

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 813 940**

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/EP2014/065118**

87 Fecha y número de publicación internacional: **29.01.2015 WO15010953**

96 Fecha de presentación y número de la solicitud europea: **15.07.2014 E 14739811 (9)**

97 Fecha y número de publicación de la concesión europea: **08.07.2020 EP 3017448**

54 Título: **Aparato, método y programa informático para decodificar una señal de audio codificada**

30 Prioridad:

22.07.2013 EP 13177346

22.07.2013 EP 13177350

22.07.2013 EP 13177353

22.07.2013 EP 13177348

18.10.2013 EP 13189382

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

25.03.2021

73 Titular/es:

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

72 Inventor/es:

**DISCH, SASCHA;
GEIGER, RALF;
HELMRICH, CHRISTIAN;
NAGEL, FREDERIK;
NEUKAM, CHRISTIAN;
SCHMIDT, KONSTANTIN y
FISCHER, MICHAEL**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 813 940 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Aparato, método y programa informático para decodificar una señal de audio codificada

- 5 La presente invención se refiere a codificación/decodificación de audio y, particularmente, a codificación de audio mediante el Llenado Inteligente de Huecos (IGF).

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 sistema de codificación de audio descrito en el presente documento 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.

30 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.

35 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 banco 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.

45 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].

55 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.

60 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).

5 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
10 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
15 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 u 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.

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 capaces de codificar señales de banda ancha utilizando los métodos de extensión de ancho de banda (BWE) [1-2]. 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. Otros esquemas que son funcionales para rellenar espacios espectrales, por ejemplo, el Relleno Inteligente de Espacios (IGF), usan mosaicos espectrales denominados vecinos para regenerar partes de los espectros de alta frecuencia (HF) de señal de audio. En general, la región de alta frecuencia (HF) consta de múltiples conexiones o mosaicos adyacentes y cada una de estas conexiones o mosaicos 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 y la colocación de mosaicos dentro de una representación de banco de filtros copiando un conjunto de coeficientes de sub-bandas adyacentes desde una región de origen hasta la región de destino. Sin embargo, para algunos contenidos de señal, el ensamblaje de la señal reconstruida a partir de la banda de baja frecuencia (LF) y conexiones adyacentes dentro de la banda de baja frecuencia (LF) puede llevar a batido de frecuencias, disonancia e irregularidades auditivas.

Por tanto, en [19], se presenta el concepto de filtrado de banda de guarda de disonancia en el contexto de un sistema de extensión de ancho de banda (BWE) basado en bancos de filtros. Se sugiere aplicar con eficiencia un filtro de muesca de aprox. 1 bark de ancho de banda en la frecuencia de cruce entre baja frecuencia (LF) y alta frecuencia (HF) regenerada con extensión de ancho de banda (BWE) para evitar la posibilidad de disonancia y sustituir el contenido espectral con ceros o ruido.

Sin embargo, la solución propuesta en [19] tiene algunos inconvenientes: En primer lugar, la sustitución estricta de contenido espectral ya sea por ceros o por ruido también puede afectar a la calidad perceptual de la señal. Además, el procesamiento propuesto no se adapta a la señal y, por tanto, puede perjudicar a la calidad perceptual en algunos casos. Por ejemplo, si la señal contiene transitorios, esto puede llevar a pre- y post-ecos.

En segundo lugar, también pueden producirse disonancias en transiciones entre conexiones de alta frecuencia (HF) consecutivas. La solución propuesta en [19] solo es funcional para remediar disonancias que se producen en frecuencias de cruce entre baja frecuencia (LF) y alta frecuencia (HF) regenerada con extensión de ancho de banda (BWE).

Por último, al contrario de los sistemas basados en filtros de bancos como el propuesto en [19], los sistemas de extensión de ancho de banda (BWE) también pueden realizarse en implementaciones basadas en transformada, como, por ejemplo, la Transformada de Coseno Discreta Modificada (MDCT). Las transformadas como la transformada de coseno discreta modificada (MDCT) son muy propensas a los denominados trinos [20] o artefactos de anillo que se producen si se copian regiones de paso de banda de coeficientes espectrales o se ajustan a cero coeficientes espectrales como los propuestos en [19].

En particular, la patente estadounidense 8.412.365 da a conocer el uso, en traslación o desdoblamiento basado en bancos de filtros, las denominadas bandas de guarda que se insertan y se componen de uno o varios canales de sub-banda ajustados a cero. Se usa un número de canales de bancos de filtro como bandas de guarda y un ancho de banda de una banda de guarda debería ser de 0,5 bark. Estas bandas de guarda de disonancia se reconstruyen parcialmente usando señales de ruido blanco aleatorias, es decir, las sub-bandas se alimentan con ruido blanco en lugar de ser cero. Las bandas de guarda se insertan independientemente de la señal actual que va a procesarse.

El documento US 2004/028244 A1 da a conocer un dispositivo decodificador que genera datos espectrales de frecuencia a partir de una corriente de datos de audio codificados introducida e incluye: una unidad decodificadora central para decodificar la corriente de datos codificados introducida y generar datos espectrales de baja frecuencia que representan una señal de audio; y una unidad decodificadora ampliada para generar, basándose en los datos espectrales de baja frecuencia, datos espectrales de frecuencia ampliada que indican una estructura armónica, que es igual que una extensión a lo largo del eje de frecuencia de la estructura armónica indicada por los datos espectrales

de baja frecuencia, en una región de frecuencia que no se representa por la corriente de datos codificados.

El documento US 2011/264454 A1 da a conocer un método para la recuperación espectral en la decodificación espectral de una señal de audio. El método comprende obtener un conjunto inicial de coeficientes espectrales que representan la señal de audio y determinar una frecuencia de transición. La frecuencia de transición se adapta al contenido espectral de la señal de audio. Los agujeros espectrales en el conjunto inicial de coeficientes espectrales por debajo de la frecuencia de transición están rellenos de ruido y el conjunto inicial de coeficientes espectrales está ampliado por ancho de banda por encima de la frecuencia de transición. También se ilustran decodificadores y codificadores que se disponen para realizar parte del o la totalidad del método.

Es el objeto de la presente invención proporcionar un concepto mejorado para decodificar una señal de audio codificada.

Este objeto se logra mediante un aparato para decodificar una señal de audio codificada de acuerdo con la reivindicación 1, un método de decodificación de una señal de audio codificada de acuerdo con la reivindicación 14, o un programa informático de acuerdo con la reivindicación 15.

La presente invención se expone en las reivindicaciones independientes 1, 14 y 15. Todas las siguientes ocurrencias de la palabra "realización (es)", si se refiere a combinaciones de características diferentes de las definidas por las reivindicaciones independientes, se refieren a ejemplos que se presentaron originalmente pero que no representan realizaciones de la invención reivindicada actualmente; estos ejemplos todavía se muestran por motivos ilustrativos solo.

De acuerdo con la presente invención, se realiza un análisis de señal del lado de decodificador usando un analizador para analizar la señal central decodificada o una señal regenerada preliminar obtenida realizando un procedimiento de regeneración de frecuencia preliminar.

Entonces, este resultado de análisis se usa por un regenerador de frecuencia para regenerar porciones espectrales no incluidas en la señal central decodificada.

Por tanto, a diferencia de un ajuste decodificador fijo, donde la interconexión o la colocación de mosaicos de frecuencia se realiza de una manera fija, es decir, donde se toma un determinado rango de origen de la señal central y se aplican determinados bordes de frecuencia fija para o bien fijar la frecuencia entre el rango de origen y el rango de reconstrucción o bien el borde de frecuencia entre dos conexiones o mosaicos de frecuencia adyacentes dentro del rango de reconstrucción, se realiza una interconexión o colocación de mosaicos dependiente de la señal, en la que, por ejemplo, la señal central puede analizarse para encontrar mínimos locales en la señal central y, entonces, el rango central se selecciona de modo que los bordes de frecuencia del rango central coinciden con mínimos locales en el espectro de señal central.

A continuación, se comentan realizaciones preferidas con referencia a los dibujos adjuntos, en los cuales:

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;

La Fig. 2a ilustra una implementación preferida del decodificador;

La Fig. 2b ilustra una implementación preferida del codificador.

La Fig. 3a ilustra una representación esquemática de un espectro generado por el decodificador de dominio espectral de la Fig. 1b.

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.

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.

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.
- La Fig. 5c ilustra una implementación del regenerador de frecuencia.
- 5 La Fig. 6a es un aparato para decodificar una señal de audio codificada de acuerdo con una implementación.
- La Fig. 6b es una realización adicional de un aparato para decodificar una señal de audio codificada.
- La Fig. 7a ilustra una implementación preferida del regenerador de frecuencia de la Fig. 6a o 6b.
- 10 La Fig. 7b ilustra una implementación adicional de una actuación conjunta entre el analizador y el regenerador de frecuencia.
- La Fig. 8 ilustra una implementación adicional del regenerador de frecuencia.
- 15 La Fig. 8b ilustra una implementación adicional de la invención.
- La Fig. 9a ilustra un decodificador con la tecnología de regeneración de frecuencia utilizando valores de potencia para el rango de frecuencia de regeneración.
- 20 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.
- 25 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.
- La Fig. 10b ilustra un diagrama de bloques para ilustrar una funcionalidad adicional de la calculadora de parámetros de la Fig. 10a.
- 30 La Fig. 10c ilustra un diagrama de bloques para ilustrar una funcionalidad adicional de la calculadora de parámetros de la Fig. 10a.
- 35 La Fig. 10d ilustra un diagrama de bloques que ilustra una funcionalidad adicional de la calculadora de parámetros de la Fig. 10a.
- La Fig. 11a ilustra un espectro de un anillo de filtro que rodea un transitorio.
- 40 La Fig. 11b ilustra un espectrograma de un transitorio después de aplicar una extensión de ancho de banda.
- La Fig. 11c ilustra un espectro de un transitorio después de aplicar una extensión de ancho de banda con reducción de anillo de filtro.
- 45 La Fig. 12a ilustra un diagrama de bloques de un aparato para decodificar una señal de audio codificada.
- La Fig. 12b ilustra espectros de magnitud (estilizados) de una señal tonal, un copiado sin adaptación de mosaico/conexión, un copiado con bordes de frecuencia cambiados y una eliminación adicional de porciones tonales de creación de artefactos.
- 50 La Fig. 12c ilustra una función de transición gradual a modo de ejemplo.
- La Fig. 13a ilustra un codificador de la técnica anterior con extensión de ancho de banda.
- 55 La Fig. 13b ilustra un decodificador de la técnica anterior con extensión de ancho de banda.
- La Fig. 14a ilustra un aparato adicional para decodificar una señal de audio codificada usando un filtro de cruce.
- La Fig. 14b ilustra una ilustración más detallada de un filtro de cruce a modo de ejemplo.
- 60 La Fig. 6a ilustra un aparato para decodificar una señal de audio codificada que comprende una señal central codificada y datos paramétricos. El aparato comprende un decodificador central 600 para decodificar la señal central codificada para obtener una señal central decodificada, un analizador 602 para analizar la señal central decodificada antes o después de realizar una operación de regeneración de frecuencia. El analizador 602 se configura para

proporcionar un resultado de análisis 603. El regenerador de frecuencia 604 se configura para regenerar porciones espectrales no incluidas en la señal central decodificada usando una porción espectral de la señal central decodificada, datos de envolvente 605 para las porciones espectrales faltantes y el resultado de análisis 603. Por tanto, a diferencia de implementaciones anteriores, la regeneración de frecuencia no se realiza en el lado de decodificador de manera independiente de la señal, sino que se realiza de manera dependiente de la señal. Esto tiene la ventaja de que, cuando no existen problemas, la regeneración de frecuencia se realiza tal como es, pero cuando existen porciones de señal problemáticas, entonces esto se detecta por el resultado de análisis 603 y el regenerador de frecuencia 604 realiza entonces una forma adaptada de regeneración de frecuencia que puede, por ejemplo, ser el cambio de un borde de frecuencia inicial entre la región central y la banda de reconstrucción o el cambio de un borde de frecuencia entre dos conexiones/mosaicos individuales dentro de la banda de reconstrucción. Al contrario de la implementación de las bandas de guarda, esto tiene la ventaja de que se realizan procedimientos específicos solo cuando se requiere y no, como en la implementación de banda de guarda, todo el tiempo sin ninguna dependencia de señal.

Preferiblemente, el decodificador central 600 se implementa como una etapa de decuantificación y decodificación por entropía (por ejemplo, decodificador aritmético o Huffman) 612 tal como se ilustra en la Fig. 6b. El decodificador central 600 emite entonces un espectro de señal central y el espectro se analiza por el analizador espectral 614 que se implementa, de manera bastante similar al analizador 602 en la Fig. 6a, como un analizador espectral en vez de cualquier analizador arbitrario que podría analizar, tal como se ilustra en la Fig. 6a, también una señal de dominio de tiempo. En la realización de la Fig. 6b, el analizador espectral se configura para analizar la señal espectral de modo que se determinan mínimos locales en la banda de origen y/o en una banda de destino, es decir, en las conexiones de frecuencia o en los mosaicos de frecuencia. Entonces, el regenerador de frecuencia 604 realiza, tal como se ilustra en 616, una regeneración de frecuencia donde los bordes de conexión se colocan en mínimos en la banda de origen y/o la banda de destino.

Posteriormente, se comenta la Fig. 7a con el fin de describir una implementación preferida del regenerador de frecuencia 604 de la Fig. 6a. Un regenerador de señal preliminar 702 recibe, como una entrada, datos de origen desde la banda de origen y, adicionalmente, información de conexión preliminar tal como frecuencias de borde preliminares. Entonces, se genera una señal regenerada preliminar 703, que se detecta por el detector 704 para detectar los componentes tonales dentro de la señal reconstruida preliminar 703. Alternativa o adicionalmente, los datos de origen 705 también pueden analizarse por el detector que corresponde al analizador 602 de la Fig. 6a. Entonces, no sería necesaria la etapa de regeneración de señal preliminar. Cuando hay un mapeo bien definido de los datos de origen a los datos de reconstrucción, entonces pueden detectarse los mínimos o las porciones tonales incluso considerando solo los datos de origen, si hay porciones tonales próximas al borde superior del rango central o en un borde de frecuencia entre dos mosaicos de frecuencia generados de manera individual tal como se comentará posteriormente con respecto a la Fig. 12b.

En el caso de que se hayan descubierto componentes tonales problemáticos cerca de bordes de frecuencia, un regulador de frecuencia de transición 706 realiza un ajuste de una frecuencia de transición tal como una frecuencia de transición o frecuencia de cruce o frecuencia de inicio de relleno de espacios entre la banda central y la banda de reconstrucción o entre porciones de frecuencia individual generadas por uno y los mismos datos de origen en la banda de reconstrucción. La señal de salida de bloque 706 se envía a un eliminador 708 de componentes tonales en los bordes. El eliminador se configura para eliminar componentes tonales restantes que están todavía allí después del ajuste de frecuencia de transición por el bloque 706. El resultado del eliminador 708 se envía entonces a un filtro de cruce 710 con el fin de abordar el problema de anillo de filtro y el resultado del filtro de cruce 710 se introduce entonces en un bloque de modelado de envolvente espectral 712 que realiza un modelado de envolvente espectral en la banda de reconstrucción.

Tal como se comenta en el contexto de la Fig. 7a, la detección de componentes tonales en el bloque 704 puede realizarse tanto en unos datos de origen 705 como en una señal reconstruida preliminar 703. Esta realización se ilustra en la Fig. 7b, donde se crea una señal regenerada preliminar tal como se muestra en el bloque 718. La señal que corresponde a la señal 703 de la Fig. 7a se envía entonces a un detector 720 que detecta componentes de creación de artefacto. Aunque el detector 720 puede configurarse para ser un detector para detectar componentes tonales en bordes de frecuencia tal como se ilustra en 704 en la Fig. 7a, el detector también puede implementarse para detectar otros componentes de creación de artefacto. Tales componentes espectrales pueden ser incluso otros componentes distintos de los componentes tonales y puede realizarse una detección de si un artefacto se ha creado probando diferentes regeneraciones y comparando los diferentes resultados de regeneración con el fin de encontrar el que ha proporcionado componentes de creación de artefacto.

El detector 720 controla ahora un manipulador 722 para manipular la señal, es decir, la señal regenerada preliminar. Esta manipulación puede realizarse procesando de hecho la señal regenerada preliminar por la línea 723 o realizando nuevamente una regeneración, pero ahora con, por ejemplo, las frecuencias de transición reformadas tal como se ilustra por la línea 724.

Una implementación del procedimiento de manipulación es que la frecuencia de transición se ajusta tal como se ilustra

en 706 en la Fig. 7a. Una implementación adicional se ilustra en la Fig. 8a, que puede realizarse en lugar del bloque 706 o junto con el bloque 706 de la Fig. 7a. Se proporciona un detector 802 para detectar frecuencias de inicio y de fin de una porción tonal problemática. Entonces, un interpolador 804 se configura para interpolar y, preferiblemente interpolar de manera compleja entre el inicio y el fin de la porción tonal dentro del rango espectral. Entonces, tal como se ilustra en la Fig. 8a por el bloque 806, la porción tonal se sustituye por el resultado de interpolación.

Una implementación alternativa se ilustra en la Fig. 8a por los bloques 808, 810. En lugar de realizar una interpolación, se realiza una generación aleatoria de líneas espectrales 808 entre el inicio y el fin de la porción tonal. Entonces, se realiza un ajuste de potencia de las líneas espectrales generadas de manera aleatoria tal como se ilustra en 810 y la potencia de las líneas espectrales generadas de manera aleatoria se ajusta de modo que la potencia es similar a las porciones espectrales no tonales adyacentes. Entonces, la porción tonal se sustituye por líneas espectrales generadas de manera aleatoria ajustadas a la envolvente. Las líneas espectrales pueden generarse de manera aleatoria o generarse de manera pseudo aleatoria con el fin de proporcionar una señal de sustitución que, en la medida de lo posible, está libre de artefacto.

Se ilustra una implementación adicional en la Fig. 8b. Se ilustra un generador de mosaico de frecuencia situado dentro del regenerador de frecuencia 604 de la Fig. 6a en el bloque 820. El generador de mosaico de frecuencia usa bordes de frecuencia predeterminados. Entonces, el analizador analiza la señal generada por el generador de mosaico de frecuencia y el generador de mosaico de frecuencia 820 se configura preferiblemente para realizar múltiples operaciones de colocación de mosaicos para generar múltiples mosaicos de frecuencia. Entonces, el manipulador 824 en la Fig. 8b manipula el resultado del generador de mosaico de frecuencia de acuerdo con el resultado de análisis emitido por el analizador 822. La manipulación puede ser el cambio de bordes de frecuencia o la atenuación de porciones individuales. Entonces, un regulador de envolvente espectral 826 realiza un ajuste de envolvente espectral usando la información paramétrica 605 tal como se comentó ya en el contexto de la Fig. 6a.

Entonces, la señal ajustada de manera espectral emitida por el bloque 826 se introduce en un convertidor frecuencia-tiempo que, adicionalmente, recibe las primeras porciones espectrales, es decir, una representación espectral de la señal de salida del decodificador central 600. La salida del convertidor frecuencia-tiempo 828 puede usarse entonces para almacenamiento o para transmitir a un altavoz para renderización de audio.

La presente invención puede aplicarse o bien a procedimientos de regeneración de frecuencia conocidos tales como los ilustrados en las Figs. 13a, 13b o bien puede aplicarse preferiblemente dentro del contexto de relleno inteligente de espacios, que se describe posteriormente con respecto a las Figs. 1a a 5b y 9a a 10d.

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 102 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.

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

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.

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.

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.

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

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

El analizador espectral 226 aplica preferentemente un enmascaramiento de tonalidad. Esta etapa de estimación de enmascaramiento de tonalidad se utiliza para separar los componentes tonales de los componentes 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.

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, $nSrc$ 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 mosaico en el decodificador. El módulo de relleno inteligente de espacios (IGF) también se encarga de no sobrescribir los componentes tonales en el espectro ya que los componentes tonales se conservan utilizando el enmascaramiento de tonalidad. Se utiliza un parámetro de 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 en comparación con ASR (Reemplazo Espectral Preciso) [2-4] es la ausencia de una etapa de síntesis de la señal que crea las porciones importantes de la señal en el decodificador. En cambio, esta tarea es asumida por el codificador central, lo que permite la conservación de los componentes importantes del espectro. Otra ventaja del sistema propuesto es la escalabilidad continua que ofrecen las características. Solo el uso de $tileNum[idx_tar]$ y $xcorr_lag = 0$, para cada 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.

$$NrgMedia[k] = NrgIzquierda[k] + NrgDerecha[k];$$

$$NrgLateral[k] = NrgIzquierda[k] - NrgDerecha[k];$$

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:

$$MosaicoMedio[k] = 0.5 \cdot (MosaicoIzquierdo[k] + MosaicoDerecho[k])$$

$$MosaicoLateral[k] = 0.5 \cdot (MosaicoIzquierdo[k] - MosaicoDerecho[k])$$

Ajuste de potencia:

$$MosaicoMedio[k] = MosaicoMedio[k] * NrgMedia[k];$$

$$MosaicoLateral[k] = MosaicoLateral[k] * NrgLateral[k];$$

Transformación conjunta de estéreo -> LR:

Si no se codifica ningún parámetro de predicción adicional:

$$MosaicoIzquierdo[k] = MosaicoMedio[k] + MosaicoLateral[k]$$

$$MosaicoDerecho[k] = MosaicoMedio[k] - MosaicoLateral[k]$$

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

$$MosaicoLateral[k] = MosaicoLateral[k] - Coefpredicción \cdot MosaicoMedio[k]$$

$$MosaicoIzquierdo[k] = MosaicoMedio[k] + MosaicoLateral[k]$$

$$MosaicoDerecho[k] = MosaicoMedio[k] + MosaicoLateral[k]$$

5 Si la dirección señalada es del lateral al medio:

$$MosaicoMedio[k] = MosaicoLateral[k] - Coefpredicción \cdot MosaicoLateral[k]$$

$$MosaicoIzquierdo[k] = MosaicoMedio[k] + MosaicoLateral[k]$$

10

$$MosaicoDerecho[k] = MosaicoMedio[k] + MosaicoLateral[k]$$

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.

15

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.

20

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.

30

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).

40

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).

50

55

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

60

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.

5 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 .

20 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 sí 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.

45 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.

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.

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

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

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.

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 6 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.

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.

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.

5 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
10 corresponder al tamaño del mosaico de destino (TT) asociado. Esto se demuestra utilizando el siguiente ejemplo. 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 I (SCB tiene una longitud de 7 + 8). Entonces, el mosaico de origen (ST) para lo
15 anterior tiene una longitud de 15 intervalos en lugar de 10 intervalos como para TT[0].

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
20 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.

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
30 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.

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
40 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.

45 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".
50 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
60 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.

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.

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).

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.

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)

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.

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.

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.

5 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
10 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
6, 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.

15 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
20 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
25 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
30 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
35 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
40 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,
45 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.

50 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
55 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
60 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.

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.

La Fig. 12a ilustra una implementación adicional del aparato para decodificar. Se recibe una corriente de bits por un decodificador central 1200 que puede ser, por ejemplo, un decodificador de codificación avanzada de audio AAC. El resultado se configura en una fase para realizar una colocación de mosaicos o interconexión de extensión de ancho de banda 1202 que corresponde al regenerador de frecuencia 604, por ejemplo. Entonces, se realiza un procedimiento de adaptación de mosaico/conexión y procesamiento posterior, y, cuando se ha realizado una adaptación de mosaico, el regenerador de frecuencia 1202 se controla para realizar una regeneración de frecuencia adicional, pero ahora con, por ejemplo, bordes de frecuencia ajustados. Además, cuando se realiza un procesamiento de mosaico tal como por la eliminación o atenuación de líneas tonales, el resultado se envía entonces un bloque 1206 que realiza el modelado de envolvente de ancho de banda basado en parámetros tal como, por ejemplo, también se comenta en el contexto del bloque 712 o 826. El resultado se envía entonces a un bloque de transformada de síntesis 1208 para realizar una transformada en el dominio de salida final que es, por ejemplo, un dominio de salida de modulación de impulsos codificados (PCM) tal como se ilustra en la Fig. 12a.

Las principales características de realizaciones de la invención son las siguientes:

La realización preferida se basa en la transformada de coseno discreta modificada (MDCT) que muestra los artefactos de trino mencionados anteriormente si se reducen las zonas espectrales tonales por la elección desafortunada de márgenes de mosaico y/o frecuencia de cruce o los componentes tonales consiguen colocarse en las inmediaciones en bordes de conexión.

La Fig. 12b muestra cómo la técnica propuesta nuevamente reduce los artefactos encontrados en métodos de extensión de ancho de banda (BWE) de la técnica anterior. En el panel (2) de la Fig. 12, se muestra el espectro de

magnitud estilizado de la salida de un método de extensión de ancho de banda (BWE) actual. En este ejemplo, la señal se ve afectada perceptualmente por el batido de frecuencias provocado por dos tonos cercanos y también por la división de un tono. Ambas zonas espectrales problemáticas están marcadas, cada una, con un círculo.

5 Para superar estos problemas, la nueva técnica detecta en primer lugar la ubicación espectral de los componentes tonales contenidos en la señal. Entonces, de acuerdo con un aspecto de la invención, se intenta ajustar las frecuencias de transición entre baja frecuencia (LF) y todas las conexiones por movimientos individuales (dentro de límites dados) de manera que se minimiza la división o el batido de frecuencias de componentes tonales. Por ese motivo, la frecuencia de transición tiene que coincidir preferiblemente con un mínimo espectral local. Esta etapa se muestra en el panel (2) y el panel (3) de la Fig. 12b, donde la frecuencia de transición f_{x2} se mueve hacia frecuencias más altas, dando como resultado f'_{x2} .

15 De acuerdo con otro aspecto de la invención, si permanece el contenido espectral problemático en las regiones de transición, se elimina al menos uno de los componentes tonales mal colocados para reducir o bien el artefacto de batido de frecuencias en las frecuencias de transmisión o bien el trino. Esto se hace mediante filtrado/interpolación o extrapolación espectral, tal como se muestra en el panel (3) de la figura 2. De ese modo, se elimina un componente tonal de punto-pie a punto-pie, es decir, desde su mínimo local izquierdo hasta su mínimo local derecho. El espectro resultante después de la aplicación de la tecnología de la invención se muestra en el panel (4) de la Fig. 12b.

20 Dicho de otro modo, la Fig. 12b ilustra, en la esquina superior izquierda, es decir, en el panel (1), la señal original. En la esquina superior derecha, es decir, en el panel (2), se muestra una señal extendida de ancho de banda de comparación con zonas problemáticas marcadas por las elipses 1220 y 1221. En la esquina inferior izquierda, es decir, en el panel (3), se ilustran dos características de procesamiento de mosaico de frecuencia o conexión preferidas. La división de porciones tonales se ha abordado aumentando el borde de frecuencia f'_{x2} de modo que ya no hay un recorte de la porción tonal correspondiente. Además, se aplican funciones de ganancia 1030 para eliminar la porción tonal 1031 y 1032 o, alternativamente, se indica una interpolación ilustrada por 1033. Finalmente, la esquina inferior derecha de la Fig. 12b, es decir, el panel (4) representa la señal mejorada que resulta de una combinación de frecuencia de conexión/mosaico que se ajusta por una parte y eliminación o al menos atenuación de porciones tonales problemáticas.

30 El panel (1) de la Fig. 12b ilustra, tal como se comenta anteriormente, el espectro original y el espectro original tiene un rango de frecuencia central hasta la frecuencia de inicio de relleno de espacios o de cruce f_{x1} .

35 Por tanto, una frecuencia f_{x1} ilustra una frecuencia de borde 1250 entre el rango de origen 1252 y un rango de reconstrucción 1254 que se extiende entre la frecuencia de borde 1250 y una frecuencia máxima que es más pequeña que o igual a la frecuencia de Nyquist $f_{Nyquist}$. En el lado de codificador, se asume que una señal está limitada por ancho de banda en f_{x1} o, cuando se aplica la tecnología relacionada con el relleno inteligente de espacios, se asume que f_{x1} corresponde con la frecuencia de inicio de relleno de espacios 309 de la Fig. 3a. Dependiendo de la tecnología, el rango de reconstrucción anterior f_{x1} estará vacío (en caso de la implementación de la Fig. 13a, 13b) o comprenderá determinadas primeras porciones espectrales que van a codificarse con una alta resolución tal como se comenta en el contexto de la Fig. 3a.

45 El panel (2) de la Fig. 12b ilustra una señal regenerada preliminar, por ejemplo, generada por el bloque 702 de la Fig. 7a que tiene dos porciones problemáticas. Una porción problemática se ilustra en 1220. La distancia de frecuencia entre la porción tonal dentro de la región central ilustrada en 1220a y la porción tonal en el inicio del mosaico de frecuencia ilustrada en 1220b es demasiado pequeña de modo que se crearía un artefacto de batido de frecuencias. El problema adicional es que en el borde superior del primer mosaico de frecuencia generado por la primera operación de interconexión u operación de colocación de mosaicos de frecuencia ilustrada en 1225 es una porción tonal dividida o recortada a la mitad 1226. Cuando esta porción tonal 1226 se compara con las otras porciones tonales en la Fig. 12b, se vuelve evidente que el ancho es más pequeño que el ancho de una porción tonal normal y esto significa que esta porción tonal se ha dividido ajustando el borde de frecuencia entre el primer mosaico de frecuencia 1225 y el segundo mosaico de frecuencia 1227 en el lugar incorrecto en el rango de origen 1252. Con el fin de abordar este problema, la frecuencia de borde f_{x2} se ha modificado para volverse un poco mayor tal como se ilustra en el panel (3) en la Fig. 12b, de modo que no se produce un recorte de esta porción tonal.

55 Por otra parte, este procedimiento, en el que se ha cambiado f'_{x2} , no aborda con eficiencia el problema de batido de frecuencias que, por tanto, se aborda por una eliminación de los componentes tonales por filtrado o interpolación o cualquier otro procedimiento tal como se comenta en el contexto del bloque 708 de la Fig. 7a. Por tanto, la Fig. 12b ilustra una aplicación secuencial del ajuste de frecuencia de transición 706 y la eliminación de componentes tonales en bordes ilustrados en 708.

Otra opción hubiera sido ajustar el borde de transición f_{x1} de modo que es un poco más bajo de modo que la porción tonal 1220a ya no está en el intervalo central. Entonces, la porción tonal 1220a también se ha retirado o eliminado ajustando la frecuencia de transición f_{x1} en un valor inferior.

Este procedimiento también hubiera funcionado para abordar el problema con el componente tonal problemático 1032. Ajustando f'_{x2} incluso más alta, la porción espectral donde se sitúa la porción tonal 1032 podría haberse regenerado dentro de la primera operación de interconexión 1225 y, por tanto, no se hubieran producido dos porciones tonales vecinas o adyacentes.

Básicamente, el problema de batido de frecuencias depende de las amplitudes y la distancia en la frecuencia de porciones tonales adyacentes. El detector 704, 720, o denominado de manera más general, el analizador 602 se configura preferiblemente de tal manera que se analiza un análisis de la porción espectral inferior situada en la frecuencia por debajo de la frecuencia de transición tal como f_{x1} , f_{x2} , f'_{x2} con el fin de localizar cualquier componente tonal. Además, el rango espectral por encima de la frecuencia de transición también se analiza con el fin de detectar un componente tonal. Cuando la detección da como resultado dos componentes tonales, uno a la izquierda de la frecuencia de transición con respecto a la frecuencia y uno a la derecha (con respecto a la frecuencia ascendente), entonces se activa el eliminador de componentes tonales en los bordes ilustrados en 708 en la Fig. 7a. La detección de componentes tonales se realiza en un determinado rango de detección que se extiende, desde la frecuencia de transición, en ambas direcciones al menos un 20% con respecto al ancho de banda de la banda correspondiente y preferiblemente solo se extiende hasta un 10% hacia abajo a la izquierda de la frecuencia de transición y hacia arriba a la derecha de la frecuencia de transición relacionada con el ancho de banda correspondiente, es decir, el ancho de banda del rango de origen por una parte y el rango de reconstrucción por otra parte o, cuando la frecuencia de transición es la frecuencia de transición entre dos mosaicos de frecuencia 1225, 1227, una cantidad correspondiente a un 10% del mosaico de frecuencia correspondiente. En una realización adicional, el ancho de banda de detección predeterminado es un bark. Debería ser posible eliminar porciones tonales dentro de un rango de 1 bark alrededor de un borde de conexión, de modo que el rango de detección completo es de 2 bark, es decir, un bark en la banda inferior y un bark en la banda superior, donde el un bark en la banda inferior es inmediatamente adyacente a un bark en la banda superior.

De acuerdo con otro aspecto de la invención, para reducir el artefacto de anillo de filtro, se aplica un filtro de cruce en el dominio de frecuencia a dos regiones espectrales consecutivas, es decir, entre la banda central y la primera conexión o entre dos conexiones. Preferiblemente, el filtro de cruce es adaptable a la señal.

El filtro de cruce consiste en dos filtros, un filtro de fundido de salida h_{out} , que se aplica a la región espectral inferior, y un filtro de fundido de entrada h_{in} , que se aplica a la región espectral superior.

Cada uno de los filtros tiene una longitud N .

Además, la inclinación de ambos filtros se caracteriza por un valor adaptable a la señal llamado $Xbias$ que determina la característica de muesca del filtro de cruce, con $0 \leq Xbias \leq N$:

Si $Xbias = 0$, entonces la suma de ambos filtros es igual a 1, es decir, no hay características de filtro de muesca en el filtro resultante.

Si $Xbias = N$, entonces ambos filtros son completamente cero.

El diseño básico de los filtros de cruce se limita a las siguientes ecuaciones:

$$h_{out}(k) = h_{in}(N - 1 - k), \forall Xbias$$

$$h_{out}(k) + h_{in}(k) = 1, Xbias = 0$$

siendo $k = 0, 1, \dots, N - 1$ el índice de frecuencia. La Fig. 12c muestra un ejemplo de un filtro de cruce de este tipo.

En este ejemplo, la siguiente ecuación se usa para crear el filtro h_{out} :

$$h_{out}(k) = 0.5 + 0.5 \cdot \cos\left(\frac{k}{N - 1 - Xbias} \cdot \pi\right), k = 0, 1, \dots, N - 1 - Xbias$$

La siguiente ecuación describe cómo se aplican entonces los filtros h_{in} y h_{out} ,

$$Y(k_t - (N - 1) + k) = LF(k_t - (N - 1) + k) \cdot h_{out}(k) + HF(k_t - (N - 1) + k) \cdot h_{in}(k), \quad k = 0, 1, \dots, N - 1$$

indicando Y el espectro ensamblado, siendo k_t la frecuencia de transición, siendo LF el contenido de baja frecuencia y siendo HF el contenido de alta frecuencia.

A continuación, se presentarán pruebas del beneficio de esta técnica. La señal original en los siguientes ejemplos es una señal de tipo transitorio, en particular una versión filtrada de paso bajo de la misma, con una frecuencia de corte de 22 kHz. En primer lugar, este transitorio está limitado por banda a 6 kHz en el dominio de transformada. Posteriormente, el ancho de banda de la señal original filtrada de paso bajo se extiende a 24 kHz. La extensión de ancho de banda se consigue copiando la banda de baja frecuencia (LF) tres veces para rellenar por completo el rango de frecuencia que está disponible por encima de 6 kHz dentro de la transformada.

La Fig. 11a muestra el espectro de esta señal, que puede considerarse un espectro normal de un artefacto de anillo de filtro que rodea de manera espectral el transitorio debido a dicha característica de pared de ladrillo de la transformada (picos de voz 1100). Aplicando el enfoque de la invención, se reduce el anillo de filtro en aprox. 20 dB en cada frecuencia de transición (picos de voz reducidos).

El mismo efecto, aunque en una ilustración diferente, se muestra en la Fig. 11b, 11c. La Fig. 11b muestra el espectrograma de la señal de tipo transitorio mencionada con el artefacto de anillo de filtro que precede y sucede temporalmente el transitorio después de aplicar la técnica de extensión de ancho de banda (BWE) descrita anteriormente sin ninguna reducción de anillo de filtro. Cada una de las líneas horizontales representa el anillo de filtro en la frecuencia de transición entre conexiones consecutivas. La figura 6 muestra la misma señal después de aplicar el enfoque de la invención dentro de la extensión de ancho de banda (BWE). A través de la aplicación de la reducción de anillo, el anillo de filtro se reduce en aprox. 20 dB en comparación con la señal visualizada en la figura anterior.

Posteriormente, se comentan las Figs. 14a, 14b con el fin de ilustrar adicionalmente el aspecto de la invención de filtro de cruce ya comentado en el contexto con la característica de analizador. Sin embargo, el filtro de cruce 710 también puede implementarse de manera independiente de la invención comentada en el contexto de Figs. 6a-7b.

La Fig. 14a ilustra un aparato para decodificar una señal de audio codificada que comprende una señal central codificada e información sobre datos paramétricos. El aparato comprende un decodificador central 1400 para decodificar la señal central codificada para obtener una señal central decodificada. La señal central decodificada puede estar limitada por ancho de banda en el contexto de la implementación de la Fig. 13a, la Fig. 13b o el decodificador central puede ser un codificador de tasa completa o rango de frecuencia completo en el contexto de las Figs. 1 a 5c o 9a-10d.

Además, se genera un generador de mosaico 1404 para regenerar uno o más mosaicos espectrales que tienen frecuencias no incluidas en la señal central decodificada usando una porción espectral de la señal central decodificada. Los mosaicos pueden reconstruir segundas porciones espectrales dentro de una banda de reconstrucción tal como, por ejemplo, se ilustra en el contexto de la Fig. 3a o que puede incluir primeras porciones espectrales que van a reconstruirse con una alta resolución pero, alternativamente, los mosaicos espectrales también pueden comprender bandas de frecuencia completamente vacías cuando el codificador ha realizado una limitación de banda dura tal como se ilustra en la Fig. 13a.

Además, se proporciona un filtro de cruce 1406 para filtrado de cruce espectral de la señal central decodificada y un primer mosaico de frecuencia que tiene frecuencias que se extienden desde una frecuencia de relleno de espacios 309 hasta una primera frecuencia de fin de mosaico o para filtrado de cruce espectral de un primer mosaico de frecuencia 1225 y un segundo mosaico de frecuencia 1221, teniendo el segundo mosaico de frecuencia una frecuencia de borde inferior que es adyacente a la frecuencia a una frecuencia de borde superior del primer mosaico de frecuencia 1225.

En una implementación adicional, la señal de salida de filtro de cruce 1406 se alimenta a un regulador de envolvente 1408 que aplica información de envolvente espectral paramétrica incluida en una señal de audio codificada como información de lado paramétrica para obtener finalmente una señal regenerada ajustada a la envolvente. Pueden implementarse elementos 1404, 1406, 1408 como un regenerador de frecuencia tal como, por ejemplo, se ilustra en la Fig. 13b, la Fig. 1b o la Fig. 6a, por ejemplo.

La Fig. 14b ilustra una implementación adicional del filtro de cruce 1406. El filtro de cruce 1406 comprende un sub-filtro de fundido de salida que recibe una primera señal de entrada IN1 y un segundo sub-filtro de fundido de entrada 1422 que recibe una segunda entrada IN2 y los resultados o salidas de ambos filtros 1420 y 1422 se proporcionan a un combinador 1424 que es, por ejemplo, un mezclador. El mezclador o el combinador 1424 emite los valores espectrales para los intervalos de frecuencia. La Fig. 12c ilustra una función de transición gradual a modo de ejemplo que comprende la característica de sub-filtro de fundido de salida 1420a y la característica de sub-filtro de fundido de entrada 1422a. Ambos filtros tienen un determinado solapamiento de frecuencia en el ejemplo en la Fig. 12c igual a 21, es decir, $N=21$. Por tanto, no se ven influidos otros valores de frecuencia de, por ejemplo, la región de origen 1252. Solo los intervalos de frecuencia más altos 21 del rango de origen 1252 se ven influidos por la función de fundido de salida 1420a.

Por otra parte, solo las líneas de frecuencia más bajas 21 del primer mosaico de frecuencia 1225 se ven influidas por

la función de fundido de entrada 1422a.

Adicionalmente, se vuelve evidente a partir de las funciones de transición gradual que las líneas de frecuencia entre 9 y 13 se ven influidas, pero la función de fundido de entrada de hecho no influye a las líneas de frecuencia entre 1 y 9 y la función de fundido de salida 1420a no influye a las líneas de frecuencia entre 13 y 21. Esto significa que solo sería necesario un solapamiento entre las líneas de frecuencia 9 y 13 y la frecuencia de cruce tal como f_{x1} se situaría en la muestra de frecuencia o el intervalo de frecuencia 11. Por tanto, solo se requeriría un solapamiento de dos intervalos de frecuencia o valores de frecuencia entre el rango de origen y el primer mosaico de frecuencia con el fin de implementar la función de transición gradual o de cruce.

Dependiendo de la implementación específica, puede aplicarse un solapamiento inferior o superior y, adicionalmente, pueden usarse otras funciones de fundido aparte de una función de coseno. Además, tal como se ilustra en la Fig. 12c, se prefiere aplicar una determinada muesca en el rango de cruce. Dicho de otra manera, la potencia en los rangos de borde se reducirá debido al hecho de que ambas funciones de filtro no añaden a la unidad tal como sería el caso en una función de transición gradual libre de muesca. Esta pérdida de potencia para los bordes del mosaico de frecuencia, es decir, el primer mosaico de frecuencia se atenuará en el borde inferior y en el borde superior, las potencias se concentraron más en la mitad de las bandas. Debido al hecho de que, sin embargo, el ajuste de envolvente espectral tiene lugar después del procesamiento por el filtro de cruce, la frecuencia general no se toca, pero se define por los datos de envolvente espectral tal como los factores de escala correspondientes tal como se comenta en el contexto de la Fig. 3a. Dicho de otro modo, la calculadora 918 de la Fig. 9b calcularía entonces el "rango de destino en bruto ya generado", que es la salida del filtro de cruce. Además, la pérdida de potencia debido a la eliminación de una porción tonal por interpolación también se compensaría debido al hecho de que esta eliminación resulta entonces en una potencia de mosaico más baja y el factor de ganancia para la banda de reconstrucción completa se volverá más alto. Por otra parte, sin embargo, la frecuencia de cruce resulta en una concentración de potencia más en la mitad de un mosaico de frecuencia y esto, al final, reduce con eficiencia los artefactos, en particular, provocados por transitorios tal como se comenta en el contexto de Figs. 11a-11c.

La Fig. 14b ilustra diferentes combinaciones de entrada. Para un filtrado en el borde entre el rango de frecuencia de origen y el mosaico de frecuencia, la entrada 1 es la porción espectral superior del rango central y la entrada 2 es la porción espectral inferior del primer mosaico de frecuencia o del único mosaico de frecuencia, cuando solo existe un único mosaico de frecuencia. Además, la entrada puede ser el primer mosaico de frecuencia y la frecuencia de transición puede ser el borde de frecuencia superior del primer mosaico y la entrada en el sub-filtro 1422 será la porción inferior del segundo mosaico de frecuencia. Cuando existe un tercer mosaico de frecuencia adicional, entonces una frecuencia de transición adicional será el borde de frecuencia entre el segundo mosaico de frecuencia y el tercer mosaico de frecuencia y la entrada en el sub-filtro de fundido de salida 1421 será el rango espectral superior del segundo mosaico de frecuencia tal como se determina por un parámetro de filtro, cuando se usa la característica de la Fig. 12c y la entrada en el sub-filtro de fundido de entrada 1422 será la porción inferior del tercer mosaico de frecuencia y, en el ejemplo de la Fig. 12c, las líneas espectrales inferiores 21.

Tal como se ilustra en la Fig. 12c, se prefiere tener el parámetro N igual para el sub-filtro de fundido de salida y el sub-filtro de fundido de entrada. Sin embargo, esto no es necesario. Los valores para N pueden variar y el resultado será entonces que la "muesca" de filtro será asimétrica entre el rango superior y el inferior. Adicionalmente, las funciones de fundido de salida/entrada no tienen que estar necesariamente en la misma característica que en la Fig. 12c. En su lugar, también pueden usarse características asimétricas.

Además, se prefiere hacer que la característica del filtro de cruce sea adaptable a la señal. Por tanto, basándose en un análisis de señal, se adapta la característica de filtro. Debido al hecho de que el filtro de cruce es particularmente útil para señales de transitorios, se detecta si se producen señales de transitorios. Cuando se producen señales de transitorios, entonces podría usarse una característica de filtro tal como se ilustra en la Fig. 12c. Cuando se detecta, sin embargo, una señal de no transitorios, se prefiere cambiar la característica de filtro para reducir la influencia del filtro de cruce. Esto podría obtenerse, por ejemplo, ajustando N a cero o ajustando X_{bias} a cero de modo que la suma de ambos filtros es igual a 1, es decir, no hay característica de filtro de muesca en el filtro resultante. Alternativamente, el filtro de cruce 1406 podría simplemente omitirse en caso de señales de no transitorios. Preferiblemente, sin embargo, se prefiere una característica de filtro de cambio relativamente bajo cambiando parámetros N, X_{bias} con el fin de evitar artefactos obtenidos por las características de filtro de cambio rápido. Además, se prefiere un filtro de paso bajo para permitir solo tales cambios de característica de filtro relativamente pequeño aunque la señal está cambiando más rápidamente a medida que se detecta por un determinado detector de tonalidad/transitorio. El detector se ilustra en 1405 en la Fig. 14a. Puede recibirse una señal de entrada en un generador de mosaico o una señal de salida del generador de mosaico 1404 o puede incluso conectarse al decodificador central 1400 con el fin de obtener una información de transitorio/no transitorio tal como una indicación de bloque corto desde la decodificación AAC, por ejemplo. Naturalmente, también puede usarse cualquier otro filtro de cruce diferente del mostrado en la Fig. 12c.

Entonces, basándose en la detección de transitorios o basándose en una detección de tonalidad o basándose en cualquier otra detección de característica de señal, se cambia la característica de filtro de cruce 1406 tal como se

comenta.

5 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.

15 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.

20 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.

25 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.

30 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.

35 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.

40 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.

45 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.

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.

50 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.

55 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.

60 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.

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.

5

Lista de Referencias

[1] M. Dietz, L. Liljeryd, K. Kjörling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112^o AES Convention, Múnich, mayo de 2002.

10 [2] Ferreira, D. Sinha, "Accurate Spectral Replacement", Audio Engineering Society Convention, Barcelona, España 2005.

[3] D. Sinha, A. Ferreira¹ and E. Harinarayanan, "A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)", 15 Audio Engineering Society Convention, Paris, Francia 2006.

[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.

20 [5] T. Żernicki, M. Bartkowiak, "Audio bandwidth extension by frequency scaling of sinusoidal partials", Audio Engineering Society Convention, San Francisco, EE.UU. 2008.

[6] J. Herre, D. Schulz, Extending the MPEG-4 AAC Codec by Perceptual Noise Substitution, 104th AES 25 Convention, Amsterdam, 1998, Preprint 4720.

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

30 [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 35 sinusoidal representation", Proceedings of the International Computer Music Conference, 1987.

[10] Purnhagen, H.; Meine, Nikolaus, "HILN-the MPEG-4 parametric audio coding tools," *Circuits and Systems, 2000*. Proceedings. ISCAS 2000 Ginebra. *The 2000 IEEE International Symposium on*, vol.3, no., pp.201, 204 vol.3, 2000

40 [11] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio", Ginebra, 1998.

[12] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, 45 Oikawa: "MPEG-2 Advanced Audio Coding", 101st AES Convention, Los Angeles 1996

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

[14] J. Herre, "Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial 50 introduction", 17th AES International Conference on High Quality Audio Coding, Agosto de 1999

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

[16] International Standard ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 55 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

60 [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

[19] Liljeryd, Lars; Ekstrand, Per; Henn, Fredrik; Kjörling, Kristofer: Spectral translation/folding in the subband domain, patente estadounidense 8.412.365, 2 de abril de 2013.

[20] Daudet, L.; Sandler, M.; "MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction," Speech and Audio Processing, IEEE Transactions on , vol. 12, n.º 3, págs. 302- 312, mayo de 2004.

REIVINDICACIONES

1. Aparato para decodificar una señal de audio codificada que comprende una señal central codificada y datos paramétricos (605), que comprende:
 - 5 un decodificador central (600) para decodificar la señal central codificada para obtener una señal central decodificada;
 - 10 un analizador (602, 614, 704, 720) para analizar la señal central decodificada o una señal regenerada preliminar (703) obtenida realizando un procedimiento de regeneración de frecuencia preliminar (702, 718) para proporcionar un resultado de análisis (603); y
 - 15 un regenerador de frecuencia (604) para regenerar porciones espectrales no incluidas en la señal central decodificada usando una o más porciones espectrales de la señal central decodificada, los datos paramétricos (605) y el resultado de análisis (603),
 - 20 en el que el analizador (602) se configura para detectar una división de una porción espectral de pico en la una o más porciones espectrales de la señal central decodificada o en la señal regenerada preliminar (703) en un borde de frecuencia de la señal central decodificada o en un borde de frecuencia entre dos porciones espectrales regeneradas en la señal regenerada preliminar (703) generada usando la misma o porciones espectrales diferentes de la señal central decodificada o en un borde de frecuencia máxima de la señal regenerada preliminar (703),
 - 25 en el que el regenerador de frecuencia (604) se configura para realizar el procedimiento de regeneración preliminar (702, 718),
 - 30 en el que el regenerador de frecuencia (604) se configura para cambiar (706, 722) el borde de frecuencia entre la señal central decodificada y la señal regenerada preliminar (703) o para cambiar el borde de frecuencia entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o para cambiar el borde de frecuencia máxima, y
 - 35 en el que el regenerador de frecuencia (604) se configura para realizar una reconstrucción con el borde de frecuencia cambiado entre la señal central decodificada y la señal regenerada preliminar (703) o con el borde de frecuencia cambiado entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o con el borde de frecuencia máxima cambiado, de modo que la división se reduce o se elimina.
2. Aparato de la reivindicación 1,
 - 40 en el que el analizador (614) se configura para analizar la señal central decodificada para situar uno o más mínimos espectrales locales en la señal central decodificada,
 - 45 en el que el regenerador de frecuencia (604) se configura para ajustar el borde de frecuencia cambiado entre la señal central decodificada y la señal regenerada preliminar (703) o el borde de frecuencia cambiado entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o el borde de frecuencia máxima cambiado en el uno o más mínimos espectrales locales.
- 50 3. Aparato de acuerdo con la reivindicación 1 o 2,
 - 55 en el que el regenerador de frecuencia (604) comprende además un manipulador de señal (722, 824) para realizar el cambio del borde de frecuencia entre la señal central decodificada y la señal regenerada preliminar (703) o para realizar el cambio del borde de frecuencia entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o para realizar el cambio del borde de frecuencia máxima para obtener datos de control manipulados (724), y
 - 60 en el que el regenerador de frecuencia (604) se configura, generando nuevamente una señal regenerada en la reconstrucción con el borde de frecuencia cambiado, para usar los datos de control manipulados que son diferentes de los datos de control usados para generar la señal regenerada preliminar (703).
4. Aparato de la reivindicación 3,
 - en el que el regenerador de frecuencia (604) comprende un generador de mosaico de frecuencia (820) que

se configura para derivar las porciones espectrales de la señal regenerada preliminar (703) usando la una o más porciones espectrales de la señal central decodificada,

5 en el que el manipulador de señal (722, 824) se configura para manipular el generador de mosaico de frecuencia (820) para obtener porciones espectrales generadas nuevamente por la reconstrucción con el borde de frecuencia cambiado, y

10 en el que el regenerador de frecuencia (604) comprende además un regulador de envolvente espectral (826) configurado para realizar un ajuste de envolvente de las porciones espectrales generadas nuevamente usando los datos paramétricos (605).

5. Aparato de acuerdo con una de las reivindicaciones 3 o 4,

15 en el que el analizador (602) se configura para detectar porciones de señal tonal situadas en un rango de detección de frecuencia, extendiéndose el rango de detección de frecuencia desde un borde de frecuencia de un rango de reconstrucción o entre mosaicos de frecuencia adyacentes dentro del rango de reconstrucción

20 por un ancho de banda de detección predeterminado que es de menos del 20% del ancho de banda de un rango de frecuencia de origen, o

por un ancho de banda de detección predeterminado que es de menos del 20% del ancho de banda del rango de frecuencia de reconstrucción, o

25 por un ancho de banda de detección predeterminado que es de un bark.

6. Aparato de la reivindicación 5,

30 en el que el manipulador de señal (722, 824) se configura para atenuar o eliminar (708) partes espectrales que comprenden porciones tonales, en el que las porciones tonales se sitúan, en una señal regenerada obtenida por la reconstrucción con el borde de frecuencia cambiado, dentro del ancho de banda de detección predeterminado.

7. Aparato de la reivindicación 6, en el que el manipulador de señal (722, 824) se configura

35 para determinar (802) una porción espectral de inicio situada en frecuencia en una frecuencia de inicio de las porciones tonales y una porción espectral de fin situada en frecuencia en una frecuencia de fin de la porción tonal,

40 para interpolar (804) entre la frecuencia de inicio y la frecuencia de fin para obtener una porción de señal interpolada, y

para sustituir (806) la porción tonal entre la frecuencia de inicio y la frecuencia de fin por la porción de señal interpolada.

45 8. Aparato de la reivindicación 6,

50 en el que el manipulador de señal (722, 824) se configura para generar de manera aleatoria o no aleatoria (808) líneas espectrales con una potencia determinada (810) por una porción de señal no tonal de la señal central decodificada o una porción de señal no tonal de porciones espectrales regeneradas obtenidas por la reconstrucción con el borde de frecuencia cambiado.

9. Aparato de acuerdo con la reivindicación 3,

55 en el que el manipulador de señal (722, 824) se configura para controlar un generador de mosaico de frecuencia (820), en el que el generador de mosaico de frecuencia (820) se configura para cambiar el borde de frecuencia entre la señal central decodificada y la señal regenerada preliminar (703) o para cambiar el borde de frecuencia entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o para cambiar el borde de frecuencia máxima.

60 10. Aparato de acuerdo con la reivindicación 1, en el que el regenerador de frecuencia (604) comprende un generador de mosaico de frecuencia (820),

en el que el generador de mosaico de frecuencia (820) se configura para generar un primer mosaico de frecuencia para una primera porción espectral para obtener una primera porción espectral regenerada de las

dos porciones espectrales regeneradas en la señal regenerada preliminar (703), y para generar un segundo mosaico de frecuencia para una segunda porción espectral para obtener una segunda porción espectral regenerada de las dos porciones espectrales regeneradas en la señal regenerada preliminar (703),

5 en el que un borde de frecuencia inferior del segundo mosaico de frecuencia coincide con un borde de frecuencia superior del primer mosaico de frecuencia,

en el que el analizador (602) se configura para detectar, si la porción espectral de pico

10 se divide por el borde de frecuencia inferior del segundo mosaico de frecuencia, o

se divide por el borde de frecuencia superior del primer mosaico de frecuencia, o

15 se divide por un borde de frecuencia inferior del primer mosaico de frecuencia, o

se divide por una frecuencia de inicio de relleno de espacios predeterminada (309) de la señal central decodificada, y

20 en el que un manipulador de señal (722, 824) se configura para controlar el generador de mosaico de frecuencia (820) de modo que el generador de mosaico de frecuencia (820) genera mosaicos de frecuencia modificados que tienen bordes de frecuencia de fin o de inicio modificados, en el que un borde de frecuencia de fin o de inicio modificado representa de manera correspondiente el borde de frecuencia cambiado entre la señal central decodificada y la señal regenerada preliminar (703) o el borde de frecuencia cambiado entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o el borde de frecuencia máxima cambiado.

25

11. Aparato de acuerdo con una de las reivindicaciones anteriores,

30 en el que el decodificador central (600) se configura para obtener, como la señal central decodificada, porciones espectrales decodificadas de dominio de frecuencia que comprenden un primer conjunto de primeras porciones espectrales que se representan por valores espectrales diferentes de una representación cero para valores espectrales y en el que un segundo conjunto de segundas porciones espectrales se representa por la representación cero para valores espectrales,

35 en el que se proporcionan los datos paramétricos (605) para el segundo conjunto de segundas porciones espectrales,

40 en el que el regenerador de frecuencia (604) se configura para usar porciones espectrales del primer conjunto de primeras porciones espectrales para realizar la reconstrucción con el borde de frecuencia cambiado para obtener porciones espectrales regeneradas nuevamente, y

45 en el que el aparato comprende además un convertidor frecuencia-tiempo (828) para convertir las porciones espectrales regeneradas nuevamente y el primer conjunto de primeras porciones espectrales en una representación de tiempo.

12. Aparato de acuerdo con la reivindicación 11,

50 en el que el decodificador central (600) se configura para emitir valores espectrales de transformada de coseno discreta modificada (MDCT), y

en el que el convertidor frecuencia-tiempo (828) comprende un procesador para realizar una transformada MDCT inversa (512, 514, 516) aplicando un procesamiento de solapamiento/adición a cuadros MDCT obtenidos posteriormente.

55 13. Aparato de acuerdo con una de las reivindicaciones anteriores,

en el que el analizador (602, 614, 704, 720) se configura para detectar (704) componentes tonales en la señal regenerada preliminar (703),

60 en el que el regenerador de frecuencia (604) se configura además para eliminar (708) componentes tonales situados en un rango de detección alrededor de los bordes de frecuencia cambiados en porciones espectrales regeneradas nuevamente obtenidas por la reconstrucción con los bordes de frecuencia cambiados;

en el que el regenerador de frecuencia (604) comprende además un filtro de cruce (710) para filtrado de cruce

de una señal con componentes tonales eliminados en un rango de cruce alrededor de los bordes de frecuencia cambiados; y

5 en el que el regenerador de frecuencia (604) comprende además un modelador de envolvente espectral (712) para envolvente espectral que modela un resultado del filtrado de cruce usando los datos paramétricos (605).

14. Método de decodificación de una señal de audio codificada que comprende una señal central codificada y datos paramétricos (605), que comprende:

10 decodificar (600) la señal central codificada para obtener una señal central decodificada;

15 analizar (602, 614, 704, 720) la señal central decodificada o una señal regenerada preliminar (703) obtenida realizando un procedimiento de regeneración de frecuencia preliminar (702, 718) para proporcionar un resultado de análisis (603); y

regenerar (604) porciones espectrales no incluidas en la señal central decodificada usando una o más porciones espectrales de la señal central decodificada, los datos paramétricos (605), y el resultado de análisis (603),

20 en el que el análisis (602) comprende detectar una división de una porción espectral de pico en la una o más porciones espectrales de la señal central decodificada o en la señal regenerada preliminar (703) en un borde de frecuencia de la señal central decodificada o en un borde de frecuencia entre dos porciones espectrales regeneradas en la señal regenerada preliminar (703) generada usando la misma o porciones espectrales diferentes de la señal central decodificada o en un borde de frecuencia máxima de la señal regenerada preliminar (703),

25 en el que la regeneración (604) comprende cambiar el borde de frecuencia entre la señal central decodificada y la señal regenerada preliminar (703) o cambiar el borde de frecuencia entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o cambiar el borde de frecuencia máxima, y

30 en el que la regeneración (604) comprende realizar una reconstrucción con el borde de frecuencia cambiado entre la señal central decodificada y la señal regenerada preliminar (703) o con el borde de frecuencia cambiado entre las dos porciones espectrales regeneradas en la señal regenerada preliminar (703) o con el borde de frecuencia máxima cambiado, de modo que la división se reduce o se elimina.

35 15. Programa informático que comprende instrucciones, que, cuando el programa informático se ejecuta por un ordenador o un procesador, hacen que el ordenador o el procesador lleve a cabo el método de la reivindicación 14.

40

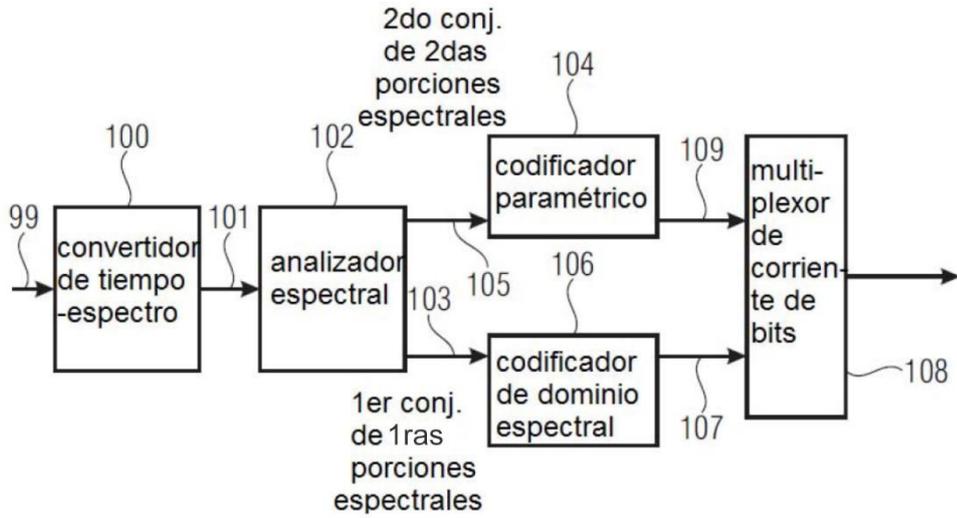


FIG 1A

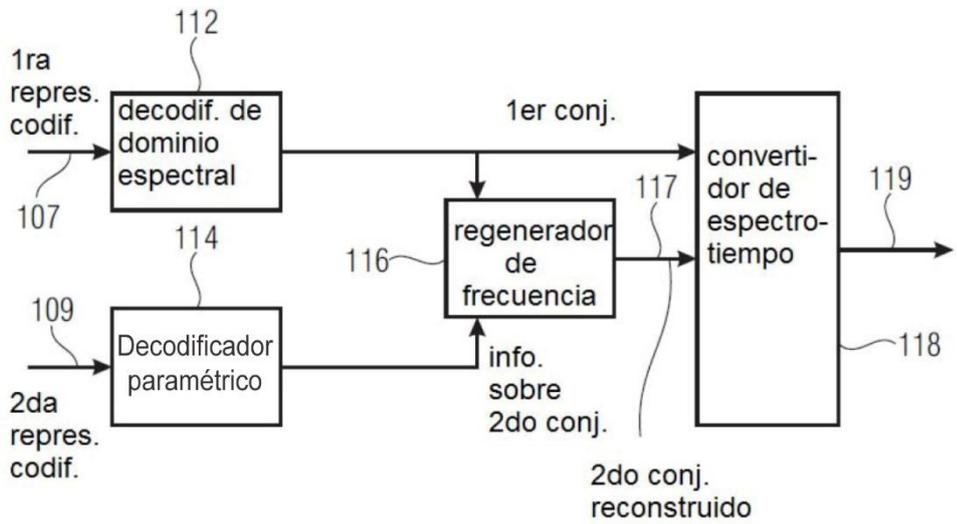


FIG 1B

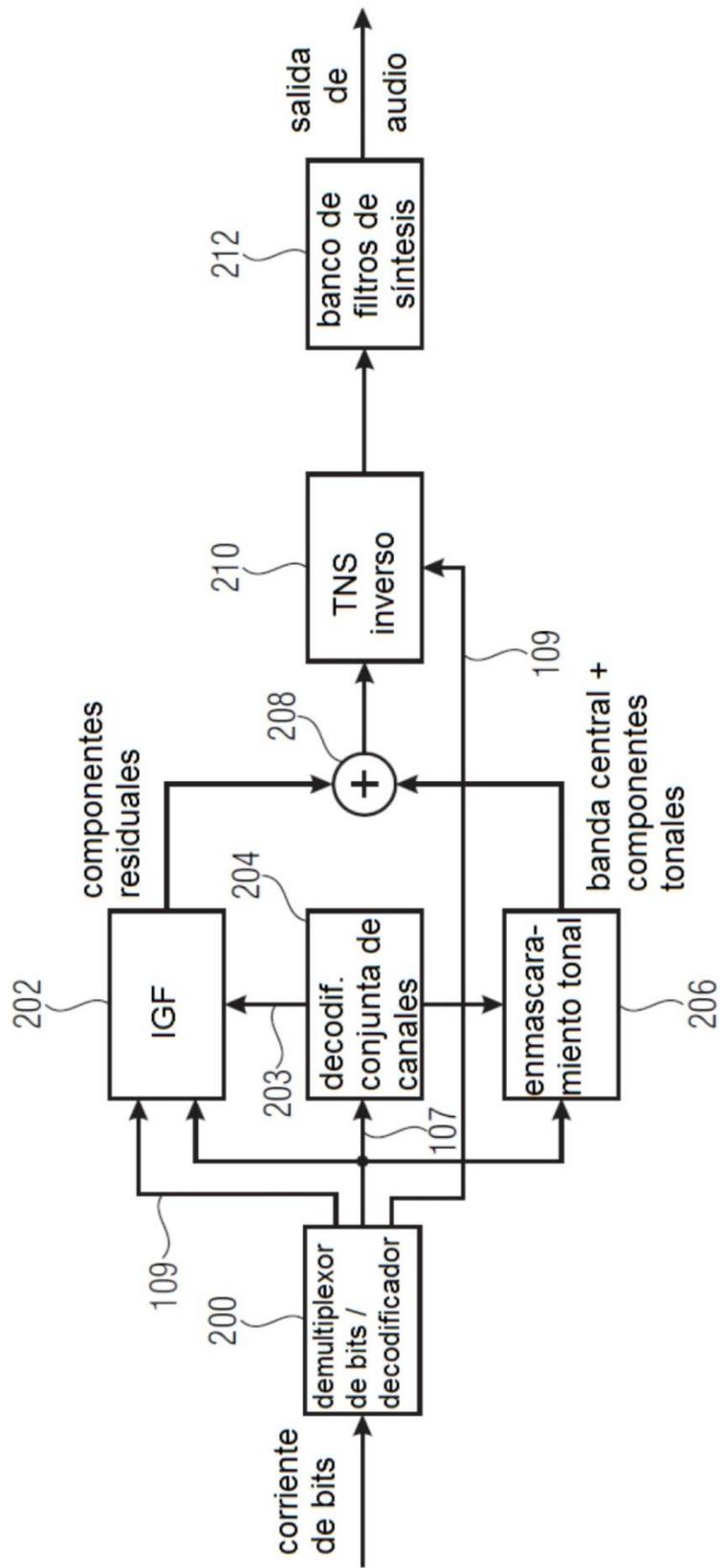


FIG 2A

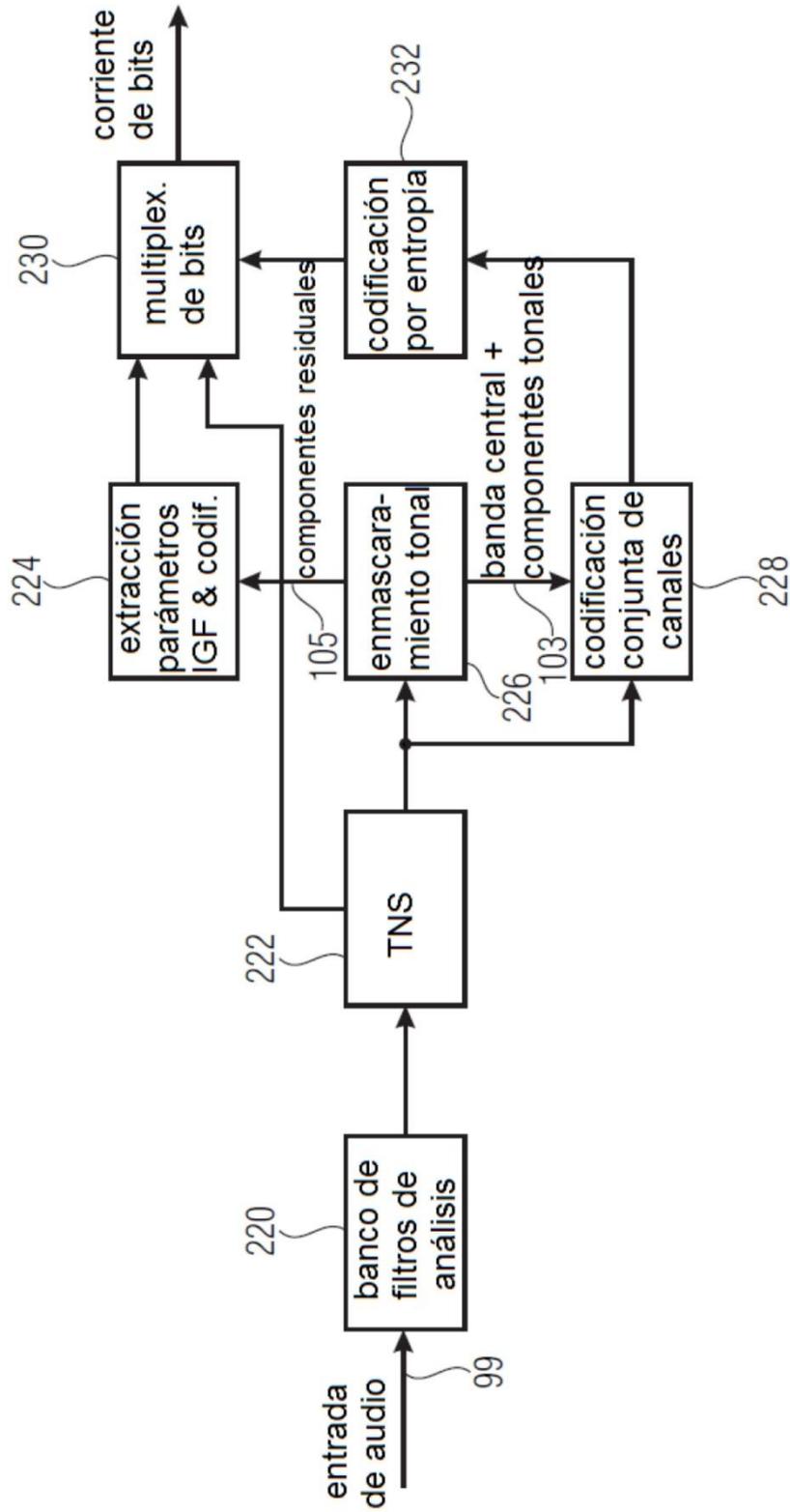


FIG 2B

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

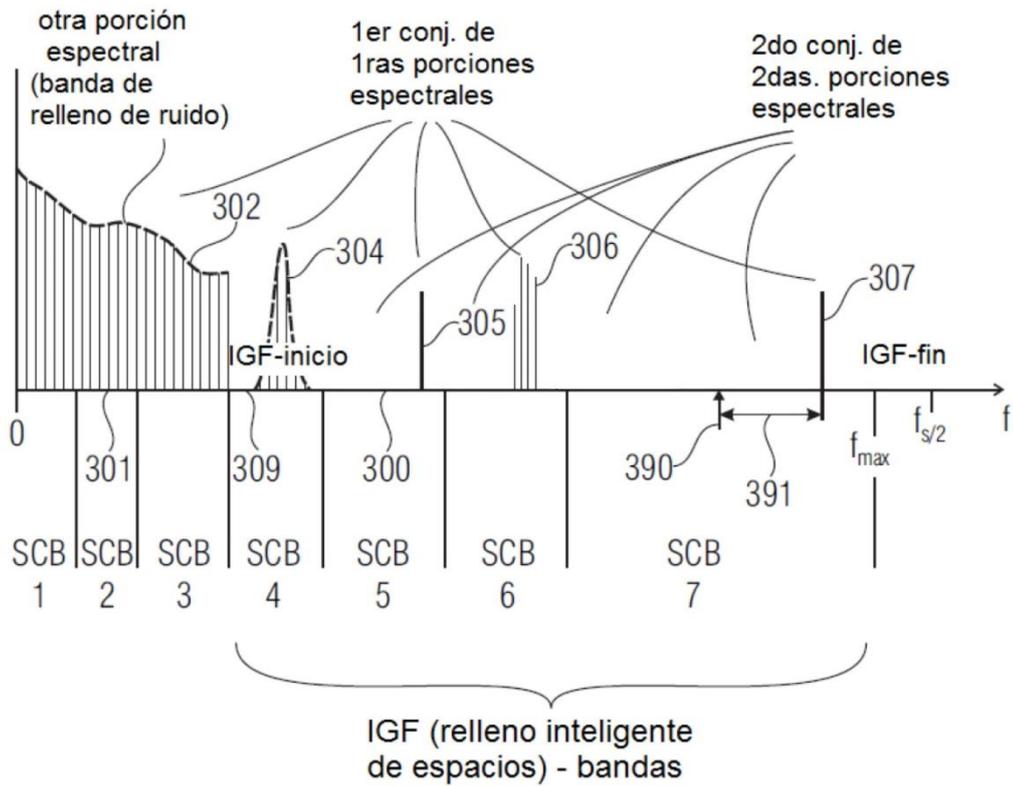


FIG 3A

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

308
310
312

FIG 3B

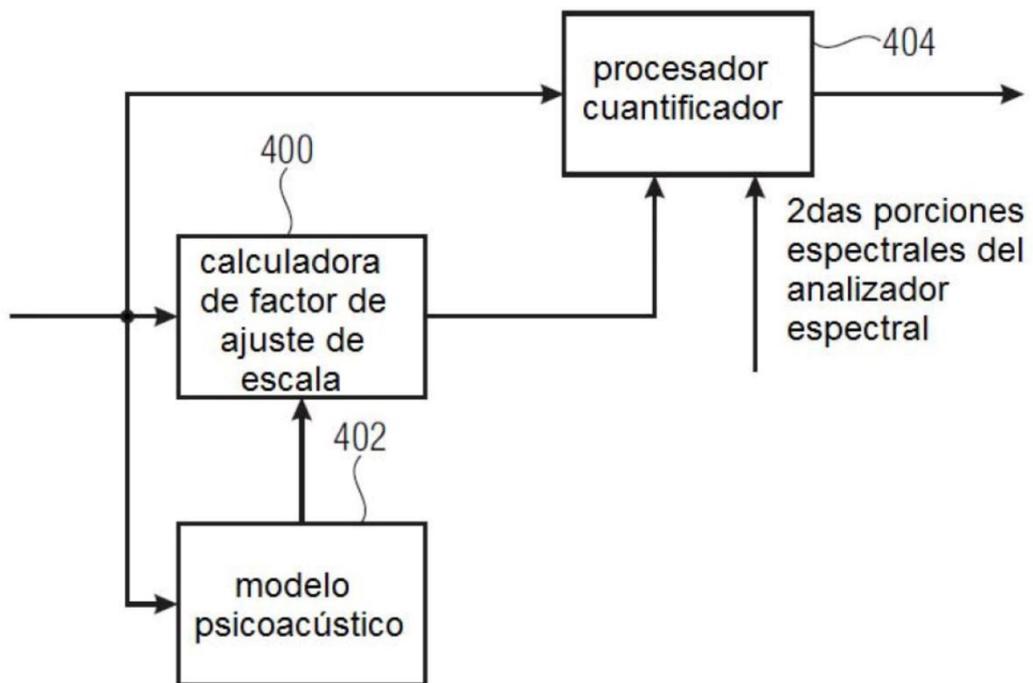


FIG 4A

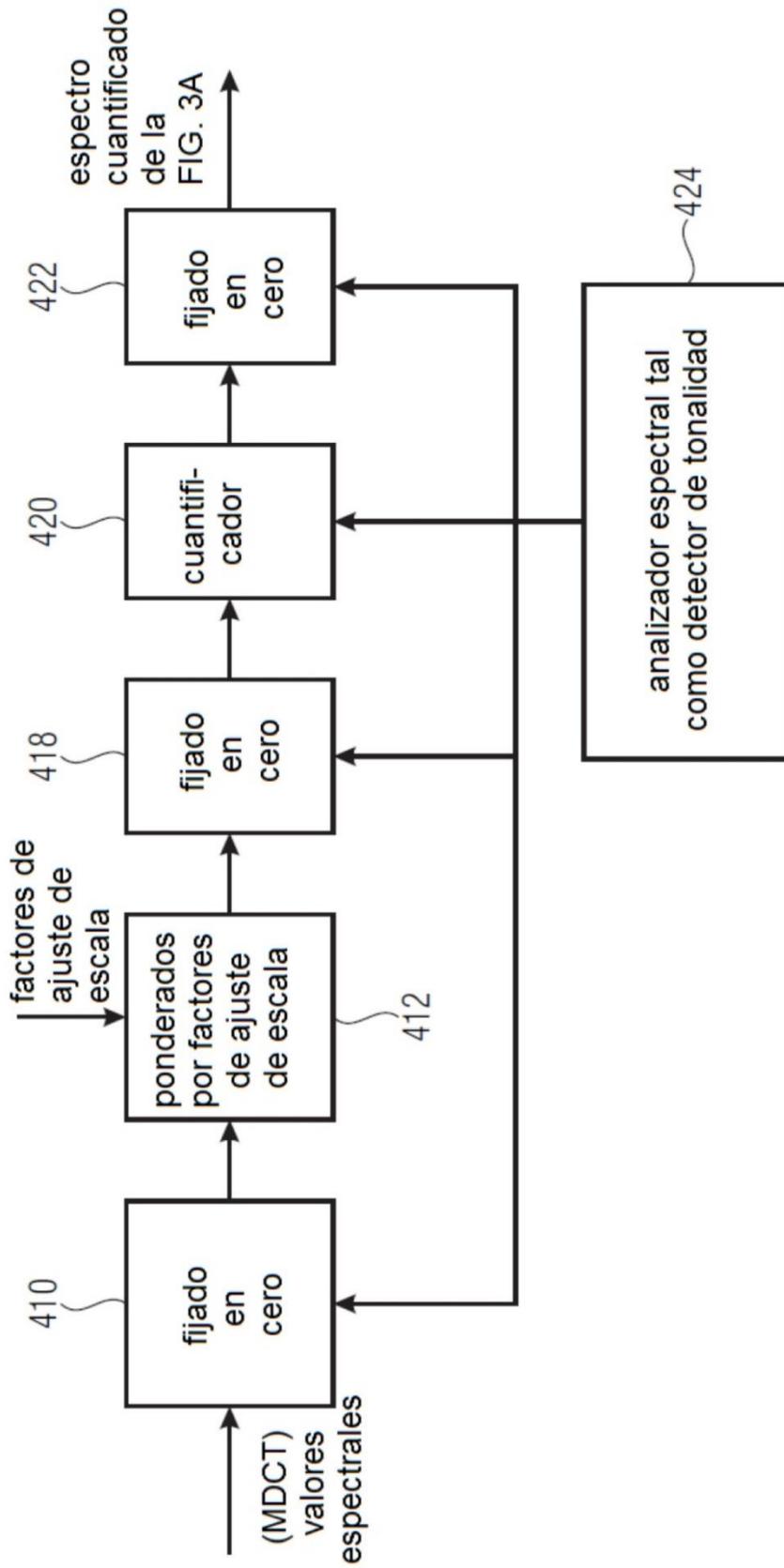


FIG 4B
(PROCESADOR CUANTIFICADOR)

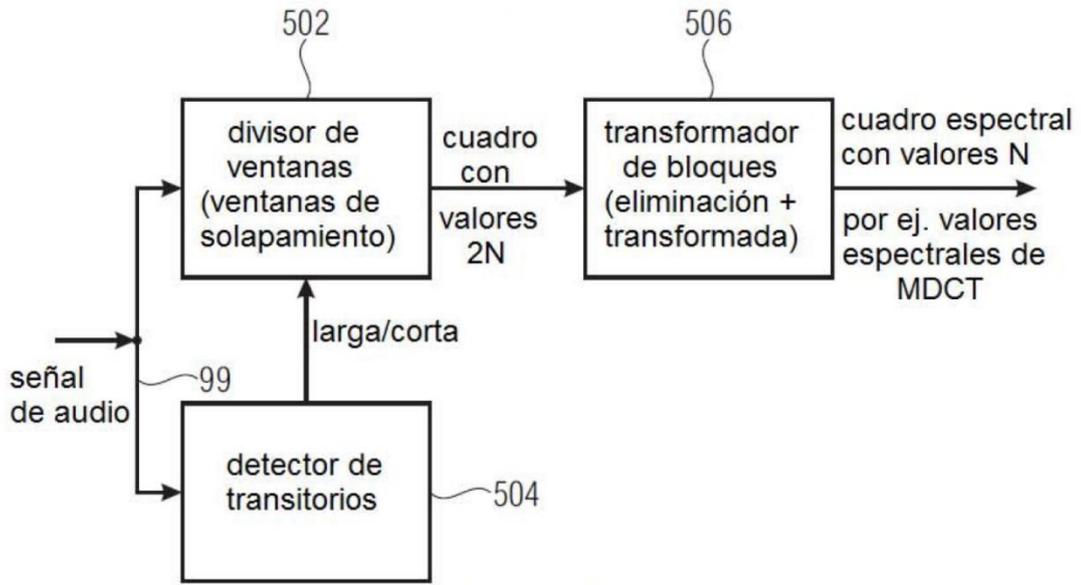


FIG 5A
(OTRAS PORCIONES ESPECTRALES)

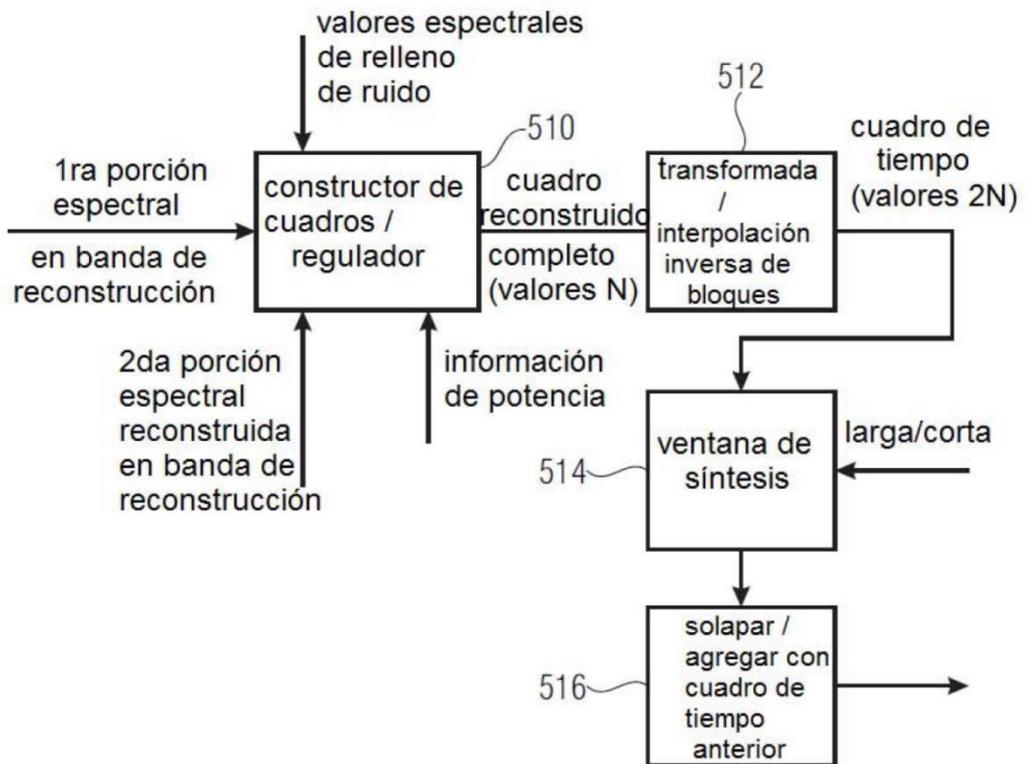


FIG 5B

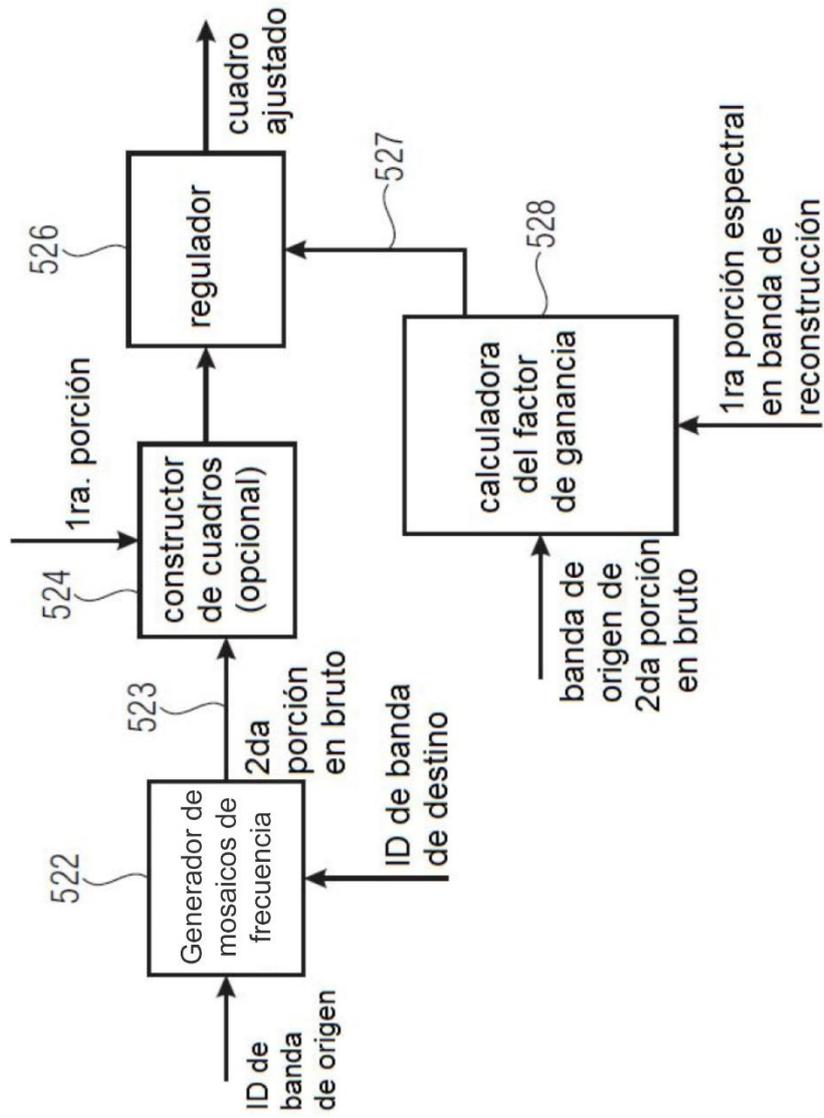


FIG 5C

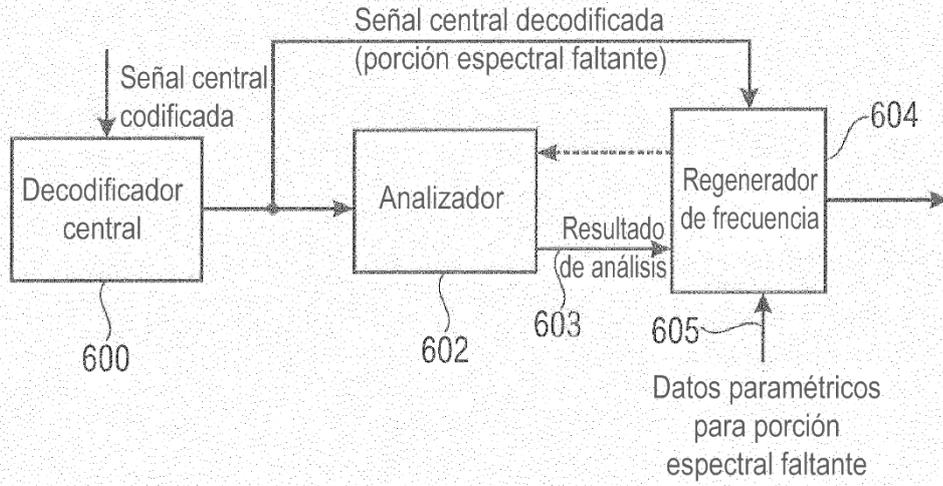


FIG 6A

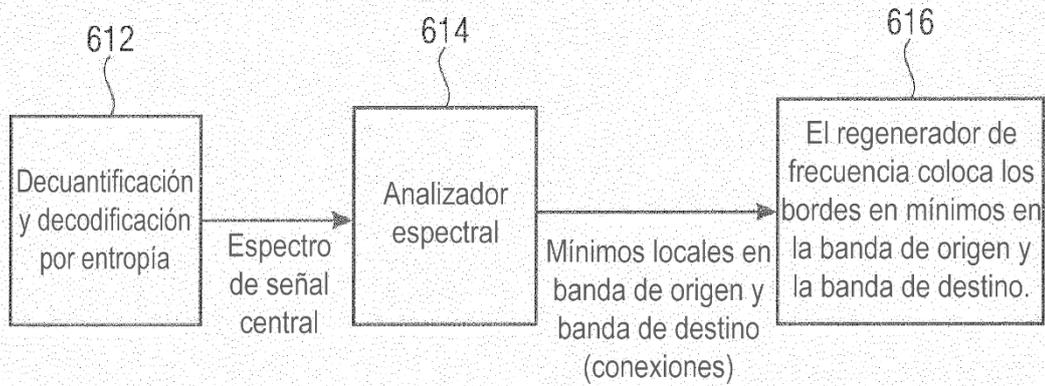


FIG 6B

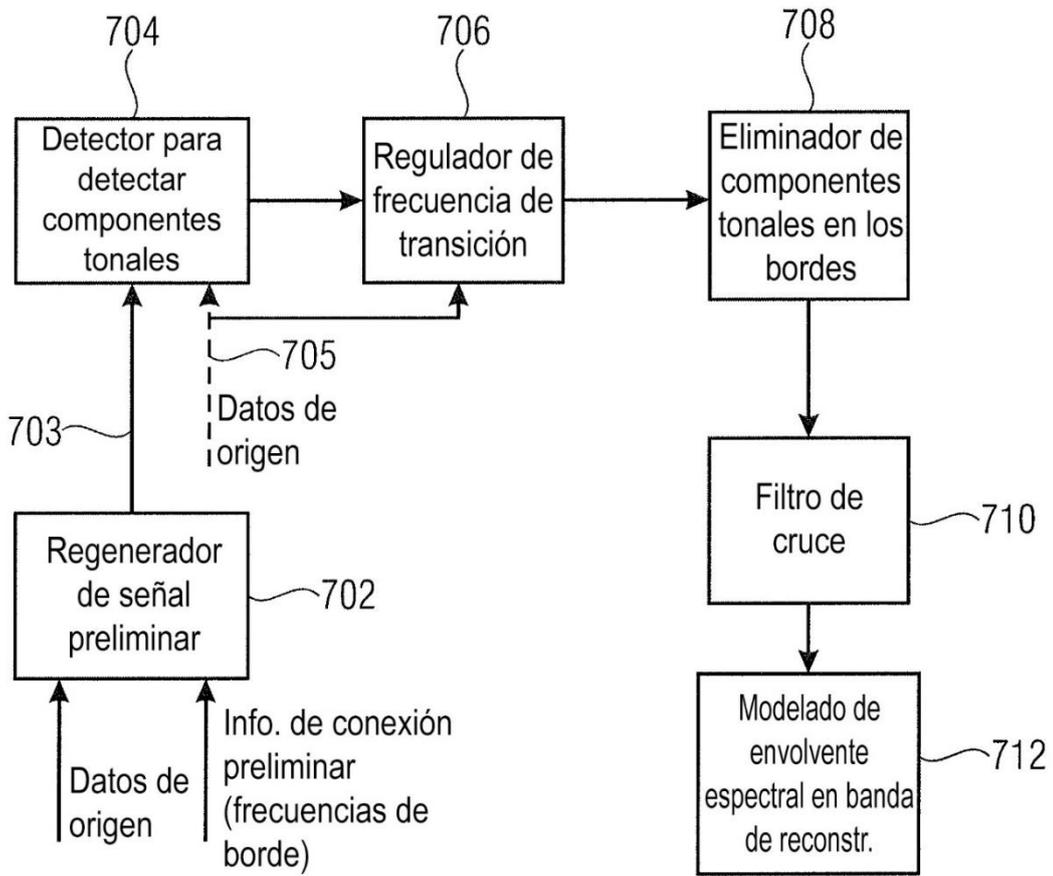


FIG 7A

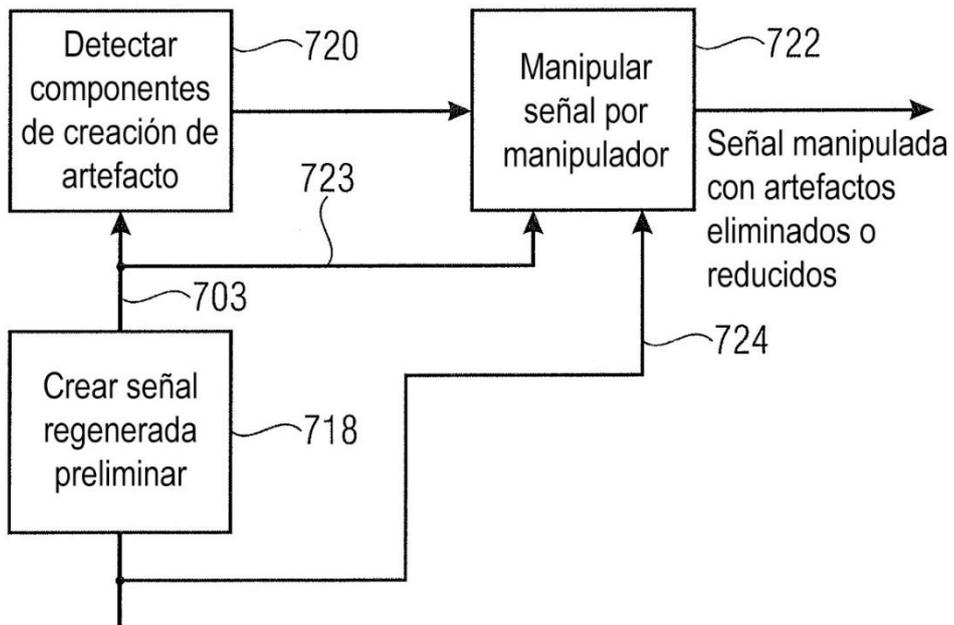


FIG 7B

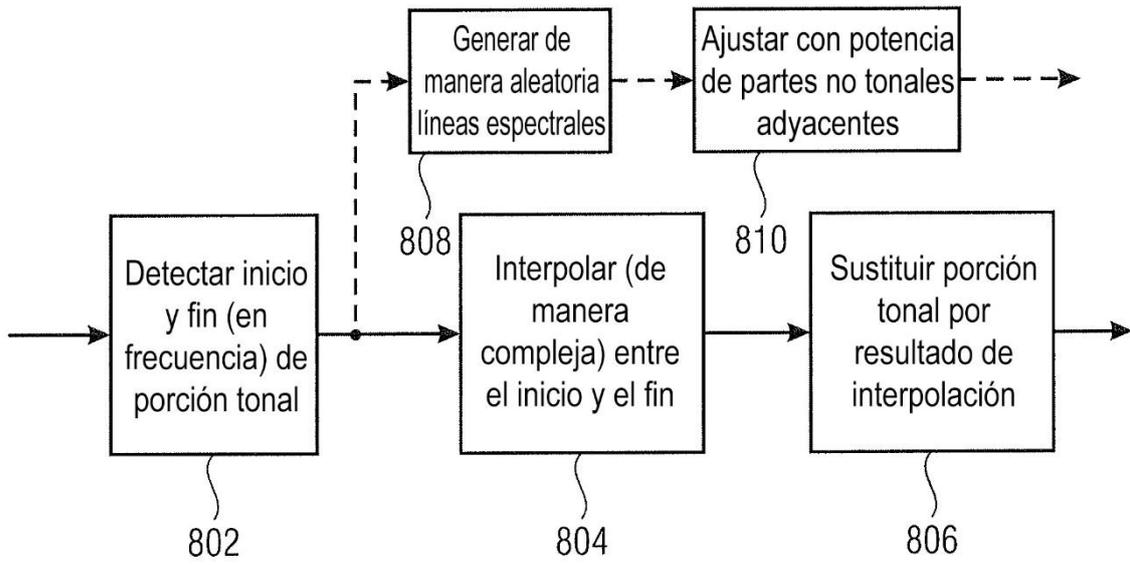


FIG 8A

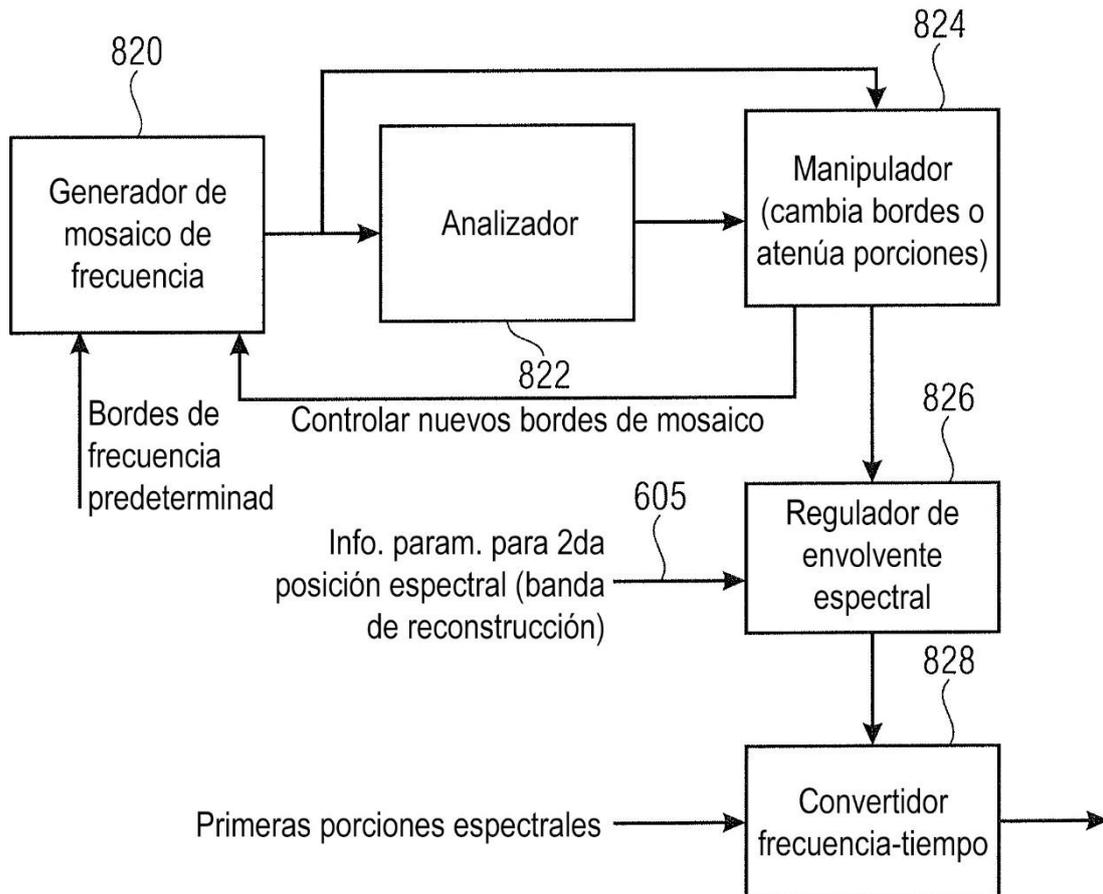


FIG 8B

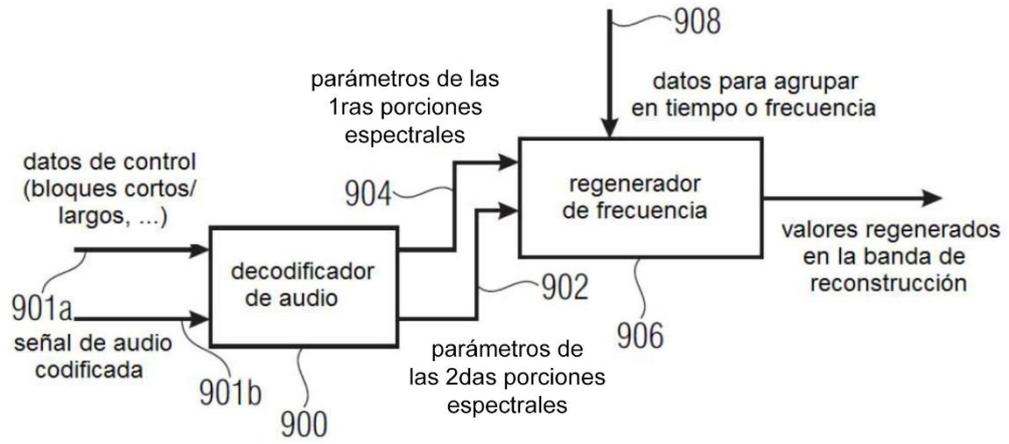


FIG 9A

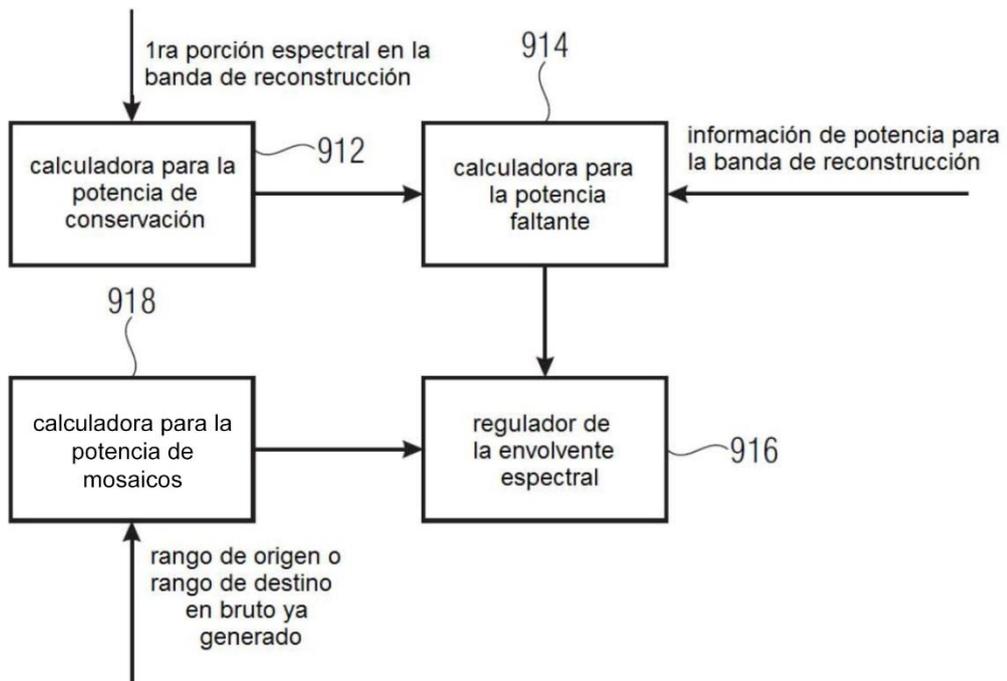
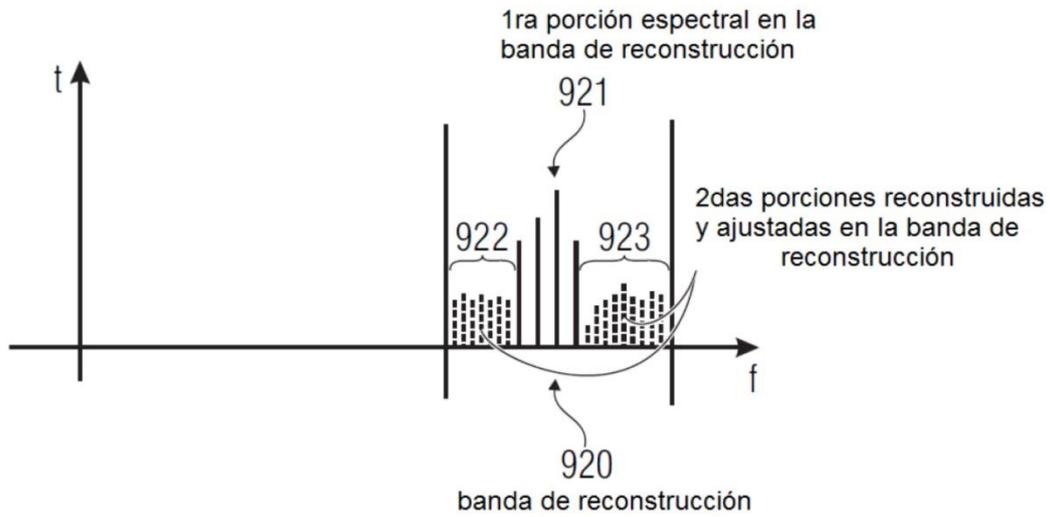


FIG 9B



por ej.

- potencia de conservación: 5 unidades
- valor de la potencia para la banda de reconstr.: 10 unidades (abarca 1ra y 2da porciones espectrales en la banda de reconstrucción)
- potencia de datos del rango de origen o datos del rango de destino en bruto: 8 unidades
- potencia faltante: 5 unidades
- factor de ganancia: $g = \sqrt{\frac{mE_k}{pE_k}} = 0.79$

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

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

FIG 9C

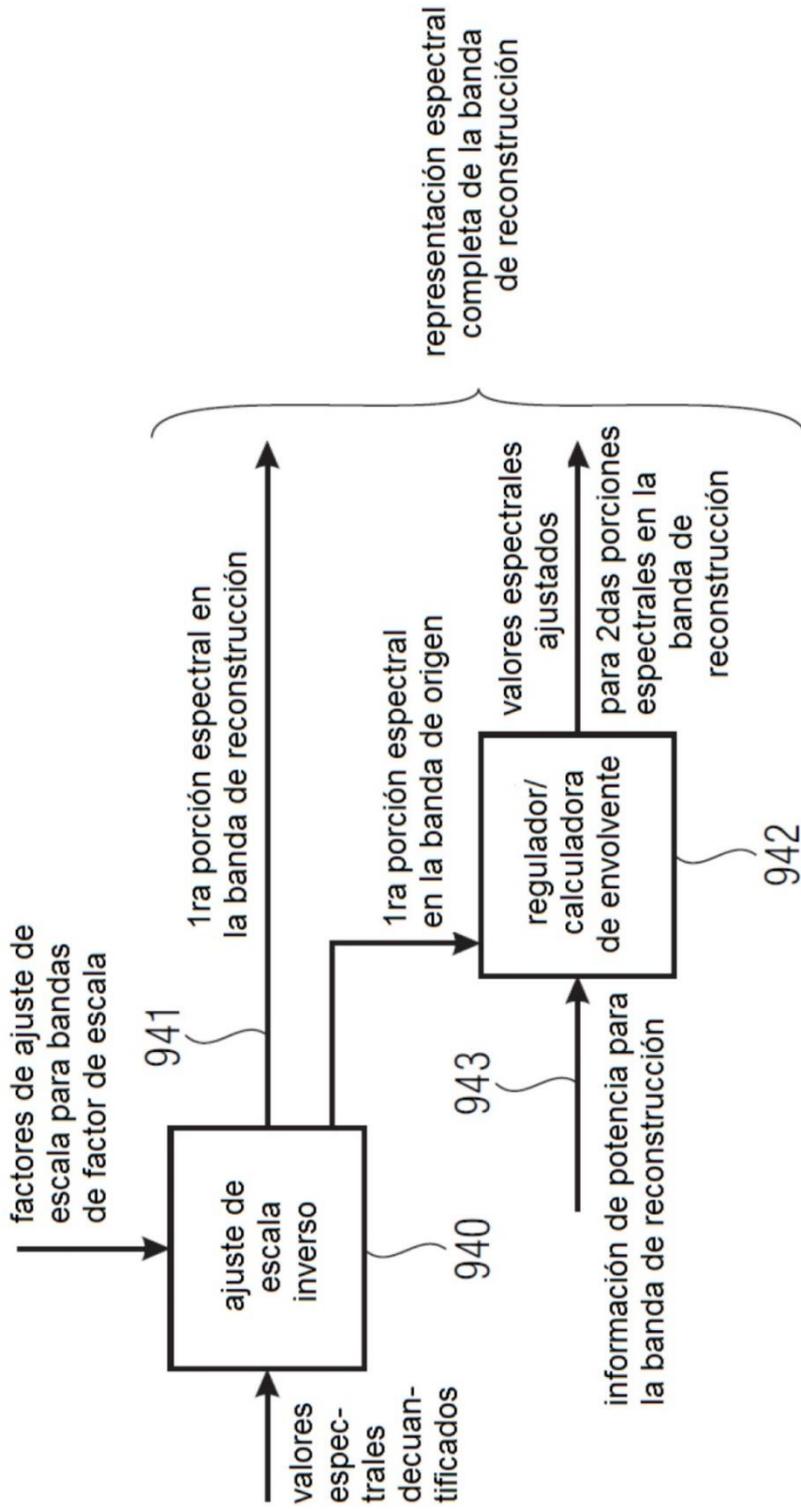


FIG 9D

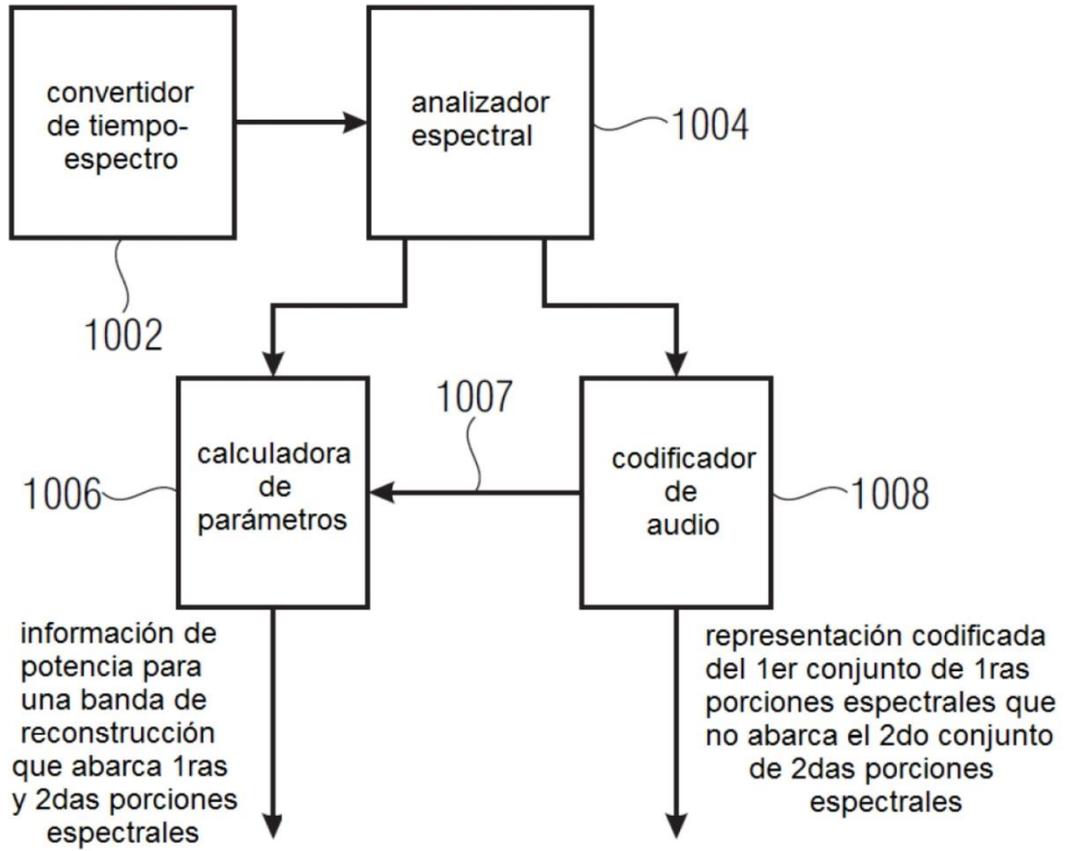


FIG 10A

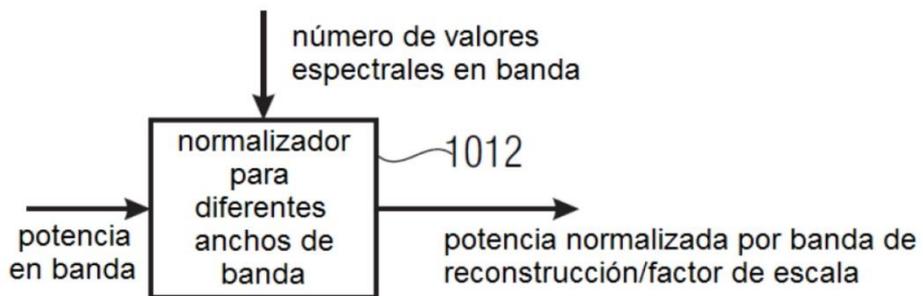


FIG 10B

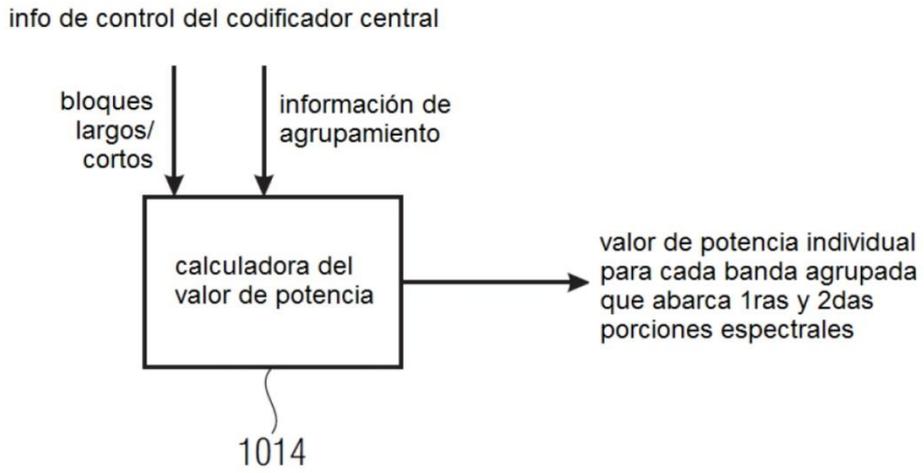


FIG 10C

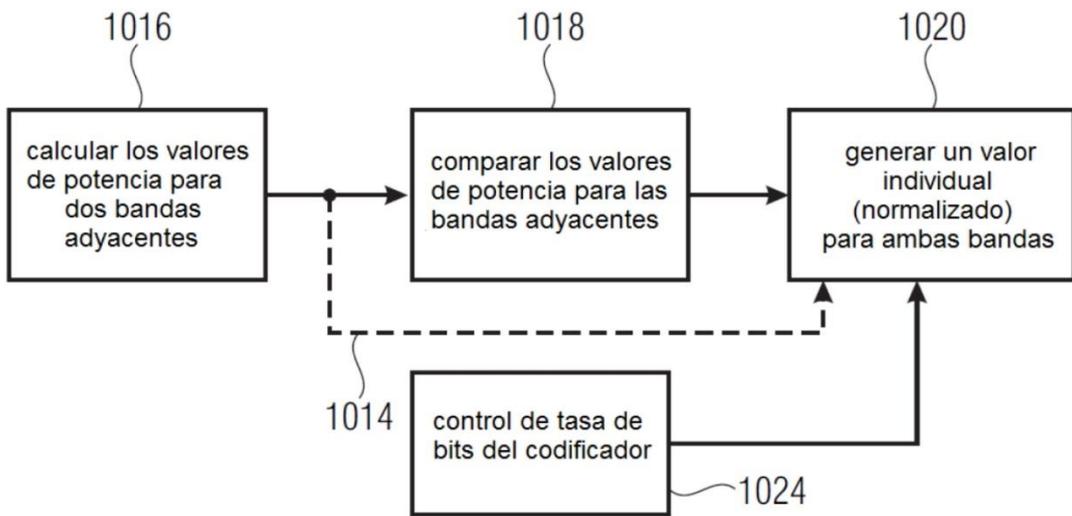


FIG 10D

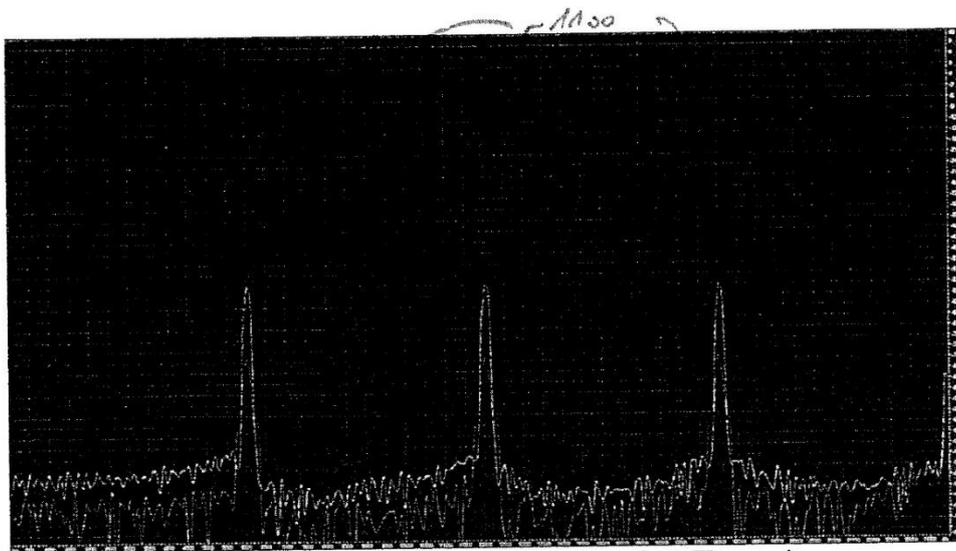
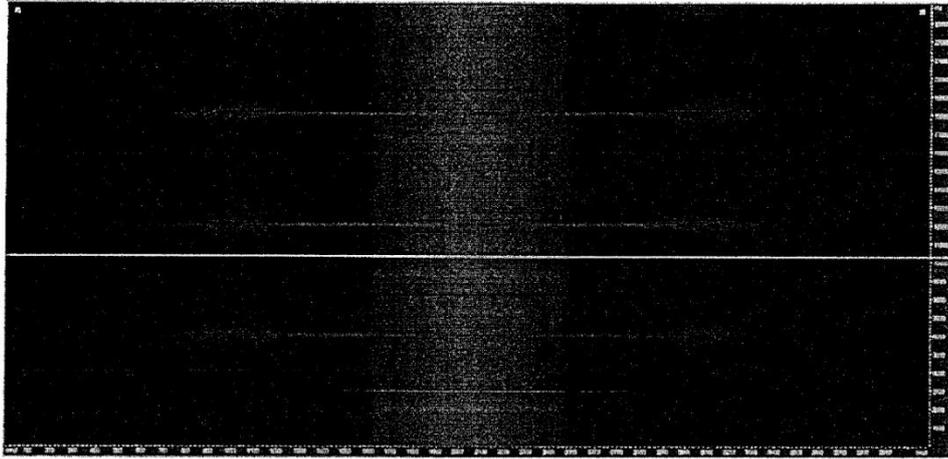
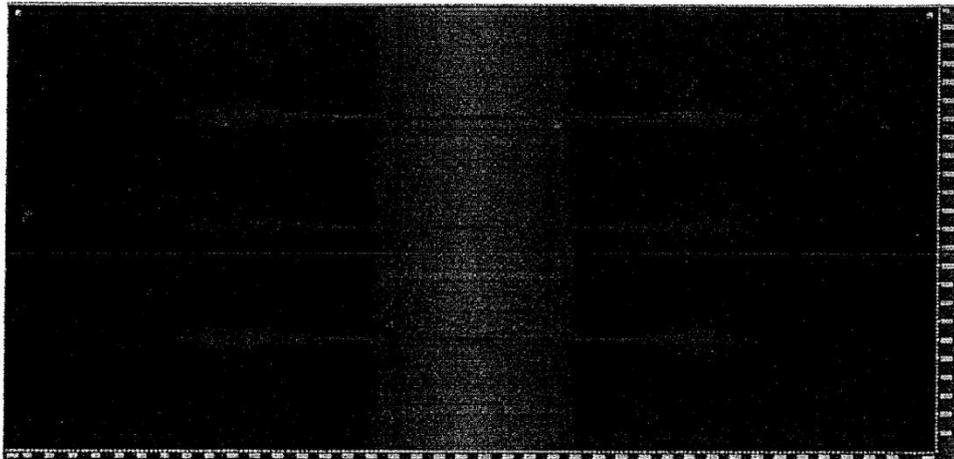


Fig. 11a



Espectrograma de un transitorio después de aplicar BWE. El eje x representa tiempo, el eje y, frecuencia.

Fig. 11c



Espectrograma de un transitorio después de aplicar BWE. El eje x representa tiempo, el eje y, frecuencia. A través de la aplicación de reducción de anillo de filtro, el anillo de filtro se reduce en aprox. 20dB.

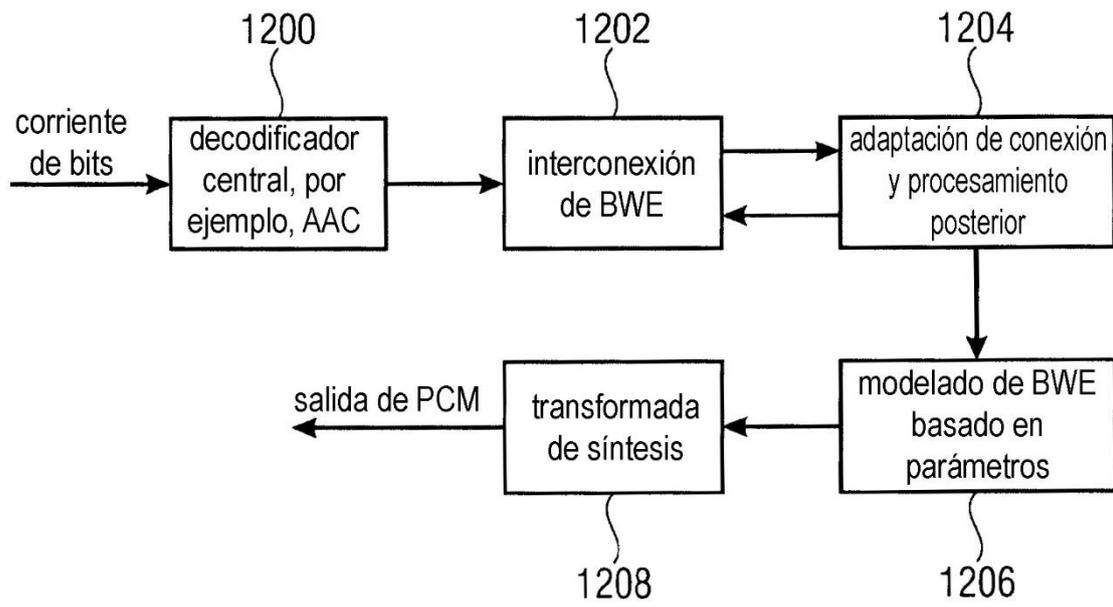
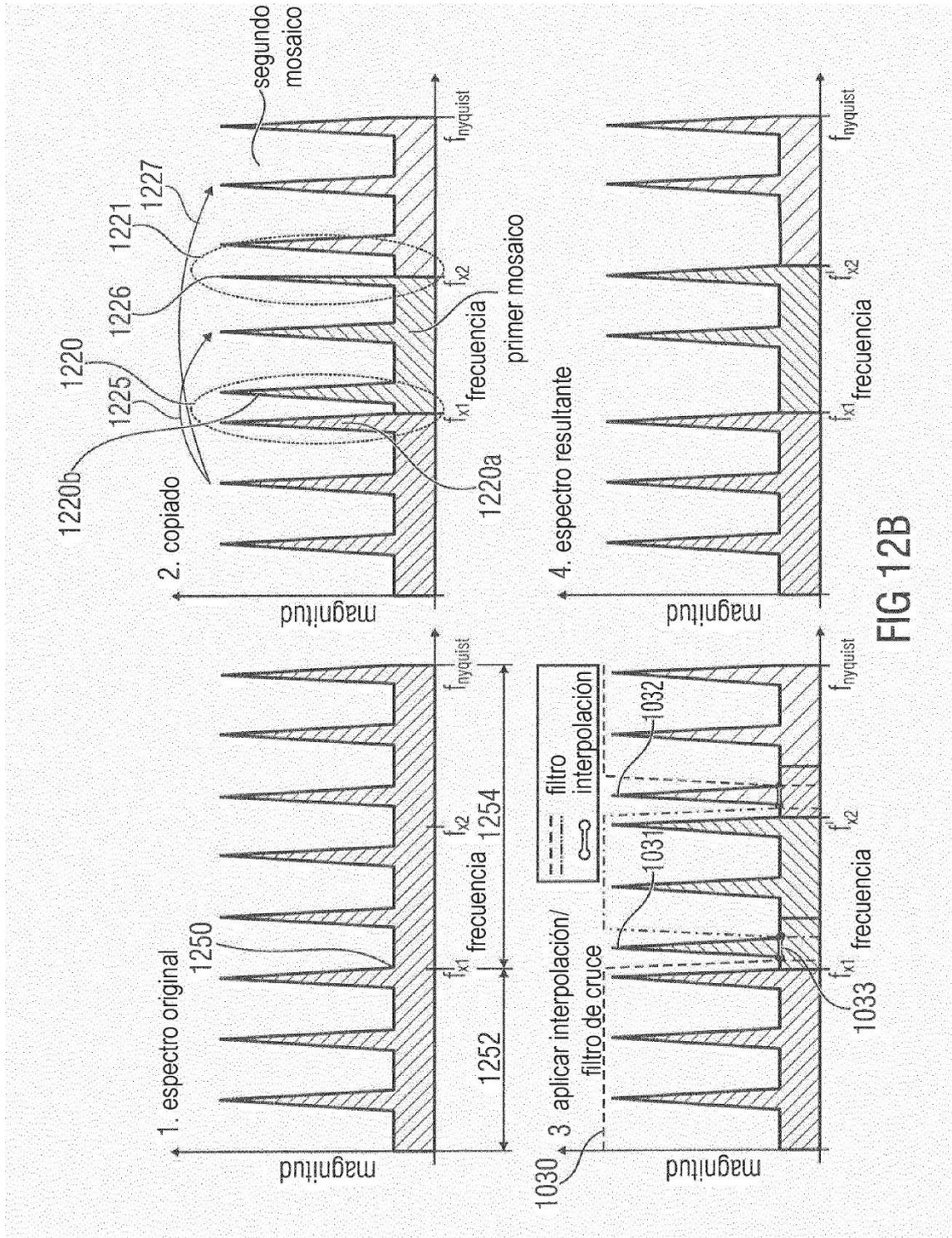
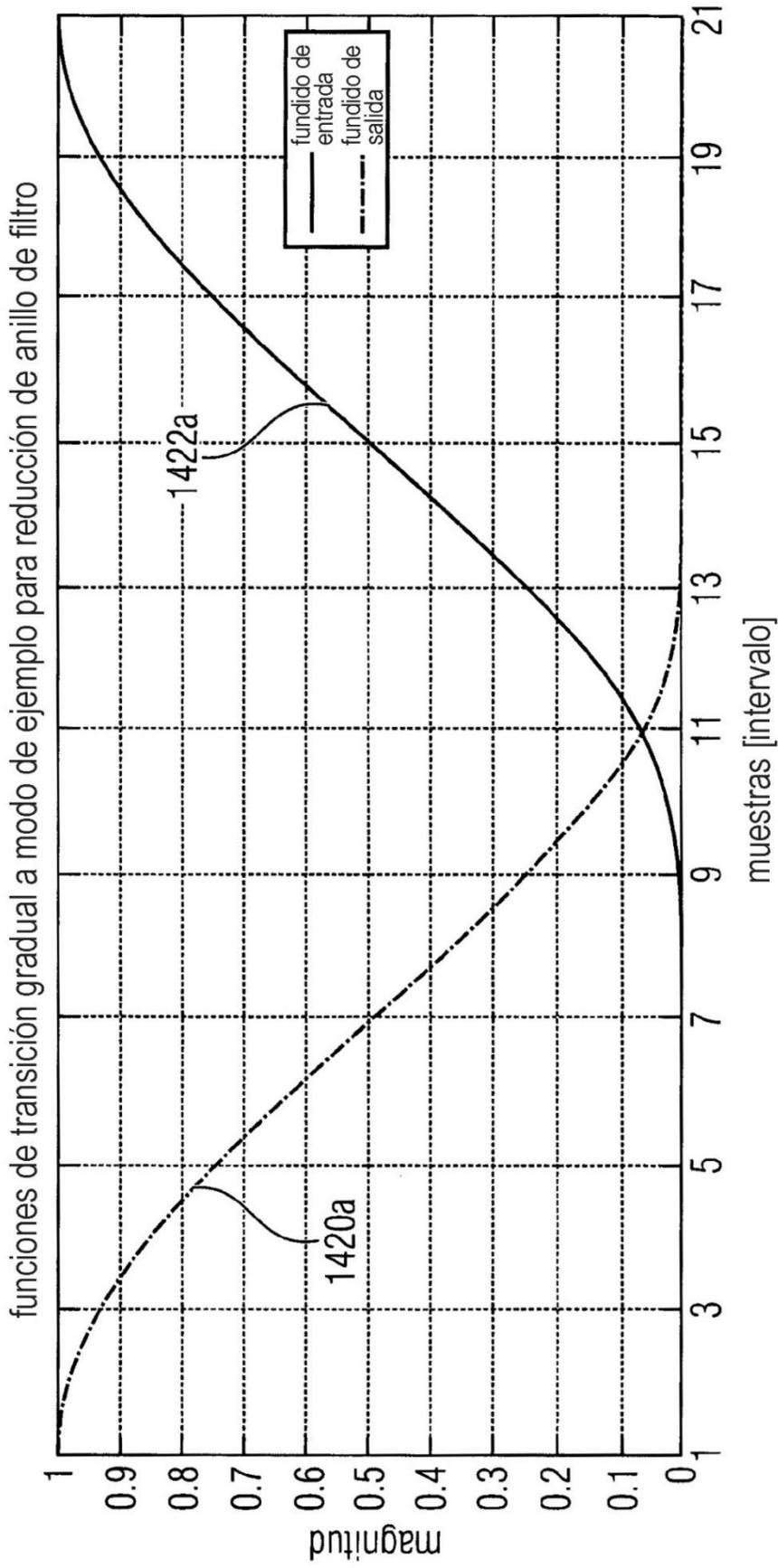


FIG 12A





funciones de transición gradual a modo de ejemplo, $N = 21$, $X_{bias} = 8$.

FIG 12C

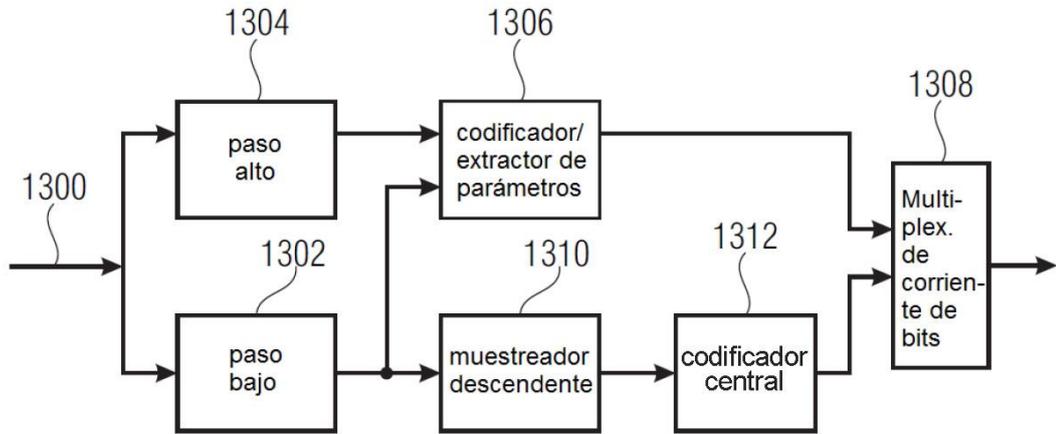


FIG 13A
(Técnica Anterior)

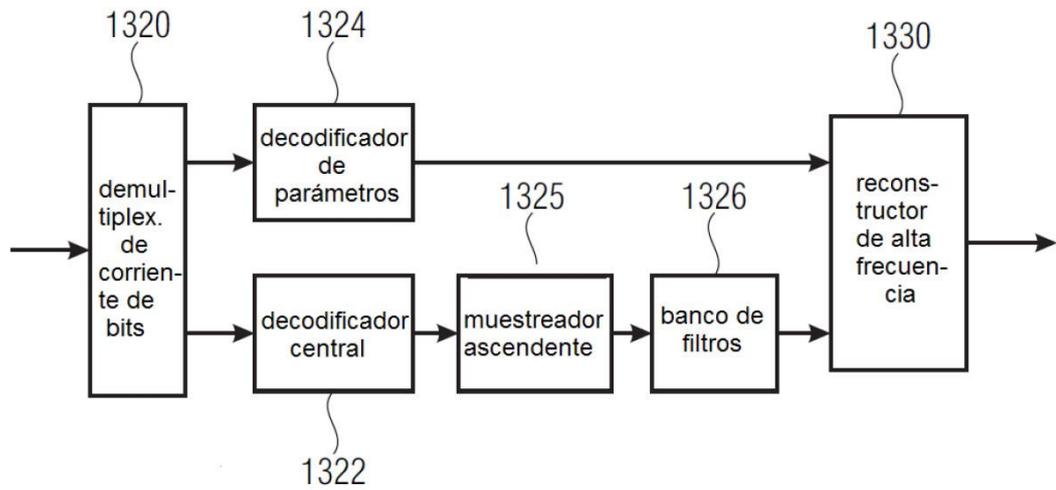


FIG 13B
(Técnica Anterior)

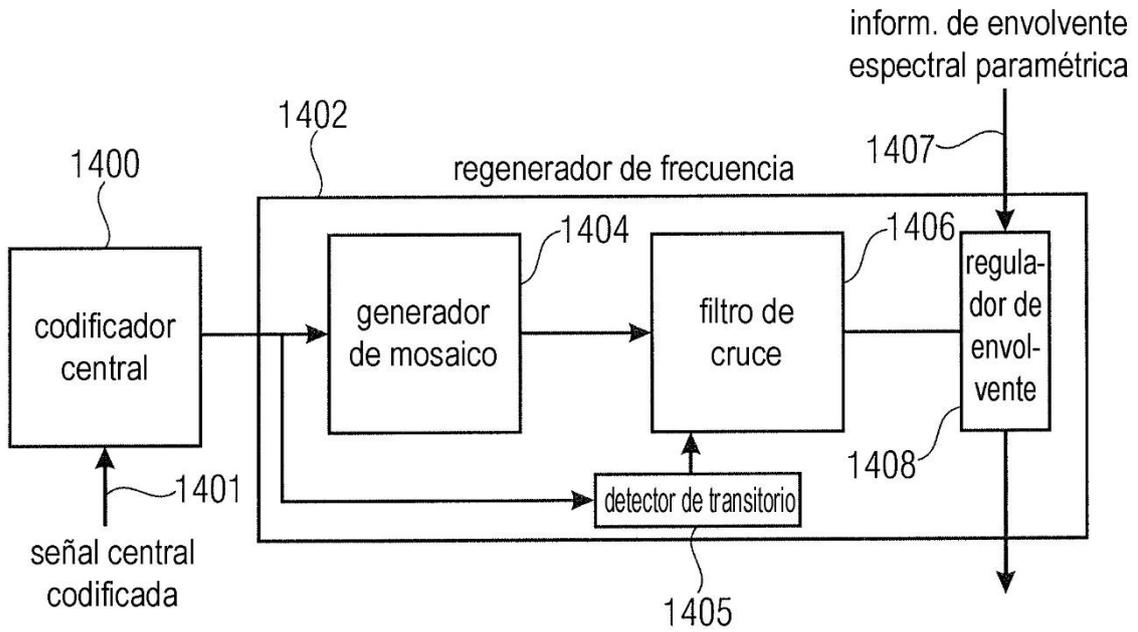


FIG 14A

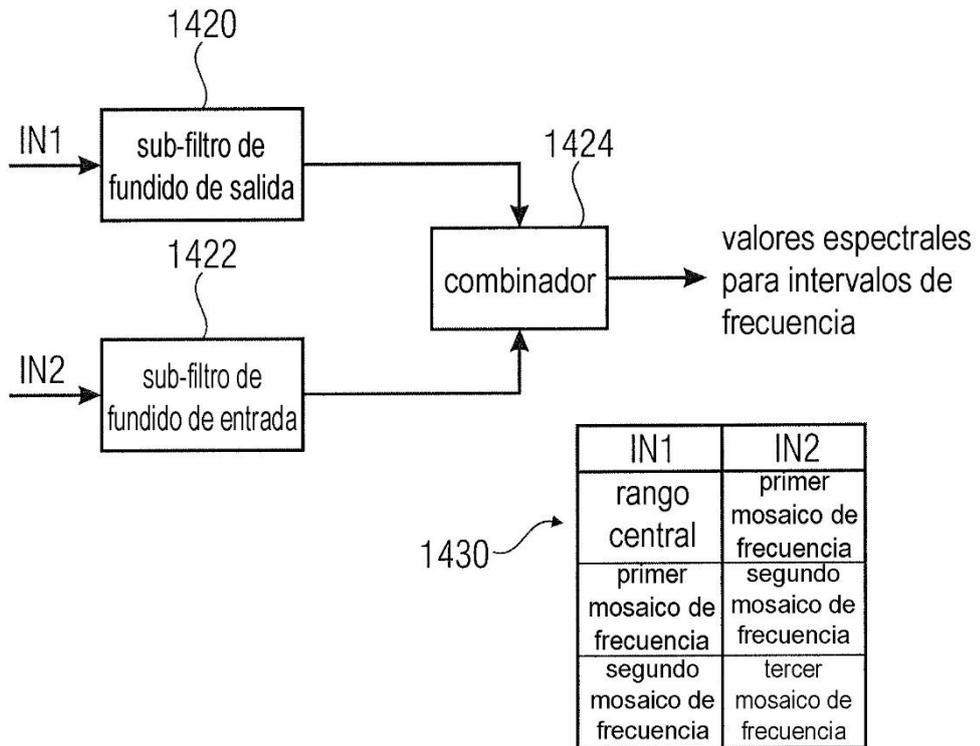


FIG 14B