cuando una trama de señal está clasificada para pertenecer a una primera clase de señales y utiliza un segundo modelo estadístico diferente (602), cuando la trama está clasificada en una segunda clase de señales diferente.
15. Procedimiento de generación de una señal codificada (1212), que comprende:
- codificación (1200) de una señal original (1206) para obtener una señal de audio codificada (1208) que tiene información sobre un número menor de bandas de frecuencia en comparación con una señal original (1206);
- 15 generación (1202) de información lateral de selección (1210) que indica una alternativa de representación paramétrica definida (702-708) proporcionada por un modelo estadístico en respuesta a una característica (112) extraída de la señal original (1206) o de la señal de audio codificada (1208) o de una versión decodificada de la señal de audio codificada (1208);
- emisión como salida (1204) de la señal codificada (1212), comprendiendo la señal codificada la señal de audio codificada (1208) y la información lateral de selección (1210); y
- 20 clasificación (606) de una trama de la señal de núcleo (100), en la que la generación (108) utiliza un primer modelo estadístico (600), cuando una trama de señal está clasificada para pertenecer a una primera clase de señales y utiliza un segundo modelo estadístico diferente (602), cuando la trama está clasificada en una segunda clase de señales diferente.
- 25 16. Programa informático para llevar a cabo, cuando se ejecuta en un ordenador o un procesador, el procedimiento de la reivindicación 14 o el procedimiento de la reivindicación 15.

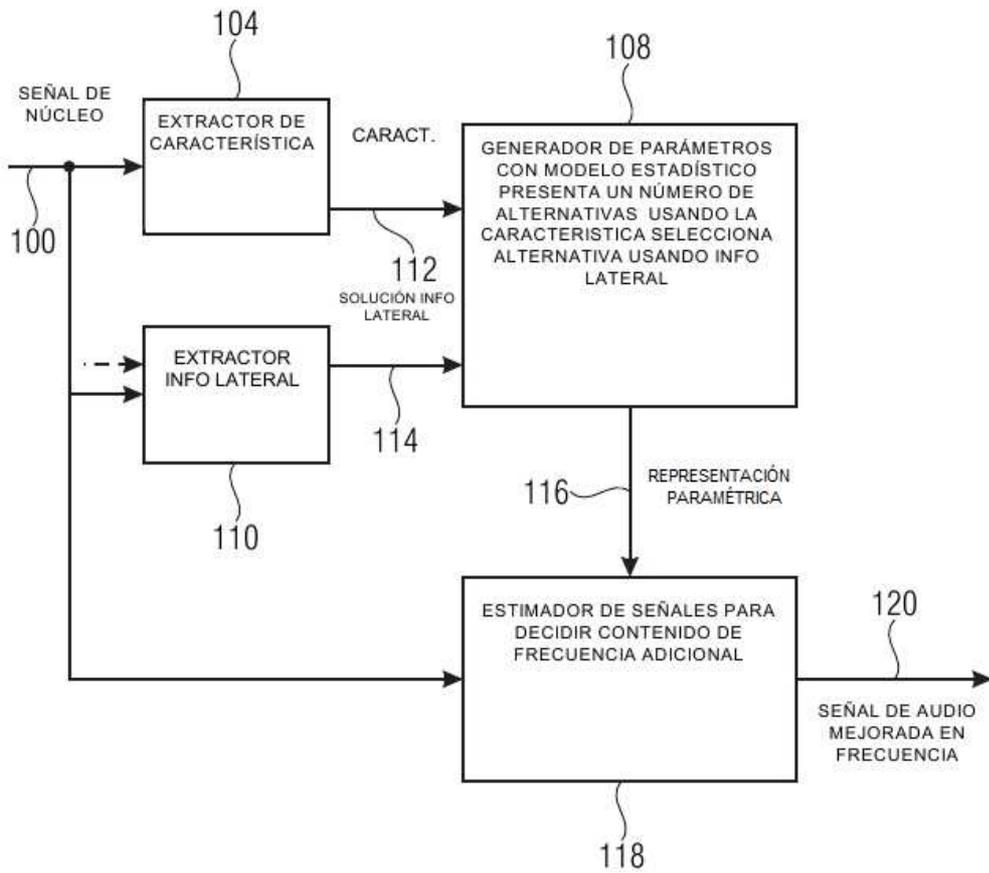


FIG 1

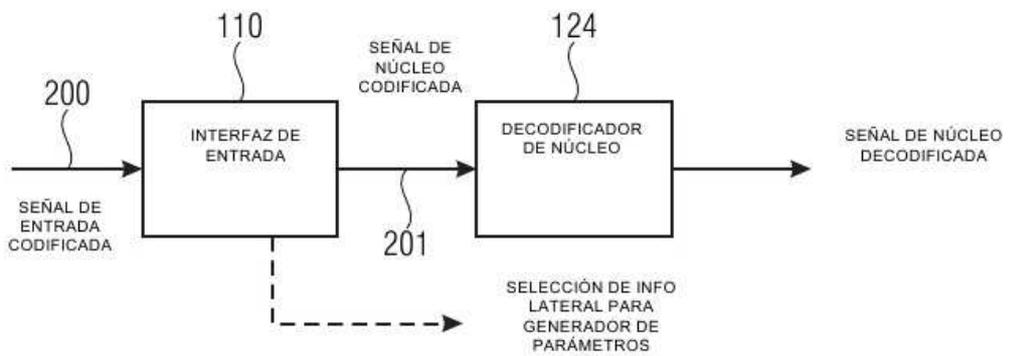


FIG 2

| N° DE BITS DE INFO LATERAL SELEC. | N° DE ALTERN. DE REPRES. PARAMÉT. |
|-----------------------------------|-----------------------------------|
| 1 | 2 |
| 2 | 4 |
| 3 | 8 |
| 4 | 16 |
| 5 | 32 |

FIG 3

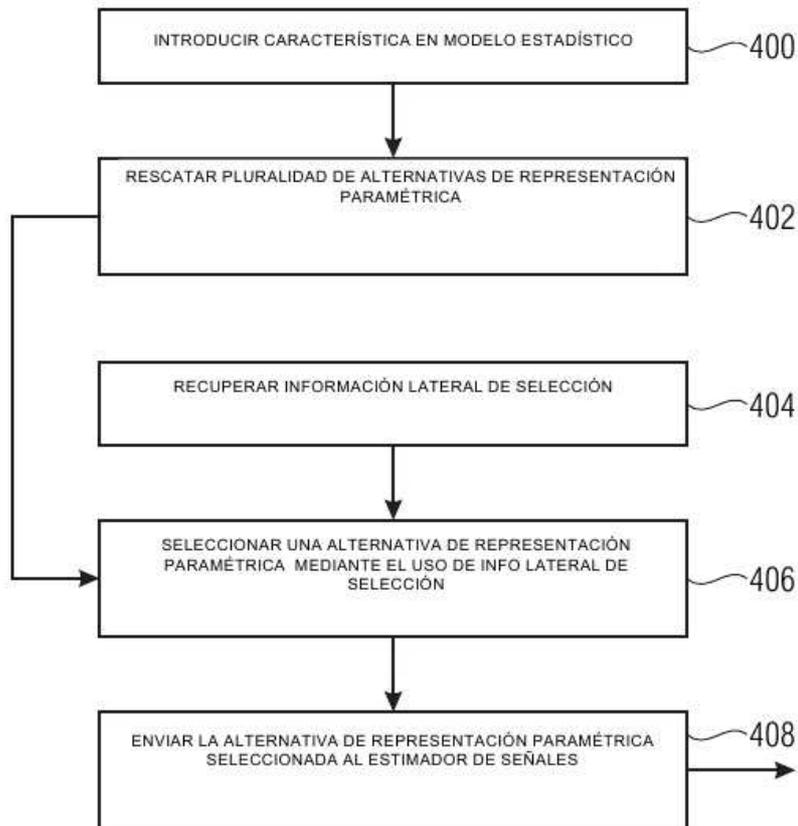


FIG 4

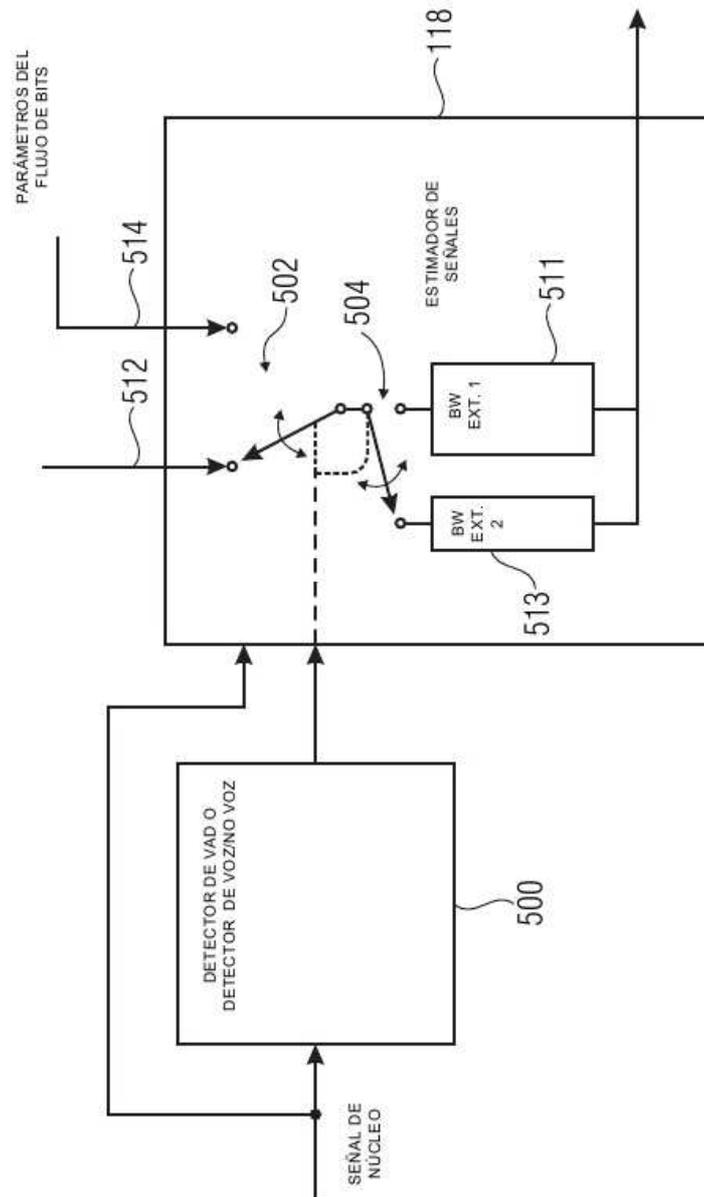


FIG 5

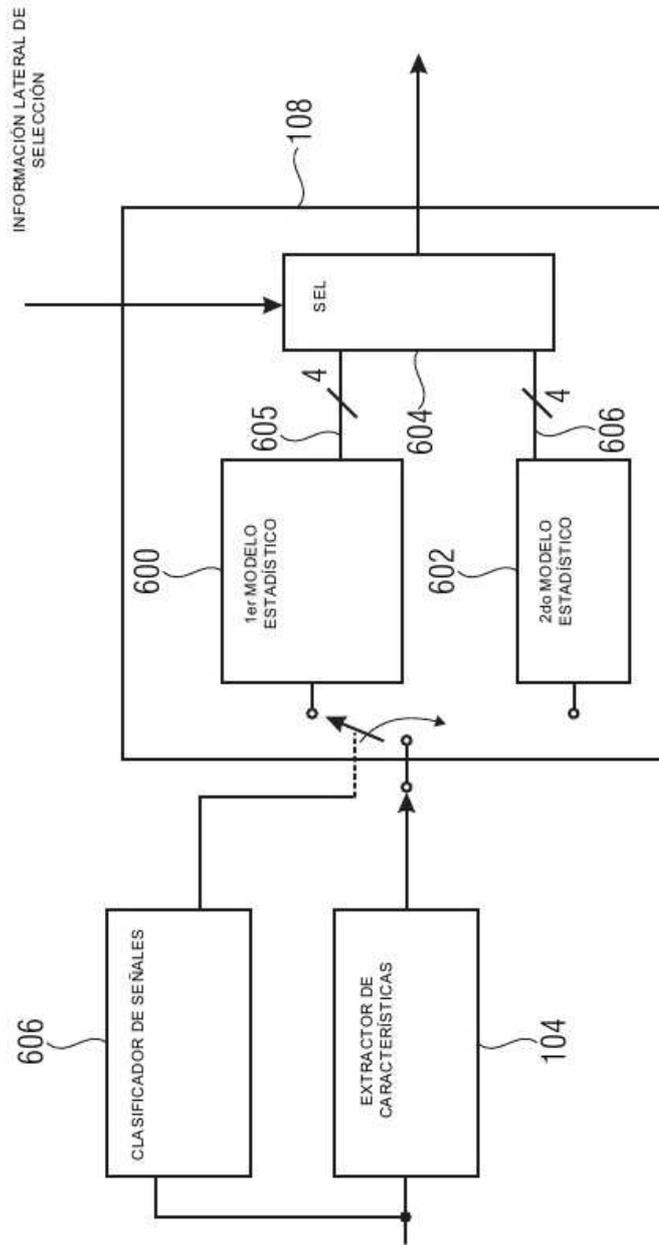


FIG 6

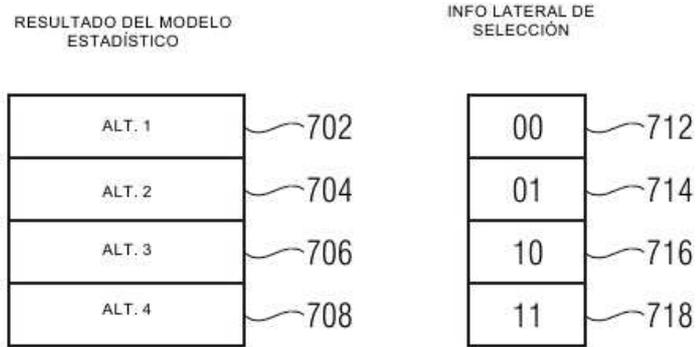


FIG 7

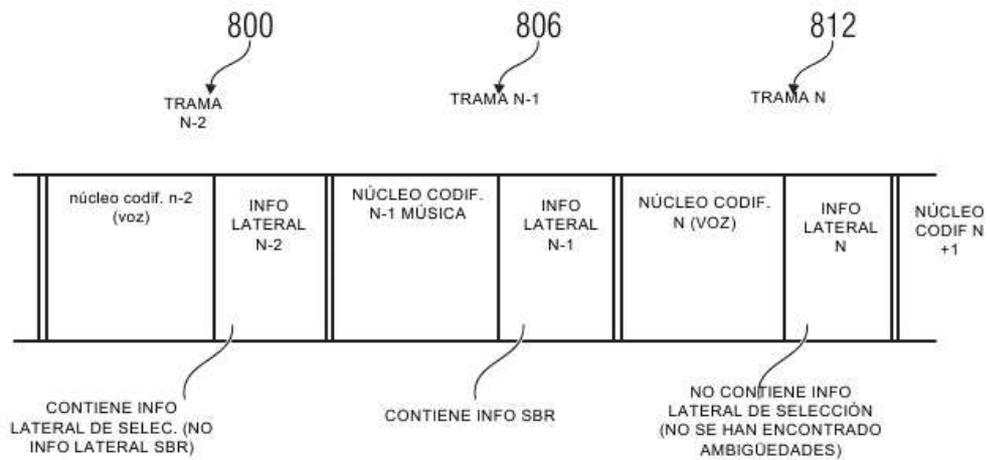


FIG 8

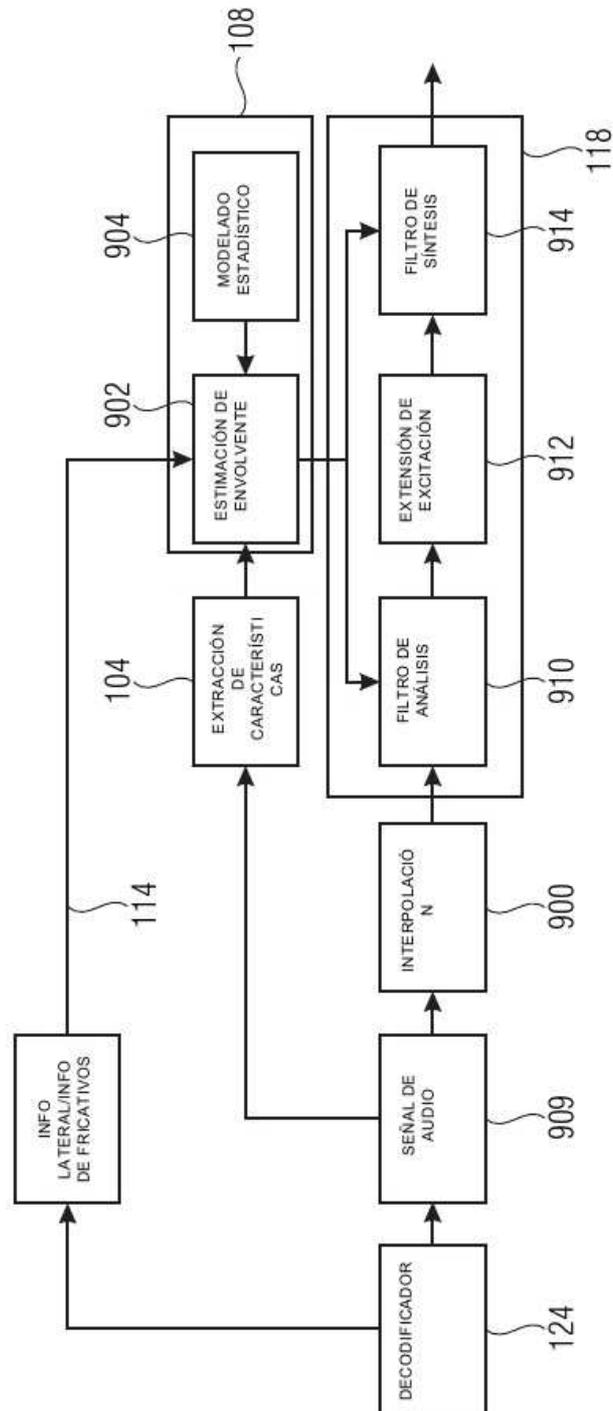


FIG 9

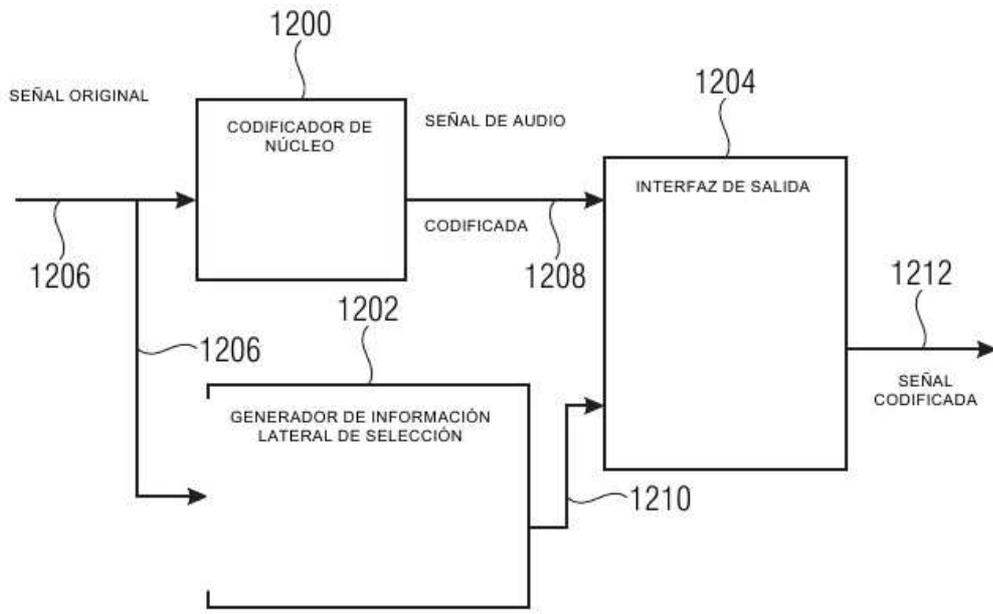


FIG 12

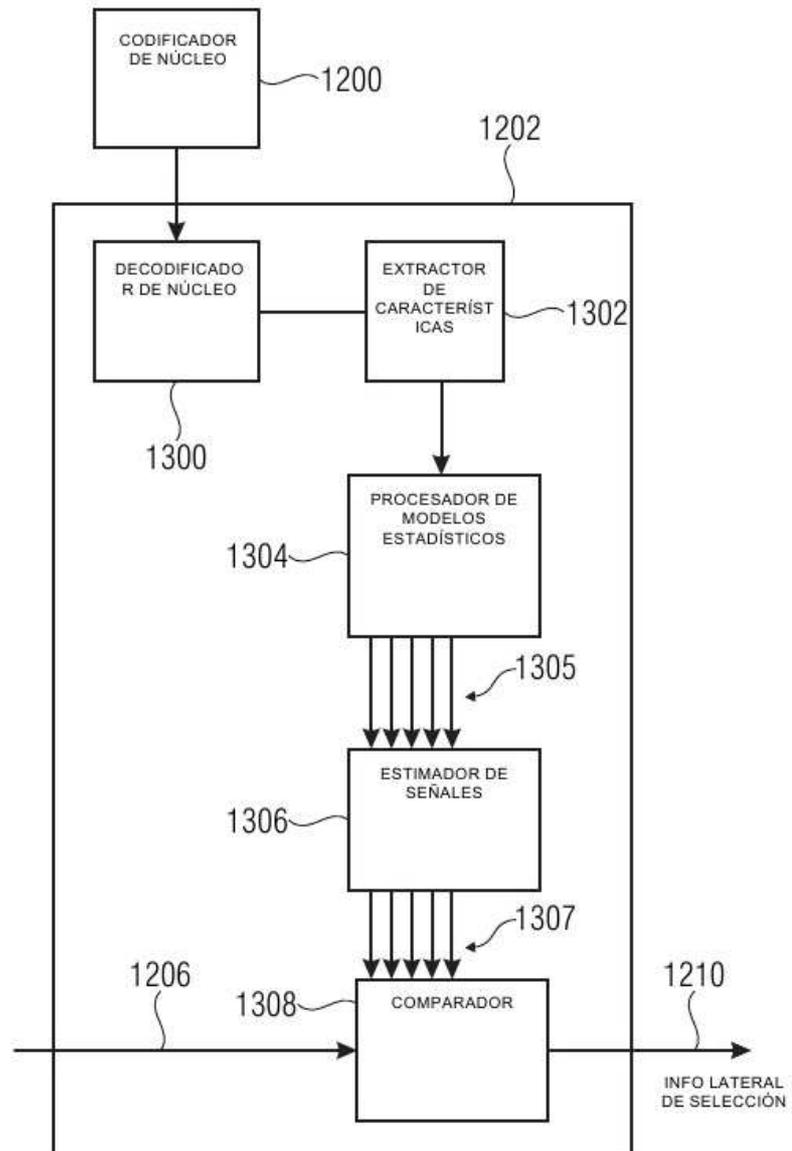


FIG 13

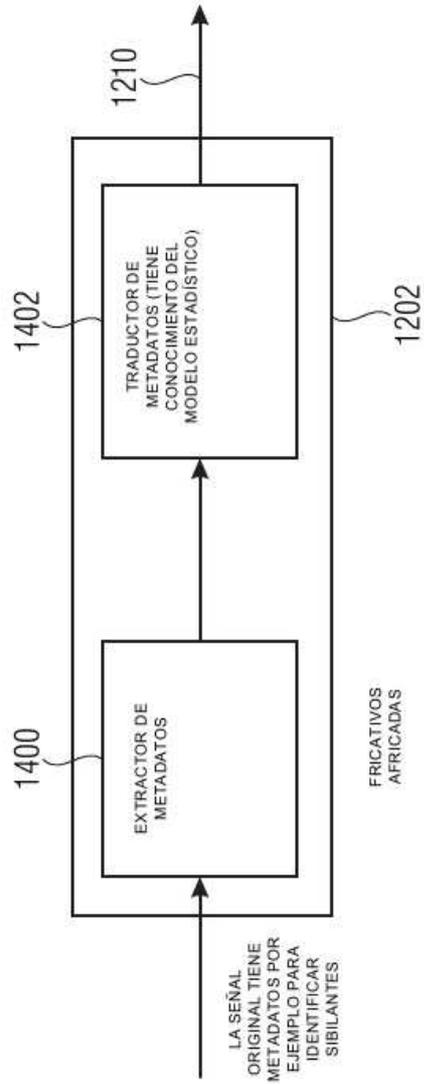


FIG 14

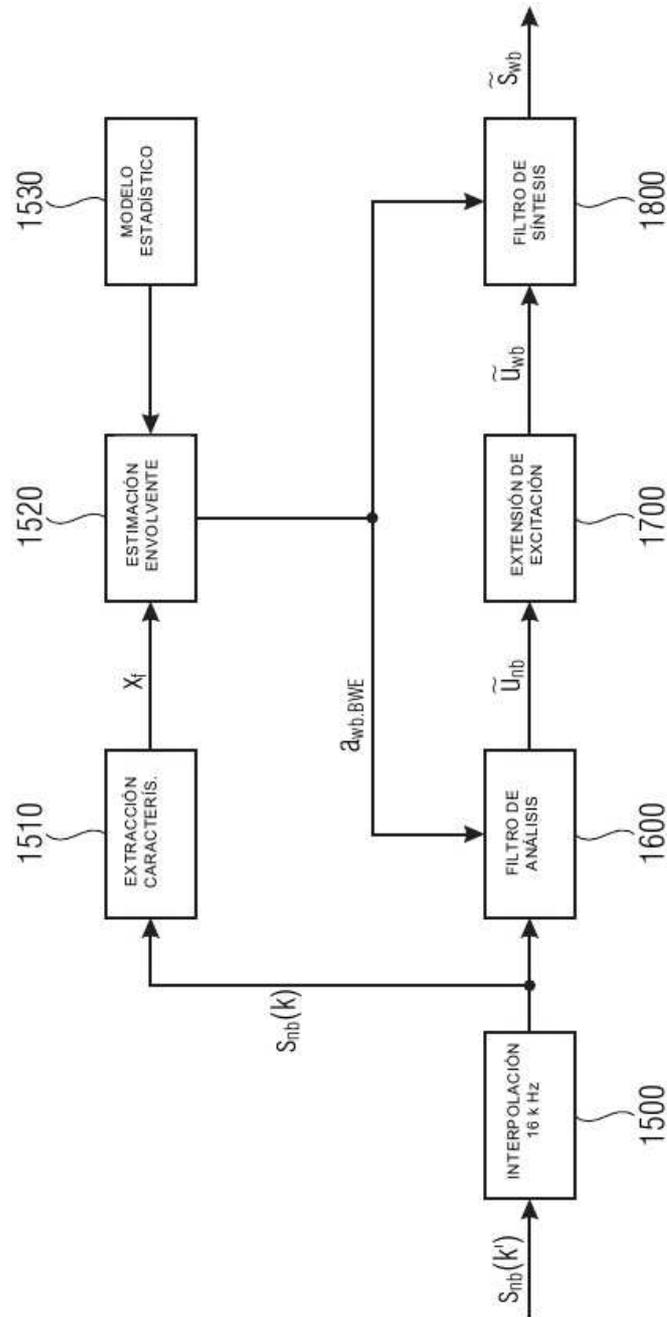


FIG 15
(TÉCNICA ANTERIOR)

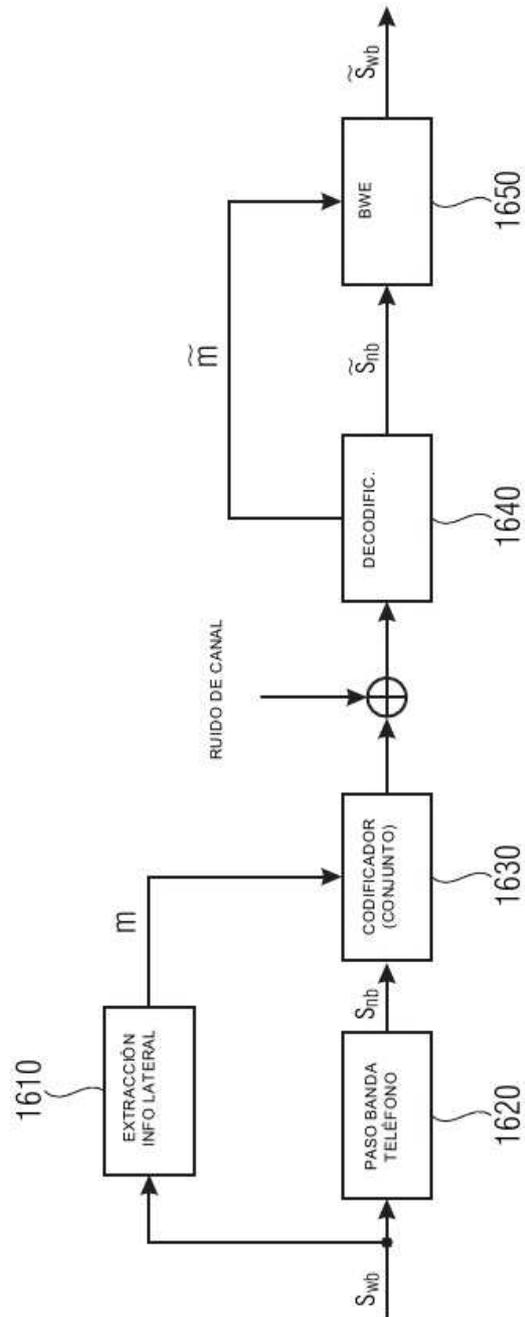


FIG 16
(TÉCNICA ANTERIOR)