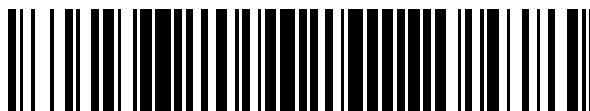


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 599 007**

51 Int. Cl.:

G10L 21/0388 (2013.01)

G10L 19/02 (2013.01)

G10L 19/03 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.07.2014** **PCT/EP2014/065123**

87 Fecha y número de publicación internacional: **29.01.2015** **WO15010954**

96 Fecha de presentación y número de la solicitud europea: **15.07.2014** **E 14738857 (3)**

97 Fecha y número de publicación de la concesión europea: **17.08.2016** **EP 2883227**

54 Título: **Aparato y método para codificar y decodificar una señal de audio codificada utilizando modelado de ruido/parche temporal**

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 13189358

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
31.01.2017

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;
THOSKHAHNA, BALAJI NAGENDRAN;
SCHMIDT, KONSTANTIN;
BAYER, STEFAN;
NEUKAM, CHRISTIAN;
EDLER, BERND y
HELMRICH, CHRISTIAN**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 599 007 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).

Aparato y método para codificar y decodificar una señal de audio codificada utilizando modelado de ruido/parche temporal

DESCRIPCIÓN

5

Campo de la invención

10 La presente invención se refiere a la codificación/decodificación de audio y, en particular, a la codificación de audio que utiliza Relleno Inteligente de Espacios (IGF).

15 La codificación de audio es el dominio de la compresión de la señal que trata el aprovechamiento de la redundancia y la irrelevancia de señales de audio utilizando el conocimiento psicoacústico. Actualmente, los códec de audio generalmente necesitan alrededor de 60 kbps/canal para la codificación perceptual transparente de casi cualquier tipo de señal de audio. Los códec más nuevos tienen como objetivo reducir la tasa de bits de codificación aprovechando las similitudes espectrales en la señal y utilizando técnicas tales como la extensión de ancho de banda (BWE, por sus siglas en inglés). Un esquema de extensión de ancho de banda (BWE) utiliza un parámetro bajo de tasa de bits establecido para representar los componentes de alta frecuencia (HF) de una señal de audio. El espectro de HF se rellena con el 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. Estos métodos de extensión de ancho de banda (BWE) permiten que los códec de audio mantengan buena calidad a tasas de bits incluso bajas de alrededor de 24 kbps/canal.

25 El almacenamiento o la transmisión de señales de audio a menudo están sometidos a estrictas limitaciones de tasas de bits. En el pasado, los codificadores se vieron obligados a reducir drásticamente el ancho de banda de audio transmitida cuando solo estaba disponible una tasa de bits muy baja.

30 Los códec de audio modernos ahora son capaces de codificar señales de banda ancha utilizando los 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) codificada en forma de onda de la señal decodificada por medio de transposición a la región espectral de alta frecuencia (HF) ("emparchar") y la aplicación de un procesamiento posterior basado en parámetros. En los esquemas de extensión de ancho de banda (BWE), la reconstrucción de la región espectral de alta frecuencia (HF) por encima de una denominada frecuencia de cruce determinada se basa a menudo en el emparchado espectral. En general, la región de alta frecuencia (HF) está compuesta de múltiples parches adyacentes y cada uno de estos parches se obtiene de regiones de paso de banda (BP) del espectro de baja frecuencia (LF) por debajo de la frecuencia de cruce determinada. Los sistemas del estado actual de la técnica desempeñan con eficiencia el emparchado dentro de una representación de bancos de filtros, por ejemplo Banco de Filtros Espejo en Cuadratura (QMF, por sus siglas en inglés), copiando un conjunto de coeficientes de subbandas adyacentes desde una región de origen a la región de destino.

45 Otra técnica que se encuentra en los códec de audio actuales que aumenta la eficiencia de compresión y permite de esta manera el ancho de banda de audio extendido en tasas de bits bajas es el reemplazo sintético basado en parámetros de partes apropiadas de los espectros de audio. Por ejemplo, las partes de la señal similar a 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 lateral. Un ejemplo es la herramienta de Sustitución de Ruido Perceptual herramienta (PNS, por sus siglas en inglés) contenida en la Codificación Avanzada de Audio MPEG-4 (AAC, por sus siglas en inglés) [5].

50 Otra disposición que también permite un ancho de banda de audio extendido en tasas de bits bajas es la técnica de relleno de ruido contenida en el Sistema Unificado de Codificación de Voz y Audio MPEG-D (USAC, por sus siglas en inglés) [7]. Los espacios espectrales (ceros) que se deducen por la zona muerta del cuantificador debido a una cuantificación demasiado basta, posteriormente se llenan de ruido artificial en el decodificador y se ajustan a escala por un procesamiento posterior basado en parámetros.

55 Otro sistema del estado actual de la técnica se denomina Reemplazo Espectral Preciso (ASR, por sus siglas en inglés) [2-4]. Además de un códec de forma de onda, el reemplazo espectral preciso (ASR) emplea una etapa de síntesis de señal especializada que restaura perceptualmente porciones sinusoidales importantes de la señal en el decodificador. Asimismo, un sistema descrito en [5] se basa en el modelado sinusoidal en la región de alta frecuencia (HF) de un codificador de forma de onda para permitir que el ancho de banda de audio extendido tenga una calidad perceptual adecuada 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, por sus siglas en inglés) y también etapas de análisis/síntesis bastante complejas para la conservación de componentes sinusoidales de alta frecuencia (HF).

La Fig. 13a ilustra un diagrama esquemático de un codificador de audio para una tecnología de extensión de ancho de banda como, por ejemplo, la que se utiliza en la Codificación Avanzada de Audio de Alta Eficiencia (HE-AAC, por sus siglas en inglés). Una señal de audio en la línea 1300 se introduce en un sistema de filtro que comprende un paso bajo 1302 y un paso alto 1304. La señal emitida por el filtro de paso alto 1304 se introduce en un extractor/codificador de parámetros 1306. El extractor/codificador de parámetros 1306 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 de corriente de bits 1308. La señal de salida de paso bajo se introduce en un procesador que generalmente comprende la funcionalidad de un muestreador descendente 1310 y un codificador central 1312. El paso bajo 1302 restringe el ancho de banda para que se codifique en 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 central solo tienen que operar 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 de paso bajo 1302 tiene a modo de ejemplo un ancho de banda de 4 kHz, con el fin de cumplir el teorema de muestreo, es teóricamente suficiente que la señal posterior al muestreador descendente tenga una frecuencia de muestreo de 8 kHz, que es una reducción sustancial de la tasa de muestreo requerida para la señal de audio 1300 que tiene que ser de al menos 40 kHz.

La Fig. 13b ilustra un diagrama esquemático de un decodificador de extensión de ancho de banda respectivo. El decodificador comprende un multiplexor de corriente de bits 1320. El demultiplexor de corriente de bits 1320 extrae una señal de entrada para un decodificador central 1322 y una señal de entrada para un decodificador de parámetros 1324. Una señal de salida del decodificador central 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 de alta frecuencia 1330 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 ascendente 1325 y un banco de filtros 1326. El reconstructor de alta frecuencia 1330 recibe entonces la señal de baja frecuencia analizada por frecuencia emitida por el banco de filtros 1326 y reconstruye el rango de frecuencias definido por el filtro de paso alto 1304 de la Fig. 13a utilizando la representación paramétrica de la banda de alta frecuencia. El reconstructor de alta frecuencia 1330 tiene varias funcionalidades tales como la regeneración del rango de frecuencias superior que utiliza el rango de origen en el rango de baja frecuencia, un ajuste de envolvente espectral, una funcionalidad de adición de ruido y una funcionalidad para introducir armónicos faltantes en el rango de frecuencia superior y, si se aplica y calcula en el codificador de la Fig. 13a, una operación de filtrado inverso con el fin de tener en cuenta el hecho de que el rango de frecuencia superior normalmente no es tan tonal como el rango de frecuencia inferior. En la Codificación Avanzada de Audio de Alta Eficiencia (HE-AAC), los armónicos faltantes se re-sintetizan en el lado del decodificador y se colocan exactamente en el medio de una banda de reconstrucción. Por lo tanto, todas las líneas de armónicos faltantes que se han determinado en una cierta banda de reconstrucción no se colocan en los valores de frecuencia en donde estaban ubicadas en la señal original. En cambio, dichas líneas de armónicos faltantes se colocan en frecuencias en el centro de la banda determinada. Por lo tanto, cuando una línea de armónico faltante en la señal original se colocó 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 está cerca del 50 % de la banda de reconstrucción individual, para la que se han generado y transmitido parámetros.

Además, a pesar de que los codificadores centrales de audio típicos operan en el dominio espectral, el decodificador central, no obstante, genera una señal de dominio temporal que, a continuación, es convertida nuevamente en un dominio espectral por la funcionalidad del banco de filtros 1326. 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 nuevamente la transformación en generalmente un dominio de frecuencia diferente y, por supuesto, esto también requiere una cantidad sustancial de complejidad computacional y, por lo tanto, energía eléctrica, que representa básicamente un problema cuando se aplica la tecnología de extensión de ancho de banda en dispositivos móviles como, por ejemplo, teléfonos móviles, tabletas o computadoras portátiles, etc.

Los códec de audio actuales realizan la codificación de audio de baja tasa de bits utilizando la extensión de ancho de banda (BWE) como parte integral del esquema de codificación. Sin embargo, las técnicas de extensión de ancho de banda (BWE) se limitan a reemplazar solo contenido de alta frecuencia (HF). Asimismo, no permiten la codificación de forma de onda del contenido perceptivamente importante por encima de una frecuencia de cruce determinada. Por lo tanto, los códec de audio contemporáneos, pierden detalle de alta frecuencia (HF) o timbre cuando se implementa la extensión de ancho de banda (BWE), ya que la alineación exacta de los armónicos tonales de la señal no se tiene en cuenta en la mayoría de los sistemas.

Otra desventaja de los sistemas de extensión de ancho de banda (BWE) del estado actual de la técnica es la

necesidad de transformación de la señal de audio en un nuevo dominio para la implementación de la BWE (por ejemplo, transformación de la Transformada de Coseno Discreta Modificada (MDCT) al dominio de Filtros Espejo en Cuadratura (QMF). Esto conduce a complicaciones de sincronización, complejidad computacional adicional y aumento de requisitos de memoria.

5 En particular, si se implementa un sistema de extensión de ancho de banda en un banco de filtros o el dominio de la transformada de tiempo-frecuencia, solo hay una posibilidad limitada para controlar la forma temporal de la señal de extensión de ancho de banda. Generalmente, la granularidad temporal está limitada por el tamaño de salto utilizado entre ventanas adyacentes de la transformada. Esto puede conducir a pre o post-ecos no deseados en el rango
10 espectral de la extensión de ancho de banda. Con el fin de aumentar la granularidad temporal se pueden utilizar cuadros de extensión de ancho de banda más cortos o con tamaños de salto más cortos, si bien esto produce una sobrecarga de tasa de bits debido a que, durante un cierto período de tiempo, se debe transmitir un número más grande de parámetros, generalmente un cierto conjunto de parámetros para cada cuadro de tiempo. De lo contrario, si los cuadros de tiempo individuales se hacen demasiado grandes, entonces se generan pre y post-ecos
15 particularmente para las porciones de transición de una señal de audio.

Un objetivo de la presente invención consiste en proporcionar un concepto mejorado de codificación/decodificación.

20 Este objetivo se logra mediante un aparato para decodificar una señal de audio codificada de acuerdo con la reivindicación 1, un aparato para codificar una señal de audio de acuerdo con la reivindicación 10, un método de decodificación de acuerdo con la reivindicación 16, un método de codificación de acuerdo con la reivindicación 18 o un programa informático de acuerdo con la reivindicación 19.

25 La presente invención se basa en el descubrimiento de que una mejora de la calidad y la reducción de la tasa de bits específicamente para señales que comprenden porciones de transición, como ocurre muy a menudo en las señales de audio, se obtiene combinando la tecnología de Modelado de Ruido Temporal (TNS, por sus siglas en inglés) (véase, por ejemplo, el documento US 2006/0122828 A1) o Modelado de Recuadro Temporal (TTS, por sus siglas en inglés) con reconstrucción de alta frecuencia. El procesamiento de Modelado de Ruido Temporal (TNS)/Modelado de Recuadro Temporal (TTS) en el lado del codificador implementado por una predicción sobre la frecuencia
30 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 rango de frecuencia no solo abarcando el rango de frecuencia de origen sino también el rango de frecuencia de destino para reconstruirse en un decodificador de regeneración de frecuencia, la envolvente temporal se aplica no solo a la señal de audio central hasta una frecuencia de inicio de relleno de espacios, sino que también se aplica la envolvente temporal a los rangos
35 espectrales de segundas porciones espectrales reconstruidas. Por lo tanto, los pre-ecos o los post-ecos que se producirían sin el modelado de recuadro temporal se reducen o eliminan. Esto se logra aplicando una predicción inversa sobre la frecuencia no solo dentro del rango de frecuencia central hasta una cierta frecuencia de inicio de relleno de espacios sino también dentro de un rango de frecuencia por encima del rango de frecuencia central. Para este fin, la regeneración de frecuencia o la generación de recuadros de frecuencia se realizan en el lado del
40 decodificador antes de aplicar una predicción sobre la frecuencia. Sin embargo, la predicción sobre la frecuencia se puede aplicar ya sea antes o después del modelado de la envolvente espectral dependiendo de si el cálculo de información de energía se ha realizado sobre los valores espectrales residuales posteriores al filtrado o a los valores espectrales (completos) antes del modelado de la envolvente.

45 El procesamiento del Modelado de Recuadro Temporal (TTS) sobre uno o más recuadros de frecuencia establece adicionalmente una continuidad de correlación entre el rango de origen y el rango de reconstrucción o en dos rangos de reconstrucción adyacentes o recuadros de frecuencia.

50 En una implementación se prefiere utilizar el filtrado complejo de Modelado de Ruido Temporal (TNS)/Modelado de Recuadro Temporal (TTS). De este modo se evitan los artefactos de solapamiento (temporales) de una representación real muestreada críticamente, como la Transformada de Coseno Discreta Modificada (MDCT). Un filtro complejo de Modelado de Ruido Temporal (TNS) se puede calcular en el lado del codificador aplicando no solo una transformada de coseno discreta modificada sino también una transformada sinusoidal discreta modificada para obtener adicionalmente una transformada modificada compleja. Sin embargo, solo se transmiten los valores de la
55 transformada de coseno discreta modificada, es decir, la parte real de la transformada compleja. Sin embargo, en el lado del decodificador, es posible estimar la parte imaginaria de la transformada utilizando los espectros de la MDCT de cuadros anteriores o posteriores de modo que, en el lado del decodificador, el filtro complejo se puede aplicar nuevamente en la predicción inversa sobre la frecuencia y, específicamente, la predicción sobre el límite entre el rango de origen y el rango de reconstrucción y también sobre el límite entre los recuadros de frecuencia adyacentes
60 de la frecuencia dentro del rango de reconstrucción.

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 central por otro lado se pueden abordar y superar realizando la extensión de ancho de banda en el mismo dominio espectral en el que opera el decodificador central.

Por lo tanto, se proporciona un decodificador central de tasa completa que codifica y decodifica todo el rango de la señal de audio. Esto no genera la necesidad de un muestreador descendente en el lado del codificador y un muestreador ascendente en el lado del decodificador. En cambio, todo el procesamiento se realiza 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, la señal de audio se analiza con el fin de encontrar un primer conjunto de primeras porciones espectrales que tiene que codificarse con una alta resolución, en donde este primer conjunto de primeras porciones espectrales puede incluir, en una forma de realización, porciones tonales de la señal de audio. Por otra parte, los componentes no tonales o con ruido en la señal de audio que constituyen un segundo conjunto de segundas porciones espectrales se codifican paramétricamente con baja resolución espectral. Entonces, la señal de audio codificada solo requiere que el primer conjunto de primeras porciones espectrales esté codificado de manera que conserve la forma de onda con una alta resolución espectral y, adicionalmente, que el segundo conjunto de segundas porciones espectrales esté codificado paramétricamente con una baja resolución utilizando "recuadros" de frecuencia obtenidos del primer conjunto. En el lado del decodificador, el decodificador central, que es un decodificador de banda completa, reconstruye el primer conjunto de primeras porciones espectrales de manera que conserve la forma de onda, es decir, desconociendo si hay una regeneración de frecuencia adicional. Sin embargo, el espectro así generado tiene un montón de espacios espectrales. Estos espacios se rellenan posteriormente con la tecnología de relleno inteligente de espacios (IGF, por sus siglas en inglés) de la invención utilizando una regeneración de frecuencia que aplica datos paramétricos por un lado, y que utiliza un rango espectral de origen, es decir, primeras porciones espectrales reconstruidas por el decodificador de audio de tasa completa, por otro lado.

En otras formas de realización, las porciones espectrales, que se reconstruyen por relleno de ruido solamente 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 opera en un dominio único para la codificación/decodificación central por un lado, y la regeneración de frecuencia por otro lado, el relleno inteligente de espacios (IGF) no solo se limita a llenar un rango de frecuencia superior sino que puede llenar rangos de frecuencia inferiores, ya sea por relleno de ruido sin regeneración de frecuencia o por regeneración de frecuencia utilizando un recuadro de frecuencia en un rango de frecuencia diferente.

Asimismo, se insiste en 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 solo un valor de energía sino también un valor de amplitud (por ejemplo, absoluto), un valor de nivel o cualquier otro valor, del que se puede obtener un valor de energía final. Por lo tanto, la información sobre una energía puede comprender, por ejemplo, el valor de energía propiamente dicho, y/o un valor de un nivel y/o de una amplitud y/o de una amplitud absoluta.

Un aspecto adicional se basa en el descubrimiento de que la situación de correlación no solo es importante para el rango de origen, sino también es importante para el rango de destino. Además, la presente invención reconoce el hecho de que pueden ocurrir diferentes situaciones de correlación en el rango de origen y en el rango de destino. Cuando se considera, por ejemplo, una señal de voz con ruido de alta frecuencia, puede ocurrir que la banda de baja frecuencia que comprende la señal de voz con un pequeño número de armónicos superiores esté altamente correlacionada en el canal izquierdo y el canal derecho, cuando el altavoz se coloca en el medio. La porción de alta frecuencia, sin embargo, puede estar fuertemente correlacionada debido al hecho de que podría haber un ruido de alta frecuencia diferente en el lado izquierdo en comparación con otro ruido de alta frecuencia o ningún ruido de alta frecuencia en el lado derecho. Por lo tanto, cuando se realiza una operación simple de relleno de espacios que ignora esta situación, entonces la porción de alta frecuencia estaría correlacionada también, y esto podría generar graves artefactos de segregación espacial en la señal reconstruida. Con el fin de abordar esta cuestión, los datos paramétricos para una banda de reconstrucción o, en general, para el segundo conjunto de segundas porciones espectrales que tiene que reconstruirse utilizando un primer conjunto de primeras porciones espectrales, se calculan para identificar, ya sea una primera o una segunda representación de dos canales, diferente para la segunda porción espectral o, dicho de otra manera, para la banda de reconstrucción. Por lo tanto, en el lado del codificador se calcula una identificación de dos canales para las segundas porciones espectrales, es decir, para las porciones para las que se calcula adicionalmente la información de energía para las bandas de reconstrucción. A continuación, un regenerador de frecuencia en el lado del decodificador regenera una segunda porción espectral en función de una primera porción del primer conjunto de primeras porciones espectrales, es decir, el rango de origen y los datos paramétricos para la segunda porción tal como la información de energía de la envolvente espectral o cualquier otro dato de la envolvente espectral y, adicionalmente, dependiendo de la identificación de dos canales para la segunda porción, es decir, para esta banda de reconstrucción bajo reconsideración.

La identificación de dos canales se transmite preferentemente como una etiqueta para cada banda de reconstrucción y estos datos se transmiten desde un codificador a un decodificador y el decodificador a continuación decodifica la señal central indicada preferentemente por etiquetas calculadas para las bandas centrales. A continuación, en una implementación, la señal central se almacena en ambas representaciones estéreo (por

ejemplo, izquierda/derecha y media/lateral) y, para el relleno de recuadros de frecuencia de relleno inteligente de espacios (IGF), se elige la representación de recuadros de origen para adaptar la representación de recuadros de destino indicada por las etiquetas de identificación de dos canales para el relleno inteligente de espacios o las bandas de reconstrucción, es decir, para el rango de destino.

5 Se hace hincapié en que este procedimiento no solo funciona para las señales estéreo, es decir, por un canal izquierdo y el canal derecho, sino que también funciona para las señales multicanal. En el caso de las señales multicanal se pueden procesar varios pares de canales diferentes, por ejemplo como un canal izquierdo y un canal derecho como un primer par, un canal envolvente izquierdo y un canal envolvente derecho como el segundo par y
10 un canal central y un canal LFE como el tercer par. Otras formaciones de pares se pueden determinar para los formatos de canal de salida superiores tales como 7.1, 11.1 y así sucesivamente.

Un aspecto adicional se basa en el descubrimiento de que ciertas 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
15 este fin se lleva a cabo un análisis en el lado del codificador con el fin de encontrar el mejor candidato de región de origen de adaptación para una cierta región de destino. Se genera y transmite como información lateral al decodificador una información de adaptación que identifica para una región de destino una determinada región de origen junto con opcionalmente alguna información adicional. El decodificador aplica entonces una operación de relleno de recuadros de frecuencia utilizando la información de adaptación. Para este fin, el decodificador lee la
20 información de adaptación a partir de la corriente de datos o archivo de datos transmitidos y accede a la región de origen identificada para una determinada banda de reconstrucción y, si está indicado en la información de adaptación, realiza además algún procesamiento de estos datos de región de origen para generar datos espectrales en bruto para la banda de reconstrucción. Entonces, este resultado de la operación de llenado de recuadros de frecuencia, es decir, los datos espectrales en bruto para la banda de reconstrucción, se modela utilizando
25 información de la envolvente espectral con el fin de obtener finalmente una banda de reconstrucción que comprenda las primeras porciones espectrales así como también las porciones tonales. Estas porciones tonales, sin embargo, no se generan por el esquema de relleno de recuadros adaptativo, sino que estas primeras porciones espectrales se emiten por el decodificador de audio o el decodificador central directamente.

30 El esquema de selección de recuadros espectral adaptativo puede operar con baja granularidad. En esta implementación, una región de origen se subdivide en regiones de origen generalmente superpuestas y la región de destino o las bandas de reconstrucción están dadas por las regiones de destino de frecuencia que no se superponen. Entonces, las similitudes entre cada región de origen y cada región de destino se determinan en el lado del codificador y el mejor par de adaptación de una región de origen y la región de destino se identifica por la
35 información de adaptación y, en el lado del decodificador, la región de origen identificada en la información de adaptación se utiliza para la generación de datos espectrales en bruto 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 cierto retardo cuando las similitudes son máximas. Este retardo puede ser tan preciso como un intervalo de
40 frecuencia y permite incluso una mejor adaptación entre una región de origen y la región de destino.

Asimismo, además de solo identificar un mejor par de adaptación, este retardo de correlación también puede transmitirse dentro de la información de adaptación y, adicionalmente, incluso puede transmitirse un signo. Cuando
45 se determina que la señal es negativa en el lado del codificador, a continuación también se transmite una etiqueta de signo correspondiente dentro de la información de adaptación y, en el lado del decodificador, los valores espectrales región de la región de origen se multiplican por "-1" o, en una representación compleja, se "hacen girar" a 180 grados.

Una implementación adicional de la presente invención aplica una operación de blanqueo de recuadros. El blanqueo
50 de un espectro quita la información de envolvente espectral en bruto y enfatiza la estructura precisa 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 solo se blanquea el recuadro utilizando un procedimiento predefinido, se transmite una etiqueta de blanqueo que indica al decodificador que se aplicará el mismo proceso de blanqueo predefinido al recuadro de frecuencia dentro
55 de IGF.

En cuanto a la selección de recuadros, se prefiere utilizar 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
60 recuadro se modula adicionalmente a través de la multiplicación por una secuencia temporal alternativa de -1/1 para compensar la representación de frecuencia inversa de cualquier otra banda dentro de la Transformada de Coseno Discreta Modificada (MDCT). Además, el signo del resultado de la correlación se aplica cuando se genera el recuadro de frecuencia.

Asimismo, se prefiere utilizar el recorte y la estabilización de recuadros con el fin de asegurar que se eviten los artefactos creados por el cambio rápido de regiones de origen para la misma región de reconstrucción o la región de destino. Para ello se realiza 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

- 5 recuadro de origen puede retirarse del conjunto de recuadros de origen potenciales, ya que está muy correlacionado con otros recuadros de origen. Además, como un tipo de estabilización de selección de recuadros, se prefiere mantener el orden del cuadro anterior si ninguno de los recuadros de origen en el cuadro actual se correlaciona (mejor que un umbral determinado) con los recuadros de destino en el cuadro actual.
- 10 El sistema de codificación de audio codifica de manera eficiente las señales de audio arbitrarias en un rango amplio de tasas de bits. Por consiguiente, para las tasas altas de bits, el sistema de la invención converge con la transparencia, para tasas bajas de bits se reduce al mínimo la molestia perceptual. Por lo tanto, la parte principal de tasa de bits disponible se utiliza para codificar en forma de onda solo la estructura perceptualmente más relevante de la señal en el codificador, y los espacios espectrales resultantes se rellenan en el decodificador con el contenido
- 15 de la señal que se aproxima en líneas generales al espectro original. Un presupuesto de bits muy limitado se consume para controlar el denominado Relleno Inteligente de Espacios (IGF) basado en parámetros por información lateral especializada transmitida desde el codificador al decodificador.

A continuación, las formas de realización preferidas de la presente invención se describirán con referencia a los dibujos adjuntos, en los cuales:

- 20 La Fig. 1a ilustra un aparato para codificar una señal de audio;
La Fig. 1b ilustra un decodificador para decodificar una señal de audio codificada que concuerda con el codificador de la Fig. 1a;
- 25 La Fig. 2a ilustra una implementación preferida del decodificador;
La Fig. 2b ilustra una implementación preferida del codificador;
La Fig. 3a ilustra una representación esquemática de un espectro generado por el decodificador de dominio espectral de la Fig. 1b;
- 30 La Fig. 3b ilustra una tabla que indica la relación entre los factores de ajuste de escala para las bandas de factor de escala y las energías para las bandas de reconstrucción y la información de relleno de ruido para una banda de relleno de ruido;
- La Fig. 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;
- 35 La Fig. 4b ilustra una implementación de la funcionalidad de la Fig. 4a;
La Fig. 5a ilustra una funcionalidad de un codificador de la Transformada de Coseno Discreta Modificada (MDCT);
- La Fig. 5b ilustra una funcionalidad del decodificador con una tecnología de MDCT;
- 40 La Fig. 5c ilustra una implementación del regenerador de frecuencia;
La Fig. 6a ilustra un codificador de audio con la funcionalidad de modelado de ruido temporal/modelado de recuadro temporal;
- La Fig. 6b ilustra un decodificador con la tecnología de modelado de ruido temporal/modelado de recuadro temporal;
- La Fig. 6c ilustra una funcionalidad adicional de modelado de ruido temporal/modelado de recuadro temporal con un orden diferente del filtro de predicción espectral y el modelador espectral;
- 45 La Fig. 7a ilustra una implementación de la funcionalidad del modelado de recuadro temporal (TTS);
La Fig. 7b ilustra una implementación del decodificador que se adapta a la implementación del codificador de la Fig. 7a;
- La Fig. 7c ilustra un espectrograma de una señal original y una señal extendida sin Modelado de Recuadro Temporal (TTS);
- 50 La Fig. 7d ilustra una representación de frecuencia que ilustra la correspondencia entre las frecuencias de relleno inteligente de espacios y las energías de modelado de recuadro temporal;
- La Fig. 7e ilustra un espectrograma de una señal original y una señal extendida con Modelado de Recuadro Temporal (TTS);
- La Fig. 8a ilustra un decodificador de dos canales con regeneración de frecuencia;
- 55 La Fig. 8b ilustra una tabla que ilustra combinaciones diferentes de las representaciones y los rangos de origen/destino;
- La Fig. 8c ilustra un diagrama de flujo que ilustra la funcionalidad del decodificador de dos canales con regeneración de frecuencia de la Fig. 8a;
- La Fig. 8d ilustra una implementación más detallada del decodificador de la Fig. 8a;
- 60 La Fig. 8e ilustra una implementación de un codificador para el procesamiento de dos canales que se decodificará por el decodificador de la Fig. 8a;
- La Fig. 9a ilustra un decodificador con la tecnología de regeneración de frecuencia utilizando valores de energía para el rango de frecuencia de regeneración;
- La Fig. 9b ilustra una implementación más detallada del regenerador de frecuencia de la Fig. 9a;

	La Fig. 9c	ilustra un esquema que ilustra la funcionalidad de la Fig. 9b;
	La Fig. 9d	ilustra una implementación adicional del decodificador de la Fig. 9a;
	La Fig. 10a	ilustra un diagrama de bloques de un codificador que concuerda con el decodificador de la Fig. 9a;
5	La Fig. 10b	ilustra un diagrama de bloques para ilustrar una funcionalidad adicional del calculador de parámetros de la Fig. 10a;
	La Fig. 10c	ilustra un diagrama de bloques para ilustrar una funcionalidad adicional del calculador de parámetros de la Fig. 10a;
	La Fig. 10d	ilustra un diagrama de bloques que ilustra una funcionalidad adicional del calculador de parámetros de la Fig. 10a;
10	La Fig. 11a	ilustra un decodificador adicional que tiene una identificación específica del rango de origen para una operación de relleno de recuadros espectrales en el decodificador;
	La Fig. 11b	ilustra la funcionalidad adicional del regenerador de frecuencia de la Fig. 11a;
	La Fig. 11c	ilustra un codificador utilizado para cooperar con el decodificador de la Fig. 11a;
15	La Fig. 11d	ilustra un diagrama de bloques de una implementación del calculador de parámetros de la Fig. 11c;
	La Fig. 12a y 12b	ilustran esquemas de frecuencia para ilustrar un rango de origen y un rango de destino;
	La Fig. 12c	ilustra un esquema de ejemplo de correlación de dos señales;
	La Fig. 13a	ilustra un codificador de la técnica anterior con extensión de ancho de banda; y
20	La Fig. 13b	ilustra un decodificador de la técnica anterior con extensión de ancho de banda.

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

Generalmente, una primera porción espectral tal como 306 de la Fig. 3a estará rodeada por dos segundas porciones espectrales tales como 307a, 307b. Esto no se aplica en la Codificación Avanzada de Audio de Alta Eficiencia (HE AAC), en donde el rango de frecuencia del codificador central es de banda limitada

La Fig. 1b ilustra un decodificador que concuerda con el codificador de la Fig. 1a. La primera representación codificada 107 se introduce en un decodificador de audio de dominio espectral 112 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 codificada 109 se introduce en un decodificador paramétrico 114 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 de frecuencia 116 para regenerar una segunda porción espectral reconstruida que tiene la primera resolución espectral que utiliza una primera porción espectral. El regenerador de frecuencia 116 realiza una operación de relleno de recuadros, es decir utiliza un recuadro o una porción del primer conjunto de primeras porciones espectrales y copia este primer conjunto de primeras porciones espectrales en el rango de reconstrucción o banda de reconstrucción que tiene la segunda porción espectral y generalmente lleva a cabo el modelado de la envolvente espectral u otra operación indicada por la segunda representación decodificada emitida por el decodificador paramétrico 114, es decir utilizando 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 de frecuencia 116 en la línea 117 se introduce en un convertidor de espectro-tiempo 118 configurado para convertir la primera representación decodificada y la segunda porción espectral reconstruida en una representación de tiempo 119, teniendo la representación de tiempo una tasa alta de muestreo determinada.

La Fig. 2b ilustra una implementación del codificador de la Fig. 1a. Una señal de entrada de audio 99 se introduce en

un banco de filtros de análisis 220 correspondiente al convertidor de espectro de tiempo 100 de la Fig. 1a. A continuación se realiza la operación de modelado de ruido temporal en el bloque de modelado de ruido temporal (TNS) 222. Por lo tanto, la entrada en el analizador espectral 102 de la Fig. 1a correspondiente a un enmascaramiento tonal de bloque 226 de la Fig. 2b pueden ser valores espectrales completos, cuando no se aplica la operación de modelado de ruido temporal/modelado de recuadro temporal o pueden ser valores residuales espectrales, cuando se aplica la operación de modelado de ruido temporal (TNS) como se ilustra en la Fig. 2b, bloque 222. Para las señales de dos canales o señales multicanal se puede realizar además una codificación conjunta de canal 228, por lo que el codificador de dominio espectral 106 de la Fig. 1a puede comprender el bloque de codificación conjunta de canal 228. Asimismo, se proporciona un codificador por entropía 232 para realizar una compresión de datos sin pérdidas que es también una porción del codificador de dominio espectral 106 de la Fig. 1a.

El analizador espectral/enmascaramiento tonal 226 separa la salida del bloque de modelado de ruido temporal (TNS) 222 en la banda central y los componentes tonales correspondientes al primer conjunto de primeras porciones espectrales 103 y los componentes residuales correspondientes al segundo conjunto de segundas porciones espectrales 105 de la Fig. 1a. El bloque 224 indicado como la codificación de extracción de parámetros de relleno inteligente de espacios (IGF) corresponde al codificador paramétrico 104 de la Fig. 1a y el multiplexor de la corriente de bits 230 corresponde al multiplexor de corriente de bits 108 de la Fig. 1a.

Preferentemente, el banco de filtros de análisis 222 se implementa como una MDCT (banco de filtros de la transformada de coseno discreta modificada) y la MDCT se utiliza para transformar la señal 99 en un dominio de frecuencia temporal con la transformada de coseno discreta modificada que actúa como la herramienta de análisis de frecuencia.

El analizador espectral 226 aplica preferentemente un enmascaramiento de tonalidad. Esta etapa de estimación de enmascaramiento de tonalidad se utiliza para separar los componentes tonales de los componentes similares a ruido en la señal. Esto permite que el codificador central 228 codifique todos los componentes tonales con un módulo psicoacústico. La etapa de estimación de enmascaramiento de tonalidad se puede implementar de muchas maneras diferentes y se implementa preferentemente de manera similar en su funcionalidad a la etapa de estimación de pista sinusoidal utilizada 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 utiliza una implementación que es fácil de implementar sin la necesidad de mantener trayectorias de nacimiento-muerte, pero también se puede utilizar cualquier otro detector de tonalidad o ruido.

El módulo de relleno inteligente de espacios (IGF) calcula la similitud que existe entre una región de origen y una región de destino. La región de destino estará representada por el espectro de la región de origen. La medida de la similitud entre las regiones de origen y de destino se realiza utilizando un enfoque de correlación cruzada. La región de destino se divide en n_{Tar} recuadros de frecuencia que no se superponen. Para cada recuadro en la región de destino, n_{Src} se crea un recuadro de origen a partir de una frecuencia fija de inicio. Estos recuadros de origen se superponen por un factor entre 0 y 1, en donde 0 significa el 0 % de solapamiento y 1 significa el 100 % de solapamiento. Cada uno de estos recuadros de origen está correlacionado con el recuadro de destino en diversos retardos para encontrar el recuadro de origen que se adapte mejor al recuadro de destino. El mejor número de recuadro de adaptación 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, el recuadro de origen debe multiplicarse por -1 antes del proceso de relleno de recuadro en el decodificador. El módulo de relleno inteligente de espacios (IGF) también se encarga de no sobrescribir los componentes tonales en el espectro ya que los componentes tonales se conservan utilizando el enmascaramiento de tonalidad. Se utiliza un parámetro de energía por bandas para almacenar la energía de la región de destino que permita reconstruir el espectro con precisión.

Este método tiene ciertas ventajas en comparación con la SBR clásica [1] en donde la cuadrícula de armónicos de una señal multitono se conserva por el codificador central en tanto que solo los espacios entre las sinusoides se llenan con el mejor "ruido modelado" de adaptación 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 la señal que crea las porciones importantes de la señal en el decodificador. En cambio, esta tarea es asumida por el codificador central, lo que permite la conservación de los componentes importantes del espectro. Otra ventaja del sistema propuesto es la escalabilidad continua que ofrecen las características. Solo el uso de $tileNum[idx_tar]$ y $xcorr_lag = 0$, para cada recuadro se denomina adaptación de granularidad en bruto y se puede utilizar para tasas bajas de bits mientras que el uso de la variable $xcorr_lag$ para cada recuadro permite adaptar mejor los espectros de destino y de origen.

Además, se propone una técnica de estabilización de elección de recuadros que elimina los artefactos de dominio de frecuencia tales como trino 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 rango de destino, la señal puede ser una fuente de sonido en panorámica y 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 adaptan a las regiones de destino, la imagen espacial puede sufrir debido a las regiones de origen no correlacionadas. El codificador analiza cada banda de energía de la región de destino, por lo general realizando una correlación cruzada de los valores espectrales y si se supera un determinado umbral, establece una etiqueta conjunta 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 esta etiqueta conjunta de estéreo. En caso de que se establezca la etiqueta de estéreo conjunto, tanto las energías como el emparchado se realizan en el dominio de estéreo conjunto. La información de estéreo conjunto para las regiones de relleno inteligente de espacios (IGF) se señala de manera similar a la información conjunta de estéreo para la codificación central, que incluye una etiqueta que indica en el caso de la predicción si la dirección 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.

$$midNrg[k] = leftNrg[k] + rightNrg[k];$$

$$sideNrg[k] = leftNrg[k] - rightNrg[k];$$

en donde k es 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 las bandas en donde el estéreo conjunto está activo, por lo que no es necesaria la transformación de energía adicional en el lado del decodificador.

Los recuadros de origen se crean siempre de acuerdo con la Matriz Media/Lateral:

$$midTile[k] = 0,5 \cdot (leftTile[k] + rightTile[k])$$

$$sideTile[k] = 0,5 \cdot (leftTile[k] - rightTile[k])$$

Ajuste de energía:

$$midTile[k] = midTile[k] * midNrg[k];$$

$$sideTile[k] = sideTile[k] * sideNrg[k];$$

Transformación conjunta de estéreo → 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 la dirección señalada es del medio al lateral:

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

Si la dirección señalada es del lateral al medio:

$$\begin{aligned} midTile1[k] &= midTile[k] - predictionCoeff \cdot sideTile[k] \\ leftTile[k] &= midTile1[k] - sideTile[k] \\ rightTile[k] &= midTile1[k] + sideTile[k] \end{aligned}$$

Este procesamiento asegura que a partir de los recuadros utilizados para regenerar regiones de destino altamente correlacionadas y regiones de destino en panorámica, los canales izquierdo y derecho resultantes siguen representando una fuente de sonido correlacionada y en panorámica incluso si las regiones de origen no están correlacionadas, conservando la imagen estéreo para dichas regiones.

En otras palabras, en la corriente de bits se transmiten etiquetas conjuntas de estéreo que indican si se utilizará L/R o M/S como un ejemplo para la codificación conjunta de estéreo general. En el decodificador, en primer lugar, la señal central se decodifica como lo indican las etiquetas conjuntas de estéreo para las bandas centrales. En segundo lugar, la señal central se almacena en ambas representaciones L/R y M/S. Para el relleno de recuadros del relleno inteligente de espacios (IGF), la representación de recuadros de origen se selecciona para ajustar la representación de recuadros de destino como lo indica la información conjunta de estéreo para las bandas de IGF.

El modelado de ruido temporal (TNS) es una técnica estándar y forma parte de la Codificación Avanzada de Audio (AAC) [11 - 13]. El modelado de ruido temporal (TNS) se puede considerar como una extensión del esquema básico de un codificador perceptual, mediante la inserción de un paso de procesamiento opcional entre el banco de filtros y la etapa de cuantificación. La tarea principal del módulo de modelado de ruido temporal (TNS) consiste en ocultar el ruido de cuantificación producido en la región de enmascaramiento temporal de señales de transición y, por lo tanto, produce un esquema de codificación más eficiente. En primer lugar, el modelado de ruido temporal (TNS) calcula un conjunto de coeficientes de predicción utilizando "predicción directa" en el dominio de la transformada, por ejemplo, la Transformada de Coseno Discreta Modificada (MDCT). Estos coeficientes se utilizan a continuación para aplanar la envolvente temporal de la señal. A medida que la cuantificación afecta el espectro filtrado del modelado de ruido temporal (TNS), también el ruido de cuantificación es temporalmente plano. Mediante la aplicación del filtrado inverso del modelado de ruido temporal (TNS) en el lado del decodificador, el ruido de cuantificación se modela de acuerdo con la envolvente temporal del filtro de TNS y, por lo tanto, el ruido de cuantificación se enmascara por el transitorio.

El relleno de inteligente de espacios (IGF) se basa en una representación de MDCT. Preferentemente, para la codificación eficiente se tienen que utilizar bloques largos de aproximadamente 20 ms. Si la señal dentro de dicho bloque largo contiene transitorios, en las bandas espectrales del relleno inteligente de espacios (IGF) ocurren pre- y post-ecos audibles debido al relleno de recuadros. La Fig. 7c muestra un efecto de pre-eco típico antes del inicio del transitorio debido al relleno inteligente de espacios (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 extendida de ancho de banda sin filtrado del modelado de ruido temporal (TNS).

Este efecto de pre-eco se reduce utilizando TNS en el contexto del relleno inteligente de espacios (IGF). En esta instancia, el TNS se utiliza como una herramienta de modelado de recuadro temporal (TTS) ya que la regeneración espectral en el decodificador se lleva a cabo sobre la señal residual del TNS. Los coeficientes de predicción del TTS requeridos se calculan y aplican utilizando el espectro completo en el lado del codificador como es habitual. Las frecuencias de inicio y fin del modelado de ruido temporal (TNS)/modelado de recuadro temporal (TTS) no resultan afectadas por la frecuencia de inicio del relleno inteligente de espacios (IGF) $f_{IGFstart}$ de la herramienta de IGF. En comparación con el modelado de ruido temporal (TNS) de la técnica anterior, la frecuencia de fin del modelado de recuadro temporal (TTS) aumenta a la frecuencia de fin de la herramienta de relleno inteligente de espacios (IGF), que es mayor que $f_{IGFstart}$. En el lado del decodificador se aplican los coeficientes de TNS/TTS sobre el espectro completo nuevamente, es decir el espectro central más el espectro regenerado más los componentes tonales del mapa de tonalidad (véase la Fig. 7e). La aplicación de modelado de recuadro temporal (TTS) es necesaria para formar la envolvente temporal del espectro regenerado para adaptarse a la envolvente de la señal original nuevamente. Por lo tanto, los pre-ecos ilustrados se reducen. Adicionalmente, todavía modela el ruido de cuantificación en la señal por debajo $f_{IGFstart}$ como es habitual en el modelado de ruido temporal (TNS).

En los decodificadores de la técnica anterior, el emparchado espectral en una señal de audio altera la correlación espectral en los límites de parche y, por lo tanto, afecta la envolvente temporal de la señal de audio introduciendo dispersión. Por lo tanto, otra ventaja de la aplicación del relleno de recuadros del relleno inteligente de espacios (IGF) en la señal residual es que, después de la aplicación del filtro de modelado, los límites del recuadro se correlacionan perfectamente, lo que resulta en una reproducción temporal más fiel de la señal.

En un codificador de la invención, el espectro que ha sido sometido al filtrado de TNS/TTS, el procesamiento de enmascaramiento de tonalidad y la estimación de parámetros de relleno inteligente de espacios (IGF), carece de cualquier señal por encima de la frecuencia de inicio de IGF excepto los componentes tonales. Este espectro disperso está codificado ahora por el codificador central utilizando los principios de codificación aritmética y codificación predictiva. Estos componentes codificados junto con los bits de señalización forman la corriente de bits del audio.

La Fig. 2a ilustra la implementación correspondiente del decodificador. La corriente de bits en la Fig. 2a

correspondiente a la señal de audio codificada se introduce en el demultiplexor/decodificador que estaría conectado, con respecto a la Fig. 1b, a los bloques 112 y 114. El demultiplexor de corriente de bits separa la señal de audio de entrada en la primera representación codificada 107 de la Fig. 1b y la segunda representación codificada 109 de la Fig. 1b. La primera representación codificada que tiene el primer conjunto de primeras porciones espectrales se introduce en el bloque de decodificación conjunta de canales 204 correspondiente al decodificador de dominio espectral 112 de la Fig. 1b. La segunda representación codificada se introduce en el decodificador paramétrico 114 que no se ilustra en la Fig. 2a y a continuación se introduce en el bloque de relleno inteligente de espacios (IGF) 202 correspondiente al regenerador de frecuencia 116 de la Fig. 1b. El primer conjunto de primeras porciones espectrales necesario para la regeneración de frecuencia se introduce en el bloque de IGF 202 a través de la línea 203. Asimismo, después de la decodificación conjunta de canales 204, la decodificación central específica se aplica en el bloque de enmascaramiento tonal 206 de manera que la salida del enmascaramiento tonal 206 corresponda a la salida del decodificador de dominio espectral 112. A continuación, el combinador 208 lleva a cabo una combinación, es decir una construcción de cuadros en donde la salida del combinador 208 tiene ahora el espectro de rango completo, pero todavía en el dominio filtrado de modelado de ruido temporal (TNS)/modelado de recuadro temporal (TTS). Posteriormente, en el bloque 210 se realiza una operación inversa de TNS/TTS utilizando información de filtro de TNS/TTS proporcionada a través de la línea 109, es decir la información lateral de TTS está incluida preferentemente en la primera representación codificada generada por el codificador de dominio espectral 106 que puede ser, por ejemplo, un codificador central de codificación avanzada de audio (AAC) directa o codificación unificada de voz y audio (USAC), o puede estar incluida también 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 rango 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 de filtros de síntesis 212 para obtener finalmente la señal de salida de audio.

La Fig. 3a ilustra una representación esquemática del espectro. El espectro se subdivide en bandas de factor de escala SCB en donde hay siete bandas de factor de escala SCB1 a SCB7 en el ejemplo ilustrado de la Fig. 3a. Las bandas de factor de escala pueden ser bandas de factor de escala de codificación avanzada de audio (AAC) que están definidas en la norma AAC y tienen un ancho de banda cada vez mayor hasta frecuencias superiores como se ilustra en la Fig. 3a esquemáticamente. Se prefiere realizar el relleno inteligente de espacios (IGF) 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 central se extiende desde la frecuencia más baja a la frecuencia de inicio de IGF. Por encima de la frecuencia de inicio de IGF se aplica el análisis de espectro para separar los componentes espectrales de alta resolución 304, 305, 306, 307 (el primer conjunto de primeras porciones espectrales) de componentes de baja resolución representados por el segundo conjunto de segundas porciones espectrales. La Fig. 3a ilustra un espectro que se introduce a modo de ejemplo en el codificador de dominio espectral 106 o en el codificador conjunto de canales 228, es decir, el codificador central opera en todo el rango pero codifica una cantidad significativa de valores espectrales cero, es decir estos valores espectrales cero se cuantifican a cero o se fijan en cero antes de la cuantificación o después de la cuantificación. De todos modos, el codificador central opera en el rango completo, es decir, como si el espectro fuera como el ilustrado, es decir, el decodificador central no necesariamente tiene que estar al tanto de cualquier relleno inteligente de espacios del segundo conjunto de segunda porciones espectrales con una resolución espectral inferior.

Preferentemente, la alta resolución está definida por una codificación por líneas de líneas espectrales, tales como las líneas de la transformada de coseno directa modificada (MDCT), mientras que la segunda resolución o baja resolución se define, por ejemplo, calculando solo un único valor espectral por banda de factor de escala, en donde una banda de factor de escala abarca varias líneas de frecuencia. Por lo tanto, la segunda resolución baja, con respecto a su resolución espectral, mucho menor que la primera o alta resolución definida por la codificación por líneas se aplica generalmente por el codificador central tal como un codificador central de codificación avanzada de audio (AAC) o codificación unificada de voz y audio (USAC).

En cuanto al factor de ajuste de escala o cálculo de energía, la situación se ilustra en la Fig. 3b. Debido al hecho de que el codificador es un codificador central y debido al hecho de que puede haber, pero no necesariamente, componentes del primer conjunto de porciones espectrales en cada banda, el codificador central calcula un factor de ajuste de escala para cada banda no solo en el rango central por debajo de la frecuencia de inicio de relleno inteligente de espacios (IGF) 309, 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 tonales codificadas 302, 304, 305, 306, 307 de la Fig. 3a y, en esta forma de realización junto con los factores de ajuste 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 ajuste de escala SF4 a SF7.

En particular, cuando el codificador central se encuentra en una condición de baja tasa de bits se puede aplicar además la operación de relleno de ruido adicional en la banda central, es decir una frecuencia inferior a la frecuencia de inicio del relleno inteligente de espacios (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 del decodificador, estos valores espectrales cuantificados a cero se re-sintetizan y los valores espectrales re-sintetizados se ajustan en su magnitud utilizando una energía de relleno de ruido tal como NF_2 ilustrada en 308 en la Fig. 3b. La energía de relleno de ruido, que se puede dar en términos absolutos o en términos relativos particularmente con respecto al factor de ajuste de escala como en la codificación unificada de voz y audio (USAC) corresponde a la energía del conjunto de valores espectrales cuantificados a cero. Estas líneas espectrales de relleno de ruido también pueden considerarse un tercer conjunto de terceras porciones espectrales que se regeneran por la síntesis de relleno de ruido simple sin ninguna operación de relleno inteligente de espacios (IGF) basada en la regeneración de frecuencia utilizando recuadros de frecuencia de otras frecuencias para la reconstrucción de recuadros de frecuencia utilizando valores espectrales de un rango 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 formas de realización 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 solo se transmite un único valor de información de energía, pero incluso en esta forma de 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 nuevos cálculos o cálculos de sincronización, y esto puede tener sentido en función de la aplicación determinada.

Preferentemente, el codificador de dominio espectral 106 de la Fig. 1a es un codificador activado psicoacústicamente como se ilustra en la Fig. 4a. Generalmente, como se ilustra por ejemplo en la norma MPEG2/4 AAC o MPEG1/2, Capa 3, la señal de audio para codificar después de haberse transformado en el rango espectral (401 en la Figura. 4a) se envía a un calculador de factor de escala 400. El calculador de factor de ajuste de escala se controla mediante un modelo psicoacústico que recibe adicionalmente la señal de audio para cuantificar 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, los factores de ajuste de escala a continuación se ajustan, por la cooperación bien conocida de los bucles de iteración interna y externa o por cualquier otro procedimiento de codificación adecuado de manera que se cumplan determinadas condiciones de tasas de bits. A continuación, los valores espectrales para cuantificar por un lado, y los factores de ajuste de escala calculados por otro lado se introducen en un procesador cuantificador 404. En la operación de codificación de audio simple, los valores espectrales para cuantificar se ponderan por los factores de ajuste de escala y los valores espectrales ponderados a continuación se introducen en un cuantificador fijo que generalmente tiene una funcionalidad de compresión para rangos de amplitud superiores. Entonces, en la salida del procesador cuantificador sí existen índices de cuantificación que a continuación se envían a un codificador por entropía que generalmente tiene codificación específica y muy eficiente para un conjunto de índices de cuantificación cero para valores de frecuencia adyacentes o, como también se denomina en la técnica, una "serie" de valores cero.

En el codificador de audio de la Fig. 1a, sin embargo, el procesador cuantificador generalmente recibe información sobre las segundas porciones espectrales del analizador espectral. Por lo tanto, el procesador cuantificador 404 se asegura de que, en la salida del procesador cuantificador 404, las segundas porciones espectrales identificadas por el analizador espectral 102 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 cero en el espectro.

La Fig. 4b ilustra una implementación del procesador cuantificador. Los valores espectrales de la transformada de coseno discreta modificada (MDCT) se pueden introducir en un bloque fijado en cero 410. Posteriormente, las segundas porciones espectrales ya se fijan en cero antes de realizar una ponderación por los factores de ajuste de escala en el bloque 412. En una implementación adicional, el bloque 410 no se proporciona, pero la cooperación fijada en cero se realiza en el bloque 418 posterior al bloque de ponderación 412. Incluso en otra implementación, la operación fijada en cero también se puede realizar en un bloque fijado en cero 422 posterior a una cuantificación en el bloque cuantificador 420. En esta implementación, los bloques 410 y 418 no estarían presentes. En general se proporcionan 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 Fig. 3a. Este espectro cuantificado se introduce entonces en un codificador por entropía tal como 232 en la Fig. 2b que puede ser un codificador Huffman o un codificador aritmético como se define, por ejemplo, en la norma de codificación unificada de voz y audio (USAC).

Los bloques fijados en cero 410, 418, 422, que se proporcionan alternativamente entre sí o en paralelo se controlan por el analizador espectral 424. 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 para codificar con una alta resolución y componentes para codificar con una baja

resolución. Otros de estos algoritmos implementados en el analizador espectral pueden ser un detector de actividad de voz, un detector de ruido, un detector de voz o cualquier otro detector que determine, en función de la información espectral o metadatos asociados, los requisitos de resolución para diferentes porciones espectrales.

5 La Fig. 5a ilustra una implementación preferida del convertidor de espectro de tiempo 100 de la Fig. 1a como se implementa, por ejemplo, en la codificación avanzada de audio (AAC) o en la codificación unificada de voz y audio (USAC). El convertidor de espectro de tiempo 100 comprende un divisor de ventanas 502 controlado por un detector de transitorios 504. Cuando el detector de transitorios 504 detecta un transitorio, entonces señala un intercambio de
10 ventanas largas a ventanas cortas al divisor de ventanas. A continuación, el divisor de ventanas 502 calcula para los bloques superpuestos, cuadros divididos en ventanas, teniendo cada cuadro dividido en ventanas normalmente dos valores N como, por ejemplo, los valores 2048. A continuación se realiza una transformación dentro de un transformador de bloques 506, y generalmente este transformador de bloques proporciona además una eliminación de manera que realiza una eliminación/transformada combinada para obtener un cuadro espectral con valores N tales como los valores espectrales de la transformada de coseno discreta modificada (MDCT). Por lo tanto, para una
15 operación de ventanas largas, el cuadro en la entrada del bloque 506 comprende dos valores N como, por ejemplo, 2048 valores y un cuadro espectral entonces tiene 1024 valores. Sin embargo, a continuación, se realiza un intercambio en los bloques cortos, es decir cuando se realizan ocho bloques cortos teniendo cada bloque corto $1/8$ valores de dominio temporal 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 esta decimación se combina
20 con una operación del 50 % de solapamiento del divisor de ventanas, el espectro es una versión muestreada críticamente de la señal de audio de dominio temporal 99.

Posteriormente se hace referencia a la Fig. 5b que ilustra una implementación específica del regenerador de frecuencia 116 y el convertidor de tiempo de espectro 118 de la Fig. 1b, o de la operación combinada de los bloques
25 208, 212 de la Fig. 2a. En la Fig. 5b se ilustra una banda de reconstrucción específica tal como la banda de factor de escala de la Fig. 3a. La primera porción espectral en esta banda de reconstrucción, es decir, la primera porción espectral 306 de la Fig. 3a se introduce en el bloque constructor/regulador de cuadros 510. Asimismo, una segunda porción espectral reconstruida para la banda de factor de escala 6 se introduce también en el constructor/regulador de cuadros 510. Además, la información de la energía tal como E_3 de la Fig. 3b para una banda de factor de escala
30 6 también se introduce en el bloque 510. La segunda porción espectral reconstruida en la banda de reconstrucción ya ha sido generada por el relleno de recuadros de frecuencia utilizando un rango de origen y la banda de reconstrucción a continuación corresponde al rango de destino. En esta instancia se lleva a cabo un ajuste de energía del cuadro para obtener finalmente el cuadro reconstruido completo que tiene los valores N como, por ejemplo, los que se obtienen en la salida del combinador 208 de la Fig. 2a. A continuación, en el bloque 512 se lleva a cabo una transformada/interpolación inversa de bloques para obtener 248 valores de dominio temporal para los
35 124 valores espectrales, por ejemplo, en la entrada del bloque 512. A continuación, en el bloque 514 se realiza una operación de síntesis de división de ventanas que está controlada nuevamente por una indicación de ventana larga/ventana corta transmitida como información lateral en la señal de audio codificada. A continuación, en el bloque 516 se realiza una operación de solapamiento/adición con un cuadro de tiempo anterior. Preferentemente, la transformada de coseno discreta modificada (MDCT) aplica un solapamiento del 50 % de manera que, para cada nuevo cuadro de tiempo de valores $2N$ se emiten finalmente los valores de dominio temporal N . Se prefiere un solapamiento del 50 % debido al hecho de que proporciona un muestreo crítico y un cruce continuo de un cuadro al cuadro siguiente debido a la operación de solapamiento/adición del bloque 516.

45 Tal como se ilustra en 301 en la Fig. 3a se puede aplicar adicionalmente una operación de relleno de ruido, no solo por debajo de la frecuencia de inicio de relleno inteligente de espacios (IGF) sino también por encima de la frecuencia de inicio de IGF como para la banda de reconstrucción contemplada coincidiendo con la banda de factor de escala 6 de la Fig. 3a. A continuación, los valores espectrales de relleno de ruido también se pueden introducir en el constructor/regulador de cuadros 510 y el ajuste de los valores espectrales de relleno de ruido también se puede
50 aplicar dentro de este bloque o los valores espectrales de relleno de ruido ya se pueden ajustar utilizando la energía de relleno de ruido antes de introducirse en el constructor/regulador de cuadros 510.

Preferentemente, una operación de IGF, es decir una operación de relleno de recuadros de frecuencia que utiliza valores espectrales de otras porciones se puede aplicar en el espectro completo. Por lo tanto, una operación de
55 relleno de recuadros espectrales no solo se puede aplicar en la banda alta por encima de una frecuencia de inicio de relleno inteligente de espacios (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 solo por debajo de la frecuencia de inicio de relleno inteligente de espacios (IGF) sino también por encima de la frecuencia de inicio de IGF. Sin embargo, se ha descubierto que la alta calidad y la alta eficiencia de la codificación de audio se pueden obtener cuando la
60 operación de relleno de ruido está limitada al rango de frecuencia por debajo de la frecuencia de inicio de IGF y cuando la operación de relleno de recuadros de frecuencia está limitada al rango de frecuencia por encima de la frecuencia de inicio de IGF, como se ilustra en la Fig. 3a.

Preferentemente, los recuadros de destino (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 recuadros de origen (ST) debe corresponder al tamaño del recuadro de destino (TT) asociado. Esto se demuestra utilizando el siguiente ejemplo.

5 TT[0] tiene una longitud de 10 Intervalos de MDCT. Esto corresponde exactamente a la longitud de dos SBC posteriores (tal como 4 + 6). Entonces, todos los recuadros de origen (ST) posibles que deben correlacionarse con TT[0], también tienen una longitud de 10 intervalos. Un segundo recuadro de destino TT[1] que es adyacente a TT[0] tiene una longitud de 15 intervalos 1 (SCB tiene una longitud de 7 + 8). Entonces, el recuadro de origen (ST) para lo anterior tiene una longitud de 15 intervalos en lugar de 10 intervalos como para TT[0].

10 En caso de que no se pueda encontrar un recuadro de destino (TT) para un recuadro de origen (ST) con la longitud del recuadro de destino (por ejemplo, cuando la longitud del TT es mayor que el rango de origen disponible), entonces no se calcula una correlación y el rango de origen se copia un número de veces en este TT (la copia se realiza una después de la otra de manera que una línea de frecuencia para la frecuencia más baja de la segunda copia sigue inmediatamente - en la frecuencia - la línea de frecuencia para la frecuencia más alta de la primera copia), hasta que el recuadro de destino (TT) se rellene completamente.

Posteriormente se hace referencia a la Fig. 5c que ilustra una forma de realización preferida adicional del regenerador de frecuencia 116 de la Fig. 1b o el bloque de relleno inteligente de espacios (IGF) 202 de la Fig. 2a. El bloque 522 es un generador de recuadros de frecuencia que no solo recibe una ID de la banda de destino sino que además recibe una ID de la banda de origen. A modo de ejemplo, se ha determinado en el lado del codificador que la banda de factor de escala 3 de la Fig. 3a es muy adecuada para la reconstrucción de la banda de factor de escala 7. Por lo tanto, la ID de la banda de origen sería 2 y la ID de la banda de destino sería 7. Basándose en esta información, el generador de recuadros de frecuencia 522 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 en bruto de los componentes espectrales 523. La segunda porción en bruto de los componentes espectrales tiene una resolución de frecuencia idéntica a la resolución de frecuencia incluida en el primer conjunto de primeras porciones espectrales.

30 Entonces, la primera porción espectral de la banda de reconstrucción tal como 307 de la Fig. 3a se introduce en un constructor de cuadros 524 y la segunda porción en bruto 523 se introduce también en el constructor de cuadros 524. A continuación, el cuadro reconstruido se ajusta por el regulador 526 utilizando un factor de ganancia para la banda de reconstrucción calculada por el calculador de factor de ganancia 528. Es importante destacar, sin embargo, que la primera porción espectral en el cuadro no resulta afectada por el regulador 526, sino que solo la segunda porción en bruto para el cuadro de reconstrucción resulta afectada por el regulador 526. Para este fin, el calculador del factor de ganancia 528 analiza la banda de origen o la segunda porción en bruto 523 y además analiza la primera porción espectral en la banda de reconstrucción para encontrar finalmente el factor de ganancia correcto 527 de manera que la energía del cuadro ajustado emitido por el regulador 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 la codificación avanzada de audio de alta eficiencia (HE-AAC). Esto se explica con respecto a la banda de factor de escala 7 en la Fig. 3a. Se supone que un codificador de la técnica anterior ilustrado en la Fig. 13a detectaría la porción espectral 307 a codificar con una alta resolución como un "armónico faltante". Entonces, la energía de este componente espectral se transmitiría junto con una información de la envolvente espectral para la banda de reconstrucción tal como la banda de factor de escala 7 al decodificador. A continuación, el decodificador recrearía el armónico faltante. Sin embargo, el valor espectral, en el que el armónico faltante 307 se reconstruiría por el decodificador de la técnica anterior de la Fig. 13b estaría en el medio de la banda 7 a una frecuencia indicada por la frecuencia de reconstrucción 390. Por lo tanto, la presente invención evita un error de frecuencia 391 que se introduciría por el decodificador de la técnica anterior de la Fig. 13d.

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 rango de reconstrucción una primera porción espectral que se adapte a la segunda porción espectral tanto como sea posible. Entonces, en esta implementación de rango de origen/rango de destino variable, el codificador paramétrico introducirá además en la segunda representación codificada una información de adaptación que indica un rango de origen de adaptación para cada rango de destino. En el lado del descodificador, esta información después podría utilizarse por un generador de recuadros de frecuencia 522 de la Fig. 5c que ilustra una generación de una segunda porción en bruto 523 basándose en una ID de la banda de origen y una ID de la banda de destino.

Asimismo, tal como se ilustra en la Fig. 3a, el analizador espectral está configurado para analizar la representación espectral hasta una frecuencia máxima de análisis que es solo 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.

Como se ilustra, el codificador opera sin reducción de muestreo y el decodificador opera sin muestreo ascendente. En otras palabras, 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.

Asimismo, tal como se ilustra en la Fig. 3a, el analizador espectral está configurado para analizar la representación espectral que se inicia con una frecuencia de relleno de espacios y que termina con una frecuencia máxima representada por una frecuencia máxima incluida en la representación espectral, en donde una porción espectral que se extiende desde una frecuencia mínima hasta la frecuencia de inicio de relleno de espacios pertenece al primer conjunto de porciones espectrales y en donde otra porción espectral tal como 304, 305, 306, 307 que tiene valores de frecuencia por encima de la frecuencia de relleno de espacios, está incluida adicionalmente en el primer conjunto de primeras porciones espectrales.

Como se explicó anteriormente, el decodificador de audio de dominio espectral 112 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 donde 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 ajuste de escala para la banda de factor de escala, que se genera y transmite sin importar si todos los valores espectrales en esta banda de factor de escala se fijan en cero o no, como se describe en el contexto de las Figs. 3a y 3b.

Por lo tanto, la invención es ventajosa 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 similar a ruido), por lo que la invención permite una reproducción de frecuencia precisa de componentes tonales. Hasta la fecha, ningún método del estado actual de la técnica aborda la representación paramétrica eficiente del contenido arbitrario de la señal por relleno de espacios espectrales sin la restricción de una división fija a priori en la banda baja (LF) y en la banda alta (HF).

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

El sistema en general consiste en:

- codificación central de banda completa
- relleno inteligente de espacios (relleno de recuadros o relleno de ruido)
- partes tonales dispersas en núcleo, seleccionadas por enmascaramiento tonal
- codificación conjunta de par de estéreo para la banda completa, incluyendo el relleno de recuadros
- TNS en el recuadro
- blanqueo espectral en el rango de relleno inteligente de espacios (IGF)

Un primer paso hacia un sistema más eficiente consiste en eliminar la necesidad de transformar datos espectrales en un segundo dominio de transformada diferente del dominio del codificador central. Como la mayoría de los códec de audio tal como, por ejemplo, la codificación avanzada de audio (AAC), utilizan la transformada de coseno discreta modificada (MDCT) como transformada básica, también es útil realizar la extensión de ancho de banda (BWE) en el dominio de la MDCT. Un segundo requisito para el sistema de BWE sería la necesidad de conservar la cuadrícula tonal mediante la cual se conservan incluso componentes tonales de alta frecuencia (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 extensión de ancho de banda (BWE) se propone un nuevo sistema denominado Relleno Inteligente de Espacios (IGF). La Fig. 2b muestra el diagrama de bloques del sistema propuesto en el lado del codificador y la Fig. 2a muestra el sistema en el lado del decodificador.

La Fig. 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 de audio de dominio espectral 602 para generar una primera representación decodificada de un primer conjunto de porciones espectrales y como el regenerador de frecuencia 604 conectado corriente abajo del decodificador de audio de dominio espectral 602 para generar una segunda porción espectral reconstruida utilizando 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 de predicción espectral 606. Este filtro de predicción inversa está configurado para realizar una predicción inversa sobre la

frecuencia utilizando los valores residuales espectrales para el primer conjunto de la primera frecuencia y las segundas porciones espectrales reconstruidas. El filtro de predicción inversa espectral 606 está configurado por la información de filtro incluida en la señal de audio codificada. La Fig. 6b ilustra una implementación más detallada de la forma de realización de la Fig. 6a. Los valores residuales de predicción espectral 603 se introducen en un generador de recuadros de frecuencia 612 que genera valores espectrales en bruto para una banda de reconstrucción o para una determinada segunda porción de frecuencia y estos datos en bruto que ahora tienen la misma resolución que la primera representación espectral de alta resolución se introducen en el modelador espectral 614. El modelador espectral ahora modela el espectro utilizando información de la envolvente transmitida en la corriente de bits y los datos modelados espectralmente a continuación se aplican al filtro de predicción espectral 616 generando finalmente un cuadro de valores espectrales completos utilizando la información de filtro 607 transmitida desde el codificador al decodificador a través de la corriente de bits.

En la Fig. 6b se supone que, en el lado del codificador, el cálculo de la información de filtro transmitida a través de la corriente de bits y utilizada a través de la línea 607 se realiza con posterioridad al cálculo de la información de la envolvente. Por lo tanto, en otras palabras, un codificador que concuerda con el decodificador de la Fig. 6b primero calcularía los valores residuales espectrales y a continuación calcularía la información de la envolvente con los valores residuales espectrales tal como se ilustra, por ejemplo, en la Fig. 7a. Sin embargo, la otra implementación también es útil para ciertas implementaciones, cuando la información de la envolvente se calcula antes de realizar el filtrado de modelado de ruido temporal (TNS) o de modelado de recuadro temporal (TTS) en el lado del codificador. A continuación, el filtro de predicción espectral 622 se aplica antes de realizar el modelado espectral en el bloque 624. Por lo tanto, en otras palabras, los valores espectrales (completos) se generan antes de aplicar la operación de modelado espectral 624.

Preferentemente se calcula un filtro de TNS o un filtro de TTS de valor complejo. Esto se ilustra en la Fig. 7a. La señal de audio original se introduce en un bloque de transformada de coseno discreta modificada (MDCT) compleja 702. A continuación se realiza 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 lateral del relleno inteligente de espacios (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 activado por modelo psicoacústico que se ilustra en 708 para obtener el primer conjunto de primeras porciones espectrales indicado en $X(k)$ en la Fig. 7a y todos estos datos se envían al multiplexor de corriente de bits 710.

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

A continuación, el bloque 724 se utiliza 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 espectros 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 el cuadro de banda completa obtenido completamente y relleno de recuadros se realiza la operación inversa de modelado de recuadro temporal (TTS) 728 y, en el lado del decodificador se realiza una operación inversa final de MDCT compleja en el bloque 730. Por lo tanto, el uso de la información de filtro de TNS complejo permite generar automáticamente, cuando se aplica no solo dentro de la banda central o dentro de las bandas de recuadros por separado sino también sobre los límites centrales/de recuadros o sobre los límites de recuadros/recuadros, un procesamiento de límites de recuadros que finalmente 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 realizando un ajuste de la envolvente espectral en estos datos en bruto de los recuadros de frecuencia.

La Fig. 7c ilustra una comparación de una señal original (panel izquierdo) y una señal extendida sin Modelado de Recuadro Temporal (TTS). Se puede observar que hay artefactos intensos ilustrados por las porciones ampliadas en el rango de frecuencias superior ilustrado en 750. Esto, sin embargo, no ocurre en la Fig. 7e cuando la misma porción espectral en 750 se compara con el componente relacionado con los artefactos 750 de la Fig. 7c.

Las formas de realización o el sistema de codificación de audio de la invención utilizan la parte principal de tasa de bits disponible para codificar en forma de onda solo la estructura perceptualmente más relevante de la señal en el codificador, y los espacios espectrales resultantes se rellenan en el decodificador con el contenido de la señal que se aproxima en líneas generales al espectro original. Un presupuesto de bits muy limitado se consume para controlar el así denominado Relleno de Espacios Inteligente (IGF) basado en parámetros por información lateral especializada transmitida desde el codificador al decodificador.

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 vieron obligados a reducir drásticamente el ancho de banda de audio transmitido cuando solo estaba disponible una tasa de bits muy baja. Los códec de audio modernos ahora son

capaces de codificar señales de banda ancha utilizando los métodos de extensión de ancho de banda (BWE) tales como la Replicación de Ancho de Banda Espectral (SBR, por sus siglas en inglés) [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) codificada en forma de onda de la señal decodificada por medio de transposición a la región espectral de alta frecuencia (HF) ("emparchado") y la aplicación de un procesamiento posterior basado en parámetros. En los esquemas de extensión de ancho de banda (BWE), la reconstrucción de la región espectral de alta frecuencia (HF) por encima de una denominada frecuencia de cruce determinada se basa a menudo en el emparchado espectral. En general, la región de alta frecuencia (HF) consiste en múltiples conexiones adyacentes y cada una de estas conexiones se obtiene de regiones de paso de banda (BP) del espectro de baja frecuencia (LF) por debajo de la frecuencia de cruce determinada. Los sistemas del estado actual de la técnica desempeñan con eficiencia el emparchado dentro de una representación de bancos de filtros copiando un conjunto de coeficientes de subbandas adyacentes desde una región de origen a la región de destino.

Si se implementa un sistema de BWE en un banco de filtros o el dominio de la transformada de tiempo-frecuencia, solo hay una posibilidad limitada para controlar la forma temporal de la señal de extensión de ancho de banda. Generalmente, la granularidad temporal está limitada por el tamaño de salto utilizado entre ventanas adyacentes de la transformada. Esto puede conducir a pre o post-ecos no deseados en el rango espectral de la extensión de ancho de banda (BWE).

En la codificación de audio perceptual se sabe que la forma de la envolvente temporal de una señal de audio se puede restaurar utilizando técnicas de filtrado espectral como el Modelado de la Envolvente Temporal (TNS) [14]. Sin embargo, el filtro del TNS conocido del estado actual de la técnica es un filtro de valor real en los espectros de valor real. Dicho filtro de valor real en los espectros de valor real puede verse seriamente afectado por artefactos de solapamiento, especialmente si la transformada real subyacente es una Transformada de Coseno Discreta Modificada (MDCT).

El modelado de recuadros de la envolvente temporal aplica filtrado complejo en espectros de valor complejo, tales como los obtenidos, por ejemplo, a través de una Transformada de Coseno Discreta Modificada Compleja (CMDCT). De esta manera se evitan los artefactos de solapamiento.

El modelado de recuadros temporal consiste en

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

La invención extiende la técnica del estado actual 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 de la dirección de la frecuencia, para el uso en una forma 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 Espacios (IGF), pero emplea una transformada de valor complejo (CMDCT) sobremuestreada en oposición a la configuración estándar de relleno inteligente de espacios (IGF) que se basa en una representación de la transformada de coseno discreta modificada (MDCT) críticamente muestreada de valor real de una señal. La CMDCT puede observarse como la combinación de los coeficientes de la MDCT en la parte real y los coeficientes de la transformada sinusoidal discreta modificada (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 utilizarse en combinación con cualquier método de extensión de ancho de banda (BWE) que se basa en una representación de bancos de filtros de la señal de audio.

En este nuevo contexto, la predicción lineal a lo largo de la dirección de la frecuencia no se utiliza 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 los componentes de la señal rellenos de recuadros se modelan temporalmente por TTS en comparación con el modelado de ruido de cuantificación por TNS en los códec de la transformada perceptual del estado actual de la técnica.

La Fig. 7a muestra un diagrama de bloques de un codificador de BWE que utiliza 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 temporal $x(n)$ para obtener la señal de dominio de frecuencia $X(k)$
- calcular el filtro de TTS de valor complejo

- obtener la información lateral para la BWE y eliminar la información espectral que tiene que replicarse por el decodificador
- aplicar la cuantificación utilizando el módulo psicoacústico (PAM, por sus siglas en inglés)
- almacenar/transmitir los datos, solo se transmiten los coeficientes de MDCT de valor real

5 La Fig. 7b muestra el decodificador correspondiente. Este invierte principalmente los pasos realizados en el codificador.

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

- 10
- estimar los coeficientes de la transformada sinusoidal discreta modificada (MDST) partir de los valores de la transformada de coseno discreta modificada (MDCT) (este procesamiento agrega un retardo de decodificador de bloques) y combinar los coeficientes de la MDCT y la MDST en coeficientes de la transformada de coseno discreta modificada compleja (CMDCT) de valor complejo
- 15
- realizar el relleno de recuadros con su procesamiento posterior
 - aplicar el filtrado de modelado de recuadro temporal (TTS) inverso con los coeficientes transmitidos del filtro de TTS
 - calcular la CMDCT inversa

20 Como alternativa, cabe destacar que 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 coherentemente en el codificador.

25 Para una codificación de transformada eficiente se deben utilizar preferentemente los denominados “bloques largos” de aproximadamente 20 ms para lograr una ganancia de transformada razonable. Si la señal dentro de dicho bloque largo contiene transitorios, en las bandas espectrales reconstruidas ocurren pre- y post-ecos audibles debido al relleno de recuadros. La Fig. 7c muestra efectos típicos de pre- y post-ecos que alteran los transitorios debido al relleno inteligente de espacios (IGF). En el panel izquierdo de la Fig. 7c se muestra el espectrograma de la señal original y en el panel derecho se muestra el espectrograma de la señal de relleno de recuadros sin el filtrado del modelado de recuadro temporal (TTS) de la invención. En este ejemplo, la frecuencia de inicio de IGF $f_{IGFstart}$ o f_{split} entre la banda central y la banda rellena de recuadros se selecciona como $f_s/4$. En el panel derecho de la Fig. 7c se observan distintos pre- y post-ecos alrededor de los transitorios, especialmente prominentes en el extremo espectral superior de la región de frecuencia replicada.

35 La tarea principal del módulo de TTS consiste en restringir estos componentes de señal no deseados en estrecha proximidad alrededor de un transitorio y de ese modo ocultarlos en la región temporal gobernada por el efecto de enmascaramiento temporal de la percepción humana. Por lo tanto, los coeficientes de predicción necesarios de TTS se calculan y aplican utilizando “predicción directa” en el dominio de la CMDCT.

40 En una forma de 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 se filtre completamente por un filtro de TTS (filtro de aplanamiento o modelado) o no. 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 Fig. 7d muestra un ejemplo de áreas operativas de TTS e IGF para un conjunto de tres filtros de TTS.

45 La frecuencia de fin del modelado de recuadro temporal (TTS) se ajusta a la frecuencia de fin de la herramienta de relleno inteligente de espacios (IGF), que es mayor que $f_{IGFstart}$. Si el modelado de recuadro temporal (TTS) utiliza más de un filtro tiene que asegurarse que la frecuencia de cruce entre dos filtros de TTS tiene que coincidir con la frecuencia dividida del relleno inteligente de espacios (IGF). De lo contrario, un sub-filtro de TTS se excederá del límite $f_{IGFstart}$ lo que producirá artefactos no deseados como, por ejemplo, sobremodelado.

50 En la variante de implementación representada en la Fig. 7a y en la Fig. 7b, se debe tener especial cuidado de que en ese decodificador, las energías de IGF estén ajustadas correctamente. Esto ocurre especialmente si, en el curso 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 de destino (como un filtro de modelado que no es la contrapartida exacta de dicho filtro de aplanamiento) de un recuadro del IGF. En este caso, la relación 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 relación.

60 En la variante de implementación alternativa, el orden de procesamiento posterior de IGF y TTS se invierte. En el decodificador, esto significa que el ajuste de energía por el procesamiento posterior de IGF se calcula después del filtrado de TTS y, de este modo, es el paso 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 se ajusta siempre correctamente por el procesamiento de IGF.

En el lado del decodificador se aplican los coeficientes de filtro de TTS en todo el espectro completo nuevamente, es decir el espectro central extendido por el espectro regenerado. La aplicación de modelado de recuadro temporal (TTS) es necesaria para formar la envolvente temporal del espectro regenerado para adaptarse a la envolvente de la señal original nuevamente. Por lo tanto, los pre-ecos ilustrados se reducen. Adicionalmente, todavía modela temporalmente el ruido de cuantificación en la señal por debajo $f_{IGFstart}$ como es habitual en el modelado de ruido temporal (TNS) de la técnica anterior.

En los codificadores de la técnica anterior, el emparchado espectral de una señal de audio (por ejemplo, la Replicación de Ancho de Banda Espectral (SBR)) altera la correlación espectral en los límites de emparchado y, por lo tanto, afecta la envolvente temporal de la señal de audio introduciendo dispersión. Por lo tanto, otra ventaja de la aplicación del relleno de recuadros del relleno inteligente de espacios (IGF) en la señal residual es que, después de la aplicación del filtro de modelado de recuadro temporal (TTS), los límites del recuadro se correlacionan perfectamente, lo que da como resultado una reproducción temporal más fiel de la señal.

El resultado de la señal procesada correspondiente se muestra en la Fig. 7e. En comparación, la versión sin filtrar (Fig. 7c, panel derecho) la señal filtrada de TTS muestra una buena reducción de pre- y post-ecos no deseados (Fig. 7e, panel derecho).

Asimismo, de acuerdo con la descripción, la Fig. 7a ilustra un codificador que concuerda con el decodificador de la Fig. 7b o el decodificador de la Fig. 6a. Básicamente, un aparato para codificar una señal de audio comprende un convertidor de espectro de tiempo 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 realizar una predicción sobre la frecuencia para generar valores residuales espectrales, en donde el filtro de predicción 704 se define por la información del filtro de predicción obtenida de la señal de audio y enviada a un multiplexor de corriente de bits 710, como se ilustra en 714 en la Fig. 7a. Asimismo, se proporciona un codificador de audio tal como el codificador de audio activado psicoacústicamente 704. 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 706 en la Fig. 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. Por último, tal como se ilustra en la Fig. 7a, se proporciona una interfaz de salida para emitir la señal codificada que comprende el segundo conjunto paramétricamente codificado de segundas porciones espectrales, el primer conjunto codificado de primeras porciones espectrales y la información del filtro ilustrada como "información lateral de modelado de recuadro temporal (TTS)" en 714 en la Fig. 7a.

Preferentemente, el filtro de predicción 704 comprende un calculador de información de filtro configurado para utilizar 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 utilizando los mismos valores espectrales de la representación espectral utilizada para calcular la información de filtro.

Preferentemente, el filtro de TTS 704 está configurado de la misma manera conocida para los codificadores de audio de la técnica anterior que aplican la herramienta de modelado de ruido temporal (TNS) de acuerdo con la norma de codificación avanzada de audio (AAC).

Posteriormente, una aplicación adicional que utiliza la decodificación de dos canales se analiza en el contexto de las Figs. 8a a 8e. Además, se hace referencia a la descripción de los elementos correspondientes en el contexto de las Figs. 2a, 2b (codificación conjunta de canales 228 y decodificación conjunta de canales 204).

La Fig. 8a ilustra un decodificador de audio para generar una señal decodificada de dos canales. El decodificador de audio comprende cuatro decodificadores de audio 802 para decodificar una señal codificada de dos canales para obtener un primer conjunto de primeras porciones espectrales y adicionalmente un decodificador paramétrico 804 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 diferente de dos canales para las segundas porciones espectrales. Adicionalmente, se proporciona un regenerador de frecuencia 806 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 Fig. 8b ilustra diferentes combinaciones para las representaciones de dos canales en el rango de origen y en el rango de destino. El rango de origen puede estar en la primera representación de dos canales y el rango de destino también puede estar en la primera representación de dos canales. Como alternativa, el rango de origen puede estar en la primera representación de dos canales y el rango de destino puede estar en la segunda representación de dos canales. Además, el rango de origen puede estar en la

segunda representación de dos canales y el rango de destino puede estar en la primera representación de dos canales como se indica en la tercera columna de la Fig. 8b. Por último, tanto el rango de origen como el rango de destino pueden estar en la segunda representación de dos canales. En una forma de realización, la primera representación de dos canales es una representación de dos canales por separado, en donde los dos canales de la señal de dos canales están representados individualmente. Entonces, la segunda representación de dos canales es una representación conjunta en donde los dos canales de la representación de dos canales están representados en forma conjunta, es decir, cuando un procesamiento posterior o la transformada de representación son necesarios para recalcular una representación de dos canales por separado que es necesaria para la salida a los 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 conjunta de estéreo. Sin embargo, otras representaciones de dos canales además de izquierda/derecha o M/S o la predicción estéreo se pueden aplicar y utilizar para la presente invención.

La Fig. 8c ilustra un diagrama de flujo para las operaciones realizadas por el decodificador de audio de la Fig. 8a. En un paso 812, el decodificador de audio 802 realiza una decodificación del rango de origen. El rango de origen puede comprender, con respecto a la Fig. 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, el paso 812 puede resultar en diferentes representaciones para diferentes bandas. Entonces, en el paso 814, el regenerador de frecuencia 806 está configurado para seleccionar un rango de origen para una regeneración de frecuencia. En el paso 816, el regenerador de frecuencia 806 comprueba entonces la representación del rango de origen y en el bloque 818, el regenerador de frecuencia 806 compara la representación de dos canales del rango de origen con la representación de dos canales del rango de destino. Si ambas representaciones son idénticas, el regenerador de frecuencia 806 proporciona una frecuencia de regeneración por separado para cada canal de la señal de dos canales. Cuando, sin embargo, ambas representaciones detectadas en el bloque 818 no son idénticas, entonces se toma el flujo de la señal 824 y el bloque 822 calcula la otra representación de dos canales del rango de origen y utiliza esta otra representación calculada de dos canales para la regeneración del rango de destino. Por lo tanto, el decodificador de la Fig. 8a hace que sea posible regenerar un rango de destino que se indica que tiene la segunda identificación de dos canales utilizando un rango de origen que está en la primera representación de dos canales. Naturalmente, la presente invención permite regenerar, además, un rango de destino utilizando un rango de origen que tiene la misma identificación de dos canales. Y, adicionalmente, la presente invención permite regenerar un rango de destino 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, necesaria para el almacenamiento o la transmisión a los altavoces correspondientes para la señal de dos canales.

Se hace hincapié en 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 descrito en el contexto de la Fig. 8a a 8e se puede realizar cuando 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). Cualquier otra formación de pares se puede utilizar con el fin de representar, por ejemplo, seis canales de entrada por tres procedimientos de procesamiento de dos canales.

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

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

El generador de recuadros de frecuencia 836 realiza esta operación para cada rango de destino 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. A continuación un regulador de envolventes 838 realiza un ajuste de la envolvente. El ajuste de envolvente se realiza en el dominio de dos canales identificado por la identificación de dos canales. Para este fin se requieren parámetros de ajuste de la envolvente y estos parámetros se transmiten desde el codificador al decodificador en la misma representación de dos canales descrita. Cuando la identificación de dos canales en el rango de destino para procesar por el regulador de envolventes tiene una identificación de dos canales que indica una representación de dos canales diferentes de los datos de la envolvente para este rango de destino, a continuación un transformador de parámetros 840 transforma los parámetros de la envolvente en la representación de dos canales requerida. Cuando, por ejemplo, la identificación de dos canales para una banda indica la codificación conjunta de estéreo y cuando los parámetros para este rango de destino se han transmitido como parámetros de la envolvente L/R, entonces el transformador de parámetros calcula los parámetros conjuntos de la envolvente estéreo a partir de los parámetros de la envolvente L/R descrita de manera que la representación paramétrica correcta se utiliza para el ajuste de la envolvente espectral de un rango de destino.

En otra forma de realización preferida, los parámetros de la envolvente ya se transmiten como parámetros conjuntos de estéreo cuando se utiliza el estéreo conjunto en una banda de destino.

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

Preferentemente, las segundas porciones espectrales corresponden a las bandas de frecuencia, y la identificación de dos canales se proporciona como una matriz de etiquetas correspondientes a la tabla de la Fig. 8b, en donde existe una etiqueta para cada banda de frecuencia. Entonces, el decodificador paramétrico está configurado para comprobar si la etiqueta se ha fijado o no y para controlar el regenerador de frecuencia 106 de acuerdo con una etiqueta para utilizar, ya sea una primera representación o una segunda representación de la primera porción espectral.

En una forma de realización, solo el rango de reconstrucción que se inicia con la frecuencia de inicio del relleno inteligente de espacios (IGF) 309 de la Fig. 3a tiene identificaciones de dos canales para diferentes bandas de reconstrucción. En otra forma de realización, esto también se aplica para el rango de frecuencia por debajo de la frecuencia de inicio de IGF 309.

En una forma de realización adicional, la identificación de la banda de origen y la identificación de la banda de destino 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 rango de origen a un rango de destino. Un rango de origen se puede utilizar para recrear, con respecto a la frecuencia, un rango de destino 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 utilizando dos o más operaciones de relleno de recuadros de frecuencia similares al procesamiento para múltiples parches conocidos a partir del procesamiento de codificación avanzada de audio (AAC) de alta eficiencia.

La Fig. 8e ilustra un codificador de audio para codificar una señal de audio de dos canales. El codificador comprende un convertidor de espectro de tiempo 860 para convertir la señal de audio de dos canales en una representación espectral. Asimismo, un analizador espectral 866 para convertir la señal de audio de canal de audio de dos canales en una representación espectral. Además se proporciona un analizador espectral 866 para realizar un análisis con el fin de determinar las porciones espectrales que se codificarán con una alta resolución, es decir para descubrir el primer conjunto de primeras porciones espectrales y para descubrir adicionalmente el segundo conjunto de segundas porciones espectrales.

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

Dependiendo del resultado del analizador de dos canales, una banda en la segunda representación espectral se parametriza utilizando la primera representación de dos canales o la segunda representación de dos canales, y esto

se realiza mediante un codificador de parámetros 868. El rango de frecuencia central, es decir la banda de frecuencia por debajo de la frecuencia de inicio del relleno inteligente de espacios (IGF) 309 de la Fig. 3a se codifica por un codificador central 870. El resultado de los bloques 868 y 870 se introduce en una interfaz de salida 872. Como se indicó anteriormente, el analizador de dos canales proporciona una identificación de dos canales para cada

5 banda, ya sea por encima de la frecuencia de inicio de IGF o para toda el rango de frecuencia, y esta identificación de dos canales también se envía a la interfaz de salida 872 de manera que estos datos también están incluidos en una señal codificada 873 emitida por la interfaz de salida 872.

Asimismo se prefiere que el codificador de audio comprenda un transformador por bandas 862. Basándose en la

10 decisión del analizador de dos canales 862, la señal de salida del convertidor de espectro de tiempo 862 se transforma en una representación indicada por el analizador de dos canales y, en particular, por la ID de dos canales 835. Por lo tanto, una salida del transformador por bandas 862 es un conjunto de bandas de frecuencia en donde cada banda de frecuencia puede estar en la primera representación de dos canales o en la segunda representación diferente de dos canales. Cuando se aplica la presente invención en banda completa, es decir cuando ambos

15 rangos, el rango de origen y el rango de reconstrucción, se procesan por el transformador por bandas, el analizador espectral 860 puede analizar esta representación. Como alternativa, sin embargo, el analizador espectral 860 también puede analizar la salida de la señal por el convertidor de espectro de tiempo indicado por la línea de control 861. Por lo tanto, el analizador espectral 860 puede aplicar el análisis de tonalidad preferido en la salida del transformador por bandas 862 o la salida del convertidor de espectro de tiempo 860 antes de haberse procesado por

20 el transformador por bandas 862. Asimismo, el analizador espectral puede aplicar la identificación del mejor rango de origen de adaptación para un cierto rango de destino, ya sea en el resultado del transformador por bandas 862 o en el resultado del convertidor de espectro de tiempo 860.

Posteriormente se hace referencia a las Figs. 9a a 9d para ilustrar un cálculo preferido de los valores de información de energía ya analizados en el contexto de la Fig. 3a y en la Fig. 3b.

25

Los codificadores de audio modernos del estado de la técnica aplican diversas técnicas para reducir al mínimo la cantidad de datos que representan una señal de audio determinada. Los codificadores de audio como, por ejemplo, la codificación unificada de voz y audio (USAC) [1] aplican una transformación de tiempo a frecuencia como la

30 transformada de coseno discreta modificada (MDCT) para obtener una representación espectral de una señal de audio determinada. Estos coeficientes de la MDCT se cuantifican aprovechando los aspectos psicoacústicos del sistema auditivo humano. Si se reduce la tasa de bits disponible, la cuantificación se vuelve más basta introduciendo un gran número de valores espectrales reducidos a cero que generan artefactos audibles en el lado del decodificador. Para mejorar la calidad de percepción, los decodificadores del estado de la técnica llenan estas partes

35 espectrales reducidas a cero con ruido aleatorio. El método de relleno inteligente de espacios (IGF) recolecta recuadros de la señal no cero restante para llenar esos espacios en el espectro. Es crucial para la calidad perceptual de la señal de audio decodificada que se conserve la envolvente espectral y la distribución de energía de los coeficientes espectrales. El método de ajuste de energía presentado en la presente invención utiliza la información lateral transmitida para reconstruir la envolvente espectral de la MDCT de la señal de audio.

Dentro de la replicación de ancho de banda espectral (eSBR) [15] la señal de audio se submuestra 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 por técnicas paramétricas, eSBR, en el lado del decodificador. La eSBR implica el uso de una

45 transformada adicional, la transformación de filtros espejo en cuadratura (QMF) que se utiliza para sustituir la parte de alta frecuencia vacía y para volver a muestrear la señal de audio [17]. Esto agrega complejidad computacional y consumo de memoria a un codificador de audio.

El codificador de USAC [15] ofrece la posibilidad de llenar huecos espectrales (líneas espectrales reducidas a cero) con ruido aleatorio pero presenta los siguientes inconvenientes: el ruido aleatorio no puede conservar la estructura

50 precisa temporal de una señal transitoria y no se puede conservar la estructura armónica de una señal tonal.

El área en donde opera la eSBR en el lado del decodificador fue completamente eliminada por el codificador [1]. Por lo tanto, la eSBR tiende a eliminar líneas tonales en la región de alta frecuencia o distorsionar las estructuras armónicas de la señal original. Como la resolución de frecuencia de filtros espejo en cuadratura (QMF) de la

55 replicación de ancho de banda espectral (eSBR) es muy baja y la reinserción de componentes sinusoidales solo es posible en la resolución gruesa del banco de filtros subyacente, la regeneración de componentes tonales en la eSBR en el rango de frecuencia replicado tiene muy poca precisión.

La eSBR utiliza técnicas para ajustar las energías de las áreas emparchadas, el ajuste de la envolvente espectral [1]. Esta técnica utiliza los valores de energía transmitidos en una cuadrícula de tiempo de frecuencia de QMF para remodelar la envolvente espectral. Este estado de la técnica no se ocupa de espectros parcialmente eliminados y debido a la alta resolución temporal tiende a necesitar una cantidad relativamente grande de bits para transmitir

60 valores de energía apropiados o para aplicar una cuantificación basta a los valores de energía.

El método de IGF no necesita una transformación adicional, ya que utiliza la transformación de la MDCT de la

técnica anterior que se calcula como se describe en [15].

El método de ajuste de energía presentado en la presente invención utiliza la información lateral generada por el codificador para reconstruir la envolvente espectral de la señal de audio. Esta información lateral se genera por el codificador como se indica a continuación:

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

b) Aplicar modelado de ruido temporal (TNS)/modelado de recuadro temporal (TTS) en los coeficientes de la MDCT [15, sección 7.8]

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

d) Cuantificar los valores medios de energía

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

Los valores calculados en el paso c) y d) están codificados sin pérdidas y se transmiten como información lateral con la corriente de bits al decodificador.

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

a) Descuantificar los valores transmitidos de la MDCT

b) Aplicar el relleno de ruido de la codificación unificada de voz y audio (USAC) de la técnica anterior si está indicado

c) Aplicar el relleno de recuadros del relleno inteligente de espacios (IGF)

d) Descuantificar los valores de energía transmitidos

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

f) Aplicar TNS/TTS si está indicado

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

TNS en $\hat{x}$.

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 .

Los límites de cada scb_k con $k = 0, 1, 2, \dots, max_sfb$ están definidos por una matriz swb_offset (16, 4.6.2), en donde $swb_offset[k]$ 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:

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

Si la herramienta de IGF se utiliza por el codificador, el usuario define una frecuencia de inicio de IGF y una frecuencia de fin de IGF. Estos dos valores se asignan al índice de banda de factor de escala de ajuste óptimo $igfStartSfb$ e $igfStopSfb$. Ambos se envían en la corriente de bits al decodificador.

[16] describe una transformación de bloque largo y de bloque corto. Para los bloques largos solo 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 guardar la tasa de bits, los factores de escala de dichas ocho ventanas de bloques cortos se agrupan por el codificador.

En el caso del relleno inteligente de espacios (IGF), el método presentado en esta invención utiliza bandas de factor de escala de la técnica anterior para agrupar valores espectrales que son transmitidos al decodificador:

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

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

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

5 Todos los valores $\hat{E}_k$ se transmiten al decodificador.

Se supone que el codificador decide agrupar los conjuntos de factor de escala num_window_group .

10 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 -ésimo subconjunto de w , donde l indica el índice del grupo de ventana, $0 \leq l < num_window_group$.

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

El cálculo de la energía de IGF utiliza 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

$$20 \quad \hat{E}_{k,l} = nINT(4\log_2(E_{k,l}))$$

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

Las fórmulas de codificación mencionadas anteriormente operan utilizando solo coeficientes de la MDCT de valor real $\hat{x}$. Para obtener una distribución de energía más estable en el rango de IGF, es decir para reducir las

25 fluctuaciones de amplitud temporal, se puede utilizar un método alternativo para calcular los valores $\hat{E}_k$:

Siendo $\hat{x}_r \in \mathbb{R}^N$ la transformada MDCT, la representación espectral de valor real de una señal de audio dividida en ventanas de longitud de ventana $2N$, y $\hat{x}_i \in \mathbb{R}^N$ la representación espectral de la transformada MDST de valor real de la misma porción de la señal de audio. La representación espectral de la transformada sinusoidal

30 discreta modificada (MDST) $\hat{x}_i$ podría calcularse o estimarse exactamente 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 su parte real y $\hat{x}_i$ como su parte imaginaria. El codificador aplica opcionalmente modelado de ruido temporal (TNS) en $\hat{x}_r$ y $\hat{x}_i$.

35 Ahora la energía de la señal en el rango de IGF se puede medir con

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

Las energías de valor real y complejo de la banda de reconstrucción, es decir el recuadro que se debe utilizar en el lado del decodificador en la reconstrucción del rango de IGF scb_k , se calcula con:

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

40 donde tr_k es un conjunto de índices - el rango de recuadro de origen asociado, en función de scb_k . En las dos

fórmulas anteriores, en lugar del conjunto de índices scb_k se podría utilizar 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

5

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

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

Con:

10

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

ahora se calcula una versión más estable de E_k , ya que un cálculo de E_k con los valores de MDCT solo 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. $\hat{E}_k$ se calcula como se indicó anteriormente.

15

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

20

Una vez más podría aplicarse la versión alternativa descrita anteriormente para calcular una versión más estable de $E_{k,l}$. Con las definiciones de $\hat{c} = (\hat{x}_r, \hat{x}_i) \in \mathbb{C}^N$, $\hat{x}_r \in \mathbb{R}^N$ que es la transformada de coseno discreta modificada (MDCT) y $\hat{x}_i \in \mathbb{R}^N$ que es la señal de audio dividida en ventanas de longitud $2N$ de la transformada sinusoidal discreta modificada (MDST), calcular

25

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

Calcular de manera análoga

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

30

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

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

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

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

35

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

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

40

Específicamente, el calculador de parámetros 1006 está configurado para calcular la información de energía para la banda de reconstrucción utilizando información sobre la energía de la banda de reconstrucción y utilizando, además, la información sobre una energía de un rango de origen para utilizarse para la reconstrucción de la banda de reconstrucción.

45

Asimismo, el calculador de parámetros 1006 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 rango de origen de una parte de valor real del espectro complejo de la señal original para utilizarse para reconstruir la banda de reconstrucción, y donde el calculador de parámetros está configurado para calcular la información de energía para la banda de reconstrucción utilizando la información de energía (E_{ok}) y la

información de energía adicional (E_{rk}).

Además, el calculador de parámetros 1006 está configurado para determinar una primera información de energía (E_{ok}) en una banda de factor de escala que debe reconstruirse de un espectro complejo de la señal original, para determinar una segunda información de energía (E_{tk}) en un rango de origen del espectro complejo de la señal original para utilizarse para reconstruir la banda de factor de escala que debe reconstruirse, para determinar una tercera información de energía (E_{rk}) en un rango de origen de una parte de valor real del espectro complejo de la señal original para utilizarse para reconstruir la banda de factor de escala que debe 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 utilizando la información de ponderación para obtener una información de energía ponderada y para utilizar la información de energía ponderada como la información de energía para la banda de reconstrucción.

A continuación se presentan ejemplos para los cálculos si bien muchos otros ejemplos pueden quedar a criterio de los expertos en la técnica 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 confirman que aunque solo se procesan valores reales de la MDCT en el lado del decodificador, el cálculo real es - debido al solapamiento y la adición - del procedimiento de cancelación de solapamiento de dominio temporal realizado implícitamente utilizando números complejos. Sin embargo, en particular, la determinación 918 de la información de energía de recuadro de las porciones espectrales adicionales 922, 923 de la banda de reconstrucción 920, para valores de frecuencia diferentes de la primera porción espectral 921 que tiene frecuencias en la banda de reconstrucción 920, se basa en valores reales de la 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.

En el lado del decodificador, si la herramienta de relleno inteligente de espacios (IGF) se señala como ACTIVADA, los valores transmitidos $\hat{E}_k$ se obtienen a partir de la corriente de bits y se descuantificarán con

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

para todos $k = ifgStartSfb, 1 + ifgStartSfb, 2 + ifgStartSfb, \dots, ifgEndSfb$.

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

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

donde k está en el rango definido anteriormente.

Indicamos que $\overline{scb_k} = \{i | i \in scb_k \wedge x_i = 0\}$. 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 sub-banda de relleno inteligente de espacios (IGF) (no descrito en la presente invención) se utiliza para llenar espacios espectrales que resultan de una cuantificación basta de los valores espectrales de la MDCT en el lado del codificador utilizando valores no cero de la MDCT transmitida. x contendrá adicionalmente los valores que reemplazan a todos los valores anteriores reducidos a cero. La energía del recuadro se calcula por:

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

donde k está en el rango definido anteriormente.

La energía faltante en la banda de reconstrucción se calcula por:

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

Y el factor de ganancia para el ajuste se obtiene por:

$$g := \begin{cases} \sqrt{\frac{mE_k}{tE_k}} & \text{if } (mE_k > 0 \wedge tE_k > 0) \\ \text{O de lo contrario} \end{cases}$$

Con:

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

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

$$x_i := g' x_i$$

para todos los $i \in \overline{scb_k}$ y k está en el rango definido anteriormente.

Esto remodela la envolvente espectral de x a la forma de la envolvente espectral original x .

En principio, con la secuencia de ventanas cortas, todos los cálculos definidos anteriormente permanecen igual, 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 de la corriente de bits. Calcular

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

y

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

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

Calcular

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

y

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

Con

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

5 Aplicar

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

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

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

10

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

donde $k = \text{igfStartSfb}, 2 + \text{igfStartSfb}, 4 + \text{igfStartSfb}, \dots, \text{igfEndSfb}$. De nuevo, después de la descuantificación, todos los valores $E_{k \gg 1}$ se transmiten al decodificador.

15

La Fig. 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 indica las energías espectrales para un segundo conjunto de segundas porciones espectrales. El primer conjunto de primeras porciones espectrales se indica en 901a en la Fig. 9a, y la representación codificada de los datos paramétricos se indica en 901b en la Fig. 9a. Un decodificador de audio 900 se proporciona para decodificar la representación codificada 901a del primer conjunto de primeras porciones espectrales para obtener un primer conjunto decodificado de primeras porciones espectrales 904 y para decodificar la representación codificada de los datos paramétricos para obtener datos paramétricos decodificados 902 para el segundo conjunto de segundas porciones espectrales que indican las energías individuales para las bandas de reconstrucción, en donde las segundas porciones espectrales están ubicadas en las bandas de reconstrucción. Además, se proporciona un regenerador de frecuencia 906 para reconstruir valores espectrales de una banda de reconstrucción que comprende una segunda porción espectral. El regenerador de frecuencia 906 utiliza una primera porción espectral del primer conjunto de primeras porciones espectrales y una información de energía individual para la banda de reconstrucción, donde la banda de reconstrucción comprende una primera porción espectral y la segunda porción espectral. El regenerador de frecuencia 906 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 la reconstrucción. Asimismo, el regenerador de frecuencia 906 comprende un calculador 918 para determinar una información de energía de recuadro de otras porciones espectrales de la banda de reconstrucción y para valores de frecuencia que son diferentes de la primera porción espectral, en donde estos valores de frecuencia tienen frecuencias en la banda de reconstrucción, en donde las otras porciones espectrales deben generarse por la regeneración de frecuencia utilizando una primera porción espectral diferente de la primera porción espectral en la banda de reconstrucción.

El regenerador de frecuencia 906 comprende además un calculador 914 para una energía faltante en la banda de reconstrucción, y el calculador 914 funciona utilizando 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 regulador de la envolvente espectral 916 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 recuadros generada por el bloque 918.

45

Con referencia a la Fig. 9c, se ilustra una cierta banda de reconstrucción 920. La banda de reconstrucción comprende una primera porción espectral en la banda de reconstrucción tal como la primera porción espectral 306 en la Fig. 3a ilustrada esquemáticamente en 921. Asimismo, el resto de los valores espectrales en la banda de reconstrucción 920 se debe generar utilizando una región de origen, por ejemplo, de la banda de factor de escala 1, 2, 3 por debajo de la frecuencia de inicio del relleno inteligente de espacios 309 de la Fig. 3a. El regenerador de frecuencia 906 está configurado para generar valores espectrales en bruto para las segundas porciones espectrales

50

922 y 923. A continuación se calcula un factor de ganancia g como se ilustra en la Fig. 9c con el fin de ajustar finalmente los valores espectrales en bruto en las bandas de frecuencia 922, 923 con el fin de obtener las segundas porciones espectrales reconstruidas y ajustadas en la banda de reconstrucción 920, que ahora tienen la misma resolución espectral, es decir, la misma distancia de línea que la primera porción espectral 921. Es importante entender que la primera porción espectral en la banda de reconstrucción ilustrada en 921 en la Fig. 9c está decodificada por el decodificador de audio 900 y no se ve influenciada por el ajuste de la envolvente realizado por el bloque 916 de la Fig. 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 de ancho de banda completa o de audio de tasa completa 900 a través de la línea 904.

A continuación se analizará un ejemplo determinado 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 la energía E_3 para la banda de reconstrucción que corresponde a la banda de factor de escala 6 de la Fig. 3b o la Fig. 3a es igual a 10 unidades. Es importante destacar que el valor de la energía no solo comprende la energía de las porciones espectrales 922, 923, sino también la energía total de la banda de reconstrucción 920 calculada en el lado del codificador, es decir, antes de realizar el análisis espectral, utilizando, por ejemplo, el enmascaramiento 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 rango de origen para los bloques 922, 923 o de los datos en bruto del rango de destino para el bloque 922, 923 es igual a ocho unidades de energía. Por lo tanto, se calcula una energía faltante de cinco unidades.

Se calcula un factor de ganancia de 0,79 basándose en la energía faltante dividida por la energía de recuadro tEk . Entonces, las líneas espectrales en bruto para las segundas porciones espectrales 922, 923 se multiplican por el factor de ganancia calculado. De este modo, solo se ajustan los valores espectrales para las segundas porciones espectrales 922, 923 y las líneas espectrales para la primera porción espectral 921 no se ven influenciadas por este ajuste de la envolvente. Después de la multiplicación de los valores espectrales en bruto para las segundas porciones espectrales 922, 923 se ha calculado una banda de reconstrucción completa que consta de las primeras porciones espectrales en la banda de reconstrucción, y que consta de líneas espectrales en las segundas porciones espectrales 922, 923 en la banda de reconstrucción 920.

Preferentemente, el rango de origen para generar los datos espectrales en bruto en las bandas 922, 923 está, con respecto a la frecuencia, por debajo de la frecuencia de inicio del relleno inteligente de espacios (IGF) 309 y la banda de reconstrucción 920 está por encima de la frecuencia de inicio de IGF 309.

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 forma de realización, el tamaño de las bandas de factor de escala respectivas del decodificador de audio central o se dimensiona de manera que, cuando se aplica la formación de pares de energía, un valor de energía para una banda de reconstrucción proporcione 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 realiza para la banda de factor de escala 4, la banda de factor de escala 5 y la banda de factor de escala, entonces el límite de frecuencia inferior de la banda de reconstrucción 920 es igual al límite inferior de la banda de factor de escala 4 y el límite de energía superior de la banda de reconstrucción 920 coincide con el límite superior de la banda de factor de escala 6.

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

Posteriormente se hace referencia a las Figs. 10a a 10b para explicar las formas de realización 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 de tiempo/espectro 1002 que alimenta un analizador espectral 1004, y el analizador espectral 1004 está conectado a un calculador de parámetros 1006 por un lado y a un codificador de audio 1008 por otro lado. El codificador de audio 1008 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 la energía para una banda de reconstrucción que abarca las primeras y las segundas porciones espectrales. Asimismo, el codificador de audio 1008 está configurado para generar una primera representación codificada del primer conjunto de primeras porciones espectrales que tiene la primera resolución espectral, en donde el codificador de audio 1008 proporciona factores de ajuste de escala para todas las bandas de la representación espectral generada por el bloque 1002. Adicionalmente, tal como se ilustra en la Fig. 3b, el codificador proporciona información de la energía al menos para las bandas de reconstrucción ubicadas, con respecto a la frecuencia, por encima de la frecuencia de inicio de IGF 309 como se ilustra en la Fig. 3a. Por lo tanto, para que las bandas de reconstrucción coincidan preferentemente con las bandas de factor de escala o con grupos de bandas de factor de escala, se proporcionan dos valores, es decir, el factor de ajuste de escala correspondiente del codificador de audio 1008 y, adicionalmente, la información de la energía emitida por el calculador de parámetros 1006.

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 a continuación emite una energía normalizada por banda de reconstrucción/banda de factor de escala.

Además, el calculador paramétrico 1006a de la Fig. 10a comprende un calculador de valor de la energía que recibe información de control del codificador de audio o central 1008 como se ilustra en la línea 1007 en la Fig. 10a. Esta información de control puede comprender información sobre los bloques largos/cortos utilizados por el codificador de audio y/o información de agrupamiento. Por consiguiente, mientras que la información sobre los 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 del valor de energía 1014 emite un único valor de energía para cada banda agrupada que abarca una primera y una segunda porción espectral cuando solo se han agrupado las porciones espectrales.

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

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

En cuanto al cálculo de la información de la energía, los valores espectrales en la banda de la reconstrucción se acumulan sobre dos o más ventanas cortas. En otras palabras, 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 solo se transmite un valor único de información de la energía para esta banda de reconstrucción que abarca dos bloques cortos. A continuación, en el lado del decodificador, el ajuste de la envolvente que se describe en la Fig. 9a a 9d no se lleva a cabo individualmente para cada bloque corto, pero se lleva a cabo conjuntamente para el conjunto de ventanas cortas agrupadas.

A continuación se aplica nuevamente la normalización correspondiente de manera que, aunque se haya realizado cualquier agrupamiento en la frecuencia o agrupamiento temporal, la normalización permite fácilmente que, para el cálculo de la información del valor de energía en el lado del decodificador, solo debe conocerse el valor de la 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 extensión de ancho de banda (BWE) del estado de la técnica, la reconstrucción de la región espectral de alta frecuencia (HF) por encima de una denominada frecuencia de cruce determinada se basa a menudo en el emparchado espectral. En general, la región de alta frecuencia (HF) consta de múltiples parches adyacentes y cada una de estos parches se obtiene de regiones de paso de banda (BP) del espectro de baja frecuencia (LF) por debajo de la frecuencia de cruce determinada. Dentro de una representación de banco de filtros de la señal, dichos sistemas copian un conjunto de coeficientes de sub-bandas adyacentes del espectro de baja frecuencia (LF) en la región de destino. Los límites de los conjuntos seleccionados suelen depender del sistema y no dependen de la señal. Para algunos contenidos de la señal, esta selección de emparchado estática puede provocar un timbre desagradable y la coloración de la señal reconstruida.

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

El enfoque de la invención se denomina Relleno Inteligente de Espacios (IGF) y, en su configuración preferida, se aplica en un sistema de extensión de ancho de banda (BWE) basándose en una transformada de frecuencia temporal como, por ejemplo, la Transformada de Coseno Discreta Modificada (MDCT). Sin embargo, las enseñanzas de la invención son de aplicación general, por ejemplo, de manera análoga dentro de un sistema basado en el Banco de Filtros Espejo en Cuadratura (QMF).

Una ventaja de la configuración de IGF basado en la MDCT es la integración perfecta en los codificadores de audio basados en la MDCT, por ejemplo la Codificación Avanzada de Audio (AAC) de MPEG. Compartiendo la misma transformada para la codificación de audio de forma de onda y para BWE reduce significativamente la complejidad computacional general para el códec de audio.

Por otra parte, la invención proporciona una solución para los problemas inherentes de estabilidad que se encuentran en los esquemas de emparchado adaptativo del estado de la técnica.

El sistema propuesto se basa en la observación de que para algunas señales, una selección de emparchado sin guía puede generar cambios de timbre y coloraciones en la señal. Si una señal que es tonal en la región espectral de origen (SSR) pero es similar a ruido en la región espectral de destino (STR), el emparchado de la STR similar a ruido por la SSR tonal puede generar un timbre antinatural. El timbre de la señal también puede cambiar ya que la estructura tonal de la señal podría desalinearse o incluso destruirse por el proceso de emparchado.

El sistema de IGF propuesto realiza una selección inteligente de recuadros utilizando 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 utilizar para ajustar con precisión el desplazamiento espectral del espectro copiado para que esté tan cerca como sea posible de 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 la señal instantánea y, al mismo tiempo, la estabilidad temporal. La provisión de estabilidad temporal es especialmente importante para las señales que tienen poca similitud de SSR y STR y que, por lo tanto, exhiben valores bajos de correlación cruzada o cuando se emplean medidas de similitud que son ambiguas. En dichos casos, la estabilización impide el comportamiento pseudo-aleatorio de la selección adaptativa de recuadros.

Por ejemplo, una clase de señales que a menudo plantea problemas para la extensión de ancho de banda del estado de la técnica se caracteriza por una concentración distinta de la energía en regiones espectrales arbitrarias, tal como se muestra en la Fig. 12a (a la izquierda). Aunque hay métodos disponibles para ajustar la envolvente espectral y la tonalidad del espectro reconstruido en la región de destino, para algunas señales, estos métodos no son capaces de conservar bien el timbre como se muestra en la Fig. 12a (a la derecha). En el ejemplo ilustrado en la Fig. 12a, la magnitud del espectro en la región de destino de la señal original por encima de una así denominada frecuencia de cruce $f_{\text{crossover}}$ (Figura 12a, a la izquierda) disminuye casi linealmente. Por el contrario, en el espectro reconstruido (Fig. 12a, a la derecha) hay un conjunto distinto de pendientes y picos que se percibe como una artefacto de coloración del timbre.

Un paso importante del nuevo enfoque consiste en definir un conjunto de recuadros entre los que puede tener lugar la elección basada en la similitud posterior. En primer lugar, los límites de los recuadros, tanto de la región de origen

como de la región de destino, tienen que estar definidos unos con otros. Por lo tanto, la región de destino entre la frecuencia de inicio de IGF del codificador central $f_{IGFstart}$ y una frecuencia más alta disponible $f_{IGFstop}$ se divide en un número entero arbitrario $nTar$ de recuadros, cada uno de los cuales tiene un tamaño individual predefinido. Entonces, para cada recuadro de destino $tar[idx_tar]$, se genera un conjunto de recuadros de origen de igual tamaño $src[idx_src]$. Por lo anterior se determina el grado básico de libertad del sistema de IGF. El número total de recuadros de origen $nSrc$ está determinado por el ancho de banda de la región de origen,

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

en donde f_{IGFmin} es la frecuencia más baja disponible para la selección de recuadros de manera que un número entero $nSrc$ de recuadros de origen se adapta 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 superponen entre sí por un factor de solapamiento entre 0 y 1, en donde 0 significa ningún solapamiento y 1 significa el 100 % de solapamiento. El caso del 100 % de solapamiento implica que solo uno o ningún recuadro de origen está disponible.

La Fig. 12b muestra un ejemplo de los límites de recuadros de un conjunto de recuadros. En este caso, todos los recuadros de destino 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 de destino, la correlación cruzada se calcula con varios recuadros de origen en retardos de hasta $xcorr_maxLag$ intervalos. Para un recuadro de destino determinado idx_tar y un recuadro de origen idx_src , $xcorr_val[idx_tar][idx_src]$ proporciona el valor máximo de la correlación cruzada entre los recuadros, mientras que $xcorr_lag[idx_tar][idx_src]$ proporciona el retardo en el que este máximo ocurre 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 utiliza para controlar la proximidad de la coincidencia entre el recuadro de origen y el recuadro de destino. Este parámetro da lugar a una reducción de artefactos y ayuda a conservar mejor el timbre y el color de la señal.

En algunos casos puede ocurrir que el tamaño de un recuadro de destino específico sea 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 llenar completamente el recuadro de destino específico. Todavía es posible realizar la correlación cruzada entre el recuadro grande de destino y el recuadro de origen más pequeño con el fin de obtener la mejor posición del recuadro de origen en el recuadro de destino en términos del retardo de la correlación cruzada $xcorr_lag$ y el signo $xcorr_sign$.

La correlación cruzada de los recuadros espectrales en bruto y la señal original pueden no ser la medida de similitud más adecuada aplicada a los espectros de audio con una estructura fuerte de formantes. El blanqueo de un espectro quita la información de la envolvente en bruto y, por lo tanto, enfatiza la estructura precisa 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 región espectral de destino (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.

En otras configuraciones, solo se blanquea el recuadro utilizando un procedimiento predefinido. Una etiqueta de "blanqueo" transmitida indica al decodificador que se aplicará el mismo proceso de blanqueo predefinido al recuadro de frecuencia dentro del relleno inteligente de espacios (IGF).

Para blanquear la señal, primero se calcula una estimación de la envolvente espectral. A continuación, el espectro de la MDCT se divide por la envolvente espectral. La estimación de la envolvente espectral se puede estimar en el espectro de la MDCT, las energías del espectro de la MDCT, las estimaciones del espectro de energía complejo basado en la MDCT o las estimaciones del espectro de energía. La señal en la que se estima la envolvente se llamará *señal de base* de aquí en adelante.

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

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

Existen diferentes métodos para calcular la envolvente:

- transformando la señal de base con una transformada de coseno discreta (DCT), reteniendo solo los coeficientes más bajos de la DCT (fijando la más alta en cero) y calculando a continuación una DCT inversa
 - calculando una envolvente espectral de un conjunto de Coeficientes de Predicción Lineal (LPC, por sus siglas en inglés) calculado sobre el cuadro de audio de dominio temporal
- 5 • filtrando la señal de base con un filtro de paso bajo

Preferentemente, se elige el último enfoque. Para aplicaciones que requieren baja complejidad computacional se puede realizar una cierta simplificación para el blanqueo de un espectro de la MDCT: En primer lugar, la envolvente se calcula por medio de una media móvil. Esto solo necesita dos ciclos de procesador por intervalo de la MDCT. Entonces, con el fin de evitar el cálculo de la división y la raíz cuadrada, la envolvente espectral se aproxima por 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 por la envolvente se puede realizar por otra operación de desplazamiento.

Después de calcular la correlación de cada recuadro de origen con cada recuadro de destino, para todos los recuadros de destino $nTar$ se selecciona el recuadro de origen con la correlación más alta para que lo sustituya. Para coincidir mejor con la estructura espectral original, el retardo de la correlación se utiliza para modular el espectro replicado por un número entero de intervalos de la transformada. En caso de retardos impares, el recuadro se modula adicionalmente a través de la multiplicación por una secuencia temporal alternativa de -1/1 para compensar la representación de frecuencia inversa de cualquier otra banda dentro de la Transformada de Coseno Discreta Modificada (MDCT).

La Figura 12c muestra un ejemplo de una correlación entre un recuadro de origen y un recuadro de destino. En este ejemplo, el retardo de la correlación es 5, por lo que el recuadro de origen tiene que modularse por 5 intervalos hacia los intervalos de frecuencia más altos en la etapa de copiado del algoritmo de extensión de ancho de banda (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.

Por lo tanto, la cantidad total de información lateral para transmitir desde el codificador al decodificador podría constar de los siguientes datos:

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

El recorte y la estabilización de recuadros constituyen un paso importante en el relleno inteligente de espacios (IGF). Su necesidad y ventajas se explican con un ejemplo, en el supuesto de una señal de audio tonal estacionaria como, por ejemplo, una nota de altura de tono estable. La lógica determina que se introducen menos artefactos si, para una región de destino determinada, siempre se seleccionan los recuadros de origen de la misma región de origen a través de los cuadros. A pesar de que se supone que la señal es estacionaria, esta condición no se aplicaría bien en cada cuadro ya que la medida de similitud (por ejemplo, correlación) de otra región de origen similar igual podría dominar el resultado de la similitud (por ejemplo, correlación cruzada). Esto hace que $tileNum[nTar]$ entre cuadros adyacentes dude entre dos o tres opciones muy similares. Este puede ser el origen de un artefacto similar a 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 máximamente disímiles. 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 , lo correlacionamos con todos los otros recuadros de origen, encontrando la mejor correlación entre s_i y s_j y almacenándola en una matriz S_x . En este punto, $S_x[i][j]$ contiene el valor máximo absoluto de correlación cruzada entre s_i y s_j . La adición de 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 otros recuadros de origen T .

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

En este punto, T representa una buena medida de similitud entre un recuadro de origen y otros recuadros de origen. Si, para cualquier recuadro de origen i ,

$$T > \text{umbral}$$

el recuadro de origen i puede retirarse del conjunto de fuentes potenciales, ya que está muy correlacionado con otras fuentes. El recuadro que tiene la correlación más baja del conjunto de recuadros que cumple la condición en la ecuación 1 se elige como un recuadro representativo para este subconjunto. De este modo, nos aseguramos de que los recuadros de origen sean máximamente disímiles entre sí.

5 El método de recorte de recuadros implica también una memoria del conjunto de recuadros recortados utilizados en el cuadro anterior. Los recuadros que estuvieron activos en el cuadro anterior se retienen en el siguiente cuadro si existen también candidatos alternativos para el recorte.

10 Dejar que los recuadros s_3, s_4 t s_5 estén activos a partir de los recuadros $\{s_1, s_2, \dots, s_5\}$ en el cuadro k , a continuación en el cuadro $k+1$, incluso si los recuadros s_1, s_3 y s_2 compiten por ser recortados donde s_3 está correlacionado máximamente con los otros, s_3 se retiene ya que fue un recuadro de origen útil en el cuadro anterior y, por lo tanto, su retención en el conjunto de recuadros de origen es beneficioso para reforzar la continuidad temporal en la selección de recuadros. Este método se aplica preferentemente si la correlación cruzada entre el origen i y el destino j , representada como $T_x[i][j]$ es alta.

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

Por ejemplo, si

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

25 entonces se utiliza un umbral provisional, entonces

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

30 para todos los $n\text{Tar}$ de este cuadro k .

Las dos técnicas anteriores reducen en gran medida los artefactos que se producen a partir del cambio rápido de números de recuadros fijos a través de los cuadros. Otra ventaja adicional de este recorte y estabilización de recuadros es que no se necesita enviar información extra al decodificador y tampoco se necesita un cambio de arquitectura del decodificador. Esta propuesta de recorte de recuadros resulta una manera elegante de reducir el ruido musical potencial en forma de artefactos o ruido excesivo en las regiones espectrales de los recuadros.

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

45 Asimismo, el decodificador de audio comprende un decodificador paramétrico 1104 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 de frecuencia 1106 que recibe, como una primera entrada 1101, primeras porciones espectrales decodificadas y como una segunda entrada en 1103 la información paramétrica que incluye, para cada recuadro de frecuencia de destino o banda de reconstrucción de destino, una información de rango de origen. El regenerador de frecuencia 1106 a continuación aplica la regeneración de frecuencia utilizando valores espectrales del rango de origen identificado por la información de adaptación con el fin de generar los datos espectrales para el rango de destino. A continuación, las primeras porciones espectrales 1101 y la salida del regenerador de frecuencia 1107 se introducen ambas en un convertidor de espectro-tiempo 1108 para generar finalmente la señal de audio decodificada.

50 Preferentemente, el decodificador de audio 1102 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, por ejemplo, un decodificador de audio de dominio temporal o paramétrico.

60 Como se indica en la Fig. 11b, el regenerador de frecuencia 1106 puede comprender las funcionalidades del bloque 1120 que ilustra un modulador de recuadros - selector del rango de origen para retardos impares, un filtro blanqueado 1122, cuando se proporciona una etiqueta de blanqueo 1123 y, adicionalmente, una envolvente espectral con funcionalidades de ajuste implementadas como se ilustra en el bloque 1128 utilizando datos espectrales en bruto generados por cualquiera de los bloques 1120 o 1122 o la cooperación de ambos bloques. De todos modos, el regenerador de frecuencia 1106 puede comprender un conmutador 1124 reactivo a una etiqueta de blanqueo recibida 1123. Cuando se fija la etiqueta de blanqueo, la salida del selector de rango de origen/modulador

de recuadros para retardos impares se introduce en el filtro de blanqueo 1122. Sin embargo, la etiqueta de blanqueo 1123 no se fija para una cierta banda de reconstrucción, por lo que entonces se activa una línea de desvío 1126 de manera que la salida del bloque 1120 se proporciona al bloque de ajuste de la envolvente espectral 1128 sin ningún blanqueo.

- 5 Puede haber más de un nivel de blanqueo (1123) señalizado en la corriente de bits y estos niveles pueden señalizarse por recuadro. En caso de que haya tres niveles señalizados por recuadro, los niveles se codificarán de la siguiente manera:

```

bit = readBit(1);
if(bit == 1) {
    for(tile_index = 0..nT)

        /*los mismos niveles que el último cuadro*/

        whitening_level[tile_index] = whitening_level_prev_frame[tile_index];
} else {
    /*primer recuadro:*/
    tile_index = 0;
    bit = readBit(1);
    if(bit == 1) {
        whitening_level[tile_index] = MID_WHITENING;
    } else {
        bit = readBit(1);
        if(bit == 1) {
            whitening_level[tile_index] = STRONG_WHITENING;
        } else {
            whitening_level[tile_index] = OFF; /*sin blanqueo*/
        }
    }
}
/*recuadros restantes:*/
bit = readBit(1);
if(bit == 1) {
    /*los niveles de aplanamiento para los recuadros restantes son iguales
    que para el primero.*/
    /*No tienen que leerse otros intervalos*/
    for(tile_index = 1..nT)

        whitening_level[tile_index] = whitening_level[0];
} else {
    /*bits leídos para los recuadros restantes como para el primer recuadro*/

```

```

for(tile_index = 1..nT) {
    bit = readBit(1);
    if(bit == 1) {
        whitening_level[tile_index] = MID_WHITENING;
    } else {
        bit = readBit(1);
        if(bit == 1) {
            whitening_level[tile_index] = STRONG_WHITENING;
        } else {
            whitening_level[tile_index] = OFF; /*sin blanqueo*/
        }
    }
}
}
}
}

```

MID_WHITENING y STRONG_WHITENING se refieren a distintos filtros de blanqueo (1122) que pueden diferir en la forma en que se calcula la envolvente (como se describió anteriormente).

5 El regenerador de frecuencia del lado del decodificador puede controlarse por una ID de rango de origen 1121 cuando se aplica solo un esquema de selección de recuadros espectrales en bruto. Sin embargo, cuando se aplica un esquema de selección de recuadros espectrales de sincronización precisa, entonces se proporciona además un retardo de rango de origen 1119. Asimismo, siempre que el cálculo de la correlación proporcione un resultado negativo, entonces, adicionalmente se puede aplicar también 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.

10 Por lo tanto, la presente invención tal como se describe en la Fig. 11a, 11b garantiza la obtención de una calidad óptima de audio debido al hecho de que el mejor rango de origen coincidente para un destino determinado o rango de destino se calcula en el lado del codificador y se aplica en el lado del decodificador.

15 La Fig. 11c es un codificador de audio determinado para codificar una señal de audio que comprende un convertidor de tiempo-espectro 1130, un analizador espectral conectado posteriormente 1132 y, adicionalmente, un calculador de parámetros 1134 y un codificador central 1136. El codificador central 1136 emite rangos de origen codificados y el calculador de parámetros 1134 emite información de adaptación para rangos de destino.

20 Los rangos de origen codificados se transmiten a un decodificador junto con información de adaptación para los rangos de destino de manera que el decodificador ilustrado en la Fig. 11a se encuentra en la posición para realizar una regeneración de frecuencia.

25 El calculador de parámetros 1134 está configurado 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 se adapte a la segunda porción espectral. Preferentemente, los resultados de adaptación para diferentes rangos de origen y rangos de destino, como se ilustra en las Figs. 12a, 12b para determinar un par de adaptación seleccionado comprenden la segunda porción espectral, y el calculador de parámetros está configurado para proporcionar esta información de adaptación que identifica el par de adaptación en una señal de audio codificada. Preferentemente, el calculador de parámetros 1134 de la presente invención está configurado para utilizar regiones de destino 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 Fig. 12b. Preferentemente, las regiones de destino 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 de inicio de relleno de espacios 309 de la Fig. 3a, y preferentemente, la región de destino predefinida que abarca una región espectral inferior coincide, con su límite de frecuencia inferior, con la frecuencia de inicio de relleno de espacios de manera

que cualquier rango de destino se encuentra por encima de la frecuencia de inicio de relleno de espacios y los rangos de origen se encuentran por debajo de la frecuencia de inicio de relleno de espacios.

Como se explicó anteriormente, una granularidad precisa se obtiene comparando una región de destino con una región de origen sin ningún retardo en la región de origen y la misma región de origen, pero con un cierto retardo. Estos retardos se aplican en el calculador de correlación cruzada 1140 de la Fig. 11d y la selección de pares de adaptación se realiza finalmente por el selector de recuadros 1144.

Además, se prefiere realizar un blanqueo de rangos de origen y/o rangos de destino como se ilustra en el bloque 1142. A continuación, este bloque 1142 proporciona una etiqueta de blanqueo a la corriente de bits que se utiliza para controlar el conmutador del lado del decodificador 1123 de la Fig. 11b. Asimismo, si el calculador de correlación cruzada 1140 proporciona un resultado negativo, entonces este resultado negativo también se señala a un decodificador. Por lo tanto, en una forma de realización preferida, el selector de recuadros emite una ID de rango de origen para un rango de destino, un retardo, un signo y el bloque 1142 proporciona además una etiqueta de blanqueo.

Asimismo, el calculador de parámetros 1134 está configurado para realizar un recorte de recuadros de origen 1146 reduciendo el número de rangos de origen potenciales por lo que un emparchado de origen se retira de un conjunto de recuadros de origen potenciales basándose en un umbral de similitud. Por lo tanto, cuando dos recuadros de origen son más similares o iguales a un umbral de similitud, entonces uno de estos dos recuadros de origen se retira del conjunto de fuentes potenciales y el recuadro de origen eliminado ya no se utiliza para el procesamiento posterior y, específicamente, no puede seleccionarse por el selector de recuadros 1144 o no se utiliza para el cálculo de la correlación cruzada entre diferentes rangos de origen y rangos de destino como se realizó en el bloque 1140.

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

Todos estos aspectos diferentes pueden ser de uso inventivo y son independientes entre sí pero, adicionalmente, también se pueden aplicar juntos como se ilustra básicamente en la Fig. 2a y 2b. Sin embargo, el procesamiento específico de dos canales se puede aplicar también a un esquema de codificador/decodificador ilustrado en la Fig. 13a y 13b, y esto mismo se aplica al procesamiento de TNS/TTS, al cálculo de información de energía de la envolvente y la aplicación en la banda de reconstrucción o la identificación del rango de origen de adaptación y la aplicación correspondiente en el lado del 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 rango de origen de adaptación o con otros tipos de cálculos de energía para la representación de la envolvente espectral. Por lo tanto, es evidente que las 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, donde un bloque o dispositivo corresponde a un paso del método o a una característica de un paso del método. Análogamente, los aspectos descritos en el contexto de un paso del método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato respectivo. Algunos o todos los pasos del método se pueden realizar por (o con) un aparato de hardware tal como, por ejemplo, un microprocesador, una computadora programable o un circuito electrónico. En algunas formas de realización, alguno o más de la mayoría de los pasos importantes del método se pueden realizar por dicho aparato.

En función de determinados requisitos de implementación, las formas de realización de la invención se pueden implementar en hardware o en software. La implementación se puede realizar utilizando 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 memoria ROM, una memoria PROM, y una memoria EPROM, una memoria EEPROM o una memoria FLASH, que tienen señales de control de lectura electrónica almacenadas en los mismos, que cooperan (o son capaces de cooperar) con un sistema informático programable de forma tal que se realice el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por computadora.

Algunas formas de realización de acuerdo con la invención comprenden un portador de datos que tiene señales de control de lectura electrónica, las cuales son capaces de cooperar con un sistema de computadora programable, de tal manera que uno de los métodos descritos en el presente documento se realice.

En general, las formas de realización de la presente invención se pueden implementar como un producto de

programa informático con un código de programa, cuyo código de programa es operativo para realizar uno de los métodos cuando el producto de programa informático se ejecuta en una computadora. El código del programa se puede almacenar, por ejemplo, en un portador legible por computadora.

Otras formas de realización comprenden los programas informáticos para realizar uno de los métodos descritos en la presente, almacenados en un portador legible por computadora.

En otras palabras, una forma de realización del método de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en una computadora.

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

Por lo tanto, una forma de realización adicional de la invención es una corriente de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en el presente documento. La corriente de datos o la secuencia de señales, por ejemplo, pueden estar configuradas para transferirse a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

Una forma de realización adicional comprende un medio de procesamiento, por ejemplo, una computadora o un dispositivo lógico programable configurado o adaptado para realizar uno de los métodos descritos en la presente invención.

Otra forma de realización comprende una computadora que tiene el programa informático instalado en la misma para realizar uno de los métodos descritos en el presente documento.

Otra forma de realización de acuerdo con la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, electrónicamente u ópticamente) un programa informático para realizar uno de los métodos descritos en el presente documento a un receptor. El receptor puede ser, por ejemplo, una computadora, 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 formas de realización, un dispositivo lógico programable (por ejemplo, un campo de matrices de puertas programables) se puede utilizar para realizar algunas o todas las funcionalidades de los métodos descritos en la presente invención. En algunas formas de realización, un campo de matrices de puertas programables puede cooperar con un microprocesador para realizar uno de los métodos descritos en el presente documento. En general, los métodos se realizarán, preferentemente, por cualquier aparato de hardware.

Las formas de realización anteriormente descritas son simplemente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para otros expertos en la materia. Es la intención, por lo tanto, de que la invención esté limitada solamente por el alcance de las reivindicaciones a continuación de la patente y no por los detalles específicos presentados a modo de descripción y explicación de las formas de realización de la presente.

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, Múnich, 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, París, 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, Estados Unidos 2006.

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

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

- 5 [7] M. Neuendorf, M. Multus, 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 de 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.
- 10 [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 Ginebra. The 2000 IEEE International Symposium on, vol.3, n. °, págs. 201, 204 vol.3, 2000
- 15 [11] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio”, Ginebra, 1998.
- 20 [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 Ángeles 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
- 25 [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.
- 30 [16] International Standard ISO/IEC 14496-3: 2005, Information technology - Coding of audio-visual objects - Part 3: Audio, Ginebra, 2005.
- 35 [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 International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas (Estados Unidos), abril de 2010.

REIVINDICACIONES

1. Aparato para decodificar una señal de audio codificada, que comprende:

5 un decodificador de audio de dominio espectral (602) para generar una primera representación decodificada de un primer conjunto de primeras porciones espectrales que son valores residuales de predicción espectral; un regenerador de frecuencia (604) para generar una segunda porción espectral reconstruida utilizando una primera porción espectral del primer conjunto de primeras porciones espectrales, en el que la segunda porción espectral reconstruida y el primer conjunto de primeras porciones espectrales comprenden valores residuales de predicción espectral; y
10 un filtro de predicción inversa (606, 616, 622) para realizar una predicción inversa sobre la frecuencia utilizando los valores residuales de predicción espectral para el primer conjunto de primeras porciones espectrales y la segunda porción espectral reconstruida utilizando la información del filtro de predicción (607) incluida en la señal de audio codificada.

15 2. El aparato de la reivindicación 1, que comprende además un modelador de la envolvente espectral (614) para modelar una envolvente espectral de una señal de entrada o una señal de salida del filtro de predicción inversa (606).

20 3. Aparato de la reivindicación 2, en el que la señal de audio codificada comprende información de la envolvente espectral para la segunda porción espectral, teniendo la información de la envolvente espectral una segunda resolución espectral, siendo la segunda resolución espectral más baja que una primera resolución espectral asociada con la primera representación decodificada, en el que el modelador de la envolvente espectral (624) está configurado para aplicar una operación de modelado de la envolvente espectral en la salida del filtro de predicción inversa (622), en el que la información del filtro (607) se ha determinado utilizando una señal de audio antes del filtrado de predicción, o
25 en el que el modelador de la envolvente espectral (614) está configurado para aplicar una operación de modelado de la envolvente espectral en la entrada del filtro de predicción inversa (616), cuando la información del filtro de predicción (607) se ha determinado utilizando una señal de audio después de un filtrado de predicción en un codificador.

30 4. Aparato de una de las reivindicaciones anteriores, que comprende además un convertidor de tiempo de frecuencia (212) para convertir una salida del filtro de predicción inversa (210) o una salida modelada con la envolvente del filtro de predicción inversa (210) en una representación de tiempo.

35 5. Aparato de acuerdo con una de las reivindicaciones anteriores, en el que el filtro de predicción inversa (606) es un filtro complejo definido por la información del filtro de predicción (607).

40 6. Aparato de acuerdo con una de las reivindicaciones anteriores, en el que el decodificador de audio en el dominio espectral (602) está configurado para generar la primera representación decodificada de manera que la primera representación decodificada tiene una frecuencia de Nyquist igual a una tasa de muestreo de una señal de dominio temporal generada por la conversión de tiempo-frecuencia de una salida del filtro de predicción inversa (606).

45 7. Aparato de una de las reivindicaciones anteriores, en el que el decodificador de audio en el dominio espectral (602) está configurado de modo que una frecuencia máxima representada por un valor espectral para la frecuencia máxima en la primera representación decodificada es igual a una frecuencia máxima incluida en una representación de tiempo generada por la conversión de frecuencia-tiempo de una salida del filtro de predicción (606), en el que el valor espectral para la frecuencia máxima en la primera representación es cero o diferente de cero.

50 8. Aparato de una de las reivindicaciones anteriores, en el que la primera representación decodificada del primer conjunto de primeras porciones espectrales comprende valores espectrales reales, en el que el aparato comprende además un estimador (724) para estimar valores imaginarios para el primer conjunto de primeras porciones espectrales a partir del primer conjunto de valor real de primeras porciones espectrales, y en el que el filtro de predicción inversa (728) es un filtro complejo de predicción inversa definido por la información del filtro de predicción de valor complejo (714), y en el que el aparato comprende además un convertidor de frecuencia-tiempo (730) configurado para realizar una conversión de un espectro de valor complejo en una señal de audio de dominio temporal.

60

9. Aparato de una de las reivindicaciones anteriores, en el que el filtro de predicción inversa (606, 728) está configurado para aplicar una pluralidad de subfiltros, en el que un límite de frecuencia de cada subfiltro coincide con un límite de frecuencia de una banda de reconstrucción que coincide con un recuadro de frecuencia.

5 10. Aparato para codificar una señal de audio, que comprende:
un convertidor de tiempo-espectro (100, 702) para convertir una señal de audio (99) en una representación
espectral;
10 un filtro de predicción (704) para realizar una predicción sobre la frecuencia en la representación espectral para generar valores residuales espectrales, estando definido el filtro de predicción por la información del filtro derivada de la señal de audio;
un codificador de audio (708) para codificar un primer conjunto de primeras porciones espectrales de los valores
residuales espectrales para obtener un primer conjunto codificado de primeros valores espectrales que tienen
15 una primera resolución espectral;
un codificador paramétrico (706) para codificar paramétricamente un segundo conjunto de segundas porciones
espectrales de los valores residuales espectrales o de valores de la representación espectral con una segunda
resolución espectral que es más baja que la primera resolución espectral; y
una interfaz de salida (710) para emitir una señal codificada que comprende el segundo conjunto codificado, el
20 primer conjunto codificado y la información del filtro (714).

11. Aparato de la reivindicación 10, en el que el convertidor de tiempo-espectro está configurado para realizar una transformada de coseno discreta modificada, y en el que los valores residuales espectrales son valores residuales espectrales de la transformada de
25 coseno discreta modificada.

12. Aparato de una de las reivindicaciones 10 y 11, en el que el filtro de predicción (704) comprende un calculador de información del filtro, estando configurado el
calculador de información del filtro para utilizar valores espectrales de una representación espectral para calcular la
30 información del filtro y en el que el filtro de predicción está configurado para calcular los valores residuales
espectrales utilizando valores espectrales de la representación espectral, en el que los valores espectrales para
calcular la información del filtro y los valores espectrales introducidos en el filtro de predicción se obtienen de la
misma señal de audio.

35 13. Aparato de una las reivindicaciones 10 a 12, en el que el filtro de predicción comprende un calculador de filtros para calcular la información del filtro utilizando
valores espectrales de una frecuencia de inicio de modelado de ruido temporal, TNS a una frecuencia de fin de TNS,
en el que la frecuencia de inicio de TNS es inferior a 4 kHz y la frecuencia de fin de TNS es superior a 9 kHz.

40 14. Aparato de la reivindicación 13 que comprende además un analizador (102, 706) para determinar el primer
conjunto de primeras porciones espectrales para codificarse por el codificador de audio (708), utilizando el
analizador una frecuencia de relleno de espacios, en el que las porciones espectrales por debajo de la frecuencia de
inicio de relleno de espacios son primeras porciones espectrales, y
45 en el que la frecuencia de fin de TNS es mayor que la frecuencia de relleno de espacios.

15. Aparato de una de las reivindicaciones 10 a 14, en el que el convertidor de tiempo-frecuencia (702) está configurado para proporcionar una representación espectral
compleja,
50 en el que el filtro de predicción está configurado para realizar una predicción sobre la frecuencia con la
representación espectral de valor complejo, y
en el que la información del filtro (714) está configurada para definir un filtro complejo de predicción inversa.

16. Un método para decodificar una señal de audio codificada, que comprende:

55 generar (602) una primera representación decodificada de un primer conjunto de primeras porciones espectrales
que son los valores residuales de predicción espectral;
regenerar (604) una segunda porción espectral reconstruida utilizando una primera porción espectral del primer
conjunto de primeras porciones espectrales, en el que la segunda porción espectral reconstruida y el primer
conjunto de primeras porciones espectrales comprenden valores residuales de predicción espectral; y
60 realizar (606, 616, 622), mediante un filtro de predicción inversa, una predicción inversa sobre la frecuencia
utilizando los valores residuales de predicción espectral para el primer conjunto de primeras porciones
espectrales y la segunda porción espectral reconstruida utilizando la información del filtro de predicción (607)
incluida en la señal de audio codificada.
que comprende además un modelador de la envolvente espectral (614) para modelar una envolvente espectral

de una señal de entrada o una señal de salida del filtro de predicción inversa (606).

17. Método de la reivindicación 16,

en el que la señal de audio codificada comprende información de la envolvente espectral para la segunda porción espectral, teniendo la información de la envolvente espectral una segunda resolución espectral, siendo la segunda resolución espectral más baja que una primera resolución espectral asociada con la primera representación decodificada,

en el que la regeneración (604) comprende un modelado de la envolvente espectral (624) que comprende aplicar una operación de modelado de la envolvente espectral en una salida del paso de realizar (606, 616, 622) una predicción inversa sobre la frecuencia, en el que la información del filtro (607) se ha determinado utilizando una señal de audio antes del filtrado de predicción, o

en el que la regeneración (604) comprende un modelado de la envolvente espectral (624) que comprende aplicar una operación de modelado de la envolvente espectral en una entrada del paso de realizar (606, 616, 622) una predicción inversa sobre la frecuencia, cuando la información del filtro de predicción (607) se ha determinado utilizando una señal de audio después de un filtrado de predicción en un codificador.

18. Método para codificar una señal de audio, que comprende:

convertir (100, 702) una señal de audio (99) en una representación espectral;

realizar (704) una predicción sobre la frecuencia en la representación espectral para generar valores residuales espectrales, estando definido el filtro de predicción por la información del filtro derivada de la señal de audio;

codificar (708) un primer conjunto de primeras porciones espectrales de los valores residuales espectrales para obtener un primer conjunto codificado de primeros valores espectrales que tienen una primera resolución espectral;

codificar paramétricamente (706) un segundo conjunto de segundas porciones espectrales de los valores residuales espectrales o de valores de la representación espectral con una segunda resolución espectral que es más baja que la primera resolución espectral; y

emitir (710) una señal codificada que comprende el segundo conjunto codificado, el primer conjunto codificado y la información del filtro (714).

19. Programa informático para realizar, cuando se ejecuta en una computadora o en un procesador, el método de la reivindicación 16 o el método de la reivindicación 18.

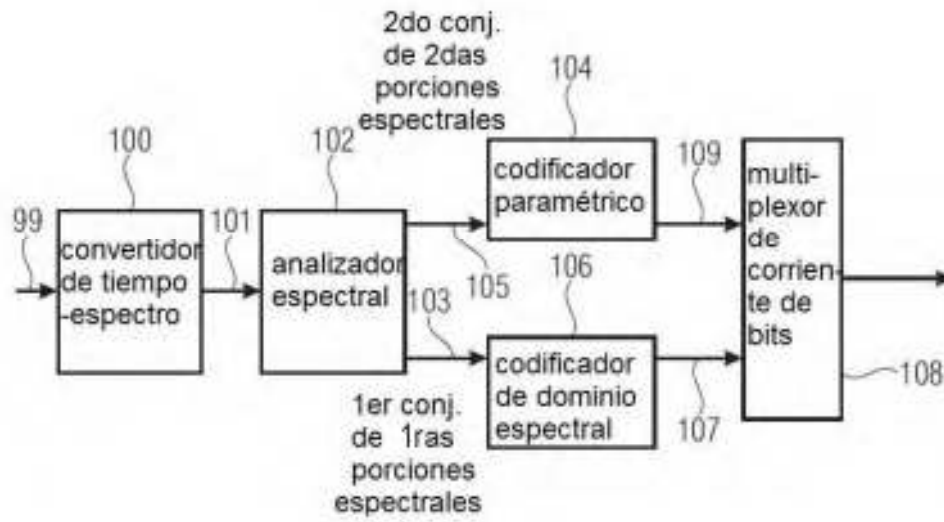


FIG 1A

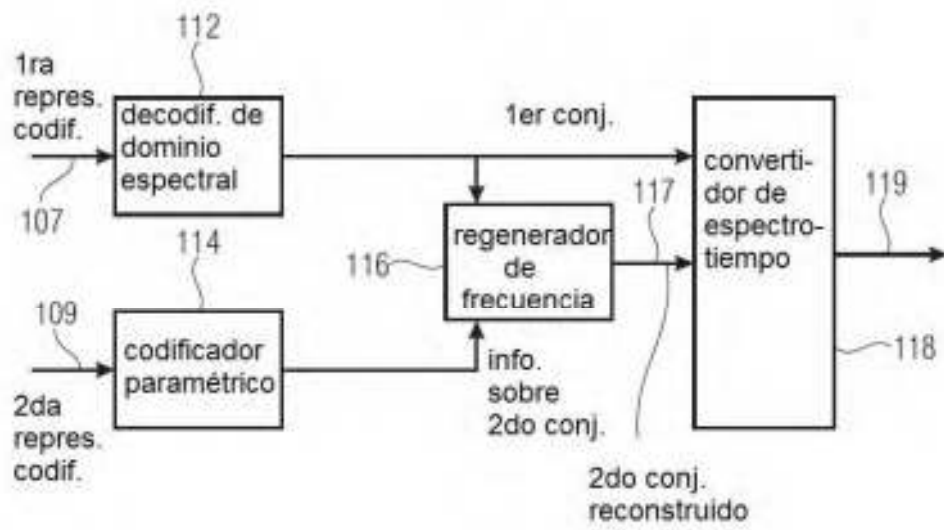


FIG 1B

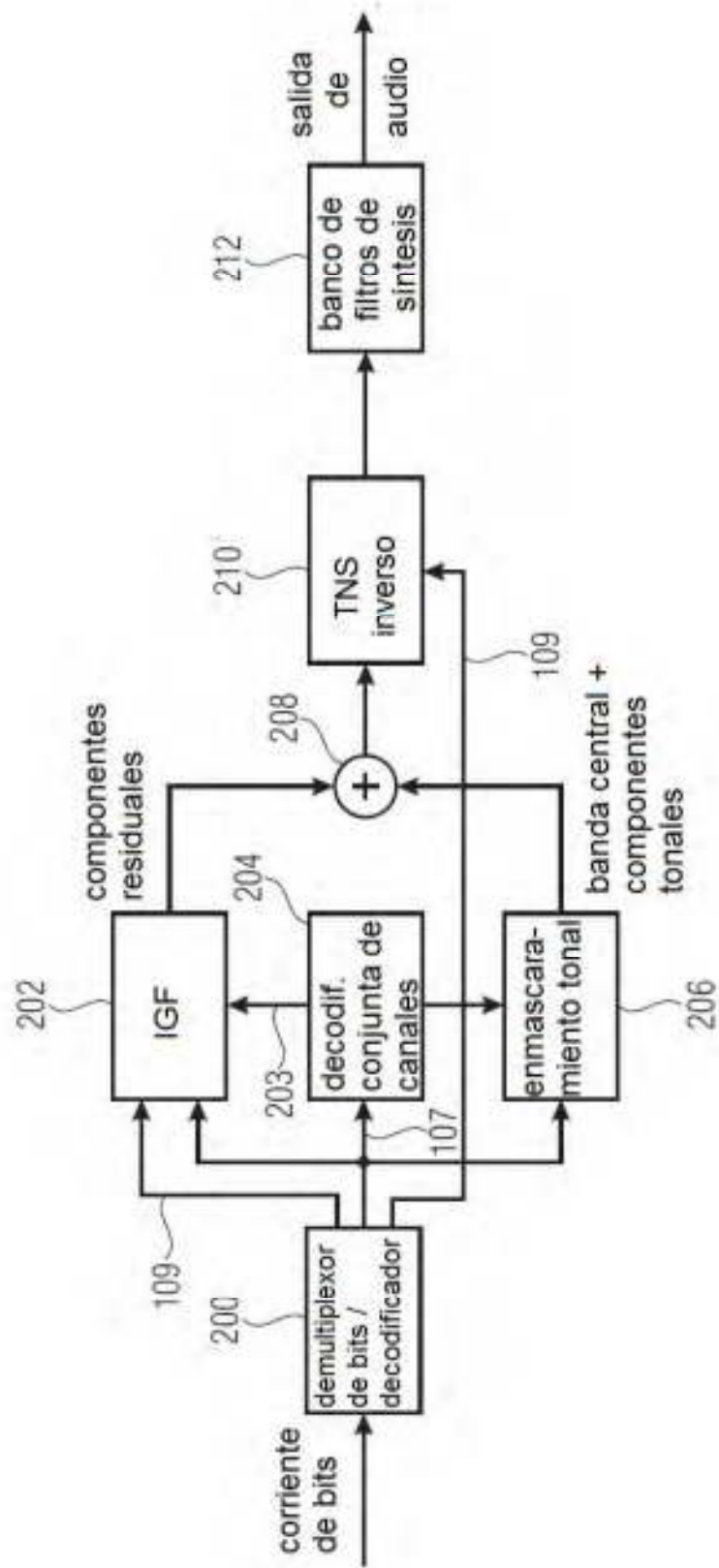


FIG 2A

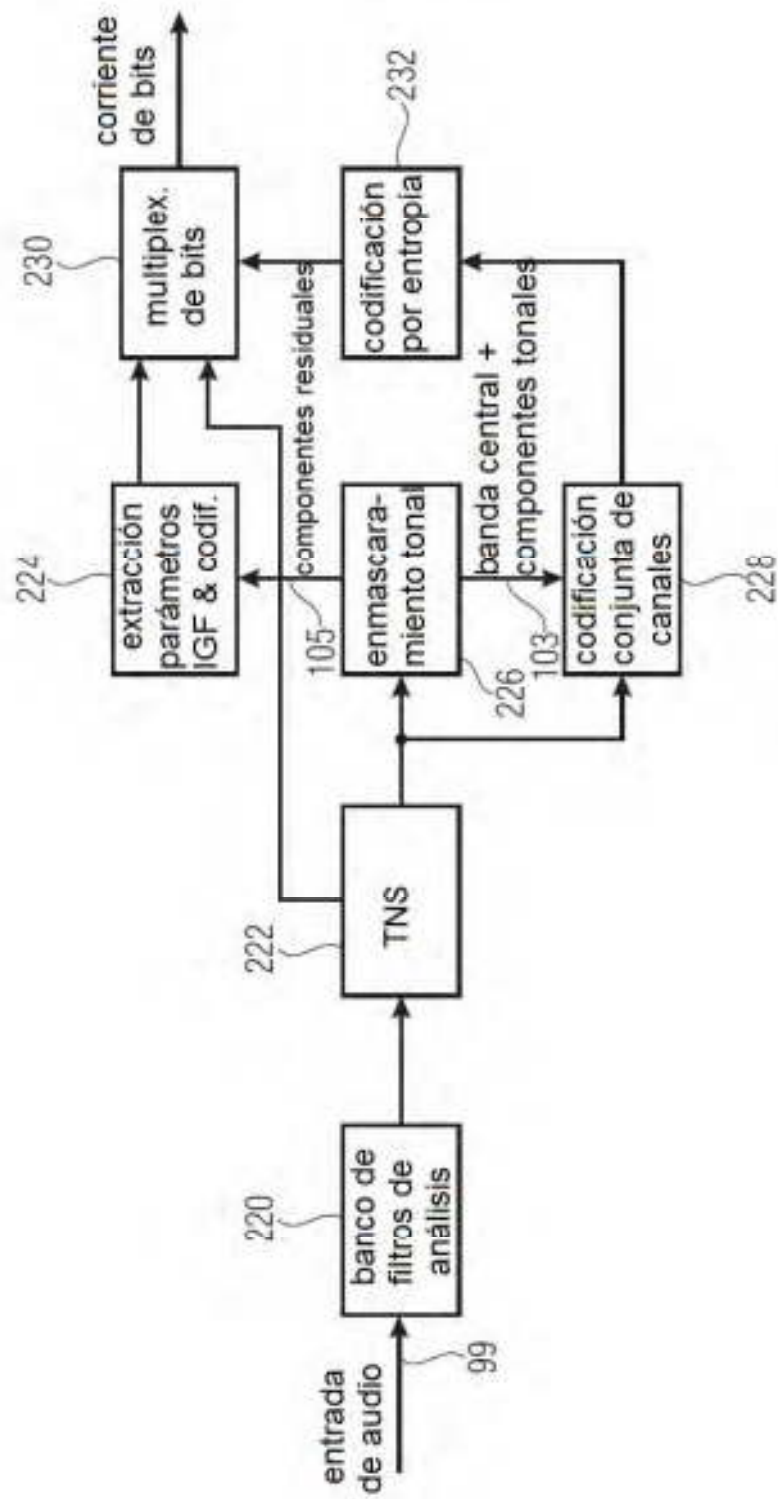


FIG 2B

- 1ra resolución (alta resolución) para la "envolvente" del 1er conjunto (codificación por líneas);
- 2da resolución (baja resolución) para la "envolvente" del 2do conjunto (factor de ajuste de escala por bandas de factor de escala (SCB));

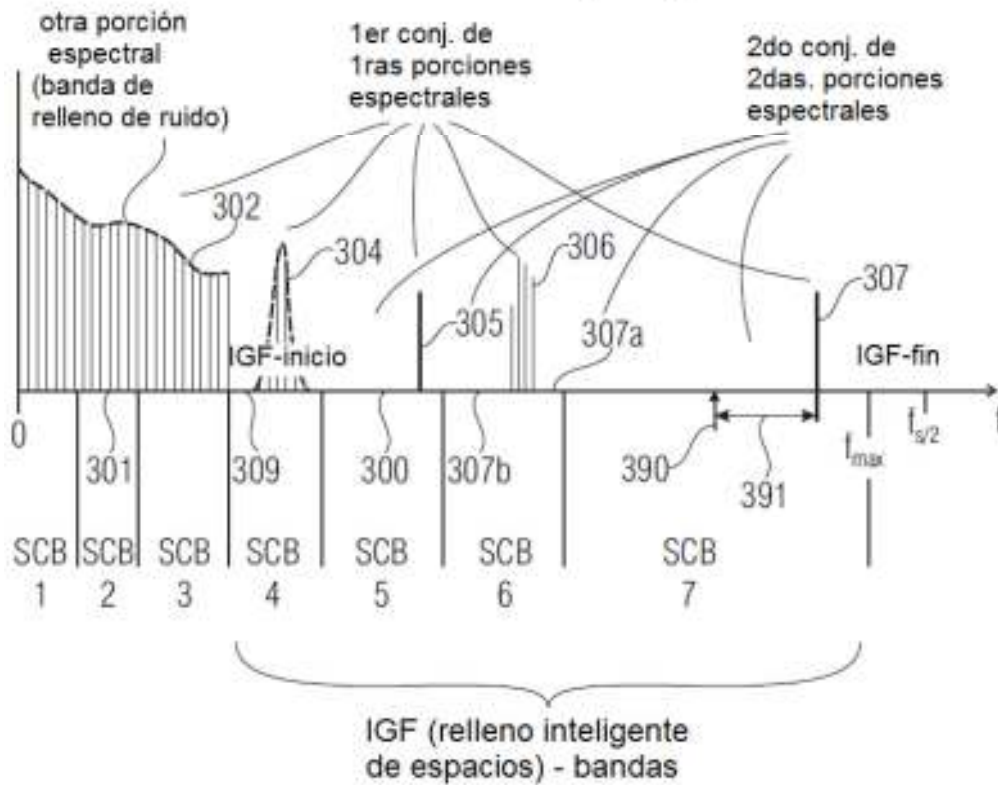


FIG 3A

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

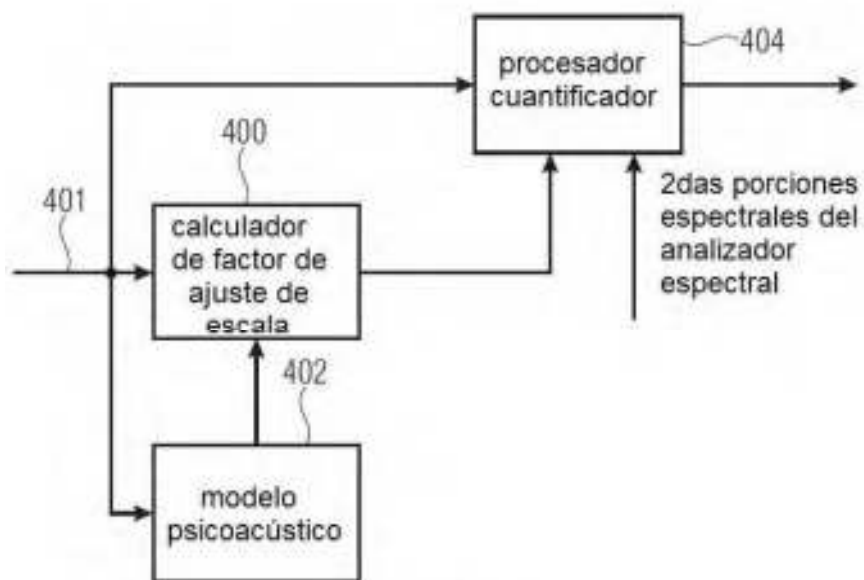


FIG 4A

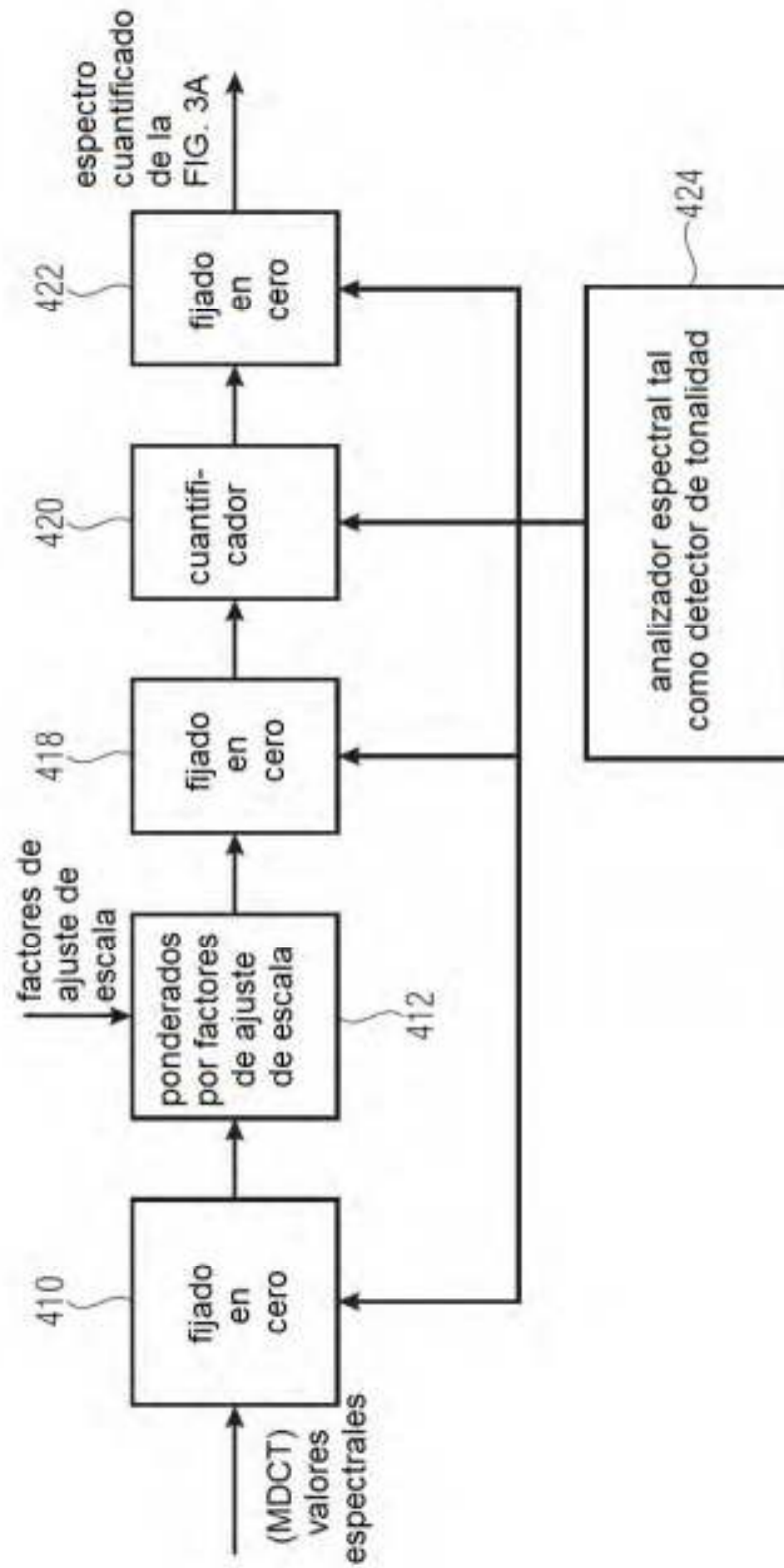


FIG 4B
(PROCESADOR CUANTIFICADOR)

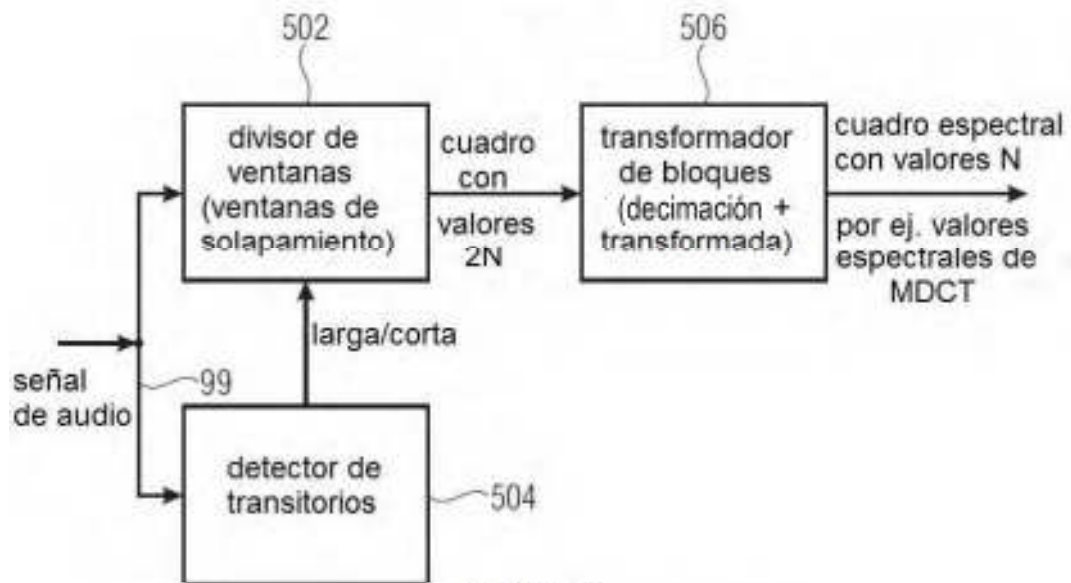


FIG 5A
(OTRAS PORCIONES ESPECTRALES)

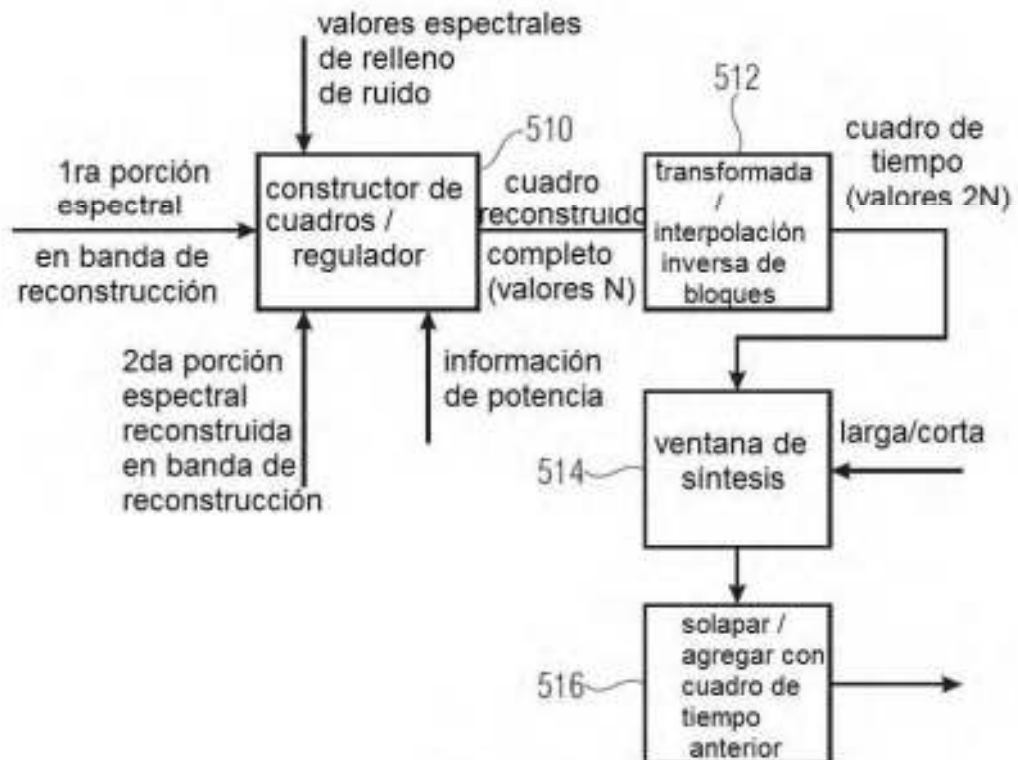


FIG 5B

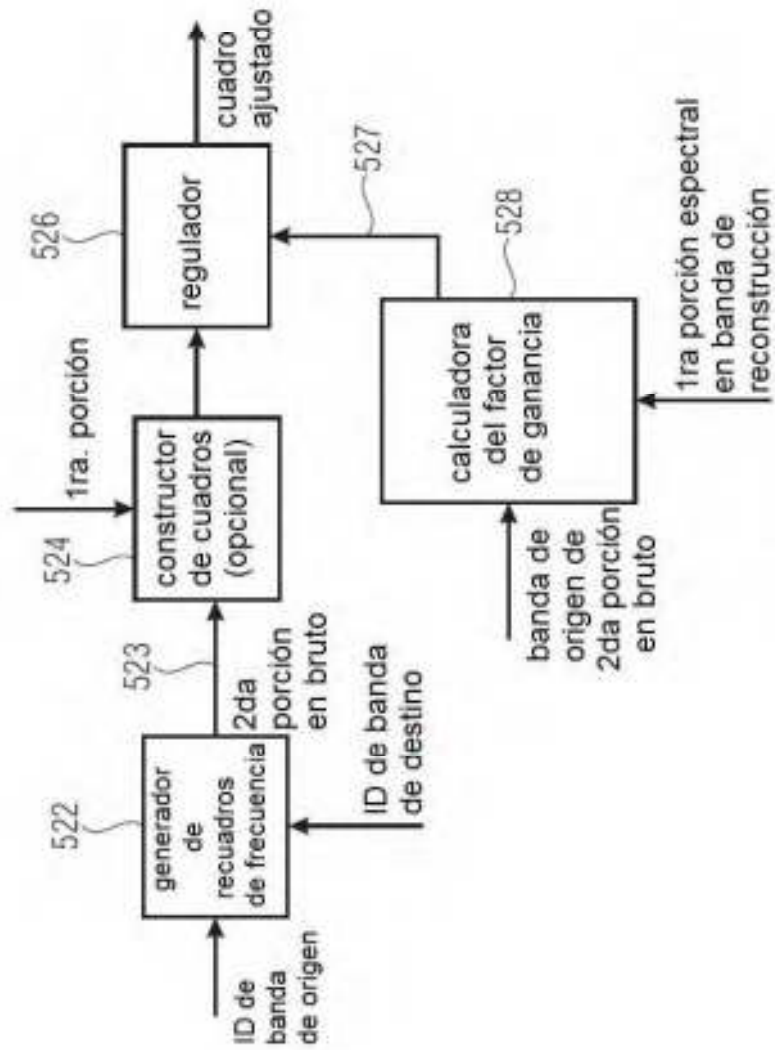


FIG 5C

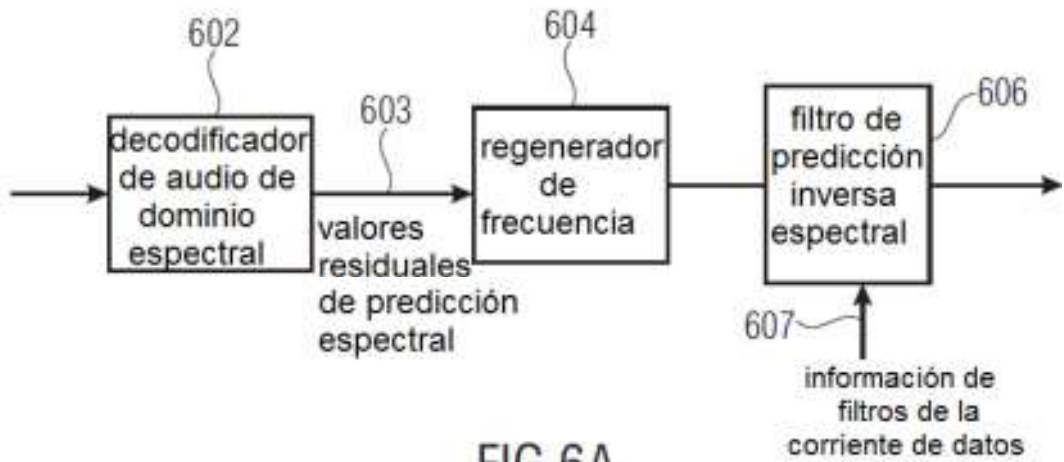


FIG 6A

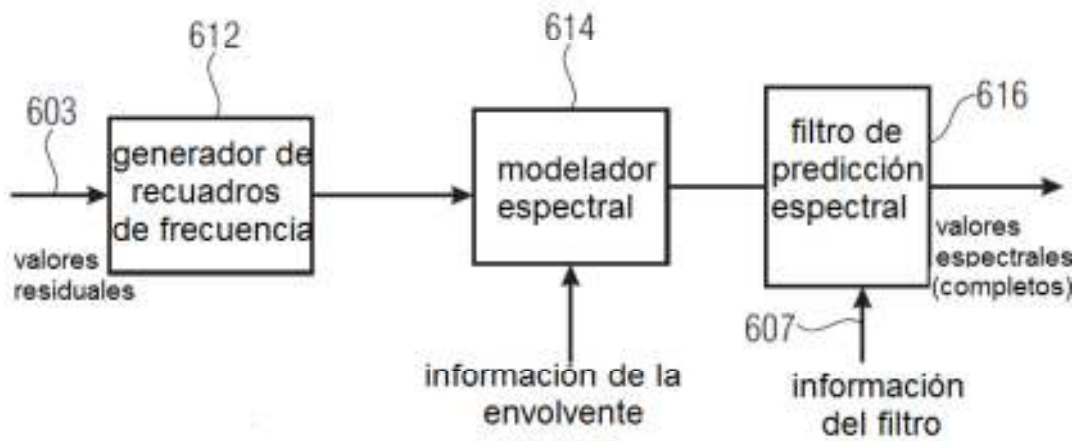


FIG 6B

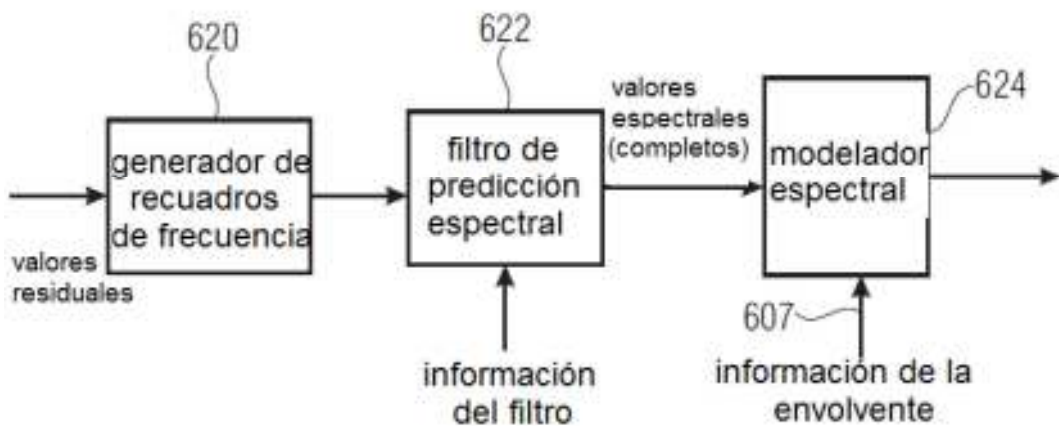
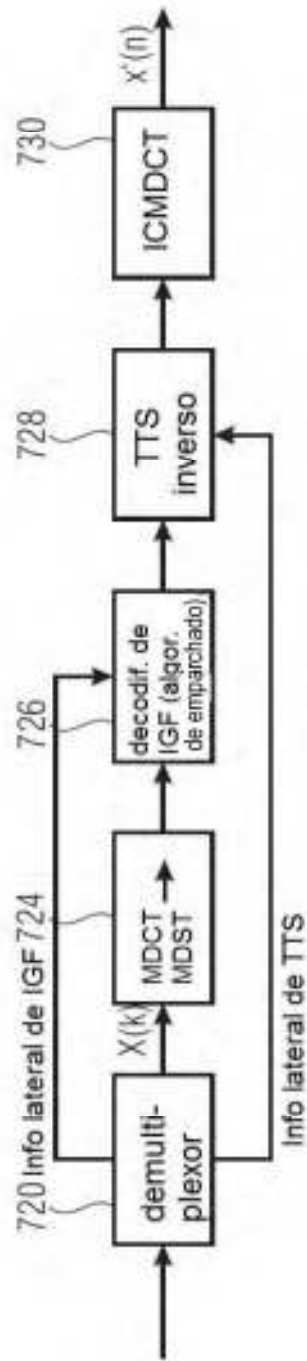
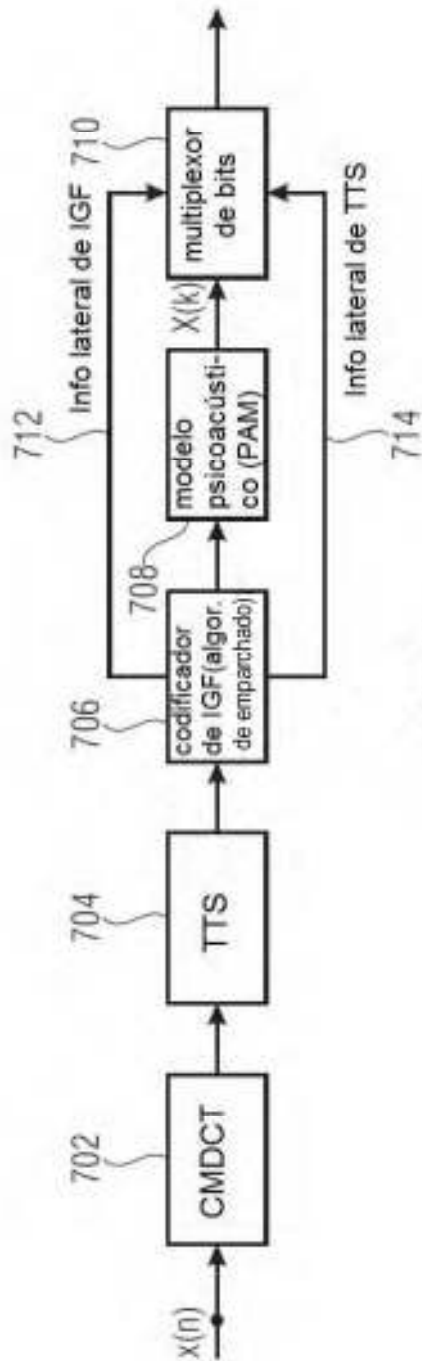
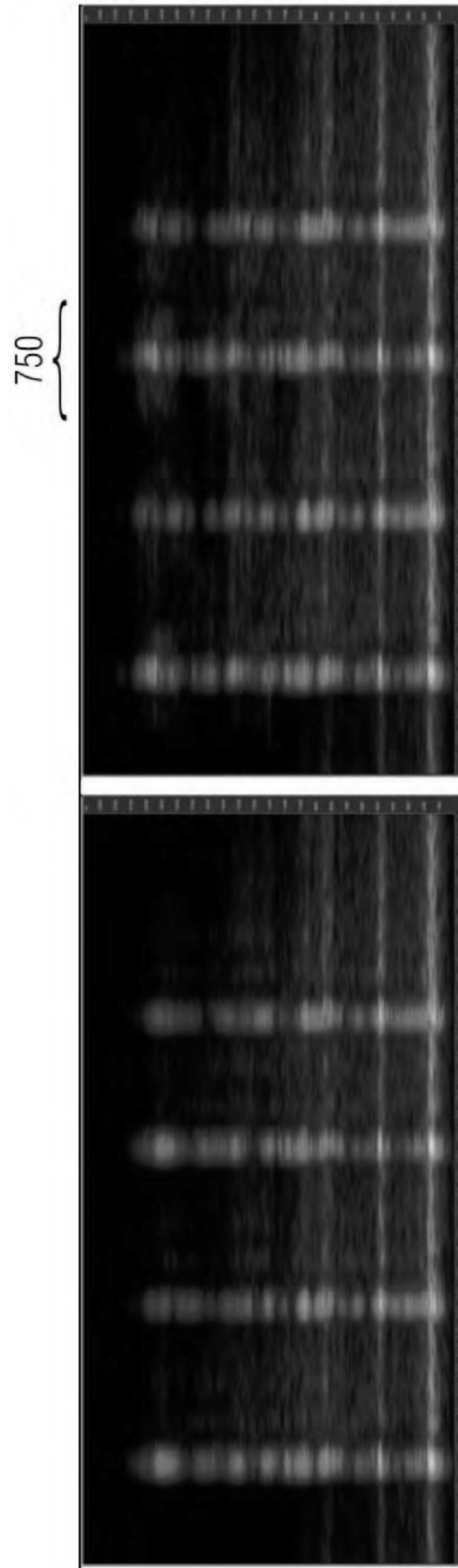


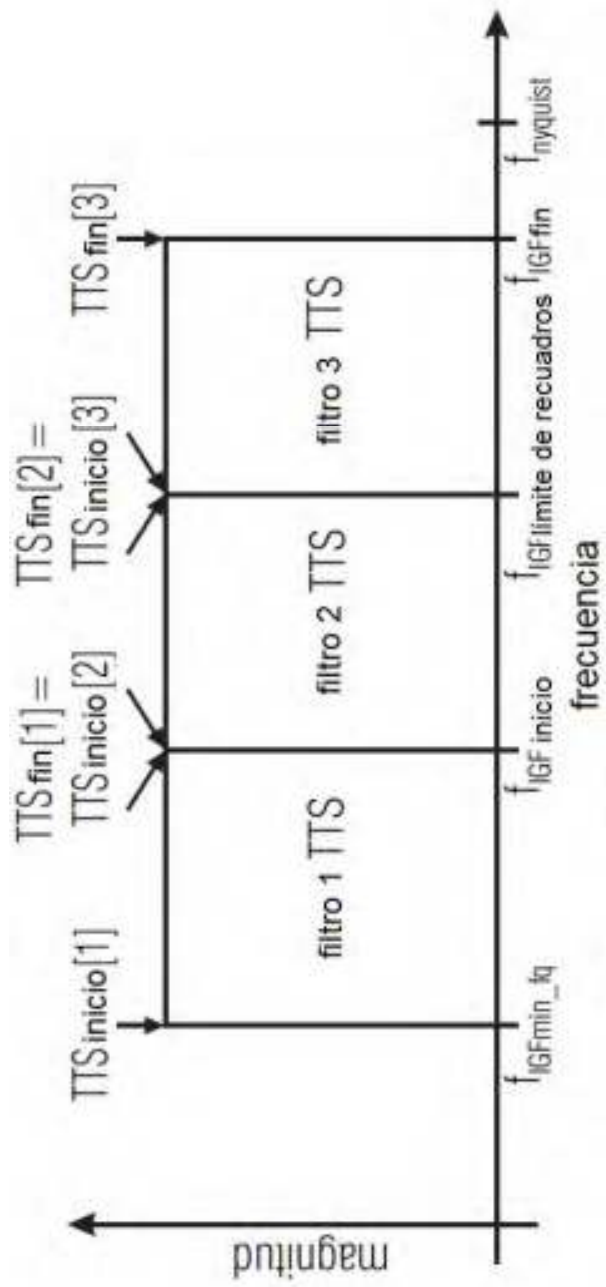
FIG 6C





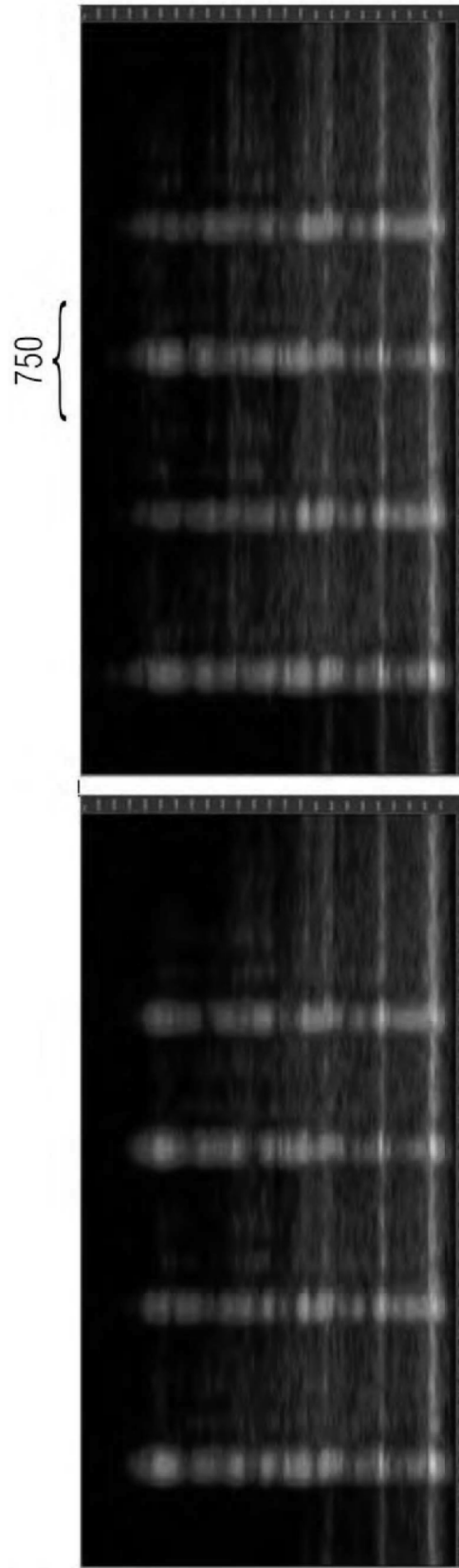
ESPECTROGRAMA DE TRANSFORMADA DISCRETA DE FOURIER (DFT) DE LA SEÑAL ORIGINAL (Panel Izquierdo);
ESPECTROGRAMA DE DFT DE LA SEÑAL EXTENDIDA SIN TTS (Panel Derecho).

FIG 7C



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

FIG 7D



ESPECTROGRAMA DE TRANSFORMADA DISCRETA DE FOURIER (DFT) DE LA SEÑAL ORIGINAL (Panel Izquierdo);
ESPECTROGRAMA DE DFT DE LA SEÑAL EXTENDIDA SIN TTS (Panel Derecho).

FIG 7E

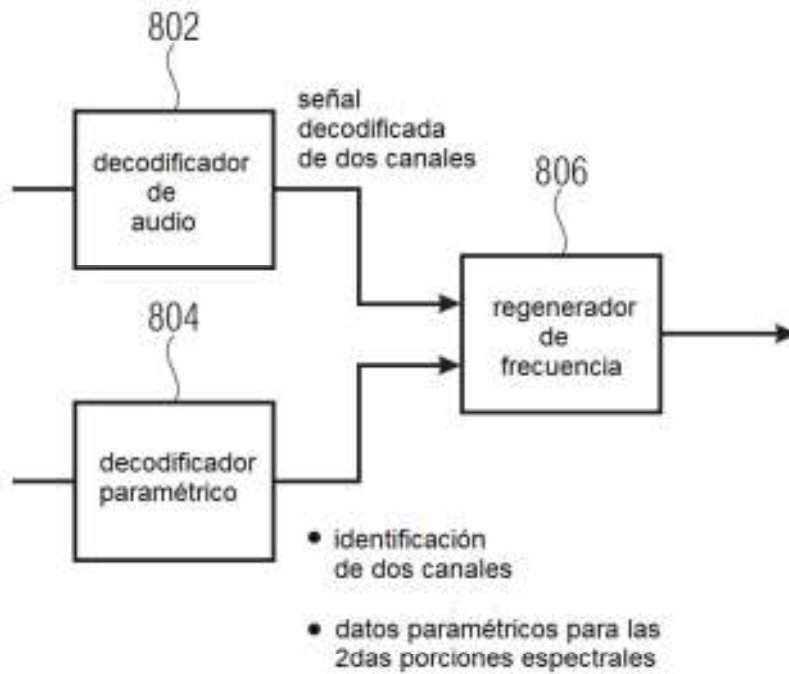


FIG 8A

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

FIG 8B

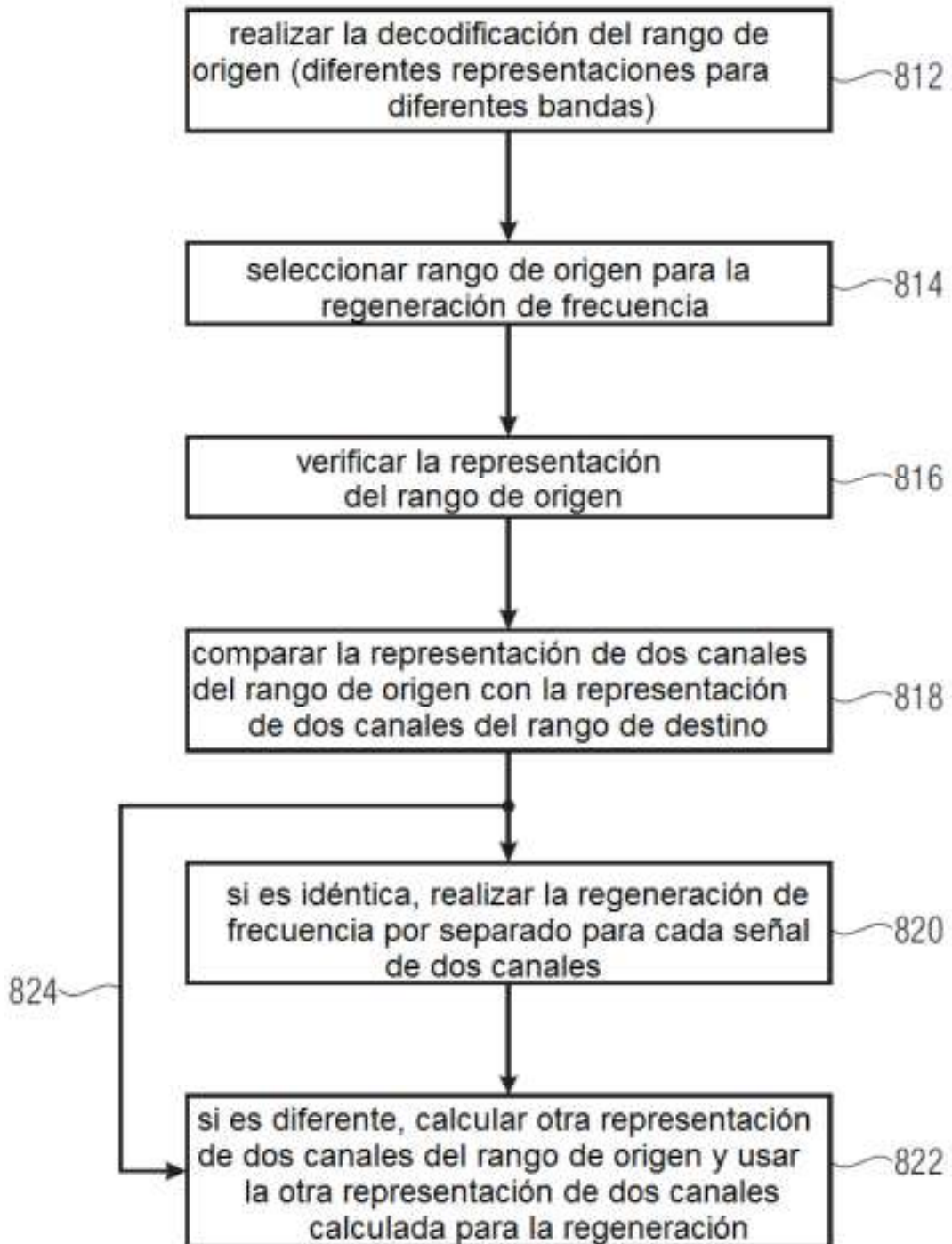


FIG 8C

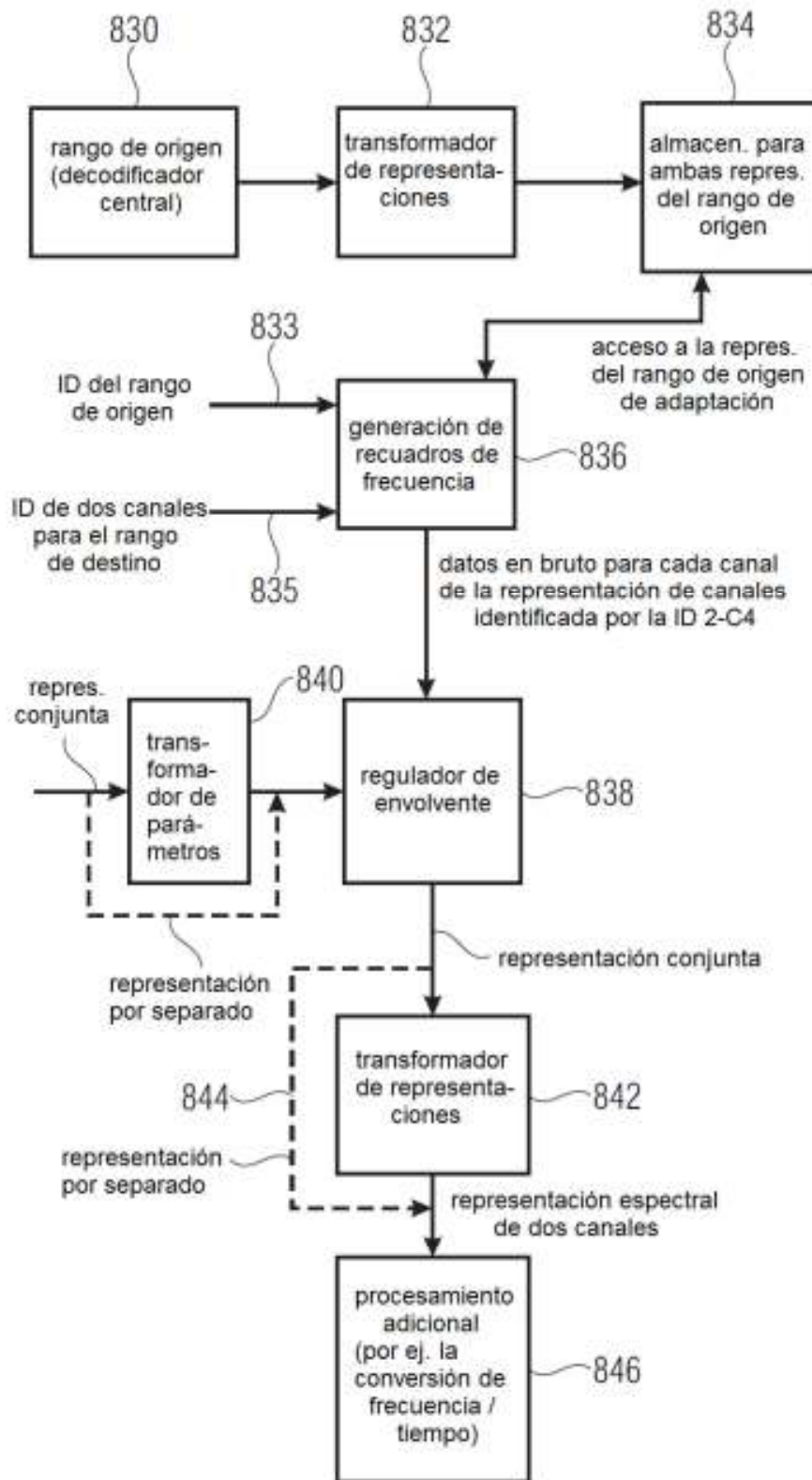


FIG 8D

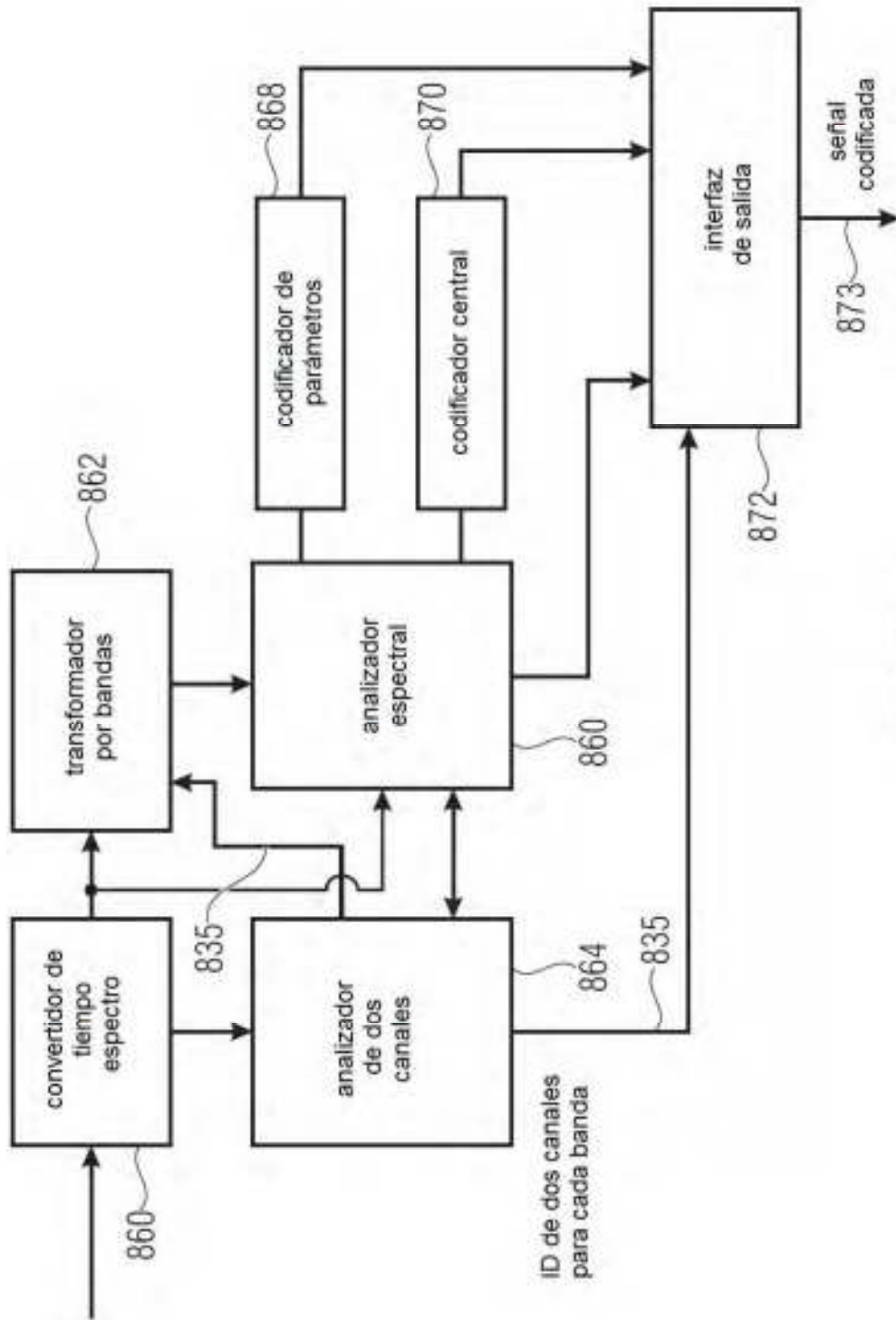


FIG 8E

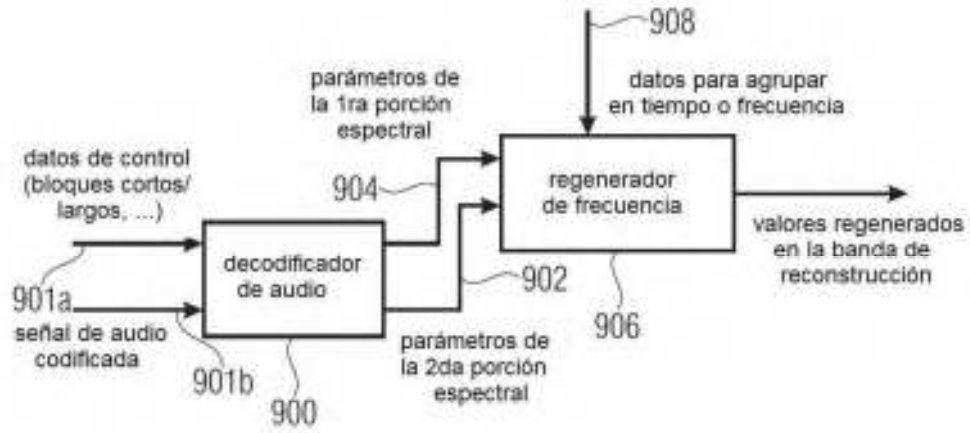


FIG 9A

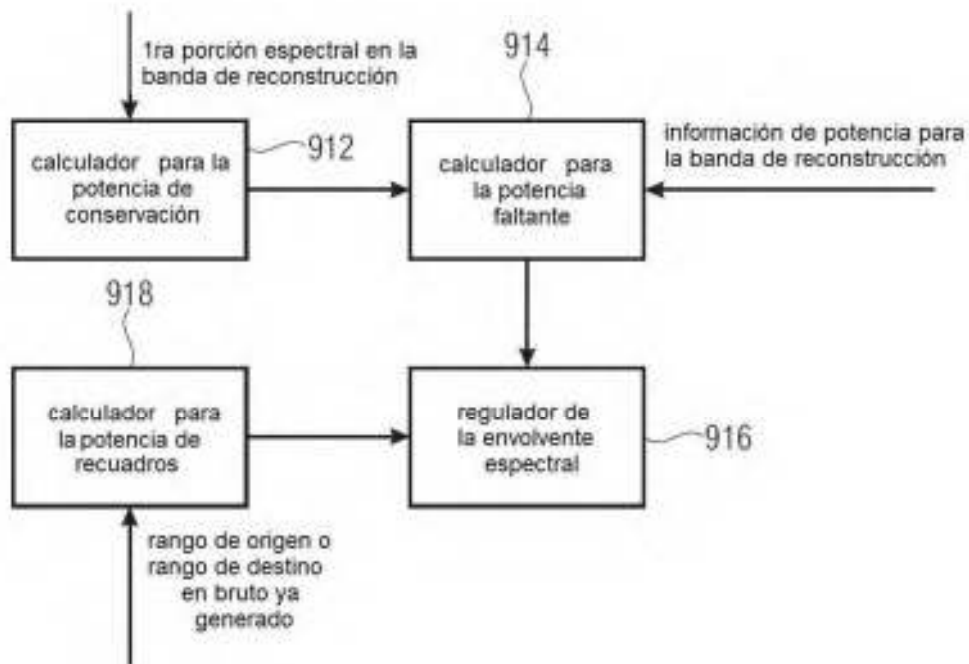
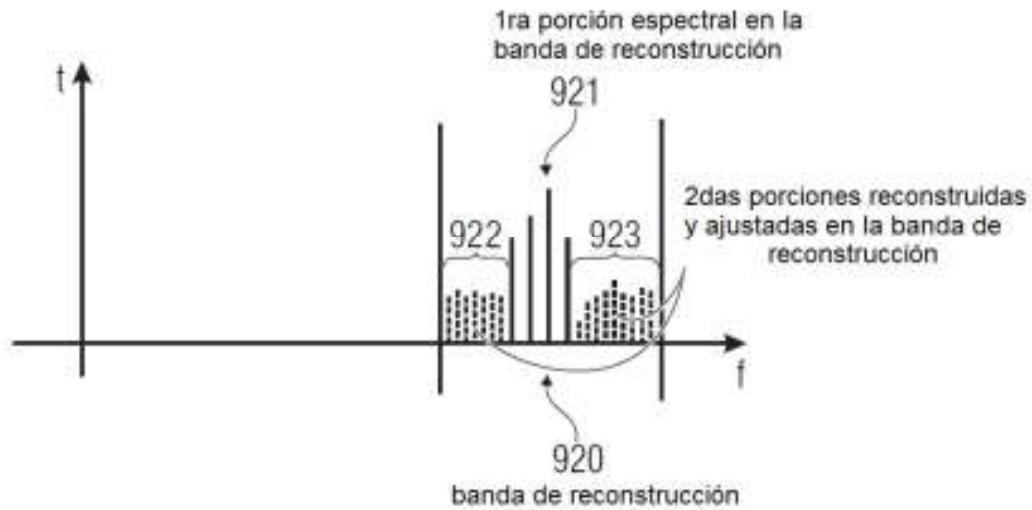


FIG 9B



por ej.

- potencia de conservación: 5 unidades
- valor de la potencia para la banda de reconstr.: 10 unidades (abarca 1ras y 2das porciones espectrales en la banda de reconstrucción)
- potencia de datos del rango de origen o datos del rango de destino en bruto: 8 unidades
- potencia faltante: 5 unidades
- factor de ganancia: $g := \sqrt{\frac{mE_k}{pE_k}} = 0,79$

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

→ la 1ra porción espectral no está influenciada por el ajuste de envolvente

FIG 9C

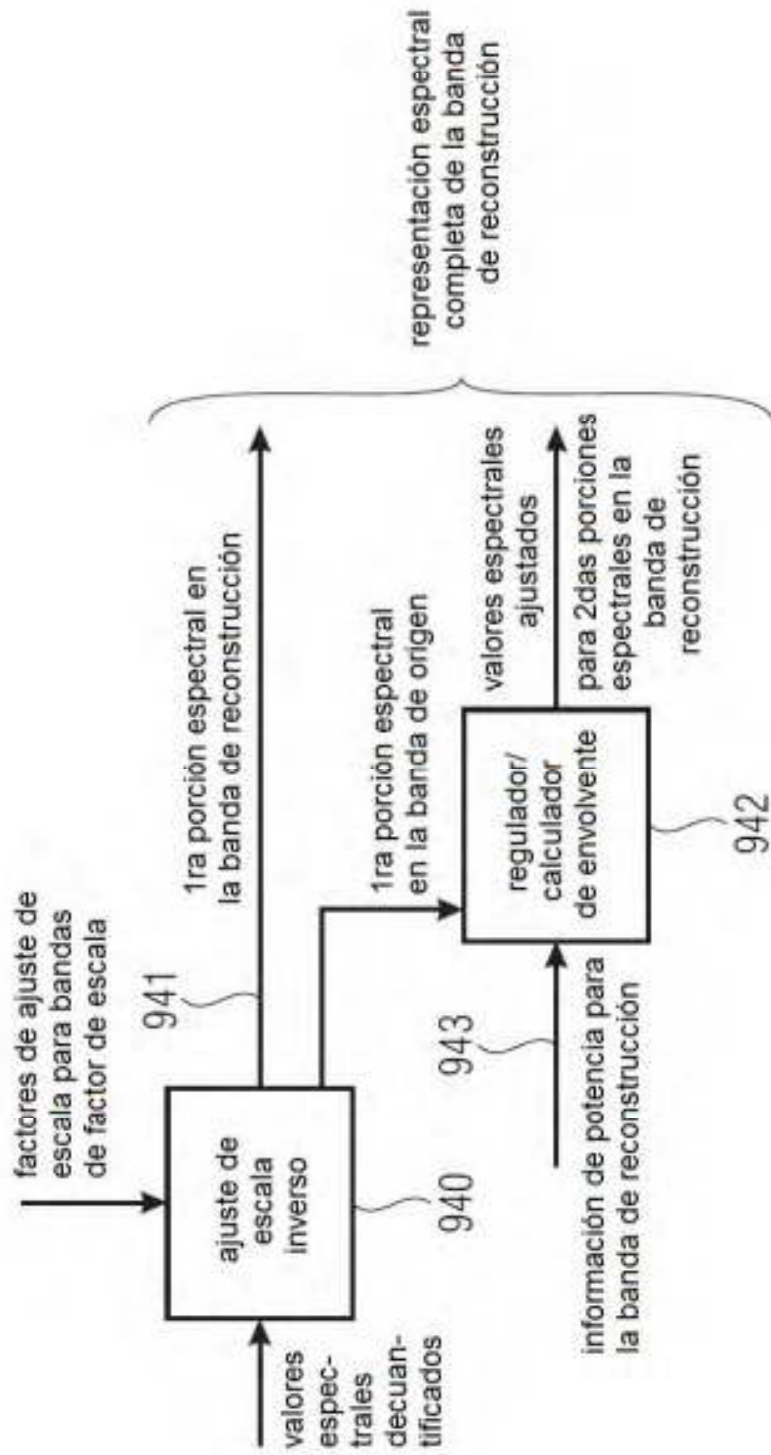


FIG 9D

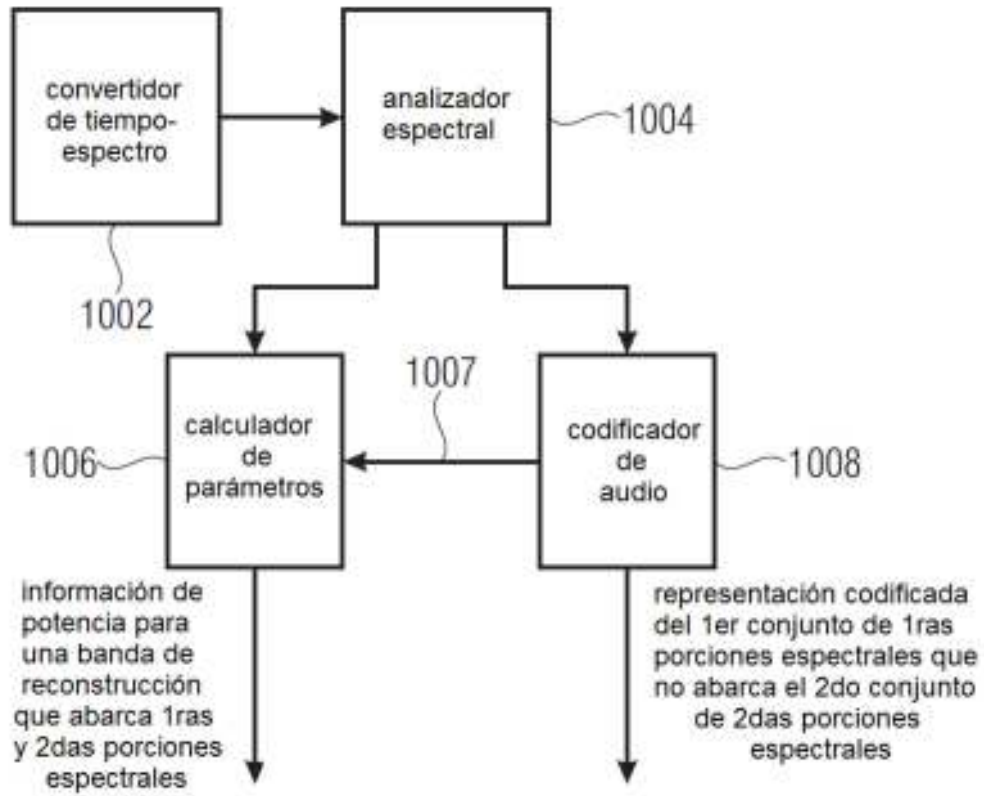


FIG 10A

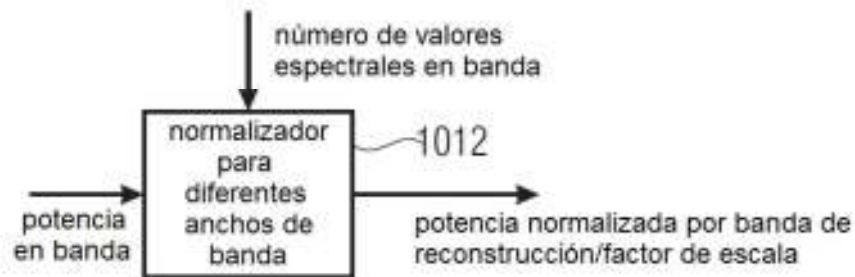


FIG 10B

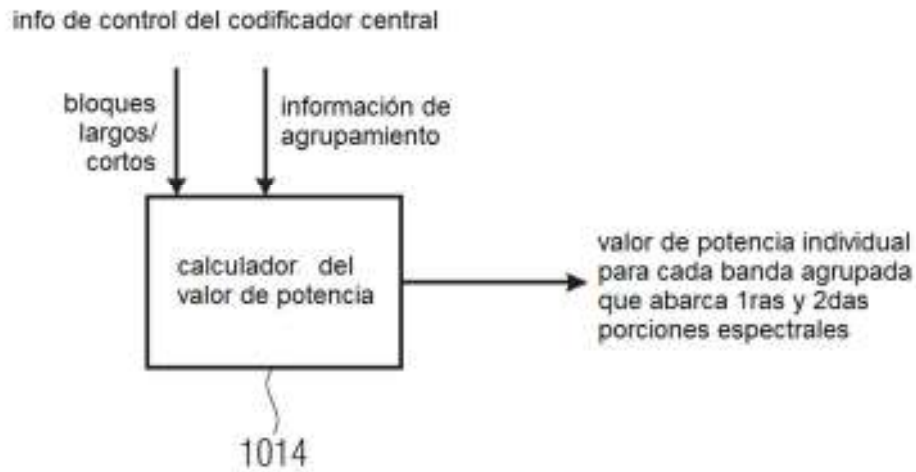


FIG 10C

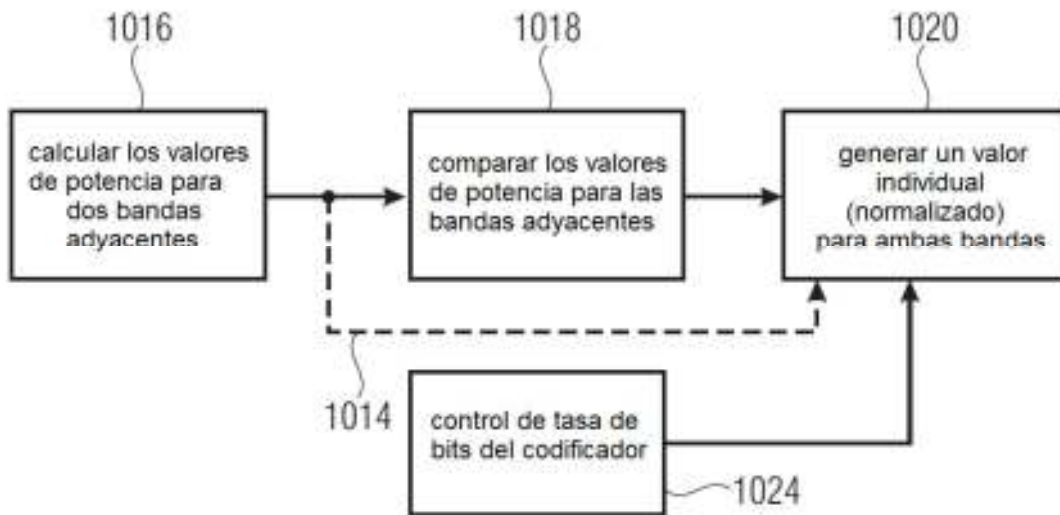


FIG 10D

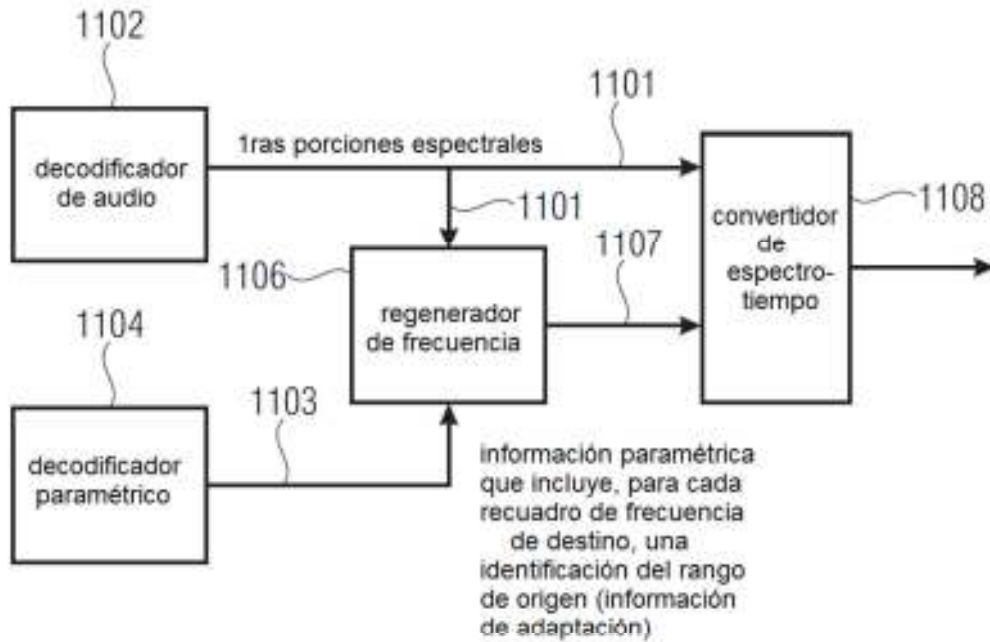


FIG 11A

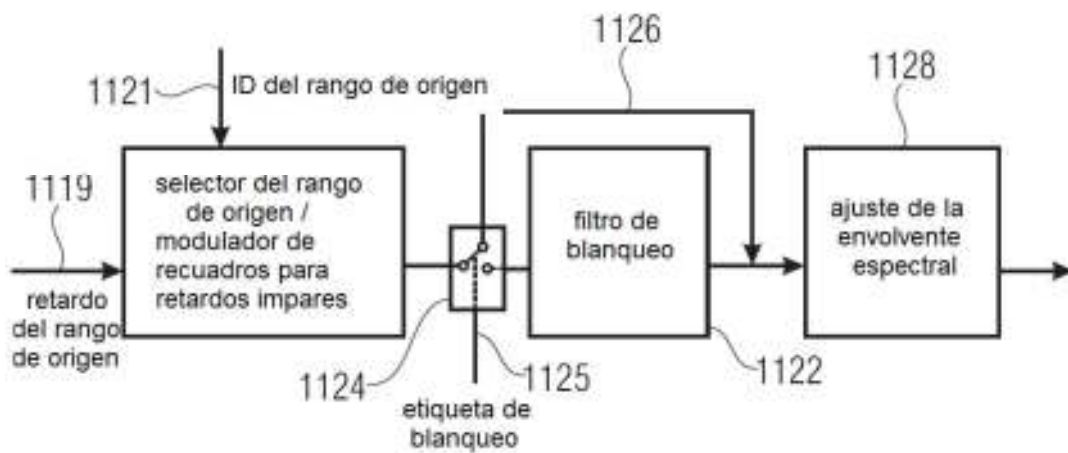


FIG 11B

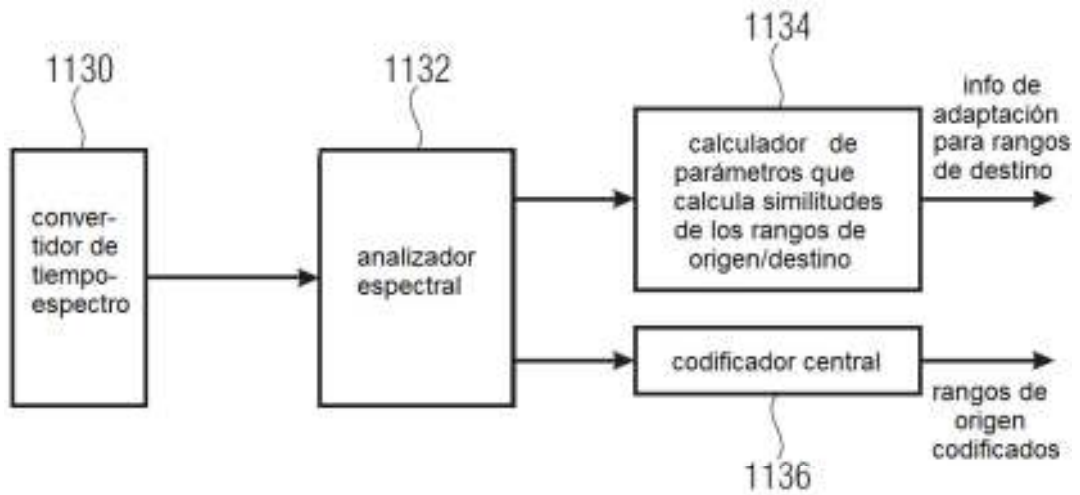


FIG 11C

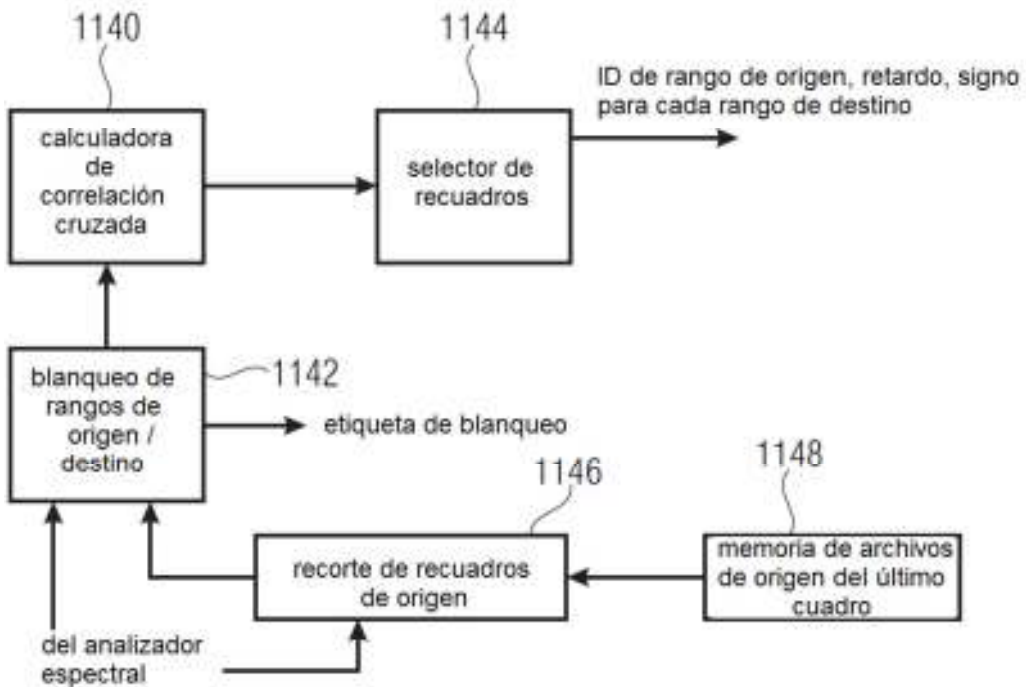


FIG 11D

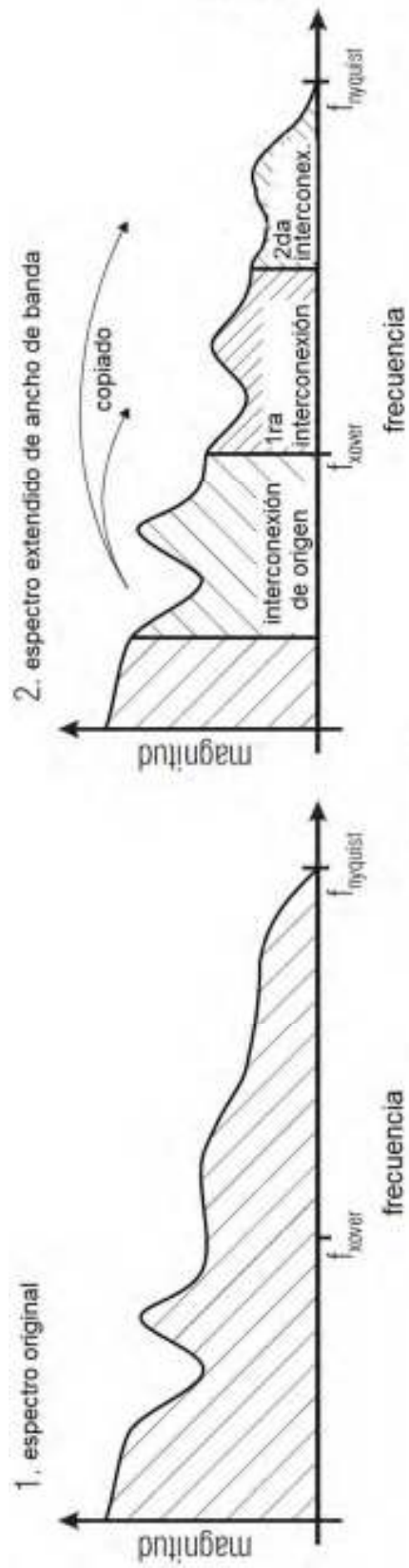


FIG 12A

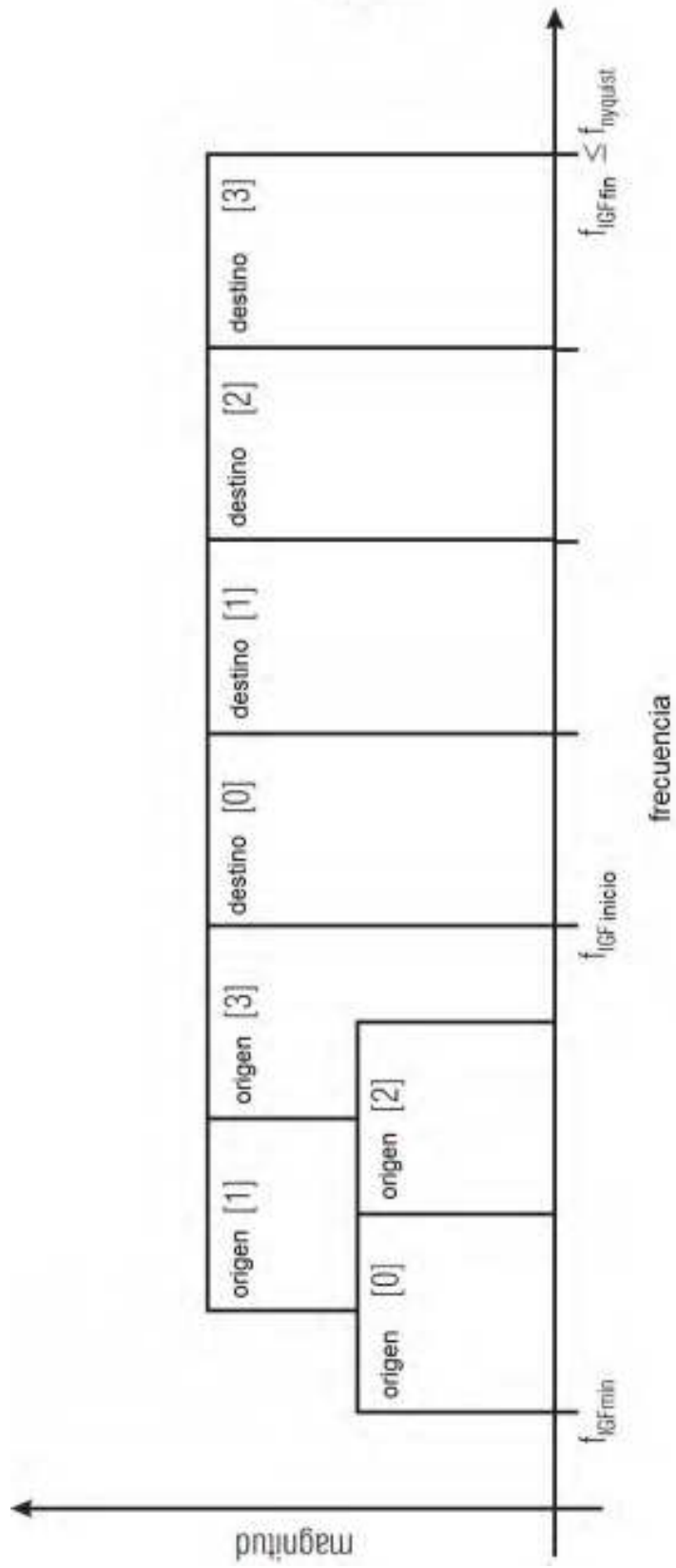


FIG 12B

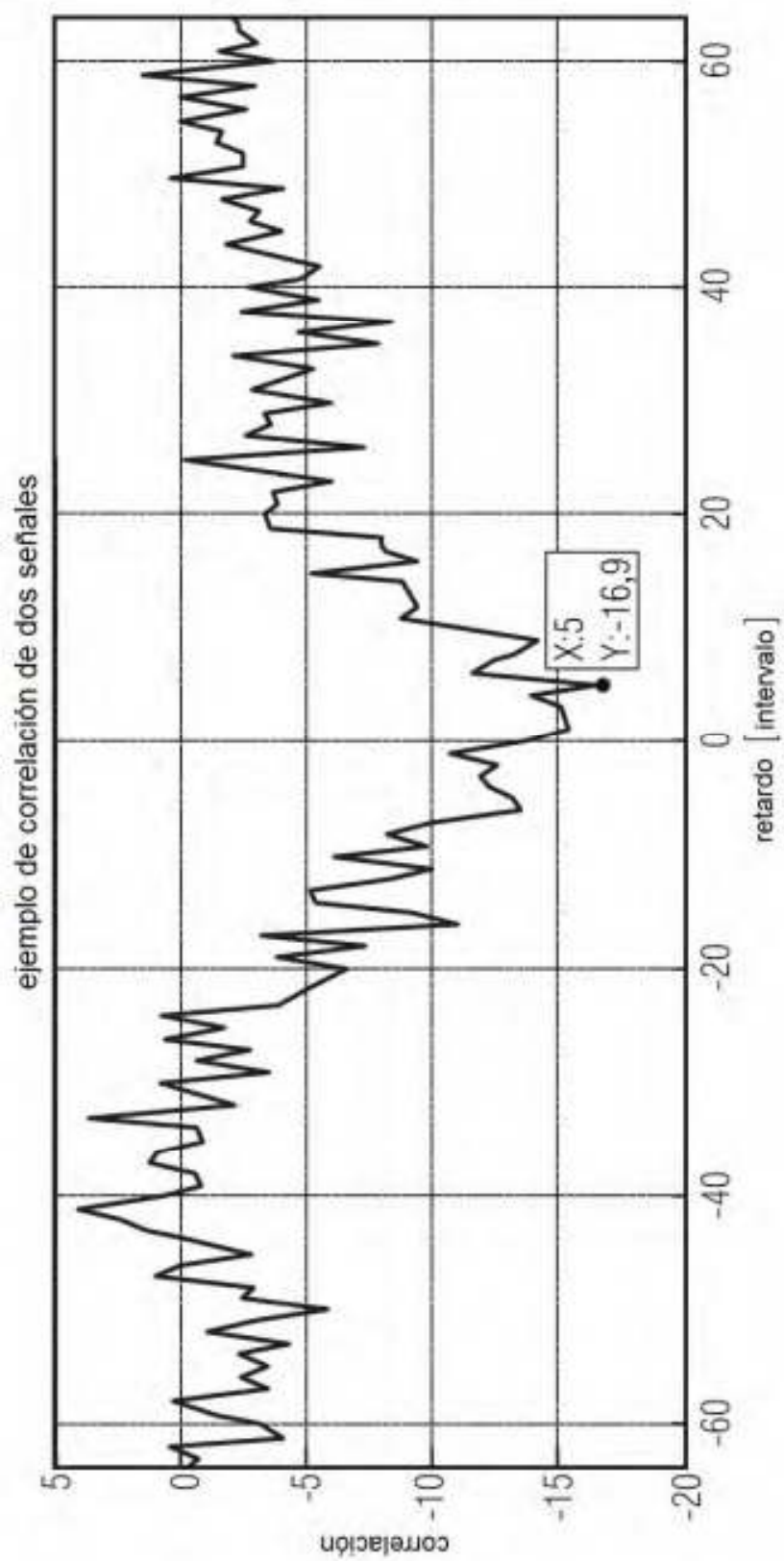


FIG 12C

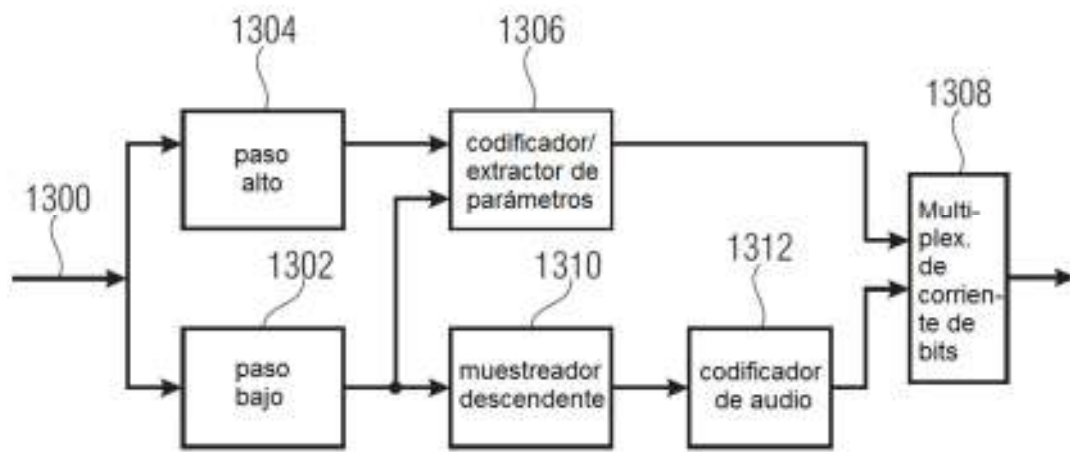


FIG 13A

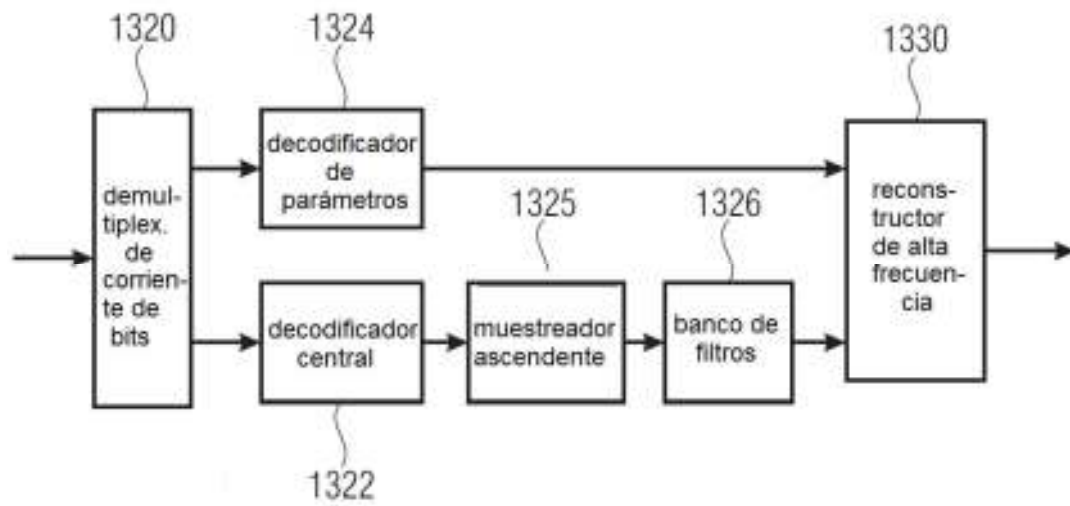


FIG 13B
(TÉCNICA ANTERIOR)