

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 667 221**

51 Int. Cl.:

**G10L 21/0388** (2013.01)

**G10L 19/02** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.07.2014 PCT/EP2014/065116**

87 Fecha y número de publicación internacional: **29.01.2015 WO15010952**

96 Fecha de presentación y número de la solicitud europea: **15.07.2014 E 14739161 (9)**

97 Fecha y número de publicación de la concesión europea: **14.02.2018 EP 3025343**

54 Título: **Aparato y método para decodificar y codificar una señal de audio utilizando selección de mosaicos espectrales adaptativos**

30 Prioridad:

**22.07.2013 EP 13177350**

**22.07.2013 EP 13177353**

**22.07.2013 EP 13177348**

**22.07.2013 EP 13177346**

**18.10.2013 EP 13189368**

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

**10.05.2018**

73 Titular/es:

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

72 Inventor/es:

**NEUKAM, CHRISTIAN;  
DISCH, SASCHA;  
NAGEL, FREDERIK;  
NIEDERMEIER, ANDREAS;  
SCHMIDT, KONSTANTIN y  
THOSKHAHNA, BALAJI NAGENDRAN**

74 Agente/Representante:

**ARIZTI ACHA, Monica**

ES 2 667 221 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 decodificar y codificar una señal de audio utilizando selección de mosaicos espectrales adaptativos

**DESCRIPCIÓN**

5 La presente invención se refiere a la codificación/decodificación de audio y, en particular, a la codificación de audio utilizando relleno inteligente de espacios.

10 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 códecs de audio generalmente necesitan alrededor de 60 kbps/canal para la codificación perceptual transparente de casi cualquier tipo de señal de audio. Los códecs 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 alta frecuencia (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 códecs de audio retengan buena calidad a tasas de bits incluso bajas de alrededor de 24 kbps/canal.

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

25 Los códecs 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) ("interconexión") 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 así denominada frecuencia de cruce determinada se basa a menudo en la interconexión espectral. En general, la región de alta frecuencia (HF) consta de 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 la interconexión 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 sub-bandas adyacentes desde una región de origen a la región de destino.

40 Otra técnica que se encuentra en los códecs de audio actuales que aumenta la eficiencia de compresión y permite así el ancho de banda de audio extendida 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 tipo ruido de la señal de audio original pueden ser reemplazadas 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 extendida 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 gruesa, posteriormente se llenan de ruido artificial en el decodificador y se ajustan a escala por un procesamiento posterior basado en parámetros.

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 dedicada 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 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).

60 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 es introducida 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 ser codificado 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 subsiguiente 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 resintetizan 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 de 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 fallos 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, potencia 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 ordenadores portátiles, etc.

Los códecs de audio actuales llevan a cabo 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 códecs de audio contemporáneos, ya sea 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 Coseno Discreta Modificada (MDCT) al dominio de Filtros Espejo en Cuadratura (QMF). Esto genera complicaciones de sincronización, complejidad computacional adicional y aumento

de requisitos de memoria.

5 Típicamente, los esquemas de extensión de ancho de banda utilizan interconexión espectral con el fin de  
reconstrucción de la región espectral de alta frecuencia por encima de una así denominada frecuencia de cruce  
determinada. La región de alta frecuencia (HF) consta de múltiples conexiones adyacentes y cada una de estas  
conexiones se obtiene de la misma región de paso de banda del espectro de baja frecuencia por debajo de la  
frecuencia de cruce determinada. Dentro de una representación de banco de filtros de las señales, dichos sistemas  
copian un conjunto de coeficientes de sub-bandas adyacentes del espectro de baja frecuencia en la región de alta  
frecuencia (HF). Los límites de los conjuntos seleccionados suelen depender del sistema y no dependen de la señal.  
10 Para algunos contenidos de la señal, esta selección de interconexión estática puede provocar un timbre  
desagradable y la coloración de la señal reconstruida.

15 Otros enfoques transfieren la señal de baja frecuencia (LF) a la región de alta frecuencia (HF) a través de una  
modulación de banda lateral única adaptativa de la señal (SSB, por sus siglas en inglés). Estos enfoques son de alta  
complejidad computacional en comparación con los procedimientos de copiado, ya que operan a alta velocidad de  
muestreo en señales de dominio temporal.

20 Por otra parte, la interconexión puede volverse inestable, especialmente para señales no tonales como, por ejemplo,  
la voz sorda. Por lo tanto, los esquemas conocidos de interconexión pueden introducir alteraciones en la señal de  
audio.

25 La publicación "Hybrid Low Bitrate Audio Coding Using Adaptive Gain Shape Vector Quantization", S. Mehrotra, et  
al. páginas 927 a 931, XP 31356759 A divulga codificación de audio a bajas tasas de bits. Se divulga un nuevo  
esquema para codificar señales de audio a bajas tasas de bits que utiliza cuantización escalar tradicional seguida  
por codificación de entropía para codificar algunas porciones del espectro. Las otras porciones del espectro se  
codifican a una baja tasa de bits usando un cuantificador de vector de forma de ganancia adaptativa, donde el libro  
de códigos para una cuantificación del vector se forma mediante versiones no modificadas o modificadas de las  
porciones del espectro que ya se han codificado.

30 El documento US 2005/0165611 A1 divulga una codificación eficiente de datos espectrales de medios usando  
similitud perceptual de amplio sentido. La codificación de manera perceptual de versiones similares de coeficientes  
espectrales omitidas se codifica, y se representan como una versión escalada de datos espectrales ya codificados.  
Los coeficientes espectrales omitidos se dividen en una serie de sub-bandas. Estas sub-bandas se codifican como  
dos parámetros, un factor de escala y un parámetro de forma.

35 Un objetivo de la presente invención consiste en proporcionar un concepto mejorado de codificación/decodificación  
que produzca una mejor calidad de audio.

40 Este objetivo se logra mediante un aparato para decodificación de acuerdo con la reivindicación 1, un aparato para  
codificación de acuerdo con la reivindicación 7, un método de decodificación de audio de acuerdo con la  
reivindicación 17, o un método de decodificación de audio de acuerdo con la reivindicación 18 o un programa  
informático de acuerdo con la reivindicación 19.

45 La presente invención 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 mosaicos de frecuencia adaptativa de señal. Para 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 mosaicos  
de frecuencia utilizando la información de adaptación. Para este fin, el decodificador lee la información de  
50 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, lleva  
a cabo 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 mosaicos de frecuencia, es  
decir, los datos espectrales en bruto para la banda de reconstrucción, se modela utilizando información de la  
55 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 mosaicos adaptativo, pero estas primeras porciones espectrales son emitidas  
por el decodificador de audio o el decodificador central directamente.

60 El esquema de selección de mosaicos 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 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.

- 5 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 fina como un intervalo de frecuencia y permite incluso una mejor adaptación entre una región de origen y la región de destino.

10 Asimismo, además de solo identificar un mejor par de adaptación, este retardo de correlación también puede ser transmitido dentro de la información de adaptación y, adicionalmente, incluso puede transmitirse un signo. Cuando 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.

15 Una implementación adicional de la presente invención aplica una operación de blanqueo de mosaicos. El blanqueo de un espectro quita la información de envolvente espectral en bruto y enfatiza la estructura fina espectral que es de interés principal para la evaluación de la similitud de mosaicos. Por lo tanto, un mosaico 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 mosaico utilizando un procedimiento predefinido, se transmite una etiqueta de blanqueo que indica al decodificador que se aplicará el mismo proceso de blanqueo predefinido al mosaico de frecuencia dentro de IGF.

20 En cuanto a la selección de mosaicos, 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 mosaico 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 Coseno Discreta Modificada (MDCT). Además, el signo del resultado de la correlación se aplica cuando se genera el mosaico de frecuencia.

30 Asimismo, se prefiere utilizar el recorte y la estabilización de mosaicos con el fin de asegurar que se evitan los fallos creadas por el cambio rápido de regiones de origen para la misma región de reconstrucción o la región de destino. Para ello se lleva a cabo un análisis de similitud entre las diferentes regiones de origen identificadas y cuando un mosaico de origen es similar a otros mosaicos de origen con una similitud por encima de un umbral, entonces este mosaico de origen puede ser retirado del conjunto de mosaicos de origen potenciales, ya que está muy correlacionado con otros mosaicos de origen. Además, como un tipo de estabilización de selección de mosaicos, se prefiere mantener el orden del cuadro anterior si ninguno de los mosaicos de origen en el cuadro actual se correlaciona (mejor que un umbral determinado) con los mosaicos de destino en el cuadro actual.

40 Un aspecto adicional se basa en el descubrimiento de que la calidad de audio de la señal reconstruida se puede mejorar a través de IGF ya que todo el espectro es accesible para el codificador central de manera que, por ejemplo, las porciones tonales perceptualmente importantes en un alto rango espectral todavía pueden ser codificadas por el codificador central en lugar de la sustitución paramétrica. Adicionalmente se lleva a cabo la operación de relleno de espacios que utiliza mosaicos de frecuencia de un primer conjunto de primeras porciones espectrales que es, por ejemplo, un conjunto de porciones tonales generalmente de un rango de frecuencia inferior, pero también de un rango de frecuencia superior, si está disponible. Sin embargo, para el ajuste de la envolvente espectral en el lado del decodificador, las porciones espectrales del primer conjunto de porciones espectrales ubicadas en la banda de reconstrucción no se procesan posteriormente otra vez por el ajuste de la envolvente espectral, por ejemplo. Solo los valores espectrales restantes en la banda de reconstrucción que no provienen del decodificador central deben ajustarse por la envolvente utilizando información de la envolvente. Preferentemente, la información de la envolvente es una información de la envolvente de banda completa que representa la potencia del primer conjunto de primeras porciones espectrales en la banda de reconstrucción y el segundo conjunto de segundas porciones espectrales en la misma banda de reconstrucción, en donde los últimos valores espectrales en el segundo conjunto de segunda porciones espectrales se indican como cero y, por lo tanto, no están codificados por el codificador central, pero están paraméricamente codificados con información de potencia de baja resolución.

50 Se ha descubierto que los valores absolutos de potencia, ya sea normalizados o no con respecto al ancho de banda de la banda correspondiente, son útiles y muy eficientes en una aplicación en el lado del decodificador. Esto se aplica especialmente cuando los factores de ganancia tienen que calcularse sobre la base de una potencia residual en la banda de reconstrucción, la potencia faltante en la banda de reconstrucción y la información de mosaicos de frecuencia en la banda de reconstrucción.

60 Asimismo, se prefiere que la corriente de bits codificada no solo abarque información de potencia para las bandas de reconstrucción sino, además, factores de ajuste de escala para bandas de factor de escala que se extienden hasta

la frecuencia máxima. Esto asegura que, para cada banda de reconstrucción, para la que está disponible una cierta porción tonal, es decir, una primera porción espectral, este primer conjunto de primera porción espectral pueda ser decodificado realmente con la amplitud correcta. Por otra parte, además del factor de escala para cada banda de reconstrucción, una potencia para esta banda de reconstrucción es generada en un codificador y transmitida a un decodificador. Además, se prefiere que las bandas de reconstrucción coincidan con las bandas de factor de escala o en el caso de agrupamiento de potencia, al menos los límites de una banda de reconstrucción coincidan con los límites de las bandas de factor de escala.

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 llevando a cabo 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 lleva a cabo en toda la tasa de muestreo o en todo el dominio de ancho de banda. Con el fin de obtener una alta ganancia de codificación, la señal de audio se analiza con el fin de encontrar un primer conjunto de primeras porciones espectrales que tiene que ser codificado con una alta resolución, en donde este primer conjunto de primeras porciones espectrales puede incluir, en una realización, porciones tonales de la señal de audio. Por otra parte, los componentes no tonales o ruidosos en la señal de audio que constituyen un segundo conjunto de segunda porciones espectrales son codificados paraméricamente 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 paraméricamente con una baja resolución utilizando "mosaicos" 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 paraméricos, 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 realizaciones, las porciones espectrales, que son reconstruidas por relleno de ruido solamente en lugar de replicación de ancho de banda o relleno de mosaicos 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 mosaico de frecuencia en un rango de frecuencia diferente.

Asimismo, se insiste en que una información sobre potencias espectrales, una información sobre potencias individuales o una información de potencia individual, una información sobre una potencia de conservación o una información de potencia de conservación, una información sobre una potencia de mosaico o una información de potencia de mosaico, o una información sobre una potencia faltante o una información de potencia faltante pueden comprender no solo un valor de potencia 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 potencia final. Por lo tanto, la información sobre una potencia puede comprender, por ejemplo, el valor de potencia 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 está 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 lleva a cabo 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 fallos de segregación espacial en la señal reconstruida. Con el fin de abordar esta cuestión, los datos paraméricos para una banda de reconstrucción o, en general, para el segundo conjunto de segundas porciones espectrales que tiene que ser reconstruido 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 potencia 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 potencia de la envolvente espectral o cualquier otro dato de la envolvente espectral y, adicionalmente, depende 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 luego 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 mosaicos de frecuencia de relleno inteligente de espacios (IGF), se elige la representación de mosaicos de origen para adaptar la representación de mosaicos 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.

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 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 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) o Modelado de Mosaico Temporal (TTS, por sus siglas en inglés) con reconstrucción de alta frecuencia. El procesamiento de Modelado de Ruido Temporal (TNS)/Modelado de Mosaico Temporal (TTS) en el lado del codificador implementado por una predicción sobre la frecuencia reconstruye la envolvente de tiempo de la señal de audio. Dependiendo de la implementación, es decir, cuando el filtro de modelado de ruido temporal se determina dentro de un rango de frecuencia no solo abarcando el rango de frecuencia de origen sino también el rango de frecuencia de destino para ser reconstruido 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 espectrales de segundas porciones espectrales reconstruidas. Por lo tanto, los pre-ecos o los post-ecos que se producirían sin el modelado de mosaico 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 mosaicos de frecuencia se llevan a cabo en el lado del 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 potencia se ha llevado a cabo sobre los valores espectrales residuales posteriores al filtrado o los valores espectrales (completos) antes del modelado de la envolvente.

El procesamiento del Modelado de Mosaico Temporal (TTS) sobre uno o más mosaicos 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 mosaicos de frecuencia.

En una implementación, se prefiere utilizar el filtrado complejo de Modelado de Ruido Temporal (TNS)/Modelado de Mosaico Temporal (TTS). De este modo se evitan los fallos de solapamiento (temporales) de una representación real muestreado críticamente, como la Transformada 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 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 mosaicos de frecuencia adyacentes de la frecuencia dentro del rango de reconstrucción.

El sistema de codificación de audio de la invención 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 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 inteligente de espacios (IGF) basado en parámetros por información lateral dedicada transmitida desde el codificador al decodificador.

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

- 10 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.
- 15 La Fig. 2a ilustra una implementación preferida del decodificador.  
La Fig. 2b ilustra una implementación preferida del codificador.
- 20 La Fig. 3a ilustra una representación esquemática de un espectro generado por el decodificador de dominio espectral de la Fig. 1b.  
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 potencias para las bandas de reconstrucción y la información de relleno de ruido para una banda de relleno de ruido.
- 25 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.  
La Fig. 4b ilustra una implementación de la funcionalidad de la Fig. 4a.
- 30 La Fig. 5a ilustra una funcionalidad de un codificador de la Transformada Coseno Discreta Modificada (MDCT).  
La Fig. 5b ilustra una funcionalidad del decodificador con una tecnología de MDCT.
- 35 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 mosaico temporal.
- 40 La Fig. 6b ilustra un decodificador con la tecnología de modelado de ruido temporal/modelado de mosaico temporal.  
La Fig. 6c ilustra una funcionalidad adicional de modelado de ruido temporal/modelado de mosaico 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 mosaico temporal (TTS).  
La Fig. 7b ilustra una implementación del decodificador que se adapta a la implementación del codificador de la Fig. 7a.
- 50 La Fig. 7c ilustra un espectrograma de una señal original y una señal extendida sin Modelado de Mosaico Temporal (TTS).  
La Fig. 7d ilustra una representación de frecuencia que ilustra la correspondencia entre las frecuencias de relleno inteligente de espacios y las potencias de modelado de mosaico temporal.
- 55 La Fig. 7e ilustra un espectrograma de una señal original y una señal extendida sin Modelado de Mosaico Temporal (TTS).  
La Fig. 8a ilustra un decodificador de dos canales con regeneración de frecuencia.  
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.
- 5 La Fig. 8d ilustra una implementación más detallada del decodificador de la Fig. 8a.
- La Fig. 8e ilustra una implementación de un codificador para el procesamiento de dos canales que será decodificado por el decodificador de la Fig. 8a.
- 10 La Fig. 9a ilustra un decodificador con la tecnología de regeneración de frecuencia utilizando valores de potencia 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.
- 15 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.
- 20 La Fig. 10b ilustra un diagrama de bloques para ilustrar una funcionalidad adicional de la calculadora de parámetros de la Fig. 10a.
- La Fig. 10c ilustra un diagrama de bloques para ilustrar una funcionalidad adicional de la calculadora de parámetros de la Fig. 10a.
- 25 La Fig. 10d ilustra un diagrama de bloques que ilustra una funcionalidad adicional de la calculadora de parámetros de la Fig. 10a.
- 30 La Fig. 11a ilustra un decodificador adicional que tiene una identificación específica del rango de origen para una operación de mosaico espectral en el decodificador.
- La Fig. 11b ilustra la funcionalidad adicional del regenerador de frecuencia de la Fig. 11a.
- 35 La Fig. 11c ilustra un codificador utilizado para cooperar con el decodificador de la Fig. 11a.
- La Fig. 11d ilustra un diagrama de bloques de una implementación de la calculadora de parámetros de la Fig. 11c.
- 40 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.
- 45 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 ser codificado con una primera resolución espectral y un segundo conjunto diferente de segundas porciones espectrales 105 para ser codificado 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 una calculadora 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, la calculadora 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.
- 50
- 55
- 60

Típicamente, 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.

5 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, en donde la representación decodificada tiene 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.

15 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 lleva a cabo una operación de relleno de mosaicos, es decir utiliza un mosaico 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, en donde la representación de tiempo tiene una tasa alta de muestreo determinada.

25 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 lleva a cabo 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 puede ser, ya sea un valor espectral completo, cuando no se aplica la operación de modelado de ruido temporal/modelado de mosaico temporal o puede ser un valor residual espectral, 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 llevar a cabo 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 llevar a cabo una compresión de datos sin pérdidas que es también una porción del codificador de dominio espectral 106 de la Fig. 1a.

40 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.

45 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.

50 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 tipo 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.

60 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  $n_{Tar}$  mosaicos de frecuencia que no se superponen. Para cada mosaico en la región de destino,  $n_{Src}$  se crea un mosaico de origen a partir de una frecuencia fija de inicio. Estos mosaicos de origen se

superponen por un factor entre 0 y 1, en donde 0 significa 0 % de solapamiento y 1 significa 100 % de solapamiento. Cada uno de estos mosaicos de origen está correlacionado con el mosaico de destino en diversos retardos para encontrar el mosaico de origen que se adapte mejor al mosaico de destino. El mejor número de mosaico 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 mosaico de origen debe ser multiplicado por -1 antes del proceso de relleno 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 potencia por bandas para almacenar la potencia 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 es conservada 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 es la 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 mosaico 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 mosaico permite adaptar mejor los espectros de destino y de origen.

Además, se propone una técnica de estabilización de elección de mosaicos que elimina los fallos 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 panoramizada 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 potencias 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 potencia de la región de destino, por lo general llevando a cabo una correlación cruzada de los valores espectrales y si se supera un determinado umbral, establece una etiqueta conjunta para esta banda de potencia. En el decodificador, las bandas de potencia de los canales izquierdo y derecho son tratadas individualmente si no se establece esta etiqueta conjunta de estéreo. En caso de que se establezca la etiqueta de estéreo conjunto, tanto las potencias como la interconexión se llevan a cabo 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 potencias pueden calcularse a partir de las potencias transmitidas en el dominio L/R.

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

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

Otra solución es calcular y transmitir las potencias 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 potencia adicional en el lado del decodificador.

Los mosaicos 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 potencia:

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

Transformación conjunta de estéreo -> LR:

Si no se codifica ningún parámetro de predicción adicional:

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

Si se codifica un parámetro de predicción adicional y si la dirección señalada es del medio al lateral:

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

Si la dirección señalada es del lateral al medio:

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

Este procesamiento asegura que a partir de los mosaicos utilizados para regenerar regiones de destino altamente correlacionadas y regiones de destino panoramizadas, los canales izquierdo y derecho resultantes siguen representando una fuente de sonido correlacionada y panoramizada 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 mosaicos del relleno inteligente de espacios (IGF), la representación de mosaicos de origen se selecciona para ajustar la representación de mosaicos 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 Coseno Discreta Modificada (MDCT). Estos coeficientes luego se utilizan 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 es enmascarado por el transitorio.

El relleno inteligente de espacios (IGF) se basa en una representación de MDCT. Preferentemente, para la codificación eficiente se tienen que utilizar bloques largos bloques 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 mosaicos. 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 mosaico 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 mosaico 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 mosaico 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 mosaico 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, la interconexión espectral en una señal de audio altera la correlación espectral en los límites de interconexión 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 mosaicos del relleno inteligente de espacios (IGF) en la señal residual es que, luego de la aplicación del filtro de modelado, los límites del mosaico 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 luego 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, luego 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 mosaico temporal (TTS). Posteriormente, en el bloque 210 se lleva a cabo 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 llevar a cabo 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 luego 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 es generalmente aplicada 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 potencia, 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 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 potencia  $E_1, E_2, E_3, E_4$ , 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 han sido cuantificadas 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 potencia de relleno de ruido tal como  $NF_2$  ilustrada en 308 en la Fig. 3b. La potencia 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 potencia del conjunto de valores espectrales cuantificados a cero. Estas líneas espectrales de relleno de ruido también pueden ser consideradas un tercer conjunto de terceras porciones espectrales que son regeneradas por la síntesis de relleno de ruido simple sin ninguna operación de relleno inteligente de espacios (IGF) basada en la regeneración de frecuencia utilizando mosaicos de frecuencia de otras frecuencias para la reconstrucción de mosaicos de frecuencia utilizando valores espectrales de un rango de origen y la información de potencia  $E_1, E_2, E_3, E_4$ .

Preferentemente, las bandas para las cuales se calcula la información de potencia coinciden con las bandas de factor de escala. En otras realizaciones se aplica un agrupamiento de valores de información de potencia de manera que, por ejemplo, para las bandas de factor de escala 4 y 5 solo se transmite un único valor de información de potencia, pero incluso en esta realización, los límites de las bandas de reconstrucción agrupadas coinciden con los límites de las bandas de factor de escala. Si se aplican diferentes separaciones de bandas, entonces se pueden aplicar 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 haber sido transformada en el rango espectral (401 en la Figura. 4a) se envía a una calculadora de factor de escala 400. La calculadora 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 luego 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 son ponderados por los factores de ajuste de escala y los valores espectrales ponderados luego 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 si existen índices de cuantificación que luego 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 "corrida" 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 ser codificada de manera muy eficiente, específicamente cuando existen "corridas" de valores cero en el espectro.

5 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 llevar a cabo 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 lleva a cabo en el bloque 418 posterior al bloque de ponderación 412. Incluso en otra implementación, la operación fijada en cero también se puede llevar a cabo 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.

15 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).

20 Los bloques fijados en cero 410, 418, 422, que se proporcionan alternativamente entre sí o en paralelo son controlados 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.

30 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 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, en donde cada cuadro dividido en ventanas normalmente tiene dos valores N como, por ejemplo, los valores 2048. Luego se lleva a cabo 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 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 lleva a cabo un intercambio en los bloques cortos, es decir cuando se llevan a cabo ocho bloques cortos en donde cada bloque corto tiene 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 eliminación se combina con una operación de 50 % de solapamiento del divisor de ventanas, el espectro es una versión muestreada críticamente de la señal de audio de dominio temporal 99.

50 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 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 potencia tal como  $E_3$  de la Fig. 3b para una banda de factor de escala 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 mosaicos de frecuencia utilizando un rango de origen y la banda de reconstrucción luego corresponde al rango de destino. En esta instancia se lleva a cabo un ajuste de potencia 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. Luego, en el bloque 512 se lleva a cabo una transformada/interpolación inversa de bloques para obtener 248 valores de dominio temporal para los 124 valores espectrales, por ejemplo, en la entrada del bloque 512. A continuación, en el bloque 514 se lleva a cabo una operación de síntesis de división de ventanas que está controlada nuevamente por una indicación de ventana larga/ventana corta transmitida como información lateral en la señal de audio codificada. Luego, en el bloque 516 se lleva a cabo 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.

5 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 aplicar dentro de este bloque o los valores espectrales de relleno de ruido ya se pueden ajustar utilizando la potencia de relleno de ruido antes de ser introducidos en el constructor/regulador de cuadros 510.

15 Preferentemente, una operación de IGF, es decir una operación de relleno de mosaicos de frecuencia que utiliza valores espectrales de otras porciones se puede aplicar en el espectro completo. Por lo tanto, una operación de relleno de mosaicos 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 mosaicos 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 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 mosaicos de frecuencia está limitada al rango de frecuencia por encima de la frecuencia de inicio de IGF, como se ilustra en la Fig. 3a.

25 Preferentemente, los mosaicos 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 mosaicos 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 mosaicos de origen (ST) debe corresponder al tamaño del mosaico de destino (TT) asociado. Esto se demuestra utilizando el siguiente ejemplo. 30 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 mosaicos de origen (ST) posibles que deben correlacionarse con TT[0], también tienen una longitud de 10 intervalos. Un segundo mosaico de destino TT[1] que es adyacente a TT[0] tiene una longitud de 15 intervalos (SCB tiene una longitud de 7 + 8). Entonces, el mosaico de origen (ST) para lo anterior tiene una longitud de 15 intervalos en lugar de 10 intervalos como para TT[0].

35 En caso de que no se pueda encontrar un mosaico de destino (TT) para un mosaico de origen (ST) con la longitud del mosaico 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 lleva a cabo 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 mosaico de destino (TT) se rellene completamente.

45 Posteriormente se hace referencia a la Fig. 5c que ilustra una 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 mosaicos 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 mosaicos de frecuencia 522 aplica una operación de copiado o de relleno de mosaicos de armónicos o cualquier otra operación de relleno de mosaicos 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.

55 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. Luego, el cuadro reconstruido es ajustado por el regulador 526 utilizando un factor de ganancia para la banda de reconstrucción calculada por la calculadora 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, la calculadora 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 potencia del cuadro ajustado emitido por el regulador 526 tiene la potencia  $E_4$  cuando se contempla una banda de factor de escala 7.

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".

5 Entonces, la potencia 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 sería reconstruido 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 sería introducido 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, sobre la base de 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 decodificador, esta información luego podría ser utilizada por un generador de mosaicos de frecuencia 522 de la Fig. 5c que ilustra una generación de una segunda porción en bruto 523 sobre la base de 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.

Tal como se ilustra, el codificador opera sin reducción de muestreo y el decodificador opera sin funciona 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.

40 Como se explicó, 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 es generado y transmitido 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.

50 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 tipo 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).

55 Realizaciones 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.

60 El sistema general consiste en

- codificación central de banda completa
- relleno inteligente de espacios (relleno de mosaicos 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 mosaicos
- TNS en el mosaico
- blanqueo espectral en el rango de relleno inteligente de espacios (IGF)

5 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 códecs 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 llevar a cabo 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.

15 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 llevar a cabo 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 realización de la Fig. 6a. Los valores residuales de predicción espectral 603 se introducen en un generador de mosaicos 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 luego 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 lleva a cabo 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 luego 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 llevar a cabo el filtrado de modelado de ruido temporal (TNS) o de modelado de mosaico temporal (TTS) en el lado del codificador. A continuación, el filtro de predicción espectral 622 se aplica antes de llevar a cabo 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. Luego se lleva a cabo el cálculo del filtro de TTS y el filtrado de TTS en el dominio complejo. A continuación, en el bloque 706 se calcula la información 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.

Entonces, 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 mosaicos

se lleva a cabo la operación inversa de modelado de mosaico temporal (TTS) 728 y, en el lado del decodificador se lleva a cabo 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 mosaicos por separado sino también sobre los límites centrales/de mosaicos o sobre los

5 límites de mosaicos/mosaicos, un procesamiento de límites de mosaicos que finalmente vuelve a introducir una correlación espectral entre los mosaicos. Esta correlación espectral sobre los límites de los mosaicos no se obtiene generando solamente mosaicos de frecuencia y llevando a cabo un ajuste de la envolvente espectral en estos datos en bruto de los mosaicos de frecuencia.

10 La Fig. 7c ilustra una comparación de una señal original (panel izquierdo) y una señal extendida sin Modelado de Mosaico Temporal (TTS). Se puede observar que hay fallos fuertes 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 fallos 750 de la Fig. 7c.

15 Las realizaciones 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

20 dedicada 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 transmitida cuando solo estaba disponible una tasa de bits muy baja. Los códecs de audio modernos ahora son

25 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) ("interconexión") y la aplicación de un procesamiento posterior basado en

30 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 así denominada frecuencia de cruce determinada se basa a menudo en la interconexión espectral. En general, la región de alta frecuencia (HF) consta de 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

35 eficiencia la interconexión dentro de una representación de bancos de filtros copiando un conjunto de coeficientes de sub-bandas 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.

40 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).

A partir de la codificación de audio perceptual, se sabe que la forma de la envolvente temporal de una señal de audio se puede restaurar 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 fallos de solapamiento, especialmente si la transformada real subyacente es una Transformada Coseno Discreta

45 Modificada (MDCT).

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

55 El modelado de mosaicos temporal consiste en

- estimación de coeficientes de filtro complejo y aplicación de un filtro de aplanamiento en el espectro de la señal original en el codificador
- transmisión de los coeficientes de filtro en la información lateral
- 60 • aplicación de un filtro de modelado en el espectro reconstruido relleno de mosaicos 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 ser vista 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 mosaico temporal (TTS). El cambio de nombre se justifica por el hecho de que los componentes de la señal rellenos de mosaicos se modelan temporalmente por TTS en comparación con el modelado de ruido de cuantificación por TNS en los códecs 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 ser replicada 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

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

Aquí, el esquema básico de decodificación funciona de la siguiente manera:

- 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
- llevar a cabo el relleno de mosaicos con su procesamiento posterior
- aplicar el filtrado de modelado de mosaico temporal (TTS) inverso con los coeficientes transmitidos del filtro de TTS
- calcular la CMDCT inversa

Alternativamente, 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 sistemáticamente en el codificador.

Para una codificación de transformada eficiente se deben utilizar preferentemente los así 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 mosaicos. 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 mosaicos sin el filtrado del modelado de mosaico 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 mosaicos 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.

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.

En una realización que combina TTS e IGF en un códec, es importante alinear determinados parámetros de TTS y

parámetros de IGF de manera que un mosaico de IGF se filtre completamente por un filtro de TTS (filtro de aplanamiento o modelado) o no. Por lo tanto, todas las frecuencias  $TTS_{start}[\dots]$  o  $TTS_{stop}[\dots]$  no estarán comprendidas dentro de un mosaico de IGF, sino más bien estarán alineadas con las frecuencias *Figf...* respectivas. La Fig. 7d muestra un ejemplo de áreas operativas de TTS e IGF para un conjunto de tres filtros de TTS.

5 La frecuencia de fin del modelado de mosaico temporal (TTS) se ajusta a la frecuencia de fin de la herramienta de relleno inteligente de espacios (IGF), que es mayor que *figfstart*. Si el modelado de mosaico temporal (TTS) utiliza más de un filtro tiene que asegurarse de 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 *figfstart*, lo que producirá fallos no deseados como, por ejemplo, sobremodelado.

10 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 potencias 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 mosaico 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 potencia por esta relación.

20 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 potencia 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 mosaico durante la codificación, la potencia final se ajusta siempre correctamente por el procesamiento de IGF.

25 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 mosaico 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 *figfstart* como es habitual en el modelado de ruido temporal (TNS) de la técnica anterior.

30 En los codificadores de la técnica anterior, la interconexión espectral de una señal de audio (por ejemplo, la Replicación de Ancho de Banda Espectral (SBR)) altera la correlación espectral en los límites de interconexión 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 mosaicos del relleno inteligente de espacios (IGF) en la señal residual es que, luego de la aplicación del filtro de modelado de mosaico temporal (TTS), los límites del mosaico se correlacionan perfectamente, lo que resulta en una reproducción temporal más fiel de la señal.

40 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).

45 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 llevar a cabo 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.

60 Finalmente, 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 mosaico temporal (TTS)" en 714 en la Fig. 7a.

Preferentemente, el filtro de predicción 704 comprende una calculadora 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.

- 5 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).

- 10 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).

15 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. Alternativamente, 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 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.

25 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.

40 La Fig. 8c ilustra un diagrama de flujo para las operaciones llevadas a cabo por el decodificador de audio de la Fig. 8a. En un paso 812, el decodificador de audio 802 lleva a cabo 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 llevar a cabo 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 mosaicos 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. Sobre la base de la ID de dos canales para el rango de destino, el generador de mosaicos 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 mosaicos 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 mosaicos 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 mosaicos de frecuencia 836 lleva a cabo esta operación para cada rango de destino y la salida del generador de mosaicos de frecuencia es tal que cada canal de la representación de canales identificada por la identificación de dos canales está presente. Luego un regulador de envolventes 838 lleva a cabo un ajuste de la envolvente. El ajuste de envolvente se lleva a cabo en el dominio de dos canales identificado por la identificación de dos canales. Para este fin se requieren parámetros de ajuste de 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 han sido transmitidos como parámetros de la envolvente L/R, entonces el transformador parámetro 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 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 es procesado 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 luego 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 un arreglo 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 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 realización, esto también se aplica para el rango de frecuencia por debajo de la frecuencia de inicio de IGF 309.

En una 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 mosaicos de frecuencia de armónicos o una operación de relleno de mosaicos de frecuencia de copiado utilizando dos o más operaciones de relleno de mosaicos de frecuencia similares al procesamiento para múltiples interconexiones conocidas 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 llevar a cabo un análisis con el fin de determinar las porciones espectrales que serán codificadas 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 es parametrizada utilizando la primera representación de dos canales o la segunda representación de dos canales, y esto se lleva a cabo 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 es codificado 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 banda, ya sea por encima de la frecuencia de inicio de IGF o para todo 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. Sobre la base de la 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 rangos, el rango de origen y el rango de reconstrucción, son procesados por el transformador por bandas, el analizador espectral 860 puede analizar esta representación. Alternativamente, 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 haber sido procesada por 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 potencia ya analizados en el contexto de la Fig. 3a y en la Fig. 3b.

Los codificadores de audio del estado moderno 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 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 gruesa introduciendo un gran número de valores espectrales reducidos a cero que generan fallos audibles en el lado del decodificador. Para mejorar la calidad de percepción, los decodificadores del estado de la técnica llenan estas partes espectrales reducidas a cero con ruido aleatorio. El método de relleno inteligente de espacios (IGF) recolecta mosaicos de la señal que no es cero restantes para llenar esos espacios en el espectro. Es crucial para la calidad perceptual de la señal de audio decodificada que la envolvente espectral y la distribución de potencia de los coeficientes espectrales se conserven. El método de ajuste de potencia 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 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 fina temporal de una señal transitoria y no se puede conservar la estructura armónica de una señal tonal.

El área 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 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 potencias de las áreas interconectadas, el ajuste de la envolvente espectral [1]. Esta técnica utiliza los valores de potencia 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 valores de potencia apropiados o para aplicar una cuantificación gruesa a los valores de potencia.

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 potencia 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 es generada 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 mosaico temporal (TTS) en los coeficientes de la MDCT [15, sección 7.8]

c) Calcular la potencia 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 potencia

$f_{IGFstart}$  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) Decuantificar 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 mosaicos del relleno inteligente de espacios (IGF)

d) Decuantificar los valores de potencia transmitidos

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

f) Aplicar TNS/TTS si está indicado

Dejar que  $\hat{x} \in \mathbb{R}^N$  sea 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  $\hat{x}$ .

5 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 un arreglo  $swb\_offset$  (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:

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

15 Si la herramienta de IGF es utilizada 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  $igfStartSfb$  y  $igfStopSfb$ . Ambos son enviados 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 son agrupados 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:

25

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

Donde  $k = igfStartSfb, 1 + igfStartSfb, 2 + igfStartSfb, \dots, igfEndSfb$ .

30 Para cuantificar, se calcula

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

35 Todos los valores  $E_{k,l}$  son transmitidos al decodificador.

Se supone que el codificador decide agrupar los conjuntos de factor de escala  $num\_window\_group$ . 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,  $0 \leq l < num\_window\_group$ .

40 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 potencia de IGF utiliza la información de agrupamiento para agrupar los valores  $E_{k,l}$ :

45

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

Para cuantificar, se calcula

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

50

Todos los valores  $\hat{E}_{k,l}$  son transmitidos 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 potencia más estable en el rango de IGF, es decir para reducir las fluctuaciones de amplitud temporal, se puede utilizar un método alternativo para calcular los valores  $\hat{E}_k$ :

Dejar que  $\hat{x}_r \in \mathbb{R}^N$  sea 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 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  $\hat{x}_i$ .

Ahora, la potencia 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 potencias de valor real y complejo de la banda de reconstrucción, es decir el mosaico que se debe utilizar en el lado del decodificador en la reconstrucción del rango de IGF  $scb_k$ , se calcula con:

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

donde  $tr_k$  es un conjunto de índices - el rango de mosaico 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  $E_r$ .

Se calcula

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

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

Con:

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

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 potencia completa de los valores espectrales.  $\hat{E}_k$  se calcula como se indicó anteriormente.

- 5 Como se indicó anteriormente, para los bloques cortos 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  $w$ , en donde  $l$  indica el índice del grupo de ventana,  $0 \leq l < num\_window\_group$ .

Una vez más podría aplicarse la versión alternativa descrita anteriormente para calcular una versión más estable de

- 10  $E_{k,l}$ . 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), se calcula

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

- 15 Se calcula de manera análoga

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

- 20 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:

25

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

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

- 30 El procedimiento que no solo utiliza la potencia 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 potencia del rango de origen proporciona una reconstrucción de potencia mejorada.

- 35 Específicamente, la calculadora de parámetros 1006 está configurado para calcular la información de potencia para la banda de reconstrucción utilizando información sobre la potencia de la banda de reconstrucción y utilizando, además, la información sobre una potencia de un rango de origen para ser utilizado para la reconstrucción de la banda de reconstrucción.

- 40 Asimismo, la calculadora de parámetros 1006 está configurado para calcular una información de potencia ( $E_{ok}$ ) en la banda de reconstrucción de un espectro complejo de la señal original, para calcular una información de potencia adicional ( $E_{rk}$ ) en un rango de origen de una parte de valor real del espectro complejo de la señal original para ser utilizada para reconstruir la banda de reconstrucción, y en el que la calculadora de parámetros está configurada para calcular la información de potencia para la banda de reconstrucción utilizando la información de potencia ( $E_{ok}$ ) y la información de potencia adicional ( $E_{rk}$ ).

- 45 Además, la calculadora de parámetros 1006 está configurada para determinar una primera información de potencia ( $E_{ok}$ ) en una banda de factor de escala que debe ser reconstruida de un espectro complejo de la señal original, para determinar una segunda información de potencia ( $E_{rk}$ ) en un rango de origen del espectro complejo de la señal original para ser utilizada para reconstruir la banda de factor de escala que debe ser reconstruida, para determinar

una tercera información de potencia ( $E_{rk}$ ) en un rango de origen de una parte de valor real del espectro complejo de la señal original para ser utilizada para reconstruir la banda de factor de escala que debe ser reconstruida, para determinar una información de ponderación sobre la base de una relación entre al menos dos de la primera información de potencia, la segunda información de potencia, y la tercera información de potencia, y para ponderar una de la primera información de potencia y la tercera información de potencia utilizando la información de ponderación para obtener una información de potencia ponderada y para utilizar la información de potencia ponderada como la información de potencia 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)

$$\begin{aligned} f_k &= E_{ok}/E_{tk}; \\ E_k &= \text{sqrt}(f_k * E_{rk}); \end{aligned}$$

B)

$$\begin{aligned} f_k &= E_{tk}/E_{ok}; \\ E_k &= \text{sqrt}((1/f_k) * E_{rk}); \end{aligned}$$

C)

$$\begin{aligned} f_k &= E_{rk}/E_{tk}; \\ E_k &= \text{sqrt}(f_k * E_{ok}) \end{aligned}$$

D)

$$\begin{aligned} f_k &= E_{tk}/E_{rk}; \\ E_k &= \text{sqrt}((1/f_k) * E_{ok}) \end{aligned}$$

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 potencia de mosaico 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 potencia transmitida al decodificador será generalmente menor que la información de potencia  $E_{ok}$  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 ON, los valores transmitidos  $\hat{E}_k$  se obtienen a partir de la corriente de bits y serán decuantificados con

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

para todos  $k = \text{igfStartSfb}, 1 + \text{igfStartSfb}, 2 + \text{igfStartSfb}, \dots, \text{igfEndSfb}$ .

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

$$sE_k := \sum_{i \in \text{scb}_k} x_i^2$$

donde  $k$  está en el rango definido anteriormente.

Indicamos que  $\overline{\text{scb}_k} = \{i | i \in \text{scb}_k \wedge x_i = 0\}$ . Este conjunto contiene todos los índices de la banda de factor de escala  $\text{scb}_k$  que han sido cuantificados a cero por el codificador.

El método de sub-banda de relleno inteligente de espacios (IGF) (no descrito en el presente documento) se utiliza para llenar espacios espectrales que resultan de una cuantificación gruesa 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 potencia del mosaico se calcula por:

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

5 donde  $k$  está en el rango definido anteriormente.

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

$$mE_k := |\overline{scb_k}| E_k^2 - sE_k$$

10

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

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

15 Con:

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

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

$$x_i := g' x_i$$

20

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  $\hat{x}$ .

25 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 potencia agrupados y decuantificados, obtenidos de la corriente de bits. Se calcula

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

30 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.

35

Se calcula

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

y

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

5 Con

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

10 Se aplica

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

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

15 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:

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

20 donde  $k = igfStartSfb, 2 + igfStartSfb, 4 + igfStartSfb, \dots, igfEndSfb$ .

Nuevamente, luego de la decuantificación, todos los valores  $E_{k \gg 1}$  son transmitidos al decodificador.

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 potencias 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 potencias 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 potencia individual para la banda de reconstrucción, en donde la banda de reconstrucción comprende una primera porción espectral y la segunda porción espectral. El regenerador de frecuencia 906 comprende una calculadora 912 para determinar una información de potencia de conservación que comprende una potencia acumulada de la primera porción espectral que tiene frecuencias en la banda de la reconstrucción. Asimismo, el regenerador de frecuencia 906 comprende una calculadora 918 para determinar una información de potencia de mosaico 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 ser generadas por la regeneración de frecuencia utilizando una primera porción espectral diferente de la primera porción espectral en la banda de reconstrucción.

45 El regenerador de frecuencia 906 comprende además una calculadora 914 para una potencia faltante en la banda

de reconstrucción, y la calculadora 914 funciona utilizando la potencia individual para la banda de reconstrucción y la potencia 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 sobre la base de la información de potencia faltante y la información de potencia de mosaicos generada por el bloque 918.

Con referencia a la Fig. 9c, allí 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 922 y 923. Luego 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 está influenciada por el ajuste de la envolvente llevado a cabo 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 es emitida 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 potencia de conservación restante calculada por el bloque 912, por ejemplo, es de cinco unidades de potencia y esta potencia es la potencia de las cuatro líneas espectrales indicadas a modo de ejemplo en la primera porción espectral 921.

Asimismo, el valor de la potencia  $E_3$  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 potencia no solo comprende la potencia de las porciones espectrales 922, 923, sino también la potencia total de la banda de reconstrucción 920 calculada en el lado del codificador, es decir, antes de llevar a cabo el análisis espectral, utilizando, por ejemplo, el enmascaramiento de la tonalidad. Por lo tanto, las diez unidades de potencia abarcan las primeras y las segundas porciones espectrales en la banda de reconstrucción. Entonces, se supone que la potencia 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 potencia. Por lo tanto, se calcula una potencia faltante de cinco unidades.

Se calcula un factor de ganancia de 0,79 sobre la base de la potencia faltante dividida por la potencia de mosaico  $tEk$ . 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 están 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 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 potencia, un valor de potencia para una banda de reconstrucción proporciona la potencia de dos o un número entero superior de bandas de factor de escala. Por lo tanto, cuando se supone que la acumulación de potencia es llevada a cabo 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 potencia 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 decuantificados 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 mosaicos de frecuencia en la banda de la reconstrucción se proporcionan al regulador/calculadora de la envolvente 942 y este bloque recibe además la información de potencia para la banda de reconstrucción proporcionada como información lateral paramétrica de la señal de audio codificada ilustrada en 943 en la Fig. 9d. Luego, el regulador/calculadora 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 realizaciones preferidas de un codificador de audio que codifica una señal de audio para proporcionar o generar una señal de audio codificada. El codificador comprende un convertidor de tiempo/espectro 1002 que alimenta un analizador espectral 1004, y el analizador espectral 1004 está conectado a una calculadora 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, la calculadora de parámetros 1006 proporciona información de la potencia 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, 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 potencia 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 potencia emitida por la calculadora 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, la calculadora paramétrica comprende un normalizador 1012 para normalizar las potencias 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 potencia en la banda y un número de valores espectrales en la banda y el normalizador 1012 luego emite una potencia normalizada por banda de reconstrucción/banda de factor de escala.

Además, la calculadora paramétrica 1006a de la Fig. 10a comprende una calculadora de valor de la potencia 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, la calculadora del valor de potencia 1014 emite un único valor de potencia 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 realización adicional para la implementación del agrupamiento espectral. Para este fin, el bloque 1016 está configurado para calcular los valores de la potencia para dos bandas adyacentes. A continuación, en el bloque 1018 se comparan los valores de la potencia para las bandas adyacentes y, cuando los valores de la potencia 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 ser controlado por un control de tasa de bits del codificador 1024. Por lo tanto, cuando la tasa de bits se debe reducir, 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 potencia.

En caso de que el codificador de audio lleve a cabo el agrupamiento de dos o más ventanas cortas, este agrupamiento se aplica también para la información de la potencia. 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 luego aplica el mismo conjunto de factores de ajuste de escala para ambas ventanas agrupadas.

En cuanto al cálculo de la información de la potencia, 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 potencia 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.

Luego se aplica nuevamente la normalización correspondiente de manera que, aunque se haya llevado a cabo cualquier agrupamiento en la frecuencia o agrupamiento temporal, la normalización permite fácilmente que, para el cálculo de la información del valor de potencia en el lado del decodificador, solo debe conocerse el valor de la información de potencia 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 así denominada frecuencia de cruce determinada se basa a menudo en la interconexión espectral. En general, la región de alta frecuencia (HF) consta de 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. 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 interconexión 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, la interconexión puede volverse inestable, especialmente para señales no tonales (por ejemplo, de voz sorda) y, por lo tanto, la interconexión adaptativa 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) sobre la base de una transformada de frecuencia temporal como, por ejemplo, la Transformada 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 interconexión adaptativa del estado de la técnica.

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

El sistema de IGF propuesto lleva a cabo una selección inteligente de mosaicos 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 mosaicos 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 mosaicos. 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 pesado-aleatorio de la selección adaptativa de mosaicos.

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 potencia 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  $f_{xover}$  (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 un fallo de coloración del timbre.

Un paso importante del nuevo enfoque consiste en definir un conjunto de mosaicos entre los que puede tener lugar la elección basada en la similitud posterior. En primer lugar, los límites de los mosaicos, 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 mosaicos, cada uno de los cuales tiene un tamaño individual predefinido. Entonces, para cada mosaico de destino  $tar[idx\_tar]$  se genera un conjunto de mosaicos 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 mosaicos de origen  $nSrc$  está determinado por el ancho de banda de la región de origen,

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

donde  $f_{IGFmin}$  es la frecuencia más baja disponible para la selección de mosaicos de manera que un número entero  $nSrc$  de mosaicos de origen se adapta en  $bw_{src}$ . El número mínimo de mosaicos de origen es 0.

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

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

Para un mosaico de destino, la correlación cruzada se calcula con varios mosaicos de origen en retardos de hasta  $xcorr\_maxLag$  intervalos. Para un mosaico de destino determinado  $idx\_src$  y un mosaico de origen  $idx\_src$ ,  $xcorr\_val[idx\_tar][idx\_src]$  proporciona el valor máximo de la correlación cruzada absoluta entre los mosaicos, mientras que  $xcorr\_lag[idx\_tar][idx\_src]$  proporciona el retardo en el que se produce este máximo y  $xcorr\_sign[idx\_tar][idx\_src]$  proporciona el signo de la correlación cruzada en  $xcorr\_lag[idx\_tar][idx\_src]$ .

El parámetro  $xcorr\_lag$  se utiliza para controlar la proximidad de la coincidencia entre el mosaico de origen y el mosaico de destino. Este parámetro da lugar a una reducción de fallos y ayuda a conservar mejor el timbre y el color de la señal.

En algunos casos puede ocurrir que el tamaño de un mosaico de destino específico es mayor que el tamaño de los mosaicos de origen disponibles. En este caso, el mosaico de origen disponible se repite tan a menudo como sea necesario para llenar completamente el mosaico de destino específico. Todavía es posible llevar a cabo la correlación cruzada entre el mosaico grande de destino y el mosaico de origen más pequeño con el fin de obtener la mejor posición del mosaico de origen en el mosaico de destino en términos del retardo de la correlación cruzada  $xcorr\_lag$  y el signo  $xcorr\_sign$ .

La correlación cruzada de los mosaicos 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 fina espectral que es de interés principal para la evaluación de la similitud de mosaicos. El blanqueo también ayuda en un modelado fácil de la envolvente fácil de la región espectral de destino (STR) en el decodificador para las regiones procesadas por IGF. Por lo tanto, opcionalmente, el mosaico y la señal de origen se blanquean antes de calcular la correlación cruzada.

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

5 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 potencias del espectro de la MDCT, las estimaciones del espectro de potencia complejo basado en la MDCT o las estimaciones del espectro de potencia. 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 potencia complejo basado en la MDCT o el espectro de potencia como señal de base tienen la ventaja de no tener fluctuación temporal en los componentes tonales.

10 Si la señal de base está en un dominio de potencia, 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:

- 15 • 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 luego calculando 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
- filtrando la señal de base con un filtro de paso bajo

20 Preferentemente, se elige el último enfoque. Para aplicaciones que requieren baja complejidad computacional se puede llevar a cabo 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  $2^n$ , en donde  $n$  es el logaritmo de número entero de la envolvente. En este dominio, la operación de raíz

25 cuadrada se convierte simplemente en una operación de desplazamiento y, además, la división por la envolvente se puede llevar a cabo por otra operación de desplazamiento.

Después de calcular la correlación de cada mosaico de origen con cada mosaico de destino, para todos los mosaicos de destino  $nTar$  se selecciona el mosaico 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 mosaico 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 Coseno Discreta Modificada (MDCT).

30

35 La Figura 12c muestra un ejemplo de una correlación entre un mosaico de origen y un mosaico de destino. En este ejemplo, el retardo de la correlación es 5, por lo que el mosaico 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 mosaico tiene que invertirse ya que el valor máximo de correlación es negativo y una modulación adicional como se describió anteriormente representa el retardo impar.

40

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

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

El recorte y la estabilización de mosaicos 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 fallos si, para una región de destino determinada, siempre se seleccionan los mosaicos 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 fallo de tipo de ruido musical molesto.

50

55

Con el fin de eliminar este tipo de fallos, el conjunto de mosaicos de origen se recortará de manera que los elementos restantes del conjunto de origen sean máximamente disimiles. Esto se logra a través de un conjunto de mosaicos de origen

60

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

de la siguiente manera. Para cualquier mosaico de origen  $s_i$ , lo correlacionamos con todos los otros mosaicos de

origen, encontrando la mejor correlación entre  $s_i$  y  $s_j$  y almacenándola en una matriz  $S_x$ . Aquí,  $S_x[i][j]$  contiene el valor máximo absoluto de correlación cruzada entre  $s_i$  y  $s_j$ . La adición de la matriz  $S_x$  a lo largo de las columnas, proporciona la suma de las correlaciones cruzadas de un mosaico de origen  $s_i$  con todos los otros mosaicos de origen  $T$ .

5

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

Aquí,  $T$  representa una buena medida de similitud entre un mosaico de origen y otros mosaicos de origen. Si, para cualquier mosaico de origen  $i$ ,

10

$$T > \text{umbral}$$

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

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

20

Dejar que los mosaicos  $s_3, s_4$  t  $s_5$  estén activos a partir de los mosaicos  $\{s_1, s_2, \dots, s_5\}$  en el cuadro  $k$ , luego en el cuadro  $k+1$ , incluso si los mosaicos  $s_1, s_3$  y  $s_2$  compiten por ser recortados en donde  $s_3$  está correlacionado máximamente con los otros,  $s_3$  se retiene ya que fue un mosaico de origen útil en el cuadro anterior y, por lo tanto, su retención en el conjunto de mosaicos de origen es beneficioso para reforzar la continuidad temporal en la selección de mosaicos. 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

30

Un método adicional para la estabilización de mosaicos consiste en retener el orden de los mosaicos del cuadro  $k-1$  anterior si ninguno de los mosaicos de origen en el cuadro  $k$  actual se correlaciona bien con los mosaicos 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  $i, j$

35

Por ejemplo, si

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

entonces se utiliza un umbral provisorio, luego

40

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

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

45

Las dos técnicas anteriores reducen en gran medida los fallos que se producen a partir del cambio rápido de números de mosaicos fijos a través de los cuadros. Otra ventaja adicional de este recorte y estabilización de mosaicos 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 mosaicos resulta una manera elegante de reducir el ruido musical potencia en forma de fallos o ruido excesivo en las regiones espectrales de los mosaicos.

50

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, en donde la representación decodificada tiene una primera resolución espectral.

55

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 mosaico de frecuencia de destino o banda de reconstrucción de destino, una información de rango de origen. El regenerador de frecuencia de 1106 luego 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.

60

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.

- 5 Como se indica en la Fig. 11b, el regenerador de frecuencia 1106 puede comprender las funcionalidades del bloque 1120 que ilustra un modulador de mosaicos - 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  
 10 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 mosaicos 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  
 15 manera que la salida del bloque 1120 se proporciona al bloque de ajuste de la envolvente espectral 1128 sin ningún blanqueo.

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

```

20 bit = readBit(1);
   if(bit == 1) {
       for(tile_index = 0..nT)
           /*same levels as last frame*/
           whitening_level[tile_index] = whitening_level_prev_frame[tile_index];
   } else {
       /*first tile:*/
       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; /*no-whitening*/
           }
       }
   }
   /*remaining tiles:*/
   bit = readBit(1);
   if(bit == 1) {
       /*flattening levels for remaining tiles same as first.*/
       /*No further bits have to be read*/
       for(tile_index = 1..nT)
           whitening_level[tile_index] = whitening_level[0];
   } else {
       /*read bits for remaining tiles as for first tile*/
       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; /*no-whitening*/
    }
}
}
}
}
}

```

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 ser controlado por una ID de rango de origen 1121 cuando se aplica solo un esquema de selección de mosaicos espectrales en bruto. Sin embargo, cuando se aplica un esquema de selección de mosaicos 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, una calculadora de parámetros 1134 y un codificador central 1136. El codificador central 1136 emite rangos de origen codificados y la calculadora 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 llevar a cabo una regeneración de frecuencia.

25 La calculadora de parámetros 1134 está configurada para calcular similitudes entre primeras porciones espectrales y segundas porciones espectrales y para determinar, sobre la base de 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 la calculadora de parámetros está configurada para proporcionar esta información de adaptación que identifica el par de adaptación en una señal de audio codificada. Preferentemente, la calculadora de parámetros 1134 de la presente invención está configurada 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.

40 Como se explicó anteriormente, una granularidad fina 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 la calculadora de correlación cruzada 1140 de la Fig. 11d y la selección de pares de adaptación es llevada a cabo finalmente por el selector de mosaicos 1144.

Además, se prefiere llevar a cabo 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 la calculadora de correlación cruzada 1140 proporciona un resultado negativo, entonces este resultado negativo también se señala a un decodificador. Por lo tanto, en una realización preferida, el selector de mosaicos 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, la calculadora de parámetros 1134 está configurada para llevar a cabo un recorte de mosaicos de origen 1146 reduciendo el número de rangos de origen potenciales por lo que una interconexión de origen se retira de un conjunto de mosaicos de origen potenciales sobre la base de un umbral de similitud. Por lo tanto, cuando dos mosaicos de origen son más similares o iguales a un umbral de similitud, entonces uno de estos dos mosaicos de origen se retira del conjunto de fuentes potenciales y el mosaico de origen eliminado ya no se utiliza para el procesamiento posterior y, específicamente, no puede ser seleccionado por el selector de mosaicos 1144 o no se utiliza para el cálculo de la correlación cruzada entre diferentes rangos de origen y rangos de destino como se llevó a cabo 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 completa 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 mosaico 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 específico de información de potencia y la aplicación, y las Figs. 11a-12c se refieren a un modo específico de selección de mosaicos.

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-13b, y esto mismo se aplica al procesamiento de TNS/TTS, al cálculo de información de potencia 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 potencia 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, en donde un bloque o dispositivo corresponde a un paso del método o a una característica de un paso del método. En forma análoga, 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 llevar a cabo por (o con) un aparato de hardware tal como, por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, alguno o más de la mayoría de los pasos importantes del método se pueden llevar a cabo por dicho aparato.

En función de determinados requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede llevar a cabo 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 ellos, que cooperan (o son capaces de cooperar) con un sistema informático programable de forma tal que se lleva a cabo el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

Algunas realizaciones 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 ordenador programable, de tal manera que uno de los métodos descritos aquí se realice.

En general, las realizaciones 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 llevar a cabo uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código del programa se puede almacenar, por ejemplo, en un portador legible por ordenador.

Otras realizaciones comprenden los programas informáticos para llevar a cabo uno de los métodos descritos en la presente, almacenados en un portador legible por ordenador.

En otras palabras, una realización del método de la invención es, por lo tanto, un programa informático que tiene un

código de programa para llevar a cabo uno de los métodos descritos en la presente, cuando el programa informático se ejecuta en un ordenador.

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

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

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

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

20 Otra realización de acuerdo con la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, por vía electrónica u óptica) un programa informático para llevar a cabo uno de los métodos descritos en la presente a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa informático al receptor.

25 En algunas realizaciones, un dispositivo lógico programable (por ejemplo, un arreglo de puerta programable de campo) se puede utilizar para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en la presente invención. En algunas realizaciones, un arreglo de puerta programable de campo puede cooperar con un microprocesador para llevar a cabo uno de los métodos descritos en la presente. En general, los métodos serán llevados a cabo, preferentemente, por cualquier aparato de hardware.

30 Las realizaciones anteriormente descritas son simplemente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de los arreglos y los detalles descritos en la presente 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 inminentes de la patente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones del presente documento.

#### Lista de Referencias

40 [1] Dietz, L. Liljeryd, K. Kjörling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112th AES Convention, Munich, Mayo de 2002.

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

45 [3] D. Sinha, A. Ferreira<sup>1</sup> and E. Harinarayanan, "A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)", Audio Engineering Society Convention, Paris, Francia 2006.

50 [4] R. Annadana, E. Harinarayanan, A. Ferreira and D. Sinha, "New Results in Low Bit Rate Speech Coding and Bandwidth Extension", Audio Engineering Society Convention, San Francisco, EE.UU. 2006.

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

55 [6] J. Herre, D. Schulz, Extending the MPEG-4 AAC Codec by Perceptual Noise Substitution, 104th AES Convention, Amsterdam, 1998, Preprint 4720.

60 [7] M. Neuendorf, M. Multrus, N. Rettelbach, et al., MPEG Unified Speech and Audio Coding-The ISO/MPEG Standard for High-Efficiency Audio Coding of all Content Types, 132nd AES Convention, Budapest, Hungary, 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.

- [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.
- 5 [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, no., pp.201, 204 vol.3, 2000
- [11] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio", Ginebra, 1998.
- 10 [12] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Oikawa: "MPEG-2 Advanced Audio Coding", 101st AES Convention, Los Angeles 1996
- [13] J. Herre, "Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction", 17th AES International Conference on High Quality Audio Coding, Agosto de 1999
- 15 [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.
- 20 [16] International Standard ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 3: Audio, Ginebra, 2005.
- [17] P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication", in Proceedings of 1st IEEE Benelux Workshop on MPCA, Leuven, Noviembre de 2002
- 25 [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 (EE.UU.), Abril de 2010
- 30

**REIVINDICACIONES**

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

5 un decodificador de audio (1102) para decodificar una representación codificada de un primer conjunto de primeras porciones espectrales para obtener un primer conjunto decodificado de primeras porciones espectrales (1101);  
 un decodificador paramétrico (1104) para decodificar una representación paramétrica codificada de un segundo conjunto de segundas porciones espectrales para obtener una representación decodificada de la representación paramétrica (1103), en el que la representación decodificada de la representación paramétrica (1103) incluye, para cada mosaico de frecuencia de destino, una identificación de la región de origen (1121) como una información de adaptación; y  
 10 un regenerador de frecuencia (1106) para regenerar un mosaico de frecuencia de destino utilizando una región de origen a partir del primer conjunto de primeras porciones espectrales (1101) identificadas por la información de adaptación,  
 en el que el regenerador de frecuencia (1106) comprende un filtro de blanqueo controlable (1122, 1123), en el que la representación decodificada de la representación paramétrica (1103) comprende información de blanqueo (1123),  
 en el que el regenerador de frecuencia (1106) es configurado para aplicar el filtro de blanqueo (1122) a una región de origen seleccionada de acuerdo con la información de adaptación, antes de efectuar un ajuste de envolvente espectral (1128), cuando la información de blanqueo, para la región de origen, indica que la región de origen va a ser blanqueada, y  
 20 en el que la aplicación del filtro de blanqueo comprende calcular una estimación de envolvente espectral de la región de origen y dividir un espectro de la región de origen por una envolvente espectral indicada por la estimación de la envolvente espectral.  
 25

2. Aparato de la reivindicación 1, en el que el decodificador de audio (1102) es un decodificador de audio de dominio espectral, y  
 en el que el aparato comprende además un convertidor de espectro-tiempo (1108) para convertir una representación espectral de las primeras porciones espectrales y segundas porciones espectrales reconstruidas en una representación temporal.  
 30

3. Aparato de la reivindicación 1, en el que la información de blanqueo comprende, para un mosaico o un grupo de mosaicos, una información de nivel de blanqueo que indica un nivel de blanqueo para aplicar a un mosaico de frecuencia de origen cuando se regenera el mosaico de frecuencia de destino, y  
 35 en el que el regenerador de frecuencia (1106) está configurado para aplicar un filtro de blanqueo seleccionado de un grupo de diferentes filtros de blanqueo en respuesta a la información de blanqueo.

4. Aparato de acuerdo con una de las reivindicaciones anteriores,  
 en el que el regenerador de frecuencia (1106) comprende un modificador de la región de origen,  
 en el que la representación decodificada de la representación paramétrica (1103) comprende, además de la identificación de la región de origen (1121), una información del signo, y  
 en el que el modificador de la región de origen se configura para aplicar una operación para obtener un desplazamiento de fase de los valores espectrales de la región de origen de acuerdo con la información del signo.  
 45

5. Aparato de acuerdo con una de las reivindicaciones anteriores,  
 en el que el regenerador de frecuencia (1106) comprende un modulador de mosaicos (1120),  
 en el que la representación decodificada de la representación paramétrica (1103) comprende un retardo de correlación, además de la identificación de la región de origen (1121), y  
 en el que el modulador de mosaicos está configurado para aplicar una modulación de mosaicos de acuerdo con el retardo de correlación asociado con la identificación de la región de origen (1121).  
 50

6. Aparato de acuerdo con cualquiera de las reivindicaciones anteriores,  
 en el que el regenerador de frecuencia (1106) comprende un modulador de mosaicos (1120),  
 en el que la representación decodificada de la representación paramétrica (1103) comprende un retardo de correlación, además de la identificación de la región de origen (1121), y  
 en el que el modulador de mosaicos (1120) está configurado para aplicar una modulación de mosaicos utilizando una secuencia temporal alternativa de  $-1/1$  cuando el retardo de la correlación es un número impar.  
 60

7. Aparato para codificar una señal de audio, que comprende:

un convertidor de tiempo-espectro (1130) para convertir una señal de audio en una representación espectral;  
 un analizador espectral para analizar la representación espectral para determinar un primer conjunto de primeras

- porciones espectrales que deben ser codificadas con una primera resolución espectral y un segundo conjunto de segundas porciones espectrales que deben ser codificadas con una segunda resolución espectral, en el que la segunda resolución espectral es inferior a una primera resolución espectral;
- 5 una calculadora de parámetros (1134) para calcular similitudes entre regiones de origen predefinidas y regiones de destino, comprendiendo una región de origen porciones espectrales y comprendiendo una región de destino segundas porciones espectrales,
- en el que la calculadora de parámetros (1134) está configurada para comparar resultados de adaptación para diferentes pares de una primera porción espectral y una segunda porción espectral para determinar un par de adaptación seleccionado y para proporcionar información de adaptación que identifique el par de adaptación; y
- 10 un codificador central (1136) para codificar el primer conjunto de primeras porciones espectrales, en el que el primer conjunto de primeras porciones espectrales comprende las regiones de origen predefinidas y porciones espectrales diferentes de las regiones de origen predefinidas,
- en el que la calculadora de parámetros está configurada para copiar un mosaico de origen múltiples veces a un mosaico de destino, hasta que el mosaico de destino se llena con múltiples copias del mosaico de origen,
- 15 cuando el mosaico de destino es mayor que el mosaico de origen, o en el que la calculadora de parámetros (1134) está configurada para blanquear espectralmente (1142) la primera o la segunda porción espectral de los pares, antes de efectuar un procesamiento de correlación para obtener la identificación de adaptación, en el que el blanqueo espectral (1142) comprende calcular una estimación de envolvente espectral de la primera o la segunda porción espectral y dividir un espectro de la primera o la segunda porción espectral, respectivamente,
- 20 por una envolvente espectral indicada por la estimación de envolvente espectral.
8. Aparato de acuerdo con la reivindicación 7, en el que la calculadora de parámetros (1134) está configurada 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.
- 25 9. Aparato de acuerdo con la reivindicación 7, en el que la calculadora de parámetros (1134) está configurada de manera que las regiones de destino predefinidas no se solapan, o las regiones de origen predefinidas se solapan, o en el que las regiones de origen predefinidas son un subconjunto del primer conjunto de las primeras porciones espectrales por debajo de una frecuencia de inicio de relleno de espacios (309), o en el que una región de destino predefinida abarca una región espectral más baja que coincide con la frecuencia de inicio de relleno de espacios (309).
- 30 10. Aparato de acuerdo con una de las reivindicaciones 7, 8 o 9, en el que la calculadora de parámetros (1134) es configurada para comparar pares de una región de destino y una región de origen y un par de la región de destino y la misma región de origen, en el que la misma región de origen es desplazada por un retardo de correlación para proporcionar información sobre el retardo de la correlación de un par seleccionado como información de adaptación adicional.
- 35 11. Aparato de acuerdo con cualquiera de las reivindicaciones 7 a 10, en el que la calculadora de parámetros (1134) está configurada para llevar a cabo un procesamiento de correlación para obtener un resultado de adaptación para un par de la primera porción espectral y la segunda porción espectral, teniendo el resultado de adaptación un signo negativo, y en el que la calculadora de parámetros (1134) está configurada para proporcionar una información sobre el signo negativo como información de adaptación adicional.
- 40 45 12. Aparato de acuerdo con cualquiera de las reivindicaciones 7 a 11, en el que la calculadora de parámetros (1134) está configurada para determinar un número entero de mosaicos de destino y para determinar una pluralidad de mosaicos de origen de igual tamaño para cada mosaico de destino.
- 50 13. Aparato de acuerdo con cualquiera de las reivindicaciones 7 a 12, en el que la calculadora de parámetros (1134) está configurada para calcular la envolvente espectral de la primera porción espectral o la segunda porción espectral utilizando al menos uno de los siguientes procedimientos:
- 55 transformar el espectro con una transformada de coseno discreta (DCT), reteniendo los coeficientes de menor frecuencia de la DCT fijando los coeficientes más altos de la DCT a cero y calculando una DCT inversa;
- calcular una envolvente espectral de un conjunto de coeficientes de predicción lineal calculado sobre un cuadro de audio de dominio temporal; o
- 60 filtrar un espectro de una transformada de coseno discreta modificada (potencia) con un filtro de paso bajo.
14. Aparato de acuerdo con una de las reivindicaciones 7 a 13, que comprende además una operación de recorte de mosaicos de origen (1146) y una memoria (1148) para almacenar información de mosaicos de origen de un cuadro anterior que precede a un cuadro actual.

15. Aparato de acuerdo con una de las reivindicaciones 7 a 14,  
 en el que una operación de recorte de origen (1146) comprende analizar una pluralidad de mosaicos de origen con  
 respecto a su similitud y eliminar un mosaico de origen, que tiene una similitud con un mosaico de origen diferente  
 que es mayor que un umbral predefinido, de un conjunto de mosaicos potenciales utilizados para un cálculo de  
 5 correlación cruzada (1140).
16. Aparato de acuerdo con una de las reivindicaciones 11 a 15,  
 en el que la calculadora de parámetros está configurada para retener un conjunto de información de adaptación para  
 cada región de destino de un cuadro anterior, cuando ninguna de las regiones de origen en un cuadro actual se  
 10 correlaciona mejor con las regiones de destino del cuadro actual con respecto al cuadro anterior que un umbral  
 predefinido.
17. Método para decodificar una señal de audio codificada, que comprende:
- 15 decodificar (1102) una representación codificada de un primer conjunto de primeras porciones espectrales para  
 obtener un primer conjunto decodificado de primeras porciones espectrales (1101);  
 decodificar (1104) una representación paramétrica codificada de un segundo conjunto de segundas porciones  
 espectrales para obtener una representación decodificada de la representación paramétrica (1103), en el que la  
 representación codificada de la representación paramétrica (1103) incluye, para cada mosaico de frecuencia de  
 20 destino, una identificación de la región de origen como una información de adaptación; y  
 regenerar (1106) un mosaico de frecuencia de destino utilizando una región de origen del primer conjunto de  
 primeras porciones espectrales (1101) identificadas por la información de adaptación,  
 en el que la regeneración (1106) comprende usar un filtro de blanqueo controlable (1122, 1123),  
 en el que la representación decodificada de la representación paramétrica (1103) comprende información de  
 25 blanqueo (1123),  
 en el que la regeneración (1106) comprende aplicar la información de blanqueo (1122) a una región de origen  
 seleccionada de acuerdo con la información de adaptación, antes de efectuar un ajuste de envolvente espectral  
 (1128), cuando la información de blanqueo para la región de origen indica que la región de origen va a ser  
 30 en el que la aplicación del filtro de blanqueo comprende calcular una estimación de envolvente espectral de la  
 región de origen y dividir un espectro de la región de origen por una envolvente espectral indicada por la  
 estimación de envolvente espectral.
18. Método para codificar una señal de audio, que comprende:
- 35 convertir (1130) una señal de audio en una representación espectral;  
 analizar (1132) la representación espectral para determinar un primer conjunto de primeras porciones  
 espectrales que deben ser codificadas con una primera resolución espectral y un segundo conjunto de segundas  
 porciones espectrales que deben ser codificadas con una segunda resolución espectral, en el que la segunda  
 40 resolución espectral es inferior a la primera resolución espectral;  
 calcular (1134) similitudes entre las regiones de origen predefinidas y las regiones de destino, en el que una  
 región de origen comprende porciones espectrales y una región de destino comprende segundas porciones  
 espectrales,  
 en el que el cálculo (1134) comprende comparar los resultados de adaptación para diferentes pares de una  
 45 primera porción espectral y una segunda porción espectral para determinar un par de adaptación seleccionado y  
 para proporcionar información de adaptación que identifica el par de adaptación; y  
 codificar (1136) el primer conjunto de primeras porciones espectrales, en el que el primer conjunto de primeras  
 porciones espectrales comprende las regiones de origen predefinidas y porciones espectrales diferentes de las  
 regiones de origen predefinidas,  
 50 en el que el cálculo (1134) copiar un mosaico de origen múltiples veces a un mosaico de destino, hasta que el  
 mosaico de destino es llenado con múltiples copias del mosaico de origen, cuando el mosaico de destino es  
 mayor que el mosaico de origen o en el que el cálculo (1134) comprende blanquear espectralmente (1142) la  
 primera o la segunda porción espectral de los pares, antes de efectuar un procesamiento de correlación para  
 obtener la identificación de adaptación, en el que el blanqueo espectral (1142) comprende calcular una  
 55 estimación de envolvente espectral de la primera o la segunda porción espectral y dividir el espectro de la  
 primera o la segunda porción espectral, respectivamente, por una envolvente espectral indicada por la  
 estimación de envolvente espectral.
19. Programa informático para realizar, cuando se ejecuta en un ordenador o un procesador, cualquiera de los  
 60 métodos de acuerdo con la reivindicación 17 o 18.

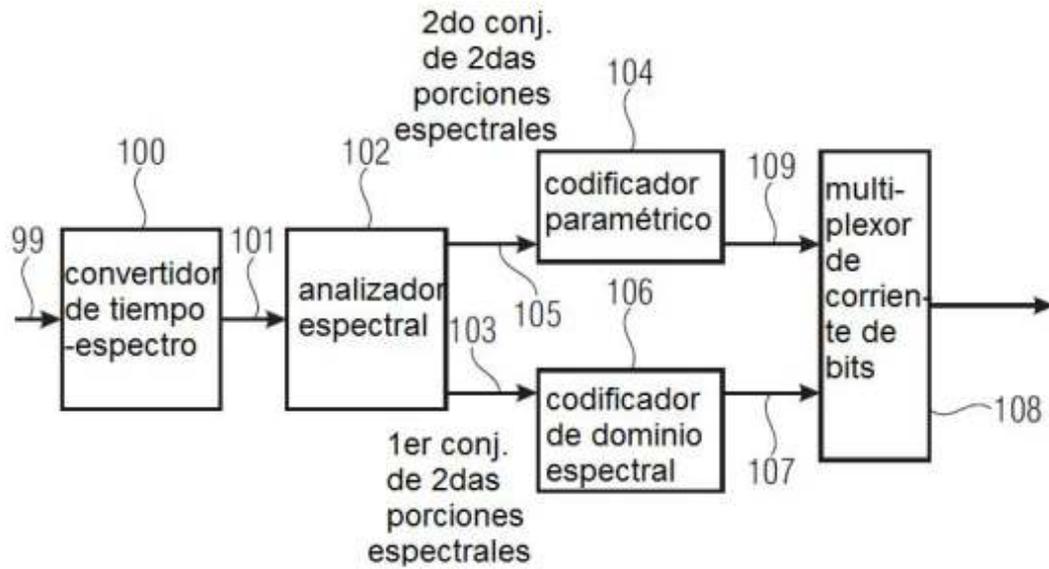


FIG 1A

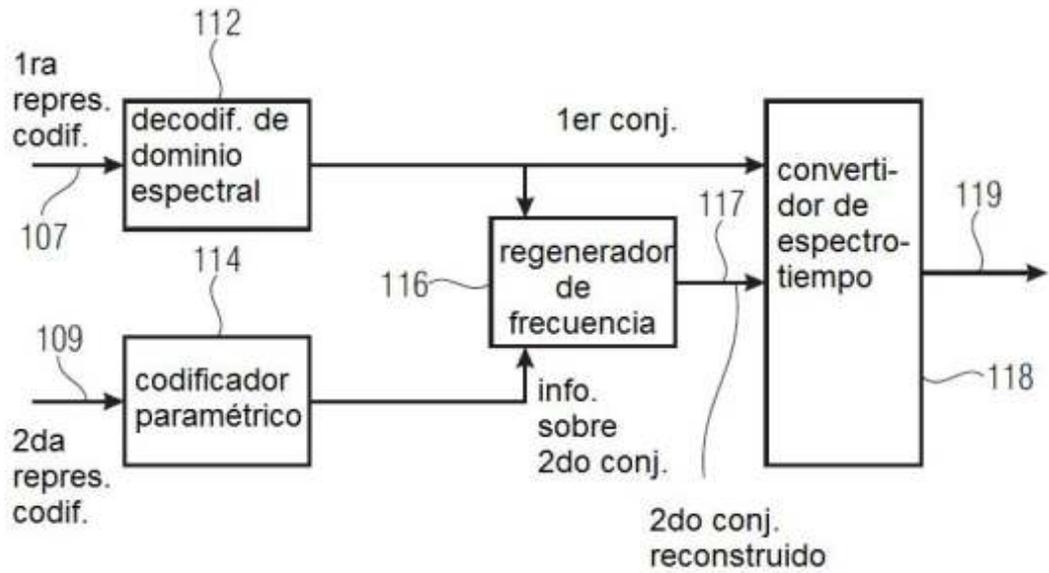


FIG 1B

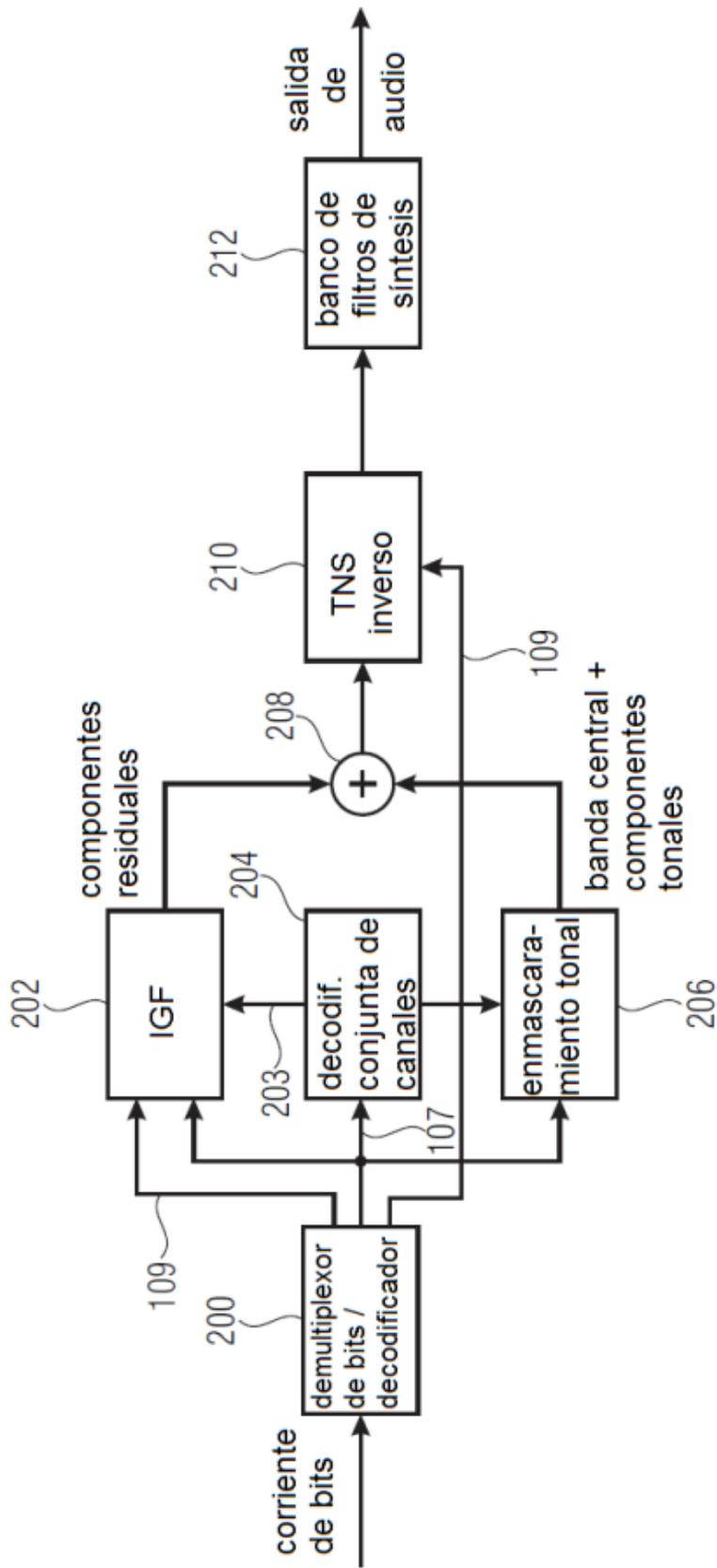


FIG 2A

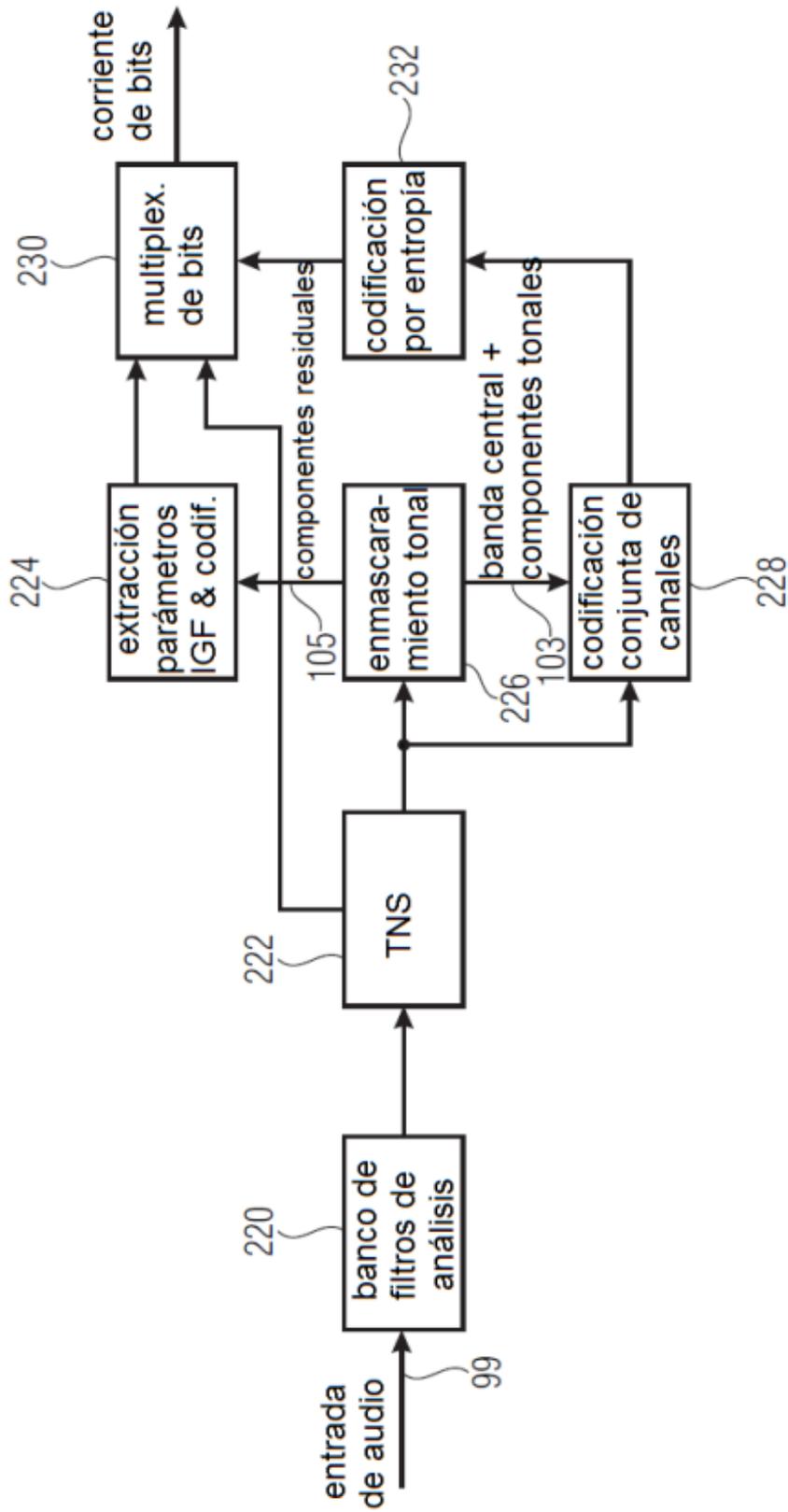


FIG 2B

- 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));

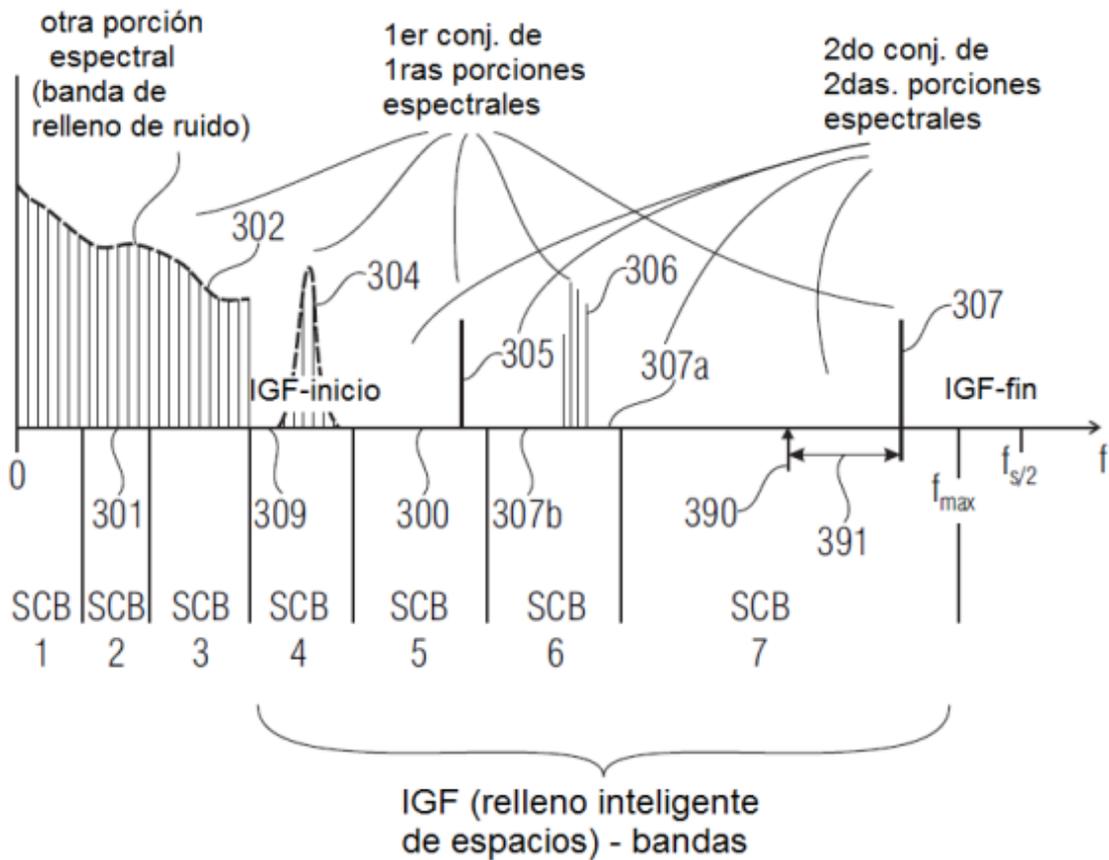


FIG 3A

SCB: Banda de Factor de Escala

SF: Factor de Ajuste de Escala

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

308
310
312

FIG 3B

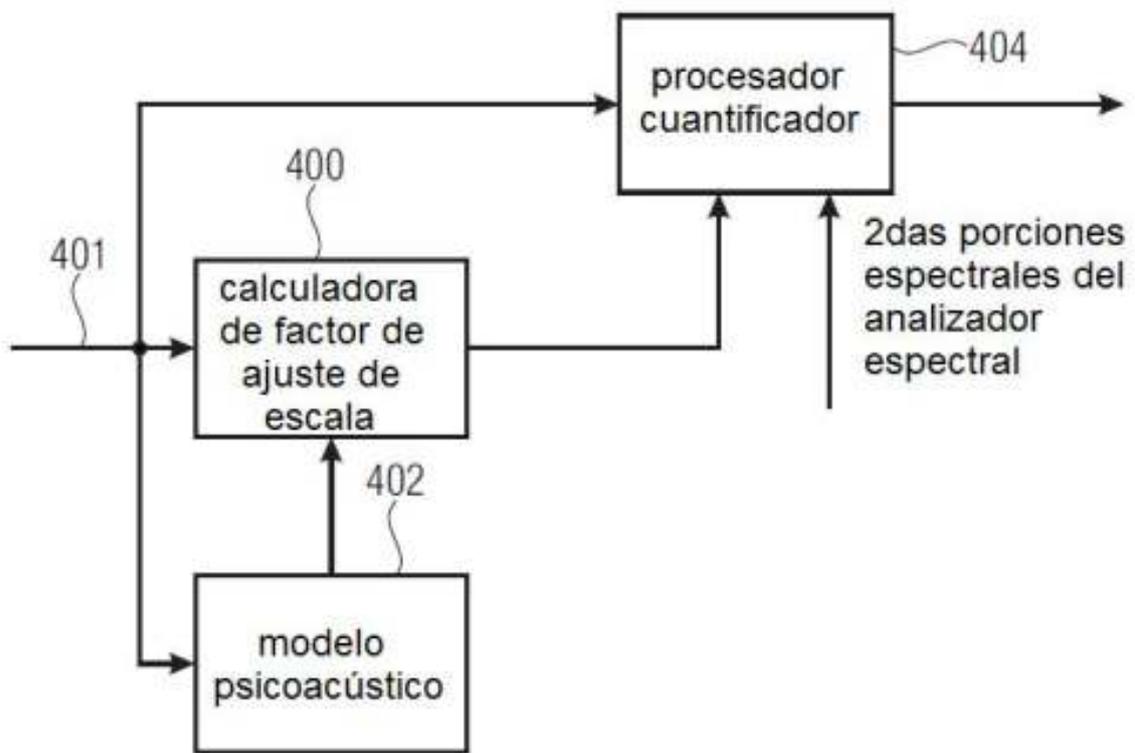


FIG 4A

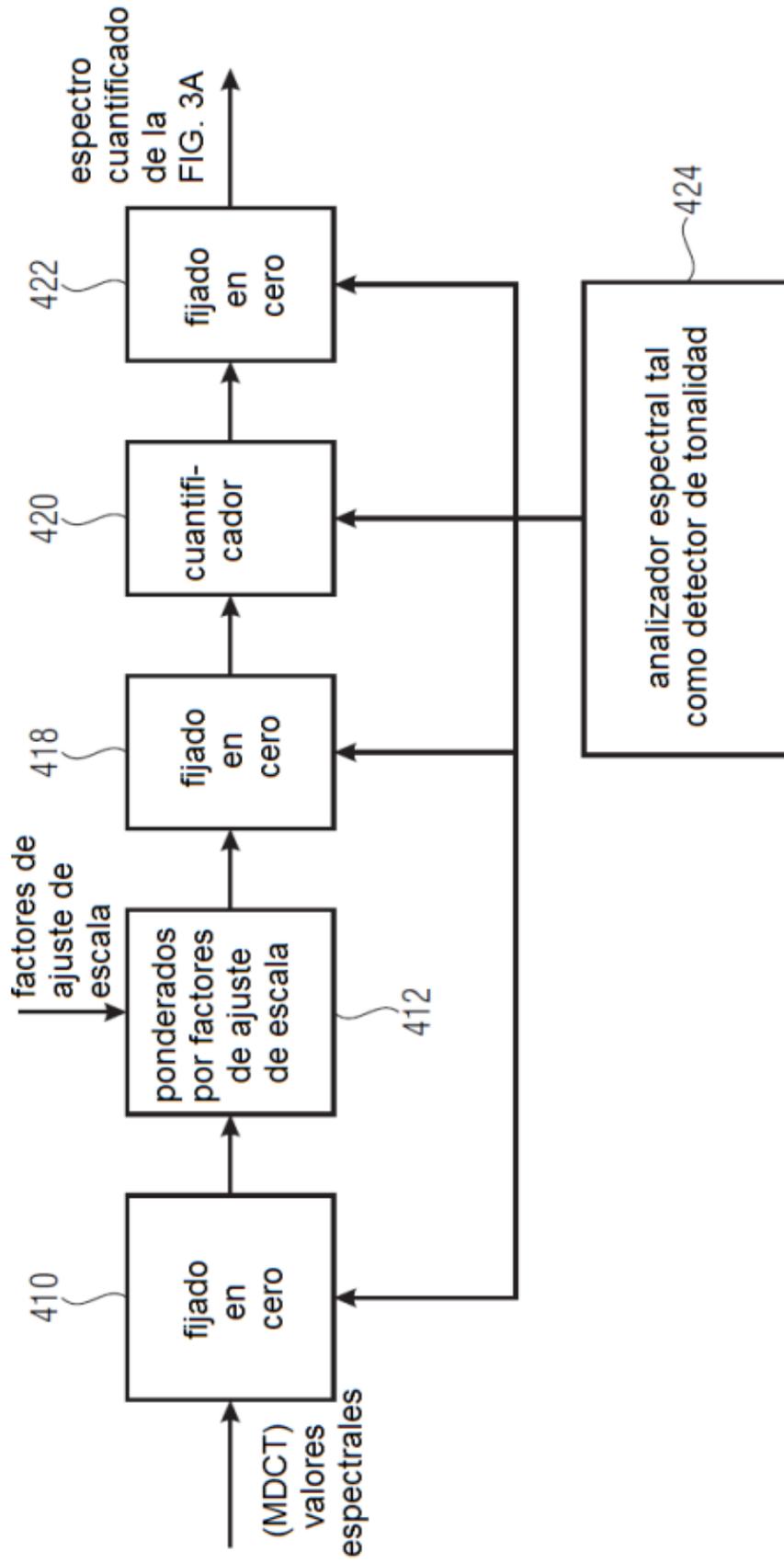


FIG 4B  
(PROCESADOR CUANTIFICADOR)

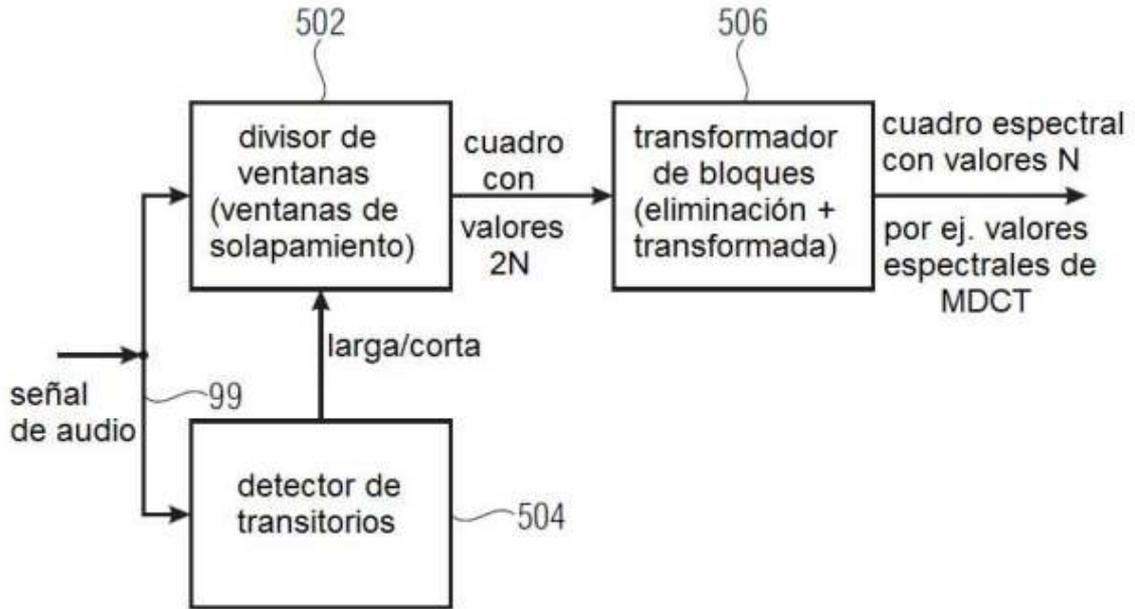


FIG 5A  
(OTRAS PORCIONES ESPECTRALES)

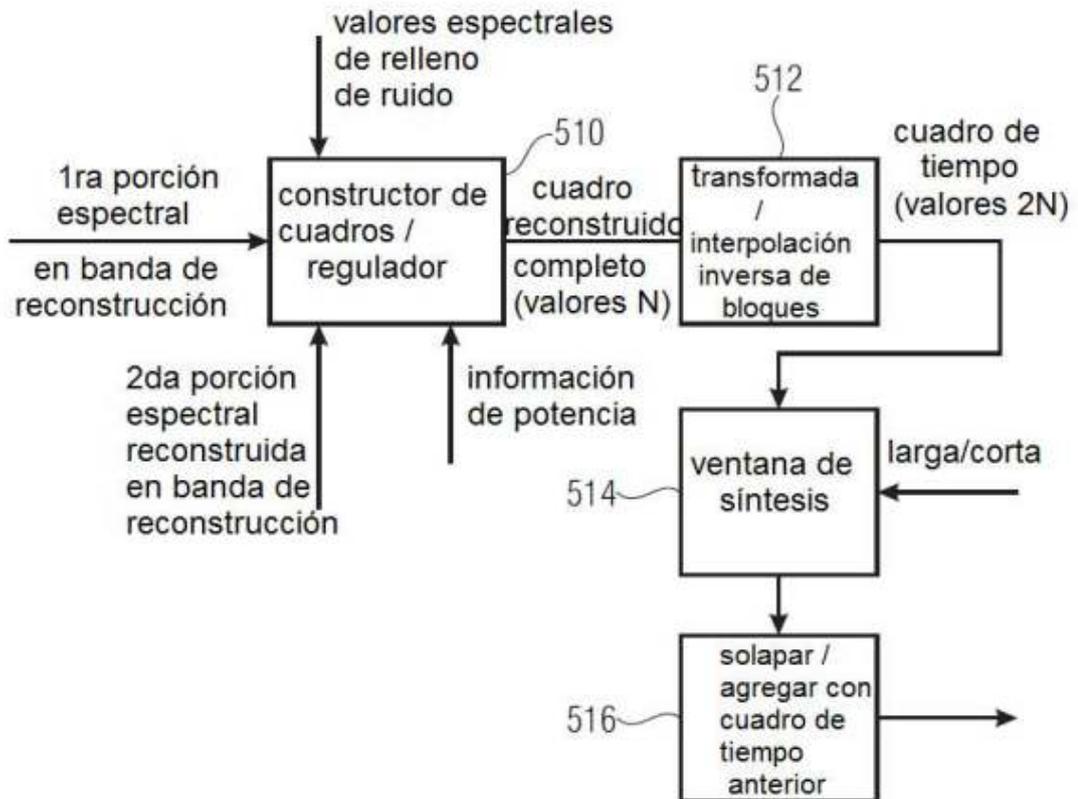


FIG 5B

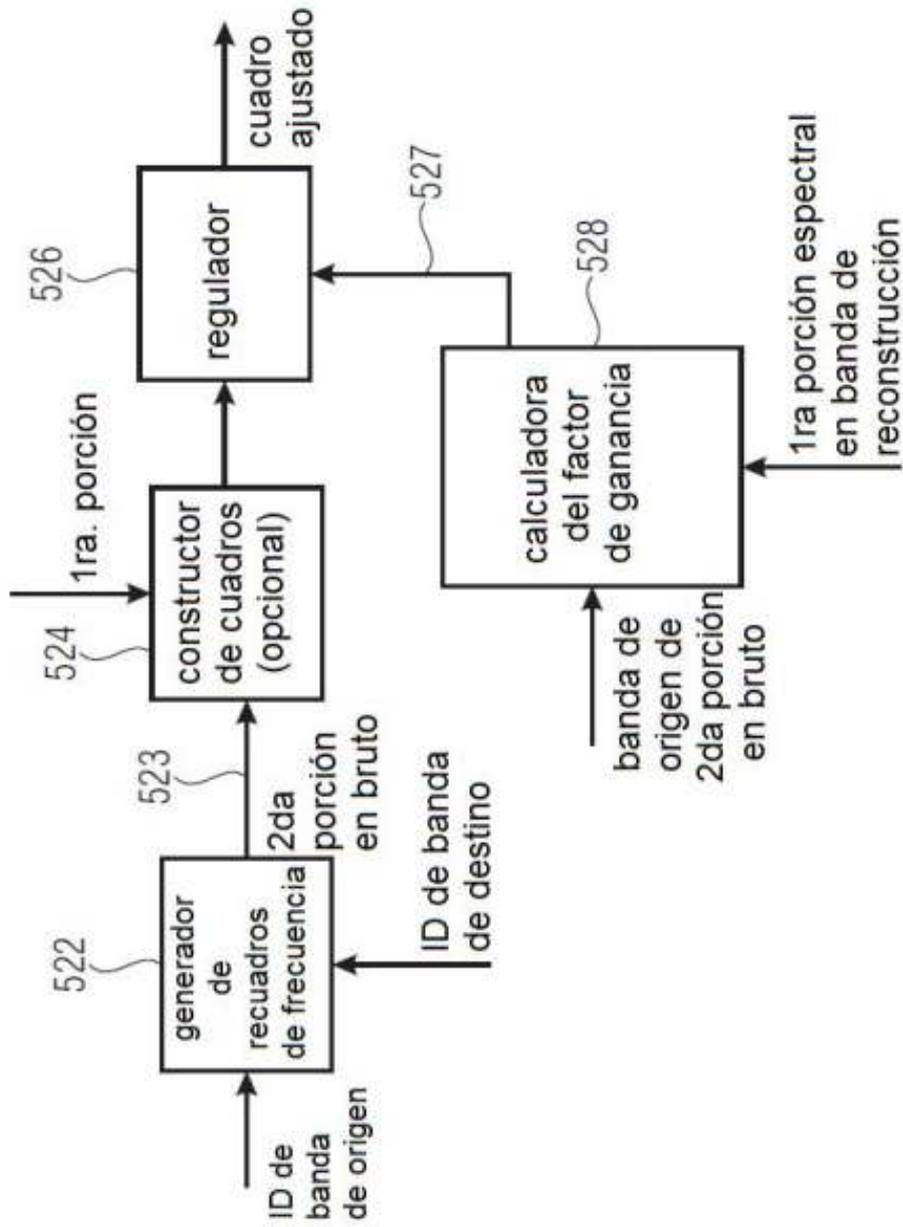


FIG 5C

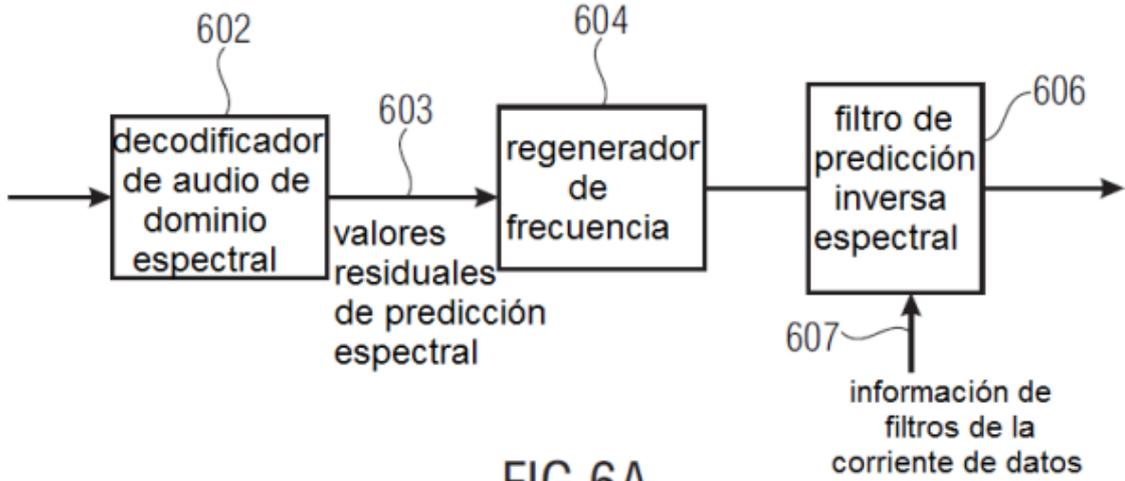


FIG 6A

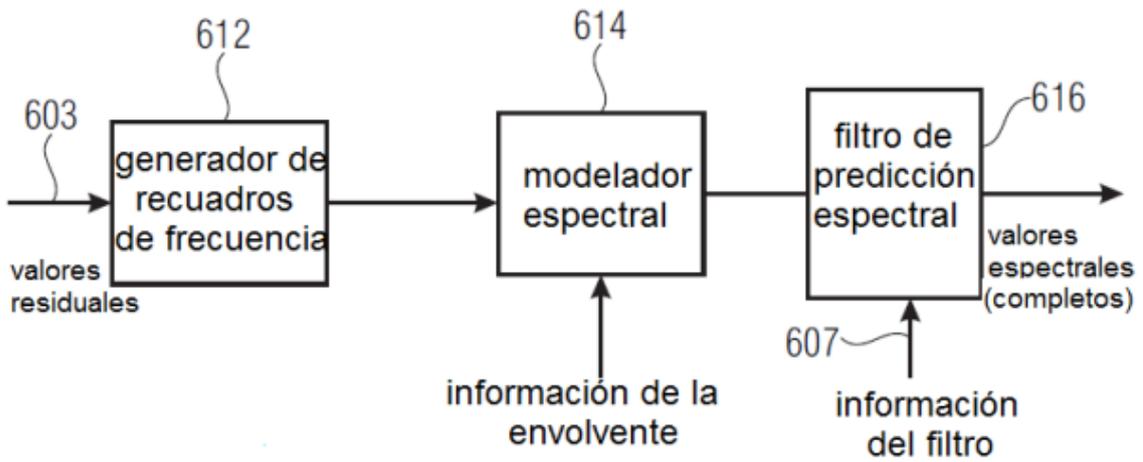


FIG 6B

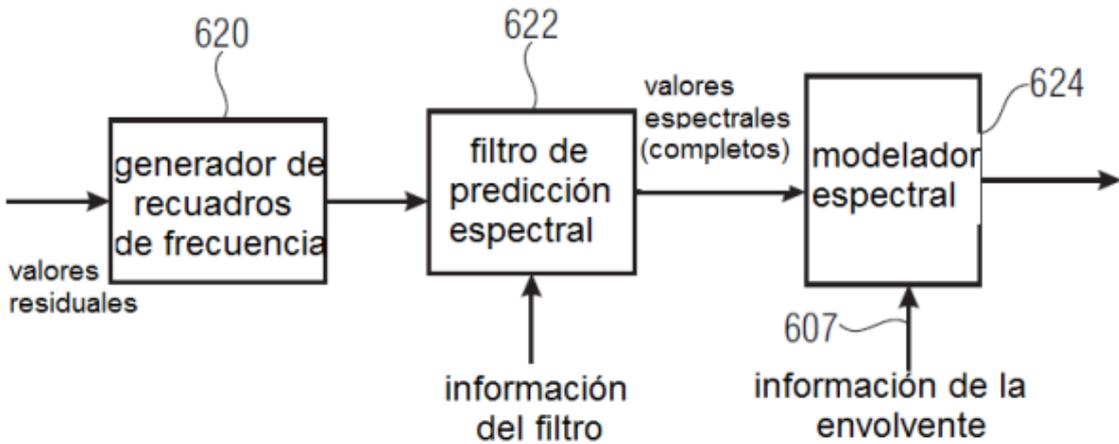


FIG 6C

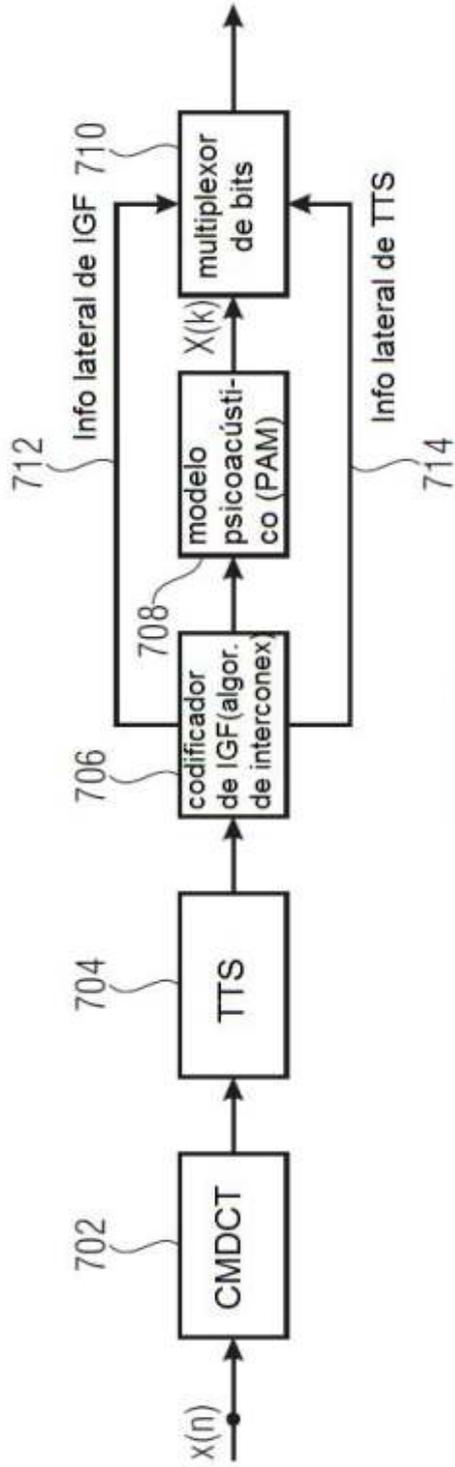


FIG 7A  
(CODIFICADOR)

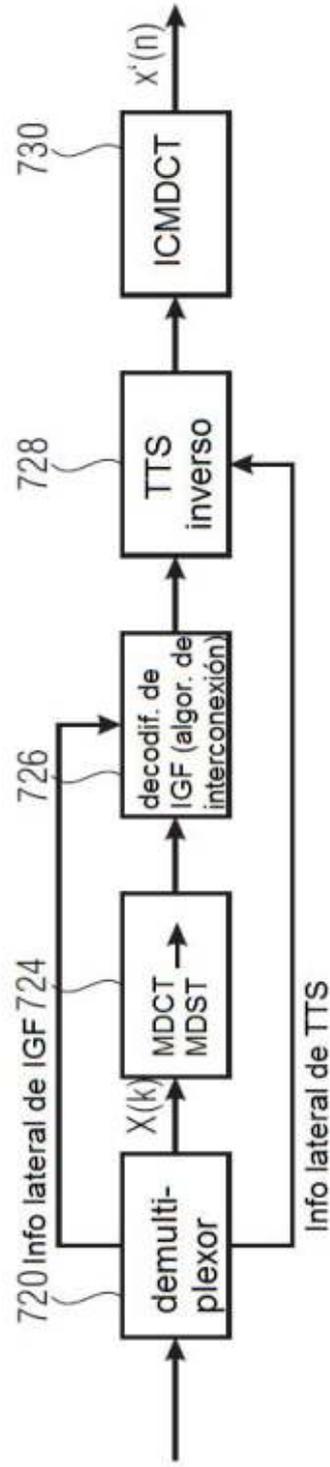
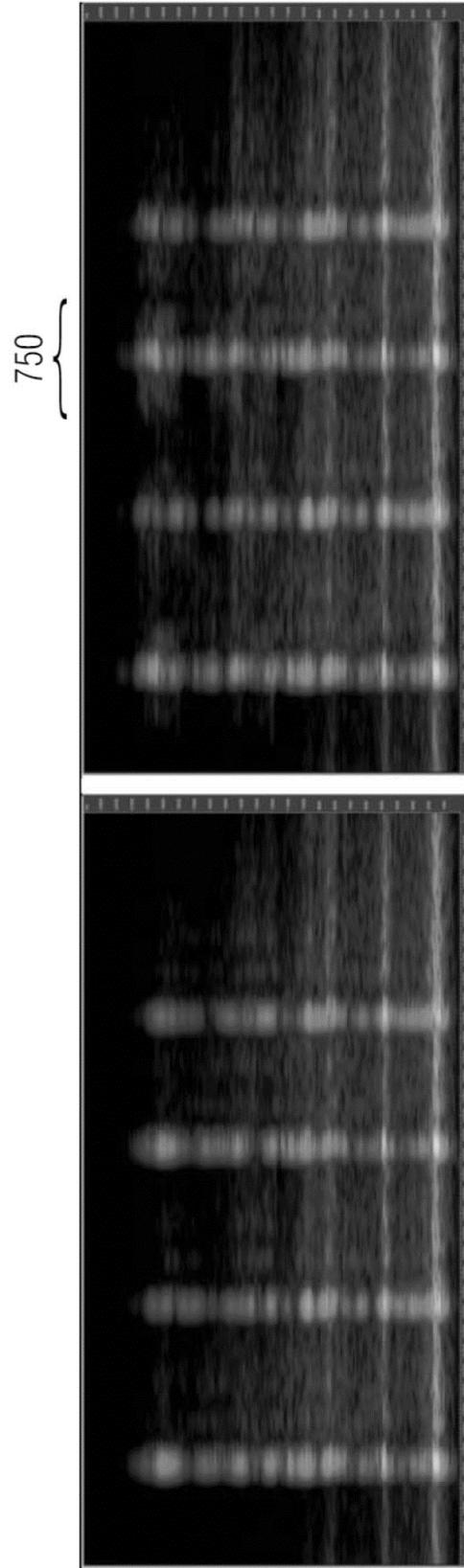
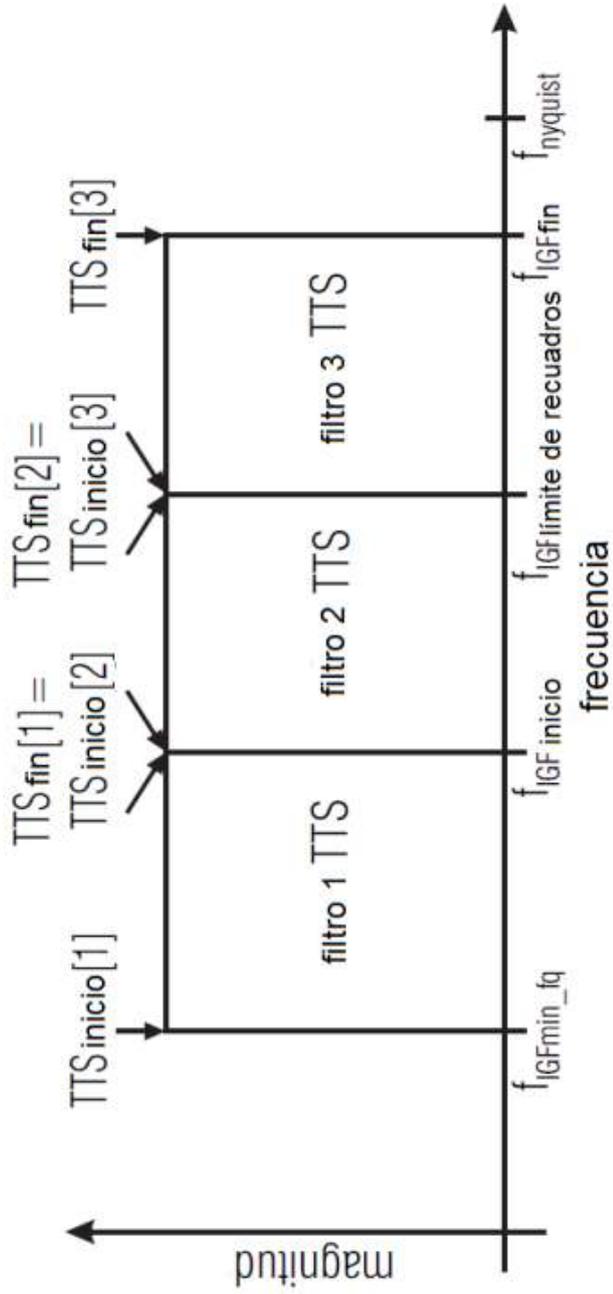


FIG 7B  
(DECODIFICADOR)



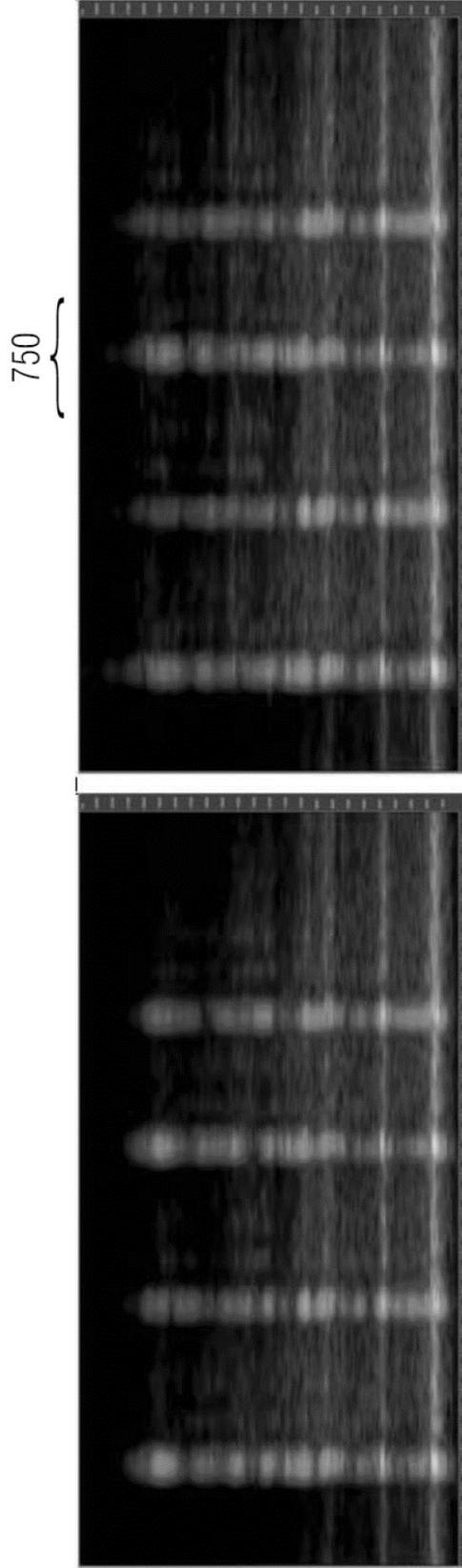
ESPECTROGRAMA DE TRANSFORMADA DISCRETA DE FOURIER (DFT) DE LA SEÑAL ORIGINAL (Panel Izquierdo);  
ESPECTROGRAMA DE DFT DE LA SEÑAL EXTENDIDA SIN TTS (Panel Derecho).

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

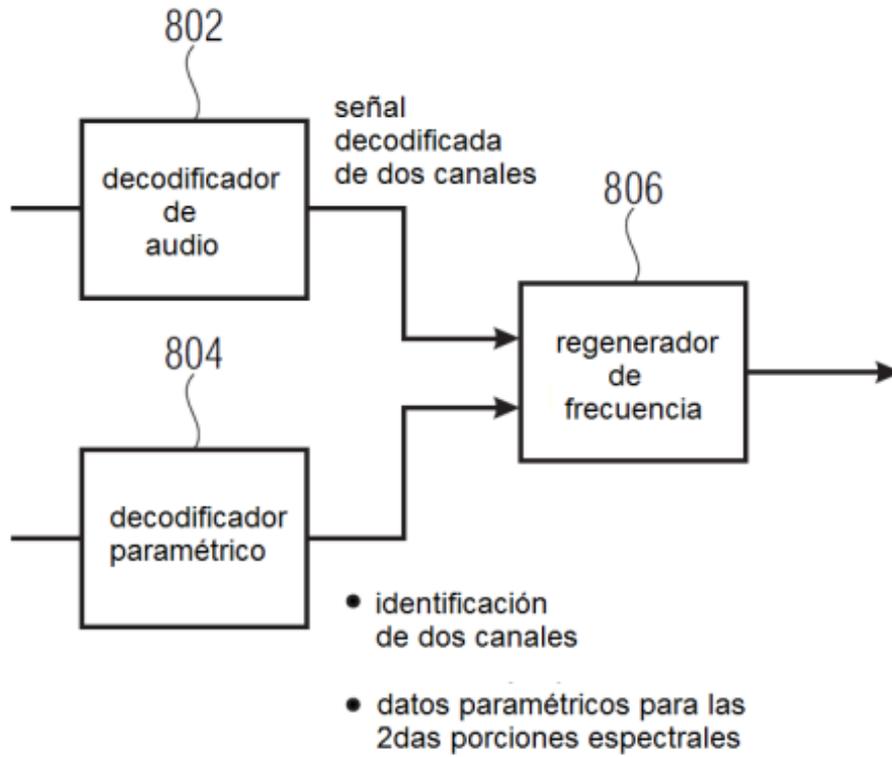


FIG 8A

rango de origen	1	1	2	2
rango de destino	1	2	1	2

1: primera representación de dos canales (por ej. L/R)

2: segunda representación de dos canales (por ej. M/S)

FIG 8B

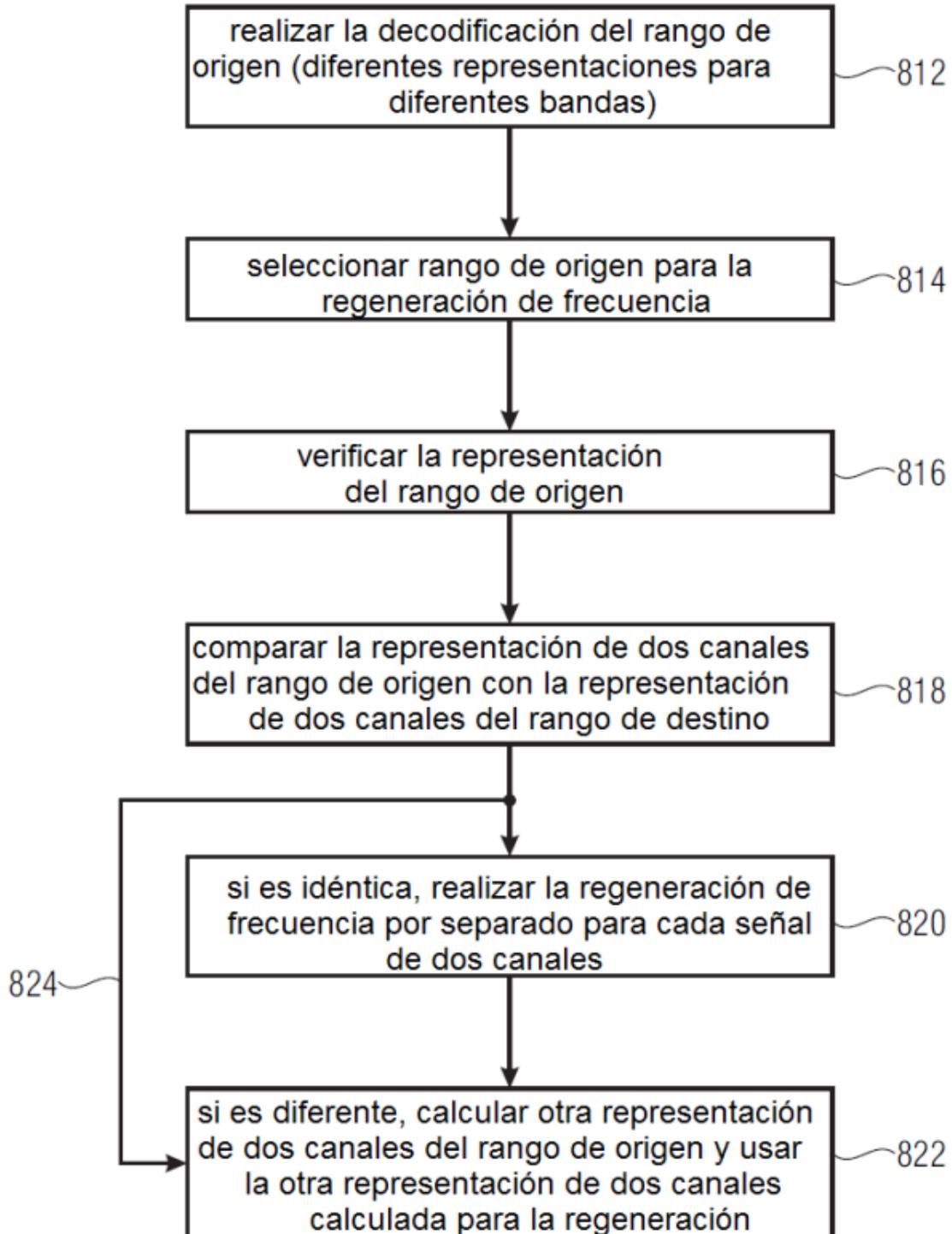


FIG 8C

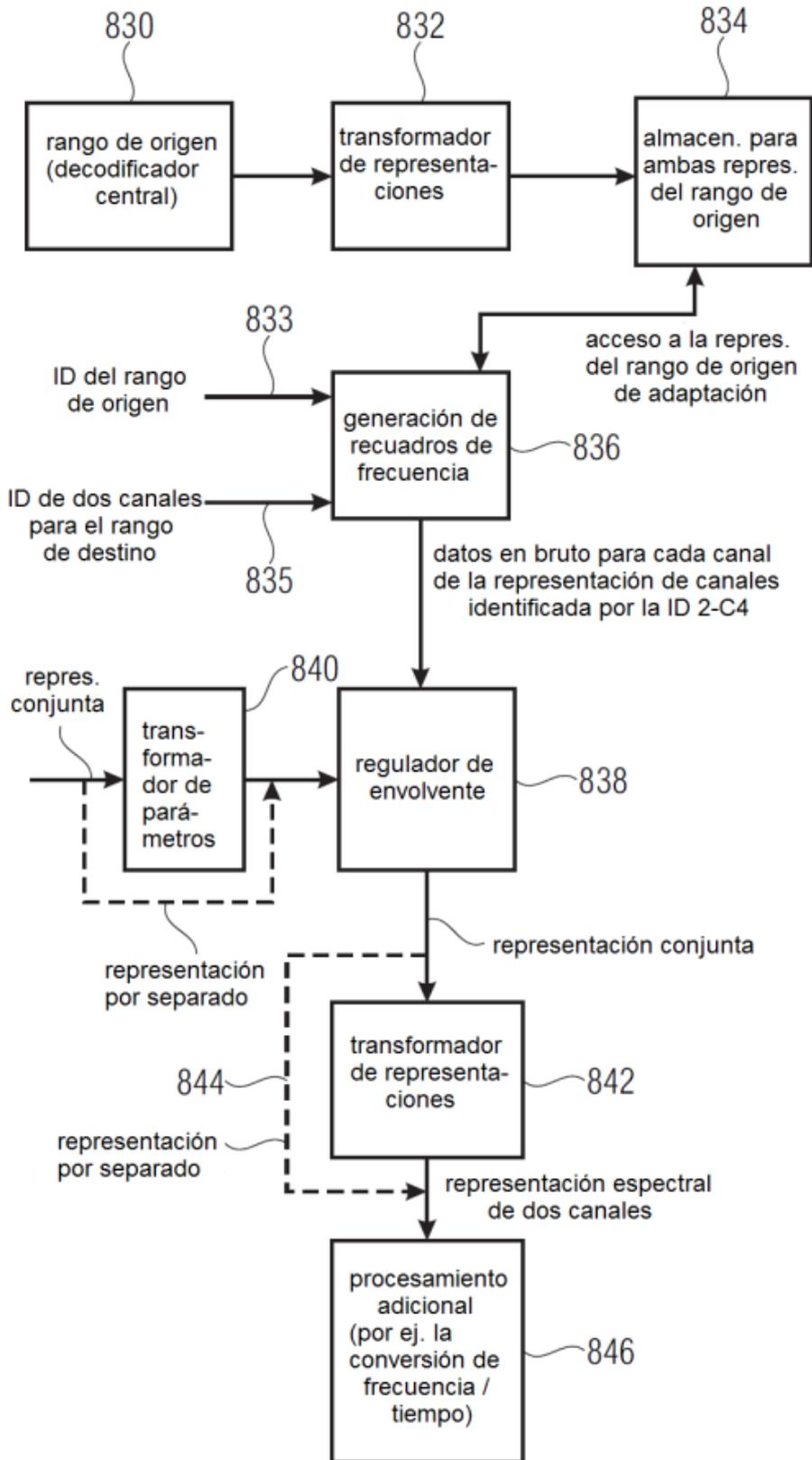


FIG 8D

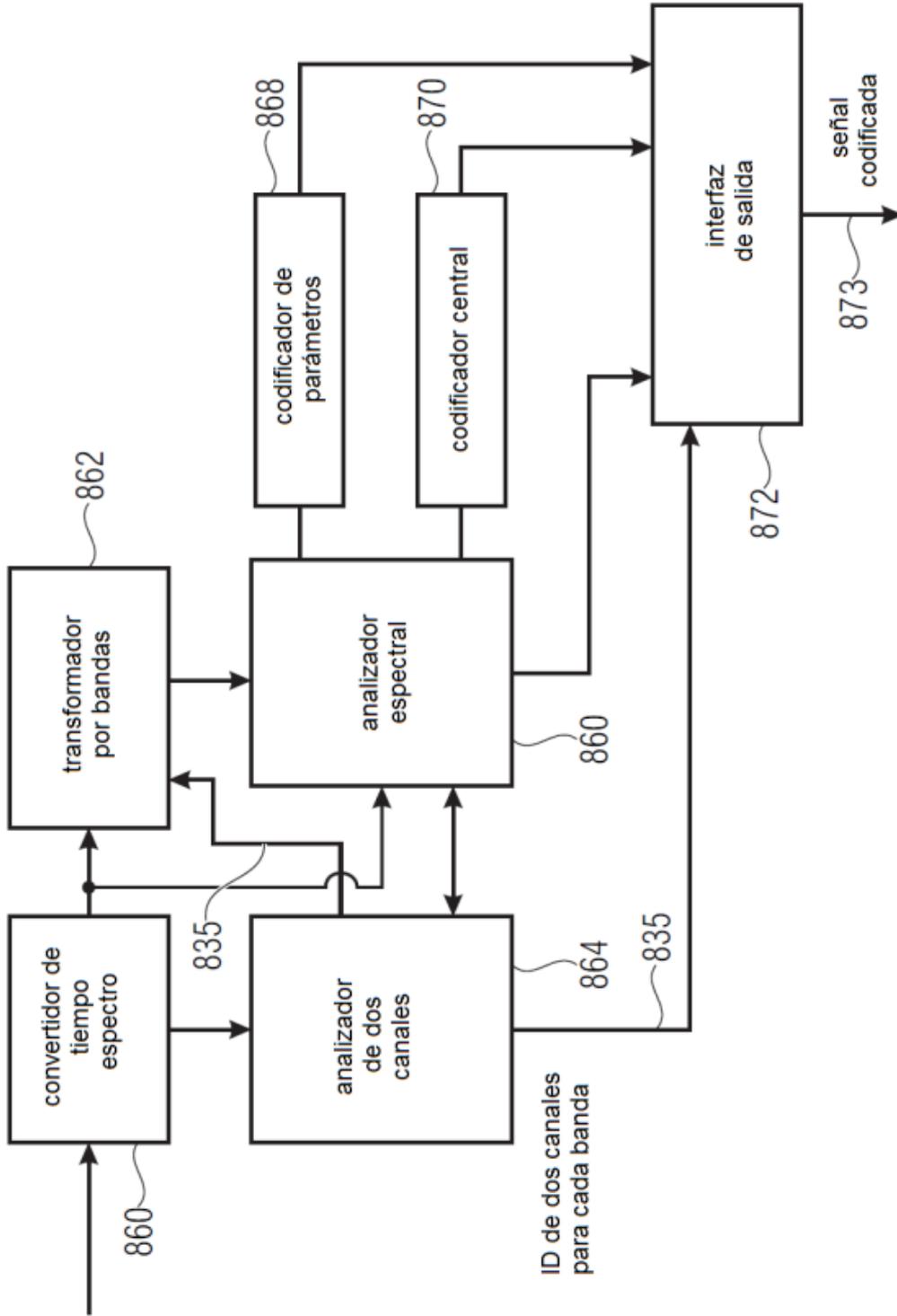


FIG 8E

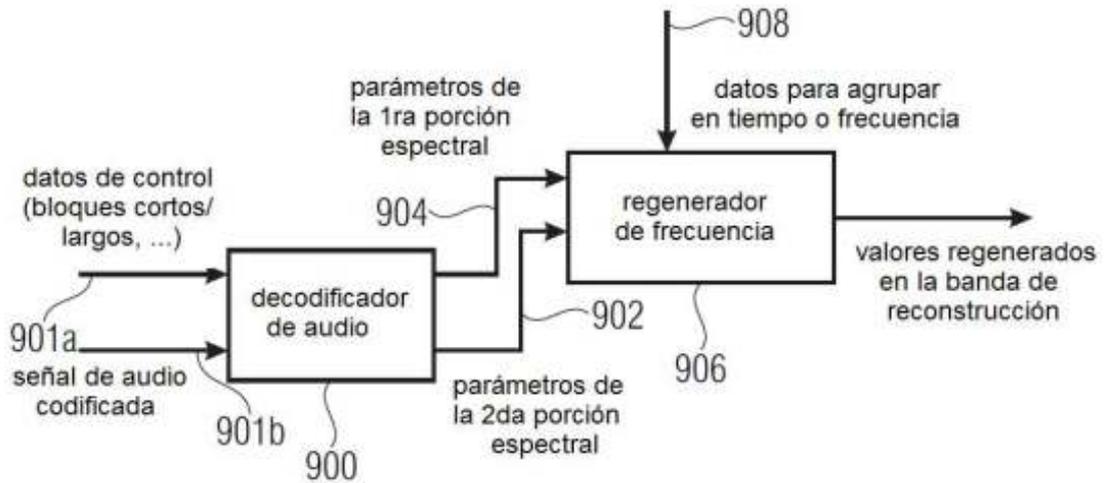


FIG 9A

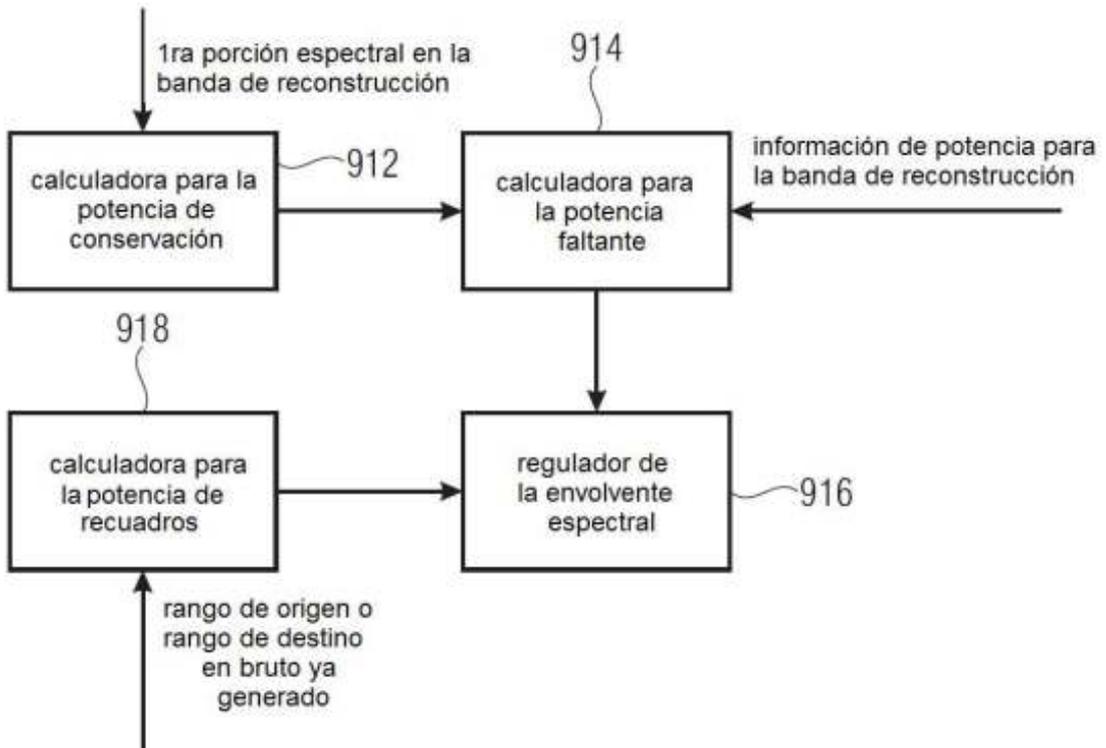
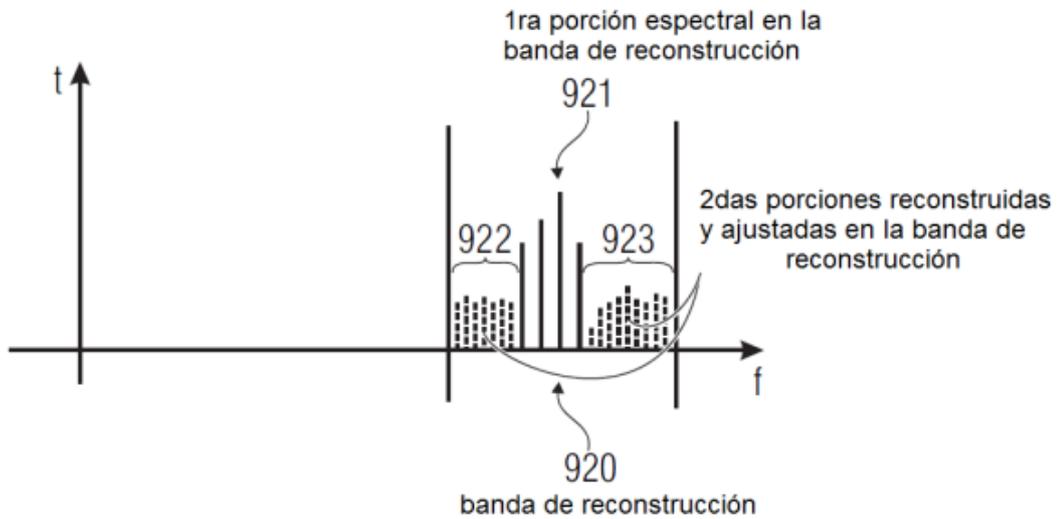


FIG 9B



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

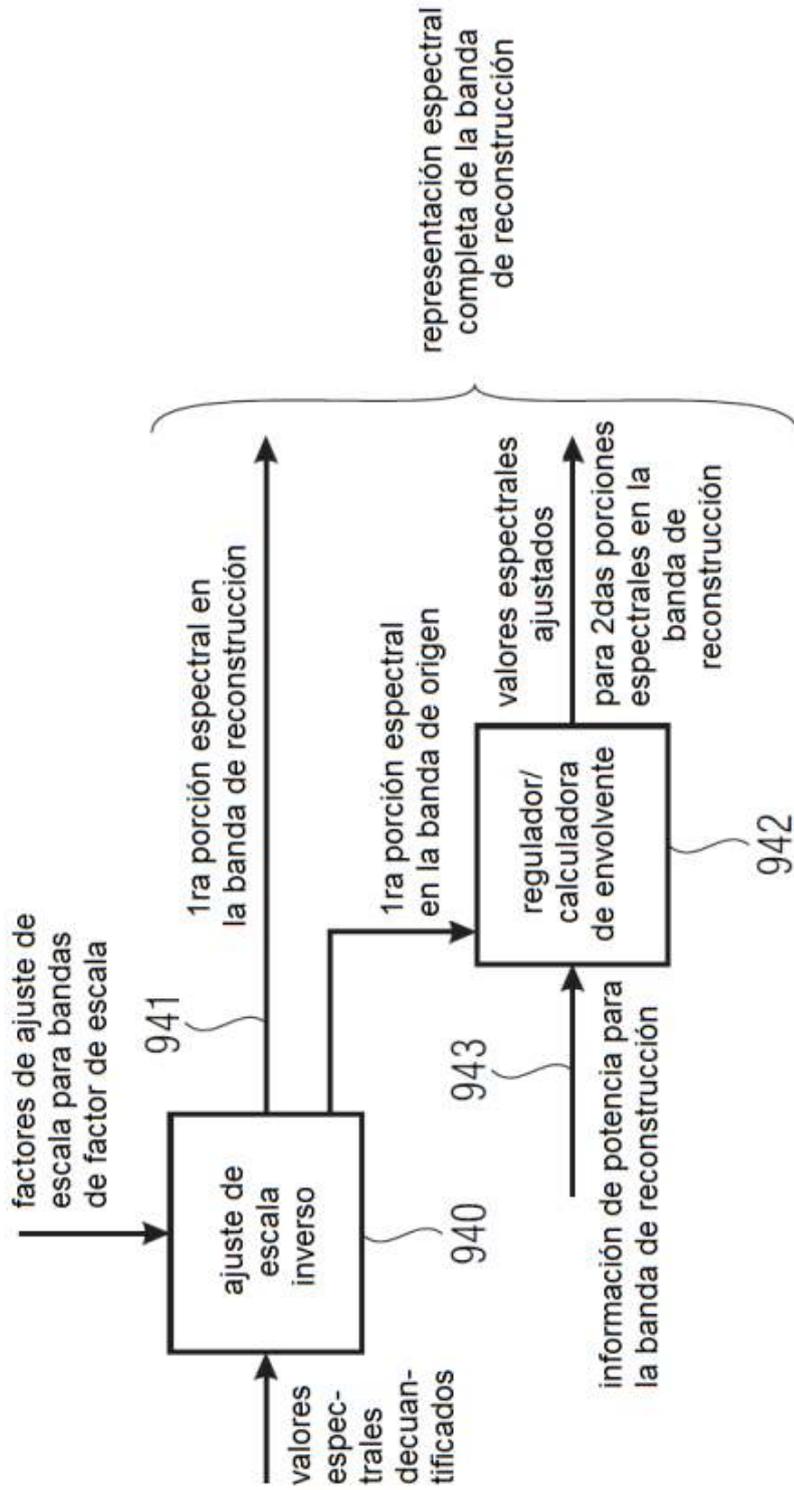


FIG 9D

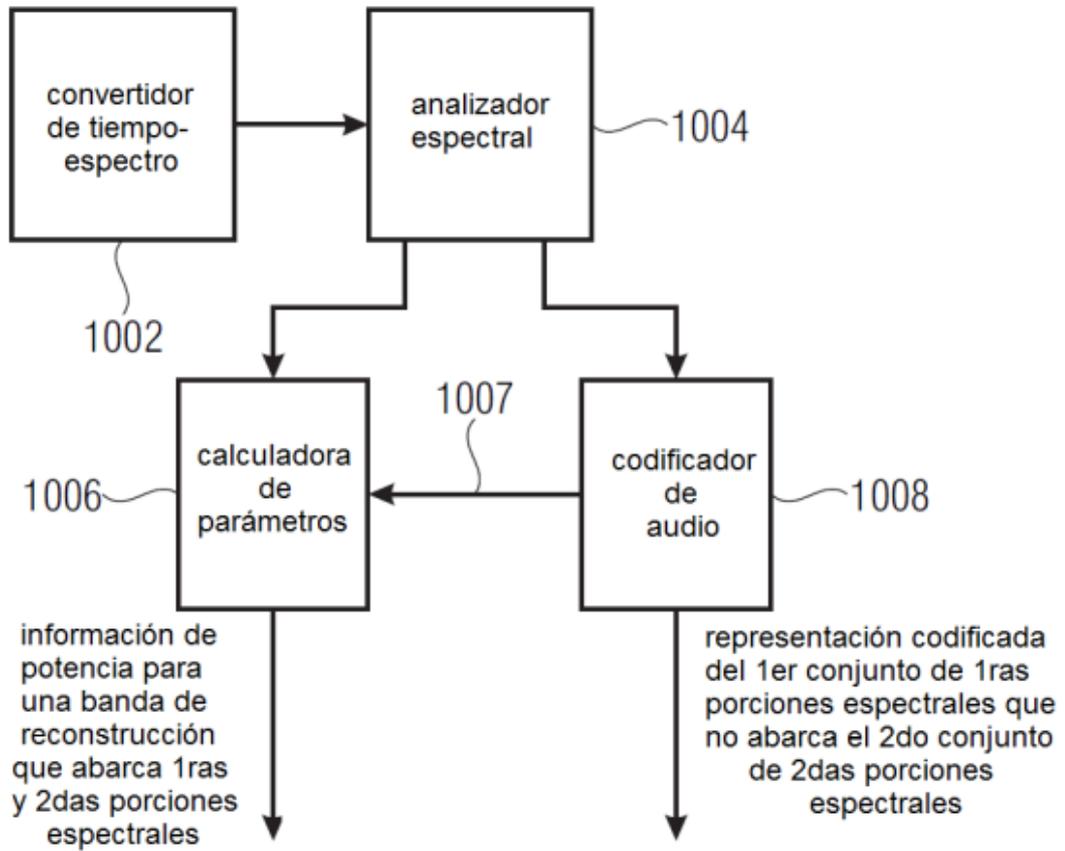


FIG 10A

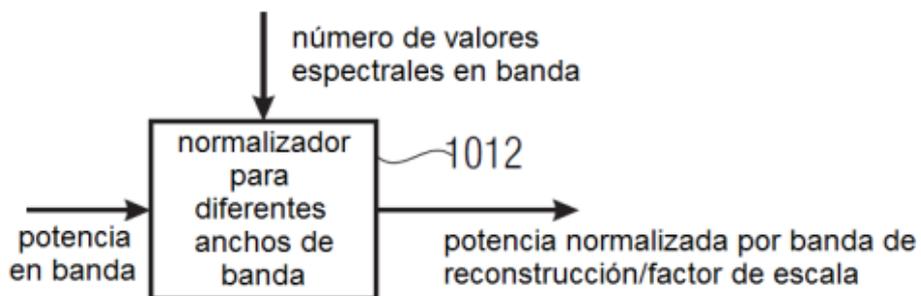


FIG 10B

info de control del codificador central

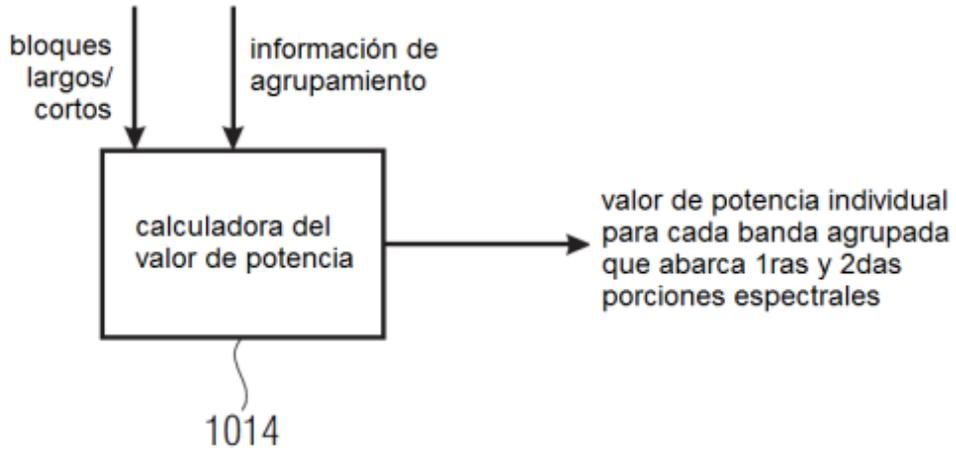


FIG 10C

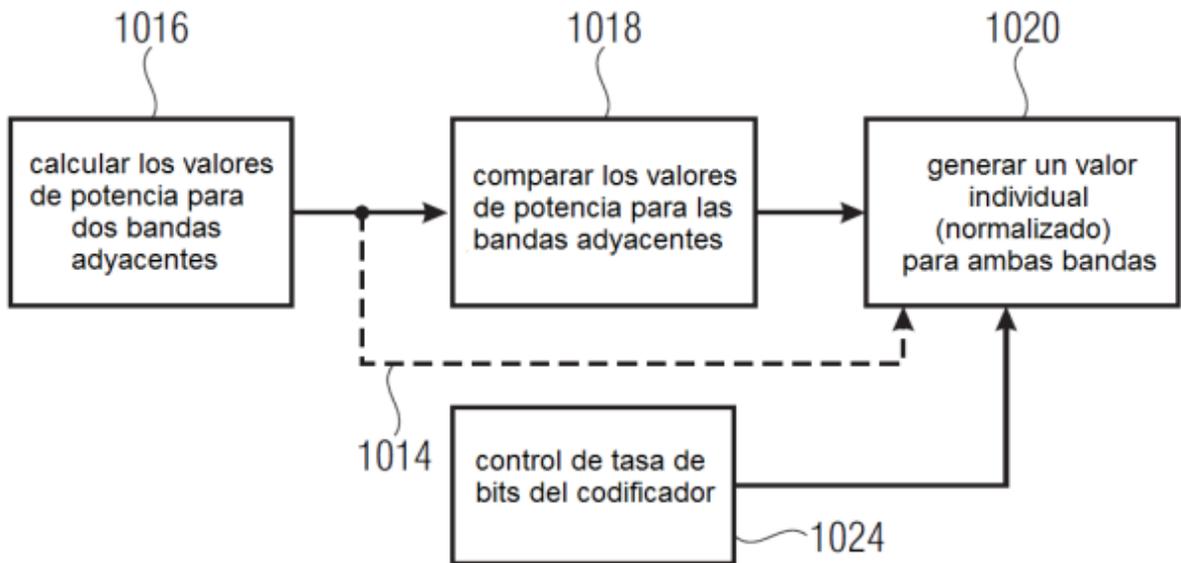


FIG 10D

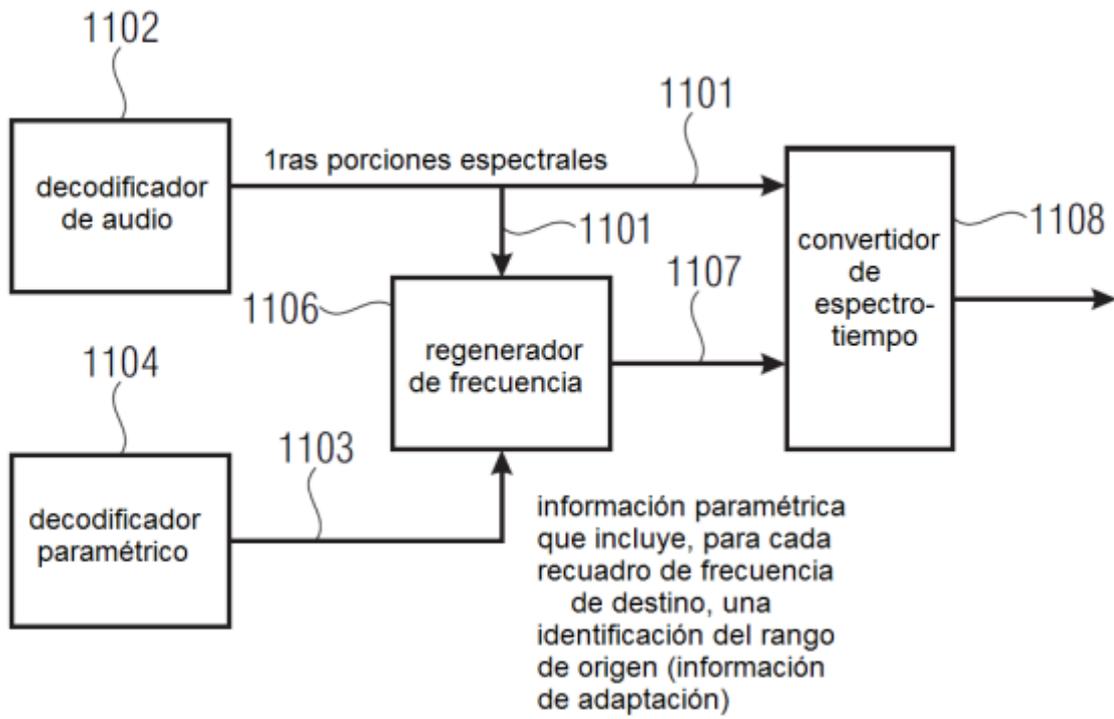


FIG 11A

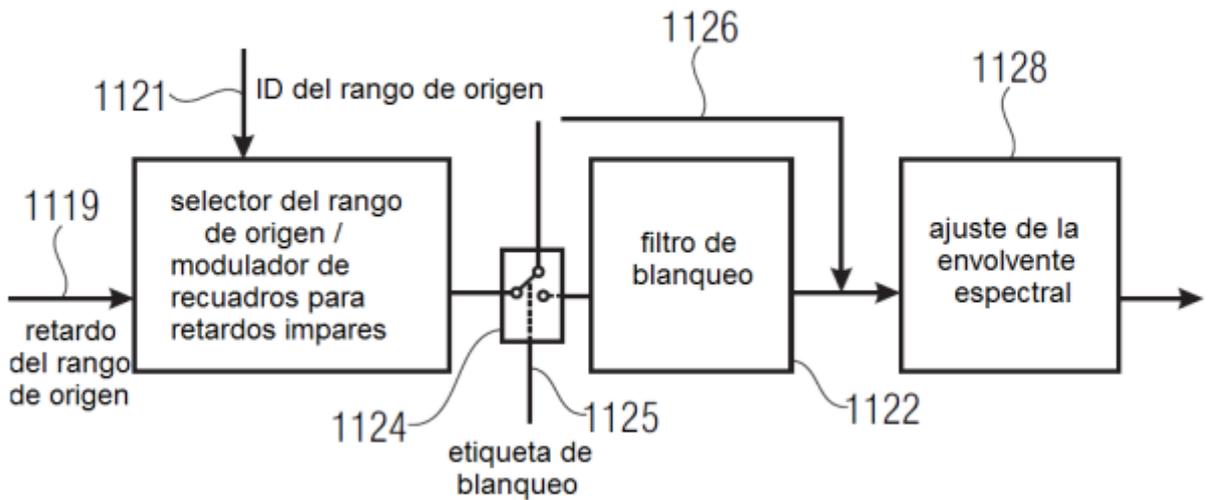


FIG 11B

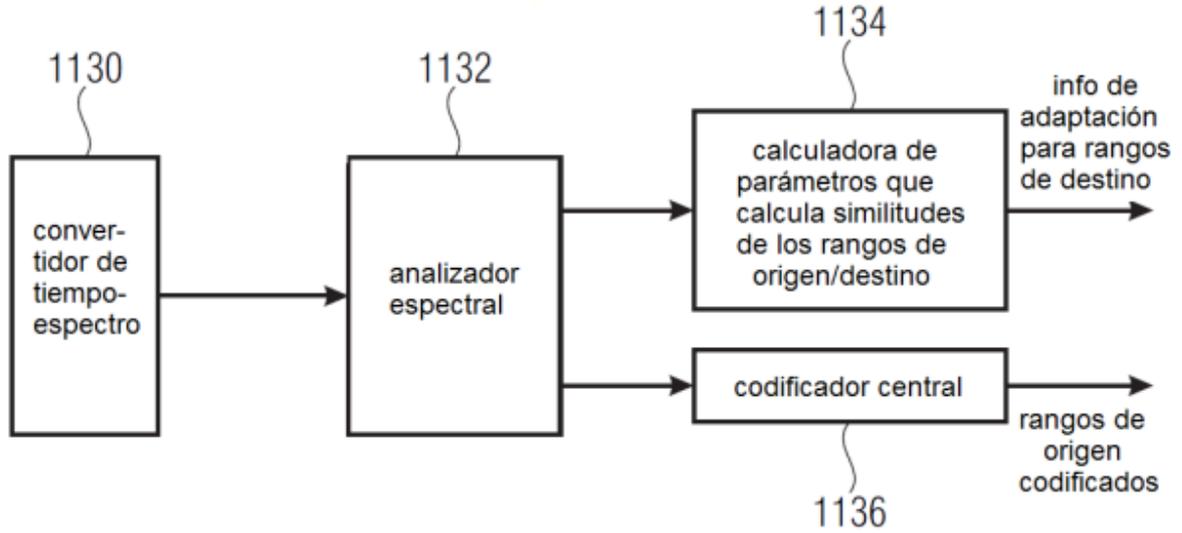


FIG 11C

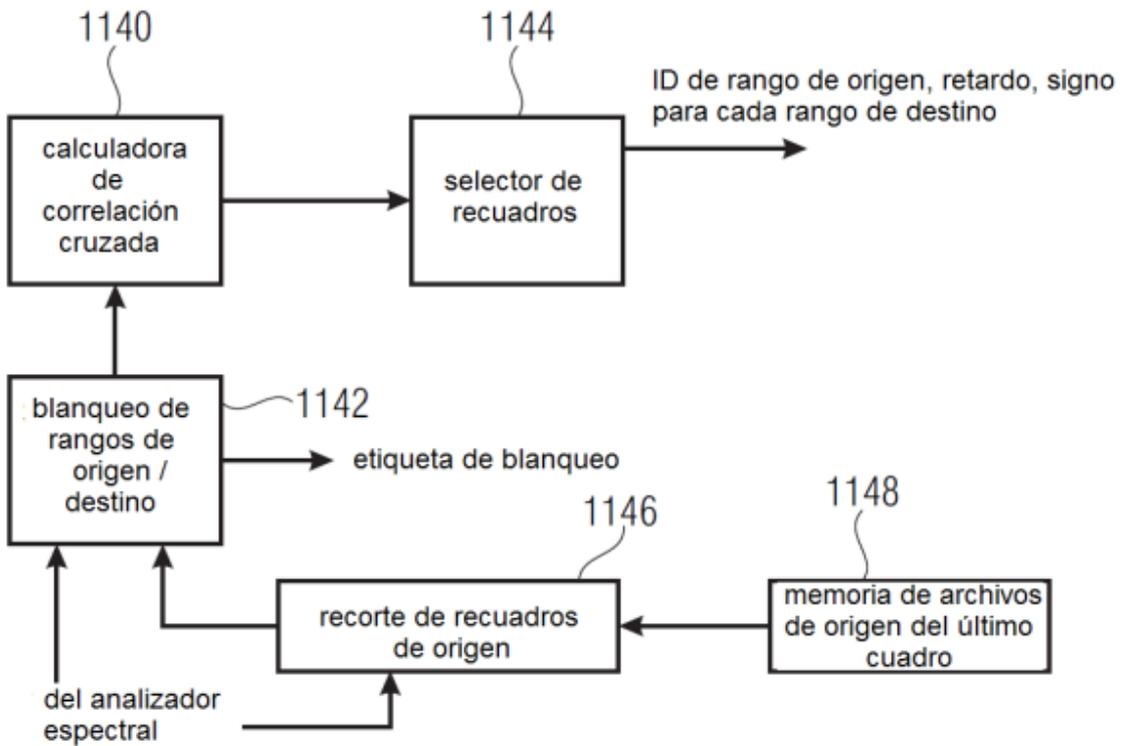


FIG 11D

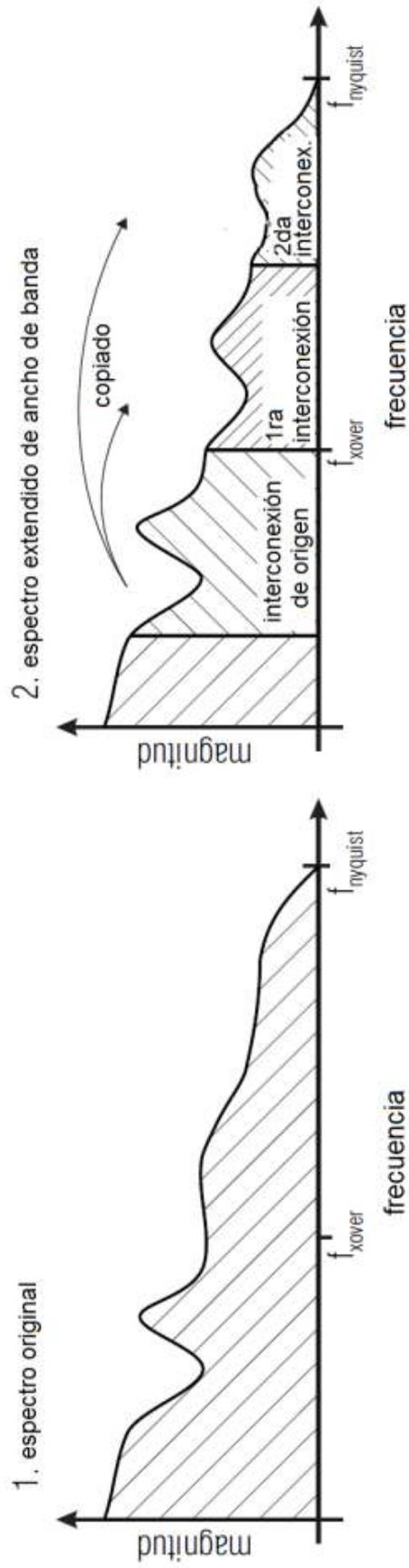


FIG 12A

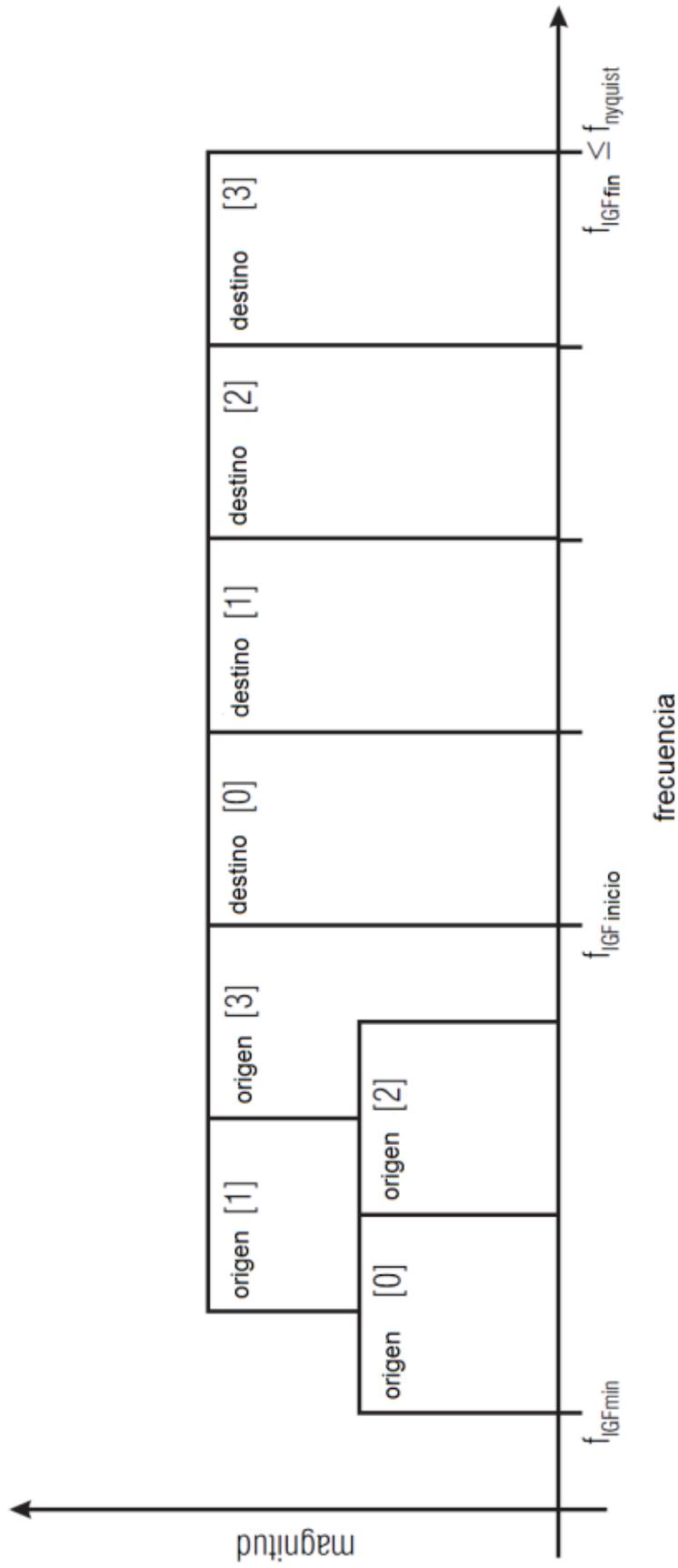


FIG 12B

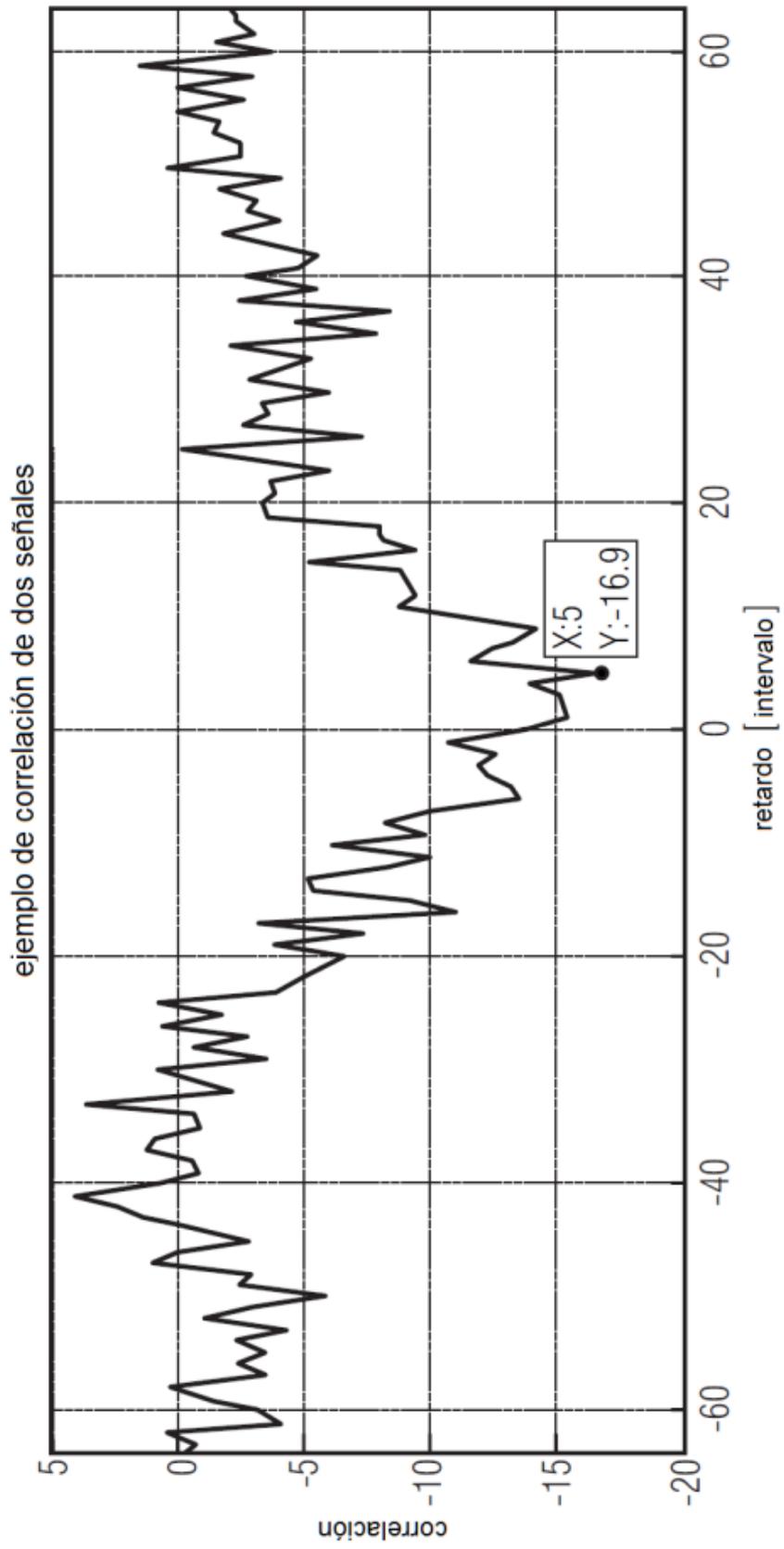


FIG 12C

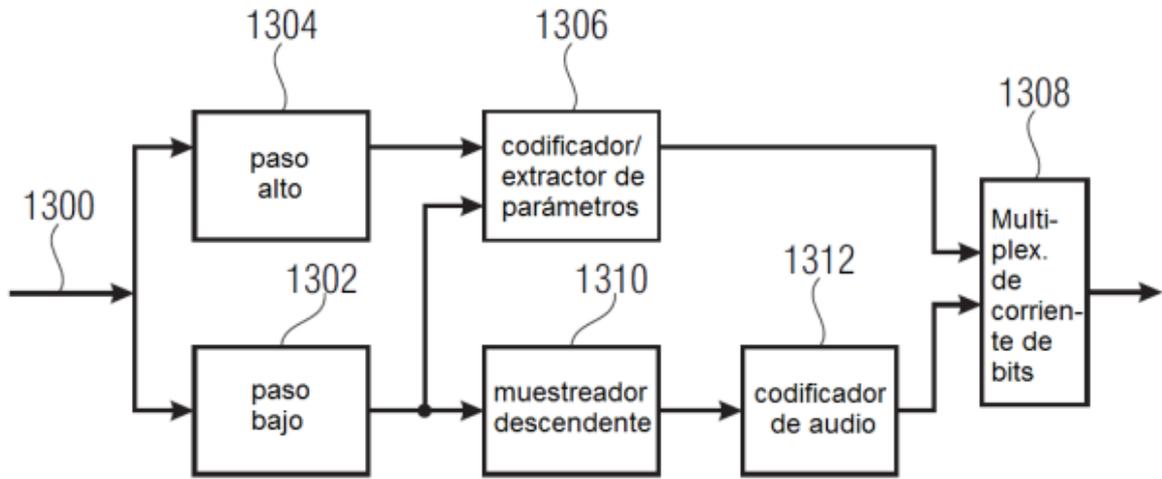


FIG 13A

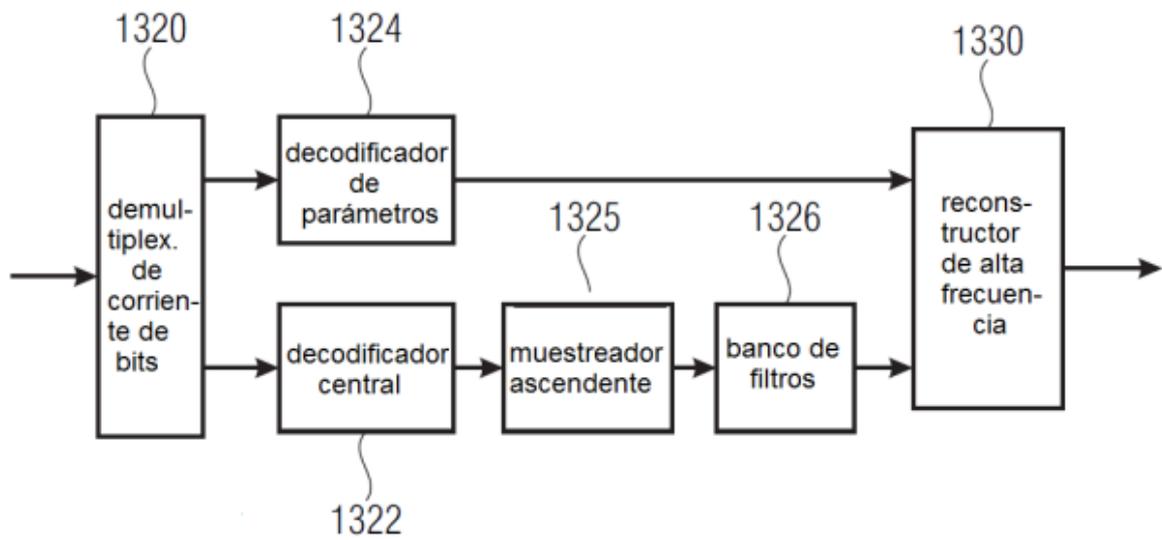


FIG 13B