

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 698 023**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 21/038 (2013.01)

G10L 19/18 (2013.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.07.2014 PCT/EP2014/065106**

87 Fecha y número de publicación internacional: **29.01.2015 WO15010947**

96 Fecha de presentación y número de la solicitud europea: **15.07.2014 E 14739160 (1)**

97 Fecha y número de publicación de la concesión europea: **01.08.2018 EP 3025328**

54 Título: **Decodificador de audio y método relacionado que usan procesamiento de dos canales dentro de un marco de relleno inteligente de huecos**

30 Prioridad:

22.07.2013 EP 13177353

22.07.2013 EP 13177350

22.07.2013 EP 13177348

22.07.2013 EP 13177346

18.10.2013 EP 13189366

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.01.2019

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)**

**Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**DISCH, SASCHA;
NAGEL, FREDERIK;
GEIGER, RALF;
THOSKAHNA, BALAJI NAGENDRAN;
SCHMIDT, KONSTANTIN;
BAYER, STEFAN;
NEUKAM, CHRISTIAN;
EDLER, BERND y
HELMRICH, CHRISTIAN**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 698 023 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 de audio y método relacionado que usan procesamiento de dos canales dentro de un marco de relleno inteligente de huecos

5

La presente invención se refiere a la codificación/decodificación de audio y, en particular, a la codificación de audio usando relleno inteligente de huecos (IGF).

10 La codificación de audio es el campo de la compresión de señales que trata el aprovechamiento de la redundancia y la irrelevancia en señales de audio usando el conocimiento psicoacústico. Actualmente, los códecs de audio generalmente necesitan alrededor de 60 kbps/canal para la codificación perceptualmente transparente de casi cualquier tipo de señal de audio. Los códecs más nuevos tienen como objetivo reducir la tasa de bits de codificación aprovechando similitudes espectrales en la señal usando técnicas tales como la extensión de ancho de banda (BWE). Un esquema de BWE usa un conjunto de parámetros de tasa de bits baja para representar las componentes de alta frecuencia (HF) de una señal de audio. El espectro de HF se rellena con contenido espectral de regiones de baja frecuencia (LF) y la forma espectral, pendiente y continuidad temporal se ajustan para mantener el timbre y el color de la señal original. Tales métodos de BWE permiten que los códecs de audio conserven una buena calidad incluso a tasas de bits bajas de alrededor de 24 kbps/canal.

20

El almacenamiento o la transmisión de señales de audio a menudo están sujetos a estrictas limitaciones de tasas de bits. En el pasado, los codificadores se veían obligados a reducir drásticamente el ancho de banda de audio transmitido cuando sólo estaba disponible una tasa de bits muy baja.

25 Los códecs de audio modernos son ahora capaces de codificar señales de banda ancha usando métodos de extensión de ancho de banda (BWE) [1]. Estos algoritmos se basan en una representación paramétrica del contenido de alta frecuencia (HF), que se genera a partir de la parte de baja frecuencia (LF) con forma de onda codificada de la señal decodificada por medio de transposición a la región espectral de HF ("interconexión") y la aplicación de un procesamiento posterior impulsado por parámetros. En los esquemas de BWE, la reconstrucción de la región espectral de HF por encima de una denominada frecuencia de cruce dada se basa a menudo en la interconexión espectral. En general, la región de HF está compuesta por múltiples interconexiones adyacentes y cada una de estas interconexiones se origina a partir de regiones de paso de banda (BP) del espectro de LF por debajo de la frecuencia de cruce dada. Los sistemas del estado de la técnica realizan de manera eficiente la interconexión dentro de una representación de banco de filtros, por ejemplo banco de filtros de espejo en cuadratura (QMF), copiando un conjunto de coeficientes de subbandas adyacentes desde una región de origen a la región objetivo.

40 Otra técnica que se encuentra en los códecs de audio actuales que aumenta la eficiencia de compresión y permite así un ancho de banda de audio extendido a tasas de bits bajas es el reemplazo sintético impulsado por parámetros de partes apropiadas de los espectros de audio. Por ejemplo, las porciones de señal tipo ruido de la señal de audio original pueden reemplazarse sin pérdida sustancial de calidad subjetiva por ruido artificial generado en el decodificador y ajustado a escala por parámetros de información secundaria. Un ejemplo es la herramienta de sustitución de ruido perceptual (PNS) contenida en la codificación avanzada de audio MPEG-4 (AAC) [5].

45 Una disposición adicional que también permite un ancho de banda de audio extendido a tasas de bits bajas es la técnica de relleno de ruido contenida en la codificación unificada de voz y audio MPEG-D (USAC) [7]. Los huecos espectrales (ceros) que se deducen por la zona muerta del cuantificador debido a una cuantificación demasiado gruesa, se rellenan posteriormente con ruido artificial en el decodificador y se ajustan a escala mediante un procesamiento posterior impulsado por parámetros.

50

Otro sistema del estado de la técnica se denomina reemplazo espectral preciso (ASR) [2-4]. Además de un códec de forma de onda, ASR emplea una etapa de síntesis de señal dedicada que restaura porciones sinusoidales perceptualmente importantes de la señal en el decodificador. Asimismo, un sistema descrito en [5] se basa en el modelado sinusoidal en la región de HF de un codificador de forma de onda para permitir un ancho de banda de audio extendido que tiene una calidad perceptual decente a tasas de bits bajas. Todos estos métodos implican la transformación de los datos en un segundo dominio aparte de la transformada de coseno discreta modificada (MDCT) y también etapas de análisis/síntesis bastante complejas para la conservación de componentes sinusoidales de HF.

60 La figura 13a ilustra un diagrama esquemático de un codificador de audio para una tecnología de extensión de ancho de banda como se usa, por ejemplo, en la codificación avanzada de audio de alta eficiencia (HE-AAC). Una señal de audio en la línea 1300 se introduce en un sistema de filtro que comprende un paso 1302 bajo y un paso 1304 alto. La señal emitida por el filtro 1304 de paso alto se introduce en un extractor/codificador 1306 de parámetros. El extractor/codificador 1306 de parámetros está configurado para calcular y codificar parámetros tales

como, por ejemplo, un parámetro de envolvente espectral, un parámetro de adición de ruido, un parámetro de armónicos faltantes o un parámetro de filtrado inverso. Estos parámetros extraídos se introducen en un multiplexor 1308 de flujo de bits. La señal de salida de paso bajo se introduce en un procesador que generalmente comprende la funcionalidad de un muestreador 1310 descendente y un codificador 1312 principal. El paso 1302 bajo restringe el ancho de banda que va a codificarse a un ancho de banda significativamente menor que el producido en la señal de audio de entrada original en la línea 1300. Esto proporciona una ganancia de codificación significativa debido al hecho de que la totalidad de las funcionalidades que se producen en el codificador principal sólo tienen que funcionar en una señal con un ancho de banda reducido. Cuando, por ejemplo, el ancho de banda de la señal de audio en la línea 1300 es de 20 kHz y cuando el filtro 1302 de paso bajo tiene a modo de ejemplo un ancho de banda de 4 kHz, con el fin de cumplir el teorema de muestreo, teóricamente es suficiente con que la señal tras el muestreador descendente tenga una frecuencia de muestreo de 8 kHz, lo cual es una reducción sustancial de la tasa de muestreo requerida para la señal 1300 de audio que tiene que ser de al menos 40 kHz.

La figura 13b ilustra un diagrama esquemático de un decodificador de extensión de ancho de banda correspondiente. El decodificador comprende un multiplexor 1320 de flujo de bits. El demultiplexor 1320 de flujo de bits extrae una señal de entrada para un decodificador 1322 principal y una señal de entrada para un decodificador 1324 de parámetros. Una señal de salida del decodificador principal tiene, en el ejemplo anterior, una tasa de muestreo de 8 kHz y, por lo tanto, un ancho de banda de 4 kHz mientras que, para una reconstrucción completa de ancho de banda, la señal de salida de un reconstructor 1330 de alta frecuencia debe ser de 20 kHz lo que requiere una tasa de muestreo de al menos 40 kHz. Con el fin de hacer esto posible, se requiere un procesador de decodificador que tenga la funcionalidad de un muestreador 1325 ascendente y un banco 1326 de filtros. El reconstructor 1330 de alta frecuencia recibe entonces la señal de baja frecuencia con frecuencia analizada emitida por el banco 1326 de filtros y reconstruye el intervalo de frecuencia definido por el filtro 1304 de paso alto de la figura 13a usando la representación paramétrica de la banda de alta frecuencia. El reconstructor 1330 de alta frecuencia tiene varias funcionalidades tales como la regeneración del intervalo de frecuencia superior usando el intervalo de origen en el intervalo de baja frecuencia, un ajuste de envolvente espectral, una funcionalidad de adición de ruido y una funcionalidad para introducir armónicos faltantes en el intervalo de frecuencia superior y, si se aplica y calcula en el codificador de la figura 13a, una operación de filtrado inverso con el fin de tener en cuenta el hecho de que el intervalo de frecuencia superior normalmente no es tan tonal como el intervalo de frecuencia inferior. En HE-AAC, los armónicos faltantes se re-sintetizan en el lado de decodificador y se colocan exactamente en el centro de una banda de reconstrucción. Por lo tanto, todas las líneas de armónicos faltantes que se han determinado en una determinada banda de reconstrucción no se colocan a los valores de frecuencia en los que estaban ubicadas en la señal original. En cambio, dichas líneas de armónicos faltantes se colocan a frecuencias en el centro de la banda determinada. Por lo tanto, cuando una línea de armónico faltante en la señal original estaba colocada muy cerca del límite de la banda de reconstrucción en la señal original, el error en la frecuencia introducida al colocar esta línea de armónico faltante en la señal reconstruida en el centro de la banda es próxima al 50% de la banda de reconstrucción individual, para la que se han generado y transmitido parámetros.

Además, a pesar de que los codificadores principales de audio típicos funcionan en el dominio espectral, el decodificador principal genera no obstante una señal de dominio de tiempo que después se convierte de nuevo en un dominio espectral por la funcionalidad de banco 1326 de filtros. Esto introduce retardos de procesamiento adicionales, puede introducir artefactos debido al procesamiento en tándem de la transformación en primer lugar del dominio espectral en el dominio de frecuencia y la transformación de nuevo generalmente en un dominio de frecuencia diferente y, por supuesto, esto también requiere una cantidad sustancial de complejidad computacional y, por lo tanto, potencia eléctrica, lo cual es específicamente un problema cuando se aplica la tecnología de extensión de ancho de banda en dispositivos móviles como teléfonos móviles, tabletas u ordenadores portátiles, etc.

Los códecs de audio actuales llevan a cabo la codificación de audio a tasa de bits baja usando BWE como parte integral del esquema de codificación. Sin embargo, las técnicas de BWE se limitan a reemplazar sólo contenido de alta frecuencia (HF). Asimismo, no permiten la codificación de forma de onda de contenido perceptualmente importante por encima de una frecuencia de cruce dada. Por lo tanto, los códecs de audio actuales pierden o bien detalle de HF o bien timbre cuando se implementa la BWE, ya que en la mayoría de los sistemas no se tiene en cuenta la alineación exacta de los armónicos tonales de la señal.

Otra desventaja de los sistemas de BWE del estado de la técnica actual es la necesidad de transformación de la señal de audio en un nuevo dominio para la implementación de la BWE (por ejemplo, transformación del dominio de MDCT al de QMF). Esto genera complicaciones de sincronización, complejidad computacional adicional y aumento de requisitos de memoria.

En el caso de pares de dos canales, existen básicamente varias representaciones de canales tales como una representación de canales conjunta o una representación de canales independiente. Una representación conjunta bien conocida es una representación central/lateral en la que el canal central es la suma del canal izquierdo y derecho y en la que el canal lateral es la diferencia entre el canal izquierdo y derecho.

Otra representación es un canal de mezclado descendente y un canal residual y un coeficiente de predicción adicional que permite recrear el canal izquierdo y derecho a partir de la mezcla descendente y el residual. En este caso, la representación independiente sería el canal independiente izquierdo o derecho o, de manera general, el primer canal y el segundo canal.

5 Adicionalmente, existe una situación en la que un intervalo de origen para operaciones de relleno de huecos podrían mostrar una fuerte correlación mientras que el intervalo objetivo no muestra esta fuerte correlación. Cuando el intervalo de origen se codifica, en esta realización, usando una primera representación en estéreo tal como una representación central/lateral con el fin de reducir la tasa de bits para la porción de frecuencia principal, entonces se genera una imagen de dos canales errónea para la porción de reconstrucción o el intervalo objetivo. Por otro lado, cuando el intervalo de origen no muestra ninguna correlación o sólo tiene una pequeña correlación y el intervalo objetivo tiene una pequeña correlación o ninguna correlación, entonces nuevamente una operación de relleno de huecos directa daría como resultado artefactos.

15 El documento US 2004/008615 A1 da a conocer un método y un aparato para reducir la cantidad de cálculo en un procesamiento posterior de audio y, más particularmente, un método y un aparato para realizar un procesamiento posterior de audio que usan información de modo que indica el grado de similitud entre una señal de canal derecho y una señal de canal izquierdo en datos de audio de capa 3 de MPEG-1. Si la diferencia entre las dos señales de canal es pequeña, se usa un primer modo en el que se recupera la componente de alta frecuencia de tan sólo un canal y la componente de alta frecuencia recuperada se usa para recuperar la componente de alta frecuencia del otro canal. Si la diferencia entre las dos señales de canal es grande, se selecciona un segundo modo en el que se recupera la componente de alta frecuencia en tan sólo una de cada dos tramas de manera alternante en el canal izquierdo y el canal derecho y la componente de alta frecuencia de cada una de las tramas omitidas se interpola basándose en las componentes de alta frecuencia de la trama anterior y la trama siguiente. Haciendo esto, el nuevo método y aparato de decodificación de audio recuperan componentes de alta frecuencia con una pequeña cantidad de cálculo. El método reduce la cantidad de cálculo hasta menos de la mitad de la cantidad de cálculo usada en la técnica anterior en la recuperación de componentes de alta frecuencia.

30 El documento US 2005/0165611 A1 da a conocer un codificador de audio que usa similitud perceptual de amplio sentido. El codificador de audio mejora la calidad codificando una versión perceptualmente similar de los coeficientes espectrales omitidos, representada como una versión ajustada a escala de un espectro ya codificado. Los coeficientes espectrales omitidos se dividen en varias subbandas. Las subbandas se codifican como dos parámetros: un factor de escala, que puede representar la energía en la banda; y un parámetro de forma, que puede representar una forma de la banda. El parámetro de forma puede estar en forma de un vector de movimiento que apunta a una porción del espectro ya codificado, un índice con respecto a una forma espectral en un libro de códigos fijo, o un vector de ruido aleatorio. Por tanto, la codificación representa eficazmente una versión ajustada a escala de una porción de forma similar del espectro que va a copiarse en la decodificación.

40 Por lo tanto un objetivo de la presente invención es proporcionar un concepto mejorado de codificación/decodificación para representaciones de dos canales.

Este objetivo se logra mediante un decodificador de audio según la reivindicación 1, un método de decodificación de audio según la reivindicación 13 o programa informático según la reivindicación 14.

45 La presente invención se basa en el descubrimiento de que la situación de correlación no sólo es importante para el intervalo de origen sino que también es importante para el intervalo objetivo. Adicionalmente, la presente invención reconoce la situación de que pueden producirse diferentes situaciones de correlación en el intervalo de origen y el intervalo objetivo. Cuando, por ejemplo, se considera una señal de voz con ruido de alta frecuencia, la situación puede ser que la banda de baja frecuencia que comprende la señal de voz con un pequeño número de sobretonos está altamente correlacionada en el canal izquierdo y el canal derecho, cuando la persona que habla se encuentra situada en el medio. Sin embargo, la porción de alta frecuencia puede estar fuertemente no correlacionada debido al hecho de que puede haber un ruido de alta frecuencia diferente en el lado izquierdo en comparación con otro ruido de alta frecuencia o nada de ruido de alta frecuencia en el lado derecho. Por tanto, cuando se realiza una operación de relleno de huecos directa que ignora esta situación, entonces también se correlacionará la porción de alta frecuencia, y esto puede generar graves artefactos de segregación espacial en la señal reconstruida. Con el fin de abordar este problema, se calculan datos paramétricos para una banda de reconstrucción o, generalmente, para el segundo conjunto de segundas porciones espectrales que tienen que reconstruirse usando un primer conjunto de primeras porciones espectrales para identificar ya sea una primera o una segunda representación de dos canales diferente para la segunda porción espectral o, dicho de otro modo, para la banda de reconstrucción. Por tanto, en el lado de codificador se calcula una identificación de dos canales para las segundas porciones espectrales, es decir, para las porciones para las que, adicionalmente, se calcula información de energía para las bandas de reconstrucción. Un regenerador de frecuencia en el lado de codificador regenera entonces una segunda porción espectral dependiendo de una primera porción del primer conjunto de primeras porciones espectrales, es decir, el intervalo de origen y datos paramétricos para la segunda porción tales como información de energía de envolvente

espectral o cualquier otro dato de envolvente espectral y, adicionalmente, dependiendo de la identificación de dos canales para la segunda porción, es decir, para esta banda de reconstrucción en consideración.

5 La identificación de dos canales se transmite de manera preferente como indicador para cada banda de reconstrucción y estos datos se transmiten desde un codificador a un decodificador y entonces el decodificador decodifica la señal principal según se indica por indicadores preferiblemente calculados para las bandas principales. Después, en una implementación, la señal principal se almacena en ambas representaciones en estéreo (por ejemplo izquierda/derecha y central/lateral) y, para el relleno de recuadros de frecuencia IGF, se elige la representación de recuadro de origen para ajustarse a la representación de recuadro objetivo según se indica por los indicadores de identificación de dos canales para el relleno inteligente de huecos o las bandas de reconstrucción, es decir, para el intervalo objetivo.

15 Se enfatiza que este procedimiento no sólo funciona para las señales estéreo, es decir para un canal izquierdo y el canal derecho, sino que también funciona para señales multicanal. En el caso de señales multicanal se pueden procesar varios pares de canales diferentes de esa forma tales como un canal izquierdo y uno derecho como primer par, un canal envolvente izquierdo y un canal envolvente derecho como segundo par y un canal central y un canal LFE como tercer par. Pueden determinarse otros emparejamientos para formatos de canales de salida superiores tales como 7.1, 11.1 y así sucesivamente.

20 Un aspecto adicional se basa en el descubrimiento de que los problemas relacionados con la separación de la extensión de ancho de banda por un lado y la codificación principal por otro lado pueden abordarse y superarse llevando a cabo la extensión de ancho de banda en el mismo dominio espectral en el que funciona el decodificador principal. Por lo tanto, se proporciona un decodificador principal de tasa completa que codifica y decodifica todo el intervalo de señal de audio. Esto no requiere la necesidad de un muestreador descendente en el lado de codificador y un muestreador ascendente en el lado de decodificador. En cambio, todo el procesamiento se lleva a cabo en toda la tasa de muestreo o en todo el dominio de ancho de banda. Con el fin de obtener una alta ganancia de codificación, se analiza la señal de audio con el fin de encontrar un primer conjunto de primeras porciones espectrales que tiene que codificarse con una alta resolución, en el que este primer conjunto de primeras porciones espectrales puede incluir, en una realización, porciones tonales de la señal de audio. Por otra parte, las componentes no tonales o ruidosas en la señal de audio que constituyen un segundo conjunto de segundas porciones espectrales se codifican paramétricamente con una baja resolución espectral. Entonces, la señal de audio codificada sólo requiere el primer conjunto de primeras porciones espectrales codificado de manera que conserva la forma de onda con una alta resolución espectral y, adicionalmente, el segundo conjunto de segundas porciones espectrales codificado paramétricamente con una baja resolución usando "recuadros" de frecuencia obtenidos a partir del primer conjunto. En el lado de decodificador, el decodificador principal, que es un decodificador de banda completa, reconstruye el primer conjunto de primeras porciones espectrales de manera que conserva la forma de onda, es decir, sin ningún conocimiento de que hay una regeneración de frecuencia adicional. Sin embargo, el espectro así generado tiene muchos huecos espectrales. Estos huecos se rellenan posteriormente con la tecnología de relleno inteligente de huecos (IGF) de la invención usando una regeneración de frecuencia que aplica datos paramétricos por un lado, y que usa un intervalo espectral de origen, es decir, primeras porciones espectrales reconstruidas por el decodificador de audio de tasa completa, por otro lado.

45 En realizaciones adicionales, porciones espectrales, que se reconstruyen solamente mediante relleno de ruido en lugar de replicación de ancho de banda o relleno de recuadros de frecuencia, constituyen un tercer conjunto de terceras porciones espectrales. Debido al hecho de que el concepto de codificación funciona en un único dominio para la codificación/decodificación principal por un lado, y la regeneración de frecuencia por otro lado, el IGF no sólo se limita a rellenar un intervalo de frecuencia superior sino que puede rellenar intervalos de frecuencia inferiores, ya sea por relleno de ruido sin regeneración de frecuencia o por regeneración de frecuencia usando un recuadro de frecuencia en un intervalo de frecuencia diferente.

50 Además, se enfatiza que una información sobre energías espectrales, una información sobre energías individuales o una información de energía individual, una información sobre una energía de conservación o una información de energía de conservación, una información sobre una energía de recuadro o una información de energía de recuadro, o una información sobre una energía faltante o una información de energía faltante, pueden comprender no sólo un valor de energía sino también un valor de amplitud (por ejemplo, absoluta), un valor de nivel o cualquier otro valor, del que se puede derivar un valor de energía final. Por lo tanto, la información sobre una energía puede comprender, por ejemplo, el propio valor de energía y/o un valor de un nivel y/o de una amplitud y/o de una amplitud absoluta.

60 Un aspecto adicional se basa en el descubrimiento de que la calidad del audio de la señal reconstruida se puede mejorar a través de IGF puesto que el espectro completo es accesible para el codificador principal de modo que, por ejemplo, porciones tonales perceptualmente importantes en un intervalo espectral alto aún pueden codificarse por el codificador principal en lugar de la sustitución paramétrica. Adicionalmente, se lleva a cabo una operación de relleno de huecos que usa recuadros de frecuencia de un primer conjunto de primeras porciones espectrales que es, por ejemplo, un conjunto de porciones tonales normalmente de un intervalo de frecuencia inferior, pero también de un

intervalo de frecuencia superior si está disponible. Sin embargo, para el ajuste de envolvente espectral en el lado de decodificador, las porciones espectrales del primer conjunto de porciones espectrales situadas en la banda de reconstrucción no se someten a procesamiento posterior adicional, por ejemplo, mediante el ajuste de envolvente espectral. Solamente los valores espectrales restantes en la banda de reconstrucción que no se original del decodificador principal deben someterse a ajuste de envolvente usando información de envolvente. De manera preferente la información de envolvente es una información de envolvente de banda completa que representa la energía del primer conjunto de primeras porciones espectrales en la banda de reconstrucción y el segundo conjunto de segundas porciones espectrales en la misma banda de reconstrucción, en la que se indica que estos últimos valores espectrales en el segundo conjunto de segundas porciones espectrales son cero y, por lo tanto, no se codifican por el codificador principal, sino que se codifican paramétricamente con información de energía de baja resolución.

Se ha encontrado que los valores de energía absolutos, ya sea normalizados con respecto al ancho de banda de la banda correspondiente o no normalizados, son útiles y muy eficientes en una aplicación en el lado de decodificador. Esto se aplica especialmente cuando tienen que calcularse factores de ganancia basándose en una energía residual en la banda de reconstrucción, la energía faltante en la banda de reconstrucción e información de recuadro de frecuencia en la banda de reconstrucción.

Adicionalmente, se prefiere que el flujo de bits codificado no sólo cubra la información de energía para las bandas de reconstrucción sino, adicionalmente, factores de escala para bandas de factor de escala que se extienden hasta la frecuencia máxima. Esto garantiza que para cada banda de reconstrucción, para la que está disponible una determinada porción tonal, es decir, una primera porción espectral, el primer conjunto de primeras porciones espectrales se puede decodificar realmente con la amplitud correcta. Adicionalmente, además del factor de escala para cada banda de reconstrucción, se genera una energía para esa banda de reconstrucción en un codificador y se transmite a un decodificador. Adicionalmente, se prefiere que las bandas de reconstrucción coincidan con las bandas de factor de escala o, en caso del agrupamiento de energía, al menos los límites de una banda de reconstrucción coincidan con los límites de bandas del factor de escala.

Un aspecto adicional se basa en el descubrimiento de que determinadas alteraciones en la calidad del audio se pueden resolver mediante la aplicación de un esquema de relleno de recuadros de frecuencia adaptativa de la señal. Para ello se lleva a cabo un análisis en el lado de codificador con el fin de encontrar el candidato de región de origen con mejor coincidencia para una determinada región objetivo. Se genera y transmite como información secundaria al decodificador una información de coincidencia que identifica para una región objetivo una determinada región de origen opcionalmente junto con algo de información adicional. El decodificador aplica entonces una operación de relleno de recuadros de frecuencia usando la información de coincidencia. Para ello, el decodificador lee la información de coincidencia a partir del flujo de datos transmitido o archivo de datos y accede a la región de origen identificada para una determinada banda de reconstrucción y, si está indicado en la información de coincidencia, lleva a cabo además algún procesamiento de estos datos de región de origen para generar datos espectrales sin procesar para la banda de reconstrucción. Entonces, este resultado de la operación de relleno de recuadros de frecuencia, es decir, los datos espectrales sin procesar para la banda de reconstrucción, se modela usando información de envolvente espectral con el fin de obtener finalmente una banda de reconstrucción que comprende también las primeras porciones espectrales tales como las porciones tonales. Sin embargo, estas porciones tonales no se generan mediante el esquema de relleno de recuadros adaptativo, sino que estas primeras porciones espectrales se emiten directamente mediante el decodificador de audio o el decodificador principal.

El esquema de selección de recuadros espectral adaptativo puede funcionar con baja granularidad. En esta implementación, una región de origen se subdivide en regiones de origen generalmente solapadas y la región objetivo o las bandas de reconstrucción se facilitan mediante regiones objetivo de frecuencia no solapadas. Entonces, las similitudes entre cada región de origen y cada región objetivo se determinan en el lado de codificador y el par de mejor coincidencia de una región de origen y la región objetivo se identifica mediante la información de coincidencia y, en el lado de decodificador, la región de origen identificada en la información de coincidencia se usa para la generación de los datos espectrales sin procesar para la banda de reconstrucción.

Con el fin de obtener una granularidad más alta, a cada región de origen se le permite cambiar con el fin de obtener un determinado retardo cuando las similitudes son máximas. Este retardo puede ser tan fino como un intervalo de frecuencia y permite una coincidencia incluso mejor entre una región de origen y la región objetivo.

Adicionalmente, además de identificar únicamente un par de mejor coincidencia, este retardo de correlación también puede transmitirse dentro de la información de coincidencia y, adicionalmente, incluso puede transmitirse un signo. Cuando se determina que el signo es negativo en el lado de codificador, entonces también se transmite un indicador de signo correspondiente dentro de la información de coincidencia y, en el lado de decodificador, los valores espectrales de la región de origen se multiplican por "-1" o, en una representación compleja, se "rotan" 180 grados.

Una implementación adicional de esta invención aplica una operación de blanqueo de recuadros. El blanqueo de un

espectro elimina la información de envolvente espectral gruesa y enfatiza la estructura fina espectral que es de interés principal para la evaluación de la similitud de recuadros. Por lo tanto, un recuadro de frecuencia por un lado y/o la señal de origen por otro lado se blanquean antes de calcular una medida de correlación cruzada. Cuando sólo se blanquea el recuadro usando un procedimiento predefinido, se transmite un indicador de blanqueo que indica al decodificador que se aplicará el mismo proceso de blanqueo predefinido al recuadro de frecuencia dentro de IGF.

En cuanto a la selección de recuadros, se prefiere usar el retardo de la correlación para desplazar espectralmente el espectro regenerado por un número entero de intervalos de transformación. Dependiendo de la transformación subyacente, el desplazamiento espectral puede requerir correcciones de adición. En caso de retardos impares, el recuadro se modula adicionalmente mediante multiplicación por una secuencia temporal alternante de $-1/1$ para compensar la representación de frecuencia invertida de una banda de cada dos dentro de la MDCT. Además, se aplica el signo del resultado de la correlación cuando se genera el recuadro de frecuencia.

Además, se prefiere usar el recorte y la estabilización de recuadros con el fin de garantizar que se evitan los artefactos creados por el cambio rápido de regiones de origen para la misma región de reconstrucción o región objetivo. Para ello se lleva a cabo un análisis de similitud entre las diferentes regiones de origen identificadas y cuando un recuadro de origen es similar a otros recuadros de origen con una similitud por encima de un umbral, entonces este recuadro de origen puede retirarse del conjunto de posibles recuadros de origen, ya que está altamente correlacionado con otros recuadros de origen. Además, como un tipo de estabilización de selección de recuadros, se prefiere mantener el orden de la trama anterior si ninguno de los recuadros de origen en la trama actual se correlaciona (mejor que un umbral determinado) con los recuadros objetivo en la trama actual.

Un aspecto adicional se basa en el descubrimiento de que se obtienen una calidad mejorada y tasa de bits reducida específicamente para señales que comprenden porciones transitorias, tal como se producen muy frecuentemente en señales de audio, al combinar la tecnología de modelado de ruido temporal (TNS) o de modelado de recuadro temporal (TTS) con la reconstrucción de alta frecuencia. El procesamiento TNS/TTS en el lado de codificador que se implementa mediante una predicción sobre la frecuencia reconstruye la envolvente de tiempo de la señal de audio. Dependiendo de la implementación, es decir, cuando el filtro de modelado de ruido temporal se determina dentro de un intervalo de frecuencia que no sólo cubre el intervalo de frecuencia de origen sino también el intervalo de frecuencia objetivo que se va a reconstruir en un decodificador de regeneración de frecuencia, la envolvente temporal no sólo se aplica a la señal de audio principal hasta una frecuencia de inicio de relleno de huecos, sino que la envolvente temporal también se aplica a los intervalos espectrales de segundas porciones espectrales reconstruidas. De esta manera, se reducen o se eliminan los ecos previos o ecos posteriores que se producirían sin modelado de recuadro temporal. Esto se logra al aplicar una predicción inversa sobre frecuencia no sólo dentro del intervalo de frecuencia principal hasta una determinada frecuencia de inicio de relleno de huecos sino también dentro de un intervalo de frecuencia por encima del intervalo de frecuencia principal. Para este fin, la regeneración de frecuencia o generación de recuadro de frecuencia se lleva a cabo en el lado de codificador antes de aplicar una predicción sobre frecuencia. Sin embargo, la predicción sobre frecuencia se puede aplicar ya sea antes o después del modelado de envolvente espectral dependiendo de si el cálculo de información de energía se ha llevado a cabo en los valores residuales espectrales tras el filtrado o a los valores espectrales (completos) antes del modelado de envolvente.

El procesamiento TTS sobre uno o más recuadros de frecuencia establece adicionalmente una continuidad de correlación entre el intervalo de origen y el intervalo de reconstrucción o en dos intervalos de reconstrucción adyacentes o recuadros de frecuencia.

En una implementación, se prefiere usar el filtrado TNS/TTS complejo. De esta manera, se evitan los artefactos de distorsión (temporales) de una representación real muestreada de manera crítica, tal como MDCT. Un filtro TNS complejo se puede calcular en el lado de codificador aplicando no sólo una transformada de coseno discreta modificada sino también además una transformada de seno discreta modificada para obtener una transformada modificada compleja. No obstante, sólo se transmiten los valores de la transformada de coseno discreta modificada, es decir, la parte real de la transformada compleja. Sin embargo, en el lado de codificador es posible estimar la parte imaginaria de la transformada usando espectros de MDCT de tramas anteriores o posteriores de modo que, en el lado de codificador, puede aplicarse de nuevo el filtro complejo en la predicción inversa sobre frecuencia y, especialmente, la predicción sobre el límite entre el intervalo de origen y el intervalo de reconstrucción y también sobre el límite entre recuadros de frecuencia adyacentes a la frecuencia dentro del intervalo de reconstrucción.

El sistema de codificación de audio de la invención codifica de manera eficiente señales de audio arbitrarias a un amplio intervalo de tasas de bits. Por consiguiente, para tasas de bits altas, el sistema de la invención converge hacia la transparencia, para tasas de bits bajas se minimiza la molestia perceptual. Por lo tanto, se usa la parte principal de tasa de bits disponible para someter a codificación de forma de onda sólo la estructura perceptualmente más relevante de la señal en el codificador, y los huecos espectrales resultantes se rellenan en el decodificador con contenido de señal que se asemeja aproximadamente al espectro original. Se consume un presupuesto de bits muy limitado para controlar el denominado relleno inteligente de huecos (IGF) impulsado por parámetros mediante

información secundaria dedicada transmitida del codificador al decodificador.

A continuación se describen realizaciones preferidas de la presente invención con referencia a los dibujos adjuntos, en los que:

- 5 la figura 1a ilustra un aparato para codificar una señal de audio;
- la figura 1b ilustra un decodificador para decodificar una señal de audio codificada que coincide con el codificador de la figura 1a;
- 10 la figura 2a ilustra una implementación preferida del decodificador;
- la figura 2b ilustra una implementación preferida del codificador;
- 15 la figura 3a ilustra una representación esquemática de un espectro generado por el decodificador de dominio espectral de la figura 1b;
- la figura 3b ilustra una tabla que indica la relación entre factores de escala para las bandas de factor de escala y energías para bandas de reconstrucción e información de relleno de ruido para una banda de relleno de ruido;
- 20 la figura 4a ilustra la funcionalidad del codificador de dominio espectral para aplicar la selección de porciones espectrales en el primer y el segundo conjunto de porciones espectrales;
- la figura 4b ilustra una implementación de la funcionalidad de la figura 4a;
- 25 la figura 5a ilustra una funcionalidad de un codificador de MDCT;
- la figura 5b ilustra una funcionalidad del decodificador con una tecnología de MDCT;
- 30 la figura 5c ilustra una implementación del regenerador de frecuencia;
- la figura 6a ilustra un codificador de audio con funcionalidad de modelado de ruido temporal/modelado de recuadro temporal;
- 35 la figura 6b ilustra un decodificador con tecnología de modelado de ruido temporal/modelado de recuadro temporal;
- la figura 6c ilustra una funcionalidad adicional de la funcionalidad de modelado de ruido temporal/modelado de recuadro temporal con un orden diferente del filtro de predicción espectral y el modelador espectral;
- 40 la figura 7a ilustra una implementación de la funcionalidad de modelado de recuadro temporal (TTS);
- la figura 7b ilustra una implementación de decodificador que coincide con la implementación de codificador de la figura 7a;
- 45 la figura 7c ilustra un espectrograma de una señal original y una señal extendida sin TTS;
- la figura 7d ilustra una representación de frecuencia que ilustra la correspondencia entre frecuencias de relleno inteligente de huecos y energías de modelado de recuadro temporal;
- 50 la figura 7e ilustra un espectrograma de una señal original y una señal extendida sin TTS;
- la figura 8a ilustra un decodificador de dos canales con regeneración de frecuencia;
- la figura 8b ilustra una tabla que ilustra diferentes combinaciones de representaciones e intervalos de origen/destino;
- 55 la figura 8c ilustra un diagrama de flujo que ilustra la funcionalidad del decodificador de dos canales con regeneración de frecuencia de la figura 8a;
- la figura 8d ilustra una implementación más detallada del decodificador de la figura 8a;
- 60 la figura 8e ilustra una implementación de un codificador para el procesamiento de dos canales que va a decodificarse por el decodificador de la figura 8a;
- la figura 9a ilustra un decodificador con tecnología de regeneración de frecuencia que usa valores de energía para el

intervalo de frecuencia de regeneración;

la figura 9b ilustra una implementación más detallada del regenerador de frecuencia de la figura 9a;

5 la figura 9c ilustra un esquema que ilustra la funcionalidad de la figura 9b;

la figura 9d ilustra una implementación adicional del decodificador de la figura 9a;

10 la figura 10a ilustra un diagrama de bloques de un codificador que coincide con el decodificador de la figura 9a;

la figura 10b ilustra un diagrama de bloques para ilustrar una funcionalidad adicional del calculador de parámetros de la figura 10a;

15 la figura 10c ilustra un diagrama de bloques para ilustrar una funcionalidad adicional del calculador paramétrico de la figura 10a;

la figura 10d ilustra un diagrama de bloques que ilustra una funcionalidad adicional del calculador paramétrico de la figura 10a;

20 la figura 11a ilustra un decodificador adicional que tiene una identificación de intervalo de origen específica para una operación de relleno de recuadros espectrales en el decodificador;

la figura 11b ilustra la funcionalidad adicional del regenerador de frecuencia de la figura 11a;

25 la figura 11c ilustra un codificador usado para cooperar con el decodificador de la figura 11a;

la figura 11d ilustra un diagrama de bloques de una implementación del calculador de parámetros de la figura 11c;

30 las figuras 12a y 12b ilustran esquemas de frecuencia para ilustrar un intervalo de origen y un intervalo objetivo;

la figura 12c ilustra un gráfico de una correlación a modo de ejemplo de dos señales;

la figura 13a ilustra un codificador de la técnica anterior con extensión de ancho de banda; y

35 la figura 13b ilustra un decodificador de la técnica anterior con extensión de ancho de banda.

La figura 1a ilustra un aparato para codificar una señal 99 de audio. La señal 99 de audio se introduce en un convertidor 100 de tiempo-espectro para convertir una señal de audio que tiene una tasa de muestreo en una representación 101 espectral emitida por el convertidor de tiempo-espectro. El espectro 101 se introduce en un analizador 102 espectral para analizar la representación 101 espectral. El analizador 101 espectral está configurado para determinar un primer conjunto de primeras porciones 103 espectrales que van a codificarse con una primera resolución espectral y un segundo conjunto diferente de segundas porciones 105 espectrales que van a codificarse con una segunda resolución espectral. La segunda resolución espectral es más baja que la primera resolución espectral. El segundo conjunto de segundas porciones 105 espectrales se introduce en un calculador de parámetros o codificador 104 paramétrico para calcular información de envolvente espectral que tiene la segunda resolución espectral. Además, se proporciona un codificador 106 de audio de dominio espectral para generar una primera representación 107 codificada del primer conjunto de primeras porciones espectrales que tiene la primera resolución espectral. Además, el calculador de parámetros/codificador 104 paramétrico está configurado para generar una segunda representación 109 codificada del segundo conjunto de segundas porciones espectrales. La primera representación 107 codificada y la segunda representación 109 codificada se introducen en un multiplexor de flujo de bits o formador 108 de flujo de bits y el bloque 108 emite finalmente la señal de audio codificada para su transmisión o almacenamiento en un dispositivo de almacenamiento.

55 Generalmente, una primera porción espectral tal como 306 de la figura 3a estará rodeada por dos segundas porciones espectrales tales como 307a, 307b. Esto no es así en HE AAC, en la que el intervalo de frecuencia del codificador principal está limitado en cuanto a la banda.

60 La figura 1b ilustra un decodificador que coincide con el codificador de la figura 1a. La primera representación 107 codificada se introduce en un decodificador 112 de audio de dominio espectral para generar una primera representación decodificada de un primer conjunto de primeras porciones espectrales, teniendo la representación decodificada una primera resolución espectral. Además, la segunda representación 109 codificada se introduce en un decodificador 114 paramétrico para generar una segunda representación decodificada de un segundo conjunto de segundas porciones espectrales que tiene una segunda resolución espectral que es más baja que la primera resolución espectral.

El decodificador comprende además un regenerador 116 de frecuencia para regenerar una segunda porción espectral reconstruida que tiene la primera resolución espectral usando una primera porción espectral. El regenerador 116 de frecuencia lleva a cabo una operación de relleno de recuadros, es decir usa un recuadro o una porción del primer conjunto de primeras porciones espectrales y copia este primer conjunto de primeras porciones espectrales en el intervalo de reconstrucción o banda de reconstrucción que tiene la segunda porción espectral y generalmente lleva a cabo el modelado de envolvente espectral u otra operación indicada por la segunda representación decodificada emitida por el decodificador 114 paramétrico, es decir usando la información sobre el segundo conjunto de segundas porciones espectrales. El primer conjunto decodificado de primeras porciones espectrales y el segundo conjunto reconstruido de porciones espectrales indicado en la salida del regenerador 116 de frecuencia en la línea 117 se introduce en un convertidor 118 de espectro-tiempo configurado para convertir la primera representación decodificada y la segunda porción espectral reconstruida en una representación 119 de tiempo, teniendo la representación de tiempo una alta tasa de muestreo determinada.

15 La figura 2b ilustra una implementación del codificador de la figura 1a. Una señal 99 de entrada de audio se introduce en un banco 220 de filtros de análisis correspondiente al convertidor 100 de tiempo-espectro de la figura 1a. A continuación se lleva a cabo una operación de modelado de ruido temporal en el bloque 222 de TNS. Por lo tanto, la entrada en el analizador 102 espectral de la figura 1a correspondiente a una máscara 226 tonal de bloque de la figura 2b o bien puede ser valores espectrales completos, cuando no se aplica la operación de modelado de ruido temporal/modelado de recuadro temporal, o bien puede ser valores residuales espectrales, cuando se aplica la operación de TNS como se ilustra en la figura 2b, bloque 222. Para señales de dos canales o señales multicanal se puede llevar a cabo además una codificación 228 de canales conjunta, de modo que el codificador 106 de dominio espectral de la figura 1a puede comprender el bloque 228 de codificación de canales conjunta. Asimismo, se proporciona un codificador 232 por entropía para llevar a cabo una compresión de datos sin pérdidas que también es una parte del codificador 106 de dominio espectral de la figura 1a.

El analizador espectral/máscara 226 tonal separa la salida del bloque 222 de TNS para dar la banda principal y las componentes tonales correspondientes al primer conjunto de primeras porciones 103 espectrales y las componentes residuales correspondientes al segundo conjunto de segundas porciones 105 espectrales de la figura 1a. El bloque 224 indicado como codificación de extracción de parámetros de IGF corresponde al codificador 104 paramétrico de la figura 1a y el multiplexor 230 de flujo de bits corresponde al multiplexor 108 de flujo de bits de la figura 1a.

Preferentemente, el banco 222 de filtros de análisis se implementa como una MDCT (banco de filtros de transformada de coseno discreta modificada) y la MDCT se usa para transformar la señal 99 en un dominio de tiempo-frecuencia actuando la transformada de coseno discreta modificada como herramienta de análisis de frecuencia.

El analizador 226 espectral aplica preferentemente una máscara de tonalidad. Esta etapa de estimación de máscara de tonalidad se usa para separar las componentes tonales de las componentes tipo ruido en la señal. Esto permite que el codificador 228 principal codifique todas las componentes tonales con un módulo psicoacústico. La etapa de estimación de máscara de tonalidad se puede implementar de muchas maneras diferentes y se implementa preferentemente de manera similar en cuanto a su funcionalidad a la etapa de estimación de pista sinusoidal usada en el modelado sinusoidal y de ruido para la codificación de voz/audio [8, 9] o un codificador de audio basado en el modelo HILN descrito en [10]. Preferentemente se usa una implementación que es fácil de implementar sin la necesidad de mantener trayectorias de nacimiento-muerte, pero también se puede usar cualquier otro detector de tonalidad o ruido.

El módulo de IGF calcula la similitud que existe entre una región de origen y una región objetivo. La región objetivo estará representada por el espectro de la región de origen. La medida de la similitud entre las regiones de origen y objetivo se realiza usando un enfoque de correlación cruzada. La región objetivo se divide en n_{Tar} recuadros de frecuencia no solapantes. Para cada recuadro en la región objetivo, se crean n_{Src} recuadros de origen a partir de una frecuencia de inicio fija. Estos recuadros de origen se solapan por un factor de entre 0 y 1, en el que 0 significa el 0% de solapamiento y 1 significa el 100% de solapamiento. Cada uno de estos recuadros de origen se correlaciona con el recuadro objetivo en diversos retardos para encontrar el recuadro de origen que coincide mejor con el recuadro objetivo. El número de recuadro de mejor coincidencia se almacena en `tileNum[idx_tar]`, el retardo en el que se correlaciona mejor con el objetivo se almacena en `xcorr_lag[idx_tar][idx_src]` y el signo de la correlación se almacena en `xcorr_sign[idx_tar][idx_src]`. En caso de que la correlación sea muy negativa, se necesita multiplicar el recuadro de origen por -1 antes del proceso de relleno de recuadros en el decodificador. El módulo de IGF también se encarga de no sobrescribir las componentes tonales en el espectro ya que las componentes tonales se conservan usando la máscara de tonalidad. Se usa un parámetro de energía por bandas para almacenar la energía de la región objetivo que permite reconstruir el espectro con precisión.

Este método tiene ciertas ventajas en comparación con la SBR clásica [1] ya que la cuadrícula de armónicos de una

señal multitono se conserva por el codificador principal mientras que sólo se rellenan los huecos entre las sinusoides con el "ruido modelado" de mejor coincidencia de la región de origen. Otra ventaja de este sistema en comparación con ASR (reemplazo espectral preciso) [2-4] es la ausencia de una etapa de síntesis de señal que crea las porciones importantes de la señal en el decodificador. En vez de esto, esta tarea se asume por el codificador principal, lo que permite la conservación de componentes importantes del espectro. Otra ventaja del sistema propuesto es la capacidad de ajuste a escala continuo que ofrecen las características. Usar sólo $tileNum[idx_tar]$ y $xcorr_lag = 0$, para cada recuadro se denomina adaptación de granularidad en bruto y se puede usar para tasas de bits bajas mientras que usar $xcorr_lag$ variable para cada recuadro permite hacer coincidir mejor los espectros objetivo y de origen.

Además, se propone una técnica de estabilización de elección de recuadros que elimina los artefactos de dominio de frecuencia tales como vibración y ruido musical.

En caso de pares de canales estéreo se aplica un procesamiento de estéreo conjunto adicional. Esto es necesario, ya que para un determinado intervalo de destino, la señal puede ser una fuente de sonido panoramizada altamente correlacionada. En caso de que las regiones de origen elegidas para esta región en particular no estén bien correlacionadas, a pesar de que las energías se hagan coincidir para las regiones de destino, la imagen espacial puede resentirse debido a las regiones de origen no correlacionadas. El codificador analiza cada banda de energía de la región de destino, llevando a cabo normalmente una correlación cruzada de los valores espectrales y si se supera un determinado umbral, establece un indicador conjunto para esta banda de energía. En el decodificador, las bandas de energía de los canales izquierdo y derecho se tratan individualmente si no se establece este indicador de estéreo conjunto. En caso de establecerse el indicador de estéreo conjunto, tanto las energías como la interconexión se llevan a cabo en el dominio de estéreo conjunto. La información de estéreo conjunto para las regiones de IGF se señala de manera similar a la información de estéreo conjunto para la codificación principal, que incluye un indicador que indica en el caso de la predicción si el sentido de la predicción es de mezcla descendente a residual o viceversa.

Las energías pueden calcularse a partir de las energías transmitidas en el dominio L/R.

$$\begin{aligned} midNrg[k] &= leftNrg[k] + rightNrg[k]; \\ sideNrg[k] &= leftNrg[k] - rightNrg[k]; \end{aligned}$$

siendo k el índice de frecuencia en el dominio de la transformada.

Otra solución consiste en calcular y transmitir las energías directamente en el dominio de estéreo conjunto para bandas en las que el estéreo conjunto está activo, por lo que no es necesaria la transformación de energía adicional en el lado de decodificador.

Los recuadros de origen se crean siempre según la matriz central/lateral:

$$\begin{aligned} midTile[k] &= 0.5 \cdot (leftTile[k] + rightTile[k]) \\ sideTile[k] &= 0.5 \cdot (leftTile[k] - rightTile[k]) \end{aligned}$$

Ajuste de energía:

$$\begin{aligned} midTile[k] &= midTile[k] * midNrg[k]; \\ sideTile[k] &= sideTile[k] * sideNrg[k]; \end{aligned}$$

Transformación de estéreo conjunto -> LR:

Si no se codifica ningún parámetro de predicción adicional:

$$\begin{aligned} leftTile[k] &= midTile[k] + sideTile[k] \\ rightTile[k] &= midTile[k] - sideTile[k] \end{aligned}$$

Si se codifica un parámetro de predicción adicional y si el sentido señalado es del centro al lado:

$$sideTile[k] = sideTile[k] - predictionCoeff \cdot midTile[k]$$

$$leftTile[k] = midTile[k] + sideTile[k]$$

$$rightTile[k] = midTile[k] - sideTile[k]$$

Si el sentido señalado es del lado al centro:

5

$$midTile1[k] = midTile[k] - predictionCoeff \cdot sideTile[k]$$

$$leftTile[k] = midTile1[k] - sideTile[k]$$

$$rightTile[k] = midTile1[k] + sideTile[k]$$

Este procesamiento garantiza que a partir de los recuadros usados para regenerar regiones de destino altamente correlacionadas y regiones de destino panoramizadas, los canales izquierdo y derecho resultantes siguen representando una fuente de sonido correlacionada y panoramizada aunque las regiones de origen no estén correlacionadas, conservando la imagen estéreo para dichas regiones.

10

Dicho de otro modo, en el flujo de bits se transmiten indicadores de estéreo conjunto que indican si se usará L/R o M/S como ejemplo para la codificación de estéreo conjunto general. En primer lugar, en el decodificador, se decodifica la señal principal como se indica por los indicadores de estéreo conjunto para las bandas principal. En segundo lugar, se almacena la señal principal en representaciones tanto L/R como M/S. Para el relleno de recuadros por IGF, se elige la representación de recuadros de origen para adaptarse a la representación de recuadros objetivo como se indica por la información de estéreo conjunto para las bandas de IGF.

15

El modelado de ruido temporal (TNS) es una técnica convencional y forma parte de AAC [11 - 13]. El TNS se puede considerar como una extensión del esquema básico de un codificador perceptual, mediante la inserción de una etapa de procesamiento opcional entre el banco de filtros y la etapa de cuantificación. La tarea principal del módulo de TNS consiste en ocultar el ruido de cuantificación producido en la región de máscara temporal de señales de tipo transitorio y, por lo tanto, conduce a un esquema de codificación más eficiente. En primer lugar, el TNS calcula un conjunto de coeficientes de predicción usando "predicción directa" en el dominio de transformada, por ejemplo, MDCT. Después se usan estos coeficientes para aplanar la envolvente temporal de la señal. Dado que la cuantificación afecta el espectro filtrado por TNS, el ruido de cuantificación también es temporalmente plano. Mediante la aplicación del filtrado de TNS inverso en el lado de decodificador, el ruido de cuantificación se modela según la envolvente temporal del filtro de TNS y, por lo tanto, el ruido de cuantificación se enmascara por el transitorio.

25

30

El IGF se basa en una representación de MDCT. Para una codificación eficiente se tienen que usar bloques preferiblemente largos de aproximadamente 20 ms. Si la señal dentro de un bloque largo de este tipo contiene transitorios, se producen ecos previos y posteriores audibles en las bandas espectrales de IGF debido al relleno de recuadros. La figura 7c muestra un efecto de eco previo típico antes de la aparición de transitorio debido a IGF. En el lado izquierdo se muestra el espectrograma de la señal original y en el lado derecho se muestra el espectrograma de la señal con ancho de banda extendido sin filtrado de TNS.

35

Este efecto de eco previo se reduce usando TNS en el contexto de IGF. En este caso, el TNS se usa como herramienta de modelado de recuadro temporal (TTS) ya que la regeneración espectral en el decodificador se lleva a cabo sobre la señal residual de TNS. Los coeficientes de predicción de TTS requeridos se calculan y aplican usando el espectro completo en el lado de codificador como es habitual. Las frecuencias de inicio y fin de TNS/ TTS no resultan afectadas por la frecuencia de inicio de IGF $f_{IGFstart}$ de la herramienta de IGF. En comparación con el TNS tradicional, la frecuencia de fin de TTS se aumenta hasta la frecuencia de fin de la herramienta de IGF, que es mayor que $f_{IGFstart}$. En el lado de decodificador se aplican nuevamente los coeficientes de TNS/TTS sobre el espectro completo, es decir el espectro principal más el espectro regenerado más las componentes tonales del mapa de tonalidad (véase la figura 7e). La aplicación de TTS es necesaria para formar la envolvente temporal del espectro regenerado para coincidir nuevamente con la envolvente de la señal original. Por lo tanto, se reducen los ecos previos mostrados. Adicionalmente, todavía modela el ruido de cuantificación en la señal por debajo $f_{IGFstart}$ como es habitual con TNS.

40

45

50

En los decodificadores tradicionales, la interconexión espectral en una señal de audio corrompe la correlación espectral en los límites de interconexión y, por lo tanto, afecta a la envolvente temporal de la señal de audio introduciendo dispersión. Por lo tanto, otro beneficio de la realización de relleno de recuadros por IGF en la señal residual es que, tras la aplicación del filtro de modelado, los límites del recuadro se correlacionan perfectamente, lo que da resultado una reproducción temporal más fiel de la señal.

55

En un codificador de la invención, el espectro que se ha sometido a filtrado de TNS/TTS, procesamiento de máscara de tonalidad y estimación de parámetros de IGF, carece de cualquier señal por encima de la frecuencia de inicio de IGF excepto las componentes tonales. Este espectro escaso se codifica ahora por el codificador principal usando principios de codificación aritmética y codificación predictiva. Estas componentes codificadas junto con los bits de señalización forman el flujo de bits del audio.

La figura 2a ilustra la implementación correspondiente del decodificador. El flujo de bits en la figura 2a correspondiente a la señal de audio codificada se introduce en el demultiplexor/decodificador que estará conectado, con respecto a la figura 1b, a los bloques 112 y 114. El demultiplexor de flujo de bits separa la señal de audio de entrada en la primera representación 107 codificada de la figura 1b y la segunda representación 109 codificada de la figura 1b. La primera representación codificada que tiene el primer conjunto de primeras porciones espectrales se introduce en el bloque 204 de decodificación de canales conjunta correspondiente al decodificador 112 de dominio espectral de la figura 1b. La segunda representación codificada se introduce en el decodificador 114 paramétrico que no se ilustra en la figura 2a y luego se introduce en el bloque 202 de IGF correspondiente al regenerador 116 de frecuencia de la figura 1b. El primer conjunto de primeras porciones espectrales requerido para la regeneración de frecuencia se introduce en el bloque 202 de IGF a través de la línea 203. Asimismo, tras la decodificación 204 de canales conjunta, se aplica la decodificación principal específica en el bloque 206 de máscara tonal de manera que la salida del máscara 206 tonal corresponde a la salida del decodificador 112 de dominio espectral. A continuación, el combinador 208 lleva a cabo una combinación, es decir una construcción de tramas en la que la salida del combinador 208 tiene ahora el espectro de intervalo completo, pero todavía en el dominio filtrado por TNS/TTS. Posteriormente, en el bloque 210 se lleva a cabo una operación de TNS/TTS inversa usando información de filtro de TNS/TTS proporcionada a través de la línea 109, es decir la información secundaria de TTS se incluye preferentemente en la primera representación codificada generada por el codificador 106 de dominio espectral que puede ser, por ejemplo, un codificador principal de AAC directa o USAC, o también puede incluirse en la segunda representación codificada. En la salida del bloque 210 se proporciona un espectro completo hasta la frecuencia máxima que es la frecuencia de intervalo completo definida por la tasa de muestreo de la señal de entrada original. A continuación se lleva a cabo una conversión de espectro/tiempo en el banco 212 de filtros de síntesis para obtener finalmente la señal de salida de audio.

La figura 3a ilustra una representación esquemática del espectro. El espectro se subdivide en bandas de factor de escala SCB en las que hay siete bandas de factor de escala SCB1 a SCB7 en el ejemplo ilustrado de la figura 3a. Las bandas de factor de escala pueden ser bandas de factor de escala de AAC que están definidas en la norma AAC y tienen un ancho de banda creciente hasta frecuencias superiores como se ilustra esquemáticamente en la figura 3a. Se prefiere llevar a cabo el relleno inteligente de huecos no desde el comienzo del espectro, es decir a bajas frecuencias, sino iniciar la operación de IGF a una frecuencia de inicio de IGF ilustrada en 309. Por lo tanto, la banda de frecuencia principal se extiende desde la frecuencia más baja hasta la frecuencia de inicio de IGF. Por encima de la frecuencia de inicio de IGF se aplica el análisis de espectro para separar las componentes 304, 305, 306, 307 espectrales de alta resolución (el primer conjunto de primeras porciones espectrales) de componentes de baja resolución representadas por el segundo conjunto de segundas porciones espectrales. La figura 3a ilustra un espectro que se introduce a modo de ejemplo en el codificador 106 de dominio espectral o en el codificador 228 conjunto de canales, es decir, el codificador principal funciona en todo el intervalo pero codifica una cantidad significativa de valores espectrales cero, es decir estos valores espectrales cero se cuantifican a cero o se establecen a cero antes de la cuantificación o tras la cuantificación. De todos modos, el codificador principal funciona en el intervalo completo, es decir, como si el espectro fuera tal como se ilustra, es decir, el decodificador principal no tiene que estar necesariamente al tanto de ningún relleno inteligente de huecos del segundo conjunto de segundas porciones espectrales con una resolución espectral inferior.

Preferentemente, la alta resolución se define mediante una codificación por líneas de líneas espectrales, tales como líneas de MDCT, mientras que la segunda resolución o baja resolución se define, por ejemplo, calculando sólo un único valor espectral por banda de factor de escala, en el que una banda de factor de escala abarca varias líneas de frecuencia. Por lo tanto, la segunda resolución baja, con respecto a su resolución espectral, es mucho menor que la primera o alta resolución definida por la codificación por líneas generalmente aplicada por el codificador principal tal como un codificador principal de AAC o USAC.

En cuanto al cálculo de energía o factor de escala, la situación se ilustra en la figura 3b. Debido al hecho de que el codificador es un codificador principal y debido al hecho de que puede haber, pero no tiene que haber necesariamente, componentes del primer conjunto de porciones espectrales en cada banda, el codificador principal calcula un factor de escala para cada banda no sólo en el intervalo principal por debajo de la frecuencia 309 de IGF, sino también por encima de la frecuencia de inicio de IGF hasta la frecuencia máxima $f_{IGFstop}$ que es menor o igual a la mitad de la frecuencia de muestreo, es decir, $f_s/2$. Por lo tanto, las porciones 302, 304, 305, 306, 307 tonales codificadas de la figura 3a y, en esta realización junto con los factores de escala SCB1 a SCB7, corresponden a los datos espectrales de alta resolución. Los datos espectrales de baja resolución se calculan a partir de la frecuencia de inicio de IGF y corresponden a los valores de información de energía E_1, E_2, E_3, E_4 , que se transmiten junto con

los factores de escala SF4 a SF7.

En particular, cuando el codificador principal se encuentra en una condición de tasa de bits baja se puede aplicar además una operación de relleno de ruido adicional en la banda principal, es decir una frecuencia inferior a la frecuencia de inicio de IGF, es decir, en las bandas de factor de escala SCB1 a SCB3. En el relleno de ruido, existen varias líneas espectrales adyacentes que se han cuantificado a cero. En el lado de decodificador, estos valores espectrales cuantificados a cero se re-sintetizan y la magnitud de los valores espectrales re-sintetizados se ajusta usando una energía de relleno de ruido tal como NF_2 ilustrada en 308 en la figura 3b. La energía de relleno de ruido, que se puede dar en términos absolutos o en términos relativos particularmente con respecto al factor de escala como en USAC, corresponde a la energía del conjunto de valores espectrales cuantificados a cero. También puede considerarse que estas líneas espectrales de relleno de ruido son un tercer conjunto de terceras porciones espectrales que se regeneran mediante síntesis de relleno de ruido directa sin ninguna operación de IGF basándose en la regeneración de frecuencia usando recuadros de frecuencia de otras frecuencias para la reconstrucción de recuadros de frecuencia usando valores espectrales de un intervalo de origen y la información de energía E_1 , E_2 , E_3 , E_4 .

Preferentemente, las bandas para las cuales se calcula la información de energía coinciden con las bandas de factor de escala. En otras realizaciones se aplica un agrupamiento de valores de información de energía de manera que, por ejemplo, para las bandas de factor de escala 4 y 5 sólo se transmite un único valor de información de energía, pero incluso en esta realización, los límites de las bandas de reconstrucción agrupadas coinciden con los límites de las bandas de factor de escala. Si se aplican diferentes separaciones de bandas, entonces se pueden aplicar determinados nuevos cálculos o cálculos de sincronización, y esto puede tener sentido en función de la implementación determinada.

Preferentemente, el codificador 106 de dominio espectral de la figura 1a es un codificador impulsado psicoacústicamente como se ilustra en la figura 4a. Generalmente, como se ilustra por ejemplo en la norma MPEG2/4 AAC o la norma MPEG1/2, capa 3, la señal de audio que va a codificarse después de haberse transformado en el intervalo espectral (401 en la figura 4a) se retransmite a un calculador 400 de factor de escala. El calculador de factor de escala se controla mediante un modelo psicoacústico que recibe adicionalmente la señal de audio que va a cuantificarse o recibe, como en la norma MPEG1/2 capa 3 o MPEG AAC, una representación espectral compleja de la señal de audio. El modelo psicoacústico calcula, para cada banda de factor de escala, un factor de escala que representa el umbral psicoacústico. Adicionalmente, luego se ajustan los factores de escala, por la cooperación de los bucles de iteración interno y externo bien conocidos o por cualquier otro procedimiento de codificación adecuado, de manera que se cumplen determinadas condiciones de tasas de bits. A continuación, los valores espectrales que van a cuantificarse por un lado, y los factores de escala calculados por otro lado, se introducen en un procesador 404 cuantificador. En el funcionamiento del codificador de audio directo, los valores espectrales que van a cuantificarse se ponderan mediante los factores de escala y luego se introducen los valores espectrales ponderados en un cuantificador fijo que tiene generalmente una funcionalidad de compresión hasta intervalos de amplitud superiores. Entonces, en la salida del procesador cuantificador sí existen índices de cuantificación que luego se retransmiten a un codificador por entropía que tiene generalmente codificación específica y muy eficiente para un conjunto de índices de cuantificación de cero para valores de frecuencia adyacentes o, como también se denomina en la técnica, una "serie" de valores de cero.

Sin embargo, en el codificador de audio de la figura 1a el procesador cuantificador recibe generalmente información sobre las segundas porciones espectrales del analizador espectral. Por lo tanto, el procesador 404 cuantificador se asegura de que, en la salida del procesador 404 cuantificador, las segundas porciones espectrales identificadas por el analizador 102 espectral son cero o tienen una representación reconocida por un codificador o un decodificador como una representación cero que puede codificarse de manera muy eficiente, específicamente cuando existen "series" de valores de cero en el espectro.

La figura 4b ilustra una implementación del procesador cuantificador. Los valores espectrales de MDCT se pueden introducir en un bloque 410 de establecimiento a cero. Posteriormente, las segundas porciones espectrales ya se establecen a cero antes de llevar a cabo una ponderación por los factores de escala en el bloque 412. En una implementación adicional, no se proporciona el bloque 410, pero la cooperación de establecimiento a cero se lleva a cabo en el bloque 418 posterior al bloque 412 de ponderación. En aún una implementación adicional, la operación de establecimiento a cero también se puede llevar a cabo en un bloque 422 de establecimiento a cero posterior a una cuantificación en el bloque 420 cuantificador. En esta implementación, los bloques 410 y 418 no estarán presentes. En general se proporciona al menos uno de los bloques 410, 418, 422 dependiendo de la implementación específica.

Entonces, en la salida del bloque 422 se obtiene un espectro cuantificado correspondiente a lo que se ilustra en la figura 3a. Este espectro cuantificado se introduce entonces en un codificador por entropía tal como 232 en la figura 2b que puede ser un codificador Huffman o un codificador aritmético como se define, por ejemplo, en la norma de USAC.

Los bloques 410, 418, 422 de establecimiento a cero, que se proporcionan alternativamente unos con respecto a otros o en paralelo, se controlan por el analizador 424 espectral. El analizador espectral comprende preferentemente cualquier implementación de un detector de tonalidad bien conocido o comprende cualquier tipo diferente de detector operativo para separar un espectro en componentes que van a codificarse con una alta resolución y componentes que van a codificarse con una baja resolución. Otros algoritmos de este tipo implementados en el analizador espectral pueden ser un detector de actividad de voz, un detector de ruido, un detector de habla o cualquier otro detector que decide, en función de información espectral o metadatos asociados, los requisitos de resolución para diferentes porciones espectrales.

La figura 5a ilustra una implementación preferida del convertidor 100 de tiempo-espectro de la figura 1a como se implementa, por ejemplo, en AAC o USAC. El convertidor 100 de tiempo-espectro comprende un divisor 502 de ventanas controlado por un detector 504 de transitorios. Cuando el detector 504 de transitorios detecta un transitorio, entonces indica un cambio de ventanas largas a ventanas cortas al divisor de ventanas. A continuación, el divisor 502 de ventanas calcula, para bloques solapados, tramas divididas en ventanas, en las que cada trama dividida en ventanas tiene normalmente dos N valores como, por ejemplo, 2048 valores. Luego se lleva a cabo una transformación dentro de un transformador 4 de bloques, y generalmente este transformador de bloques proporciona además una disminución de manera que se realiza una disminución/transformada combinadas para obtener una trama espectral con N valores tales como valores espectrales de MDCT. Por lo tanto, para una operación de ventanas largas, la trama en la entrada del bloque 506 comprende dos N valores como 2048 valores y una trama espectral tiene entonces 1024 valores. Sin embargo, a continuación, se lleva a cabo un cambio a bloques cortos, es decir cuando se llevan a cabo ocho bloques cortos en los que cada bloque corto tiene 1/8 valores de dominio de tiempo divididos en ventanas en comparación con una ventana larga y cada bloque espectral tiene 1/8 valores espectrales en comparación con un bloque largo. Por lo tanto, cuando se combina esta disminución con una operación de solapamiento del 50% del divisor de ventanas, el espectro es una versión muestreada críticamente de la señal 99 de audio de dominio de tiempo.

A continuación, se hace referencia a la figura 5b que ilustra una implementación específica del regenerador 116 de frecuencia y el convertidor 118 de espectro-tiempo de la figura 1b, o del funcionamiento combinado de los bloques 208, 212 de la figura 2a. En la figura 5b se considera una banda de reconstrucción específica tal como la banda de factor de escala 6 de la figura 3a. La primera porción espectral en esta banda de reconstrucción, es decir, la primera porción 306 espectral de la figura 3a se introduce en el bloque 510 constructor/ajustador de tramas. Asimismo, una segunda porción espectral reconstruida para la banda de factor de escala 6 se introduce también en el constructor/ajustador 510 de tramas. Además, también se introduce información de energía tal como E_3 de la figura 3b para una banda de factor de escala 6 en el bloque 510. La segunda porción espectral reconstruida en la banda de reconstrucción ya se ha generado por el relleno de recuadros de frecuencia usando un intervalo de origen y la banda de reconstrucción corresponde entonces al intervalo objetivo. Ahora se lleva a cabo un ajuste de energía de la trama para después obtener finalmente la trama reconstruida completa que tiene los N valores como se obtiene, por ejemplo, en la salida del combinador 208 de la figura 2a. Luego, en el bloque 512 se lleva a cabo una transformada/interpolación inversa de bloque para obtener 248 valores de dominio de tiempo para los 124 valores espectrales, por ejemplo, en la entrada del bloque 512. A continuación, en el bloque 514 se lleva a cabo una operación de síntesis de división de ventanas que nuevamente se controla por una indicación de ventana larga/ventana corta transmitida como información secundaria en la señal de audio codificada. Luego, en el bloque 516 se lleva a cabo una operación de solapamiento/adición con una trama de tiempo anterior. Preferentemente, la MDCT aplica un solapamiento del 50% de manera que, para cada nueva trama de tiempo de 2N valores, se emiten finalmente N valores de dominio de tiempo. Se prefiere en gran medida un solapamiento del 50% debido al hecho de que proporciona un muestreo crítico y un cruce continuo de una trama a la siguiente trama debido a la operación de solapamiento/adición del bloque 516.

Tal como se ilustra en 301 en la figura 3a se puede aplicar adicionalmente una operación de relleno de ruido, no sólo por debajo de la frecuencia de inicio de IGF sino también por encima de la frecuencia de inicio de IGF, tal como para la banda de reconstrucción contemplada que coincide con la banda de factor de escala 6 de la figura 3a. Entonces, también pueden introducirse valores espectrales de relleno de ruido en el constructor/ajustador 510 de tramas y el ajuste de los valores espectrales de relleno de ruido también puede aplicarse dentro de este bloque o los valores espectrales de relleno de ruido ya se pueden ajustar usando la energía de relleno de ruido antes de introducirse en el constructor/ajustador 510 de tramas.

Preferentemente, una operación de IGF, es decir una operación de relleno de recuadros de frecuencia que usa valores espectrales de otras porciones, se puede aplicar en el espectro completo. Por lo tanto, una operación de relleno de recuadros espectrales no sólo se puede aplicar en la banda alta por encima de una frecuencia de inicio de IGF sino que también se puede aplicar en la banda baja. Asimismo, el relleno de ruido sin relleno de recuadros de frecuencia también se puede aplicar no sólo por debajo de la frecuencia de inicio de IGF sino también por encima de la frecuencia de inicio de IGF. Sin embargo, se ha encontrado que puede obtenerse codificación de audio de alta calidad y alta eficiencia cuando la operación de relleno de ruido se limita al intervalo de frecuencia por debajo de la

frecuencia de inicio de IGF y cuando la operación de relleno de recuadros de frecuencia se limita al intervalo de frecuencia por encima de la frecuencia de inicio de IGF, como se ilustra en la figura 3a.

5 Preferentemente, los recuadros objetivo (TT) (que tienen frecuencias superiores a la frecuencia de inicio de IGF) están sujetos a los límites de la banda de factor de escala del codificador de tasa completa. Los recuadros de origen (ST), de los cuales se obtiene información, es decir para frecuencias inferiores a la frecuencia de inicio de IGF, no están sujetos a los límites de la banda de factor de escala. El tamaño de los ST debe corresponder al tamaño del TT asociado. Esto se ilustra usando el siguiente ejemplo. TT[0] tiene una longitud de 10 intervalos de MDCT. Esto
10 corresponde exactamente a la longitud de dos SBC posteriores (tal como 4 + 6). Entonces, todos los ST posibles que deben correlacionarse con TT[0] también tienen una longitud de 10 intervalos. Un segundo recuadro objetivo TT[1] que es adyacente a TT[0] tiene una longitud de 15 intervalos (1 SBC que tienen una longitud de 7 + 8). Entonces, el ST para lo anterior tiene una longitud de 15 intervalos en lugar de 10 intervalos como para TT[0].

15 En caso de que no se pueda encontrar un TT para un ST con la longitud del recuadro objetivo (por ejemplo, cuando la longitud del TT es mayor que el intervalo de origen disponible), entonces no se calcula una correlación y el intervalo de origen se copia varias veces en este TT (la copia se lleva a cabo una después de otra de manera que una línea de frecuencia para la frecuencia más baja de la segunda copia sigue inmediatamente, en la frecuencia, a la línea de frecuencia para la frecuencia más alta de la primera copia), hasta que se rellena completamente el recuadro objetivo TT.

20 A continuación, se hace referencia a la figura 5c que ilustra una realización preferida adicional del regenerador 116 de frecuencia de la figura 1b o el bloque 202 de IGF de la figura 2a. El bloque 522 es un generador de recuadros de frecuencia que no sólo recibe una ID de banda objetivo sino que además recibe una ID de banda de origen. A modo de ejemplo, se ha determinado en el lado de codificador que la banda de factor de escala 3 de la figura 3a es muy adecuada para la reconstrucción de la banda de factor de escala 7. Por lo tanto, la ID de banda de origen será 2 y la
25 ID de banda objetivo será 7. Basándose en esta información, el generador 522 de recuadros de frecuencia aplica una operación de copiado o de relleno de recuadros de armónicos o cualquier otra operación de relleno de recuadros para generar la segunda porción sin procesar de componentes 523 espectrales. La segunda porción sin procesar de componentes espectrales tiene una resolución de frecuencia idéntica a la resolución de frecuencia incluida en el primer conjunto de primeras porciones espectrales.

Entonces, la primera porción espectral de la banda de reconstrucción tal como 307 de la figura 3a se introduce en un constructor 524 de tramas y la segunda 523 porción sin procesar también se introduce en el constructor 524 de tramas. Luego, la trama reconstruida se ajusta mediante el ajustador 526 usando un factor de ganancia para la
35 banda de reconstrucción calculada por el calculador 528 de factor de ganancia. Sin embargo, de manera importante, la primera porción espectral en la trama no se ve influida por el ajustador 526, sino que sólo la segunda porción sin procesar para la trama de reconstrucción se ve influida por el ajustador 526. Para ello, el calculador 528 de factor de ganancia analiza la banda de origen o la segunda 523 porción sin procesar y además analiza la primera porción espectral en la banda de reconstrucción para encontrar finalmente el factor 527 de ganancia correcto de manera que la energía de la trama ajustada emitida por el ajustador 526 tiene la energía E_4 cuando se contempla una banda de factor de escala 7.

40 En este contexto, es muy importante evaluar la precisión de la reconstrucción de alta frecuencia de la presente invención en comparación con HE-AAC. Esto se explica con respecto a la banda de factor de escala 7 en la figura 3a. Se supone que un codificador de la técnica anterior como se ilustra en la figura 13a detectará la porción 307 espectral que va a codificarse con una alta resolución como un "armónico faltante". Entonces, la energía de esta componente espectral se transmitirá junto con una información de envolvente espectral para la banda de reconstrucción tal como la banda de factor de escala 7 al decodificador. A continuación, el decodificador recreará el armónico faltante. Sin embargo, el valor espectral, al que se reconstruirá el armónico faltante 307 por el
50 decodificador de la técnica anterior de la figura 13b, estará en el medio de la banda 7 a una frecuencia indicada por la frecuencia 390 de reconstrucción. Por lo tanto, la presente invención evita un error 391 de frecuencia que se introducirá por el decodificador de la técnica anterior de la figura 13d.

55 En una implementación, el analizador espectral también se implementa para el cálculo de similitudes entre primeras porciones espectrales y segundas porciones espectrales y para determinar, basándose en las similitudes calculadas, para una segunda porción espectral en un intervalo de reconstrucción una primera porción espectral que coincide lo más posible con la segunda porción espectral. Entonces, en esta implementación de intervalo de origen/intervalo de destino variable, el codificador paramétrico introducirá además en la segunda representación codificada una información de coincidencia que indica un intervalo de origen coincidente para cada intervalo de destino. En el lado
60 de decodificador, esta información se usará después por un generador 522 de recuadros de frecuencia de la figura 5c que ilustra una generación de una segunda porción 523 sin procesar basándose en una ID de banda de origen y una ID de banda objetivo.

Asimismo, tal como se ilustra en la figura 3a, el analizador espectral está configurado para analizar la representación

espectral hasta una frecuencia de análisis máxima que sólo está una pequeña cantidad por debajo de la mitad de la frecuencia de muestreo y que es preferentemente al menos un cuarto de la frecuencia de muestreo o generalmente superior.

- 5 Tal como se ilustra, el codificador funciona sin muestreo descendente y el decodificador funciona sin muestreo ascendente. Dicho de otro modo, el codificador de audio de dominio espectral está configurado para generar una representación espectral que tiene una frecuencia de Nyquist definida por la tasa de muestreo de la señal de audio introducida originalmente.
- 10 Asimismo, tal como se ilustra en la figura 3a, el analizador espectral está configurado para analizar la representación espectral comenzando con una frecuencia de relleno de huecos y terminando con una frecuencia máxima representada por una frecuencia máxima incluida en la representación espectral, en el que una porción espectral que se extiende desde una frecuencia mínima hasta la frecuencia de inicio de relleno de huecos pertenece al primer conjunto de porciones espectrales y en el que una porción espectral adicional tal como 304, 305, 306, 307 que tiene valores de frecuencia por encima de la frecuencia de relleno de huecos se incluye adicionalmente en el primer conjunto de primeras porciones espectrales.

20 Como se explicó resumidamente, el decodificador 112 de audio de dominio espectral está configurado de manera que una frecuencia máxima representada por un valor espectral en la primera representación decodificada es igual a una frecuencia máxima incluida en la representación de tiempo que tiene la tasa de muestreo, en el que el valor espectral para la frecuencia máxima en el primer conjunto de primeras porciones espectrales es cero o diferente de cero. De todos modos, para esta frecuencia máxima en el primer conjunto de componentes espectrales existe un factor de escala para la banda de factor de escala, que se genera y se transmite independientemente de si todos los valores espectrales en esta banda de factor de escala se establecen a cero o no, como se comentó en el contexto de las figuras 3a y 3b.

30 Por lo tanto, la invención es ventajosa ya que, con respecto a otras técnicas paramétricas para aumentar la eficiencia de compresión, por ejemplo, la sustitución de ruido y el relleno de ruido (estas técnicas son exclusivamente para la representación eficiente de contenido de señal local de tipo ruido), la invención permite una reproducción de frecuencia precisa de componentes tonales. Hasta la fecha, ninguna técnica del estado de la técnica aborda la representación paramétrica eficiente de contenido de señal arbitrario mediante relleno de huecos espectrales sin la restricción de una división fijada a priori en la banda baja (LF) y en la banda alta (HF).

35 Las realizaciones del sistema de la invención mejoran los enfoques del estado de la técnica y, por lo tanto, proporcionan una alta eficiencia de compresión, ninguna o sólo una pequeña molestia perceptual y un ancho de banda de audio completo incluso para tasas de bits bajas.

El sistema general consiste en:

- 40
- codificación principal de banda completa
 - relleno inteligente de huecos (relleno de recuadros o relleno de ruido)
 - partes tonales escasas en núcleo, seleccionadas por máscara tonal
- 45
- codificación conjunta de par de estéreo para la banda completa, incluyendo el relleno de recuadros
 - TNS en el recuadro
- 50
- blanqueo espectral en el intervalo de IGF

55 Un primer paso hacia un sistema más eficiente es eliminar la necesidad de transformar datos espectrales en un segundo dominio de transformada diferente de aquel del codificador principal. Como la mayoría de los códecs de audio, tales como, por ejemplo, AAC, usan MDCT como transformada básica, también es útil llevar a cabo la BWE en el dominio de MDCT. Un segundo requisito para el sistema de BWE será la necesidad de conservar la cuadrícula tonal mediante lo cual se conservan componentes tonales incluso de HF y, por lo tanto, la calidad del audio codificado es superior a los sistemas existentes. Para tener en cuenta ambos requisitos mencionados anteriormente para un esquema de BWE se propone un nuevo sistema denominado relleno inteligente de huecos (IGF). La figura 2b muestra el diagrama de bloques del sistema propuesto en el lado de codificador y la figura 2a muestra el sistema en el lado de decodificador.

60

La figura 6a ilustra un aparato para decodificar una señal de audio codificada en otra implementación de la presente invención. El aparato para decodificar comprende un decodificador 602 de audio de dominio espectral para generar una primera representación decodificada de un primer conjunto de porciones espectrales y como el regenerador 604

de frecuencia conectado aguas abajo del decodificador 602 de audio de dominio espectral para generar una segunda porción espectral reconstruida usando una primera porción espectral del primer conjunto de primeras porciones espectrales. Como se ilustra en 603, los valores espectrales en la primera porción espectral y en la segunda porción espectral son valores residuales de predicción espectral. Con el fin de transformar estos valores residuales de predicción espectral en una representación espectral completa se proporciona un filtro 606 de predicción espectral. Este filtro de predicción inversa está configurado para llevar a cabo una predicción inversa sobre frecuencia usando los valores residuales espectrales para el primer conjunto de la primera frecuencia y las segundas porciones espectrales reconstruidas. El filtro 606 de predicción inversa espectral se configura mediante información de filtro incluida en la señal de audio codificada. La figura 6b ilustra una implementación más detallada de la realización de la figura 6a. Los valores 603 residuales de predicción espectral se introducen en un generador 612 de recuadros de frecuencia que genera valores espectrales sin procesar para una banda de reconstrucción o para una determinada segunda porción de frecuencia y estos datos sin procesar que ahora tienen la misma resolución que la primera representación espectral de alta resolución se introducen en el modelador 614 espectral. El modelador espectral modela ahora el espectro usando información de envolvente transmitida en el flujo de bits y los datos modelados espectralmente se aplican entonces al filtro 616 de predicción espectral generando finalmente una trama de valores espectrales completos usando la información 607 de filtro transmitida del codificador al decodificador a través del flujo de bits.

En la figura 6b se supone que, en el lado de codificador, el cálculo de la información de filtro transmitida a través del flujo de bits y usada a través de la línea 607 se lleva a cabo con posterioridad al cálculo de la información de envolvente. Por lo tanto, dicho de otro modo, un codificador que coincide con el decodificador de la figura 6b calculará en primer lugar los valores residuales espectrales y luego calculará la información de envolvente con los valores residuales espectrales tal como se ilustra, por ejemplo, en la figura 7a. Sin embargo, la otra implementación también es útil para determinadas implementaciones, en las que la información de envolvente se calcula antes de llevar a cabo el filtrado de TNS o TTS en el lado de codificador. A continuación, se aplica el filtro 622 de predicción espectral antes de llevar a cabo el modelado espectral en el bloque 624. Por lo tanto, dicho de otro modo, los valores espectrales (completos) se generan antes de aplicar la operación 624 de modelado espectral.

Preferentemente se calcula un filtro de TNS o un filtro de TTS de valor complejo. Esto se ilustra en la figura 7a. La señal de audio original se introduce en un bloque 702 de MDCT compleja. Luego se lleva a cabo el cálculo del filtro de TTS y el filtrado de TTS en el dominio complejo. A continuación, en el bloque 706, se calcula la información secundaria de IGF y también se calcula cualquier otra operación tal como el análisis espectral para la codificación, etc. Posteriormente, el primer conjunto de primera porción espectral generado por el bloque 706 se codifica con un codificador impulsado por modelo psicoacústico que se ilustra en 708 para obtener el primer conjunto de primeras porciones espectrales indicado en $X(k)$ en la figura 7a y todos estos datos se retransmiten al multiplexor 710 de flujo de bits.

En el lado de decodificador, los datos codificados se introducen en un demultiplexor 720 para separar información secundaria de IGF por un lado, información secundaria de TTS por otro lado y la representación codificada del primer conjunto de primeras porciones espectrales.

Entonces, se usa el bloque 724 para calcular un espectro complejo a partir de uno o más espectros de valor real. A continuación, tanto los espectros de valor real como los complejos se introducen en el bloque 726 para generar valores de frecuencia reconstruidos en el segundo conjunto de segundas porciones espectrales para una banda de reconstrucción. Entonces, en la trama de banda completa obtenida completamente y sometida a relleno de recuadros se lleva a cabo la operación 728 de TTS inversa y, en el lado de decodificador, se lleva a cabo una operación de MDCT compleja inversa final en el bloque 730. Por lo tanto, el uso de la información de filtro de TNS complejo permite generar automáticamente, cuando se aplica no sólo dentro de la banda principal o dentro de las bandas de recuadros por separado sino que también se aplica sobre los límites núcleo/recuadro o sobre los límites recuadro/recuadro, un procesamiento de límites de recuadros que, en última instancia, vuelve a introducir una correlación espectral entre los recuadros. Esta correlación espectral sobre los límites de los recuadros no se obtiene generando solamente recuadros de frecuencia y llevando a cabo un ajuste de la envolvente espectral en estos datos sin procesar de recuadros de frecuencia.

La figura 7c ilustra una comparación de una señal original (panel izquierdo) y una señal extendida sin TTS. Se puede observar que hay fuertes artefactos ilustrados por las porciones ensanchadas en el intervalo de frecuencia superior ilustrado en 750. Sin embargo, esto no se produce en la figura 7e cuando se compara la misma porción espectral en 750 con la componente 750 relacionada con artefactos de la figura 7c.

Las realizaciones o el sistema de codificación de audio de la invención usan la parte principal de tasa de bits disponible para someter a codificación de forma de onda sólo la estructura perceptualmente más relevante de la señal en el codificador, y los huecos espectrales resultantes se rellenan en el decodificador con contenido de señal que se aproxima en líneas generales al espectro original. Se consume un presupuesto de bits muy limitado para controlar el denominado relleno inteligente de huecos (IGF) espectral impulsado por parámetros mediante

información secundaria dedicada transmitida del codificador al decodificador.

El almacenamiento o la transmisión de señales de audio a menudo están sujetos a estrictas restricciones de tasa de bits. En el pasado, los codificadores estaban forzados a reducir drásticamente el ancho de banda de audio transmitida cuando sólo estaba disponible una tasa de bits muy baja. Los códecs de audio modernos ahora son capaces de codificar señales de banda ancha usando métodos de extensión de ancho de banda (BWE) tales como replicación de ancho de banda espectral (SBR) [1]. Estos algoritmos se basan en una representación paramétrica del contenido de alta frecuencia (HF), que se genera a partir de la parte de baja frecuencia (LF) sometida a codificación de forma de onda de la señal decodificada por medio de transposición a la región espectral de HF (“interconexión”) y la aplicación de un procesamiento posterior impulsado por parámetros. En los esquemas de BWE, la reconstrucción de la región espectral de HF por encima de una denominada frecuencia de cruce dada se basa a menudo en la interconexión espectral. En general, la región de HF está compuesta por múltiples interconexiones adyacentes y cada una de estas interconexiones se obtiene a partir de regiones de paso de banda (BP) del espectro de LF por debajo de la frecuencia de cruce dada. Los sistemas del estado de la técnica desempeñan con eficiencia la interconexión dentro de una representación de bancos de filtros copiando un conjunto de coeficientes de subbandas adyacentes de una región de origen a la región objetivo.

Si se implementa un sistema de BWE en un banco de filtros o dominio de la transformada de tiempo-frecuencia, sólo hay una posibilidad limitada de controlar la forma temporal de la señal de extensión de ancho de banda. Generalmente, la granularidad temporal está limitada por el tamaño de salto usado entre ventanas de transformada adyacentes. Esto puede conducir a ecos previos o posteriores no deseados en el intervalo espectral de BWE.

A partir de la codificación de audio perceptual se sabe que la forma de la envolvente temporal de una señal de audio se puede restaurar usando técnicas de filtrado espectral como el modelado de envolvente temporal (TNS) [14]. Sin embargo, el filtro de TNS conocido a partir del estado de la técnica es un filtro de valor real en espectros de valor real. Un filtro de valor real en espectros de valor real de este tipo puede verse gravemente afectado por artefactos de distorsión, especialmente si la transformada real subyacente es una transformada de coseno discreta modificada (MDCT).

El modelado de recuadro de envolvente temporal aplica filtrado complejo en espectros de valor complejo, tales como los obtenidos, por ejemplo, a partir de una transformada de coseno discreta modificada compleja (CMDCT). De esta manera se evitan artefactos de distorsión.

El modelado de recuadro temporal consiste en

- estimación de coeficientes de filtro complejo y aplicación de un filtro de aplanamiento en el espectro de señal original en el codificador
- transmisión de los coeficientes de filtro en la información secundaria
- aplicación de un filtro de modelado en el espectro reconstruido con relleno de recuadros en el decodificador

La invención amplía la técnica del estado de la técnica conocida a partir de la codificación por transformada de audio, específicamente el modelado de ruido temporal (TNS) por predicción lineal a lo largo del sentido de la frecuencia, para el uso de una manera modificada en el contexto de extensión de ancho de banda.

Asimismo, el algoritmo de extensión de ancho de banda de la invención se basa en un relleno inteligente de huecos (IGF), pero emplea una transformada de valor complejo (CMDCT) sobremuestreada, en contraposición a la configuración convencional de IGF que se basa en una representación de MDCT críticamente muestreada de valor real de una señal. La CMDCT puede considerarse como la combinación de los coeficientes de MDCT en la parte real y los coeficientes de MDST en la parte imaginaria de cada coeficiente espectral de valor complejo.

Aunque el nuevo enfoque se describe en el contexto de IGF, el procesamiento de la invención puede usarse en combinación con cualquier método de BWE que se basa en una representación de bancos de filtros de la señal de audio.

En este contexto novedoso, la predicción lineal a lo largo del sentido de la frecuencia no se usa como modelado de ruido temporal, sino más bien como una técnica de modelado de recuadro temporal (TTS). El cambio de nombre se justifica por el hecho de que las componentes de señal sometida a relleno de recuadros se modelan temporalmente por TTS en contraposición al modelado de ruido de cuantificación por TNS en códecs de transformada perceptual del estado de la técnica.

La figura 7a muestra un diagrama de bloques de un codificador de BWE que usa IGF y el nuevo enfoque de TTS.

Por lo tanto, el esquema básico de codificación funciona de la siguiente manera:

- calcular la CMDCT de una señal de dominio de tiempo $x(n)$ para obtener la señal de dominio de frecuencia $X(k)$
- 5 - calcular el filtro de TTS de valor complejo
- obtener la información secundaria para la BWE y eliminar la información espectral que tiene que replicarse por el decodificador
- 10 - aplicar la cuantificación usando el módulo psicoacústico (PAM)
- almacenar/transmitir los datos, sólo se transmiten los coeficientes de MDCT de valor real

La figura 7b muestra el decodificador correspondiente. Este invierte principalmente las etapas realizadas en el codificador.

En este caso, el esquema básico de decodificación funciona de la siguiente manera:

- 20 - estimar los coeficientes de MDST partir de los valores de MDCT (este procesamiento añade un retardo de decodificador de un bloque) y combinar los coeficientes de MDCT y de MDST para dar coeficientes de CMDCT de valor complejo
- llevar a cabo el relleno de recuadros con su procesamiento posterior
- 25 - aplicar el filtrado de TTS inverso con los coeficientes de filtro de TTS transmitidos
- calcular la CMDCT inversa

Obsérvese que, alternativamente, el orden de la síntesis de TTS y el procesamiento posterior de IGF también se puede invertir en el decodificador si el análisis de TTS y la estimación de parámetros de IGF se invierten sistemáticamente en el codificador.

Para una codificación de transformada eficiente se deben usar preferentemente los denominados “bloques largos” de aproximadamente 20 ms para lograr una ganancia de transformada razonable. Si la señal dentro de un bloque largo de este tipo contiene transitorios, se producen ecos previos y posteriores audibles en las bandas espectrales reconstruidas debido al relleno de recuadros. La figura 7c muestra efectos típicos de ecos previos y posteriores que alteran los transitorios debido a IGF. En el panel izquierdo de la figura 7c se muestra el espectrograma de la señal original y en el panel derecho se muestra el espectrograma de la señal sometida a relleno de recuadros sin filtrado de TTS de la invención. En este ejemplo, la frecuencia de inicio de IGF $f_{IGFstart}$ o f_{split} entre la banda principal y la banda con relleno de recuadros se elige para ser $f_s/4$. En el panel derecho de la figura 7c ecos previos y posteriores claros son visibles alrededor de los transitorios, especialmente prominentes en el extremo espectral superior de la región de frecuencia replicada.

La tarea principal del módulo de TTS es confinar estas componentes de señal no deseadas en estrecha proximidad alrededor de un transitorio y de ese modo ocultarlas en la región temporal gobernada por el efecto de enmascaramiento temporal de la percepción humana. Por lo tanto, los coeficientes de predicción de TTS requeridos se calculan y aplican usando “predicción directa” en el dominio de CMDCT.

En una realización que combina TTS e IGF en un códec, es importante alinear determinados parámetros de TTS y parámetros de IGF de manera que un recuadro de IGF o bien se filtra completamente por un filtro de TTS (filtro de aplanamiento o modelado) o bien no se filtra. Por lo tanto, todas las frecuencias $TTSstart[.]$ o $TTSstop[.]$ no estarán comprendidas dentro de un recuadro de IGF, sino más bien estarán alineadas con las frecuencias $f_{IGF...}$ respectivas. La figura 7d muestra un ejemplo de áreas operativas de TTS e IGF para un conjunto de tres filtros de TTS.

La frecuencia de fin de TTS se ajusta a la frecuencia de fin de la herramienta de IGF, que es mayor que $f_{IGFstart}$. Si el TTS usa más de un filtro, tiene que garantizarse que la frecuencia de cruce entre dos filtros de TTS tiene que coincidir con la frecuencia de división de IGF. De lo contrario, un sub-filtro de TTS funcionará a lo largo de $f_{IGFstart}$ dando como resultado artefactos no deseados como, por ejemplo, sobremodelado.

En la variante de implementación representada en la figura 7a y en la figura 7b, se debe tener especial cuidado de que en ese decodificador las energías de IGF estén correctamente ajustadas. Este es el caso especialmente si, en el transcurso del procesamiento de TTS e IGF, diferentes filtros de TTS que tienen diferentes ganancias de predicción se aplican a la región de origen (como un filtro de aplanamiento) y a la región espectral objetivo (como un filtro de modelado que no es el homólogo exacto de dicho filtro de aplanamiento) de un recuadro de IGF. En este

caso, la razón de la ganancia de predicción de los dos filtros de TTS aplicados ya no es igual a uno y, por lo tanto, debe aplicarse un ajuste de energía por esta razón.

5 En la variante de implementación alternativa, se invierte el orden de procesamiento posterior de IGF y TTS. En el decodificador, esto significa que el ajuste de energía mediante procesamiento posterior de IGF se calcula después del filtrado de TTS y, de este modo, es la etapa de procesamiento final antes de la transformada de síntesis. Por lo tanto, independientemente de las diferentes ganancias de filtro de TTS aplicadas a un recuadro durante la codificación, la energía final siempre se ajusta correctamente mediante el procesamiento de IGF.

10 En el lado de decodificador se aplican de nuevo los coeficientes de filtro de TTS en todo el espectro completo, es decir el espectro principal extendido por el espectro regenerado. La aplicación de TTS es necesaria para formar la envolvente temporal del espectro regenerado para coincidir de nuevo con la envolvente de la señal original. Por lo tanto, se reducen los ecos previos mostrados. Adicionalmente, todavía modela temporalmente el ruido de cuantificación en la señal por debajo $f_{IGFstart}$ como es habitual en el TNS tradicional.

15 En los codificadores tradicionales, la interconexión espectral de una señal de audio (por ejemplo, SBR) corrompe la correlación espectral en los límites de interconexión y, por lo tanto, afecta a la envolvente temporal de la señal de audio introduciendo dispersión. Por lo tanto, otro beneficio de la realización de relleno de recuadros de IGF en la señal residual es que, tras la aplicación del filtro de modelado de TTS, los límites del recuadro se correlacionan perfectamente, lo que da como resultado una reproducción temporal más fiel de la señal.

20 En la figura 7e se muestra el resultado de la señal procesada de manera correspondiente. En comparación con la versión sin filtrar (figura 7c, panel derecho) la señal filtrada de TTS muestra una buena reducción de ecos previos y posteriores no deseados (figura 7e, panel derecho).

25 Asimismo, tal como se comenta, la figura 7a ilustra un codificador que coincide con el decodificador de la figura 7b o el decodificador de la figura 6a. Básicamente, un aparato para codificar una señal de audio comprende un convertidor de tiempo-espectro tal como 702 para la conversión de una señal de audio en una representación espectral. La representación espectral puede ser una representación espectral de valor real o, como se ilustra en el bloque 702, una representación espectral de valor complejo. Además, se proporciona un filtro de predicción tal como 704 para llevar a cabo una predicción sobre la frecuencia para generar valores residuales espectrales, en el que el filtro 704 de predicción se define mediante información del filtro de predicción derivada de la señal de audio y retransmitida a un multiplexor 710 de flujo de bits, como se ilustra en 714 en la figura 7a. Asimismo, se proporciona un codificador de audio tal como el codificador 7004 de audio impulsado psicoacústicamente. El codificador de audio está configurado para codificar un primer conjunto de primeras porciones espectrales de los valores residuales espectrales para obtener un primer conjunto codificado de primeros valores espectrales. Adicionalmente, un codificador paramétrico tal como el que se ilustra en la figura 7a se proporciona para codificar un segundo conjunto de segundas porciones espectrales. Preferentemente, el primer conjunto de primeras porciones espectrales se codifica con resolución espectral superior en comparación con el segundo conjunto de segundas porciones espectrales.

40 Finalmente, tal como se ilustra en la figura 7a, se proporciona una interfaz de salida para emitir la señal codificada que comprende el segundo conjunto paraméricamente codificado de segundas porciones espectrales, el primer conjunto codificado de primeras porciones espectrales y la información de filtro ilustrada como "información secundaria de TTS" en 714 en la figura 7a.

45 Preferentemente, el filtro 704 de predicción comprende un calculador de información de filtro configurado para usar los valores espectrales de la representación espectral para calcular la información de filtro. Asimismo, el filtro de predicción está configurado para calcular los valores residuales espectrales usando los mismos valores espectrales de la representación espectral usada para calcular la información de filtro.

Preferentemente, el filtro 704 de TTS está configurado de la misma manera conocida para codificadores de audio de la técnica anterior que aplican la herramienta de TNS según la norma de AAC.

55 A continuación, se comenta una implementación adicional que usa la decodificación de dos canales en el contexto de las figuras 8a a 8e. Además, se hace referencia a la descripción de los elementos correspondientes en el contexto de las figuras 2a, 2b (codificación 228 de canales conjunta y decodificación 204 de canales conjunta).

60 La figura 8a ilustra un decodificador de audio para generar una señal decodificada de dos canales. El decodificador de audio comprende cuatro decodificadores 802 de audio para decodificar una señal codificada de dos canales para obtener un primer conjunto de primeras porciones espectrales y adicionalmente un decodificador 804 paramétrico para proporcionar datos paramétricos para un segundo conjunto de segundas porciones espectrales y, adicionalmente, una identificación de dos canales que identifica ya sea una primera o una segunda representación de dos canales diferente para las segundas porciones espectrales. Adicionalmente, se proporciona un regenerador

806 de frecuencia para regenerar una segunda porción espectral en función de una primera porción espectral del primer conjunto de primeras porciones espectrales y datos paramétricos para la segunda porción y la identificación de dos canales para la segunda porción. La figura 8b ilustra diferentes combinaciones para representaciones de dos canales en el intervalo de origen y en el intervalo de destino. El intervalo de origen puede estar en la primera representación de dos canales y el intervalo de destino también puede estar en la primera representación de dos canales. Alternativamente, el intervalo de origen puede estar en la primera representación de dos canales y el intervalo de destino puede estar en la segunda representación de dos canales. Además, el intervalo de origen puede estar en la segunda representación de dos canales y el intervalo de destino puede estar en la primera representación de dos canales según se indica en la tercera columna de la figura 8b. Por último, tanto el intervalo de origen como el intervalo de destino pueden estar en la segunda representación de dos canales. En una realización, la primera representación de dos canales es una representación de dos canales por separado, en la que los dos canales de la señal de dos canales están representados individualmente. Entonces, la segunda representación de dos canales es una representación conjunta en la que los dos canales de la representación de dos canales están representados en forma conjunta, es decir, en la que se requiere un procesamiento posterior o la transformada de representación para recalcular una representación de dos canales por separado tal como se requiere para su emisión a altavoces correspondientes.

En una implementación, la primera representación de dos canales puede ser una representación izquierda/derecha (L/R) y la segunda representación de dos canales es una representación de estéreo conjunto. Sin embargo, otras representaciones de dos canales además de izquierda/derecha o M/S o la predicción estéreo se pueden aplicar y usar para la presente invención.

La figura 8c ilustra un diagrama de flujo para operaciones llevadas a cabo por el decodificador de audio de la figura 8a. En una etapa 812, el decodificador 802 de audio lleva a cabo una decodificación del intervalo de origen. El intervalo de origen puede comprender, con respecto a la figura 3a, bandas de factor de escala SCB1 a SCB3. Asimismo, puede haber una identificación de dos canales para cada banda de factor de escala y la banda de factor de escala 1 puede estar, por ejemplo, en la primera representación (tal como L/R) y la tercera banda de factor de escala puede estar en la segunda representación de dos canales tal como M/S o predicción de mezcla descendente/residual. Por lo tanto, la etapa 812 puede dar como resultado diferentes representaciones para diferentes bandas. Entonces, en la etapa 814, el regenerador 806 de frecuencia está configurado para seleccionar un intervalo de origen para una regeneración de frecuencia. En la etapa 816, el regenerador 806 de frecuencia comprueba entonces la representación del intervalo de origen y en el bloque 818 el regenerador 806 de frecuencia compara la representación de dos canales del intervalo de origen con la representación de dos canales del intervalo objetivo. Si ambas representaciones son idénticas, el regenerador 806 de frecuencia proporciona una regeneración de frecuencia por separado para cada canal de la señal de dos canales. Sin embargo, cuando ambas representaciones detectadas en el bloque 818 no son idénticas, entonces se toma el flujo 824 de señal y el bloque 822 calcula la otra representación de dos canales del intervalo de origen y usa esta otra representación de dos canales calculada para la regeneración del intervalo objetivo. Por lo tanto, el decodificador de la figura 8a hace posible regenerar un intervalo de destino que se indica que tiene la segunda identificación de dos canales usando un intervalo de origen que está en la primera representación de dos canales. Naturalmente, la presente invención permite adicionalmente regenerar un intervalo objetivo usando un intervalo de origen que tiene la misma identificación de dos canales. Y, adicionalmente, la presente invención permite regenerar un intervalo objetivo que tiene una identificación de dos canales que indica una representación conjunta de dos canales y a continuación transformar esta representación en una representación de canales por separado, requerida para el almacenamiento o la transmisión a altavoces correspondientes para la señal de dos canales.

Se enfatiza que los dos canales de la representación de dos canales pueden ser dos canales estéreo, tales como el canal izquierdo y el canal derecho. Sin embargo, la señal también puede ser una señal multicanal que tiene, por ejemplo, cinco canales y un canal de altavoz de graves o que tiene incluso más canales. Entonces, un procesamiento de dos canales por pares tal como se comentó en el contexto de las figuras 8a a 8e se puede llevar a cabo en el que los pares pueden ser, por ejemplo, un canal izquierdo y un canal derecho, canal envolvente izquierdo y un canal envolvente derecho, y un canal central y un canal LFE (altavoz de graves). Puede usarse cualquier otro emparejamiento con el fin de representar, por ejemplo, seis canales de entrada mediante tres procedimientos de procesamiento de dos canales.

La figura 8d ilustra un diagrama de bloques de un decodificador de la invención correspondiente a la figura 8a. Un intervalo de origen o un decodificador 830 principal pueden corresponder al decodificador 802 de audio. Los otros bloques 832, 834, 836, 838, 840, 842 y 846 pueden ser partes del regenerador 806 de frecuencia de la figura 8a. En particular, el bloque 832 es un transformador de representación para transformar representaciones del intervalo de origen en bandas individuales de manera que, en la salida del bloque 832 está presente un conjunto completo del intervalo de origen en la primera representación por un lado y en la segunda representación de dos canales por otro lado. Estas dos representaciones completas del intervalo de origen se pueden almacenar en el almacenamiento 834 para ambas representaciones del intervalo de origen.

Entonces, el bloque 836 aplica una generación de recuadros de frecuencia usando, como entrada, una ID del intervalo de origen y usando, además, como entrada una ID de dos canales para el intervalo objetivo. Basándose en la ID de dos canales para el intervalo objetivo, el generador de recuadros de frecuencia accede al almacenamiento 834 y recibe la representación de dos canales del intervalo de origen que coincide con la ID de dos canales para el intervalo objetivo introducida en el generador de recuadros de frecuencia en 835. Por lo tanto, cuando la ID de dos canales para el intervalo objetivo indica el procesamiento de estéreo conjunto, entonces el generador 836 de recuadros de frecuencia accede al almacenamiento 834 con el fin de obtener la representación de estéreo conjunto del intervalo de origen indicado por la ID 833 del intervalo de origen.

El generador 836 de recuadros de frecuencia lleva a cabo esta operación para cada intervalo objetivo y la salida del generador de recuadros de frecuencia es tal que cada canal de la representación de canales identificada por la identificación de dos canales está presente. Luego un ajustador 838 de envolvente lleva a cabo un ajuste de envolvente. El ajuste de envolvente se lleva a cabo en el dominio de dos canales identificado por la identificación de dos canales. Para este fin se requieren parámetros de ajuste de envolvente y estos parámetros se transmiten del codificador al decodificador en la misma representación de dos canales descrita. Cuando la identificación de dos canales en el intervalo objetivo que va a procesarse por el ajustador de envolventes tiene una identificación de dos canales que indica una representación de dos canales diferente de los datos de envolvente para este intervalo objetivo, entonces un transformador 840 de parámetros transforma los parámetros de envolvente en la representación de dos canales requerida. Cuando, por ejemplo, la identificación de dos canales para una banda indica la codificación de estéreo conjunto y cuando los parámetros para este intervalo objetivo se han transmitido como parámetros de envolvente L/R, entonces el transformador de parámetros calcula los parámetros de envolvente estéreo conjunta a partir de los parámetros de envolvente L/R tal como se describe de manera que se usa la representación paramétrica correcta para el ajuste de envolvente espectral de un intervalo objetivo.

En otra realización preferida, los parámetros de envolvente ya se transmiten como parámetros de estéreo conjunto cuando se usa el estéreo conjunto en una banda objetivo.

Cuando se supone que la entrada en el ajustador 838 de envolvente es un conjunto de intervalos objetivo que tienen diferentes representaciones de dos canales, entonces la salida del ajustador 838 de envolvente también es un conjunto de intervalos objetivo en diferentes representaciones de dos canales. Cuando un intervalo objetivo tiene una representación conjunta tal como M/S, entonces este intervalo objetivo es procesado por un transformador 842 de representaciones para calcular la representación separada requerida para un almacenamiento o transmisión a los altavoces. Sin embargo, cuando un intervalo objetivo ya tiene una representación separada se toma el flujo 844 de señal y se evita el transformador 842 de representaciones. En la salida del bloque 842 se obtiene una representación espectral de dos canales que es una representación de dos canales por separado que luego se puede procesar adicionalmente como se indica por el bloque 846, en el que este procesamiento adicional puede ser, por ejemplo, una conversión de frecuencia/tiempo o cualquier otro procesamiento requerido.

Preferentemente, las segundas porciones espectrales corresponden a bandas de frecuencia, y la identificación de dos canales se proporciona como una matriz de indicadores correspondientes a la tabla de la figura 8b, en la que existe un indicador para cada banda de frecuencia. Entonces, el decodificador paramétrico está configurado para comprobar si el indicador está establecido o no y para controlar el regenerador 106 de frecuencia según un indicador para usar ya sea una primera representación o una segunda representación de la primera porción espectral.

En una realización, sólo el intervalo de reconstrucción que se inicia con la frecuencia 309 de inicio de IGF de la figura 3a tiene identificaciones de dos canales para diferentes bandas de reconstrucción. En otra realización, esto también se aplica para el intervalo de frecuencia por debajo de la frecuencia 309 de inicio de IGF.

En una realización adicional, la identificación de la banda de origen y la identificación de la banda objetivo se pueden determinar de forma adaptativa por un análisis de similitud. Sin embargo, el procesamiento de dos canales de la invención también se puede aplicar cuando hay una asociación fija de un intervalo de origen a un intervalo objetivo. Un intervalo de origen se puede usar para recrear, con respecto a la frecuencia, un intervalo objetivo más amplio, ya sea por una operación de relleno de recuadros de frecuencia de armónicos o una operación de relleno de recuadros de frecuencia de copiado usando dos o más operaciones de relleno de recuadros de frecuencia similares al procesamiento para múltiples interconexiones conocidas a partir del procesamiento de AAC de alta eficiencia.

La figura 8e ilustra un codificador de audio para codificar una señal de audio de dos canales. El codificador comprende un convertidor 860 de tiempo-espectro para convertir la señal de audio de dos canales en una representación espectral. Asimismo, un analizador 866 espectral para convertir la señal de audio de canal de audio de dos canales en una representación espectral. Además se proporciona un analizador 866 espectral para llevar a cabo un análisis con el fin de determinar qué porciones espectrales deben codificarse con una alta resolución, es decir, para hallar el primer conjunto de primeras porciones espectrales y para hallar adicionalmente el segundo conjunto de segundas porciones espectrales.

Adicionalmente, se proporciona un analizador 864 de dos canales para analizar el segundo conjunto de segundas porciones espectrales para determinar una identificación de dos canales que identifica o bien una primera representación de dos canales o bien una segunda representación de dos canales.

5 Dependiendo del resultado del analizador de dos canales, una banda en la segunda representación espectral se parametriza usando o bien la primera representación de dos canales o bien la segunda representación de dos canales, y esto se lleva a cabo mediante un codificador 868 de parámetros. El intervalo de frecuencia principal, es decir la banda de frecuencia por debajo de la frecuencia 309 de inicio de IGF de la figura 3a, se codifica por un codificador 870 principal. El resultado de los bloques 868 y 870 se introduce en una interfaz 872 de salida. Como se
10 indicó, el analizador de dos canales proporciona una identificación de dos canales para cada banda, ya sea por encima de la frecuencia de inicio de IGF o para todo el intervalo de frecuencia, y esta identificación de dos canales también se retransmite a la interfaz 872 de salida de manera que estos datos también se incluyen en una señal 873 codificada emitida por la interfaz 872 de salida.

15 Asimismo se prefiere que el codificador de audio comprenda un transformador 862 por bandas. Basándose en la decisión del analizador 862 de dos canales, la señal de salida del convertidor 862 de tiempo-espectro se transforma en una representación indicada por el analizador de dos canales y, en particular, por la ID 835 de dos canales. Por lo tanto, una salida del transformador 862 por bandas es un conjunto de bandas de frecuencia en el que cada banda de frecuencia puede estar o bien en la primera representación de dos canales o bien en la segunda representación
20 de dos canales diferente. Cuando se aplica la presente invención en banda completa, es decir cuando el intervalo de origen y el intervalo de reconstrucción se procesan ambos por el transformador por bandas, el analizador 860 espectral puede analizar esta representación. Sin embargo, alternativamente, el analizador 860 espectral también puede analizar la señal emitida por el convertidor de tiempo-espectro tal como se indica por la línea 861 de control. Por lo tanto, el analizador 860 espectral puede aplicar el análisis de tonalidad preferido en la salida del transformador 862 por bandas o la salida del convertidor 860 de tiempo-espectro antes de procesarse por el transformador 862 por bandas. Asimismo, el analizador espectral puede aplicar la identificación del intervalo de
25 origen de mejor coincidencia para un determinado intervalo objetivo, ya sea en el resultado del transformador 862 por bandas o en el resultado del convertidor 860 de tiempo-espectro.

30 A continuación, se hace referencia a las figuras 9a a 9d para ilustrar un cálculo preferido de los valores de información de energía ya comentados en el contexto de la figura 3a y la figura 3b.

Los codificadores de audio modernos del estado de la técnica aplican diversas técnicas para minimizar la cantidad de datos que representan una señal de audio determinada. Los codificadores de audio tales como USAC [1] aplican
35 una transformación de tiempo a frecuencia como MDCT para obtener una representación espectral de una señal de audio determinada. Estos coeficientes de MDCT se cuantifican aprovechando los aspectos psicoacústicos del sistema auditivo humano. Si se reduce la tasa de bits disponible, la cuantificación se vuelve más gruesa introduciendo grandes números de valores espectrales reducidos a cero lo que conduce a artefactos audibles en el lado de decodificador. Para mejorar la calidad perceptual, los decodificadores del estado de la técnica rellenan estas partes espectrales reducidas a cero con ruido aleatorio. El método de IGF recolecta recuadros a partir de la señal
40 distinta de cero restante para rellena esos huecos en el espectro. Es crucial para la calidad perceptual de la señal de audio decodificada que se conserven la envolvente espectral y la distribución de energía de los coeficientes espectrales. El método de ajuste de energía presentado en el presente documento usa información secundaria transmitida para reconstruir la envolvente de MDCT espectral de la señal de audio.

45 Dentro de eSBR [15] la señal de audio se muestrea de manera descendente al menos por un factor de dos y la parte de alta frecuencia del espectro se reduce completamente a cero [1, 17]. Esta parte eliminada se sustituye mediante técnicas paramétricas, eSBR, en el lado de decodificador. La eSBR implica el uso de una transformada adicional, la transformación de QMF que se usa para sustituir la parte de alta frecuencia vacía y para volver a muestrear la señal de audio [17]. Esto añade tanto complejidad computacional como consumo de memoria a un codificador de audio.

El codificador de USAC [15] ofrece la posibilidad de rellena huecos espectrales (líneas espectrales reducidas a cero) con ruido aleatorio pero presenta los siguientes inconvenientes: el ruido aleatorio no puede conservar la estructura fina temporal de una señal transitoria y no puede conservar la estructura armónica de una señal tonal.

55 El área en la que actúa la eSBR en el lado de decodificador se eliminó completamente por el codificador [1]. Por lo tanto, la eSBR es propensa a eliminar líneas tonales en la región de alta frecuencia o distorsionar estructuras armónicas de la señal original. Como la resolución de frecuencia de QMF de la eSBR es muy baja y la reinsertión de componentes sinusoidales sólo es posible en la resolución gruesa del banco de filtros subyacente, la regeneración de componentes tonales en la eSBR en el intervalo de frecuencia replicado tiene una precisión muy baja.
60

La eSBR usa técnicas para ajustar energías de áreas interconectadas, el ajuste de la envolvente espectral [1]. Esta técnica usa valores de energía transmitidos en una cuadrícula de frecuencia-tiempo de QMF para remodelar la envolvente espectral. Esta técnica del estado de la técnica no se ocupa de espectros parcialmente eliminados y

debido a la alta resolución temporal es propensa o bien a necesitar una cantidad relativamente grande de bits para transmitir valores de energía apropiados o bien a aplicar una cuantificación gruesa a los valores de energía.

5 El método de IGF no necesita una transformación adicional, ya que usa la transformación de MDCT tradicional que se calcula como se describe en [15].

El método de ajuste de energía presentado en el presente documento usa información secundaria generada por el codificador para reconstruir la envolvente espectral de la señal de audio. Esta información secundaria se genera por el codificador según se indica a continuación:

10 a) Aplicar una transformada de MDCT dividida en ventanas a la señal de audio de entrada [16, sección 4.6], opcionalmente calcular una MDST dividida en ventanas, o estimar una MDST dividida en ventanas a partir de la MDCT calculada.

15 b) Aplicar TNS/ TTS en los coeficientes de MDCT [15, sección 7.8]

c) Calcular la energía promedio para cada banda de factor de escala de MDCT por encima de la frecuencia de inicio de IGF ($f_{IGFstart}$) hasta la frecuencia de fin de IGF ($f_{IGFstop}$)

20 d) Cuantificar los valores de energía promedio

$f_{IGFstart}$ y $f_{IGFstop}$ son parámetros dados por el usuario.

25 Los valores calculados en las etapas c) y d) se codifican sin pérdidas y se transmiten como información secundaria con el flujo de bits al decodificador.

El decodificador recibe los valores transmitidos y los usa para ajustar la envolvente espectral.

30 a) Descuantificar los valores de MDCT transmitidos

b) Aplicar el relleno de ruido de USAC tradicional si se indica

c) Aplicar el relleno de recuadros de IGF

35 d) Descuantificar los valores de energía transmitidos

e) Ajustar la envolvente espectral por banda de factor de escala

40 f) Aplicar TNS/TTS si se indica

Sea $\hat{x} \in \mathbb{R}^N$ la representación espectral de valor real, sometida a transformada MDCT, de una señal de audio dividida en ventanas con una longitud de ventana de $2N$. Esta transformación se describe en [16]. El codificador aplica opcionalmente TNS en \hat{x} .

45 En [16, 4.6.2] se describe una partición de \hat{x} en bandas de factor de escala. Las bandas de factor de escala son un conjunto de un conjunto de índices y se indican en este texto con scb_k .

50 Los límites de cada scb_k , siendo $k = 0, 1, 2, \dots, \text{max_sfb}$ se definen por una matriz swb_offset (16, 4.6.2), en la que $swb_offset[k]$ y $swb_offset[k + 1]-1$ definen el primer y el último índice para la línea de coeficiente espectral más baja y más alta contenida en scb_k . La banda de factor de escala se indica de la siguiente manera:

$$scb_k = \{swb_offset[k], 1+ swb_offset[k], 2+ swb_offset[k], \dots, swb_offset[k+1]-1\}$$

55 Si la herramienta de IGF se usa por el codificador, el usuario define una frecuencia de inicio de IGF y una frecuencia de fin de IGF. Estos dos valores se mapean al índice de banda de factor de escala de mejor ajuste $igfStartSfb$ e $igfStopSfb$. Ambos se indican en el flujo de bits al decodificador.

[16] describe una transformación tanto de bloque largo como de bloque corto. Para los bloques largos sólo se

transmite un conjunto de coeficientes espectrales junto con un conjunto de factores de escala al decodificador. Para los bloques cortos se calculan ocho ventanas cortas con ocho conjuntos diferentes de coeficientes espectrales. Para ahorrar tasa de bits, los factores de escala de dichas ocho ventanas de bloques cortos se agrupan por el codificador.

- 5 En el caso de IGF, el método presentado en el presente documento usa bandas de factor de escala de legado para agrupar valores espectrales que se transmiten al decodificador:

$$E_k = \sqrt{\frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{x}_i^2}$$

- 10 donde $k = igfStartSfb, 1 + igfStartSfb, 2 + igfStartSfb, \dots, igfEndSfb$. Para cuantificar, se calcula

$$\hat{E}_k = nINT(4 \log_2(E_k)).$$

- 15 Todos los valores \hat{E}_k se transmiten al decodificador.

Se supone que el codificador decide agrupar num_window_group conjuntos de factor de escala. Se indica con w este agrupamiento-partición del conjunto $\{0, 1, 2, \dots, 7\}$ que son los índices de las ocho ventanas cortas. w_l indica el l-simo subconjunto de w, en el que l indica el índice del grupo de ventana, $0 \leq l < \text{num_window_group}$.

- 20 Para el cálculo de bloques cortos, se mapea la frecuencia de inicio/fin de IGF definida por el usuario a bandas de factor de escala apropiadas. Sin embargo, por razones de simplicidad también se indica para bloques cortos $k = igfStartSfb, 1 + igfStartSfb, 2 + igfStartSfb, \dots, igfEndSfb$.

- 25 El cálculo de la energía de IGF usa la información de agrupamiento para agrupar los valores $E_{k,l}$:

$$E_{k,l} := \sqrt{\frac{1}{|w_l|} \sum_{j \in w_l} \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{x}_{j,i}^2}$$

Para cuantificar, se calcula

- 30
$$\hat{E}_{k,l} = nINT(4 \log_2(E_{k,l}))$$

Todos los valores $\hat{E}_{k,l}$ se transmiten al decodificador.

- 35 Las fórmulas de codificación mencionadas anteriormente funcionan usando sólo coeficientes de MDCT de valor real \hat{x} . Para obtener una distribución de energía más estable en el intervalo de IGF, es decir para reducir las fluctuaciones de amplitud temporal, se puede usar un método alternativo para calcular los valores \hat{E}_k :

Sea $\hat{x}_r \in \mathbb{R}^N$ la representación espectral de valor real, sometida a transformada MDCT, de una señal de audio dividida en ventanas con una longitud de ventana de $2N$, y $\hat{x}_i \in \mathbb{R}^N$ la representación espectral de valor real, sometida a transformada MDST, de la misma porción de la señal de audio. La representación espectral de MDST, \hat{x}_i , puede o bien calcularse de manera exacta o bien estimarse a partir de \hat{x}_r . $\hat{c} := (\hat{x}_r, \hat{x}_i) \in \mathbb{C}^N$ indica la representación espectral compleja de la señal de audio dividida en ventanas, que tiene \hat{x}_r como parte real y \hat{x}_i como parte imaginaria. El codificador aplica opcionalmente TNS en \hat{x}_r y \hat{x}_i .

- 45 Ahora, la energía de la señal original en el intervalo de IGF se puede medir con

$$E_{ok} = \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{c}_i^2$$

- 50 Las energías de valor real y complejo de la banda de reconstrucción, es decir el recuadro que se debe usar en el

lado de decodificador en la reconstrucción del intervalo de IGF scb_k , se calculan con:

$$E_{tk} = \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{c}_i^2, \quad E_{rk} = \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{x}_{r_i}^2$$

5 donde tr_k es un conjunto de índices, el intervalo de recuadro de origen asociado, en función de scb_k . En las dos fórmulas anteriores, en lugar del conjunto de índices scb_k , se puede usar el conjunto $\overline{scb_k}$ (definido más adelante en este texto) para crear tr_k para lograr valores más precisos E_t y E_r .

Calcular

10

$$f_k = \frac{E_{ok}}{E_{tk}}$$

si $E_{tk} > 0$, de lo contrario $f_k = 0$.

15 Con:

$$E_k = \sqrt{f_k E_{rk}}$$

20 ahora se calcula una versión más estable de E_k , ya que un cálculo de E_k con los valores de MDCT sólo resulta afectado por el hecho de que los valores de MDCT no obedecen el teorema de Parseval y, por lo tanto, no reflejan la información de energía completa de los valores espectrales. \tilde{E}_k se calcula como anteriormente.

25 Como se indicó anteriormente, para bloques cortos se supone que el codificador decide agrupar num_window_group conjuntos de factor de escala. Como anteriormente, w_l indica el l -ésimo subconjunto de w , en el que l indica el índice del grupo de ventana, $0 \leq l < num_window_group$.

Una vez más, puede calcularse la versión alternativa descrita anteriormente para calcular una versión más estable de $E_{k,l}$. Con las definiciones de $\hat{c} := (\hat{x}_r, \hat{x}_i) \in \mathbb{C}^N$, $\hat{x}_r \in \mathbb{R}^N$ la señal de audio dividida en ventanas con una longitud de $2N$ sometida a transformada MDCT y siendo $\hat{x}_i \in \mathbb{R}^N$ la sometida a transformada MDST, calcular

30

$$E_{ok,l} = \frac{1}{|w_l|} \sum_{l \in w_l} \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{c}_{i,l}^2$$

Calcular de manera análoga

$$35 \quad E_{tk,l} = \frac{1}{|w_l|} \sum_{l \in w_l} \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{c}_{i,l}^2, \quad E_{rk,l} = \frac{1}{|w_l|} \sum_{l \in w_l} \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{x}_{r,l}^2$$

y continuar con el factor $f_{k,l}$

$$f_{k,l} = \frac{E_{ok,l}}{E_{tk,l}}$$

40

que se usa para ajustar $E_{rk,l}$ calculado anteriormente:

$$E_{k,l} = \sqrt{f_{k,l} E_{rk,l}}$$

45

$\hat{E}_{k,l}$ se calcula como anteriormente.

El procedimiento que no usar sólo la energía de la banda de reconstrucción, ya sea derivada de la banda de reconstrucción compleja o de los valores de MDCT, sino también usar una información de energía del intervalo de origen proporciona una reconstrucción de energía mejorada.

Específicamente, el calculador 1006 de parámetros está configurado para calcular la información de energía para la banda de reconstrucción usando información sobre la energía de la banda de reconstrucción y usando, además, información sobre una energía de un intervalo de origen para usarse para la reconstrucción de la banda de reconstrucción.

Asimismo, el calculador 1006 de parámetros está configurado para calcular una información de energía (E_{ok}) en la banda de reconstrucción de un espectro complejo de la señal original, para calcular una información de energía adicional (E_{rk}) en un intervalo de origen de una parte de valor real del espectro complejo de la señal original para usarse para reconstruir la banda de reconstrucción, y en el que el calculador de parámetros está configurado para calcular la información de energía para la banda de reconstrucción usando la información de energía (E_{ok}) y la información de energía adicional (E_{rk}).

Además, el calculador 1006 de parámetros está configurado para determinar una primera información de energía (E_{ok}) en una banda de factor de escala que va a reconstruirse de un espectro complejo de la señal original, para determinar una segunda información de energía (E_{tk}) en un intervalo de origen del espectro complejo de la señal original que va a usarse para reconstruir la banda de factor de escala que va a reconstruirse, para determinar una tercera información de energía (E_{rk}) en un intervalo de origen de una parte de valor real del espectro complejo de la señal original que va a usarse para reconstruir la banda de factor de escala que va a reconstruirse, para determinar una información de ponderación basándose en una relación entre al menos dos de la primera información de energía, la segunda información de energía y la tercera información de energía, y para ponderar una de la primera información de energía y la tercera información de energía usando la información de ponderación para obtener una información de energía ponderada y para usar la información de energía ponderada como información de energía para la banda de reconstrucción.

Los siguientes son ejemplos para los cálculos, pero a los expertos en la técnica se les pueden ocurrir muchos otros en vista del principio general anterior:

A)

$$f_k = E_{ok}/E_{tk};$$

$$E_k = \sqrt{f_k * E_{rk}};$$

B)

$$f_k = E_{tk}/E_{ok};$$

$$E_k = \sqrt{(1/f_k) * E_{rk}};$$

C)

$$f_k = E_{rk}/E_{tk};$$

$$E_k = \sqrt{f_k * E_{ok}}$$

D)

$$f_k = E_{tk}/E_{rk};$$

$$E_k = \sqrt{(1/f_k) * E_{ok}}$$

Todos estos ejemplos reconocen que aunque sólo se procesan valores reales de MDCT en el lado de decodificador, el cálculo real es, debido al solapamiento y la adición, del procedimiento de cancelación de distorsión de dominio de tiempo realizado implícitamente usando números complejos. Sin embargo, en particular, la determinación 918 de la información de energía de recuadro de las porciones 922, 923 espectrales adicionales de la banda 920 de reconstrucción para valores de frecuencia diferentes de la primera porción 921 espectral que tiene frecuencias en la banda 920 de reconstrucción, se basa en valores reales de MDCT. Por lo tanto, la información de energía transmitida al decodificador será generalmente menor que la información de energía E_{ok} sobre la banda de

reconstrucción del espectro complejo de la señal original. Por ejemplo, para el caso C anterior, esto significa que el factor f_k (información de ponderación) será menor que 1.

5 En el lado de decodificador, si la herramienta de IGF se indica como ON (activada), los valores transmitidos \hat{E}_k se obtienen a partir del flujo de bits y se descuantificarán con

$$E_k = 2^{\frac{1}{4}\hat{E}_k}$$

10 para todos $k = igfStartSfb, 1 + igfStartSfb, 2 + igfStartSfb, \dots, igfEndSfb$.

Un decodificador descuantifica los valores de MDCT transmitidos a $x \in \mathbb{R}^N$ y calcula la energía de conservación restante:

$$sE_k := \sum_{i \in scb_k} x_i^2$$

15 donde k está en el intervalo definido anteriormente.

20 Se indica $\overline{scb}_k = \{i | i \in scb_k \wedge x_i = 0\}$. Este conjunto contiene todos los índices de la banda de factor de escala scb_k que se han cuantificado a cero por el codificador.

El método de subbanda de IGF (no descrito en el presente documento) se usa para rellenar huecos espectrales que resultan de una cuantificación gruesa de valores espectrales de MDCT en el lado de codificador usando valores distintos de cero de la MDCT transmitida. x contendrá adicionalmente valores que reemplazan a todos los valores anteriores reducidos a cero. La energía del recuadro se calcula mediante:

$$25 \quad tE_k := \sum_{i \in \overline{scb}_k} x_i^2$$

donde k está en el intervalo definido anteriormente.

30 La energía que falta en la banda de reconstrucción se calcula mediante:

$$mE_k := |scb_k| E_k^2 - sE_k$$

35 Y el factor de ganancia para el ajuste se obtiene mediante:

$$g := \begin{cases} \sqrt{\frac{mE_k}{tE_k}} & \text{Si } (mE_k > 0 \wedge tE_k > 0) \\ \text{sino } 0 & \end{cases}$$

40 con:

$$g' = \min(g, 10)$$

45 El ajuste de la envolvente espectral que usa el factor de ganancia es:

$$x_i := g' x_i$$

50 para todos los $i \in \overline{scb}_k$ y k está en el intervalo definido anteriormente.

Esto remodela la envolvente espectral de x a la forma de la envolvente espectral original \hat{x} .

En principio, con la secuencia de ventanas cortas, todos los cálculos definidos anteriormente permanecen iguales, pero se tiene en cuenta el agrupamiento de bandas de factor de escala. Se indica como $E_{k,l}$ los valores de energía agrupados y descuantificados, obtenidos a partir del flujo de bits. Calcular

5

$$sE_{k,l} := \frac{1}{|w_l|} \sum_{j \in w_l} \sum_{i \in scb_{j,k}} x_{j,i}^2$$

y

10

$$pE_{k,l} := \frac{1}{|w_l|} \sum_{j \in w_l} \sum_{i \in scb_{j,k}} x_{j,i}^2$$

El índice j describe el índice de ventana de la secuencia de bloques cortos.

15 Calcular

$$mE_{k,l} := |scb_k| E_{k,l}^2 - sE_{k,l}$$

y

20

$$g := \begin{cases} \sqrt{\frac{mE_{k,l}}{pE_{k,l}}} & \text{Si } (mE_{k,l} > 0 \wedge pE_{k,l} > 0) \\ \text{sino } 0 & \end{cases}$$

25

Con

$$g' = \min(g, 10)$$

30 Aplicar

$$x_{j,i} := g' x_{j,i}$$

para todos los $i \in \overline{scb_{k,l}}$.

35

Para las aplicaciones de tasas de bits bajas es posible un agrupamiento por pares de los valores E_k sin perder demasiada precisión. Este método se aplica sólo con bloques largos:

$$E_{k \gg 1} = \sqrt{\frac{1}{|scb_k \cup scb_{k+1}|} \sum_{i \in scb_k \cup scb_{k+1}} \hat{x}_i^2}$$

40

donde $k = igfStartSfb, 2 + igfStartSfb, 4 + igfStartSfb, \dots, igfEndSfb$. Nuevamente, tras la descuantificación, todos los valores $E_{k \gg 1}$ se transmiten al decodificador.

45

La figura 9a ilustra un aparato para decodificar una señal de audio codificada que comprende una representación codificada de un primer conjunto de primeras porciones espectrales y una representación codificada de datos paramétricos que indican energías espectrales para un segundo conjunto de segundas porciones espectrales. El primer conjunto de primeras porciones espectrales se indica en 901a en la figura 9a, y la representación codificada de los datos paramétricos se indica en 901b en la figura 9a. Se proporciona un decodificador 900 de audio para decodificar la representación 901a codificada del primer conjunto de primeras porciones espectrales para obtener un

primer conjunto decodificado de primeras porciones 904 espectrales y para decodificar la representación codificada de los datos paramétricos para obtener datos 902 paramétricos decodificados para el segundo conjunto de segundas porciones espectrales que indican energías individuales para bandas de reconstrucción individuales, en el que las segundas porciones espectrales están ubicadas en las bandas de reconstrucción. Además, se proporciona un regenerador 906 de frecuencia para reconstruir valores espectrales de una banda de reconstrucción que comprende una segunda porción espectral. El regenerador 906 de frecuencia usa una primera porción espectral del primer conjunto de primeras porciones espectrales y una información de energía individual para la banda de reconstrucción, en el que la banda de reconstrucción comprende una primera porción espectral y la segunda porción espectral. El regenerador 906 de frecuencia comprende un calculador 912 para determinar una información de energía de conservación que comprende una energía acumulada de la primera porción espectral que tiene frecuencias en la banda de reconstrucción. Asimismo, el regenerador 906 de frecuencia comprende un calculador 918 para determinar una información de energía de recuadro de porciones espectrales adicionales de la banda de reconstrucción y para valores de frecuencia que son diferentes de la primera porción espectral, en el que estos valores de frecuencia tienen frecuencias en la banda de reconstrucción, en el que las porciones espectrales adicionales van a generarse mediante regeneración de frecuencia usando una primera porción espectral diferente de la primera porción espectral en la banda de reconstrucción.

El regenerador 906 de frecuencia comprende además un calculador 914 para una energía faltante en la banda de reconstrucción, y el calculador 914 funciona usando la energía individual para la banda de reconstrucción y la energía de conservación generada por el bloque 912. Además, el regenerador 906 de frecuencia comprende un ajustador 916 de envolvente espectral para el ajuste de las porciones espectrales adicionales en la banda de reconstrucción basándose en la información de energía faltante y la información de energía de recuadro generada por el bloque 918.

Se hace referencia a la figura 9c que ilustra una determinada banda 920 de reconstrucción. La banda de reconstrucción comprende una primera porción espectral en la banda de reconstrucción tal como la primera porción 306 espectral en la figura 3a ilustrada esquemáticamente en 921. Asimismo, el resto de los valores espectrales en la banda 920 de reconstrucción se deben generar usando una región de origen, por ejemplo, a partir de la banda de factor de escala 1, 2, 3 por debajo de la frecuencia 309 de inicio del relleno inteligente de huecos de la figura 3a. El regenerador 906 de frecuencia está configurado para generar valores espectrales sin procesar para las segundas porciones 922 y 923 espectrales. Luego se calcula un factor de ganancia g como se ilustra en la figura 9c con el fin de ajustar finalmente los valores espectrales sin procesar en las bandas 922, 923 de frecuencia con el fin de obtener las segundas porciones espectrales reconstruidas y ajustadas en la banda 920 de reconstrucción, que ahora tienen la misma resolución espectral, es decir, la misma distancia de línea que la primera porción 921 espectral. Es importante entender que la primera porción espectral en la banda de reconstrucción ilustrada en 921 en la figura 9c se decodifica por el decodificador 900 de audio y no se ve influida por el ajuste de envolvente llevado a cabo en el bloque 916 de la figura 9b. En cambio, la primera porción espectral en la banda de reconstrucción indicada en 921 se deja como está, ya que esta primera porción espectral se emite por el decodificador 900 de audio de tasa completa o ancho de banda completo a través de la línea 904.

A continuación se comenta un determinado ejemplo con números reales. La energía de conservación restante calculada por el bloque 912, por ejemplo, es de cinco unidades de energía y esta energía es la energía de las cuatro líneas espectrales indicadas a modo de ejemplo en la primera porción espectral 921.

Asimismo, el valor de energía E_3 para la banda de reconstrucción que corresponde a la banda de factor de escala 6 de la figura 3b o la figura 3a es igual a 10 unidades. Resulta importante que el valor de la energía no sólo comprende la energía de las porciones 922, 923 espectrales, sino también la energía total de la banda 920 de reconstrucción calculada en el lado de codificador, es decir, antes de llevar a cabo el análisis espectral usando, por ejemplo, la máscara de la tonalidad. Por lo tanto, las diez unidades de energía abarcan las primeras y las segundas porciones espectrales en la banda de reconstrucción. Entonces, se supone que la energía de los datos del intervalo de origen para los bloques 922, 923 o de los datos sin procesar de intervalo objetivo para el bloque 922, 923 es igual a ocho unidades de energía. Por lo tanto, se calcula una energía faltante de cinco unidades.

Basándose en la energía faltante dividida entre la energía de recuadro tE_k se calcula un factor de ganancia de 0,79. Entonces, las líneas espectrales sin procesar para las segundas porciones 922, 923 espectrales se multiplican por el factor de ganancia calculado. De este modo, sólo se ajustan los valores espectrales para las segundas porciones 922, 923 espectrales y las líneas espectrales para la primera porción espectral 921 no se ven influidas por este ajuste de envolvente. Después de la multiplicación de los valores espectrales sin procesar para las segundas porciones 922, 923 espectrales, se ha calculado una banda de reconstrucción completa que consiste en las primeras porciones espectrales en la banda de reconstrucción, y que consiste en líneas espectrales en las segundas porciones 922, 923 espectrales en la banda 920 de reconstrucción.

Preferentemente, el intervalo de origen para generar los datos espectrales sin procesar en las bandas 922, 923 está, con respecto a la frecuencia, por debajo de la frecuencia 309 de inicio de IGF y la banda 920 de reconstrucción está

por encima de la frecuencia 309 de inicio de IGF.

Además, se prefiere que los límites de la banda de reconstrucción coincidan con los límites de la banda de factor de escala. Por lo tanto, una banda de reconstrucción tiene, en una realización, el tamaño de bandas de factor de escala correspondientes del decodificador de audio principal o están dimensionadas de manera que, cuando se aplica el emparejamiento de energía, un valor de energía para una banda de reconstrucción proporciona la energía de dos o un número entero superior de bandas de factor de escala. Por lo tanto, cuando se supone que la acumulación de energía se lleva a cabo para la banda de factor de escala 4, la banda de factor de escala 5 y la banda de factor de escala 6, entonces el límite de frecuencia inferior de la banda 920 de reconstrucción es igual al límite inferior de la banda de factor de escala 4 y el límite de frecuencia superior de la banda 920 de reconstrucción coincide con el límite superior de la banda de factor de escala 6.

A continuación se comenta la figura 9d con el fin de mostrar funcionalidades adicionales del decodificador de la figura 9a. El decodificador 900 de audio recibe los valores espectrales descuantificados correspondientes a primeras porciones espectrales del primer conjunto de porciones espectrales y, adicionalmente, se proporcionan factores de escala para bandas de factor de escala tal como se ilustra en la figura 3b a un bloque 940 de ajuste de escala inverso. El bloque 940 de ajuste de escala inverso proporciona todos los primeros conjuntos de primeras porciones espectrales por debajo de la frecuencia 309 de inicio de IGF de la figura 3a y, adicionalmente, las primeras porciones espectrales por encima de la frecuencia de inicio de IGF, es decir, las primeras porciones 304, 305, 306, 307 espectrales de la figura 3a que están todas ubicadas en una banda de reconstrucción ilustrada en 941 en la figura 9d. Por otra parte, las primeras porciones espectrales en la banda de origen usada para el relleno de recuadros de frecuencia en la banda de reconstrucción se proporcionan al ajustador/calculador 942 de envolvente y este bloque recibe además la información de energía para la banda de reconstrucción proporcionada como información secundaria paramétrica de la señal de audio codificada ilustrada en 943 en la figura 9d. Luego, el ajustador/calculador 942 de envolvente proporciona las funcionalidades de las figuras 9b y 9c y por último emite valores espectrales ajustados para las segundas porciones espectrales en la banda de reconstrucción. Estos valores 922, 923 espectrales ajustados para las segundas porciones espectrales en la banda de reconstrucción y las primeras porciones 921 espectrales en la banda de reconstrucción indicada en la línea 941 en la figura 9d representan conjuntamente la representación espectral completa de la banda de reconstrucción.

A continuación se hace referencia a las figuras 10a a 10b para explicar realizaciones preferidas de un codificador de audio que codifica una señal de audio para proporcionar o generar una señal de audio codificada. El codificador comprende un convertidor 1002 de tiempo/espectro que alimenta un analizador 1004 espectral, y el analizador 1004 espectral está conectado a un calculador 1006 de parámetros por un lado y a un codificador 1008 de audio por otro lado. El codificador 1008 de audio proporciona la representación codificada de un primer conjunto de primeras porciones espectrales y no abarca el segundo conjunto de segundas porciones espectrales. Por otra parte, el calculador de parámetros 1006 proporciona información de energía para una banda de reconstrucción que abarca las primeras y las segundas porciones espectrales. Asimismo, el codificador 1008 de audio está configurado para generar una primera representación codificada del primer conjunto de primeras porciones espectrales que tienen la primera resolución espectral, en el que el codificador 1008 de audio proporciona factores de escala para todas las bandas de la representación espectral generada por el bloque 1002. Adicionalmente, tal como se ilustra en la figura 3b, el codificador proporciona información de energía al menos para las bandas de reconstrucción ubicadas, con respecto a la frecuencia, por encima de la frecuencia 309 de inicio de IGF como se ilustra en la figura 3a. Por lo tanto, para bandas de reconstrucción que coinciden preferentemente con bandas de factor de escala o con grupos de bandas de factor de escala, se proporcionan dos valores, es decir, el factor de escala correspondiente a partir del codificador 1008 de audio y, adicionalmente, la información de la energía emitida por el calculador 1006 de parámetros.

Preferentemente, el codificador de audio tiene bandas de factor de escala con diferentes anchos de banda de frecuencia, es decir, con un número diferente de valores espectrales. Por lo tanto, el calculador paramétrico comprende un normalizador 1012 para normalizar las energías para el ancho de banda diferente con respecto al ancho de banda de la banda de reconstrucción específica. Para este fin, el normalizador 1012 recibe, como entradas, una energía en la banda y un número de valores espectrales en la banda y el normalizador 1012 emite entonces una energía normalizada por banda de reconstrucción/de factor de escala.

Además, el calculador 1006a paramétrico de la figura 10a comprende un calculador de valor de energía que recibe información de control del codificador 1008 de audio o principal como se ilustra en la línea 1007 en la figura 10a. Esta información de control puede comprender información sobre bloques largos/cortos usados por el codificador de audio y/o información de agrupamiento. Por consiguiente, aunque la información sobre bloques largos/cortos y la información de agrupamiento sobre ventanas cortas se refieren a un agrupamiento "temporal", la información de agrupamiento puede referirse además a un agrupamiento espectral, es decir, el agrupamiento de dos bandas de factor de escala en una sola banda de reconstrucción. Por lo tanto, el calculador 1014 de valor de energía emite un único valor de energía para cada banda agrupada que abarca una primera y una segunda porción espectral cuando sólo se han agrupado porciones espectrales.

La figura 10d ilustra una realización adicional para la implementación del agrupamiento espectral. Para este fin, el bloque 1016 está configurado para calcular valores de energía para dos bandas adyacentes. A continuación, en el bloque 1018 se comparan los valores de energía para las bandas adyacentes y, cuando los valores de energía no son tan diferentes o son menos diferentes que lo definido, por ejemplo, por un umbral, entonces se genera un único valor (normalizado) para ambas bandas según se indica en el bloque 1020. Como se ilustra en la línea 1019, el bloque 1018 se puede omitir. Asimismo, la generación de un único valor para dos o más bandas que se lleva a cabo en el bloque 1020 puede controlarse mediante un control 1024 de tasa de bits de codificador. Por lo tanto, cuando la tasa de bits se debe reducir, el control 1024 de tasa de bits codificado controla al bloque 1020 para generar un único valor normalizado para dos o más bandas, aunque la comparación en el bloque 1018 no se hubiera permitido para agrupar los valores de información de energía.

En caso de que el codificador de audio lleve a cabo el agrupamiento de dos o más ventanas cortas, este agrupamiento también se aplica para la información de energía. Cuando el codificador principal lleva a cabo un agrupamiento de dos o más bloques cortos, entonces, para estos dos o más bloques, sólo se calcula y se transmite un único conjunto de factores de escala. En el lado de decodificador, el decodificador de audio aplica entonces el mismo conjunto de factores de escala para ambas ventanas agrupadas.

En cuanto al cálculo de información de energía, los valores espectrales en la banda de reconstrucción se acumulan sobre dos o más ventanas cortas. Dicho de otro modo, esto significa que los valores espectrales en una determinada banda de reconstrucción para un bloque corto y para el bloque corto posterior se acumulan y sólo se transmite un único valor de información de energía para esta banda de reconstrucción que abarca dos bloques cortos. A continuación, en el lado de decodificador, el ajuste de envolvente comentado con respecto a las figuras 9a a 9d no se lleva a cabo individualmente para cada bloque corto, sino que se lleva a cabo conjuntamente para el conjunto de ventanas cortas agrupadas.

Luego se aplica nuevamente la normalización correspondiente de manera que, aunque se haya llevado a cabo cualquier agrupamiento en la frecuencia o agrupamiento temporal, la normalización permite fácilmente que, para el cálculo de información de valor de energía en el lado de decodificador, sólo debe conocerse el valor de información de energía por un lado y la cantidad de líneas espectrales en la banda de reconstrucción o en el conjunto de bandas de reconstrucción agrupadas.

En los esquemas de BWE del estado de la técnica, la reconstrucción de la región espectral de HF por encima de una denominada frecuencia de cruce dada se basa a menudo en la interconexión espectral. En general, la región de HF está compuesta por múltiples interconexiones adyacentes y cada una de estas interconexiones se obtiene de regiones de paso de banda (BP) del espectro de LF por debajo de la frecuencia de cruce dada. Dentro de una representación de banco de filtros de la señal, dichos sistemas copian un conjunto de coeficientes de subbandas adyacentes del espectro de LF en la región objetivo. Los límites de los conjuntos seleccionados suelen depender del sistema y no dependen de la señal. Para algunos contenidos de señal, esta selección de interconexión estática puede conducir a un timbre desagradable y la coloración de la señal reconstruida.

Otros enfoques transfieren la señal de LF a la HF a través de una modulación de banda lateral única (SSB) adaptativa de la señal. Dichos enfoques son de alta complejidad computacional en comparación con [1] ya que funcionan a una alta tasa de muestreo en muestras de dominio de tiempo. Además, la interconexión puede volverse inestable, especialmente para señales no tonales (por ejemplo, habla no vocalizada) y, por lo tanto, la interconexión adaptativa de señal del estado de la técnica puede introducir alteraciones en la señal.

El enfoque de la invención se denomina relleno inteligente de huecos (IGF) y, en su configuración preferida, se aplica en un sistema de BWE basado en una transformada de tiempo-frecuencia como, por ejemplo, la transformada de coseno discreta modificada (MDCT). Sin embargo, las enseñanzas de la invención pueden aplicarse de manera general, por ejemplo, de manera análoga dentro de un sistema basado en banco de filtros espejo en cuadratura (QMF).

Una ventaja de la configuración de IGF basada en MDCT es la integración perfecta en codificadores de audio basados en MDCT, por ejemplo la codificación avanzada de audio (AAC) de MPEG. Compartir la misma transformada para la codificación de audio de forma de onda y para BWE reduce significativamente la complejidad computacional global para el códec de audio.

Además, la invención proporciona una solución para los problemas inherentes de estabilidad que se encuentran en los esquemas de interconexión adaptativa del estado de la técnica.

El sistema propuesto se basa en la observación de que para algunas señales, una selección de interconexión sin guía puede conducir a cambios de timbre y coloraciones en la señal. Si una señal es tonal en la región espectral de origen (SSR) pero es de tipo ruido en la región espectral objetivo (STR), la interconexión de la STR tipo ruido por la

SSR tonal puede conducir a un timbre antinatural. El timbre de la señal también puede cambiar ya que la estructura tonal de la señal puede desalinearse o incluso destruirse por el proceso de interconexión.

El sistema de IGF propuesto lleva a cabo una selección inteligente de recuadros usando la correlación cruzada como medida de similitud entre una SSR en particular y una STR específica. La correlación cruzada de dos señales proporciona una medida de similitud de esas señales y también el retardo de correlación máxima y su signo. Por lo tanto, el enfoque de una selección de recuadros basada en la correlación también se puede usar para ajustar con precisión el desplazamiento espectral del espectro copiado para llegar a ser lo más próximo posible a la estructura espectral original.

La contribución fundamental del sistema propuesto es la elección de una medida de similitud adecuada, y también técnicas para estabilizar el proceso de selección de recuadros. La técnica propuesta proporciona un equilibrio óptimo entre la adaptación de señal instantánea y, al mismo tiempo, la estabilidad temporal. La provisión de estabilidad temporal es especialmente importante para señales que tienen poca similitud de SSR y STR y que, por lo tanto, muestran valores bajos de correlación cruzada o si se emplean medidas de similitud que son ambiguas. En dichos casos, la estabilización impide el comportamiento pseudoaleatorio de la selección adaptativa de recuadros.

Por ejemplo, una clase de señales que a menudo plantea problemas para la BWE del estado de la técnica se caracteriza por una concentración clara de energía en regiones espectrales arbitrarias, tal como se muestra en la figura 12a (izquierda). Aunque hay métodos disponibles para ajustar la envolvente espectral y la tonalidad del espectro reconstruido en la región objetivo, para algunas señales, estos métodos no son capaces de conservar correctamente el timbre como se muestra en la figura 12a (derecha). En el ejemplo mostrado en la figura 12a, la magnitud del espectro en la región objetivo de la señal original por encima de una denominada frecuencia de cruce f_{xover} (figura 12a, izquierda) disminuye casi linealmente. Por el contrario, en el espectro reconstruido (figura 12a, derecha) hay un conjunto claro de valles y picos que se percibe como un artefacto de coloración del timbre.

Una etapa importante del nuevo enfoque es definir un conjunto de recuadros entre los que puede tener lugar la elección basada en similitud posterior. En primer lugar, los límites de recuadros, tanto de la región de origen como de la región objetivo, tienen que definirse unos con respecto a otros. Por lo tanto, la región objetivo entre la frecuencia de inicio de IGF del codificador principal $f_{IGFstart}$ y la frecuencia más alta disponible $f_{IGFstop}$ se divide en un número entero arbitrario n_{Tar} de recuadros, teniendo cada uno de ellos un tamaño individual predefinido. Entonces, para cada recuadro objetivo $tar[idx_tar]$ se genera un conjunto de recuadros de origen de igual tamaño $src[idx_src]$. Mediante esto, se determina el grado básico de libertad del sistema de IGF. El número total de recuadros de origen n_{Src} está determinado por el ancho de banda de la región de origen,

$$bw_{src} = (f_{IGFstart} - f_{IGFmin})$$

donde f_{IGFmin} es la frecuencia más baja disponible para la selección de recuadros de manera que un número entero n_{Src} de recuadros de origen caben en bw_{src} . El número mínimo de recuadros de origen es 0.

Para aumentar aún más el grado de libertad para la selección y el ajuste, se puede definir que los recuadros de origen se solapan entre sí por un factor de solapamiento de entre 0 y 1, donde 0 significa ningún solapamiento y 1 significa el 100% de solapamiento. El caso del 100% de solapamiento implica que sólo está disponible uno o ningún recuadro de origen.

La figura 12b muestra un ejemplo de límites de recuadros de un conjunto de recuadros. En este caso, todos los recuadros objetivo están correlacionados con cada uno de los recuadros de origen. En este ejemplo, los recuadros de origen se solapan en un 50%.

Para un recuadro objetivo, la correlación cruzada se calcula con varios recuadros de origen en retardos de hasta $xcorr_maxLag$ intervalos. Para un recuadro objetivo dado idx_tar y un recuadro de origen idx_src , $xcorr_val[idx_tar][idx_src]$ proporciona el valor máximo de la correlación cruzada absoluta entre los recuadros, mientras que $xcorr_lag[idx_tar][idx_src]$ proporciona el retardo en el que se produce este máximo y $xcorr_sign[idx_tar][idx_src]$ proporciona el signo de la correlación cruzada en $xcorr_lag[idx_tar][idx_src]$.

El parámetro $xcorr_lag$ se usa para controlar la proximidad de la coincidencia entre los recuadros de origen y objetivo. Este parámetro conduce a una reducción de artefactos y ayuda a conservar mejor el timbre y el color de la señal.

En algunas situaciones puede suceder que el tamaño de un recuadro objetivo específico es mayor que el tamaño de los recuadros de origen disponibles. En este caso, el recuadro de origen disponible se repite tan a menudo como sea necesario para rellenar completamente el recuadro objetivo específico. Todavía es posible llevar a cabo la correlación cruzada entre el recuadro objetivo grande y el recuadro de origen más pequeño con el fin de obtener la mejor posición del recuadro de origen en el recuadro objetivo en cuanto al retardo de la correlación cruzada

xcorr_lag y el signo xcorr_sign.

5 La correlación cruzada de los recuadros espectrales sin procesar y la señal original pueden no ser la medida de similitud más adecuada aplicada a espectros de audio con una estructura fuerte de formantes. El blanqueo de un espectro elimina la información de envolvente gruesa y, de ese modo, enfatiza la estructura fina espectral que es de interés principal para la evaluación de la similitud de recuadros. El blanqueo también ayuda en un modelado fácil de la envolvente de la STR en el decodificador para las regiones procesadas por IGF. Por lo tanto, opcionalmente, el recuadro y la señal de origen se blanquean antes de calcular la correlación cruzada.

10 En otras configuraciones, sólo se blanquea el recuadro usando un procedimiento predefinido. Un indicador de "blanqueo" transmitido indica al decodificador que se aplicará el mismo proceso de blanqueo predefinido al recuadro dentro de IGF.

15 Para blanquear la señal, primero se calcula una estimación de envolvente espectral. A continuación, el espectro de MDCT se divide entre la envolvente espectral. La estimación de envolvente espectral se puede estimar en el espectro de MDCT, las energías de espectro de MDCT, el espectro de potencia complejo basado en MDCT o las estimaciones de espectro de potencia. La señal en la que se estima la envolvente se denominará señal de base de aquí en adelante.

20 Las envolventes calculadas sobre espectro de potencia complejo basado en MDCT o estimaciones de espectro de potencia como señal de base tienen la ventaja de no tener fluctuación temporal en componentes tonales.

Si la señal de base está en un dominio de energía, el espectro de MDCT tiene que dividirse entre la raíz cuadrada de la envolvente para blanquear la señal correctamente.

25 Existen diferentes métodos para calcular la envolvente:

- transformando la señal de base con una transformada de coseno discreta (DCT), reteniendo sólo los coeficientes más bajos de DCT (fijando la más alta en cero) y luego calculando una DCT inversa
- calculando una envolvente espectral de un conjunto de coeficientes de predicción lineal (LPC) calculados sobre la trama de audio de dominio de tiempo
- filtrando la señal de base con un filtro de paso bajo

35 Preferentemente, se elige el último enfoque. Para aplicaciones que requieren baja complejidad computacional se puede llevar a cabo alguna simplificación para el blanqueo de un espectro de MDCT: En primer lugar, se calcula la envolvente por medio de un promedio móvil. Esto sólo necesita dos ciclos de procesador por intervalo de MDCT. Entonces, con el fin de evitar el cálculo de la división y la raíz cuadrada, se aproxima la envolvente espectral a 2^n , donde n es el logaritmo de número entero de la envolvente. En este dominio, la operación de raíz cuadrada se convierte simplemente en una operación de desplazamiento y, además, la división entre la envolvente se puede llevar a cabo mediante otra operación de desplazamiento.

45 Después de calcular la correlación de cada recuadro de origen con cada recuadro objetivo, para los nTar recuadros objetivo se selecciona el recuadro de origen con la correlación más alta para sustituirlo. Para coincidir mejor con la estructura espectral original, se usa el retardo de la correlación para modular el espectro replicado por un número entero de intervalos de transformada. En caso de retardos impares, el recuadro se modula adicionalmente a través de la multiplicación por una secuencia temporal alternante de -1/1 para compensar la representación de frecuencia invertida de una banda de cada dos dentro de la MDCT.

50 La figura 12c muestra un ejemplo de una correlación entre un recuadro de origen y un recuadro objetivo. En este ejemplo, el retardo de la correlación es de 5, por lo que el recuadro de origen tiene que modularse por 5 intervalos hacia intervalos de frecuencia más altos en la etapa de copiado del algoritmo de BWE. Adicionalmente, el signo del recuadro tiene que invertirse ya que el valor máximo de correlación es negativo y una modulación adicional como se describió anteriormente representa el retardo impar.

55 Por lo tanto, la cantidad total de información secundaria para transmitir del codificador al decodificador puede consistir en los siguientes datos:

- tileNum[nTar]: índice del recuadro de origen seleccionado por recuadro objetivo
- tileSign[nTar]: signo del recuadro objetivo
- tileMod[nTar]: retardo de la correlación por recuadro objetivo

El recorte y la estabilización de recuadros son una etapa importante en el IGF. Su necesidad y ventajas se explican con un ejemplo, suponiendo una señal de audio tonal estacionaria como, por ejemplo, una nota de flauta de altura de tono estable. La lógica dicta que se introducen menos artefactos si, para una región objetivo dada, siempre se seleccionan recuadros de origen de la misma región de origen a través de las tramas. A pesar de que se supone que la señal es estacionaria, esta condición no se cumplirá correctamente en cada trama ya que la medida de similitud (por ejemplo, correlación) de otra región de origen igualmente similar puede dominar el resultado de la similitud (por ejemplo, correlación cruzada). Esto hace que tileNum[nTar] entre tramas adyacentes dude entre dos o tres opciones muy similares. Este puede ser el origen de un artefacto de tipo ruido musical molesto.

Con el fin de eliminar este tipo de artefactos, el conjunto de recuadros de origen se recortará de manera que los elementos restantes del conjunto de origen sean diferentes al máximo. Esto se logra a través de un conjunto de recuadros de origen

$$S = \{s_1, s_2, \dots, s_n\}$$

de la siguiente manera. Para cualquier recuadro de origen s_i , se correlaciona con todos los demás recuadros de origen, encontrando la mejor correlación entre s_i y s_j y almacenándola en una matriz S_x . Aquí, $S_x[i][j]$ contiene el valor máximo absoluto de correlación cruzada entre s_i y s_j . Sumar la matriz S_x a lo largo de las columnas, proporciona la suma de las correlaciones cruzadas de un recuadro de origen s_i con todos los demás recuadros de origen T .

$$T[i] = S_x[i][1] + S_x[i][2] \dots + S_x[i][n]$$

Aquí, T representa una buena de lo similar que es un recuadro de origen a otros recuadros de origen. Si, para cualquier recuadro de origen i ,

$$T > \text{umbral}$$

el recuadro de origen i puede retirarse del conjunto de posibles orígenes, ya que está muy correlacionado con otros orígenes. El recuadro que tiene la correlación más baja del conjunto de recuadros que cumple la condición en la ecuación 1 se elige como recuadro representativo para este subconjunto. De este modo, se garantiza que los recuadros de origen son máximamente diferentes unos de otros.

El método de recorte de recuadros también implica una memoria del conjunto de recuadros recortados usados en la trama anterior. Los recuadros que estaban activos en la trama anterior también se conservan en la siguiente trama si existen candidatos alternativos para el recorte.

Estando los recuadros s_3 , s_4 y s_5 activos de entre los recuadros $\{s_1, s_2, \dots, s_5\}$ en la trama k , entonces en la trama $k+1$, aunque los recuadros s_1 , s_3 y s_2 compitan para recortarse estando s_3 máximamente correlacionado con los otros, se conserva s_3 ya que fue un recuadro de origen útil en la trama anterior y, por lo tanto, conservarlo en el conjunto de recuadros de origen es beneficioso para implementar continuidad temporal en la selección de recuadros. Este método se aplica preferentemente si la correlación cruzada entre el origen i y el objetivo j , representada como $T_x[i][j]$, es alta

Un método adicional para la estabilización de recuadros es conservar el orden de recuadros de la trama $k-1$ anterior si ninguno de los recuadros de origen en la trama k actual se correlaciona bien con los recuadros objetivo. Esto puede suceder si la correlación cruzada entre el origen i y el objetivo j , representada como $T_x[i][j]$, es muy baja para todos los i, j .

Por ejemplo, si

$$T_x[i][j] < 0,6$$

usándose ahora un umbral provisional, entonces

$$\text{tileNum}[n\text{Tar}]_k = \text{tileNum}[n\text{Tar}]_{k-1}$$

para todos los $n\text{Tar}$ de esta trama k .

Las dos técnicas anteriores reducen en gran medida los artefactos que se producen a partir del cambio rápido de números de recuadros establecidos a través de las tramas. Otra ventaja añadida de este recorte y estabilización de recuadros es que no se necesita enviar información adicional al decodificador y tampoco se necesita un cambio de arquitectura del decodificador. Este recorte de recuadros propuesto es una manera elegante de reducir los posibles artefactos de tipo ruido musical o ruido excesivo en las regiones espectrales de recuadros.

La figura 11a ilustra un decodificador de audio para decodificar una señal de audio codificada. El decodificador de audio comprende un decodificador 1102 (principal) de audio para generar una primera representación decodificada de un primer conjunto de primeras porciones espectrales, teniendo la representación decodificada una primera resolución espectral.

Asimismo, el decodificador de audio comprende un decodificador 1104 paramétrico para generar una segunda representación decodificada de un segundo conjunto de segundas porciones espectrales que tiene una segunda resolución espectral que es más baja que la primera resolución espectral. Además se proporciona un regenerador 1106 de frecuencia que recibe, como primera entrada 1101, primeras porciones espectrales decodificadas y, como una segunda entrada en 1103, la información paramétrica que incluye, para cada recuadro de frecuencia objetivo o banda de reconstrucción objetivo, una información de intervalo de origen. El regenerador 1106 de frecuencia aplica entonces la regeneración de frecuencia usando valores espectrales a partir del intervalo de origen identificado por la información de coincidencia con el fin de generar los datos espectrales para el intervalo objetivo. A continuación, las primeras porciones 1101 espectrales y la salida del regenerador 1107 de frecuencia se introducen ambas en un convertidor 1108 de espectro-tiempo para generar finalmente la señal de audio decodificada.

Preferentemente, el decodificador 1102 de audio es un decodificador de audio de dominio espectral, aunque el decodificador de audio también se puede implementar como cualquier otro decodificador de audio tal como un decodificador de audio de dominio de tiempo o paramétrico.

Como se indica en la figura 11b, el regenerador 1106 de frecuencia puede comprender las funcionalidades del bloque 1120 que ilustra un modulador de recuadros - selector del intervalo de origen para retardos impares, un filtro 1122 blanqueado, cuando se proporciona un indicador 1123 de blanqueo, y, adicionalmente, una envolvente espectral con funcionalidades de ajuste implementadas como se ilustra en el bloque 1128 usando los datos espectrales sin procesar generados por cualquiera del bloque 1120 o el bloque 1122 o la cooperación de ambos bloques. De todos modos, el regenerador 1106 de frecuencia puede comprender un conmutador 1124 reactivo frente a un indicador 1123 de blanqueo recibido. Cuando se establece el indicador de blanqueo, la salida del selector de intervalo de origen/modulador de recuadros para retardos impares se introduce en el filtro 1122 de blanqueo. Sin embargo, entonces el indicador 1123 de blanqueo no se establece para una determinada banda de reconstrucción, entonces se activa una línea 1126 de desvío de manera que la salida del bloque 1120 se proporciona al bloque 1128 de ajuste de envolvente espectral sin ningún blanqueo.

Puede haber más de un nivel de blanqueo (1123) señalado en el flujo de bits y estos niveles pueden señalarse por recuadro. En caso de que haya tres niveles señalados por recuadro, se codificarán de la siguiente manera:

```
bit = readBit(1);
```

```
si (bit == 1) {
```

```
    para (tile_index = 0..nT)
```

```
        /*los mismos niveles que la última trama*/
```

```
        whitening_level[tile_index] = whitening_level_prev_frame[tile_index];
```

```
    } sino {
```

```
        /*primer recuadro:*/
```

```
        tile_index = 0;
```

```
        bit = readBit(1);
```

```
        si (bit == 1) {
```

```
            whitening_level[tile_index] = MID_WHITENING;
```

```
        } sino {
```

```
            bit = readBit(1);
```

```
            si (bit == 1) {
```

```

whitening_level[tile_index] = STRONG_WHITENING;
} sino {
5 whitening_level[tile_index] = OFF; /*sin blanqueo*/
}
}
10 /*recuadros restantes:*/
bit = readBit(1);
15 si (bit == 1) {
/*los niveles de aplanamiento para los recuadros restantes son iguales que para el primero.*/
/*No tienen que leerse bits adicionales*/
20 para (tile_index = 1..nT)
whitening_level[tile_index] = whitening_level[0];
25 } sino {
/*bits leídos para los recuadros restantes como para el primer recuadro*/
para (tile_index = 1..nT) {
30 bit = readBit(1);
si(bit == 1) {
35 whitening_level[tile_index] = MID_WHITENING;
} sino {
bit = readBit(1);
40 si (bit == 1) {
whitening_level[tile_index] = STRONG_WHITENING;
45 } sino {
whitening_level[tile_index] = OFF; /*sin blanqueo*/
}
50 }
}
55 }
}
60 MID_WHITENING y STRONG_WHITENING se refieren a distintos filtros (1122) de blanqueo que pueden diferir en
cuanto a la forma en la que se calcula la envolvente (como se describió anteriormente).

```

El regenerador de frecuencia del lado de decodificador puede controlarse mediante una ID 1121 de intervalo de origen cuando se aplica sólo un esquema de selección de recuadros espectrales grueso. Sin embargo, cuando se aplica un esquema de selección de recuadros espectrales de ajuste preciso, entonces se proporciona además un

retardo 1119 de intervalo de origen. Asimismo, siempre que el cálculo de la correlación proporcione un resultado negativo, entonces, adicionalmente también se puede aplicar un signo de la correlación al bloque 1120 de manera que cada una de las líneas espectrales de datos de página se multiplican por "-1" para representar el signo negativo.

5 Por lo tanto, la presente invención tal como se comenta en las figuras 11a, 11b garantiza la obtención de una calidad óptima de audio debido al hecho de que el mejor intervalo de origen coincidente para un determinado intervalo objetivo o de destino se calcula en el lado de codificador y se aplica en el lado de decodificador.

10 La figura 11c es un determinado codificador de audio para codificar una señal de audio que comprende un convertidor 1130 de tiempo-espectro, un analizador 1132 espectral conectado posteriormente y, adicionalmente, un calculador 1134 de parámetros y un codificador 1136 principal. El codificador 1136 principal emite intervalos de origen codificados y el calculador 1134 de parámetros emite información de coincidencia para intervalos objetivo.

15 Los intervalos de origen codificados se transmiten a un decodificador junto con información de coincidencia para los intervalos objetivo de manera que el decodificador ilustrado en la figura 11a se encuentra en posición de llevar a cabo una regeneración de frecuencia.

20 El calculador 1134 de parámetros está configurada para calcular similitudes entre primeras porciones espectrales y segundas porciones espectrales y para determinar, basándose en las similitudes calculadas, para una segunda porción espectral, una primera porción espectral coincidente que coincide con la segunda porción espectral. Preferentemente, los resultados de coincidencia para diferentes intervalos de origen e intervalos objetivo, como se ilustra en las figuras 12a, 12b para determinar un par de coincidencia seleccionado que comprende la segunda porción espectral, y el calculador de parámetros está configurada para proporcionar esta información de coincidencia que identifica el par de coincidencia en una señal de audio codificada. Preferentemente, el calculador 1134 de parámetros de la presente invención está configurado para usar regiones objetivo predefinidas en el segundo conjunto de segundas porciones espectrales o regiones de origen predefinidas en el primer conjunto de primeras porciones espectrales como se ilustra, por ejemplo, en la figura 12b. Preferentemente, las regiones objetivo predefinidas no se solapan o las regiones de origen predefinidas se solapan. Cuando las regiones de origen predefinidas son un subconjunto del primer conjunto de primeras porciones espectrales por debajo de una frecuencia 309 de inicio de relleno de huecos de la figura 3a, y preferentemente, la región objetivo predefinida que abarca una región espectral inferior coincide, con su límite de frecuencia inferior, con la frecuencia de inicio de relleno de huecos de manera que cualquier intervalo objetivo se encuentra por encima de la frecuencia de inicio de relleno de huecos y los intervalos de origen se encuentran por debajo de la frecuencia de inicio de relleno de huecos.

35 Como se comentó, se obtiene una granularidad fina comparando una región objetivo con una región de origen sin ningún retardo con respecto a la región de origen y la misma región de origen, pero con un determinado retardo. Estos retardos se aplican en el calculador 1140 de correlación cruzada de la figura 11d y la selección de pares de coincidencia se lleva a cabo finalmente por el selector 1144 de recuadros.

40 Además, se prefiere llevar a cabo un blanqueo de intervalos de origen y/u objetivo como se ilustra en el bloque 1142. Entonces, este bloque 1142 proporciona un indicador de blanqueo al flujo de bits que se usa para controlar el conmutador 1123 de lado de decodificador de la figura 11b. Asimismo, si el calculador 1140 de correlación cruzada proporciona un resultado negativo, entonces este resultado negativo también se señala a un decodificador. Por lo tanto, en una realización preferida, el selector de recuadros emite una ID de intervalo de origen para un intervalo objetivo, un retardo, un signo y el bloque 1142 proporciona además un indicador de blanqueo.

50 Asimismo, el calculador de parámetros 1134 está configurado para llevar a cabo un recorte 1146 de recuadros de origen reduciendo el número de posibles intervalos de origen ya que se omite una interconexión de origen de un conjunto de posibles recuadros de origen basándose en un umbral de similitud. Por lo tanto, cuando dos recuadros de origen son más similares o igual de similares que un umbral de similitud, entonces uno de estos dos recuadros de origen se retira del conjunto de posibles orígenes y el recuadro de origen retirado ya no se usa para el procesamiento adicional y, específicamente, no puede seleccionarse por el selector 1144 de recuadros o no se usa para el cálculo de correlación cruzada entre diferentes intervalos de origen e intervalos objetivo como se lleva a cabo en el bloque 1140.

60 Se han descrito diferentes implementaciones con respecto a diferentes figuras. Las figuras 1a-5c se refieren a un esquema de codificador/decodificador de ancho de banda completo o de tasa completa. Las figuras 6a-7e se refieren a un esquema de codificador/decodificador con procesamiento de TNS o TTS. Las figuras 8a-8e se refieren a un esquema de codificador/decodificador con procesamiento específico de dos canales. Las figuras 9a-10d se refieren a un cálculo y aplicación de información de energía específica, y las figuras 11a-12c se refieren a un modo específico de selección de recuadros.

Todos estos aspectos diferentes pueden ser de uso inventivo independientemente unos de otros, pero,

adicionalmente, también se pueden aplicar juntos como se ilustra básicamente en las figuras 2a y 2b. Sin embargo, el procesamiento específico de dos canales se puede aplicar también a un esquema de codificador/decodificador ilustrado en las figuras 13a-13b, y lo mismo es cierto para el procesamiento de TNS/TTS, el cálculo y aplicación de información de energía de envolvente en la banda de reconstrucción o la identificación de intervalo de origen adaptativa y la aplicación correspondiente en el lado de decodificador. Por otro lado, el aspecto de tasa completa se puede aplicar con o sin procesamiento de TNS/TTS, con o sin procesamiento de dos canales, con o sin una identificación del intervalo de origen adaptativa o con otros tipos de cálculos de energía para la representación de envolvente espectral. Por lo tanto, es evidente que características de uno de estos aspectos individuales se pueden aplicar también en otros aspectos.

Aunque algunos aspectos se han descrito en el contexto de un aparato para codificar o decodificar, es evidente que estos aspectos también representan una descripción del método correspondiente, en el que un bloque o dispositivo corresponde a una etapa de método o a una característica de una etapa de método. De forma análoga, aspectos descritos en el contexto de una etapa de método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato respectivo. Algunas o todas las etapas de método se pueden llevar a cabo por (o usando) un aparato de hardware tal como, por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, alguna o más de las etapas de método más importantes se pueden llevar a cabo por dicho aparato.

Dependiendo de determinados requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede llevar a cabo usando un medio de almacenamiento no transitorio tal como un medio de almacenamiento digital, por ejemplo un disco flexible, un disco duro (HDD), un DVD, un Blu-Ray, un CD, una ROM, una PROM, y una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control legibles de manera electrónica almacenadas en el mismo, que cooperan (o son capaces de cooperar) con un sistema informático programable de forma que se lleva a cabo el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles de manera electrónica, las cuales son capaces de cooperar con un sistema informático programable, de tal manera que se lleva a cabo uno de los métodos descritos en el presente documento.

En general, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, siendo el código de programa operativo para llevar a cabo uno de los métodos cuando se ejecuta el producto de programa informático en un ordenador. El código del programa se puede almacenar, por ejemplo, en un soporte legible por ordenador.

Otras realizaciones comprenden el programa informático para llevar a cabo uno de los métodos descritos en el presente documento, almacenado en un soporte legible por ordenador.

Dicho de otro modo, una realización del método de la invención es, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los métodos descritos en el presente documento, cuando se ejecuta el programa informático en un ordenador.

Por lo tanto, otra realización del método de la invención es un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El soporte de datos, el medio de almacenamiento digital o el medio grabado son generalmente tangibles y/o no transitorios.

Por lo tanto, una realización adicional de la invención es un flujo de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales puede estar configurado, por ejemplo, para transferirse a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los métodos descritos en el presente documento.

Otra realización comprende un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los métodos descritos en el presente documento.

Otra realización según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, por vía electrónica u óptica) un programa informático para llevar a cabo uno de los métodos descritos en el presente documento a 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.

En algunas realizaciones, se puede usar un dispositivo lógico programable (por ejemplo, una matriz de puerta programable en el campo) para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puerta programable en el campo puede cooperar con un microprocesador para llevar a cabo uno de los métodos descritos en el presente documento. En general, los métodos se llevan a cabo, preferentemente, por cualquier aparato de hardware.

Las realizaciones anteriormente descritas son simplemente ilustrativas de los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento resultarán evidentes para otros expertos en la materia. Por lo tanto, se pretende limitarse solamente por el alcance de las reivindicaciones de patente a continuación y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

Lista de referencias

[1] Dietz, L. Liljeryd, K. Kjörling y O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112th AES Convention, Munich, mayo de 2002.

[2] Ferreira, D. Sinha, "Accurate Spectral Replacement", Audio Engineering Society Convention, Barcelona, España 2005.

[3] D. Sinha, A. Ferreira¹ y E. Harinarayanan, "A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)", Audio Engineering Society Convention, Paris, Francia 2006.

[4] R. Annadana, E. Harinarayanan, A. Ferreira y D. Sinha, "New Results in Low Bit Rate Speech Coding and Bandwidth Extension", Audio Engineering Society Convention, San Francisco, EE.UU. 2006.

[5] T. Zernicki, M. Bartkowiak, "Audio bandwidth extension by frequency scaling of sinusoidal partials", Audio Engineering Society Convention, San Francisco, EE.UU. 2008.

[6] J. Herre, D. Schulz, Extending the MPEG-4 AAC Codec by Perceptual Noise Substitution, 104th AES Convention, Amsterdam, 1998, Preprint 4720.

[7] M. Neuendorf, M. Multrus, N. Rettelbach, *et al.*, MPEG Unified Speech and Audio Coding-The ISO/MPEG Standard for High-Efficiency Audio Coding of all Content Types, 132nd AES Convention, Budapest, Hungría, abril, 2012.

[8] McAulay, Robert J., Quatieri, Thomas F. "Speech Analysis/Synthesis Based on a Sinusoidal Representation". IEEE Transactions on Acoustics, Speech, And Signal Processing, Vol 34(4), agosto de 1986.

[9] Smith, J.O., Serra, X. "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation", Proceedings of the International Computer Music Conference, 1987.

[10] Purnhagen, H.; Meine, Nikolaus, "HILN-the MPEG-4 parametric audio coding tools," Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on, vol. 3, n.º, págs. 201,204 vol.3, 2000

[11] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio", Ginebra, 1998.

[12] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Oikawa: "MPEG-2 Advanced Audio Coding", 101st AES Convention, Los Angeles 1996

[13] J. Herre, "Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction", 17th AES International Conference on High Quality Audio Coding, agosto de 1999

[14] J. Herre, "Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction", 17th AES International Conference on High Quality Audio Coding, agosto de 1999

[15] International Standard ISO/IEC 23001-3:2010, Unified speech and audio coding Audio, Ginebra, 2010.

[16] International Standard ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 3: Audio, Ginebra, 2005.

[17] P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication", in Proceedings of 1st IEEE Benelux Workshop on MPCA, Leuven, noviembre de 2002

[18] F. Nagel, S. Disch, S. Wilde, A continuous modulated single sideband bandwidth extension, ICASSP 5 International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas (EE.UU.), abril 2010

REIVINDICACIONES

1. Aparato para generar una señal de dos canales decodificada, que comprende:

5 un procesador (802) de audio para decodificar una señal de dos canales codificada para obtener un primer conjunto de primeras porciones espectrales;

10 un decodificador (804) paramétrico para proporcionar datos paramétricos para un segundo conjunto de segundas porciones espectrales y una identificación de dos canales que identifica ya sea una primera o una segunda representación de dos canales diferente para las segundas porciones espectrales; y

15 un regenerador (806) de frecuencia para regenerar una segunda porción espectral dependiendo de una primera porción espectral del primer conjunto de primeras porciones espectrales, los datos paramétricos para la segunda porción y la identificación de dos canales para la segunda porción,

20 en el que el procesador de audio está configurado para decodificar el primer conjunto de primeras porciones espectrales según una identificación de dos canales adicional para el primer conjunto de primeras porciones espectrales y para transformar (832) el primer conjunto de primeras porciones espectrales de modo que se obtienen una primera representación de dos canales del primer conjunto de primeras porciones espectrales y una segunda representación de dos canales del primer conjunto de primeras porciones espectrales, y

25 en el que el regenerador (806) de frecuencia está configurado para usar ya sea la primera representación de dos canales o la segunda representación de dos canales del primer conjunto de primeras porciones espectrales según se indica en la identificación de dos canales para la segunda porción.

2. Aparato según la reivindicación 1, en el que la identificación de dos canales identifica ya sea un procesamiento independiente de dos canales de la señal de dos canales o un procesamiento conjunto de los dos canales de la señal de dos canales y

30 en el que el regenerador (806) de frecuencia está configurado para regenerar la segunda porción espectral para un primer canal de los dos canales y la segunda porción espectral para un segundo canal de los dos canales usando la primera porción del primer canal y la primera porción del segundo canal, en el que la primera porción del primer canal y la primera porción del segundo canal están en una representación de dos canales identificada por la identificación de dos canales para la segunda porción espectral.

3. Aparato según la reivindicación 1 ó 2,

40 en el que la identificación de dos canales identifica ya sea un procesamiento independiente de dos canales de la señal de dos canales o un procesamiento conjunto de los dos canales de la señal de dos canales, y

45 en el que el regenerador (806) de frecuencia está configurado para regenerar una representación conjunta de los dos canales en la segunda porción espectral según se identifica por la identificación de dos canales, y

en el que el regenerador (806) de frecuencia comprende además un transformador (842) de representación para transformar la representación conjunta de la segunda porción espectral en una representación independiente para la segunda porción espectral.

50 4. Aparato según la reivindicación 3, en el que el transformador de representación usa parámetros de representación conjunta adicionales para la transformación de representación.

55 5. Aparato según la reivindicación 3, en el que la representación conjunta es una representación central/lateral, y en el que el transformador de representación está configurado para funcionar basándose en la siguiente ecuación:

$$\text{leftTile} = 0,5 * (\text{midTile} + \text{sideTile})$$

$$\text{rightTile} = 0,5 * (\text{midTile} - \text{sideTile})$$

60 en el que leftTile (recuadro izquierdo) y rightTile (recuadro derecho) son la representación independiente para la segunda porción espectral, y en el que midTile (recuadro central) y sideTile (recuadro lateral) son la representación conjunta para la segunda porción espectral.

6. Aparato según la reivindicación 3, en el que la representación conjunta es una representación que comprende un coeficiente de predicción adicional, y en el que el transformador de representación está configurado para funcionar basándose en la siguiente ecuación:

$$\text{sideTile}[k] = \text{sideTile}[k] - \text{predictionCoeff} \cdot \text{midTile}[k]$$

$$\text{leftTile}[k] = \text{midTile}[k] + \text{sideTile}[k]$$

$$\text{rightTile}[k] = \text{midTile}[k] - \text{sideTile}[k]$$

10 cuando un sentido de predicción es del centro al lado, o

$$\text{midTile1}[k] = \text{midTile}[k] - \text{predictionCoeff} \cdot \text{sideTile}[k]$$

$$\text{leftTile}[k] = \text{midTile1}[k] - \text{sideTile}[k]$$

$$\text{rightTile}[k] = \text{midTile1}[k] + \text{sideTile}[k]$$

20 cuando un sentido de predicción es del lado al centro,

en el que leftTile (recuadro izquierdo) y rightTile (recuadro derecho) son la representación independiente para la segunda porción espectral, y en el que midTile (recuadro central) y sideTile (recuadro lateral) son la representación conjunta para la segunda porción espectral, y en el que predictionCoeff (coeficiente de predicción) es el coeficiente de predicción adicional.

7. Aparato según una de las reivindicaciones anteriores,

en el que los datos paramétricos para el segundo conjunto de segundas porciones espectrales se proporcionan por separado para cada canal de la representación de dos canales, y

en el que el regenerador (806) de frecuencia está configurado para transformar (840) los datos paramétricos para la segunda porción espectral en una representación conjunta para la segunda porción espectral y para aplicar los datos paramétricos a una representación conjunta de la primera porción espectral, cuando la identificación de dos canales identifica la representación conjunta para la segunda porción espectral.

8. Aparato según una de las reivindicaciones anteriores, en el que las segundas porciones espectrales corresponden a las bandas de frecuencia, y en el que la identificación de dos canales es una matriz de indicadores, un indicador para cada banda de frecuencia, y en el que el decodificador (804) paramétrico está configurado para verificar si el indicador está establecido o no y para controlar la regeneración de frecuencia según el indicador para usar ya sea una primera representación de dos canales o una segunda representación de dos canales de la primera porción espectral de la señal de dos canales.

9. Aparato según una cualquiera de las reivindicaciones anteriores, en el que el decodificador (804) paramétrico está configurado para proporcionar la identificación de dos canales adicional para el primer conjunto de primeras porciones espectrales que indica ya sea una primera o una segunda representación de dos canales diferente para la primera porción espectral, y

en el que el aparato está configurado para decodificar la segunda representación de dos canales según se indica por la identificación de dos canales para la primera porción espectral, y

en el que el regenerador (806) de frecuencia está configurado para transformar (832) la segunda representación de dos canales en la primera representación de dos canales tras la decodificación (830) principal.

10. Aparato según una cualquiera de las reivindicaciones anteriores, que comprende además un combinador (846) para combinar el primer conjunto de primeras porciones espectrales generado por el procesador (802) de audio y la segunda porción espectral reconstruida generada por el regenerador (806) de frecuencia para obtener la señal de dos canales decodificada.

11. Aparato según una de las reivindicaciones anteriores,

en el que el decodificador paramétrico está configurado para proporcionar adicionalmente, para la segunda porción espectral, una identificación (1121) de banda de origen que indica una primera porción espectral

específica que se va a usar para regenerar la segunda porción espectral, y

en el que el regenerador (806) de frecuencia está configurado para regenerar la segunda porción espectral usando la primera porción espectral identificada por la identificación (833) de banda de origen.

5 12. Aparato según una de las reivindicaciones anteriores,

10 en el que el regenerador (806) de frecuencia comprende un transformador (832) de representación para proporcionar la primera y la segunda representación de dos canales de un primer conjunto de primeras porciones espectrales generado por el procesador (830) de audio,

15 en el que el regenerador (806) de frecuencia comprende además un generador (836) de recuadro de frecuencia para generar datos sin procesar para cada canal de la representación de canal identificada por la identificación de dos canales y usar una identificación (833) de intervalo de origen que indica las primeras porciones espectrales que se van a usar para generar los datos sin procesar,

20 en el que el regenerador (806) de frecuencia comprende además un transformador (840) de parámetros para transformar parámetros proporcionados en una primera representación de dos canales para dar una segunda representación de dos canales para los parámetros, cuando los datos sin procesar para cada canal se proporcionan en la segunda representación de dos canales por el generador (836) de recuadro de frecuencia,

25 en el que el regenerador (806) de frecuencia comprende además un ajustador (838) de envolvente para ajustar una envolvente de cada canal de la representación de dos canales, siendo la representación de dos canales la segunda representación de dos canales,

30 en el que el regenerador (806) de frecuencia comprende además un transformador (842) de representación para transformar la representación de dos canales de valores espectrales en la segunda porción espectral para dar la primera representación de dos canales,

en el que el aparato comprende además un convertidor (846) de frecuencia-tiempo para convertir una representación generada por el transformador (842) de representación de un dominio espectral a un dominio de tiempo.

35 13. Método para generar una señal de dos canales decodificada, que comprende:

decodificar (802) una señal de dos canales codificada para obtener un primer conjunto de primeras porciones espectrales;

40 proporcionar (804) datos paramétricos para un segundo conjunto de segundas porciones espectrales y una identificación de dos canales que identifica ya sea una primera o una segunda representación de dos canales diferente para las segundas porciones espectrales; y

45 regenerar (806) una segunda porción espectral dependiendo de una primera porción espectral del primer conjunto de primeras porciones espectrales, los datos paramétricos para la segunda porción y la identificación de dos canales para la segunda porción,

50 en el que la decodificación (802) comprende decodificar el primer conjunto de primeras porciones espectrales según una identificación de dos canales adicional para el primer conjunto de primeras porciones espectrales y transformar (832) el primer conjunto de primeras porciones espectrales de modo que se obtienen una primera representación de dos canales del primer conjunto de primeras porciones espectrales y una segunda representación de dos canales del primer conjunto de primeras porciones espectrales, y

55 en el que la regeneración (806) comprende usar ya sea la primera representación de dos canales o la segunda representación de dos canales del primer conjunto de primeras porciones espectrales según se indica en la identificación de dos canales para la segunda porción.

60 14. Programa informático para llevar a cabo, cuando se ejecuta en un ordenador o procesador, el método según la reivindicación 13.

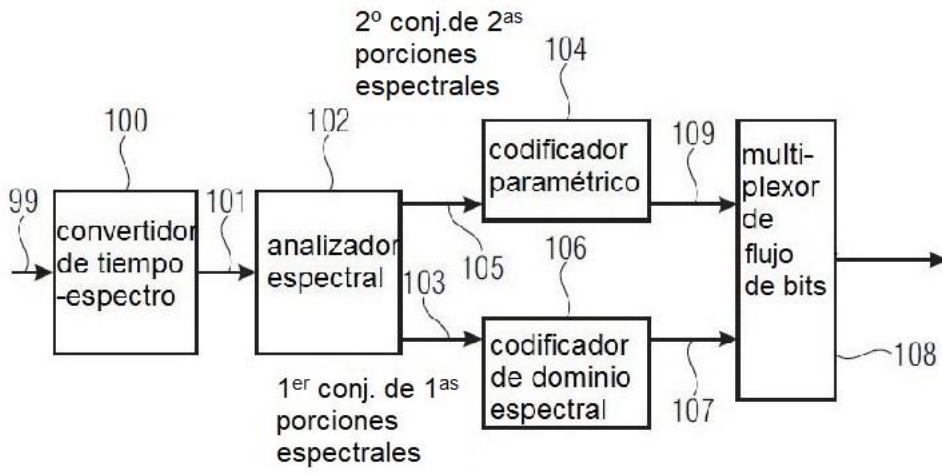


FIG 1A

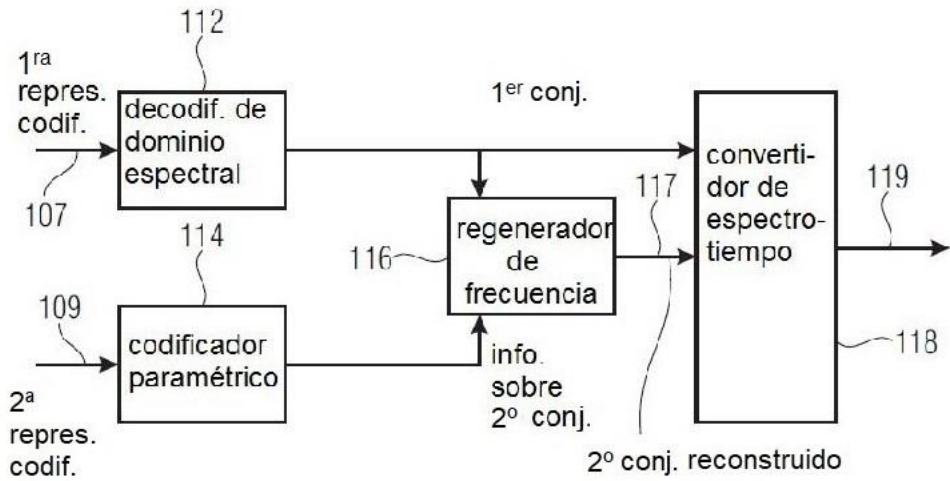


FIG 1B

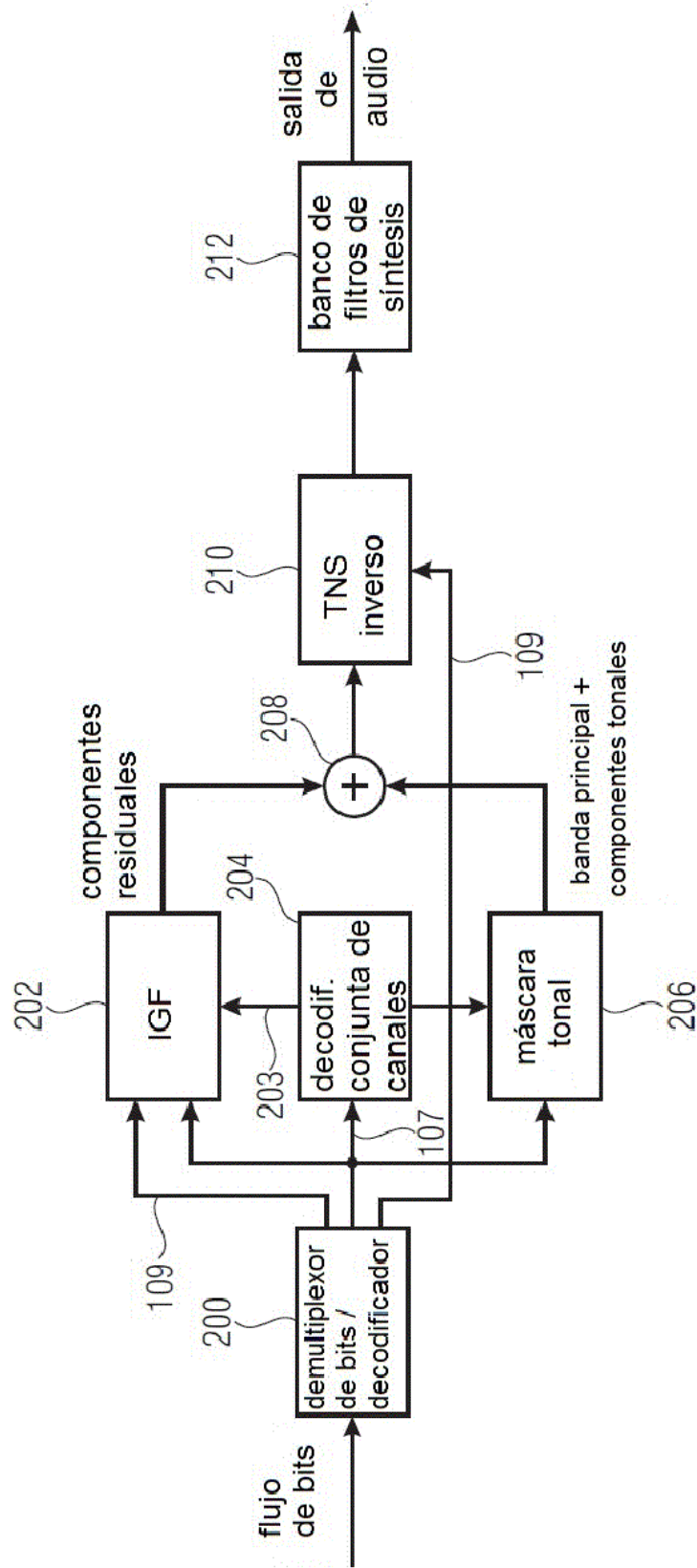


FIG 2A

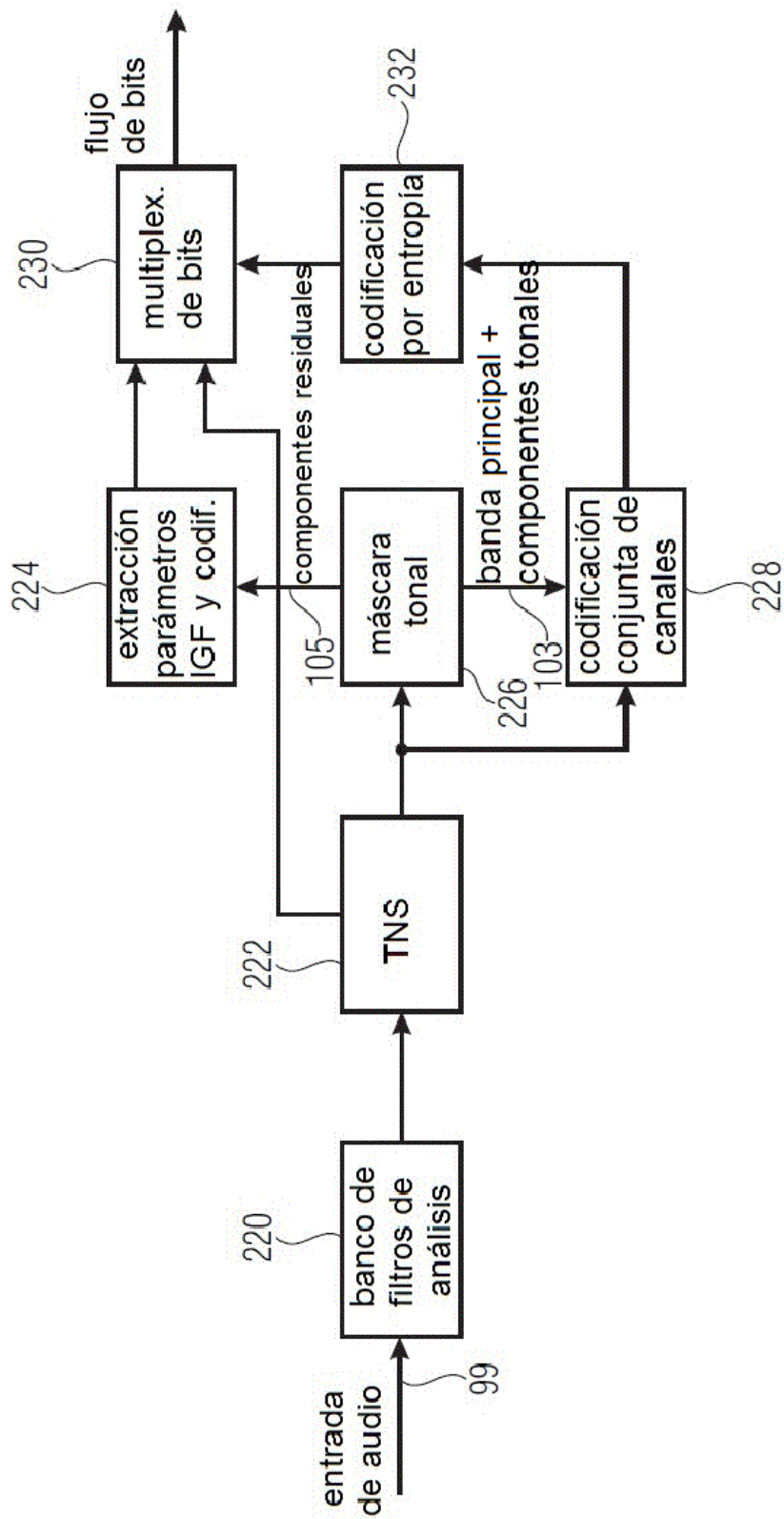


FIG 2B

- 1^{ra} resolución (alta resolución) para la "envolvente" del 1^{er} conjunto (codificación por líneas);
- 2^a resolución (baja resolución) para la "envolvente" del 2^o conjunto (factor de escala por SCB);

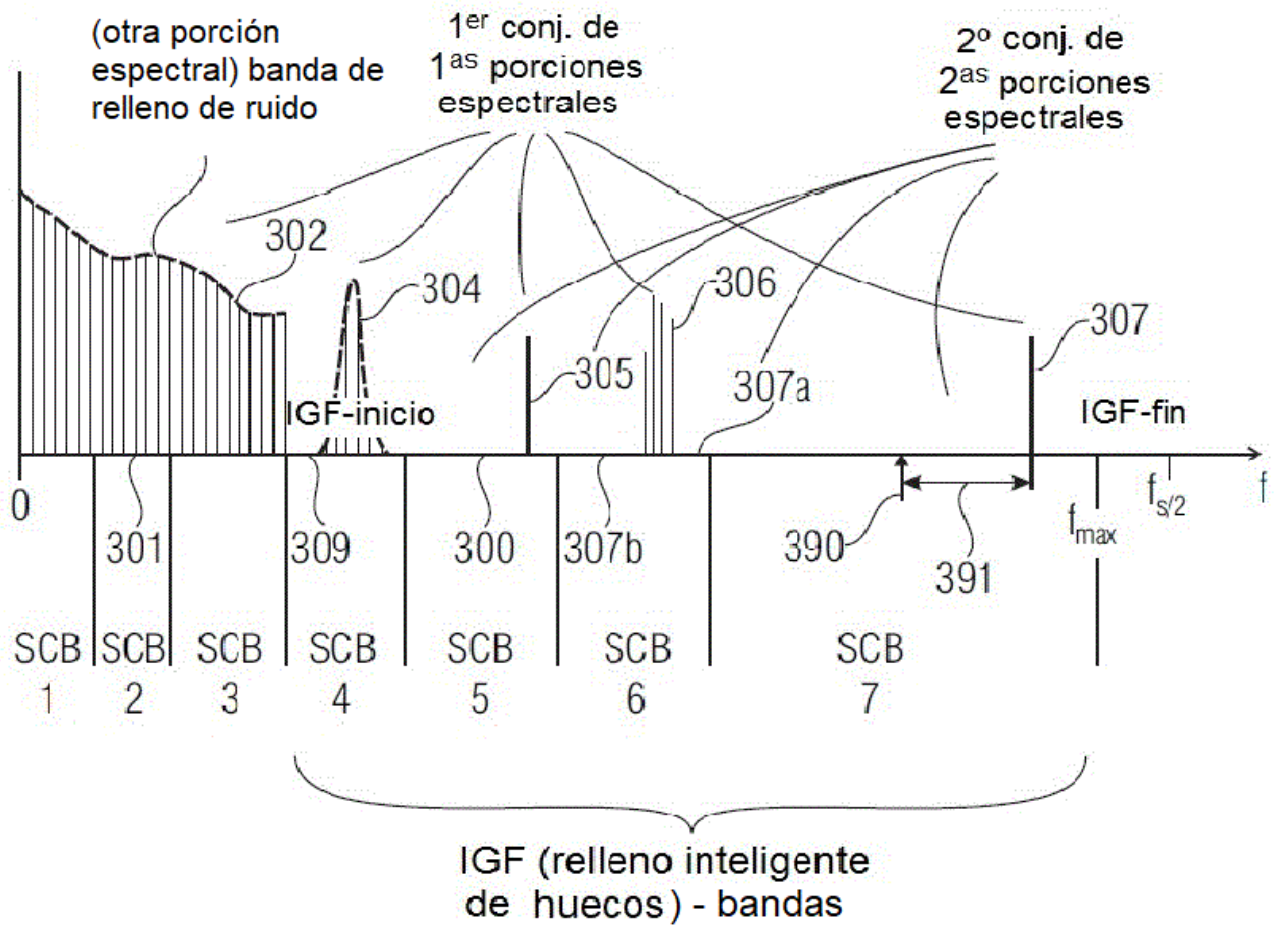


FIG 3A

SCB1	SCB2	SCB3	SCB4	SCB5	SCB6	SCB7
SF1	SF2	SF3	SF4	SF5	SF6	SF7
			E ₁	E ₂	E ₃	E ₄
	NF ₂					

308
310
312

FIG 3B

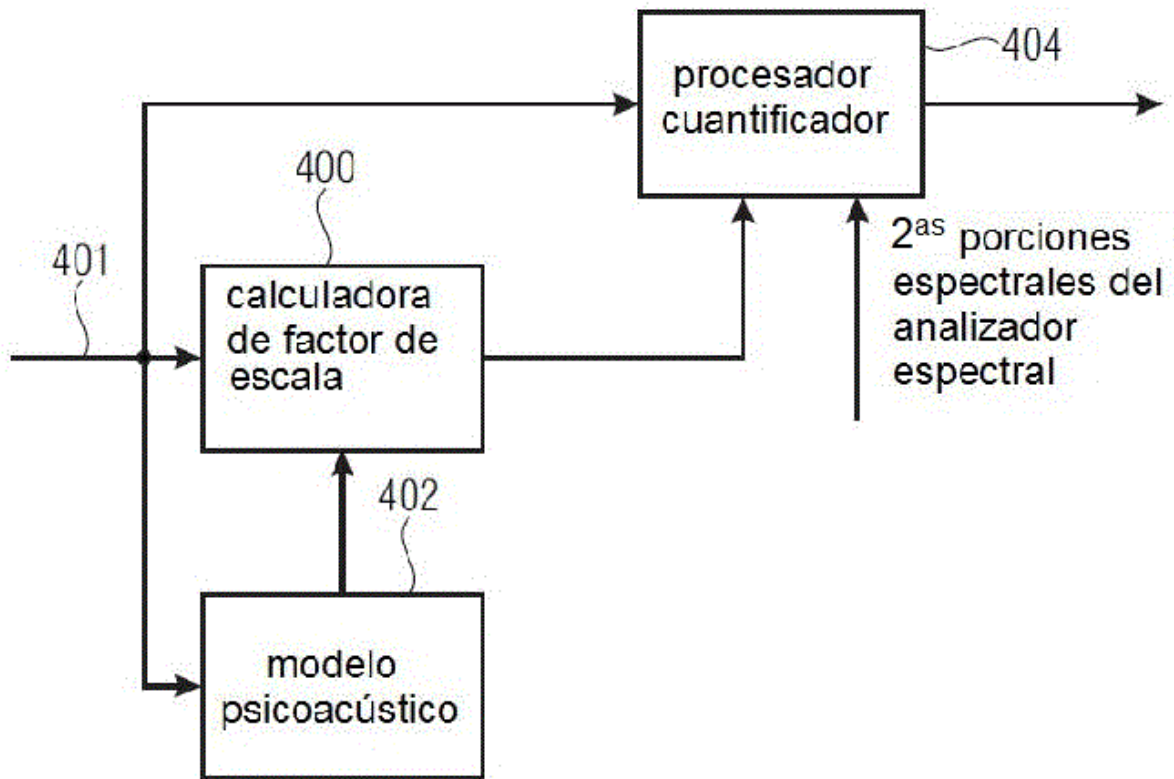


FIG 4A

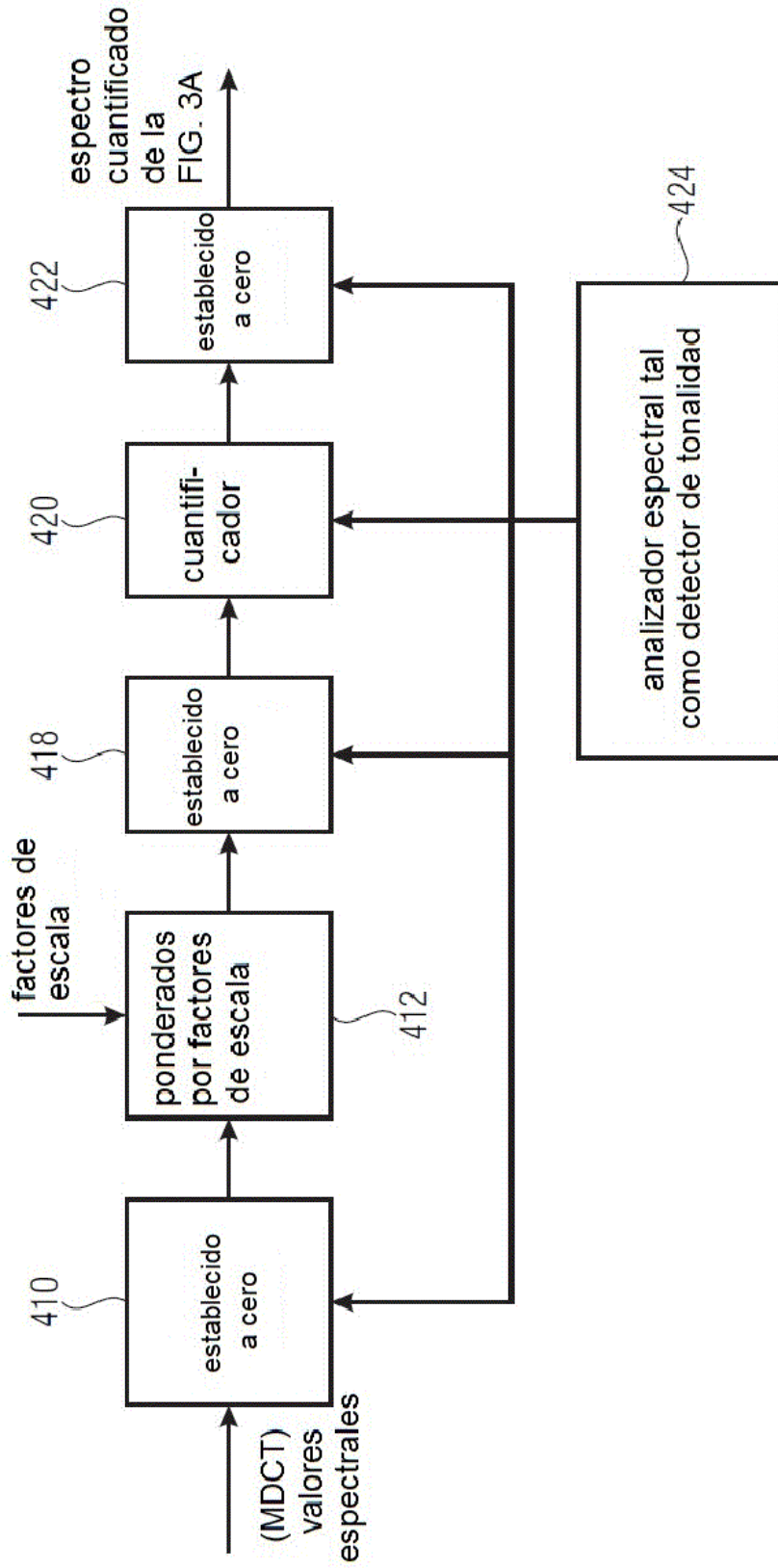


FIG 4B
(PROCESADOR CUANTIFICADOR)

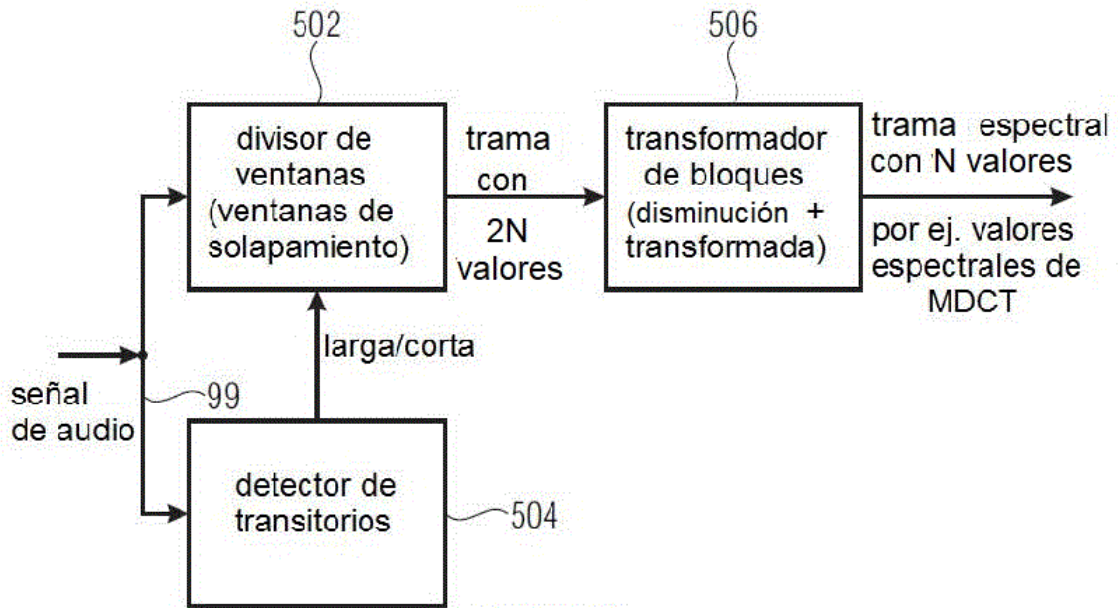


FIG 5A
(OTRAS PORCIONES ESPECTRALES)

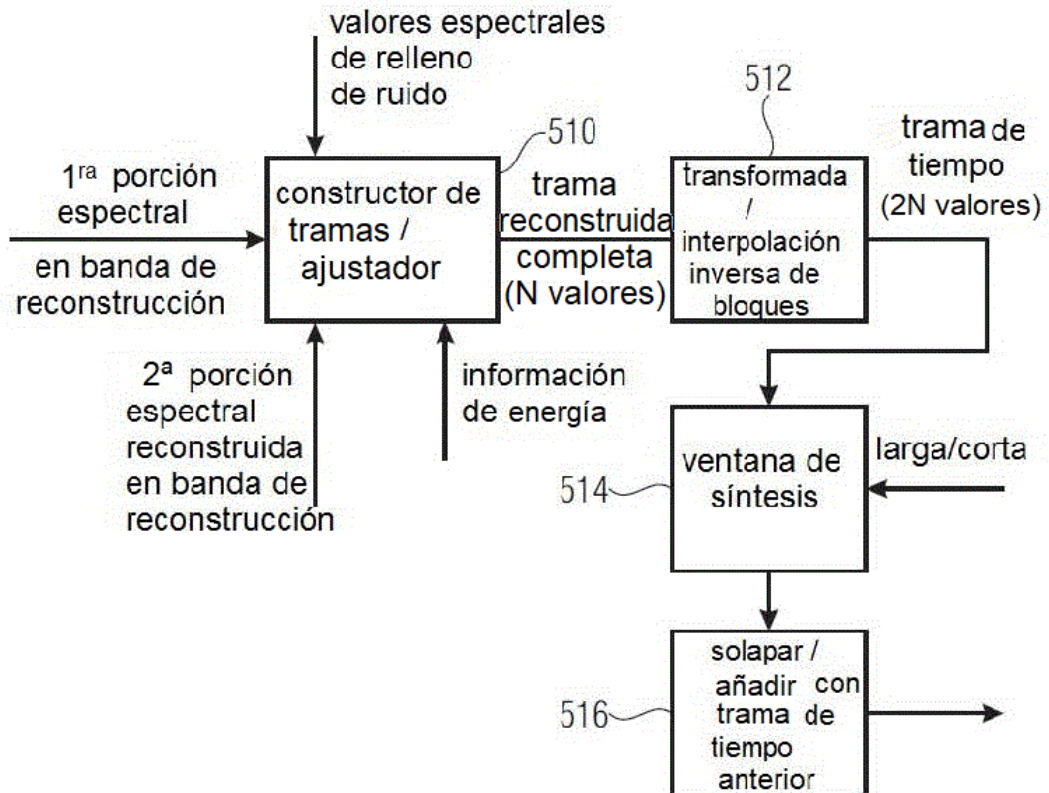


FIG 5B

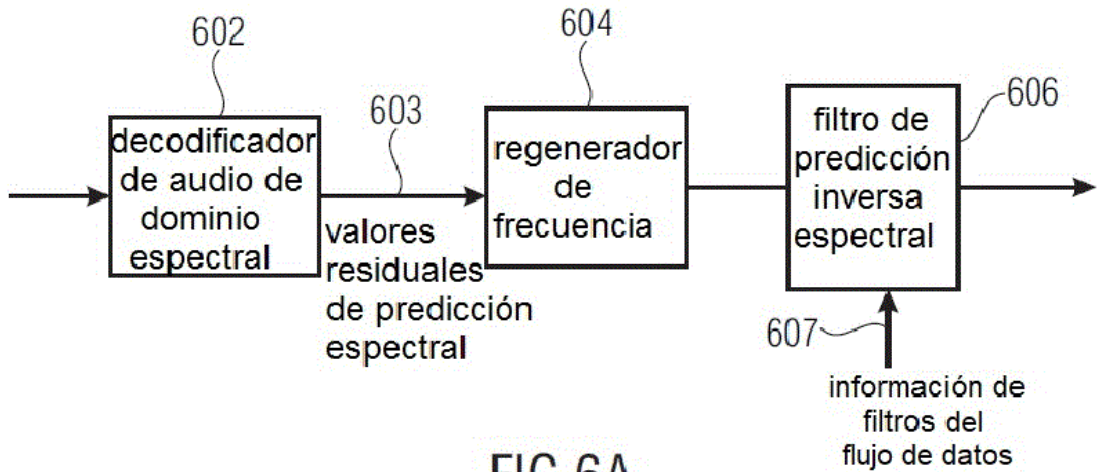


FIG 6A

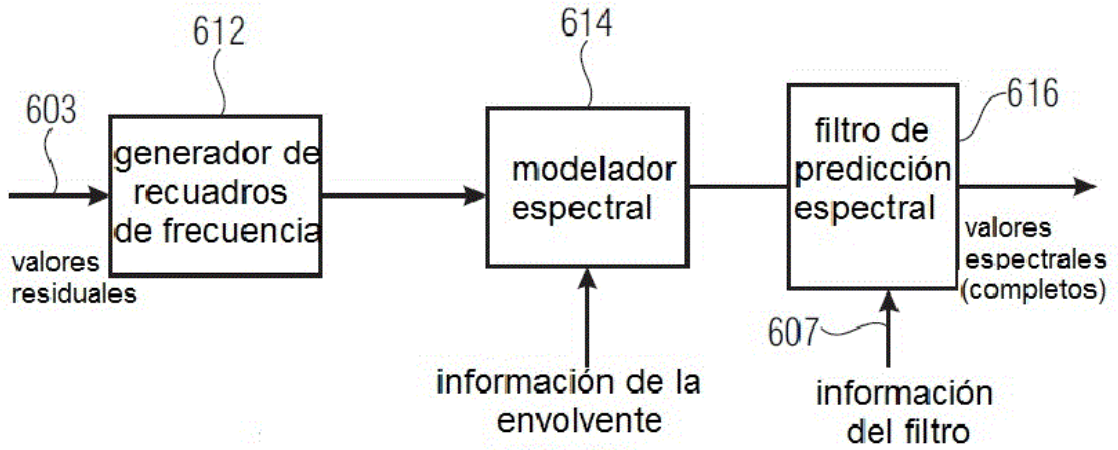


FIG 6B

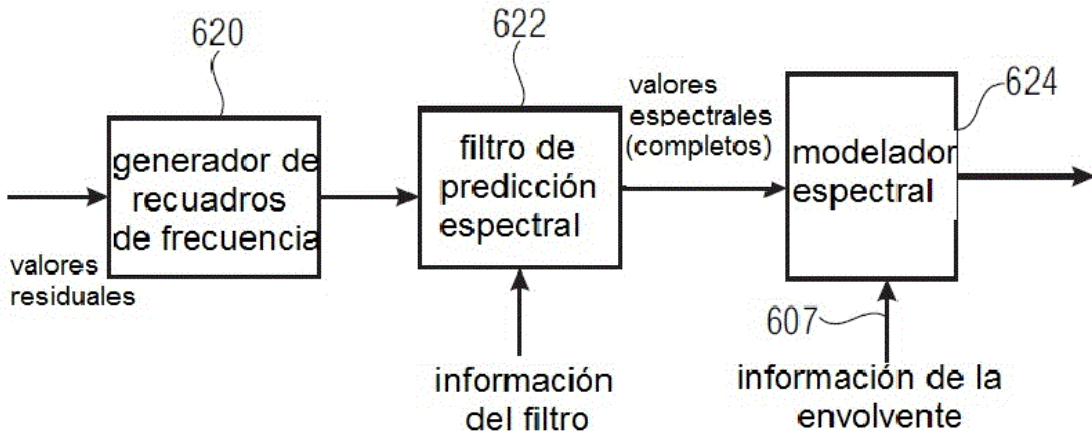
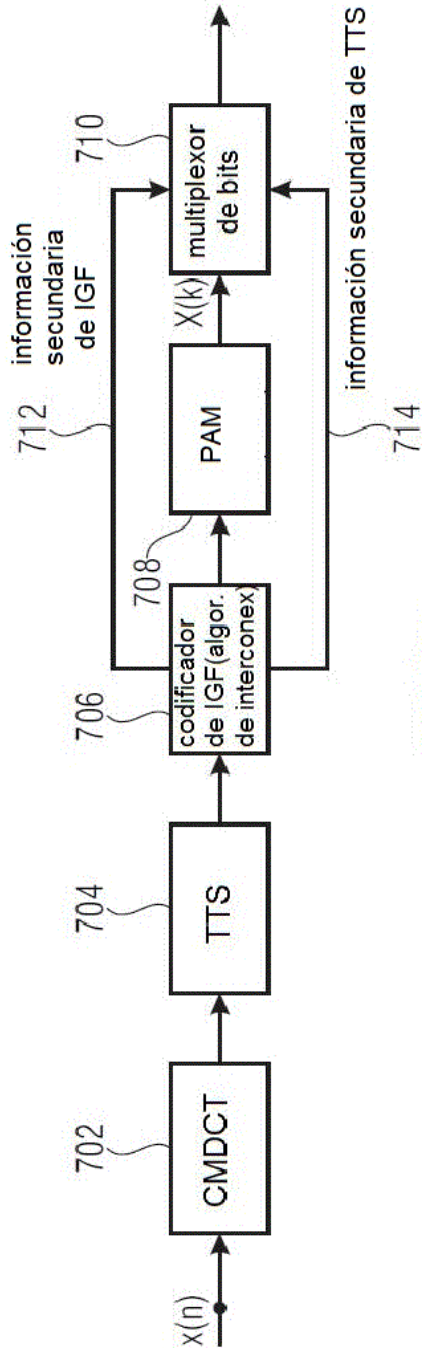
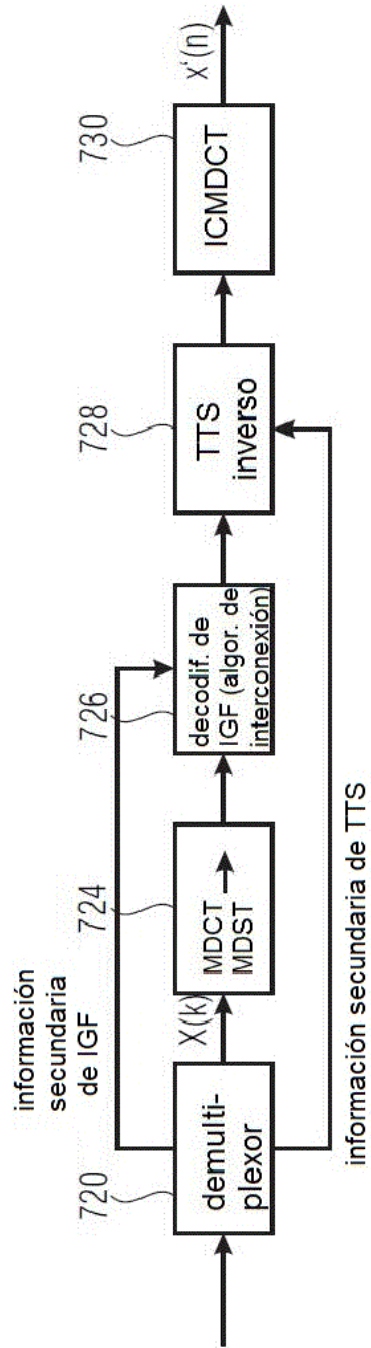


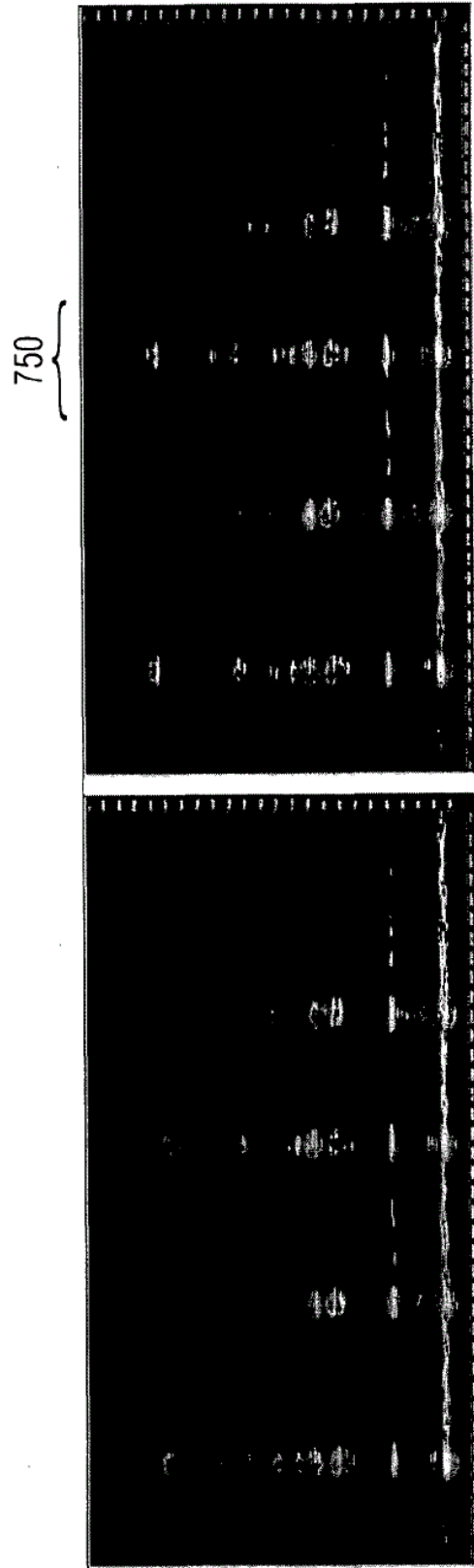
FIG 6C



**FIG 7A
(CODIFICADOR)**



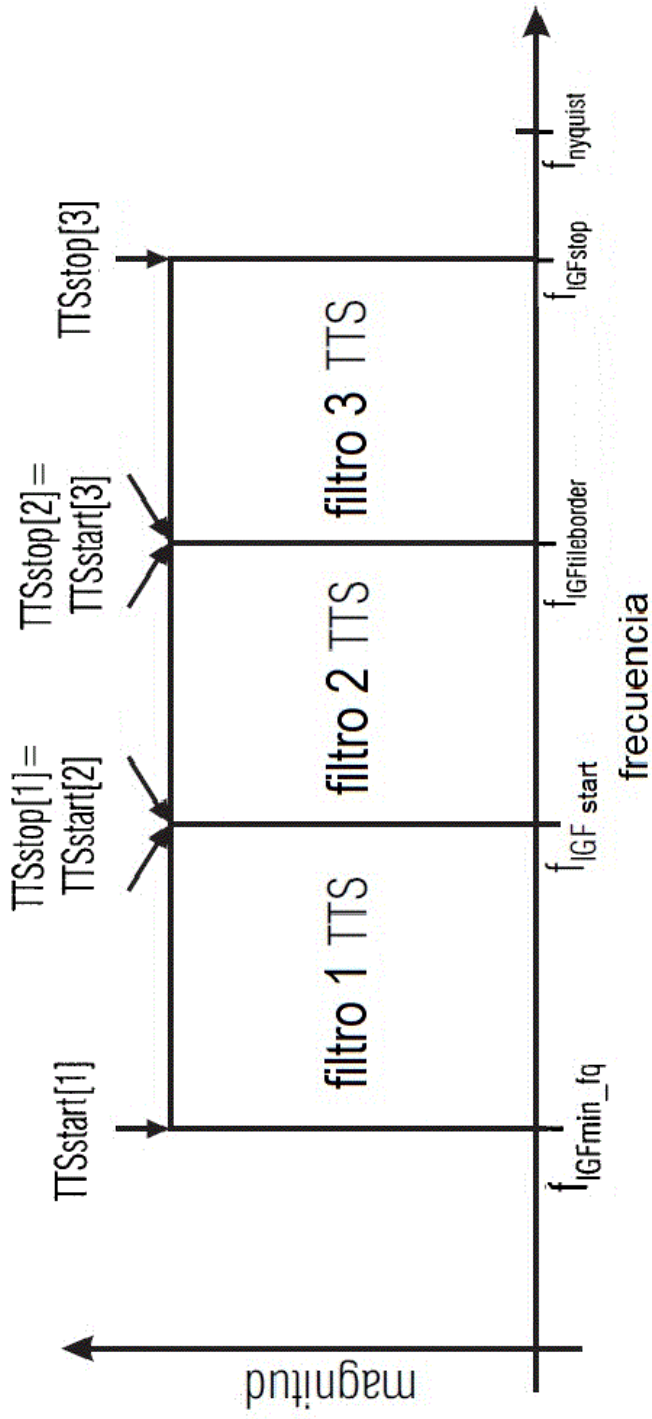
**FIG 7B
(DECODIFICADOR)**



Espectrograma de DFT de la señal original (panel izquierdo);

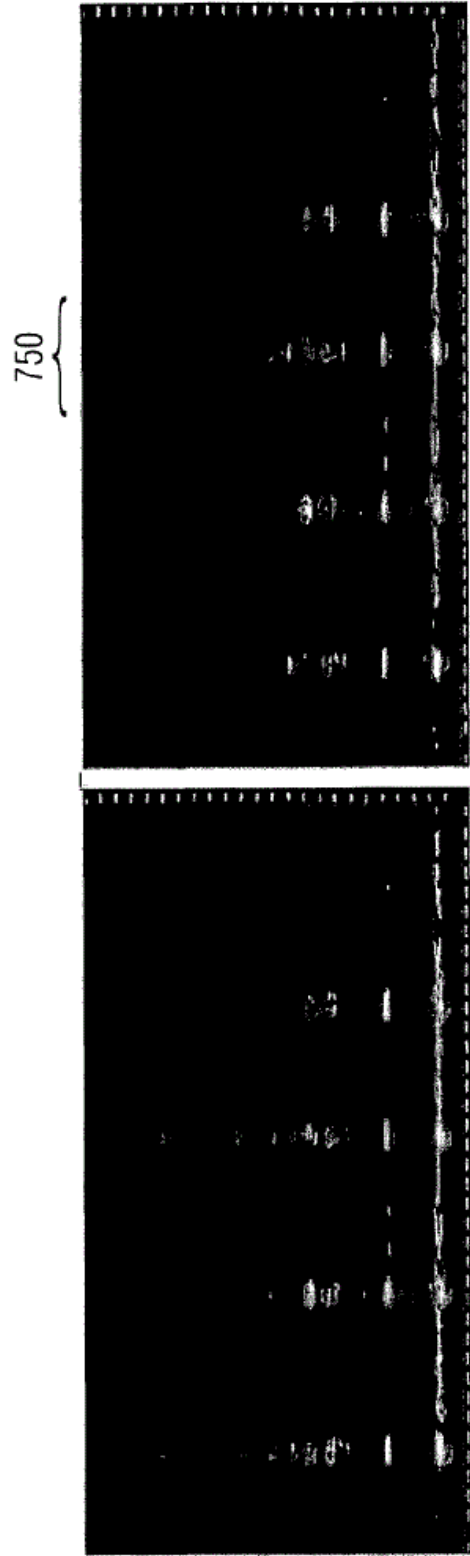
Espectrograma de DFT de la señal extendida sin TTS (panel derecho).

FIG 7C



Esquema de filtros de TTS de base. Cada recuadro marca la región espectral cubierta por un filtro en particular .

FIG 7D



Espectrograma de DFT de la señal original (panel izquierdo);
Espectrograma de DFT de la señal extendida con TTS (panel derecho).

FIG 7E

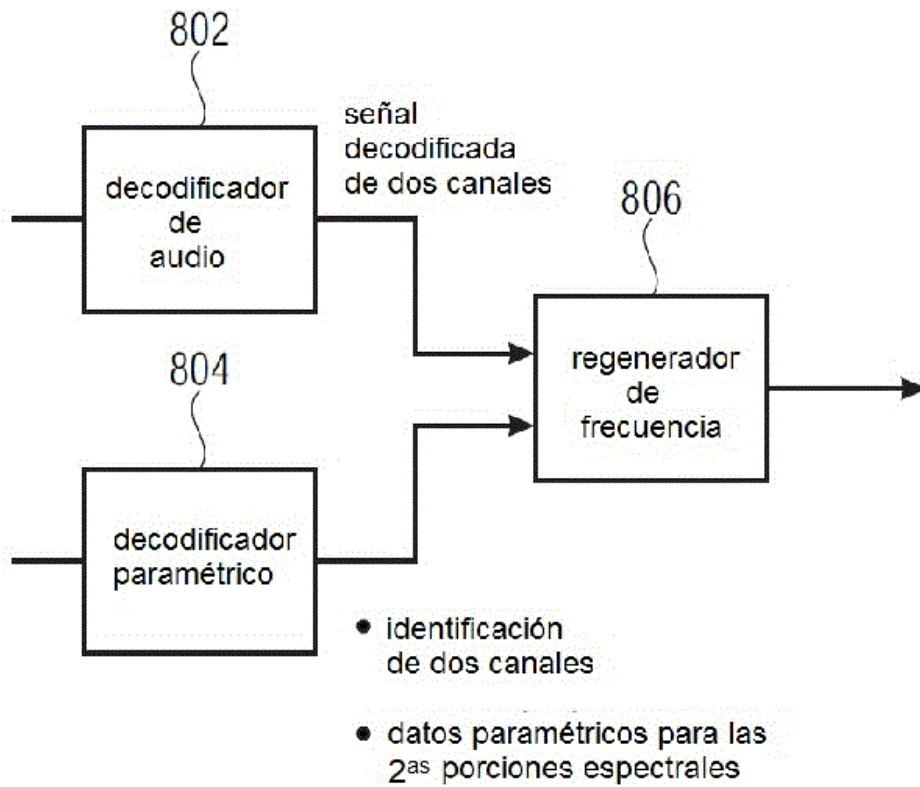


FIG 8A

intervalo de origen	1	1	2	2	<p>1: primera representación de dos canales (por ej. L/R)</p> <p>2: segunda representación de dos canales (por ej. M/S)</p>
intervalo de destino	1	2	1	2	

FIG 8B

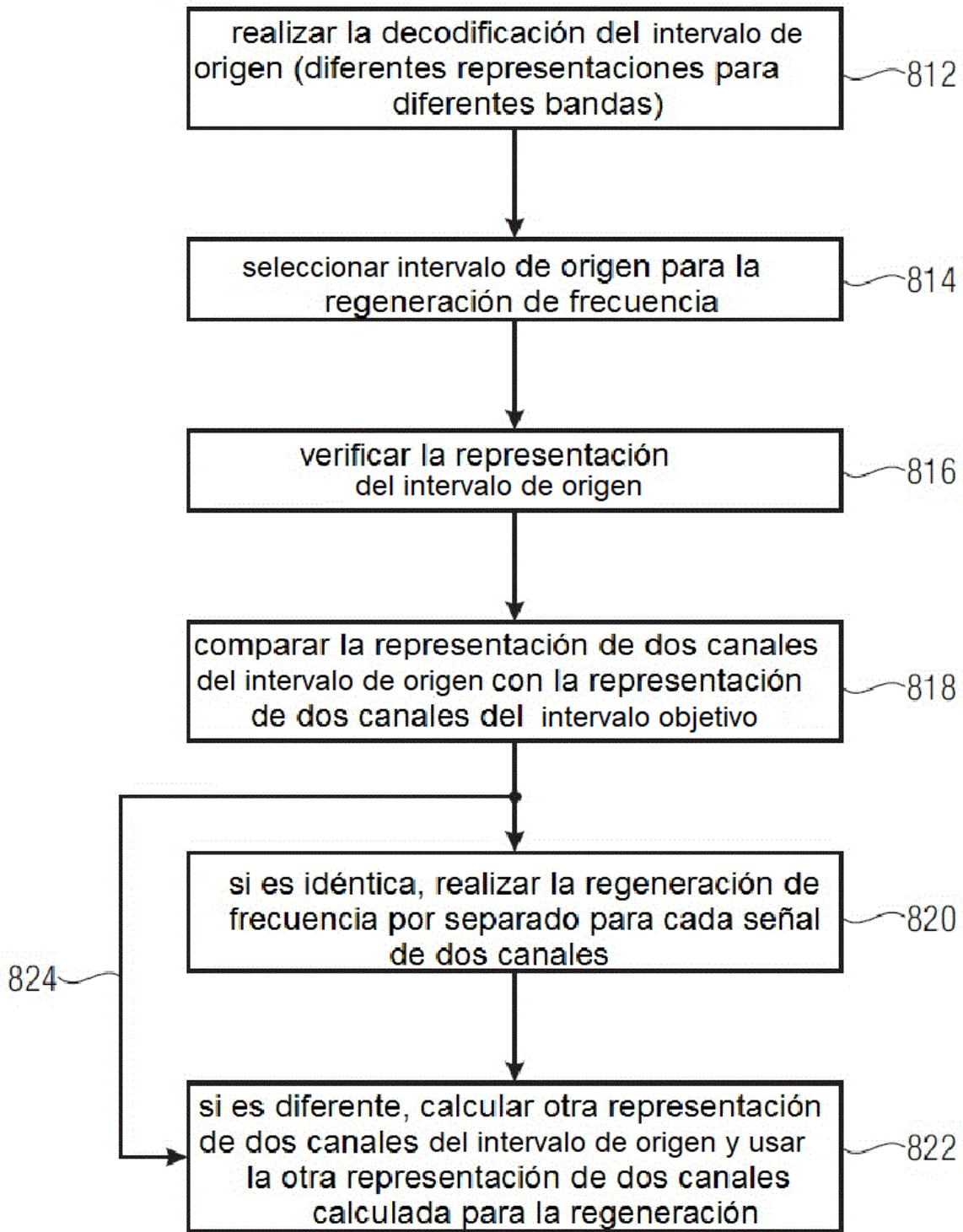


FIG 8C

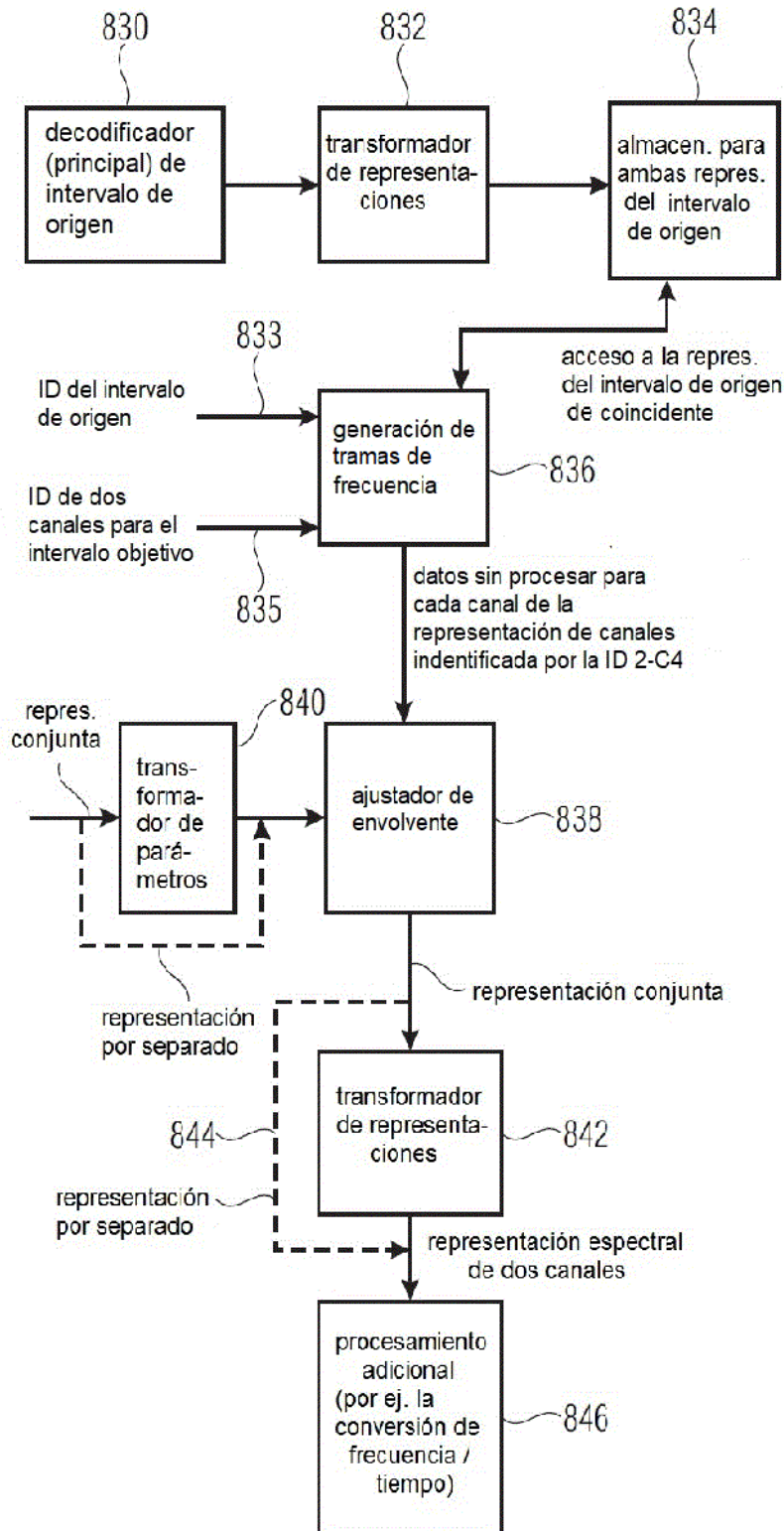


FIG 8D

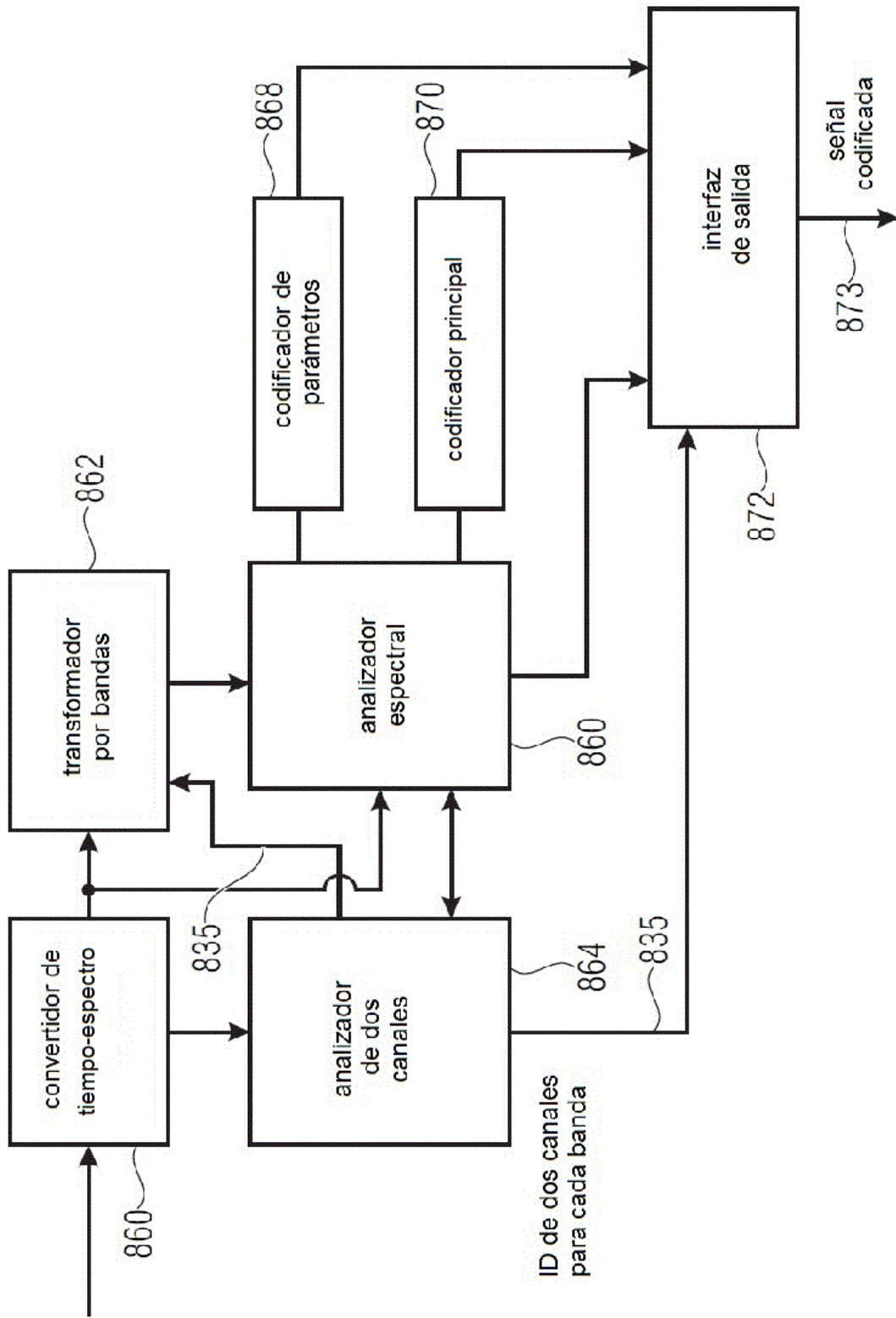


FIG 8E

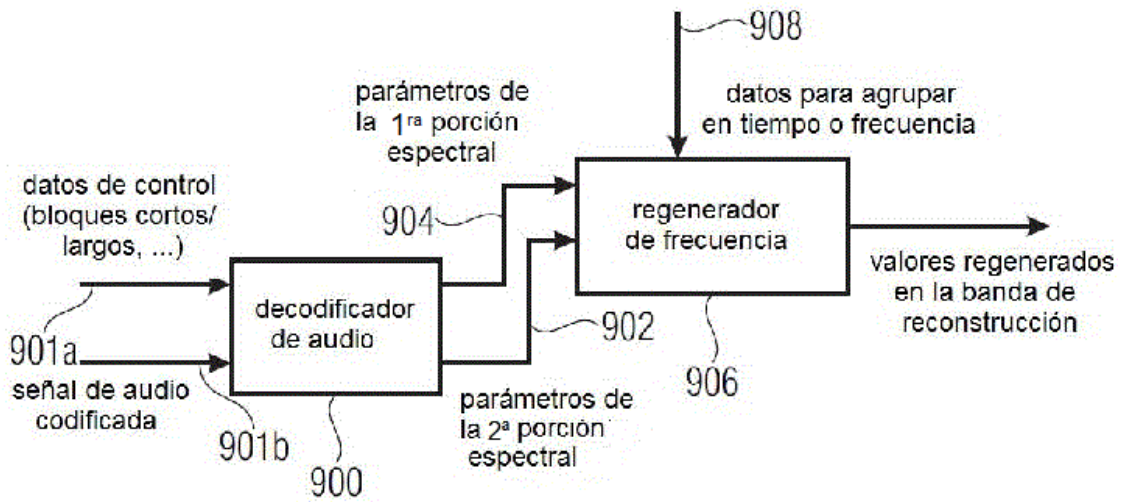


FIG 9A

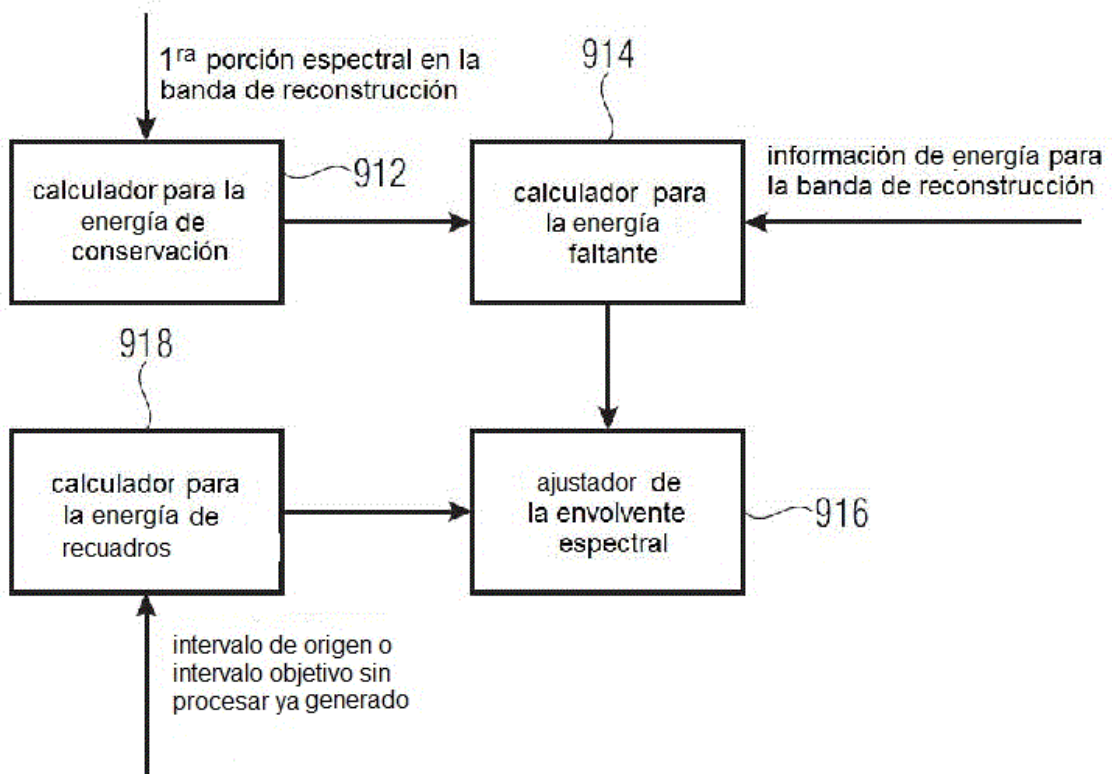
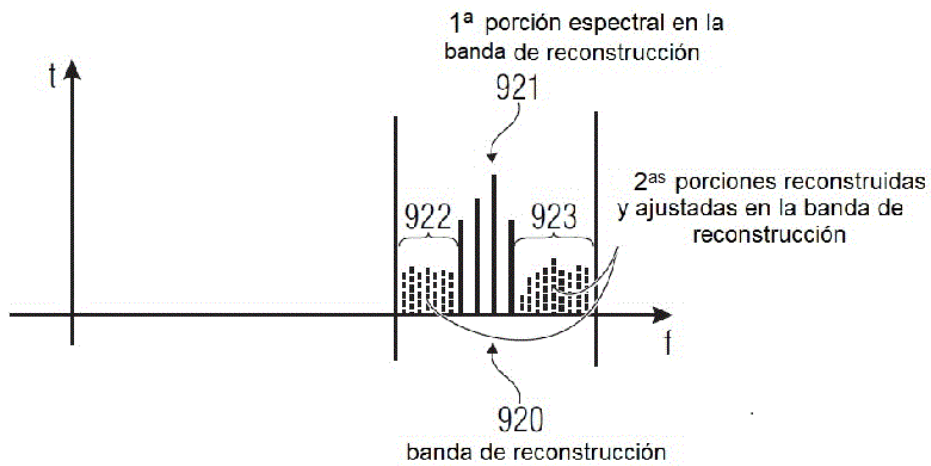


FIG 9B



por ej.

- energía de conservación: 5 unidades
- valor de la energía para la banda de reconstr.: 10 unidades (abarca 1ª y 2ª porciones espectrales en la banda de reconstrucción)
- energía de datos de intervalo de origen o datos de intervalo objetivo sin procesar: 8 unidades
- energía faltante: 5 unidades
- factor de ganancia: $g = \sqrt{\frac{mE_k}{pE_k}} = 0.79$

→ sólo se ajustan valores espectrales para las 2ª porciones espectrales

→ la 1ª porción espectral no se ve influida por el ajuste de envolvente

FIG 9C

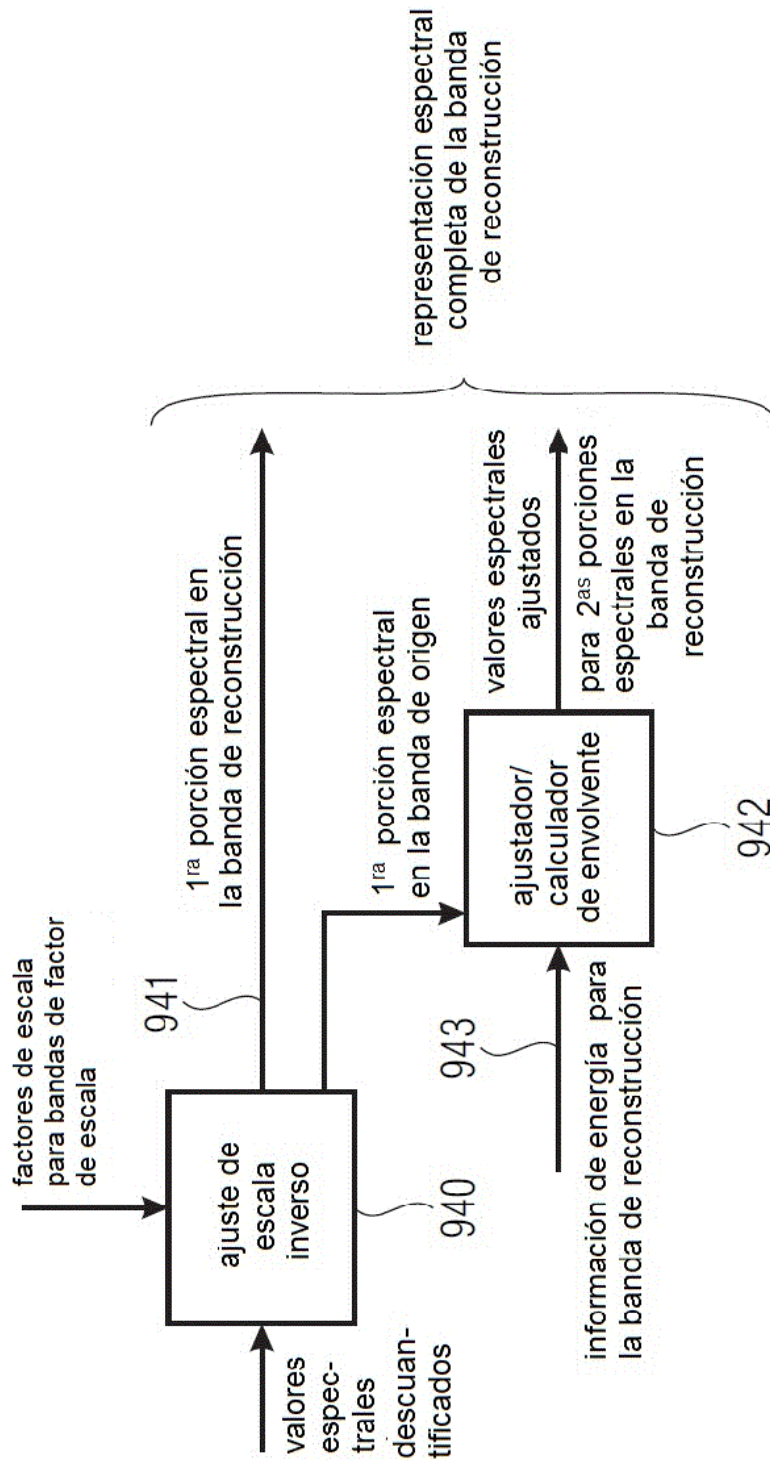


FIG 9D

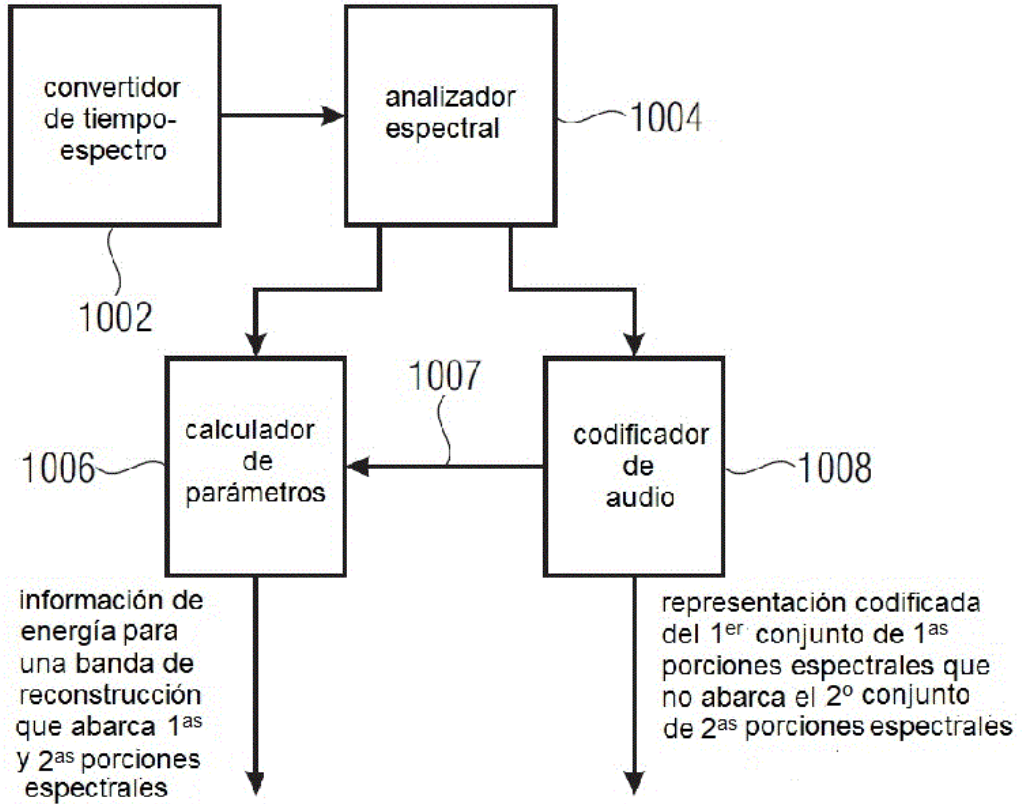


FIG 10A

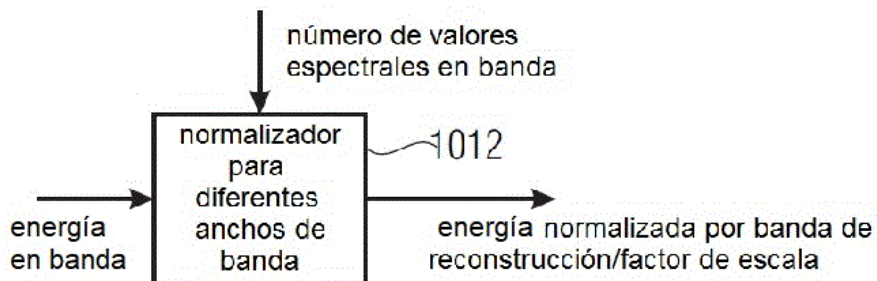


FIG 10B

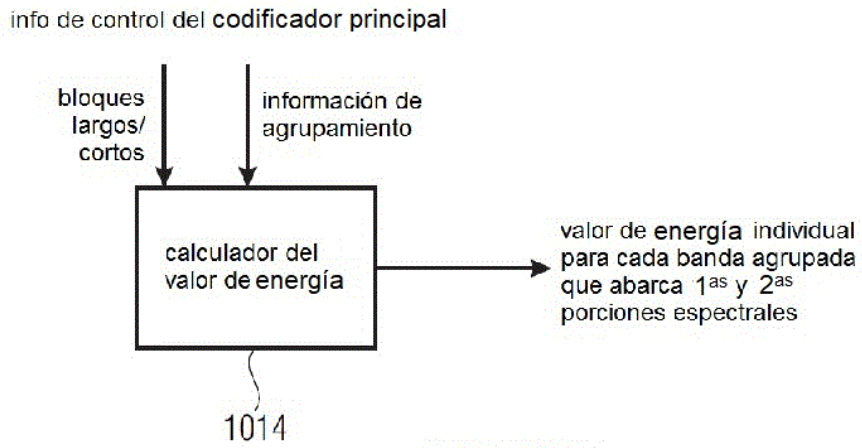


FIG 10C

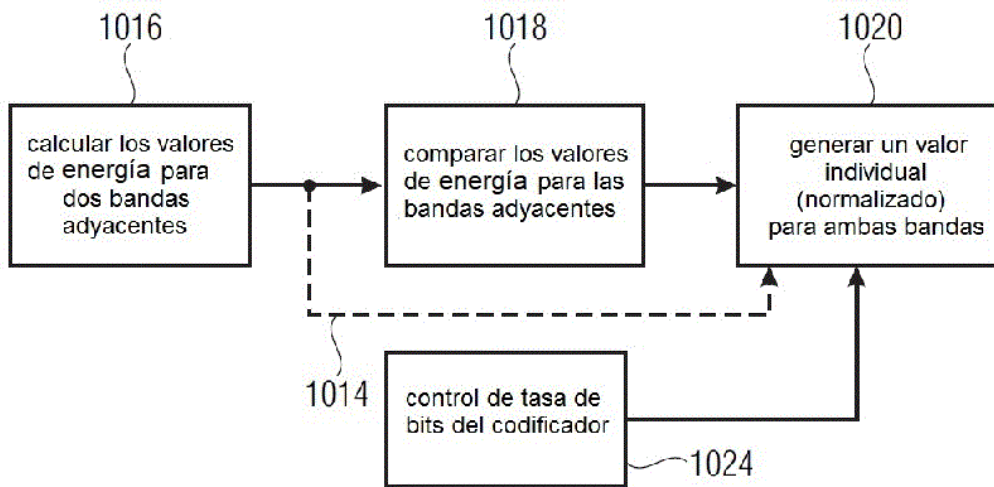


FIG 10D

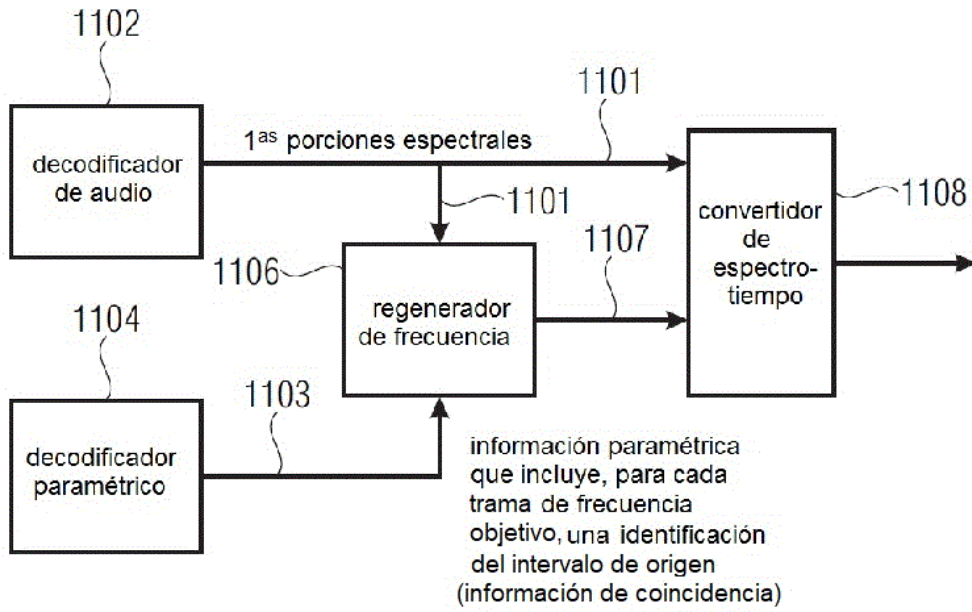


FIG 11A

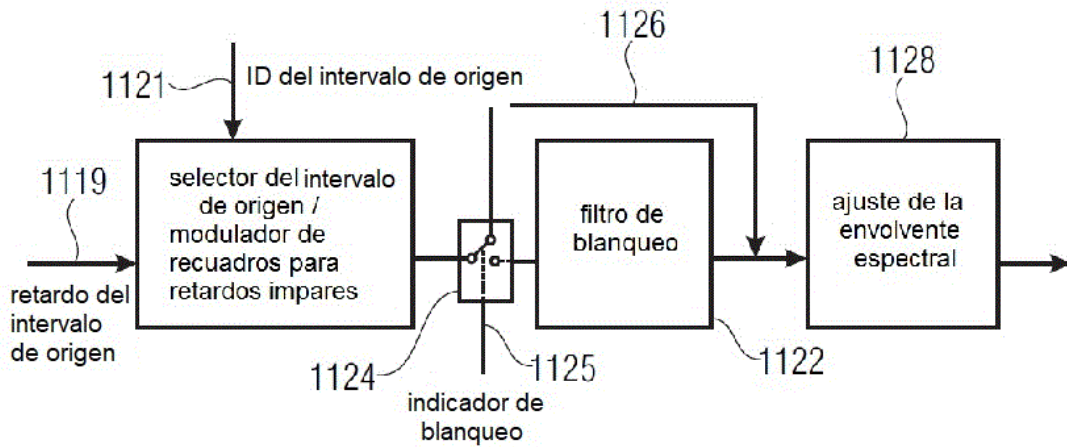


FIG 11B

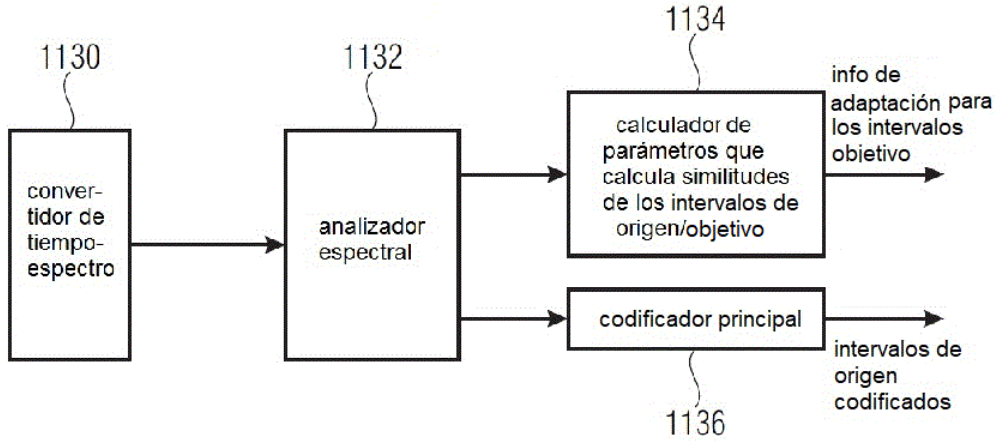


FIG 11C

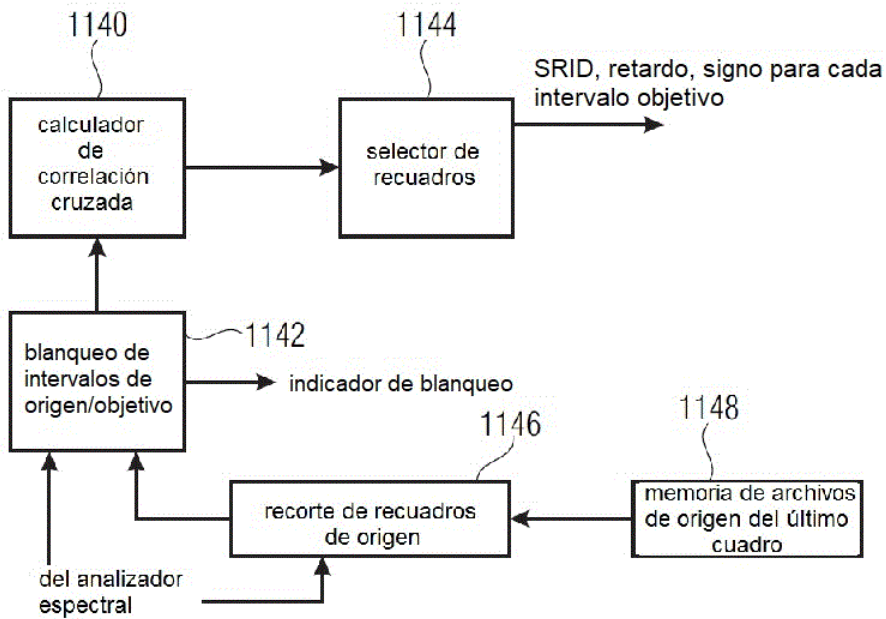


FIG 11D

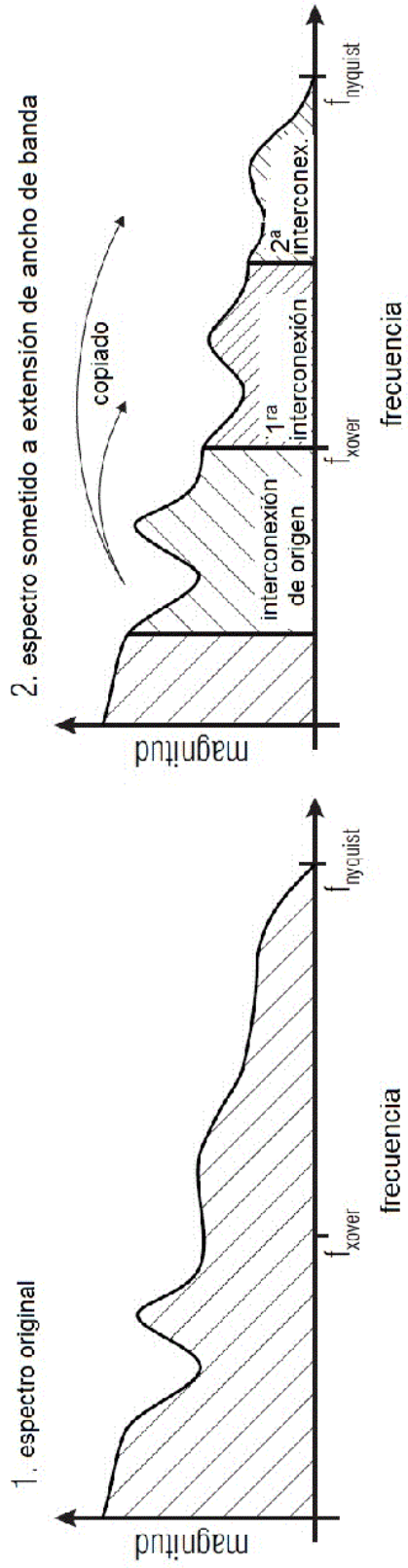


FIG 12A

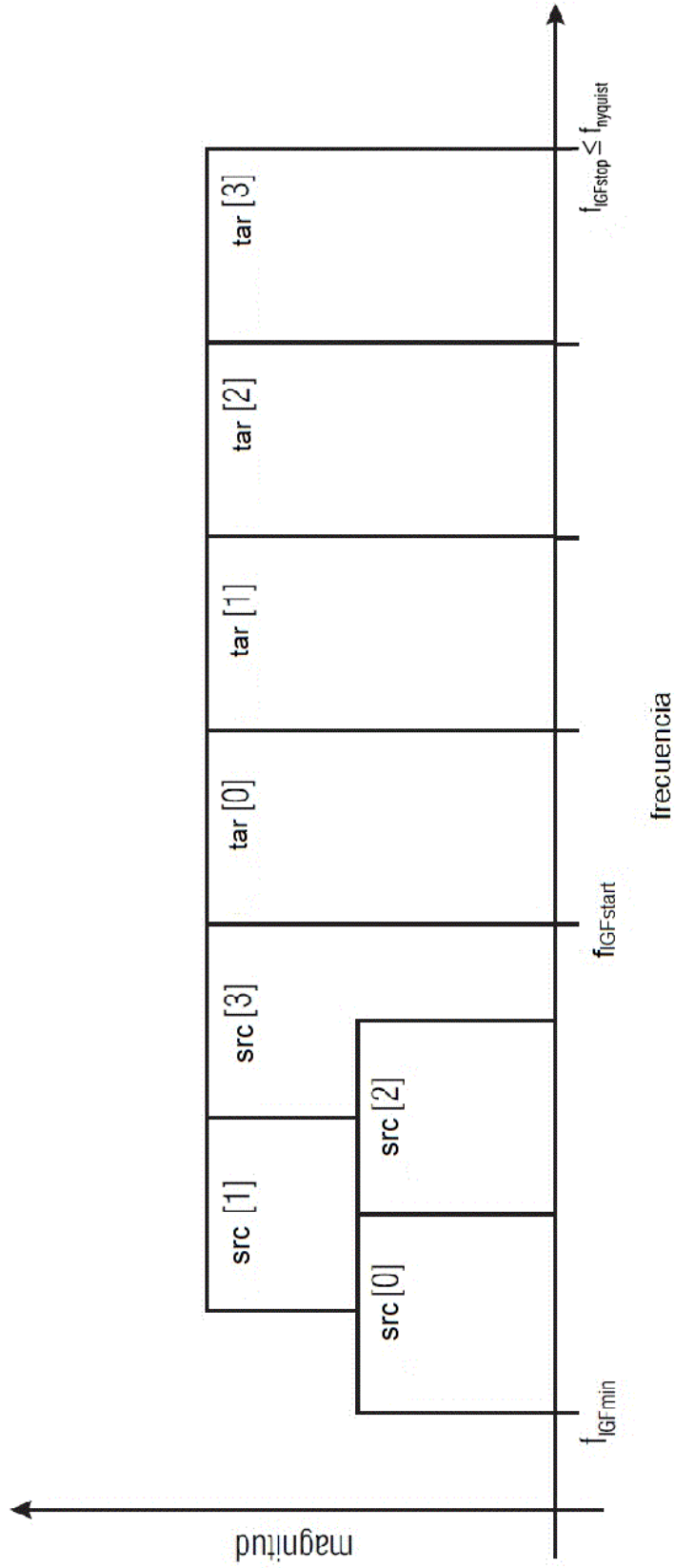


FIG 12B

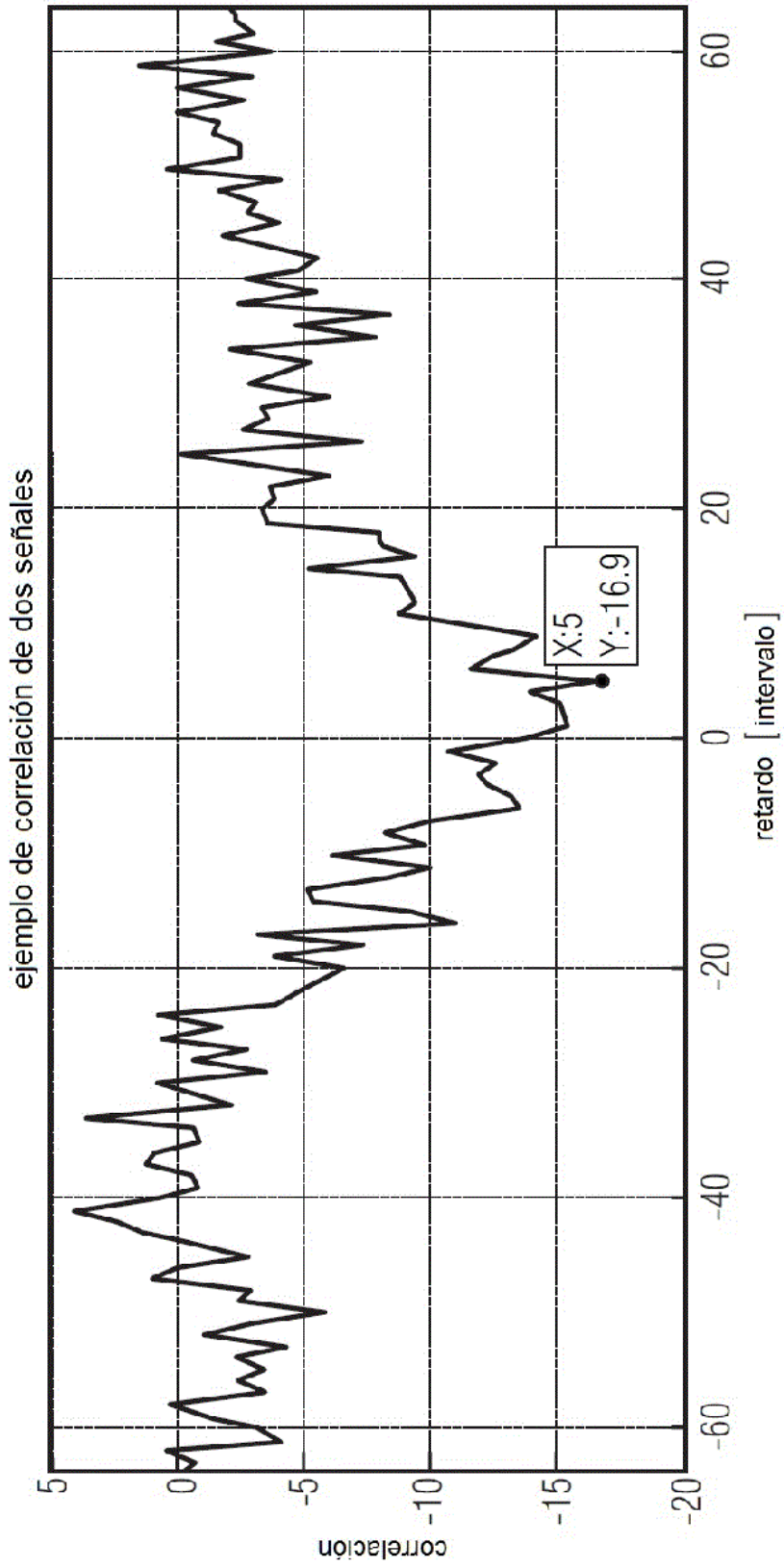


FIG 12C

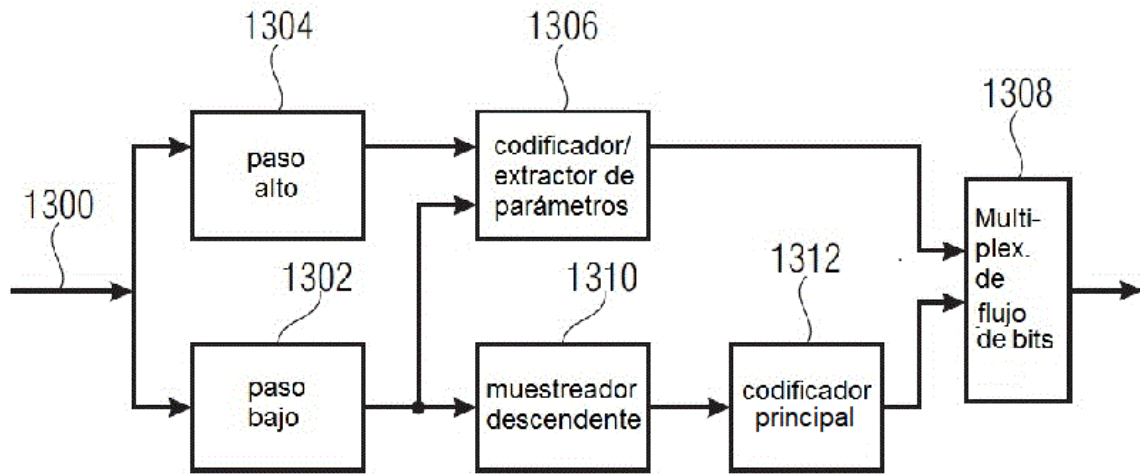


FIG 13A
(TÉCNICA ANTERIOR)

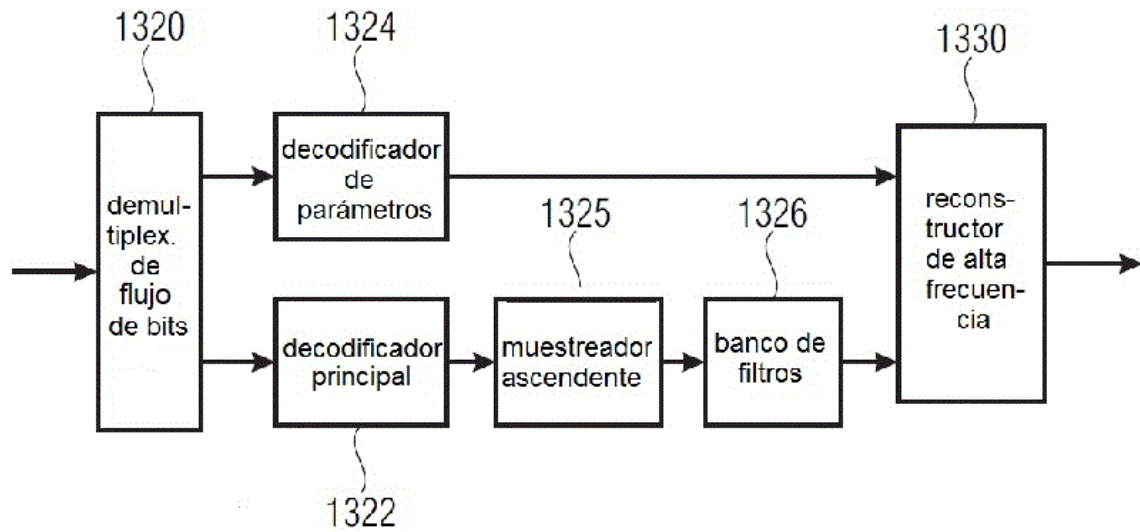


FIG 13B
(TÉCNICA ANTERIOR)