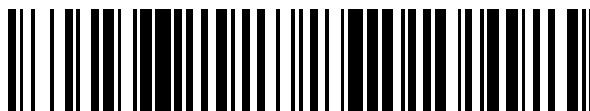


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 441 069**

51 Int. Cl.:

G10L 19/022 (2013.01)

G10L 19/20 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.10.2010 E 10760726 (9)**

97 Fecha y número de publicación de la concesión europea: **02.10.2013 EP 2471061**

54 Título: **Decodificador multimodo para señal de audio, codificador multimodo para señal de audio, procedimiento y programa de computación que usan un modelado de ruido en base a linealidad-predicción-codificación**

30 Prioridad:

08.10.2009 US 249774 P

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

31.01.2014

73 Titular/es:

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

72 Inventor/es:

**NEUENDORF, MAX;
FUCHS, GUILLAUME;
RETTELBACH, NIKOLAUS;
BAECKSTROEM, TOM;
LECOMTE, JÉRÉMIE y
HERRE, JÜRGEN**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 441 069 T3

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

DESCRIPCIÓN

Decodificador multimodo para señal de audio, codificador multimodo para señal de audio, procedimiento y programa de computación que usan un modelado de ruido en base a linealidad–predicción–codificación.

Campo Técnico

[0001] Las formas de realización de acuerdo con la presente invención se relacionan con un decodificador multimodo para señal de audio para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada del contenido de audio.

[0002] Otras formas de realización de acuerdo con la invención se relacionan con un codificador multimodo para señal de audio para proporcionar una representación codificada de un contenido de audio sobre la base de una representación de entrada del contenido de audio.

[0003] Otras formas de realización de acuerdo con la invención se relacionan con un procedimiento para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada del contenido de audio.

[0004] Otras formas de realización de acuerdo con la invención se relacionan con un procedimiento para proporcionar una representación codificada de un contenido de audio sobre la base de una representación de entrada del contenido de audio.

[0005] Otras formas de realización de acuerdo con la invención se relacionan con programas de computación que implementan dichos procedimientos.

Antecedentes de la Invención

[0006] A continuación se explicarán algunos antecedentes de la invención con el fin de facilitar la comprensión de la invención y las ventajas de la misma.

[0007] Durante la última década se ha realizado un gran esfuerzo para crear la posibilidad de almacenar y distribuir digitalmente contenidos de audio. Un logro importante en este sentido es la definición de la norma internacional ISO (Organización Internacional de Normas)/IEC (Comisión Internacional de Electrotecnia) 14496–3. La parte 3 de esta norma está relacionada con una codificación y decodificación de contenidos de audio y la sub–parte 4 de la parte 3 está relacionada con la codificación de audio en general. La parte 3, sub–parte 4 de la norma ISO/IEC 14496 define un concepto para la codificación y decodificación de contenido de audio en general. Asimismo, otras mejoras han sido propuestas con el fin de mejorar la calidad y/o reducir la velocidad de transmisión de bits necesarios.

[0008] Por otra parte, se ha comprobado que el rendimiento de los codificadores de audio basados en el dominio de frecuencia no es óptimo para los contenidos de audio que comprenden lenguaje. Recientemente se ha propuesto un codificador–decodificador (codec) unificado de voz–y–audio que combina eficazmente las técnicas de ambos mundos, es decir, la codificación de voz y la codificación de audio (véase, por ejemplo, la Referencia [1]).

[0009] En un codificador de audio de tales características, algunos cuadros de audio están codificados en el dominio de frecuencia y algunos cuadros de audio están codificados en el dominio de predicción lineal.

[0010] Sin embargo, se ha descubierto que es difícil hacer la transición entre los cuadros codificados en distintos dominios sin sacrificar una cantidad significativa de velocidad de transmisión de bits.

[0011] En vista de esta situación, hay un deseo de crear un concepto para codificar y decodificar un contenido de audio que comprende tanto voz como audio en general, lo que permite una realización eficaz de transiciones entre las porciones codificados utilizando diferentes modos.

[0012] Otro ejemplo de un codificador/decodificador (codec) de habla y de audio unificado conocido (USAC) se describe en LECOMTE Y OTROS: "Efficient Cross-Fade Windows for Transitions between LPC-Based and Non-LPC Based Audio Coding", CONVENCION AES 126; MAYO DE 2009, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK, USA, 1 de Mayo de 2009, XP040508994, en el que la conmutación sin costuras entre diferentes codecs de núcleo se logra utilizando ventanas de fundido cruzado adecuadamente diseñadas.

Resumen de la Invención

[0013] Una forma de realización de acuerdo con la invención crea un decodificador multimodo para señal de audio para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada del contenido de audio. El decodificador para señal de audio cuenta con un determinador de valor espectral configurado para obtener conjuntos de coeficientes espectrales decodificados para una pluralidad de

porciones del contenido de audio. El decodificador multimodo para señal de audio también incluye un procesador de espectro configurado para aplicar un modelado espectral a un conjunto de coeficientes espectrales decodificados, o a una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal para una porción del contenido de audio codificado en un modo de predicción lineal, y para aplicar un modelado espectral a un conjunto de coeficientes espectrales decodificados, o a una versión pre-procesada del mismo, independiente de un conjunto de parámetros de factores de escala para una porción del contenido de audio codificado en un modo de dominio de frecuencia. El decodificador multimodo para señal de audio también comprende un convertidor de dominio de frecuencia a dominio del tiempo configurado para obtener una representación de dominio de tiempo del contenido de audio sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de predicción lineal, y también para obtener una representación de dominio de tiempo del contenido de audio sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de dominio de frecuencia.

[0014] Dicho decodificador multimodo para señal de audio se basa en la constatación de que se pueden obtener transiciones eficaces entre las porciones del contenido de audio codificado en diferentes modos llevando a cabo un modelado espectral en el dominio de frecuencia, es decir, un modelado espectral de conjuntos de coeficientes espectrales decodificados, tanto para porciones de contenido de audio codificado en el modo de dominio de frecuencia como para porciones de contenido de audio codificado en el modo de predicción lineal. Por lo tanto, una representación de dominio de tiempo obtenida sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados para porción del contenido de audio codificado en el modo de predicción lineal está "en el mismo dominio" (por ejemplo, son valores de salida de transformaciones de dominio de frecuencia a dominio de tiempo del mismo tipo de transformación) como una representación de dominio de tiempo obtenida sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de dominio de frecuencia. De este modo, las representaciones de dominio de tiempo de una porción del contenido de audio codificado en el modo de predicción lineal y de una porción del contenido de audio codificado en el modo de dominio de frecuencia se pueden combinar de manera eficiente y sin artefactos inaceptables. Por ejemplo, las características de cancelación de solapamiento de los convertidores típicos de dominio de frecuencia a dominio de tiempo pueden ser aprovechadas por las señales de conversión de dominio de frecuencia a dominio de tiempo que están en el mismo dominio (por ejemplo, ambos representan un contenido de audio en un dominio de contenido de audio). Por lo tanto se pueden obtener transiciones de buena calidad entre las porciones del contenido de audio codificado en diferentes modos sin necesidad de una cantidad sustancial de tasa de bits para permitir tales transiciones.

[0015] En una realización preferida, el decodificador multimodo para señal de audio comprende además un solapador configurado para solapar y agregar una representación de dominio de tiempo de una porción del contenido de audio codificado en el modo de predicción lineal con una porción del contenido de audio codificado en el modo de dominio de frecuencia. Mediante la superposición de porciones del contenido de audio codificado en diferentes dominios se puede obtener la ventaja lograda mediante la introducción de conjuntos modelados espectralmente de coeficientes espectrales decodificados en el convertidor de dominio de frecuencia a dominio de tiempo en ambos modos del decodificador multimodo para señal de audio. Al llevar a cabo el modelado espectral antes de la conversión de dominio de frecuencia a dominio de tiempo en ambos modos del decodificador multimodo para señal de audio, el dominio del tiempo, las representaciones de dominio de tiempo de las porciones de los contenidos de audio codificado en los diferentes modos típicamente comprenden muy características adecuadas de superposición y agregado que permiten transiciones de buena calidad sin necesidad de información adicional de lado.

[0016] En una forma de realización preferida, el convertidor de dominio de frecuencia a dominio de tiempo está configurado para obtener una representación de dominio de tiempo del contenido de audio para una porción del contenido de audio codificado en el modo de predicción lineal utilizando una transformación solapada y para obtener una representación de dominio de tiempo del contenido de audio para una porción del contenido de audio codificado en el modo de dominio de frecuencia utilizando una transformación solapada. En este caso, el solapador está preferentemente configurado para solapar representaciones de dominio del tiempo de porciones posteriores del contenido de audio codificado en diferentes modos. En consecuencia se pueden obtener transiciones suaves. Debido al hecho de que un modelado espectral se aplica en el dominio de frecuencia para ambos modos, las representaciones de dominio de tiempo proporcionadas por el convertidor de dominio de frecuencia a dominio del tiempo en ambos modos son compatibles y permiten una transición de buena calidad. El uso de transformaciones solapadas aporta un mejor equilibrio entre calidad y eficiencia de tasa de bits de las transiciones, ya que las transformaciones solapadas permiten transiciones suaves, incluso en presencia de errores de cuantificación, evitando al mismo tiempo una cabecera significativa de tasa de bits.

[0017] En una forma de realización preferida, el convertidor de dominio de frecuencia a dominio de tiempo está configurado para aplicar una transformación solapada del mismo tipo de transformación para obtener la representación de dominio de tiempo de los contenidos de audio de las porciones del contenido de audio codificado en diferentes modos. En este caso, el solapador está configurado para solapar y agregar las representaciones de dominio de tiempo de porciones posteriores del contenido de audio codificado en diferentes modos, de manera que

un solapamiento de dominio de tiempo causado por la transformación solapada se reduce o elimina por el solapamiento y agregado. Este concepto se basa en el hecho de que las señales de salida de la conversión de dominio de frecuencia a dominio de tiempo están en el mismo dominio (dominio de contenido de audio) para los dos modos aplicando tanto los parámetros de factores de escala como los parámetros de dominio de predicción lineal en el dominio de frecuencia. Por consiguiente se puede aprovechar la cancelación de solapamiento que se obtiene típicamente mediante la aplicación de transformaciones solapadas del mismo tipo de transformación a porciones posteriores y parcialmente solapadas de una representación de la señal de audio.

[0018] En una forma de realización preferida, el solapador está configurado para solapar y agregar una representación de dominio de tiempo de una primera porción del contenido de audio codificado en el primero de los modos, proporcionado por una transformación solapada de síntesis asociada o una versión en escala de amplitud, pero espectralmente no alterada del mismo, y una representación de dominio de tiempo de una segunda porción posterior del contenido de audio codificado en un segundo de los modos, proporcionado por una transformación solapada de síntesis asociada o una versión en escala de amplitud, pero espectralmente no alterada del mismo. Al evitar aplicar, en las señales de salida de la transformación solapada de síntesis, cualquier procesamiento de señales (por ejemplo, un filtrado o similar) no común a todos los diferentes modos de codificación utilizados para porciones posteriores (que se solapan parcialmente) del contenido de audio, se puede aprovechar al máximo las características de cancelación de solapamiento de la transformación solapada.

[0019] En una forma de realización preferida, el convertidor de dominio de frecuencia a dominio de tiempo está configurado para proporcionar representaciones de dominio de tiempo de porciones del contenido de audio codificado indistintamente del modo, de manera tal que las representaciones de dominio de tiempo proporcionadas se encuentran en un mismo dominio y se pueden combinar linealmente sin aplicar una operación de filtrado de modelado de señales a una o ambas de las representaciones de dominio de tiempo proporcionadas. En otras palabras, las señales de salida de la conversión de dominio de frecuencia a dominio de tiempo son representaciones de dominio de tiempo del contenido de audio en sí para los dos modos (y no señales de excitación para una operación de filtrado de conversión de dominio de excitación a dominio de tiempo).

[0020] En una forma de realización preferida, el convertidor de dominio de frecuencia a dominio de tiempo está configurado para llevar a cabo una transformada inversa coseno discreta modificada para obtener, como resultado de la transformada inversa coseno discreta modificada, una representación de dominio de tiempo del contenido de audio en un dominio de señal de audio, tanto para una porción del contenido de audio codificado en el modo de predicción lineal como para una porción del contenido de audio codificado en el modo de dominio de frecuencia.

[0021] En una forma de realización preferida, el decodificador multimodo para señal de audio comprende un determinador de coeficiente de filtro LPC (codificación de predicción lineal) para obtener coeficientes decodificados de filtro LPC sobre la base de una representación codificada de los coeficientes de filtro LPC para una porción del contenido de audio codificado en un modo de predicción lineal. En este caso, el decodificador multimodo para señal de audio también comprende un transformador de coeficiente de filtro configurado para transformar los coeficientes decodificados de filtro LPC en una representación espectral, con el fin de obtener valores de ganancia asociados con frecuencias diferentes. Por lo tanto, el coeficiente de filtro LPC puede servir como parámetro de predicción lineal de dominio. El decodificador multimodo para señal de audio también comprende un determinador de factores de escala configurado para obtener valores decodificados de factores de escala (que sirven como parámetros de factores de escala) sobre la base de una representación codificada de los valores de factores de escala para una porción del contenido de audio codificado en un modo de dominio de frecuencia. El procesador espectral comprende un modificador de espectro configurado para combinar un conjunto de coeficientes espectrales decodificados asociados con una porción del contenido de audio codificado en el modo de predicción lineal, o una versión pre-procesada del mismo, con los valores de ganancia de modo de predicción lineal, con el fin de obtener una versión procesada de valor de ganancia (y, por consiguiente, modelada espectralmente) de los coeficientes espectrales (decodificados) en los que las contribuciones de los coeficientes espectrales decodificados, o de la versión pre-procesada del mismo, se ponderan dependiendo de los valores de ganancia. Asimismo, el modificador de espectro está configurado para combinar un conjunto de coeficientes espectrales decodificados asociados a una porción del contenido de audio codificado en el modo de dominio de frecuencia, o una versión pre-procesada del mismo, con los valores decodificados de factores de escala, a fin de obtener una versión procesada de factores de escala (modelada espectralmente) de los coeficientes espectrales (decodificado) en los que las contribuciones de los coeficientes espectrales decodificados, o de la versión pre-procesada del mismo, se ponderan dependiendo de los valores de factores de escala.

[0022] Mediante este enfoque se pueden obtener un modelado de ruido propio en ambos modos del decodificador multimodo para señal de audio garantizando al mismo tiempo que el convertidor de dominio de frecuencia a dominio de tiempo proporcione señales de salida con características adecuadas de transición en las transiciones entre porciones de la señal de audio codificada en diferentes modos.

[0023] En una forma de realización preferida, el transformador de coeficientes está configurado para transformar los coeficientes decodificados de filtro LPC, que representan una respuesta al impulso de dominio de tiempo de un filtro de codificación de predicción lineal (filtro LPC) en la representación espectral utilizando una transformada discreta

de Fourier de función impar. El transformador de coeficientes de filtro está configurado para obtener valores de ganancia del modo de predicción lineal a partir de la representación espectral de los coeficientes decodificados de filtro LPC, de manera tal que los valores de ganancia dependen de magnitudes de coeficientes de la representación espectral. Por lo tanto, el modelado espectral, que se realiza en el modo de predicción lineal adopta la funcionalidad de modelado de ruido de un filtro de codificación de predicción lineal. En consecuencia, el ruido de cuantificación de la representación espectral decodificada (o de la versión pre-procesada del mismo) se modifica de manera tal que el ruido de cuantificación es comparativamente pequeño para las frecuencias "importantes", para las cuales la representación espectral del coeficiente decodificado del filtro LPC es comparativamente grande.

[0024] En una forma de realización preferida, el transformador de coeficientes de filtro y el combinador están configurados de manera tal que una contribución de un coeficiente espectral decodificado determinado, o de una versión pre-procesada del mismo, a una versión procesada de ganancia del coeficiente espectral se determina por una magnitud de un valor de ganancia de modo de predicción lineal asociado al coeficiente espectral decodificado determinado.

[0025] En una forma de realización preferida, el determinador de valor espectral está configurado para aplicar una cuantificación inversa a valores espectrales cuantificados decodificados, con el fin de obtener coeficientes espectrales inversamente cuantificados. En este caso, el modificador de espectro está configurado para llevar a cabo un modelado de ruido de cuantificación ajustando un paso de cuantificación eficaz para un coeficiente espectral decodificado determinado dependiendo de una magnitud de un valor de ganancia de modo de predicción lineal asociado al coeficiente espectral decodificado determinado. En consecuencia, el modelado de ruido que se lleva a cabo en el dominio espectral está adaptado a características de señales descriptas por los coeficientes de filtro LPC.

[0026] En una forma de realización preferida, el decodificador multimodo para señal de audio está configurado para utilizar un cuadro de inicio de modo de predicción lineal intermedio con el fin de llevar a cabo una transición desde un cuadro de modo de dominio de frecuencia a un cuadro combinado de modo de predicción lineal/cuadro de modo de predicción lineal excitada por código algebraico. En este caso, el decodificador para señal de audio está configurado para obtener un conjunto de coeficientes espectrales decodificados para el cuadro de inicio de modo de predicción lineal. Además, el decodificador de audio está configurado para aplicar un modelado espectral al conjunto de coeficientes espectrales decodificados para el cuadro de inicio de modo de predicción lineal, o a una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros lineales de predicción de dominio asociados al mismo. El decodificador para señal de audio también está configurado para obtener una representación de dominio de tiempo del cuadro de inicio de modo de predicción lineal sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados. El decodificador de audio también está configurado para aplicar una ventana de inicio que tiene una pendiente de transición comparativamente larga en el lado izquierdo y una pendiente de transición comparativamente corta en el lado derecho a la representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal. Por consiguiente se crea una transición entre un cuadro de modo de dominio de frecuencia y un cuadro combinado de modo de predicción lineal/modo de predicción lineal excitada por código algebraico que comprende características adecuadas de solapamiento y agregado con el cuadro de modo de dominio de frecuencia anterior y que, al mismo tiempo, hace que los coeficientes de predicción lineal de dominio estén disponibles para ser utilizados por el cuadro combinado posterior de modo predicción lineal/modo de predicción lineal excitada por código algebraico.

[0027] En una forma de realización preferida, el decodificador multimodo para señal de audio está configurado para solapar una porción en el lado derecho de una representación de dominio de tiempo de un cuadro de modo de dominio de frecuencia anterior al cuadro de inicio del modo de predicción lineal con una porción en el lado izquierdo de una representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal, para obtener una reducción o cancelación de solapamiento del dominio de tiempo. Esta forma de realización se basa en la constatación de que se obtienen características adecuadas de cancelación de solapamiento del dominio de tiempo llevando a cabo un modelado espectral del cuadro de inicio de modo de predicción lineal en el dominio de frecuencia, debido a que un modelado espectral del cuadro de modo de dominio de frecuencia anterior también se lleva a cabo en el dominio de frecuencia.

[0028] En una forma de realización preferida, el decodificador para señal de audio está configurado para utilizar los parámetros de dominio de predicción lineal asociados al cuadro de inicio de modo de predicción lineal a fin de inicializar un decodificador de modo de predicción lineal excitada por código algebraico para decodificar por lo menos una porción del cuadro combinado de modo de predicción lineal/modo de predicción lineal excitada por código algebraico. De esta manera se elimina la necesidad de transmitir un conjunto adicional de parámetros de dominio de predicción lineal, que existe en algunos de los procedimientos convencionales. Más bien, el cuadro de inicio del modo de predicción lineal permite crear una transición adecuada a partir de un cuadro de modo de dominio de frecuencia anterior, incluso durante un período de solapamiento comparativamente largo, y para inicializar un decodificador de modo de predicción lineal excitada por código algebraico (ACELP). Por lo tanto se pueden obtener transiciones de buena calidad de audio con muy alto grado de eficiencia.

[0029] Otra forma de realización de acuerdo con la invención crea un codificador multimodo para señal de audio para proporcionar una representación codificada de un contenido de audio sobre la base de una representación de entrada del contenido de audio. El codificador de audio comprende un convertidor de dominio de tiempo a dominio de frecuencia configurado para procesar la representación de entrada del contenido de audio, para obtener una representación de dominio de frecuencia del contenido de audio. El codificador de audio comprende además un procesador espectral configurado para aplicar un modelado espectral a un conjunto de coeficientes espectrales, o una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal para una porción del contenido de audio a codificar en el dominio de predicción lineal. El procesador espectral también está configurado para aplicar un modelado espectral a un conjunto de coeficientes espectrales, o una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de factores de escala para una porción del contenido de audio a codificar en el modo de dominio de frecuencia.

[0030] El codificador multimodo para señal de audio anteriormente descrito está basado en la constatación de que se puede obtener una codificación de audio eficiente, que permite una decodificación de audio simple con bajas distorsiones, si se convierte una representación de entrada del contenido de audio en el dominio de frecuencia (también denominado dominio de frecuencia de tiempo), tanto para las porciones del contenido de audio a codificar en el modo de predicción lineal como para porciones del contenido de audio a codificar en el modo de dominio de frecuencia. Además, se ha descubierto que los errores de cuantificación se pueden reducir mediante la aplicación de un modelado espectral a un conjunto de coeficientes espectrales (o una versión pre-procesada del mismo) tanto para una porción del contenido de audio a codificar en el modo de predicción lineal como para una porción del contenido de audio a codificar en el modo de dominio de frecuencia. Si se utilizan diferentes tipos de parámetros para determinar el modelado espectral en los diferentes modos (es decir, parámetros de dominio de predicción lineal en el modo de predicción lineal y parámetros de factores de escala en el modo de dominio de frecuencia), el modelado de ruido se puede adaptar a la característica de la porción actualmente procesada del contenido de audio aplicando al mismo tiempo la conversión de dominio de tiempo a dominio de frecuencia a (porciones de) la misma señal de audio en los diferentes modos. En consecuencia, el codificador multimodo para señal de audio es capaz de proporcionar un buen desempeño de codificación para señales de audio que tienen porciones de audio en general y porciones de voz y audio aplicando selectivamente el tipo correcto de modelado espectral a los conjuntos de coeficientes espectrales. En otras palabras, un modelado espectral sobre la base de un conjunto de parámetros de dominio de predicción lineal se puede aplicar a un conjunto de coeficientes espectrales para un cuadro de audio que es reconocido como tipo voz, y se puede aplicar un modelado espectral sobre la base de un conjunto de parámetros de factores de escala a un conjunto de coeficientes espectrales para un cuadro de audio que es reconocido como tipo de audio en general, en vez de un tipo de voz.

[0031] En resumen, el codificador multimodo para señal de audio permite codificar un contenido de audio que tiene características temporalmente variables (tipo voz para algunas porciones temporales y de audio en general para otras porciones) en donde la representación de dominio de tiempo del contenido de audio se convierte en el dominio de frecuencia de la misma manera para porciones de contenido de audio a codificar en diferentes modos. Las diferentes características de las diferentes porciones del contenido de audio se consideran aplicando un modelado espectral sobre la base de diferentes parámetros (parámetros de dominio de predicción lineal versus parámetros de factores de escala), con el fin de obtener coeficientes espectrales modelados espectralmente o la cuantificación posterior.

[0032] En una forma de realización preferida, el convertidor de dominio de tiempo a dominio de frecuencia está configurado para convertir una representación de dominio de tiempo de un contenido de audio en un dominio de señal de audio, en una representación de dominio de frecuencia del contenido de audio, tanto para una porción del contenido de audio a codificar en el modo de predicción lineal como para una porción del contenido de audio a codificar en el modo de dominio de frecuencia. Al llevar a cabo la conversión de dominio de tiempo a dominio de frecuencia (en el sentido de una operación de transformación tal como, por ejemplo, una operación de transformación MDCT o una operación de separación la operación de separación de frecuencia basada en el banco de filtros) sobre la base de la misma señal de entrada tanto para el modo de dominio de frecuencia como para el modo de predicción lineal, se puede llevar a cabo una operación de solapar y agregar en el lado del decodificador con una eficacia particularmente adecuada, lo que facilita la reconstrucción de señales en el lado del decodificador y evita la necesidad de transmitir datos adicionales cada vez que hay una transición entre los diferentes modos.

[0033] En una forma de realización preferida, el convertidor de dominio de tiempo a dominio de frecuencia está configurado para aplicar una transformación solapada de análisis del mismo tipo de transformación para obtener representaciones de dominio de frecuencia para porciones del contenido de audio a codificar en diferentes modos. Una vez más, el uso de transformaciones solapadas del mismo tipo de transformación permite una simple reconstrucción del contenido de audio, evitando al mismo tiempo artefactos de bloqueo. En particular, es posible utilizar un muestreo crítico sin una carga significativa.

[0034] En una forma de realización preferida, el procesador espectral está configurado para aplicar selectivamente el modelado espectral del conjunto de coeficientes espectrales, o una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal que se obtienen utilizando un análisis basado en la correlación de una porción del contenido de audio a codificar en el modo de predicción lineal, o

dependiendo de un conjunto de parámetros de factores de escala que se obtienen utilizando un análisis de modelo psicoacústico de una porción del contenido de audio a codificar en el modo de dominio de frecuencia. De esta manera se puede lograr un modelado de ruido adecuado tanto para porciones tipo voz del contenido de audio, en las que el análisis basado en la correlación proporciona información significativa sobre el modelado de ruido, como para porciones de audio en general del contenido de audio, para las que el análisis del modelo psicoacústico proporciona información significativa sobre el modelado de ruido.

[0035] En una forma de realización preferida, el codificador para señal de audio comprende un selector de modo configurado para analizar el contenido de audio a fin de decidir si se codificará una porción del contenido de audio en el modo de predicción lineal o en el modo de dominio de frecuencia. En consecuencia se puede optar por el concepto de modelado de ruido adecuado, sin afectar el tipo de conversión de dominio de tiempo a dominio de frecuencia en algunos casos.

[0036] En una forma de realización preferida, el codificador multimodo para señal de audio está configurado para codificar un cuadro de audio, el cual está entre un cuadro de modo de dominio de frecuencia y un cuadro combinado de modo de predicción lineal/modo de predicción lineal excitada por código algebraico como un cuadro de inicio de modo de predicción lineal. El codificador multimodo para señal de audio está configurado para aplicar una ventana de inicio que tiene una pendiente de transición comparativamente larga en el lado izquierdo y una pendiente de transición comparativamente corta en el lado derecho a la representación de dominio de tiempo del cuadro de inicio de modo de predicción lineal, para obtener una ventana de representación de dominio de tiempo. El codificador multimodo para señal de audio también está configurado para obtener una representación de dominio de frecuencia de la ventana de representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal. El codificador multimodo para señal de audio también está configurado para obtener un conjunto de parámetros de dominio de predicción lineal para el cuadro de inicio de modo de predicción lineal y para aplicar un modelado espectral a la representación de dominio de frecuencia de la ventana de representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal, o a una versión pre-procesada del mismo, dependiendo del conjunto de parámetros de dominio de predicción lineal. El codificador para señal de audio también está configurado para codificar el conjunto de parámetros de dominio de predicción lineal y la representación del dominio de frecuencia modelada espectralmente de la ventana de representación de dominio de tiempo del cuadro de inicio de modo de predicción lineal. De esta manera se obtiene información codificada de un cuadro de transición de audio, cuya información codificada del cuadro de transición de audio se puede utilizar para una reconstrucción del contenido de audio, en donde la información codificada sobre el cuadro de transición de audio permite una transición suave en el lado izquierdo y al mismo tiempo permite una inicialización de un decodificador de modo ACELP para decodificar un cuadro de audio posterior. Se reduce al mínimo una cabecera causada por la transición entre los diferentes modos del codificador multimodo para señal de audio.

[0037] En una forma de realización preferida, el codificador multimodo para señal de audio está configurado para utilizar los parámetros de dominio de predicción lineal asociados al cuadro de inicio de modo de predicción lineal a fin de inicializar un codificador de modo de predicción lineal excitada por código algebraico para codificar por lo menos una porción del cuadro combinado de modo de predicción lineal/modo de predicción lineal excitada por código algebraico siguiendo el cuadro de inicio del modo de predicción lineal. Por consiguiente, los parámetros de dominio de predicción lineal, que se obtienen para el cuadro de inicio de modo de predicción lineal, y que también están codificados en una corriente de bits que representa el contenido de audio, se volverán a utilizar para la codificación de un cuadro de audio posterior, en el que se utilice el modo ACELP. Esto aumenta la eficiencia de la codificación y también permite una decodificación eficiente sin información adicional en el lado de inicialización ACELP.

[0038] En una forma de realización preferida, el codificador multimodo para señal de audio comprende un determinador de coeficientes de filtro LPC configurado para analizar una porción del contenido de audio a codificar en un modo de predicción lineal, o una versión pre-procesada del mismo, para determinar los coeficientes de filtro LPC asociados a la porción del contenido de audio a codificar en el modo de predicción lineal. El codificador multimodo para señal de audio también comprende un transformador de coeficientes de filtro configurado para transformar los coeficientes decodificados de filtro LPC en una representación espectral, con el fin de obtener valores de ganancia del modo de predicción lineal asociados a diferentes frecuencias. El codificador multimodo para señal de audio también comprende un determinador de factores de escala configurado para analizar una porción del contenido de audio a codificar en el modo de dominio de frecuencia, o una versión pre-procesada del mismo, para determinar factores de escala asociados a la porción del contenido de audio a codificar en el modo de dominio de frecuencia. El codificador multimodo para señal de audio también incluye un dispositivo combinador configurado para combinar una representación del dominio de frecuencia de una porción del contenido de audio a codificar en el modo de predicción lineal, o una versión pre-procesada del mismo, con los valores de ganancia del modo de predicción lineal, para obtener los componentes espectrales procesados de ganancia (también denominados coeficientes), en donde las contribuciones de los componentes espectrales (o coeficientes espectrales) de la representación de dominio de frecuencia del contenido de audio se ponderan dependiendo de los valores de ganancia del modo de predicción lineal. El combinador también está configurado para combinar una representación del dominio de frecuencia de una porción del contenido de audio a codificar en el modo de dominio de frecuencia, o una versión pre-procesada del mismo, con los factores de escala, para obtener componentes espectrales

procesados de ganancia, en donde las contribuciones de los componentes espectrales (o coeficientes espectrales) de la representación de dominio de frecuencia del contenido de audio se ponderan dependiendo de los factores de escala.

5 **[0039]** En esta forma de realización, los componentes espectrales procesados de ganancia forman conjuntos formados espectralmente de coeficientes espectrales (o componentes espectrales).

10 **[0040]** Otra forma de realización de acuerdo con la invención crea un procedimiento para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada del contenido de audio.

15 **[0041]** Sin embargo, otra forma de realización de acuerdo con la invención crea un procedimiento para proporcionar una representación codificada de un contenido de audio sobre la base de una representación de entrada del contenido de audio.

[0042] Incluso, otra forma de realización de acuerdo con la invención crea un programa de computadora para llevar a cabo uno o más de dichos procedimientos.

20 **[0043]** Los procedimientos y el programa de computadora se basan en las mismas conclusiones que el aparato mencionado anteriormente.

Breve Descripción de las Figuras

25 **[0044]** A continuación se describirán las formas de realización de la presente invención tomando como referencia las Figuras adjuntas, en las cuales:

La Figura 1 muestra un diagrama esquemático en bloques de un codificador para señal de audio, de acuerdo con una forma de realización de la invención;

30 La Figura 2 muestra un diagrama esquemático en bloques de un codificador para señal de audio de referencia;

La Figura 3 muestra un diagrama esquemático en bloques de un codificador para señal de audio, de acuerdo con una forma de realización de la invención;

35 La Figura 4 muestra una ilustración de una interpolación de coeficientes LPC para una ventana TCX (transformación con código de excitación);

40 La Figura 5 muestra un código de programa de computadora de una función para obtener los valores de ganancia de dominio de predicción lineal sobre la base de coeficientes decodificados de filtro LPC;

La Figura 6 muestra un código de programa de computadora para combinar un conjunto de coeficientes espectrales decodificados con los valores de ganancia del modo de predicción lineal (o valores de ganancia de dominio de predicción lineal);

45 La Figura 7 muestra una representación esquemática de diferentes cuadros e información relacionada para un codec conmutado de dominio de tiempo/ dominio de frecuencia (TD/FD) que envía lo que se denomina "LPC" como cabecera;

50 La Figura 8 muestra una representación esquemática de cuadros y parámetros relacionados para un conmutador desde el codificador de dominio de frecuencia a dominio de predicción lineal utilizando "LPC2MDCT" para las transiciones;

55 La Figura 9 muestra una representación esquemática de un codificador para señal de audio que comprende un modelado de ruido a base de LPC para TCX y un codificador de dominio de frecuencia;

La Figura 10 muestra una vista unificada de un sistema unificado de codificación de voz y audio (USAC) donde se llevó a cabo TCX MDCT (transformación con código de excitación – transformada coseno discreta modificada) en la dominio de la señal;

60 La Figura 11 muestra un diagrama esquemático en bloques de un decodificador para señal de audio, de acuerdo con una forma de realización de la invención;

La Figura 12 muestra una vista unificada de un decodificador de USAC con TCX–MDCT en el dominio de señal;

65 La Figura 13 muestra una representación esquemática de los pasos de procesamiento, los cuales se pueden llevar a cabo en los decodificadores de señal de audio de acuerdo con las Figuras 7 y 12;

La Figura 14 muestra una representación esquemática de un procesamiento de cuadros de audio posteriores en los decodificadores de audio de acuerdo con las Figuras 11 y 12;

5 La Figura 15 muestra una tabla que representa un número de coeficientes espectrales en función de una variable MOD [];

La Figura 16 muestra una tabla que representa las secuencias de ventana y las ventanas de transformación;

10 La Figura 17a muestra una representación esquemática de una transición de ventana de audio en una forma de realización de la invención;

La Figura 17b muestra una tabla que representa una transición de ventana de audio en una forma de realización ampliada de acuerdo a la invención; y

15 La Figura 18 muestra un flujo de proceso para obtener valores de ganancia de dominio de predicción lineal $g[k]$ dependiendo de un coeficiente codificado de filtro LPC.

Descripción Detallada de la Forma de Realización

20 1. Codificador para señal de audio de acuerdo con la Figura 1

[0045] A continuación se describirá un codificador para señal de audio de acuerdo con una forma de realización de la invención tomando como referencia la Figura 1, que muestra un diagrama esquemático en bloques de dicho codificador multimodo para señal de audio 100. El codificador multimodo para señal de audio 100 algunas veces también se designa en forma abreviada como codificador de audio.

[0046] El codificador de audio 100 está configurado para recibir una representación de entrada 110 de un contenido de audio, cuya representación de entrada 100 es típicamente una representación de dominio de tiempo. El codificador de audio 100 proporciona, sobre la base del mismo, una representación codificada del contenido de audio. Por ejemplo, el codificador de audio 100 proporciona una corriente de bits 112, que es una representación codificada de audio.

[0047] El codificador de audio 100 comprende un convertidor de dominio de tiempo a dominio de frecuencia 120, que está configurado para recibir la representación de entrada 110 del contenido de audio, o una versión pre-procesada de la misma 110'. El convertidor de dominio de tiempo a dominio de frecuencia 120 proporciona, sobre la base de la representación de entrada 110, 110', una representación de dominio de frecuencia 122 del contenido de audio. La representación de dominio de frecuencia 122 puede adoptar la forma de una secuencia de conjuntos de coeficientes espectrales. Por ejemplo, el convertidor de dominio de tiempo a dominio de frecuencia puede ser un convertidor de dominio de tiempo a dominio de frecuencia sobre la base de ventanas, que proporciona un primer conjunto de coeficientes espectrales sobre la base de las muestras de dominio de tiempo de un primer cuadro del contenido de entrada de audio, y para proporcionar un segundo conjunto de coeficientes espectrales sobre la base de las muestras de dominio de tiempo de un segundo cuadro del contenido de entrada de audio. El primer cuadro del contenido de entrada de audio se puede solapar, por ejemplo, en aproximadamente un 50%, con el segundo cuadro del contenido de entrada de audio. Se puede aplicar una división en ventanas de dominio de tiempo para obtener el primer conjunto de coeficientes espectrales a partir del primer cuadro de audio, y también se puede aplicar una división en ventanas para obtener el segundo conjunto de coeficientes espectrales a partir del segundo cuadro de audio. Así, el convertidor de dominio de tiempo a dominio de frecuencia puede estar configurado para llevar a cabo transformaciones solapadas de porciones divididas en ventanas (por ejemplo, cuadros de solapamiento) de la información de entrada de audio.

[0048] El codificador de audio 100 también comprende un procesador espectral 130, que está configurado para recibir la representación de dominio de frecuencia 122 del contenido de audio (o, opcionalmente, una versión post-procesada espectralmente 122' de la misma), y para proporcionar, sobre la base del mismo, una secuencia de conjuntos modelados espectralmente 132 de coeficientes espectrales. El procesador espectral 130 puede estar configurado para aplicar un modelado espectral a un conjunto 122 de coeficientes espectrales, o una versión pre-procesada 122' del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal 134 para una porción (por ejemplo, un cuadro) del contenido de audio a codificar en el modo de predicción lineal, para obtener un conjunto modelado espectralmente 132 de coeficientes espectrales. El procesador espectral 130 también puede estar configurado para aplicar un modelado espectral a un conjunto 122 de coeficientes espectrales, o a una versión pre-procesada 122' del mismo, dependiendo de un conjunto de parámetros de factores de escala 136 para una porción (por ejemplo, un cuadro) del contenido de audio a codificar en un modo de dominio de frecuencia, para obtener un conjunto modelado espectralmente 132 de coeficientes espectrales para dicha porción del contenido de audio a codificar en el modo de dominio de frecuencia. El procesador espectral 130 puede comprender, por ejemplo, un proveedor de parámetros 138, que está configurado para proporcionar el conjunto de parámetros de dominio de predicción lineal 134 y el conjunto de parámetros de factores de escala 136. Por ejemplo, el proveedor de

parámetros 138 puede proporcionar el conjunto de parámetros de dominio de predicción lineal 134 utilizando un analizador de dominio de predicción lineal, y para proporcionar el conjunto de parámetros de factores de escala 136 utilizando un procesador de modelo psicoacústico. Sin embargo también se pueden aplicar otras posibilidades para proporcionar los parámetros de dominio de predicción lineal 134 o el conjunto de parámetros de factores de escala 136.

[0049] El codificador de audio 100 también comprende un codificador de cuantificación 140, que está configurado para recibir un conjunto modelado espectralmente 132 de coeficientes espectrales (proporcionados por el procesador espectral 130) para cada porción (por ejemplo, para cada cuadro) del contenido de audio. Por otra parte, el codificador de cuantificación 140 puede recibir una versión post–procesada 132' de un conjunto modelado espectralmente 132 de coeficientes espectrales. El codificador de cuantificación 140 está configurado para proporcionar una versión codificada 142 de un conjunto modelado espectralmente coeficientes espectrales 132 (u opcionalmente, de una versión pre–procesada del mismo). El codificador de cuantificación 140 puede estar configurado, por ejemplo, para proporcionar una versión codificada 142 de un conjunto modelado espectralmente 132 de coeficientes espectrales para una porción del contenido de audio a codificar en el modo de predicción lineal, y también para proporcionar una versión codificada 142 de un conjunto modelado espectralmente 132 de coeficientes espectrales para una porción del contenido de audio a codificar en el modo de dominio de frecuencia. En otras palabras, el mismo codificador de cuantificación 140 puede ser utilizado para codificar conjuntos modelados espectralmente de coeficientes espectrales independientemente de si una porción del contenido de audio se codificará en el modo de predicción lineal o en el modo de dominio de frecuencia.

[0050] Asimismo, el codificador de audio 100 puede comprender opcionalmente un formateador de carga útil de corriente de bits 150, que está configurado para proporcionar la corriente de bits 112 sobre la base de las versiones codificadas 142 de los conjuntos modelados espectralmente de coeficientes espectrales. Sin embargo, el formateador de carga útil de corriente de bits 150 puede incluir naturalmente información codificada adicional en la corriente de bits 112, así como también información de configuración, información de control, etc. Por ejemplo, un codificador opcional 160 puede recibir el conjunto codificado 134 de parámetros de dominio de predicción lineal y/o el conjunto 136 de parámetros de factores de escala y proporcionar una versión codificada del mismo al formateador de carga útil de corriente de bits 150. En consecuencia, una versión codificada del conjunto 134 de parámetros de dominio de predicción lineal puede estar incluida en la corriente de bits 112 para una porción del contenido de audio a codificar en el modo de predicción lineal y una versión codificada del conjunto 136 de parámetros de factores de escala puede estar incluida en la corriente de bits 112 para una porción del contenido de audio a codificar en el dominio de frecuencia.

[0051] El codificador de audio 100 comprende además, opcionalmente, un controlador de modo 170, que está configurado para decidir si una porción del contenido de audio (por ejemplo, un cuadro del contenido de audio) debe ser codificada en el modo de predicción lineal o en el modo de dominio de frecuencia. Por esta razón, el controlador de modo 170 puede recibir la representación de entrada 110 del contenido de audio, la versión pre–procesada 110' de la misma o la representación de dominio de frecuencia 122 de la misma. El controlador de modo 170 puede utilizar, por ejemplo, un algoritmo de detección de voz para determinar las porciones tipo voz del contenido de audio y proporcionar una señal de control de modo 172 que indica codificar la porción del contenido de audio en el modo de predicción lineal en respuesta a la detección de una porción tipo voz. Por el contrario, si el controlador de modo descubre que una determinada porción del contenido de audio no es tipo voz, el controlador de modo 170 proporciona la señal de control de modo 172 de manera tal que la señal de control de modo 172 indica codificar dicha porción del contenido de audio en el modo de dominio de frecuencia.

[0052] A continuación se describirá detalladamente la funcionalidad general del codificador de audio 100. El codificador multimodo para señal de audio 100 está configurado para codificar de manera eficiente ambas porciones del contenido de audio que son tipo voz y las porciones del contenido de audio que no son tipo voz. Por esta razón, el codificador de audio 100 comprende por lo menos dos modos, a saber, el modo de predicción lineal y el modo de dominio de frecuencia. Sin embargo, el convertidor de dominio de tiempo a dominio de frecuencia 120 del codificador de audio 110 está configurado para transformar la misma representación de dominio de tiempo del contenido de audio (por ejemplo, la representación de entrada 110, o la versión pre–procesada 110' de la misma) en el dominio de frecuencia tanto para el modo de predicción lineal como para el modo de dominio de frecuencia. Sin embargo, una resolución de frecuencia de la representación de dominio de frecuencia 122 puede ser diferente para los diferentes modos de operación. La representación de dominio de frecuencia 122 no se cuantifica y codifica de inmediato, sino más bien se modela espectralmente antes de la cuantificación y la codificación. El modelado espectral se lleva a cabo de manera tal que un efecto del ruido de cuantificación introducido por el codificador de cuantificación 140 se mantiene suficientemente pequeño, con el fin de evitar distorsiones excesivas. En el modo de predicción lineal, el modelado espectral se lleva a cabo dependiendo de un conjunto 134 de parámetros de dominio de predicción lineal, que provienen del contenido de audio. En este caso, el modelado espectral se puede llevar a cabo, por ejemplo, de manera tal que los coeficientes espectrales se enfatizan (se ponderen más alto), si un coeficiente espectral correspondiente de una representación de dominio de frecuencia de los parámetros de dominio de predicción lineal comprende un valor comparativamente mayor. En otras palabras, los coeficientes espectrales de la representación de dominio de frecuencia 122 se ponderan de acuerdo con los coeficientes espectrales correspondientes de una representación de dominio espectral de los parámetros de dominio de predicción lineal. Por

consiguiente, los coeficientes espectrales de la representación de dominio de frecuencia 122, para los que el coeficiente espectral correspondiente de la representación de dominio espectral de los parámetros de dominio de predicción lineal adopta valores comparativamente mayores, se cuantifican con una resolución comparativamente más alta debido a la ponderación mayor en el conjunto modelado espectralmente 132 de coeficientes espectrales.

En otras palabras, hay porciones del contenido de audio para las que un modelado espectral de acuerdo con los parámetros de dominio de predicción lineal 134 (por ejemplo, de acuerdo con una representación espectral de dominio de los parámetros de dominio de predicción lineal 134) trae un modelado de ruido adecuado, debido a que los coeficientes espectrales de la representación de dominio de frecuencia 132, que son más sensibles con respecto al ruido de cuantificación, se ponderan más alto en el modelado espectral, de manera que en realidad se reduce el ruido de cuantificación eficaz introducido por el codificador de cuantificación 140.

[0053] En comparación, las porciones del contenido de audio, que están codificadas en el modo de dominio de frecuencia, experimentan un modelado espectral diferente. En este caso, los parámetros de factores de escala 136 se determinan, por ejemplo, utilizando un procesador de modelo psicoacústico. El procesador de modelo psicoacústico evalúa un enmascaramiento espectral y/o un enmascaramiento temporal de los componentes espectrales de la representación de dominio de frecuencia 122. Dicha evaluación del enmascaramiento espectral y del enmascaramiento temporal se utiliza para decidir cuales componentes espectrales (por ejemplo, los coeficientes espectrales) de la representación de dominio de frecuencia 122 se deberían codificar con una precisión de cuantificación muy eficaz y cuales componentes espectrales (por ejemplo, los coeficientes espectrales) de la representación de dominio de frecuencia 122 se pueden codificar con una precisión de cuantificación eficaz comparativamente baja. En otras palabras, el procesador de modelo psicoacústico puede determinar, por ejemplo, la relevancia psicoacústica de los diferentes componentes espectrales y puede indicar que los componentes espectrales psicoacústicamente menos importantes deberían ser cuantificados con baja o incluso muy baja precisión de cuantificación. En consecuencia, el modelado espectral (que es realizado por el procesador espectral 130), puede ponderar los componentes espectrales (por ejemplo, los coeficientes espectrales) de la representación de dominio de frecuencia 122 (o de la versión post-procesada 122' de la misma), de acuerdo con los parámetros de factores de escala 136 proporcionados por el procesador de modelo psicoacústico. Los componentes espectrales psicoacústicamente importantes reciben una alta ponderación en el modelado espectral, de manera que son eficazmente cuantificados con alta precisión de cuantificación por el codificador de cuantificación 140. Por lo tanto, los factores de escala pueden describir una relevancia psicoacústica de diferentes frecuencias o bandas de frecuencia.

[0054] En conclusión, el codificador de audio 100 es conmutable entre por lo menos dos modos diferentes, a saber, un modo de predicción lineal y un modo de dominio de frecuencia. Las porciones solapadas del contenido de audio se pueden codificar en modos diferentes. A tal efecto, las representaciones de dominio de frecuencia de diferentes (pero preferentemente solapadas) porciones de la misma señal de audio se utilizan cuando se codifican porciones posteriores (por ejemplo, inmediatamente posteriores) del contenido de audio en diferentes modos. Los componentes espectrales de dominio de la representación de dominio de frecuencia 122 son modelados espectralmente dependiendo de un conjunto de parámetros de dominio de predicción lineal para una porción del contenido de audio a codificar en el modo de dominio de frecuencia, y dependiendo de los parámetros de factores de escala para una porción del contenido de audio a codificar en el modo de dominio de frecuencia. Los diferentes conceptos que se utilizan para determinar un modelado espectral adecuado, que se lleva a cabo entre la conversión de dominio de tiempo a dominio de frecuencia y la cuantificación/codificación, permiten tener una eficiencia de codificación adecuada y un modelado de ruido con baja distorsión para diferentes tipos de contenidos de audio (tipo voz y no tipo voz).

2. Codificador de audio de acuerdo con la Figura 3

[0055] A continuación se describirá un codificador de audio 300 de acuerdo con otra forma de realización de la invención tomando como referencia la Figura 3. La Figura 3 muestra un diagrama esquemático en bloques de dicho codificador de codificación de audio 300. Cabe señalar que el codificador de audio 300 es una versión mejorada del codificador de audio de referencia 200, cuyo diagrama esquemático en bloques se ilustra en la Figura 2.

2.1 Codificador para señal de audio de referencia de acuerdo con la Figura 2

[0056] En otras palabras, para facilitar la comprensión del codificador de audio 300 de acuerdo con la Figura 3, primero se describirá el codificador de codificación unificada de voz y audio de referencia (codificador de USAC) 200 tomando como referencia el diagrama en bloques de las funciones del codificador de USAC, que se ilustra en la Figura. 2. El codificador de audio de referencia 200 está configurado para recibir una representación de entrada 210 de un contenido de audio, que es típicamente una representación de dominio de tiempo, y proporcionar, sobre la base del mismo, una representación codificada 212 del contenido de audio. El codificador de audio 200 comprende, por ejemplo, un conmutador o distribuidor 220, que está configurado para proporcionar la representación de entrada 210 del contenido de audio a un codificador de dominio de frecuencia 230 y/o a un codificador de dominio de predicción lineal 240. El codificador de dominio de frecuencia 230 está configurado para recibir la representación de entrada 210' del contenido de audio y proporcionar, sobre la base del mismo, una representación codificada espectral 232 y una información codificada de factores de escala 234. El codificador de dominio de predicción lineal

240 está configurado para recibir la representación de entrada 210" y proporcionar, sobre la base del mismo, una excitación codificada 242 y una información codificada de coeficiente de filtro LPC 244. El codificador de dominio de frecuencia 230 comprende, por ejemplo, un convertidor de dominio de tiempo a dominio de frecuencia de transformada coseno discreta modificada 230a, que proporciona una representación espectral 230b del contenido de audio. El codificador de dominio de frecuencia 230 también comprende un análisis psicoacústico 230c, que está configurado para analizar un enmascaramiento espectral y un enmascaramiento temporal del contenido de audio y proporcionar factores de escala 230d y la información codificada de factores de escala 234. El codificador de dominio de frecuencia 230 también comprende un escalímetro 230e, que está configurado para ajustar a escala los valores espectrales proporcionados por el convertidor de dominio de tiempo a dominio de frecuencia 230a de acuerdo con los factores de escala 230d, obteniendo así una representación espectral en escala 230f del contenido de audio. El codificador de dominio de frecuencia 230 también comprende un cuantificador 230g configurado para cuantificar la representación espectral en escala 230f del contenido de audio y un codificador de entropía 230h, configurado para codificar por entropía la representación espectral cuantificada en escala del contenido de audio proporcionado por el cuantificador 230g. El codificador de entropía 230h proporciona, en consecuencia, la representación espectral codificada 232.

[0057] El codificador de dominio de predicción lineal 240 está configurado para proporcionar una excitación codificada 242 y una información codificada del coeficiente de filtro LPC 244 sobre la base de la representación de entrada de audio 210". El codificador LPD 240 comprende un análisis de predicción lineal 240a, que está configurado para proporcionar coeficientes de filtro LPC 240b y la información codificada del coeficiente de filtro LPC 244 sobre la base de la representación de entrada 210" del contenido de audio. El codificador LPD 240 también comprende una codificación de excitación, que comprende dos ramales paralelos, a saber, un ramal TCX 250 y un ramal ACELP 260. Los ramales son intercambiables (por ejemplo, utilizando un conmutador 270), ya sea para proporcionar una transformación con código de excitación 252 o excitación codificada algebraica 262. El ramal TCX 250 comprende un filtro basado en LPC 250a, que está configurado para recibir tanto la representación de entrada 210" del contenido de audio y los coeficientes de filtro LPC 240b proporcionados por el análisis LP 240a. El filtro basado en LPC 250a proporciona una señal de salida de filtro, que puede describir un estímulo requerido por un filtro basado en LPC a fin de proporcionar una señal de salida que sea bastante similar a la representación de entrada 210" del contenido de audio. El ramal TCX también comprende una transformada coseno discreta modificada (MDCT) configurada para recibir la señal de estímulo 250d y proporcionar, sobre la base de la misma, una representación de dominio de frecuencia 250d de la señal de estímulo 250b. El ramal TCX también comprende un cuantificador 250e configurado para recibir la representación de dominio de frecuencia el 250b y proporcionar una versión cuantificada 250f de la misma. El ramal TCX también comprende un codificador de entropía 250g configurado para recibir la versión cuantificada 250f de la representación de dominio de frecuencia 250d de la señal de estímulo 250b y proporcionar, sobre la base del mismo, la señal de transformación con código de excitación 252.

[0058] El ramal ACELP 260 comprende un filtro basado en LPC 260a que está configurado para recibir los coeficientes de filtro LPC 240b proporcionados por el análisis LP 240a y también para recibir la representación de entrada 210" del contenido de audio. El filtro basado en LPC 260a está configurado para proporcionar, sobre la base del mismo, una señal de estímulo 260b, que describe, por ejemplo, un estímulo requerido por un filtro basado en LPC en el lado del decodificador con el fin de proporcionar una señal reconstruida que es bastante similar a la representación de entrada 210" del contenido de audio. El ramal ACELP 260 también comprende un codificador ACELP 260c configurado para codificar la señal de estímulo 260b utilizando un algoritmo de codificación algebraica adecuado.

[0059] Para resumir lo anterior, en un codec de audio de conmutación, como, por ejemplo, un codec de audio de acuerdo con el grupo de trabajo MPEG-D (grupo de expertos en imágenes en movimiento) de codificación unificada de voz y audio (USAC), que se describe en la referencia [1], los segmentos adyacentes de una entrada señal pueden ser procesados por diferentes codificadores. Por ejemplo, el codec de audio de acuerdo con el documento de trabajo de codificación unificada de voz y audio (USAC WD) puede conmutar entre un codificador de dominio de frecuencia basado en la así denominada codificación de audio avanzada (AAC), que se describe, por ejemplo, en la referencia [2], y los codificadores de dominio de predicción lineal (LPD), a saber, TCX y ACELP, basados en lo que se denomina AMR-WB + concepto, que se describe, por ejemplo, en la referencia [3]. El codificador de USAC está esquematizado en la Figura 2.

[0060] Se ha descubierto que el diseño de las transiciones entre los diferentes codificadores es un tema importante e incluso esencial para poder conmutar en forma ininterrumpida entre los diferentes codificadores. También se ha descubierto que es generalmente difícil lograr dichas transiciones debido a la distinta naturaleza de las técnicas de codificación reunidas en la estructura conmutada. Sin embargo, se ha descubierto que las herramientas comunes compartidas por los diferentes programadores pueden facilitar las transiciones.

[0061] Tomando como referencia ahora al codificador de audio de referencia 200 de acuerdo con la Figura 2, se puede observar que en la USAC, el codificador de dominio de frecuencia 230 calcula una transformada coseno discreta modificada (MDCT) en el dominio de la señal en tanto que el ramal de la excitación transformada por código (TCX) calcula una transformada coseno discreta modificada (MDCT 250c) en el dominio residual LPC (utilizando el residual LPC 250b). Asimismo, tanto los codificadores (es decir, el codificador de dominio de frecuencia 230 y el

ramal TCX 250) comparten el mismo tipo de banco de filtros, que se aplican en un dominio diferente. Por lo tanto, el codificador de audio de referencia 200 (que puede ser un codificador de audio USAC) no puede aprovechar al máximo las grandes propiedades de la MDCT, en particular la cancelación de solapamiento del dominio de tiempo (TDAC) cuando pasa desde un codificador (por ejemplo, el codificador de dominio de frecuencia 230) a otro codificador (por ejemplo, el codificador TCX 250).

[0062] Tomando como referencia nuevamente al codificador de audio de referencia 200 de acuerdo con la Figura 2, también se puede observar que el ramal TCX 250 y el ramal ACELP 260 comparten una herramienta de codificación de predicción lineal (LPC). Es una característica clave para ACELP, que es un codificador modelo de origen, en donde la LPC se utiliza para modelar el tracto vocal del lenguaje. Para TCX, la LPC se utiliza para modelar el ruido de cuantificación introducido en los coeficientes MDCT 250d. Se lleva a cabo mediante filtración (por ejemplo, utilizando el filtro basado en LPC 250a) en el dominio de tiempo de la señal de entrada 210" antes de llevar a cabo la MDCT 250c. Por otra parte, la LPC se utiliza dentro de TCX durante las transiciones a la ACELP obteniendo una señal de excitación alimentada en la codificación adaptativa de ACELP. Además permite obtener conjuntos de LPC interpolados de coeficientes para el próximo cuadro ACELP.

2.2 Codificador para señal de audio de acuerdo con la Figura 3

[0062] A continuación se describirá el codificador para señal de audio 300 de acuerdo con la Figura 3. Con este fin se hará referencia al codificador para señal de audio de referencia 200 de acuerdo con la Figura 2, ya que el codificador para señal de audio 300 de acuerdo con la Figura 3 tiene algunas similitudes con el codificador para señal de audio 200 de acuerdo con la Figura 2.

[0063] El codificador para señal de audio 300 está configurado para recibir una representación de entrada 310 de un contenido de audio, y proporcionar, sobre la base del mismo, una representación codificada 312 del contenido de audio. El codificador para señal de audio 300 está configurado para ser conmutable entre un modo de dominio de frecuencia, en el que una representación codificada de una porción del contenido de audio es proporcionada por el codificador de dominio de frecuencia 230, y un modo de predicción lineal en el que una representación codificada de una porción del contenido de audio es proporcionada por el codificador de dominio de predicción lineal 340. Las porciones del contenido de audio codificadas en diferentes modos pueden ser solapadas en algunas formas de realización, y pueden ser no ser solapadas en otras formas de realización.

[0064] El codificador de dominio de frecuencia 330 recibe la representación de entrada 310' del contenido de audio para una porción del contenido de audio a codificar en el modo de dominio de frecuencia y proporciona, sobre la base del mismo, una representación codificada espectral 332. El codificador de dominio de predicción lineal 340 recibe la representación de entrada 310" del contenido de audio para una porción del contenido de audio a codificar en el modo de predicción lineal y proporciona, sobre la base del mismo, una excitación codificada 342. El conmutador de 320 se puede utilizar, opcionalmente, para proporcionar la representación de entrada 310 al codificador de dominio de frecuencia 330 y/o al codificador de dominio de predicción lineal 340.

[0065] El codificador de dominio de frecuencia también proporciona una información codificada de factores de escala 334. El codificador de dominio de predicción lineal 340 proporciona una información codificada de coeficiente de filtro LPC 344.

[0066] El multiplexor en el lado de salida 380 es configurado para proporcionar, como la representación codificada 312 del contenido de audio, la representación espectral codificada 332 y la información codificada de factores de escala 334 para una porción del contenido de audio a codificar en el dominio de frecuencia y proporcionar, como la representación codificada 312 del contenido de audio, la excitación codificada 342 y la información codificada del coeficiente de filtro LPC 344 para una porción del contenido de audio a codificar en el modo de predicción lineal.

[0067] El codificador de dominio de frecuencia 330 comprende una transformada coseno discreta modificada 330a, que recibe la representación de dominio de tiempo 310' del contenido de audio y transforma la representación de dominio de tiempo 310' del contenido de audio, para obtener una representación de dominio de frecuencia MDCT transformada 330b del contenido de audio. El codificador de dominio de frecuencia 330 también comprende un análisis psicoacústico 330c, que está configurado para recibir la representación de dominio de tiempo 310' del contenido de audio y proporcionar, sobre la base del mismo, factores de escala 330d y la información codificada de factores de escala 334. El codificador de dominio de frecuencia 330 también comprende un combinador 330e configurado para aplicar los factores de escala 330e a la representación de dominio de frecuencia MDCT transformada 330d del contenido de audio, con el fin de ajustar a escala los diferentes coeficientes espectrales de la representación de dominio de frecuencia MDCT transformada 330b del contenido de audio con diferentes valores de factores de escala. En consecuencia se obtiene una versión modelada espectralmente 330f de la representación de dominio de frecuencia MDCT transformada 330d del contenido de audio, en donde el modelado espectral se lleva a cabo dependiendo de los factores de escala 330d, en donde las regiones espectrales, a las que están asociadas factores de escala comparativamente grandes 330e, están enfatizadas sobre regiones espectrales a las que están asociadas factores de escala comparativamente más pequeños 330e. El codificador de dominio de frecuencia 330 también comprende un cuantificador configurado para recibir la versión graduada (modelada espectralmente) 330f

de la representación de dominio de frecuencia MDCT transformada 330b del contenido de audio, y para proporcionar una versión cuantificada 330h de la misma. El codificador de dominio de frecuencia 330 también comprende un codificador de entropía 330i configurado para recibir la versión cuantificada 330h y proporcionar, sobre la base del mismo, la representación espectral codificada 332. El cuantificador 330g y el codificador de entropía 330i se pueden considerar un codificador de cuantificación.

[0068] El codificador de dominio de predicción lineal 340 comprende una ramal TCX 350 y una ramal ACELP 360. Además, el codificador LPD 340 comprende un análisis de LP 340a, que es comúnmente utilizado por el ramal TCX 350 y el ramal ACELP 360. El análisis de LP 340^a proporciona coeficientes de filtro LPC 340b y la información codificada del coeficiente de filtro LPC 344.

[0069] El ramal TCX 350 comprende una transformada MDCT 350a, que está configurada para recibir, como una entrada de transformada MDCT, la representación de dominio de tiempo 310". Es importante destacar que, la MDCT 330a del codificador de dominio de frecuencia y la MDCT 350a del ramal TCX 350 reciben (diferentes) porciones de la misma representación de dominio de tiempo del contenido de audio como señales de entrada de transformada.

[0070] En consecuencia, si porciones posteriores y solapadas (por ejemplo, cuadros) del contenido de audio están codificadas en diferentes modos, la MDCT 330a del codificador de dominio de frecuencia 330 y la MDCT 350a del ramal TCX 350 pueden recibir representaciones de dominio de tiempo que tienen un solapamiento temporal como señales de entrada de transformada. En otras palabras, la MDCT 330a del codificador de dominio de frecuencia 330 y la MDCT 350a del ramal TCX 350 reciben señales de entrada de transformada que están "en el mismo dominio", es decir, que son ambas señales de dominio de tiempo que representan el contenido de audio. Esto difiere del codificador de audio 200, en donde la MDCT 230a del codificador de dominio de frecuencia 230 recibe una representación de dominio de tiempo del contenido de audio, mientras que la MDCT 250c del ramal TCX 250 recibe una representación de dominio de tiempo residual de una señal o señal de excitación 250b de la señal, pero no una representación de dominio de tiempo del contenido de audio como tal.

[0071] El ramal TCX 350 comprende además un transformador de coeficiente de filtro 350b, que está configurado para transformar los coeficientes de filtro LPC 340b en el dominio espectral, para obtener valores de ganancia 350c. El transformador de coeficiente de filtro 350b a veces también se denomina "convertidor de predicción lineal a MDCT". El ramal TCX 350 también comprende un combinador 350d, que recibe la representación MDCT transformada del contenido de audio y los valores de ganancia 350c y proporciona, sobre la base del mismo, una versión modelada espectralmente 350e de la representación MDCT transformada del contenido de audio. Con este fin, el combinador 350d pondera coeficientes espectrales de la representación MDCT transformada del contenido de audio dependiendo de los valores de ganancia 350c con el fin de obtener la versión modelada espectralmente 350e. El ramal TCX 350 también comprende un cuantificador 350f que está configurado para recibir la versión modelada espectralmente 350e de la representación MDCT transformada del contenido de audio y para proporcionar una versión cuantificada 350g de la misma. El ramal TCX 350 también comprende un codificador de entropía 350h, que está configurado para proporcionar una versión codificada por entropía (por ejemplo, aritméticamente codificada) de la representación cuantificada 350g como la excitación codificada 342.

[0072] El ramal ACELP comprende un filtro basado en LPC 360a, que recibe los coeficientes de filtro LPC 340b proporcionados por el análisis LP 340a y la representación de dominio de tiempo 310" del contenido de audio. El filtro basado en LPC 360a adopta la misma funcionalidad que el filtro basado en LPC 260a y proporciona una señal de excitación 360b, que es equivalente a la señal de excitación 260b. El ramal ACELP 360 también comprende un codificador ACELP 360c, que es equivalente al codificador ACELP 260c. El codificador ACELP 360c proporciona una excitación codificada 342 para una porción del contenido de audio a codificar utilizando el modo ACELP (que es un sub-modo del modo de predicción lineal).

[0073] En cuanto a la funcionalidad general del codificador de audio 300, se puede decir que una porción del contenido de audio puede ser codificada ya sea en el modo de dominio de frecuencia, en el modo TCX (que es un primer sub-modo del modo de predicción lineal) o en el modo ACELP (que es un segundo sub-modo del modo de predicción lineal). Si una porción del contenido de audio está codificada en el modo de dominio de frecuencia o en el modo TCX, la porción del contenido de audio se transforma primero en el dominio de frecuencia utilizando la MDCT 330a del codificador de dominio de frecuencia o la MDCT 350a del ramal TCX. Tanto la MDCT 330a como la MDCT 350a operan en la misma representación de dominio de tiempo del contenido de audio, e incluso operan, al menos parcialmente, en porciones idénticas del contenido de audio cuando hay una transición entre el modo de dominio de frecuencia y el modo TCX. En el modo de dominio de frecuencia, el modelado espectral de la representación de dominio de frecuencia proporcionada por el transformador MDCT 330a se lleva a cabo dependiendo de los factores de escala proporcionado por el análisis psicoacústico 330c, y en el modo TCX, el modelado espectral de la representación de dominio de frecuencia proporcionada por la MDCT 350a se lleva a cabo dependiendo de los coeficientes de filtro LPC proporcionados por el análisis LP 340a. La cuantificación 330g puede ser similar a, o incluso idéntica a, la cuantificación 350f, y la codificación por entropía 330i de puede ser similar, o incluso idéntica a, la codificación por entropía 350h. Además, la transformada MDCT 330a puede ser similar, o incluso idéntica a, la transformada MDCT 350a. Sin embargo, las diferentes dimensiones de la transformada MDCT se pueden utilizar en los codificadores de dominio de frecuencia 330 y el ramal TCX 350.

[0074] Asimismo, se puede observar que los coeficientes de filtro LPC 340b son utilizados tanto por el ramal TCX 350 como por el ramal ACELP 360. Esto facilita transiciones entre las porciones del contenido de audio codificadas en el modo TCX y porciones del contenido de audio codificadas en el modo ACELP.

[0075] Para resumir lo anterior, una forma de realización de la presente invención consiste en llevar a cabo, en el contexto de la codificación de voz y audio (USAC), la MDCT 350a del TCX en el dominio de tiempo y aplicar el filtrado basado en LPC en el dominio de frecuencia (combinador 350d). El análisis LPC (por ejemplo, el análisis LP 340a) se realiza del modo anterior (por ejemplo, como en el codificador para señal de audio 200), y los coeficientes (por ejemplo, los coeficientes 340b) siguen siendo transmitidos como de costumbre (por ejemplo, en la forma de coeficientes codificados de filtro LPC 344). Sin embargo, el modelado de ruido ya no se realiza aplicando un filtro en el dominio de tiempo, sino aplicando una ponderación en el dominio de frecuencia (que es llevada a cabo, por ejemplo, el combinador 350d). El modelado de ruido en el dominio de frecuencia se logra convirtiendo los coeficientes LPC (por ejemplo, los coeficientes del filtro LPC 340b) en el dominio MDCT (que puede ser llevado a cabo por el transformador de coeficientes del filtro 350b). Para obtener más información, se hace referencia a La Figura 3, que ilustra el concepto de aplicación del modelado de ruido basado en LPC de TCX en el dominio de frecuencia.

2.3 Información sobre el Cálculo y la Aplicación de los Coeficientes LPC

[0076] A continuación se describirán el cálculo y la aplicación de los coeficientes LPC. En primer lugar se calcula un conjunto adecuado de coeficientes LPC para la presente ventana TCX, por ejemplo, utilizando el análisis LPC 340a. Una ventana TCX puede ser una porción particionada en ventanas de la representación de dominio de tiempo del contenido de audio, que se ha de codificar en el modo TCX. Las ventanas de análisis LPC están ubicadas en los límites extremos de los cuadros del codificador LPC, como se ilustra en la Figura 4.

[0077] Tomando como referencia la Figura 4 se ilustra un cuadro de TCX, es decir, un cuadro de audio a codificar en el modo TCX. Una abscisa 410 describe el tiempo, y una ordenada 420 describen los valores de magnitud de una función de la ventana.

[0078] Se lleva a cabo una interpolación para calcular el conjunto LPC de coeficientes 340b correspondientes al baricentro de la ventana de TCX. La interpolación se lleva a cabo en la frecuencia espectral de impedancia (dominio ISF), donde los coeficientes LPC suelen ser cuantificados y codificados. Los coeficientes interpolados luego se centran el medio de la ventana de TCX de $\text{sizeR} + \text{sizeM} + \text{sizeL}$.

[0079] Para más información se hace referencia a la Figura 4, que muestra una ilustración de la interpolación de coeficientes LPC para una ventana de TCX.

[0080] Los coeficientes LPC interpolados luego son ponderados como ocurre en TCX (para mayor información, véase la referencia [3]), para obtener un modelado de ruido adecuado en línea con el factor psicoacústico. Los coeficientes LPC ponderados e interpolados obtenidos (también denominados en forma abreviada `lpc_coeffs`) finalmente se convierten a factores de escala MDCT (también denominados valores de ganancia de modo de predicción lineal), utilizando un procedimiento, un pseudocódigo ilustrado en las Figuras 5 y 6.

[0081] La Figura 5 muestra un pseudocódigo de programa de una función "LPC2MDCT" para proporcionar factores de escala MDCT ("`mdct_scaleFactors`") sobre la base de coeficientes LPC de entrada ("`lpc_coeffs`"). Como se puede observar, la función "LPC2MDCT" recibe, como variables de entrada, los coeficientes LPC "`lpc_coeffs`", un valor de orden LPC "`lpc_order`" y valores de tamaño de la ventana "`sizeR`", "`sizeM`", "`sizeL`". En un primer paso, las entradas de una matriz "`InRealData [i]`" se llenan con una versión modulada de los coeficientes LPC, tal como se muestra en el número de referencia 510. Como se puede observar, las entradas de la matriz "`InRealData`" y las entradas de la matriz "`InImagData`" que tienen índices entre 0 y `lpc_order - 1` se proporcionan en los valores determinados por el correspondiente coeficiente LPC "`lpcCoeffs[i]`", modulado por un término coseno o un término seno. Las entradas de la matriz "`InRealData`" e "`InImagData`" que tienen índices $i \geq \text{lpc_order}$ se fijan en 0.

[0082] En consecuencia, las matrices "`InRealData [i]`" e "`InImagData [i]`" describen una parte real y una parte imaginaria de una respuesta de dominio de tiempo descrita por los coeficientes LPC, modulados con un término complejo de modulación ($\cos(i \cdot \pi/\text{sizeN}) - j \cdot \sin(i \cdot \pi/\text{sizeN})$).

[0083] A continuación se aplica una forma compleja de la transformada rápida de Fourier, en donde las matrices "`InRealData [i]`" e "`InImagData [i]`" describen la señal de entrada de la forma compleja de la transformada rápida de Fourier. Un resultado de la forma compleja de la transformada rápida de Fourier es proporcionado por las matrices "`OutRealData`" y "`OutImagData`". Por lo tanto, las matrices "`OutRealData`" y "`OutImagData`" describen coeficientes espectrales (que tienen índices de frecuencia i) que representan la respuesta del filtro LPC descrito por los coeficientes de filtro del dominio de tiempo.

[0084] Posteriormente se calculan los así denominados factores de escala MDCT, que tienen índices de frecuencia i , y que se denominan “mdct_scaleFactors [i]”. Un factores de escala MDCT “mdct_scaleFactors [i]” se calcula como la inversa del valor absoluto del coeficiente espectral correspondiente (descrito por las entradas “OutRealData[i]” y “OutImagData[i]”).

[0085] Cabe señalar que la operación de modulación de valores complejos ilustrada en el número de referencia 510 y la ejecución de una forma compleja de transformada rápida de Fourier ilustrada en el número de referencia 520 constituyen efectivamente una transformada discreta de Fourier impar (ODFT). La transformada discreta de Fourier impar tiene la siguiente fórmula:

$$X_0(k) = \sum_{n=0}^{n=N} x(n) e^{-j \frac{2\pi}{N} \left(k + \frac{1}{2}\right) n},$$

en donde $N = \text{sizeN}$, que es dos veces el tamaño de la MDCT

[0086] En la fórmula anterior, los coeficientes LPC $\text{lpc_coeffs}[n]$ adoptan el rol de la función de entrada de la transformada $x(n)$. La función de salida $X_0(k)$ está representada por los valores “OutRealData[k]” (parte real) y “OutImagData[k]” (parte imaginaria).

[0087] La función “complex_fft()” es una implementación rápida de una forma compleja convencional de transformada discreta de Fourier (DFT). Los factores de escala MDCT obtenidos (“mdct_scaleFactors”) son valores positivos que luego se utilizan para ajustar a escala los coeficientes MDCT (proporcionados por la MDCT 350a) de la señal de entrada. El ajuste de escala se llevará a cabo de acuerdo con el pseudocódigo ilustrado en la Figura 6.

2.4 Información sobre la División en Ventanas y el Solapamiento

[0088] La división en ventanas y el solapamiento entre los cuadros siguientes se describen en la Figura 7 y 8.

[0089] La Figura 7 ilustra una división en ventanas que es llevada a cabo por un codec conmutado de dominio de tiempo/dominio de frecuencia que envía el LPC0 como cabecera. La Figura 8 ilustra una división en ventanas que es llevada a cabo cuando se conmuta desde un codificador de dominio de frecuencia a un codificador de dominio de tiempo utilizando “lpc2mdct” para las transiciones.

[0090] Tomando como referencia ahora a la Figura 7, un primer cuadro de audio 710 está codificado en el modo de dominio de frecuencia y dividido en ventanas utilizando una ventana 712.

[0091] El segundo cuadro de audio 716, que solapa el primer cuadro de audio 710 en aproximadamente un 50%, y que está codificado en el modo de dominio de frecuencia, es dividido en ventanas utilizando una ventana 718, la cual se denomina “ventana de inicio”. La ventana de inicio tiene una pendiente de transición larga en el lado izquierdo 718a y una pendiente de transición corta en el lado derecho 718c.

[0092] Un tercer cuadro de audio 722, que está codificado en el modo de predicción lineal, está dividido en ventanas utilizando una ventana de modo de predicción lineal 724, que comprende una pendiente de transición corta en el lado izquierdo 724a que coincide con la pendiente de transición en el lado derecho 718c y una pendiente de transición corta en el lado derecho 724c. Un cuarto cuadro de audio 728, que está codificado en el modo de dominio de frecuencia, es dividido utilizando una “ventana de parada” 730 que tiene una pendiente de transición comparativamente corta en el lado izquierdo 730a y una pendiente de transición comparativamente larga en el lado derecho 730c.

[0093] Cuando se lleva a cabo una transición desde el modo de dominio de frecuencia al modo de predicción lineal, es decir, una transición entre el segundo cuadro de audio 716 y el tercer cuadro de audio 722, generalmente se envía un conjunto adicional de coeficientes LPC (también denominados “LPC0”) para asegurar una transición adecuada hacia el modo de codificación del dominio de predicción lineal.

[0094] Sin embargo, una forma de realización de acuerdo con la invención crea un codificador de audio que tiene un nuevo tipo de ventana de inicio para la transición entre el modo de dominio de frecuencia y el modo de predicción lineal. Tomando como referencia ahora a la Figura 8, se puede observar que un primer cuadro de audio 810 es dividido en ventanas utilizando la así denominada “ventana larga” 812 y codificado en el modo de dominio de frecuencia. La “ventana larga” 812 comprende una pendiente de transición comparativamente larga en el lado derecho 812b. Un segundo cuadro de audio 816 es dividido en ventanas utilizando una ventana de inicio del dominio de predicción lineal 818, que comprende una pendiente de transición comparativamente larga en el lado izquierdo 818a, que coincide con la pendiente de transición en el lado derecho 812b de la ventana 812. La ventana de inicio del dominio de predicción lineal 818 también comprende una pendiente de transición comparativamente corta en el

lado derecho 818b. El segundo cuadro de audio 816 está codificado en el modo de predicción lineal. En consecuencia, los coeficientes de filtro LPC están determinados para el segundo cuadro de audio 816, y las muestras de dominio de tiempo del segundo cuadro de audio 816 también son transformadas en la representación espectral, utilizando una MDCT. Los coeficientes de filtro LPC, que se han determinado para el segundo cuadro de audio 816, luego se aplican en el dominio de frecuencia y se utilizan para modelar espectralmente los coeficientes espectrales proporcionados por la MDCT sobre la base de la representación de dominio de tiempo del contenido de audio.

[0095] Un tercer cuadro de audio 822 es dividido en ventanas utilizando una ventana 824, que es idéntica a la ventana 724 descrita anteriormente. El tercer cuadro de audio 822 está codificado en el modo de predicción lineal. Un cuarto cuadro de audio 828 es dividido en ventanas utilizando una ventana 830, que es sustancialmente idéntica a la ventana 730.

[0096] El concepto descrito con referencia a la Figura 8 trae la ventaja de que una transición entre el cuadro de audio 810, que está codificado en el modo de dominio de frecuencia utilizando la así denominada "ventana larga" y un tercer cuadro de audio 822, que está codificado en el modo de predicción lineal utilizando la ventana 824, se lleva a cabo a través de un segundo cuadro de audio intermediario 816 (que se solapa parcialmente), que está codificado en el modo de predicción lineal utilizando la ventana 818. Ya que el segundo cuadro de audio está típicamente codificado de manera tal que el modelado espectral se lleva a cabo en el dominio de frecuencia (es decir, utilizando el transformador de coeficiente de filtro 350b), se puede obtener un solapamiento y agregado adecuado entre el cuadro de audio 810 codificado en el modo de dominio de frecuencia utilizando una ventana que tiene una pendiente de transición comparativamente larga en el lado derecho 812b y el segundo cuadro de audio 816. Adicionalmente, los coeficientes codificados de filtro LPC se transmiten para el segundo cuadro de audio 816 en lugar de los valores de factores de escala. Esto distingue la transición de la Figura 8 de la transición de la Figura 7, en donde se transmiten coeficientes adicionales de LPC (LPC0) además de los valores de factores de escala. En consecuencia, la transición entre el segundo cuadro de audio 816 y el tercer cuadro de audio 822 se puede llevar a cabo con buena calidad sin transmitir más datos adicionales como, por ejemplo, los coeficientes de LPC0 transmitidos en el caso de la Figura 7. Por lo tanto, la información que se requiere para inicializar el codec de dominio de predicción lineal utilizados en el tercer cuadro de audio 822 está disponible sin transmitir información adicional.

[0097] En resumen, en la forma de realización descrita con referencia a la Figura 8, la ventana de inicio del dominio de predicción lineal 818 puede utilizar un modelado de ruido a base de LPC en lugar de los factores de escala convencionales (que se transmiten, por ejemplo, para el cuadro de audio 716). La ventana de análisis LPC 818 corresponde a la ventana de inicio 718, y no es necesario enviar coeficientes LPC de configuración adicionales (como, por ejemplo, los coeficientes LPC0), como se describe en la Figura 8. En este caso, la codificación adaptativa de ACELP (que se puede utilizar para codificar por lo menos una porción del tercer cuadro de audio 822) puede ser introducida fácilmente con el residual LPC calculado de la ventana de inicio del codificador de dominio de predicción lineal decodificado 818.

[0098] Para resumir lo anterior, la Figura 7 ilustra una función de un codec conmutado de dominio de tiempo/dominio de frecuencia que necesita enviar un conjunto adicional de coeficientes LPC fijados denominados LP0 como cabecera. La Figura 8 ilustra una conmutación desde un codificador de dominio de frecuencia a un codificador de dominio de predicción lineal utilizando lo que se denomina "LPC2MDCT" para las transiciones.

3. Codificador para señal de audio de acuerdo con la Figura 9

[0099] A continuación se describirá un codificador para señal de audio 900 tomando como referencia la Figura 9, que está adaptada para poner en práctica el concepto descrito con referencia a la Figura 8. El codificador para señal de audio 900 de acuerdo con la Figura 9 es muy similar a la señal de audio 300 de acuerdo con la Figura 3, de manera tal que los mismos medios y señales se designan con los mismos números de referencia. En el presente contexto se omitirá un análisis de dichos medios y señales idénticos y se hará referencia al análisis del codificador para señal de audio 300.

[0100] Sin embargo, el codificador para señal de audio 900 está ampliado en comparación con el codificador para señal de audio 300 debido a que el combinador 330e del codificador de dominio de frecuencia 930 puede aplicar selectivamente los factores de escala 340d o los valores de ganancia del dominio de predicción lineal 350c para el modelado espectral. Con este fin se utiliza un conmutador 930j, que permite introducir ya sea los factores de escala 330d o los valores de ganancia del dominio de predicción lineal 350c en el combinador 330e para el modelado espectral de los coeficientes espectrales 330b. Por lo tanto, el codificador para señal de audio 900 incluso conoce tres modos de operación, a saber:

1. Modo de dominio de frecuencia: la representación del dominio de tiempo del contenido de audio se transforma en el dominio de frecuencia utilizando la MDCT 330a y un modelado espectral se aplica a la representación del dominio de frecuencia 330b del contenido de audio dependiendo de los factores de escala 330d. Una versión codificada y cuantificada 332 de la representación de dominio de frecuencia

modelada espectralmente 330f y una información codificada de factores de escala 334 están incluidas en la corriente de bits para un cuadro de audio codificado utilizando el modo de dominio de frecuencia.

2. Modo de predicción lineal: en el modo de predicción lineal, los coeficientes del filtro LPC 340b están determinados para una porción del contenido de audio y, o bien una transformación con código de excitación (primer sub-modo) o una excitación codificada de ACELP se determinan utilizando dichos coeficientes de filtro LPC 340b, dependiendo de la excitación codificada en la que parece tener una tasa de bits más eficiente. La excitación codificada 342 y la información codificada del coeficiente de filtro LPC 344 están incluidas en la corriente de bits para un cuadro de audio codificado en el modo de predicción lineal.

3. Modo de dominio de frecuencia con modelado espectral basado en los coeficientes de filtro LPC: como alternativa, en un tercer modo posible, el contenido de audio puede ser procesado por el codificador de dominio de frecuencia 930. Sin embargo, en lugar de los factores de escala 330d, los valores de ganancia del dominio de predicción lineal 350c se aplican para el modelado espectral en el combinador 330e. En consecuencia, una versión cuantificada y codificada por entropía 332 de la representación del dominio de frecuencia modelada espectralmente 330f del contenido de audio está incluida en la corriente de bits, en donde la representación del dominio de frecuencia modelada espectralmente 330f está modelada espectralmente de acuerdo con los valores de ganancia del dominio de predicción lineal 350c proporcionados por el codificador del dominio de predicción lineal 340. Adicionalmente, una información codificada del coeficiente de filtro LPC 344 está incluida en la corriente de bits para dicho cuadro de audio.

[0101] Al utilizar el tercer modo descrito, es posible lograr la transición que se ha descrito con referencia a la Figura 8 para el segundo cuadro de audio 816. Cabe señalar aquí que la codificación de un cuadro de audio utilizando el codificador de dominio de frecuencia 930 con una configuración espectral dependiendo de los valores de ganancia del dominio de predicción lineal es equivalente a la codificación del cuadro de audio 816 utilizando un codificador de dominio de predicción lineal si la dimensión de la MDCT utilizada por el codificador de dominio de frecuencia 930 corresponde a la dimensión de la MDCT utilizada por el ramal TCX 350, y si la cuantificación 330g utilizada por el codificador de dominio de frecuencia 930 corresponde a la cuantificación 350f utilizada por el ramal TCX 350 y si la codificación por entropía 330e utilizada por el codificador de dominio de frecuencia corresponde a la codificación por entropía 350h utilizada en el ramal TCX. En otras palabras, la codificación del cuadro de audio 816 se puede llevar a cabo ya sea adaptando el ramal TCX 350, de manera que la MDCT 350g adopte las características de la MDCT 330a, y de manera que la cuantificación 350f adopte las características de la cuantificación 330e y de manera que la codificación por entropía 350h adopte las características de codificación por entropía 330i, o aplicando los valores de ganancia del dominio de predicción lineal 350c en el codificador de dominio de frecuencia 930. Ambas soluciones son equivalentes y dan lugar al procesamiento de la ventana de inicio 816 como se describe con referencia a la Figura 8.

4. Decodificador para Señal de Audio de acuerdo con la Figura 10

[0102] A continuación se describirá una vista unificada de la USAC (codificación unificada de voz y audio) con TCX MDCT llevada a cabo en el dominio de señal tomando como referencia la Figura 10.

[0103] Cabe señalar aquí que en algunas formas de realización de acuerdo con la invención del ramal TCX 350 y el codificador de dominio de frecuencia 330, 930 comparten casi todas las mismas herramientas de codificación (MDCT 330a, 350a; combinador 330e, 350d, cuantificación 330g, 350f; codificador por entropía 330i, 350h) y se pueden considerar un codificador único, como se describe en la Figura 10. Por lo tanto, las formas de realización de acuerdo con la presente invención permiten una estructura más unificada de la USAC de codificador conmutado, en donde sólo dos tipos de codecs (codificador de dominio de frecuencia y codificador de dominio de tiempo) pueden estar delimitados.

[0104] Tomando como referencia ahora a la Figura 10, se puede observar que el codificador para señal de audio 1000 está configurado para recibir una representación de entrada 1010 del contenido de audio y proporcionar, sobre la base del mismo, una representación codificada 1012 del contenido de audio. La representación de entrada 1010 del contenido de audio, que suele ser una representación de dominio de tiempo, se introduce en una MDCT 1030a si una porción del contenido de audio se ha de codificar en el modo de dominio de frecuencia o en un sub-modo TCX del modo de predicción lineal. La MDCT 1030a proporciona una representación de dominio de frecuencia 1030b de la representación de dominio de tiempo 1010. La representación de dominio de frecuencia 1030b se introduce en un combinador 1030e, que combina la representación de dominio de frecuencia 1030b con los valores de modelado espectral 1040, para obtener una versión modelada espectralmente 1030f de la representación de dominio de frecuencia 1030b. La representación modelada espectralmente 1030f se cuantifica utilizando un cuantificador 1030g, para obtener una versión cuantificada 1030h de la misma, y la versión cuantificada 1030h se envía a un codificador por entropía (por ejemplo, el codificador aritmético) 1030i. El codificador por entropía 1030i proporciona una representación cuantificada y codificada por entropía de la representación de dominio de frecuencia modelada espectralmente 1030f, cuya representación cuantificada y codificada se designa con el número de referencia 1032. La MDCT 1030a, el combinador 1030e, el cuantificador 1030g y el codificador por entropía 1030i formar una ruta de

procesamiento de señal común para el modo de dominio de frecuencia y el sub–modo TCX del modo de predicción lineal.

5 **[0105]** El codificador para señal de audio 1000 comprende una ruta de procesamiento de señal ACELP 1060, que también recibe la representación de dominio de tiempo 1010 del contenido de audio y que proporciona, sobre la base del mismo, una excitación codificada 1062 utilizando una información de coeficiente de filtro LPC 1040b. La ruta de procesamiento de señal ACELP 1060, que se puede considerar como opcional, comprende un filtro basado en LPC 1060a, que recibe la representación de dominio de tiempo 1010 del contenido de audio y proporciona una señal residual o señal de excitación 1060b al codificador ACELP 1060c. El codificador ACELP proporciona la
10 excitación codificada 1062 sobre la base de la señal de excitación o señal residual 1060b.

15 **[0106]** El codificador para señal de audio 1000 también comprende un analizador de señal común 1070 que está configurado para recibir la representación de dominio de tiempo 1010 del contenido de audio y proporcionar, sobre la base del mismo, la información de modelado espectral 1040a y la información de filtro del coeficiente de filtro LPC 1040b, como así también una versión codificada de la información lateral necesaria para decodificar un cuadro de audio actual. Por lo tanto, el analizador de señal común 1070 proporciona la información de modelado espectral 1040a utilizando un análisis psicoacústico 1070a si el cuadro de audio actual está codificado en el modo de dominio de frecuencia, y proporciona una información codificada de factores de escala si el cuadro de audio actual está codificado en el modo de dominio de frecuencia. La información de factores de escala que se utiliza para el
20 modelado espectral, es proporcionada por el análisis psicoacústico 1070a, y una información codificada de factores de escala 1070b está incluida en la corriente de bits 1012 para un cuadro de audio codificado en el modo de dominio de frecuencia.

25 **[0107]** Para un cuadro de audio codificado en el sub–modo TCX del modo de predicción lineal, el analizador de señal común 1070 obtiene la información de modelado espectral 1040a a través de un análisis de predicción lineal 1070c. El análisis de predicción lineal 1070c tiene como resultado un conjunto de coeficientes de filtro LPC, que se transforman en una representación espectral mediante el bloque predicción lineal a MDCT 1070d. En consecuencia, la información de modelado espectral 1040a se obtiene a partir de los coeficientes de filtro LPC proporcionados por el análisis LP 1070c como se mencionó anteriormente. Por consiguiente, para un cuadro de audio codificado en el
30 sub–modo de transformación con código de excitación del modo de predicción lineal, el analizador de señal común 1070 proporciona información de modelado espectral 1040a sobre la base del análisis de predicción lineal 1070c (y no sobre la base del análisis psicoacústico 1070a) y también proporciona una información codificada de coeficiente de filtro LPC en lugar de una información codificada de factores de escala, para inclusión en la corriente de bits 1012.

35 **[0108]** Asimismo, para un cuadro de audio a codificar en el sub–modo ACELP del modo de predicción lineal, el análisis de predicción lineal 1070c del analizador de señal común 1070 proporciona la información del coeficiente de filtro LPC 1040b al filtro basado en LPC 1060a del ramal de procesamiento de señal ACELP 1060. En este caso, el analizador de señal común 1070 proporciona una información codificada del coeficiente de filtro LPC para inclusión en la corriente de bits 1012.
40

45 **[0109]** Para resumir lo anterior se utiliza la misma ruta de procesamiento de señal para el modo de dominio de frecuencia y para el sub–modo TCX del modo de predicción lineal. Sin embargo, la división en ventanas aplicada antes o en combinación con la MDCT y la dimensión de la MDCT 1030a pueden variar dependiendo del modo de codificación. Sin embargo, el modo de dominio de frecuencia y el sub–modo TCX del modo de predicción lineal se diferencian en que una información codificada de factores de escala está incluida en la corriente de bits en el modo de dominio de frecuencia, mientras que una información codificada de coeficiente de filtro LPC está incluida en la corriente de bits en el modo de predicción lineal.

50 **[0110]** En el sub–modo ACELP del modo de predicción lineal, una excitación codificada de ACELP y una información codificada de coeficiente de filtro LPC están incluidas en la corriente de bits.

5. Decodificador para Señal de audio de acuerdo con la Figura 11

55 5.1. Perspectiva General del Decodificador

[0111] A continuación se describirá un decodificador para señal de audio que es capaz de decodificar la representación codificada de un contenido de audio proporcionado por el codificador para señal de audio descrito anteriormente.
60

[0112] El decodificador para señal de audio 1100 de acuerdo con la Figura 11 está configurado para recibir la representación codificada 1110 de un contenido de audio y proporciona, sobre la base del mismo, una representación decodificada 1112 del contenido de audio. El codificador para señal de audio 1110 comprende un desformatador de carga útil de corriente de bits opcional 1120 que está configurado para recibir una corriente de bits que comprende la representación codificada 1110 del contenido de audio y para extraer la representación codificada del contenido de audio de dicha corriente de bits obteniendo, de ese modo, una representación codificada
65

extraída 1110' del contenido de audio. El desformateador de carga útil de corriente opcional 1120 puede extraer de la corriente de bits una información codificada de escala de factor, una información codificada de coeficiente de filtro LPC e información de control adicional o información lateral de mejoramiento de la señal.

5 **[0113]** El decodificador para señal de audio 1100 también comprende un determinador de valor espectral 1130 que está configurado para obtener una pluralidad de conjuntos 1132 de coeficientes espectrales decodificados para una pluralidad de porciones (por ejemplo, cuadros de audio solapados o no solapados) del contenido de audio. Los conjuntos de coeficientes espectrales decodificados pueden ser pre-procesados opcionalmente utilizando un pre-procesador 1140 proporcionando, de este modo, conjuntos pre-procesados 1132' de coeficientes espectrales decodificados.

15 **[0114]** El decodificador para señal de audio 1100 también comprende un procesador espectral 1150 configurado para aplicar un modelado espectral a un conjunto 1132 de coeficientes espectrales decodificados, o a una versión pre-procesada 1132" del mismo, dependiendo de un conjunto 1152 de parámetros de dominio de predicción lineal para una porción del contenido de audio (por ejemplo, un cuadro de audio) codificado en un modo de predicción lineal, y para aplicar un modelado espectral a un conjunto 1132 de coeficientes espectrales decodificados, o a una versión pre-procesada 1132" del mismo, dependiendo de un conjunto 1154 de parámetros de factores de escala para una porción del contenido de audio (por ejemplo, un cuadro de audio) codificado en un modo de dominio de frecuencia. En consecuencia, el procesador espectral 1150 obtiene conjuntos modelados espectralmente 1158 de coeficientes espectrales decodificados.

25 **[0115]** El decodificador para señal de audio 1100 también comprende un convertidor de dominio de frecuencia a dominio de tiempo 1160, que está configurado para recibir un conjunto modelado espectralmente 1158 de coeficientes espectrales decodificados y para obtener una representación de dominio de tiempo 1162 del contenido de audio sobre la base del conjunto modelado espectralmente 1158 de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de predicción lineal. El convertidor de dominio de frecuencia a dominio de tiempo 1160 también está configurado para obtener una representación de dominio de tiempo 1162 del contenido de audio sobre la base de un conjunto modelado espectralmente respectivo 1158 de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de dominio de frecuencia.

35 **[0116]** El decodificador para señal de audio 1100 también comprende un procesador de dominio de tiempo opcional 1170 que lleva a cabo, opcionalmente, un procesamiento posterior de dominio de tiempo de la representación de dominio de tiempo 1162 del contenido de audio, para obtener la representación decodificada 1112 del contenido de audio. Sin embargo, a falta del post-procesador de dominio de tiempo 1170, la representación decodificada 1112 del contenido de audio puede ser igual a la representación de dominio de tiempo 1162 del contenido de audio proporcionado por el convertidor de dominio de frecuencia a dominio de tiempo 1160.

5.2 Información Adicional

40 **[0117]** A continuación se describirán más detalles relacionados con el decodificador de audio 1100, cuya información puede ser considerada como mejoras opcionales del decodificador para señal de audio.

45 **[0118]** Cabe señalar que el decodificador para señal de audio 1100 es un decodificador multimodo para señal de audio, que es capaz de administrar una representación codificada de señal de audio en la que las porciones posteriores (por ejemplo, los cuadros de audio solapados o no solapados) del contenido de audio se codifican utilizando diferentes modos. A continuación los cuadros de audio se considerarán como un ejemplo simple de una porción del contenido de audio. Debido a que el contenido de audio está subdividido en cuadros de audio, es particularmente importante contar con transiciones suaves entre representaciones decodificadas de posteriores cuadros de audio (parcialmente solapados o no solapados) codificados en el mismo modo, y también entre los cuadros de audio posteriores (solapados o no solapados) codificados en diferentes modos. Preferentemente, el decodificador para señal de audio 1100 administra representaciones de señal de audio en las que cuadros de audio posteriores están solapados en aproximadamente un 50%, a pesar de que el solapamiento puede ser significativamente menor en algunos casos y/o para algunas transiciones.

55 **[0119]** Por esta razón, el decodificador para señal de audio 1100 comprende un solapador configurado para solapar y agregar las representaciones de dominio de tiempo de cuadros de audio posteriores codificados en diferentes modos. El solapador puede, por ejemplo, formar parte del convertidor de dominio de frecuencia a dominio de tiempo 1160, o puede estar dispuesto en la salida del convertidor de dominio de frecuencia a dominio de tiempo 1160. Con el fin de obtener una alta eficiencia y una calidad adecuada cuando se solapan cuadros de audio posteriores, el convertidor de dominio de frecuencia a dominio de tiempo está configurado para obtener una representación de dominio de tiempo de un cuadro de audio codificado en el modo de predicción lineal (por ejemplo, en el sub-modo de transformación con código de excitación del mismo) utilizando una transformación solapada, y también para obtener una representación de dominio de tiempo de un cuadro de audio codificado en el modo de dominio de frecuencia utilizando una transformación solapada. En este caso, el solapador está configurado para solapar representaciones de dominio de tiempo de los cuadros de audio posteriores codificados en diferentes modos. Al

utilizar dichas transformaciones solapadas de síntesis para las conversiones de dominio de frecuencia a dominio de tiempo, que puede ser preferentemente del mismo tipo de transformación para los cuadros de audio codificados en diferentes modos, se puede utilizar un muestreo crítico y se reduce al mínimo la cabecera causada por la operación de solapar y agregar. Al mismo tiempo, existe una cancelación de solapamiento del dominio de tiempo entre las porciones solapadas de las representaciones de dominio de tiempo de los cuadros de audio posteriores. Cabe señalar que la posibilidad de tener una cancelación de solapamiento del dominio de tiempo en la transición entre cuadros de audio posteriores codificados en diferentes modos es causada por el hecho de que una conversión de dominio de frecuencia a dominio de tiempo es aplicada en el mismo dominio en modos diferentes, de modo que una salida de una transformación solapada de síntesis llevada a cabo en un conjunto modelado espectralmente de coeficientes espectrales decodificados de un primer cuadro de audio codificado en un primer modo se puede combinar directamente (es decir, combinar sin una operación de filtrado intermedio) con una salida de una transformación solapada llevada a cabo en un conjunto modelado espectralmente de coeficientes espectrales decodificados de un cuadro de audio posterior codificado en un segundo modo. Por lo tanto se lleva a cabo una combinación lineal de la salida de la transformación solapada llevada a cabo para un cuadro de audio codificado en el primer modo y de la salida de la transformación solapada para un cuadro de audio codificado en el segundo modo. Naturalmente se puede llevar a cabo una división en ventanas de solapamiento adecuado como parte del proceso de transformación solapada o posterior al proceso de transformación solapada.

[0120] En consecuencia, la cancelación de solapamiento del dominio de tiempo se obtiene por la simple operación de solapar y agregar entre las representaciones de dominio de tiempo de cuadros de audio posteriores codificados en diferentes modos.

[0121] En otras palabras, es importante que el convertidor de dominio de frecuencia a dominio de tiempo 1160 proporcione señales de salida de dominio de tiempo, que estén en el mismo dominio para los dos modos. El hecho de que las señales de salida de la conversión de dominio de frecuencia a dominio de tiempo (por ejemplo, la transformación solapada en combinación con una división en ventanas de transición asociada) se encuentren en el mismo dominio para diferentes modos significa que las señales de salida de la conversión de dominio de frecuencia a dominio de tiempo son linealmente combinables incluso en una transición entre diferentes modos. Por ejemplo, las señales de salida de la conversión de dominio de frecuencia a dominio de tiempo son ambas representaciones de dominio de tiempo de un contenido de audio que describen una evolución temporal de una señal de altavoz. En otras palabras, las representaciones de dominio de tiempo 1162 del contenido de audio de cuadros de audio posteriores pueden ser comúnmente procesadas a fin de derivar las señales de altavoz.

[0122] Por otra parte, cabe señalar que el procesador espectral 1150 puede comprender un proveedor de parámetros 1156 que está configurado para proporcionar el conjunto 1152 de parámetros de dominio de predicción lineal y el conjunto 1154 de parámetros de factores de escala sobre la base de la información extraída de la corriente de bits 1110, por ejemplo, sobre la base de una información codificada de factores de escala y una información codificada de parámetros de filtro LPC. El proveedor de parámetros 1156 puede comprender, por ejemplo, un determinador de coeficientes de filtro LPC configurado para obtener coeficientes decodificados de filtro LPC en la base a una representación codificada de los coeficientes de filtro LPC para una porción del contenido de audio codificado en el modo de predicción lineal. Además, el proveedor de parámetros 1156 puede comprender un transformador de coeficientes de filtro configurado para transformar los coeficientes decodificados de filtro LPC en una representación espectral, con el fin de obtener valores de ganancia de modo de predicción lineal asociados a diferentes frecuencias. Los valores de ganancia del modo de predicción lineal (a veces también denominados $g[k]$) pueden constituir un conjunto 1152 de parámetros de dominio de predicción lineal.

[0123] El proveedor de parámetros 1156 puede comprender además un determinador de factores de escala configurado para obtener valores decodificados de factores de escala sobre la base de una representación codificada de los valores de factores de escala para un cuadro de audio codificado en el modo de dominio de frecuencia. Los valores decodificados de factores de escala pueden servir como un conjunto 1154 de parámetros de factores de escala.

[0124] En consecuencia, el modelado espectral, que puede considerarse como una modificación del espectro, está configurado para combinar un conjunto 1132 de coeficientes espectrales decodificados asociados a un cuadro de audio codificado en el modo de predicción lineal, o una versión pre-procesada 1132' del mismo, con valores de ganancia del modo de predicción lineal (que constituyen el conjunto 1152 de parámetros de dominio de predicción lineal), con el fin de obtener una versión procesada de ganancia (es decir, modelada espectralmente) 1158 de los coeficientes espectrales decodificados 1132 en los que las contribuciones de los coeficientes espectrales decodificados 1132, o de la versión pre-procesada 1132' del mismo, se ponderan dependiendo de los valores de ganancia de modo de predicción lineal. Además, el modificador espectral puede estar configurado para combinar un conjunto 1132 de coeficientes espectrales decodificados asociados a un cuadro de audio codificado en el modo de dominio de frecuencia, o una versión pre-procesada 1132' del mismo, con los valores de factores de escala (que constituyen el conjunto 1154 de parámetros de factores de escala) con el fin de obtener una versión procesada de factores de escala (es decir, modelada espectralmente) 1158 de los coeficientes espectrales decodificados 1132 en los que las contribuciones de los coeficientes espectrales decodificados 1132, o de la versión pre-procesada 1132' del mismo, se ponderan dependiendo de los valores de factores de escala (del conjunto 1154 de parámetros de

factores de escala). En consecuencia, un primer tipo de modelado espectral, es decir, un modelado espectral dependiendo de un conjunto 1152 de parámetros de dominio de predicción lineal, se lleva a cabo en el modo de predicción lineal, y un segundo tipo de modelado espectral, es decir, un modelado espectral dependiendo de un conjunto 1154 de parámetros de factores de escala, se lleva a cabo en el modo de dominio de frecuencia. En consecuencia, un impacto perjudicial del ruido de cuantificación en la representación de dominio de tiempo 1162 se mantiene pequeño, tanto para los cuadros de audio tipo voz (en los que el modelado espectral se lleva a cabo preferentemente dependiendo del conjunto 1152 de parámetros de dominio de predicción lineal) y para el audio en general, por ejemplo, cuadros de audio no tipo voz para los que el modelado espectral se lleva a cabo preferentemente dependiendo del conjunto 1154 de parámetros de factores de escala. Sin embargo, al llevar a cabo el modelado de ruido utilizando el modelado espectral tanto para cuadros de audio tipo voz como no tipo voz, es decir, tanto para cuadros de audio codificados en el modo de predicción lineal como para cuadros de audio codificados en el modo de dominio de frecuencia, el decodificador multimodo de audio 1100 comprende una estructura de baja complejidad y al mismo tiempo permite un solapamiento y agregado con cancelación de solapamiento de las representaciones de dominio de tiempo 1162 de cuadros de audio codificados en diferentes modos.

[0125] A continuación se analizarán otros detalles de información.

6. Decodificador para Señal de Audio de acuerdo con la Figura 12

[0126] La Figura 12 muestra un diagrama esquemático en bloques de un decodificador para señal de audio 1200, de acuerdo con una forma de realización adicional de la invención. La Figura 12 ilustra una vista unificada de un decodificador de codificación unificada de voz y audio (USAC) con una transformación con código de excitación-transformada coseno discreta modificada (TCX-MDCT) en el dominio de la señal.

[0127] El decodificador para señal de audio 1200 de acuerdo con la Figura 12 comprende un demultiplexor de corriente de bits 1210 que puede adoptar la función del desformateador de carga útil de corriente de bits 1120. El demultiplexor de corriente de bits 1210 extrae, de una corriente de bits que representa un contenido de audio, una representación codificada del contenido de audio, que puede comprender valores codificados espectrales e información adicional (por ejemplo, una información codificada de factores de escala y una información codificada de parámetro de filtro LPC).

[0128] El decodificador para señal de audio 1200 también comprende conmutadores 1216, 1218, que están configurados para distribuir los componentes de la representación codificada del contenido de audio proporcionada por el demultiplexor de corriente de bits a los diferentes bloques de procesamiento de componentes del decodificador para señal de audio 1200. Por ejemplo, el decodificador para señal de audio 1200 comprende un ramal combinado de modo de dominio de frecuencia/sub-modo de TCX 1230, que recibe desde el conmutador 1216 una representación codificada de dominio de frecuencia 1228 y proporciona, sobre la base del mismo, una representación de dominio de tiempo 1232 del contenido de audio. El decodificador para señal de audio 1200 también comprende un decodificador ACELP 1240 que está configurado para recibir desde el conmutador 1216 una información de excitación codificada ACELP 1238 y proporcionar, sobre la base del mismo, una representación de dominio de tiempo 1242 del contenido de audio.

[0129] El decodificador para señal de audio 1200 también comprende un proveedor de parámetros 1260 que está configurado para recibir desde el conmutador 1218 una información codificada de factores de escala 1254 para un cuadro de audio codificado en el modo de dominio de frecuencia y una información codificada de coeficiente de filtro LPC 1256 para un cuadro de audio codificado en el modo de predicción lineal, que comprende el sub-modo TCX y el sub-modo ACELP. El proveedor de parámetro 1260 se configura más para recibir la información de control 1258 del conmutador 1218. El proveedor de parámetros 1260 está configurado para proporcionar una información de modelado espectral 1262 para el ramal combinado de modo dominio de frecuencia/sub-modo TCX 1230. Además, el proveedor de parámetros 1260 está configurado para proporcionar una información de coeficiente de filtro LPC 1264 al decodificador ACELP 1240.

[0130] El ramal combinado de modo de dominio de frecuencia/sub-modo TCX 1230 puede comprender un decodificador por entropía 1230a, que recibe la información codificada de dominio de frecuencia 1228 y proporciona, sobre la base del mismo, una información decodificada de dominio de frecuencia 1230b, que se introduce en un cuantificador inverso 1230c. El cuantificador inverso 1230c proporciona, sobre la base de información decodificada de dominio de frecuencia 1230b, una información decodificada e inversamente cuantificada de dominio de frecuencia 1230d, por ejemplo, en forma de conjuntos de coeficientes espectrales decodificados. Un combinador 1230e está configurado para combinar la información decodificada e inversamente cuantificada de dominio de frecuencia 1230d con la información de modelado espectral 1262, para obtener la información modelada espectralmente de dominio de frecuencia 1230f. Una transformada inversa coseno discreta modificada 1230g recibe la información modelada espectralmente de dominio de frecuencia 1230f y proporciona, sobre la base del mismo, la representación de dominio de tiempo 1232 del contenido de audio.

[0131] El decodificador por entropía 1230a, el cuantificador inverso 1230c y la transformada inversa coseno discreta modificada 1230g, pueden todos recibir opcionalmente cierta información de control, que puede estar incluida en la corriente de bits o provenir de la corriente de bits mediante el proveedor de parámetros 1260.

5 **[0132]** El proveedor de parámetros 1260 comprende un decodificador de factores de escala 1260a, que recibe la información codificada de factores de escala 1254 y proporciona una información decodificada de factores de escala 1260b. El proveedor de parámetros 1260 comprende también un decodificador de coeficiente LPC 1260c, que está configurado para recibir la información codificada del coeficiente de filtro LPC 1256 y proporcionar, sobre la base del mismo, una información decodificada de coeficiente de filtro LPC 1260d a un transformador de coeficiente de filtro
10 1260e. Asimismo, el decodificador de coeficiente LPC 1260c proporciona la información de coeficiente de filtro LPC 1264 al decodificador ACELP 1240. El transformador de coeficiente de filtro 1260e está configurado para transformar los coeficientes del filtro LPC 1260d en el dominio de frecuencia (también denominado dominio espectral) y para derivar posteriormente los valores de ganancia del modo de predicción lineal 1260f desde los coeficientes del filtro LPC 1260d. Además, el proveedor de parámetros 1260 está configurado para proporcionar selectivamente, por
15 ejemplo utilizando un conmutador 1260g, los factores de escala decodificados 1260b o los valores de ganancia del modo de predicción lineal 1260f como la información de modelado espectral 1262.

[0133] Cabe señalar aquí que el codificador para señal de audio 1200 de acuerdo con la Figura 12 puede estar complementado por un número de pasos adicionales de pre-procesamiento y pasos de post-procesamiento recorridos entre las etapas. Los pasos de pre-procesamiento y los pasos de post-procesamiento pueden ser diferentes para los diferentes modos.

[0134] A continuación se describirán algunos detalles de información.

25 7. Flujo de Señales de acuerdo con la Figura 13

[0135] A continuación se describirá un flujo posible de señales tomando como referencia la Figura 13. El flujo de señales 1300 de acuerdo con la Figura 13 puede ocurrir en el decodificador para señal de audio 1200 de acuerdo con la Figura 12.

30 **[0136]** Cabe señalar que el flujo de señales 1300 de la Figura 13 sólo describe la operación en el modo de dominio de frecuencia y el sub-modo TCX del modo de predicción lineal con el fin de simplificar la explicación. Sin embargo, la descodificación en el sub-modo ACELP del modo de predicción lineal se puede llevar a cabo como se describió con referencia a la Figura 12.

35 **[0137]** El ramal común de modo de dominio de frecuencia/sub-modo TCX 1230 recibe la información codificada del dominio de frecuencia 1228. La información codificada del dominio de frecuencia 1228 puede comprender lo que se denomina datos espectrales codificados aritméticamente "ac_spectral_data", que se extraen de una corriente de canal de dominio de frecuencia ("fd_channel_stream") en el modo de dominio de frecuencia. La información codificada del dominio de frecuencia 1228 puede comprender lo que se denomina una codificación TCX ("tcx_coding"), que puede ser extraída de una corriente de canal de dominio de predicción lineal ("lpd_channel_stream") en el sub-modo TCX. Una decodificación por entropía 1330a puede ser llevada a cabo por el decodificador por entropía 1230a. Por ejemplo, la decodificación por entropía 1330a se puede llevar a cabo utilizando un decodificador aritmético. Por lo tanto, los coeficientes espectrales cuantificados "x_ac_quant" se obtienen para cuadros de audio codificados en el dominio de frecuencia, y los coeficientes espectrales cuantificados en el modo de TCX "x_tcx_quant" se obtienen para los cuadros de audio codificados en el modo TCX. Los coeficientes espectrales cuantificados en el modo de dominio de frecuencia y los coeficientes espectrales cuantificados en el modo TCX pueden ser números enteros en algunas formas de realización. La decodificación por entropía puede, por ejemplo, decodificar en forma conjunta grupos de coeficientes espectrales codificados de una manera sensible al contexto. Por otra parte, el número de bits necesarios para codificar un determinado coeficiente espectral puede variar dependiendo de la magnitud de los coeficientes espectrales, de manera tal que más bits de palabras clave son necesarios para codificar un coeficiente espectral que tiene una magnitud comparativamente más grande.

50 **[0138]** Posteriormente, la cuantificación inversa 1330c de los coeficientes espectrales cuantificados del modo de dominio de frecuencia y de los coeficientes espectrales cuantificados del modo TCX se llevará a cabo, por ejemplo, utilizando el cuantificador inverso 1230c. La cuantificación inversa puede estar descrita por la siguiente fórmula:

$$x_invquant = \text{Sign}(x_quant) \cdot |x_quant|^{\frac{4}{3}}$$

60 **[0139]** En consecuencia, los coeficientes espectrales inversamente cuantificados del modo de dominio de frecuencia ("x_ac_invquant") se obtienen para los cuadros de audio codificados en el modo de dominio de frecuencia, y los

coeficientes espectrales inversamente cuantificados del modo TCX (“x_tcx_invquant”) se obtienen para los cuadros de audio codificados en el sub–modo TCX.

7.1 Procesamiento para Cuadro de Audio Codificado en el Dominio de Frecuencia

5 **[0140]** A continuación se resumirá el procesamiento en el modo de dominio de frecuencia. En el modo de dominio de frecuencia, un relleno de ruido 1340 se aplica opcionalmente a los coeficientes espectrales inversamente cuantificados del modo de dominio de frecuencia para obtener una versión rellena de ruido 1342 de los coeficientes espectrales inversamente cuantificados del modo de dominio de frecuencia 1330d (“x_ac_invquant”). A continuación se puede llevar a cabo un ajuste de escala de la versión rellena de ruido 1342 de los coeficientes espectrales inversamente cuantificados del modo de dominio de frecuencia, en donde el ajuste de escala es designada con el número de referencia 1344. En el ajuste de escala, los parámetros de factores de escala (también brevemente designados como factores de escala o sf[g][sfb]) se aplican para ajustar a escala los coeficientes espectrales inversamente cuantificados del modo de dominio de frecuencia 1342 (“x_ac_invquant”). Por ejemplo, los diferentes factores de escala pueden estar asociados a los coeficientes espectrales de diferentes bandas de frecuencia (rangos de frecuencia o bandas de factores de escala). Por consiguiente, los coeficientes espectrales inversamente cuantificados 1342 pueden ser multiplicados con factores de escala asociados para obtener los coeficientes espectrales graduados 1346. El ajuste de escala 1344 se puede llevar a cabo preferentemente como se describe en la norma internacional ISO/IEC 14496–3, subparte 4, sub–cláusulas 4.6.2 y 4.6.3. El ajuste de escala 1344 puede llevarse a cabo, por ejemplo, utilizando el combinador 1230e. Por consiguiente se obtiene una versión graduada (y, en consecuencia, modelada espectralmente) 1346, “x_rescal” de los coeficientes espectrales del modo de dominio de frecuencia, que puede ser equivalente a la representación del dominio de frecuencia 1230f. Posteriormente, una combinación de un procesamiento medio/lateral 1348 y de un procesamiento de modelado de ruido temporal 1350 se puede llevar a cabo opcionalmente sobre la base de la versión graduada 1346 de los coeficientes espectrales del modo de dominio de frecuencia, para obtener una versión post–procesada 1352 de los coeficientes espectrales graduados del modo de dominio de frecuencia 1346. El procesamiento medio/lateral opcional 1348 puede llevarse a cabo, por ejemplo, como se describe en la norma ISO/IEC 14496–3:2005, tecnología de la información de codificación de objetos audiovisuales – parte 3: Audio, subparte 4, sub–cláusula 4.6.8.1. El modelado de ruido temporal opcional se puede llevar a cabo como se describe en la norma ISO/IEC 14496–3: 2005, tecnología de la información de codificación de objetos audiovisuales – parte 3: Audio, subparte 4, sub–cláusula 4.6.9.

35 **[0141]** Posteriormente se puede aplicar una transformada inversa coseno discreta modificada 1354 a la versión graduada 1346 de los coeficientes espectrales del modo de dominio de frecuencia o a la versión post–procesada 1352 del mismo. Por lo tanto se obtiene una representación de dominio de tiempo 1356 del contenido de audio del cuadro de audio actualmente procesado. La representación de dominio de tiempo 1356 también es designada como $x_{i,n}$. Como una simple presunción, se puede suponer que existe una representación de dominio de tiempo $x_{i,n}$ por cuadro de audio. Sin embargo, en algunos casos, en los que varias ventanas (por ejemplo, las que se denominan “ventanas cortas”) están relacionadas con un único cuadro de audio, puede haber una pluralidad de representaciones de dominio de tiempo $x_{i,n}$ por cuadro de audio.

45 **[0142]** Posteriormente se aplica una división en ventanas 1358 a la representación de dominio de tiempo 1356, para obtener una representación de dominio de tiempo dividida en ventanas 1360, la que también se designa como $z_{i,n}$. Por lo tanto, en un caso simplificado, en el que hay una ventana por cuadro de audio, se obtiene una representación de dominio de tiempo dividida en ventanas 1360 por cuadro de audio codificado en el modo de dominio de frecuencia.

7.2. Procesamiento para Cuadro de Audio Codificado en el Modo TCX

50 **[0143]** A continuación se describirá el procesamiento para un cuadro de audio codificado totalmente o parcialmente en el modo TCX. En lo que concierne a este tema, cabe señalar que un cuadro de audio puede estar dividido en una pluralidad de, por ejemplo, cuatro subcuadros, que puede estar codificados en diferentes sub–modos del modo de predicción lineal. Por ejemplo, los subcuadros de un cuadro de audio pueden estar codificados selectivamente en el sub–modo TCX del modo de predicción lineal o en el sub–modo ACELP del modo de predicción lineal. En consecuencia, cada uno de los subcuadros puede estar codificados de manera tal que se obtiene un rendimiento de codificación óptimo o un equilibrio óptimo entre la calidad de audio y la tasa de bits. Por ejemplo, una señalización que utiliza una matriz denominada “mod[]” pueden estar incluida en la corriente de bits para un cuadro de audio codificado en el modo de predicción lineal para indicar cuáles de los subcuadros de dicho cuadro de audio están codificados en el sub–modo TCX y cuáles están codificados en el sub–modo ACELP. Sin embargo, cabe señalar que el presente concepto se puede entender más fácilmente si se supone que todo el cuadro está codificado en el modo TCX. Los otros casos, en los que un cuadro de audio comprende ambos subcuadros TCX deberían considerarse como una ampliación opcional de dicho concepto.

65 **[0144]** Suponiendo ahora que todo el cuadro está codificado en el modo TCX, se puede observar que un relleno de ruido 1370 se aplica a los coeficientes espectrales inversamente cuantificados del modo TCX 1330d, que también son designados como “quant[]”. En consecuencia se obtiene un conjunto relleno de ruido de coeficientes espectrales del modo TCX 1372, que es también designado como “r[i]”. Además, lo que se denomina desmodelado espectral

1374 se aplica al conjunto relleno de ruido de los coeficientes espectrales del modo TCX 1372, para obtener un conjunto desmodelado espectral 1376 de los coeficientes espectrales del modo TCX, que es también designado como “ $r[i]$ ”. Posteriormente se aplica un modelado espectral 1378, en donde el modelado espectral se lleva a cabo dependiendo de los valores de ganancia del dominio de predicción lineal que provienen de los coeficientes codificados LPC que describen una respuesta de filtro de un filtro de Codificación de Predicción Lineal (LPC). El modelado espectral 1378 se puede llevar a cabo, por ejemplo, utilizando el combinador 1230a. Por consiguiente se obtiene un conjunto reconstruido 1380 de los coeficientes espectrales del modo TCX, también designado como “ $rr[i]$ ”. Posteriormente se lleva a cabo una MDCT inversa 1382 sobre la base del conjunto reconstruido 1380 de los coeficientes espectrales del modo TCX, para obtener una representación de dominio de tiempo 1384 de un cuadro (o, alternativamente, de un subcuadro) codificado en el modo TCX. Posteriormente se aplica un reajuste de escala 1386 a la representación de dominio de tiempo 1384 de un cuadro (o un subcuadro) codificado en el modo TCX, para obtener una representación de dominio de tiempo reajustada a escala 1388 del cuadro (o subcuadro) codificado en el modo TCX, en donde la representación de dominio de tiempo reajustada a escala también se designa como “ $x_w[i]$ ”. Cabe señalar que el reajuste de escala 1386 es típicamente un ajuste de escala igual de todos los valores de dominio de tiempo de un cuadro codificado en el modo TCX o del subcuadro codificado en el modo TCX. Por lo tanto, el reajuste de escala 1386 normalmente no ocasiona una distorsión de frecuencia debido a que no es selectiva de frecuencia.

[0145] Con posterioridad al reajuste de escala 1386 se aplica una división en ventanas 1390 a la representación de dominio de tiempo reajustada a escala 1388 de un cuadro (o un subcuadro) codificado en el modo TCX. En consecuencia se obtienen las muestras del dominio de tiempo divididas en ventanas 1392 (también designadas como “ $z_{i,n}$ ”, que representan el contenido de audio de un cuadro (o un subcuadro) codificado en el modo TCX.

7.3. Procesamiento de Solapar y Agregar

[0146] Las representaciones de dominio de tiempo 1360, 1392 de una secuencia de cuadros se combinan utilizando un procesamiento de solapar y agregar 1394. En el procesamiento de solapar y agregar, las muestras del dominio de tiempo de una porción de lado derecho (temporalmente posterior) de un primer cuadro de audio se solapan y agregan con muestras de dominio de tiempo de una porción de lado izquierdo (temporalmente anterior) de un segundo cuadro de audio posterior. Este procesamiento de solapar y agregar 1394 se lleva a cabo tanto para cuadros de audio posteriores codificados en el mismo modo y para cuadros de audio posteriores codificados en diferentes modos. Una cancelación de solapamiento del dominio de tiempo es llevada a cabo por el procesamiento de solapar y agregar 1394, incluso si los cuadros de audio posteriores están codificados en diferentes modos (por ejemplo, en el modo de dominio de frecuencia y en el modo TCX), debido a la estructura específica del decodificador de audio, lo que evita cualquier distorsión del proceso entre la salida de la MDCT inversa 1954 y el procesamiento de solapar y agregar 1394, y también entre la salida de la MDCT inversa 1382 y el procesamiento de solapar y agregar 1394. En otras palabras, no hay un procesamiento adicional entre el procesamiento de MDCT inversa 1354, 1382 y el procesamiento de solapar y agregar 1394 excepto por la división en ventanas 1358, 1390 y el reajuste de escala 1386 (y, opcionalmente, una combinación no espectralmente distorsionante de un filtrado de pre-énfasis y una operación desenfatizante).

8. Detalles de Información relativos a la TCX basada en MDCT

8.1. Descripción de la herramienta de TCX basada en MDCT

[0147] Cuando el modo central es un modo de predicción lineal (que está indicado por el hecho de que la variable de corriente de bits “core_mode” es igual a uno) y cuando uno o más de los tres modos TCX (por ejemplo, de un primer modo TCX para proporcionar una porción TCX de 512 muestras, incluyendo 256 muestras de solapamiento, un segundo modo TCX para proporcionar 768 muestras de dominio de tiempo, incluyendo 256 muestras de solapamiento) es seleccionado como la codificación de “dominio de predicción lineal”, es decir, si una de las cuatro entradas de matriz de “mod[x]” es mayor que cero (en donde cuatro entradas de matriz mod[0], mod[1], mod[2], mod[3] provienen de una variable de corriente de bits e indican los sub-modos de LPC para cuatro subcuadros del cuadro de audio actual, es decir, indican si un subcuadro está codificado en el sub-modo ACELP del modo de predicción lineal o en el sub-modo TCX del modo de predicción lineal, y si se utiliza una codificación de TCX comparativamente larga, una codificación de TCX de longitud mediana o una codificación de TCX de longitud corta), se utiliza la herramienta de TCX basada en MDCT. En otras palabras, si uno de los subcuadros del cuadro de audio actual está codificado en el sub-modo TCX del modo de predicción lineal, se utiliza la herramienta de TCX. La TCX basada en MDCT recibe los coeficientes espectrales cuantificados desde un decodificador aritmético (que se puede utilizar para implementar el decodificador por entropía 1230a o la decodificación por entropía 1330a). Los coeficientes cuantificados (o una versión inversamente cuantificada 1230b de los mismos) se completan primero con un ruido de confort (que se puede llevar a cabo mediante la operación de relleno de ruido 1370). El modelado de ruido de dominio de frecuencia basado en LPC luego se aplica a los coeficientes espectrales resultantes (por ejemplo, utilizando el combinador 1230e, o la operación de modelado espectral 1378) (o a una versión de desmodelado espectral del mismo), y una transformación MDCT inversa (que puede ser implementada por la MDCT

1230g o por la operación de MDCT inversa 1382) se lleva a cabo para obtener la señal de síntesis del dominio de tiempo.

8.2. Definiciones de TCX basada en MDCT

[0148] A continuación se proporcionarán algunas definiciones.

[0149] "lg" designa un número de coeficientes espectrales cuantificados producidos por el decodificador aritmético (por ejemplo, para un cuadro de audio codificado en el modo de predicción lineal).

[0150] La variable de corriente de bits "noise_factor" designa un índice de cuantificación de nivel de ruido.

[0151] La variable "nivel de ruido" designa un nivel de ruido introducido en el espectro reconstruido.

[0152] La variable "noise[]" designa un vector de ruido generado.

[0153] La variable de corriente de bits "global_gain" designa un índice de cuantificación de ganancia de reajuste de escala.

[0154] La variable "g" designa una ganancia de reajuste de escala.

[0155] La variable "rms" designa la raíz cuadrada media de la señal sintetizada de dominio de tiempo "x[]".

[0156] La variable "x[]" designa la señal sintetizada de dominio de tiempo.

8.3. Proceso de Decodificación

[0157] La TCX basada en MDCT requiere del decodificador aritmético 1230a un número de coeficientes espectrales cuantificados, lg, que es determinado por el valor mod[] (es decir, por el valor de la variable mod[]). Este valor (es decir, el valor de la variable mod []) también define la longitud y la forma de la ventana que se aplicarán en la MDCT inversa 1230g (o por el procesamiento de la MDCT inversa 1382 y la correspondiente división en ventanas 1390). La ventana se compone de tres partes, un solapamiento de lado izquierdo de las muestras L (también designado como pendiente de transición de lado izquierdo), una parte central de una de las muestras M y una parte derecha de solapamiento (también designada como pendiente de transición de lado derecho) de las muestras R. Para obtener una ventana de MDCT de 2*lg de longitud se agregan ceros ZL en el lado izquierdo y se agregan ceros ZR en el lado derecho.

[0158] En caso de una transición desde o hacia una ventana corta "short_window" la región correspondiente de solapamiento L o R puede necesitar reducirse a 128 (muestras) a fin de adaptarse a una probable pendiente de ventana más corta "short_window". En consecuencia, la región M y la región cero ZL o ZR correspondientes pueden necesitar expandirse en 64 muestras cada una.

[0159] En otras palabras, normalmente hay un solapamiento de 256 muestras = L = R. Esto se reduce a 128 en el caso del modo FD al modo LPD.

[0160] El diagrama de la Figura 15 muestra un número de coeficientes espectrales en función de mod[], así como también un número de muestras de dominio de tiempo de la región cero izquierda ZL de la región izquierda de solapamiento L, de la parte central M, de la región derecha de solapamiento R y de la región cero derecha ZR.

[0161] La ventana MDCT es la siguiente

$$W(n) = \begin{cases} 0 & \text{Para } 0 \leq n < ZL \\ W_{SIN_LEFT,L}(n - ZL) & \text{Para } ZL \leq n < ZL + L \\ 1 & \text{Para } ZL + L \leq n < ZL + L + M \\ W_{SIN_RIGHT,R}(n - ZL - L - M) & \text{Para } ZL + L + M \leq n < ZL + L + M + R \\ 0 & \text{Para } ZL + L + M + R \leq n < 2lg \end{cases}$$

[0162] Las definiciones de $W_{SIN_LEFT,L}$ y $W_{SIN_RIGHT,R}$ se indican a continuación. (N. del T.: SIN_LEFT = seno_izquierdo), SIN_RIGHT = seno_derecho).

[0163] La ventana MDCT $W(n)$ se aplica en el paso de división en ventanas 1390, que puede ser considerado como parte de una MDCT inversa dividida en ventanas (por ejemplo, de la MDCT inversa 1230g).

5 [0164] Los coeficientes espectrales cuantificados, también designados como "quant[]", proporcionados por el decodificador aritmético 1230a (o, alternativamente, por la cuantificación inversa 1230c) son completados por un ruido de confort. El nivel del ruido inyectado es determinado por la variable decodificada de corriente de bits "noise_factor" como se indica a continuación:

10
$$\text{noise_level} = 0,0625 \cdot (8 - \text{noise_factor})$$

[0165] Luego se calcula un vector de ruido, también designado como "noise[]", utilizando una función aleatoria, designada como "random_sign()", que proporciona aleatoriamente el valor -1 o +1. La relación es la siguiente:

15
$$\text{noise}[i] = \text{random_sign}() \cdot \text{noise_level};$$

[0166] Los vectores de "quant[]" y "noise[]" se combinan para formar el vector reconstruido de coeficientes espectrales, también designado como "r[]", de manera tal que las ejecuciones de 8 ceros consecutivos en "quant[]" se sustituyen por los componentes de "noise[]". Una ejecución de ocho ceros se detecta de acuerdo con la siguiente fórmula:

$$r[i] = \begin{cases} 1 & \text{for } i \in [0, \lg/6[\\ \sum_{k=0}^{\min(7, \lg - 8 \lfloor i/8 \rfloor - 1)} |\text{quant}[\lg/6 + 8 \lfloor i/8 \rfloor + k]|^2 & \text{for } i \in [0, 5 \cdot \lg/6[\end{cases}$$

20 [0167] El espectro reconstruido se obtiene de la siguiente manera:

$$r[i] = \begin{cases} \text{noise}[i] & \text{Si } r[i] = 0 \\ \text{quant}[i] & \text{En los demás casos} \end{cases}$$

25 [0168] El relleno de ruido anteriormente descrito se puede llevar a cabo como un post-procesamiento entre la decodificación por entropía llevada a cabo por el decodificador por entropía 1230a y la combinación llevada a cabo por el combinador 1230e.

[0169] Se aplica un desmodelado espectral al espectro reconstruido (por ejemplo, al espectro reconstruido 1376, r[i]) de acuerdo con los siguientes pasos:

- 30 1. calcular la energía E_m del bloque de ocho dimensiones a índice m para cada bloque de 8 dimensiones del primer cuarto del espectro
- 35 2. calcular la relación $R_m = \text{sqrt}(E_m/E_l)$, en donde l es el índice de bloque con el valor máximo de todas las E_m
3. si $R_m < 0,1$, entonces fijar $R_m = 0,1$
4. si $R_m < R_{m-1}$, entonces fijar $R_m = R_{m-1}$

40 [0170] Cada bloque de 8 dimensiones que pertenece al primer cuarto del espectro luego se multiplica por el factor R_m .

45 [0171] Se llevará a cabo un desmodelado espectral como un post-procesamiento dispuesto en una ruta de señales entre el decodificador por entropía 1230a y el combinador 1230e. El desmodelado espectral puede ser llevado a cabo, por ejemplo, por el desmodelado espectral 1374.

50 [0172] Antes de aplicar la MDCT inversa se recuperan los dos filtros LPC cuantificados correspondientes a ambos extremos del bloque MDCT (es decir, los puntos plegables izquierdo y derecho), se calculan sus versiones ponderadas y se calculan los espectros diezmados correspondientes (64 puntos, cualquiera sea la longitud de transformación).

5 **[0173]** En otras palabras, se obtiene un primer conjunto de coeficientes de filtro LPC para un primer período de tiempo y se determina un segundo conjunto de coeficientes de filtro LPC para un segundo período de tiempo. Los conjuntos de coeficientes de filtro LPC provienen preferentemente de una representación codificada de dichos coeficientes de filtro LPC, que está incluida en la corriente de bits. El primer período de tiempo es preferentemente en o anterior al comienzo del cuadro codificado actual de TCX (o subcuadro) y el segundo período de tiempo es preferentemente en o posterior al final del cuadro codificado de TCX o subcuadro. Por consiguiente, un conjunto efectivo de coeficientes de filtro LPC se determina formando un promedio ponderado de los coeficientes de filtro LPC del primer conjunto y de los coeficientes de filtro LPC del segundo conjunto.

10 **[0174]** Los espectros de LPC ponderados se calculan aplicando una transformada discreta de Fourier impar (ODFT) a los coeficientes de filtro LPC. Se aplica una modulación compleja a los coeficientes (de filtro) LPC antes de calcular la transformada discreta de Fourier impar (ODFT), de manera que los intervalos de frecuencia ODFT están (preferentemente perfectamente) alineados con los intervalos de frecuencia MDCT. Por ejemplo, el espectro de síntesis LPC ponderado de un filtro LPC determinado $\hat{A}(z)$ se calcula de la siguiente manera:

$$X_o[k] = \sum_{n=0}^{M-1} x_i[n] e^{-j \frac{2\pi k}{M} n}$$

15 con

$$x_i[n] = \begin{cases} \hat{w}[n] e^{-j \frac{\pi}{M} n} & \text{if } 0 \leq n < lpc_order + 1 \\ 0 & \text{if } lpc_order + 1 \leq n < M \end{cases}$$

en donde $\hat{w}[n]$, $n = 0 \dots lpc_order + 1$ son los coeficientes del filtro LPC ponderado según lo indicado por:

$$\hat{W}(z) = \hat{A}(z / \gamma_1) \text{ con } \gamma_1 = 0.92$$

20 **[0175]** En otras palabras, una respuesta de dominio de tiempo de un filtro LPC, representada por los valores $\hat{w}[n]$ siendo n entre 0 y $lpc_order - 1$, se transforma en el dominio espectral, para obtener los coeficientes espectrales $X_o[k]$. La respuesta del dominio de tiempo $\hat{w}[n]$ del filtro LPC puede provenir de los coeficientes de dominio de tiempo a_1 a a_{16} que describen el filtro de Codificación de Predicción Lineal.

25 **[0176]** Las ganancias $g[k]$ se pueden calcular a partir de la representación espectral $X_o[k]$ de los coeficientes LPC (por ejemplo, a_1 a a_{16}) de acuerdo con la siguiente ecuación:

$$g[k] = \sqrt{\frac{1}{X_o[k] X_o^*[k]}} \quad \forall k \in \{0, \dots, M-1\}$$

30 en donde $M=64$ es el número de bandas en las que se aplican las ganancias calculadas.

35 **[0177]** Posteriormente se obtiene un espectro reconstruido 1230f, 1380, $rr[i]$ dependiendo de las ganancias calculadas $g[k]$ (también designadas como valores de ganancia del modo de predicción lineal). Por ejemplo, un valor de ganancia $g[k]$ puede estar asociado con un coeficiente espectral 1230d, 1376, $r[i]$. Alternativamente, una pluralidad de valores de ganancia puede estar asociada con un coeficiente espectral 1230d, 1376, $r[i]$. Un coeficiente de ponderación $a[i]$ puede provenir de uno o más valores de ganancia $g[k]$, o el coeficiente de ponderación $a[i]$, puede incluso ser idéntico a un valor de ganancia $g[k]$ en algunas formas de realización. Por lo tanto, un coeficiente de ponderación $a[i]$ se puede multiplicar con un valor espectral asociado $r[i]$, para determinar un aporte del coeficiente espectral $r[i]$ al coeficiente espectral modelado espectralmente $rr[i]$.

40 **[0178]** Por ejemplo, la siguiente ecuación puede indicar:

$$rr[i] = g[k] \cdot r[i].$$

[0179] Sin embargo, también se pueden utilizar distintas relaciones.

5 [0180] En lo anterior, la variable k es igual a $i/(lg/64)$ para tener cuenta el hecho de que los espectros de LPC están diezmados. El espectro reconstruido rr[] se introduce en una MDCT inversa 1230g, 1382. Al llevar a cabo la MDCT inversa, que se describirá en forma detallada más adelante, los valores del espectro reconstruido rr[i] sirven como valores de frecuencia de tiempo $X_{i, k}$, o como valores de frecuencia de tiempo spec[i][k]. La siguiente relación puede indicar:

10 $X_{i, k} = rr[k];$

o

15 $spec[i][k] = rr[k].$

[0181] Cabe señalar aquí que en la explicación anterior del procesamiento espectral en el ramal TCX, la variable i es un índice de frecuencia. En comparación, en la explicación del banco de filtros MDCT y la conmutación en bloque, la variable i es un índice de ventana. Un experto en la materia reconocerá fácilmente a partir del contexto si la variable i es un índice de frecuencia o un índice de ventana.

[0182] Además, cabe señalar que un índice de ventana puede ser equivalente a un índice de cuadro, si un cuadro de audio comprende una sola ventana. Si un cuadro comprende varias ventanas, como ocurre algunas veces, puede haber varios valores de índice de ventana por cuadro.

25 [0183] La señal de salida que no es de ventana x[] se reajustará a escala por la ganancia g, obtenida por una cuantificación inversa del índice decodificado de ganancia global ("global_gain"):

$$g = \frac{10^{global_gain/28}}{2 \cdot rms}$$

[0184] En donde rms se calcula de la siguiente manera:

$$rms = \sqrt{\frac{\sum_{k=lg/2}^{3*lg/2-1} rr^2[k]}{L + M + R}}$$

30 [0185] La dominio de la señal de tiempo sintetizada y reajustada a escala es entonces igual a:

$$x_w[n] = x[n] \cdot g$$

[0186] Luego del reajuste de escala se aplica la división en ventanas y la operación de solapar-agregar. La división en ventanas se puede llevar a cabo utilizando una ventana W(n) tal como se describió anteriormente y teniendo en cuenta los parámetros de división en ventanas que se ilustran en la Figura 15. Por lo tanto, se obtiene una representación de dominio de la señal de tiempo dividida en ventanas $z_{i, n}$ de acuerdo con lo siguiente:

35 $z_{i, n} = x_w[n] \cdot W(n).$

40 [0187] A continuación se describirá un concepto que es útil si hay cuadros de audio codificados TCX (o subcuadros de audio) y cuadros de audio codificados ACELP (o subcuadros de audio). Además, cabe señalar que los coeficientes de filtro LPC, que son transmitidos para los cuadros codificados por TCX o subcuadros significa que algunas formas de realización se aplicarán con el fin de inicializar la decodificación ACELP.

45 [0188] Cabe apreciar también que la longitud de la síntesis TCX está determinada por la longitud del cuadro TCX (sin solapamiento): 256, 512 o 1024 muestras para [] mod de 1, 2 o 3, respectivamente.

[0189] A continuación se adopta la siguiente notación: $x[]$ designa la salida de la transformada inversa coseno discreta modificada, $z[]$ designa la señal decodificada dividida en ventanas en el dominio de tiempo y $out[]$ designa la señal sintetizada de dominio de tiempo.

- 5 **[0190]** La salida de la transformada inversa coseno discreta modificada luego se reajusta a escala y se divide en ventanas de la siguiente manera:

$$z[n] = x[n] \cdot w[n] \cdot g; \forall 0 \leq n < N$$

[0191] N corresponde al tamaño de ventana MDCT, es decir, $N=2lg$.

- 10 **[0192]** Cuando el modo de codificación anterior fue el modo FD o TCX basado en MDCT se aplica una operación de solapar y agregar convencional entre la señal decodificada dividida en ventanas actual $z_{i,n}$ y la señal decodificada dividida en ventanas anterior $z_{i-1,n}$, en donde el índice i representa el número de ventanas MDCT ya decodificadas. La síntesis de dominio de tiempo final out se obtiene a través de las fórmulas siguientes.

- 15 **[0193]** En el caso de que $z_{i-1,n}$ provenga del modo FD:

$$out[i_{out} + n] = \begin{cases} z_{i-1, \frac{N-l}{2}+n}; \forall 0 \leq n < \frac{N-l}{4} - \frac{L}{2} \\ z_{i, \frac{N-N-l}{4}+n} + z_{i-1, \frac{N-l}{2}+n}; \forall \frac{N-l}{4} - \frac{L}{2} \leq n < \frac{N-l}{4} + \frac{L}{2} \\ z_{i, \frac{N-N-l}{4}+n}; \forall \frac{N-l}{4} + \frac{L}{2} \leq n < \frac{N-l}{4} + \frac{N-R}{2} \end{cases}$$

[0194] N_{-1} es el tamaño de la secuencia de ventanas que proviene del modo FD. i_{out} indexa el registro intermedio

de salida y aumenta por el número $\frac{N-l}{4} + \frac{N-R}{2}$ de muestras escritas.

- 20 **[0195]** En el caso de que $z_{i-1,n}$ provenga de TCX basada en MDCT:

$$out[i_{out} + n] = \begin{cases} z_{i, \frac{N-L}{4}+n} + z_{i-1, \frac{3N-l-L}{4}+n}; \forall 0 \leq n < L \\ z_{i, \frac{N-L}{4}+n}; \forall L \leq n < \frac{N+L-R}{2} \end{cases}$$

[0196] N_{-1} es el tamaño de la ventana MDCT anterior. i_{out} indexa el registro intermedio de salida out y aumenta por el número $(N+L-R)/2$ de las muestras escritas.

- 25 **[0197]** A continuación se describirán algunas posibilidades para reducir los artefactos en una transición de un cuadro o un subcuadro codificado en el modo ACELP en un cuadro o un subcuadro codificado en el modo TCX basado en MDCT. Sin embargo, cabe señalar que también se pueden utilizar procedimientos diferentes.

- 30 **[0198]** A continuación se describirá brevemente un primer procedimiento. Cuando proviene de ACELP se puede utilizar una ventana específica para la siguiente TCX reduciendo R a 0, y luego eliminando la región de solapamiento entre los dos cuadros siguientes.

- 35 **[0199]** A continuación se describirá brevemente un segundo procedimiento (como se describe en USAC WD5 y anteriormente). Cuando proviene de ACELP, la siguiente ventana TCX se amplía aumentando M (longitud media) en 128 muestras. En el decodificador, la parte derecha de la ventana, es decir, las primeras muestras decodificadas R que no son cero, simplemente se eliminan y reemplazan por las muestras ACELP decodificadas.

[0200] La síntesis reconstruida $out[iout+n]$ luego se filtra a través del filtro de pre-énfasis . La síntesis pre-énfatisada resultante luego es filtrada por el filtro de análisis a fin de obtener la señal de excitación. La excitación calculada actualiza la codificación adaptativa ACELP y permite la conmutación desde TCX a ACELP en un cuadro posterior. Los coeficientes de filtro de análisis son interpolados en una base de subcuadro.

9. Detalles de información con respecto al Banco de Filtros y Conmutación de Bloque

[0201] A continuación se describirá más detalladamente la transformada inversa coseno discreta modificada y la conmutación de bloque, es decir, la operación de solapar y agregar llevada a cabo entre cuadros posteriores o subcuadros. Cabe señalar que la transformada inversa coseno discreta modificada que se describe a continuación se puede aplicar tanto para los cuadros de audio codificados en el dominio de frecuencia como para los cuadros de audio o subcuadros de audio codificados en el modo TCX. Si bien se ha descrito anteriormente las ventanas ($W(n)$) que se pueden utilizar en el modo TCX, las ventanas utilizadas para el modo de dominio de frecuencia se describirán a continuación: cabe señalar que la elección de las ventanas adecuadas, en particular, en la transición desde un cuadro codificado en el modo de frecuencia a un cuadro posterior codificado en el modo TCX, o viceversa, permite obtener una cancelación de solapamiento del dominio de tiempo, de manera tal que se pueden obtener transiciones con poco o ningún solapamiento sin sobrecarga de tasa de bits.

9.1. Banco de Filtros y Descripción de la Conmutación de Bloque

[0202] La representación de tiempo/frecuencia de la señal (por ejemplo, la representación de tiempo-frecuencia 1158, 1230f, 1352, 1380) se traza sobre el dominio de tiempo introduciéndola en el módulo de banco de filtros (por ejemplo, el módulo 1160, 1230g, 1354–1358–1394, 1382–1386–1390–1394). Este módulo consta de una transformada inversa coseno discreta modificada (IMDCT), y una ventana y una función de solapar-agregar. Con el fin de adaptar la resolución de tiempo/frecuencia del banco de filtros a las características de la señal de entrada se adopta también una herramienta de conmutación de bloque. N representa la longitud de la ventana, en donde N depende de la variable de corriente de bits "**window_sequence**". Para cada canal, los valores de frecuencia de tiempo $N/2 X_{i,k}$ se transforman en los valores de dominio de tiempo $N x_{i,n}$ a través de la IMDCT. Después de aplicar la función de ventana, para cada canal, la primera mitad de la secuencia $z_{i,n}$ se agrega a la segunda mitad de la secuencia dividida en ventanas de bloque anterior $z_{(i-1),n}$ para reconstruir las muestras de salida para cada canal $out_{i,n}$.

9.2. Banco de Filtros y Conmutación de Bloque – Definiciones

[0203] A continuación se proporcionarán algunas definiciones de las variables de corriente de bits.

[0204] La variable de corriente de bits "**window_sequence**" consta de dos bits que indican la secuencia de ventana (es decir, el tamaño del bloque) que se utiliza. La variable de corriente de bits "**window_sequence**" se utiliza típicamente para los cuadros de audio codificados en el dominio de frecuencia.

[0205] La variable de corriente de bits "**window_shape**" (forma de ventana) comprende un bit que indica la función de la ventana seleccionada.

[0206] La tabla de la Figura 16 muestra las once secuencias de ventanas (también designadas como **window_sequences**) basadas en las siete ventanas de transformación. (ONLY_LONG_SEQUENCE, LONG_START_SEQUENCE, EIGHT_SHORT_SEQUENCE, LONG_STOP_SEQUENCE, STOP_START_SEQUENCE).

[0207] A continuación, LPD_SEQUENCE se refiere a todas las combinaciones permitidas de ventanas/modo de codificación dentro de lo que se denomina codec de dominio de predicción lineal. En el contexto de la decodificación de un cuadro codificado de dominio de frecuencia sólo es importante saber si un cuadro siguiente está codificado en los modos de codificación de dominio LP, que está representado por LPD_SEQUENCE. Sin embargo, se tiene en cuenta la estructura exacta dentro de LPD_SEQUENCE cuando se decodifica el cuadro codificado en el dominio LP.

[0208] En otras palabras, un cuadro de audio codificado en el modo de predicción lineal puede comprender un solo cuadro codificado por TCX, una pluralidad de subcuadros codificados por TCX o una combinación de subcuadros codificados por TCX y subcuadros codificados por ACELP.

9.3. Banco de Filtros y Proceso de Decodificación–Conmutación de Bloque

9.3.1 Banco de Filtros y Conmutación de Bloque–IMDCT

[0209] La expresión analítica de la IMDCT es:

$$x_{i,n} = \frac{2}{N} \sum_{k=0}^{\frac{N}{2}-1} \text{spec}[i][k] \cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right) \quad \text{for } 0 \leq n < N$$

en donde:

n = índice de muestra

5

i = índice de ventana

k = índice de coeficientes espectrales

10 N = longitud de ventana basada en el valor window_sequence

$n_0 = (N / 2 + 1) / 2$

15 **[0210]** La longitud de la ventana de síntesis N para la transformada inversa depende del elemento de sintaxis “**window_sequence**” y del contexto algorítmico. Se define de la siguiente manera:

Longitud de ventana 2048:

$$N = \begin{cases} 2048, & \text{if ONLY_LONG_SEQUENCE} \\ 2048, & \text{if LONG_START_SEQUENCE} \\ 256, & \text{if EIGHT_SHORT_SEQUENCE} \\ 2048, & \text{if LONG_STOP_SEQUENCE} \\ 2048, & \text{if STOP_START_SEQUENCE} \end{cases}$$

20 (N. del T.: only long sequence = sólo secuencia larga; long start sequence = secuencia larga de inicio; eight short sequence = secuencia de ocho ventanas cortas; long stop sequence = secuencia larga de parada; stop start sequence = secuencia de inicio – de parada).

25 **[0212]** Una marca de verificación () en una determinada celda de la tabla de la Figura 17a o 17b indica que a una secuencia de ventanas que figura en esa fila en particular le puede seguir una secuencia de ventanas que figura en esa columna en particular.

30 **[0213]** En la Figura 17a se muestran transiciones de bloque significativas de una primera forma de realización. En la tabla de la Figura 17d se muestran transiciones de bloque significativas de una forma de realización adicional. Las transiciones de bloque adicionales en la forma realización de acuerdo con la Figura 17b se explicarán por separado más adelante.

9.3.2 Banco de Filtros y Conmutación de Bloque – División en Ventanas y Conmutación de Bloque

35 **[0214]** Dependiendo de las variables de corriente de bits (o elementos) se utilizan distintas ventanas “**window_sequence**” (secuencia de ventanas) y “**window_shape**” (forma de ventana) de transformación de elementos. Una combinación de mitades de ventanas que se describe a continuación proporciona todas las secuencias de ventanas posibles.

40 Para “**window_shape**” (forma de ventana) == 1, los coeficientes de ventana son proporcionados por la ventana derivada de Kaiser – Bessel (KBD), como se indica a continuación:

$$W_{KBD_LEFT,N}(n) = \sqrt{\frac{\sum_{p=0}^n [W'(p,\alpha)]}{\sum_{p=0}^{N/2} [W'(p,\alpha)]}} \quad \text{para } 0 \leq n < \frac{N}{2}$$

$$W_{KBD_RIGHT,N}(n) = \sqrt{\frac{\sum_{p=0}^{N-n-1} [W'(p,\alpha)]}{\sum_{p=0}^{N/2} [W'(p,\alpha)]}} \quad \text{para } \frac{N}{2} \leq n < N$$

(N. del T.: KBD _ LEFT = KBD _ izquierda).
 (N. del T.: KBD _ RIGHT = KBD _ derecho).

5 en donde:

W', Kaiser – la función de ventana núcleo de Bessel, véase también [5], está definida de la siguiente manera:

$$W'(n,\alpha) = \frac{I_0 \left[\pi\alpha \sqrt{1.0 - \left(\frac{n - N/4}{N/4} \right)^2} \right]}{I_0[\pi\alpha]} \quad \text{para } 0 \leq n \leq \frac{N}{2}$$

$$I_0[x] = \sum_{k=0}^{\infty} \left[\frac{\left(\frac{x}{2} \right)^k}{k!} \right]^2$$

$\alpha = \text{factor alfa de ventana kernel } \alpha = \begin{cases} 4 \text{ para } N = 2048 \text{ (1920)} \\ 6 \text{ para } N = 256 \text{ (240)} \end{cases}$

10 **[0215]** De lo contrario, para “window_shape” (forma de ventana) == 0, se utiliza una ventana seno de la siguiente manera:

$$W_{SIN_LEFT,N}(n) = \sin\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right)\right) \quad \text{para } 0 \leq n < \frac{N}{2}$$

$$W_{SIN_RIGHT,N}(n) = \sin\left(\frac{\pi}{N} \left(n + \frac{1}{2}\right)\right) \quad \text{para } \frac{N}{2} \leq n < N$$

(N. del T.: SIN_ LEFT = seno izquierdo; SIN _ RIGHT = seno derecho).

[0216] La longitud de ventana N puede ser 2048 (1920) o 256 (240) para la ventana KBD y seno.

[0217] En las partes a)–e) del presente inciso se explica cómo obtener las secuencias de ventanas posibles.

[0218] Para todo tipo de secuencias de ventanas, la variable “window_shape” (forma de ventana) de la mitad izquierda de la primera ventana de transformación está determinada por la forma de ventana del bloque anterior, descrita por la variable “window_shape_previous_block”. La fórmula siguiente expresa lo anterior:

$$W_{LEFT,N}(n) = \begin{cases} W_{KBD_LEFT,N}(n), & \text{Si "window_shape_previous_block" == 1} \\ W_{SIN_LEFT,N}(n), & \text{Si "window_shape_previous_block" == 0} \end{cases}$$

(N. del T.: KBD _ LEFT = KBD izquierdo; SIN _ LEFT = seno izquierdo).

en donde:

“Window_shape_previous_block” (forma de ventana del bloque anterior) es una variable, que es igual a la variable de corriente de bits “window_shape” (forma de ventana) del bloque anterior (i–1).

Para el primer bloque de datos en bruto “raw_data_block()” a decodificar, la variable “window_shape” (forma de ventana) de la mitad izquierda y derecha de la ventana son idénticas.

[0219] En caso de que el bloque anterior se hubiere codificado utilizando el modo LPD, “window_shape_previous_block” (forma de ventana del bloque anterior) se fija en 0.

a) **ONLY_LONG_SEQUENCE** (sólo secuencia larga) :

[0220] La secuencia de ventanas designada como **window_shape** == ONLY_LONG_SEQUENCE es igual a una ventana de tipo “LONG_WINDOW” (ventana larga) que tiene una longitud total de ventana de 2048 (1920).

[0221] Para **window_shape** (forma de ventana) == 1, la ventana para el valor de la variable “ONLY_LONG_SEQUENCE” (sólo secuencia larga) se describe de la siguiente manera:

$$W(n) = \begin{cases} W_{LEFT,N_I}(n), & \text{para } 0 \leq n < N_I/2 \\ W_{KBD_RIGHT,N_I}(n), & \text{para } N_I/2 \leq n < N_I \end{cases}$$

(N. del T.: LEFT = izquierdo; KBD _ RIGHT = KBD _ derecho).

[0222] Si **window_shape** (forma de ventana) == 0, la ventana para el valor de la variable “ONLY_LONG_SEQUENCE” (sólo secuencia larga) se puede describir de la siguiente manera:

$$W(n) = \begin{cases} W_{LEFT,N_I}(n), & \text{para } 0 \leq n < N_I/2 \\ W_{SIN_RIGHT,N_I}(n), & \text{para } N_I/2 \leq n < N_I \end{cases}$$

(N. del T.: LEFT = izquierdo; SIN _ RIGHT = seno derecho).

[0223] Luego de la división en ventanas, los valores del dominio de tiempo (z_{i,n}) puede expresarse de la siguiente manera:

$$z_{i,n} = w(n) \cdot x_{i,n};$$

b) **LONG_START_SEQUENCE** (secuencia larga de inicio) :

[0224] La ventana de tipo "LONG_START_SEQUENCE" se puede utilizar para obtener una operación correcta de solapar y agregar para una transición de bloque desde una ventana de tipo "ONLY_LONG_SEQUENCE" (sólo secuencia larga) a cualquier bloque con una mitad de ventana de bajo solapamiento (pendiente de ventana corta) a la izquierda (EIGHT_SHORT_SEQUENCE (secuencia de ocho ventanas cortas), LONG_STOP_SEQUENCE (secuencia larga de parada), STOP_START_SEQUENCE (secuencia de parada – inicio) o LPD_SEQUENCE (secuencia LPD).

[0225] En el caso de que la secuencia de ventanas siguiente no sea una ventana de tipo "LPD_SEQUENCE" (secuencia de LPD): la longitud de la ventana w y s se fija en 2048 (1920) y 256 (240), respectivamente.

[0226] En el caso de que la secuencia de ventanas siguiente sea una ventana de tipo "LPD_SEQUENCE" (secuencia de LPD): la longitud de la ventana w y s se fija en 2048 (1920) y 512 (480), respectivamente.

[0227] Si $window_shape$ (forma de ventana) == 1 la ventana para el tipo de ventana "LONG_START_SEQUENCE" (secuencia larga de inicio) es la siguiente:

$$W(n) = \begin{cases} W_{LEFT,N_l}(n), & \text{para } 0 \leq n < N_l/2 \\ 1.0, & \text{para } N_l/2 \leq n < \frac{3N_l - N_s}{4} \\ W_{KBD_RIGHT,N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{para } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

(N. del T.: LEFT = izquierdo; KBD _ RIGHT = KBD _ derecho).

[0228] Si $window_shape$ (forma de ventana) == 0, la ventana para el tipo de ventana "LONG_START_SEQUENCE" (secuencia larga de inicio) es la siguiente:

$$W(n) = \begin{cases} W_{LEFT,N_l}(n), & \text{para } 0 \leq n < N_l/2 \\ 1.0, & \text{para } N_l/2 \leq n < \frac{3N_l - N_s}{4} \\ W_{SIN_RIGHT,N_s}(n + \frac{N_s}{2} - \frac{3N_l - N_s}{4}), & \text{para } \frac{3N_l - N_s}{4} \leq n < \frac{3N_l + N_s}{4} \\ 0.0, & \text{para } \frac{3N_l + N_s}{4} \leq n < N_l \end{cases}$$

(N. del T.: LEFT = izquierdo; SIN _ RIGHT = seno _ derecho).

[0229] Los valores de dominio de tiempo divididos en ventanas se pueden calcular con la fórmula explicada en a).

c) EIGHT_SHORT (ocho ventanas cortas)

[0230] La secuencia de ventanas para $window_sequence$ (secuencia de ventana) == EIGHT_SHORT comprende ocho ventanas cortas (SHORT_WINDOW) solapadas y agregadas que tienen con una longitud w de 256 (240) cada una. La longitud total de $window_sequence$ (secuencia de ventanas) junto con los ceros anteriores y siguientes es de 2048 (1920). Cada uno de los ocho bloques cortos primero se divide en ventanas por separado. El número de bloque corto se indexa con la variable $j = 0, \dots, M - 1$ ($M = N_l / N_s$).

[0231] $Window_shape$ (forma de ventana) del bloque anterior influye en el primero de los ocho bloques cortos ($W_0(n)$) solamente. Si $window_shape = 1$, las funciones de la ventana pueden ser de la siguiente manera:

$$W_0(n) = \begin{cases} W_{LEFT,N_s}(n), & \text{para } 0 \leq n < N_s/2 \\ W_{KBD_RIGHT,N_s}(n) & \text{para } N_s/2 \leq n < N_s \end{cases}$$

$$W_j(n) = \begin{cases} W_{KBD_LEFT,N_s}(n), & \text{para } 0 \leq n < N_s/2 \\ W_{KBD_RIGHT,N_s}(n), & \text{para } N_s/2 \leq n < N_s \end{cases}, 0 < j \leq M-1$$

(N. del T.: KBD _ LEFT = KBD _ izquierdo; KBD _ RIGHT = KBD _ derecho).

- 5 [0232] De lo contrario, si **window_shape** (forma de ventana) == 0, las funciones de la ventana se pueden describir de la siguiente manera:

$$W_0(n) = \begin{cases} W_{LEFT,N_s}(n), & \text{para } 0 \leq n < N_s/2 \\ W_{SIN_RIGHT,N_s}(n), & \text{para } N_s/2 \leq n < N_s \end{cases}$$

$$W_j(n) = \begin{cases} W_{SIN_LEFT,N_s}(n), & \text{para } 0 \leq n < N_s/2 \\ W_{SIN_RIGHT,N_s}(n), & \text{para } N_s/2 \leq n < N_s \end{cases}, 0 < j \leq M-1$$

(N. del T.: SIN _ LEFT = seno _ izquierdo; SIN _ RIGHT = seno _ derecho).

- 10 [0233] La operación de solapar y agregar entre **EIGHT_SHORT_WINDOW_SEQUENCE** (secuencia de ocho ventanas cortas) que da lugar a los valores de dominio de tiempo dividido en ventanas $z_{i,n}$ se describe de la siguiente manera:

$$z_{i,n} = \begin{cases} 0, & \text{para } 0 \leq n < \frac{N_l - N_s}{4} \\ x_{0,n-\frac{N_l-N_s}{4}} \cdot W_0\left(n - \frac{N_l-N_s}{4}\right) & \text{para } \frac{N_l-N_s}{4} \leq n < \frac{N_l+N_s}{4} \\ x_{j-1,n-\frac{N_l+(2j-3)N_s}{4}} \cdot W_{j-1}\left(n - \frac{N_l+(2j-3)N_s}{4}\right) + x_{j,n-\frac{N_l+(2j-1)N_s}{4}} \cdot W_j\left(n - \frac{N_l+(2j-1)N_s}{4}\right), & \text{para } 1 \leq j < M, \frac{N_l+(2j-1)N_s}{4} \leq n < \frac{N_l+(2j+1)N_s}{4} \\ x_{M-1,n-\frac{N_l+(2M-3)N_s}{4}} \cdot W_{M-1}\left(n - \frac{N_l+(2M-3)N_s}{4}\right), & \text{para } \frac{N_l+(2M-1)N_s}{4} \leq n < \frac{N_l+(2M+1)N_s}{4} \\ 0, & \text{para } \frac{N_l+(2M+1)N_s}{4} \leq n < N_l \end{cases}$$

d) **LONG_STOP_SEQUENCE** (secuencia larga de parada)

- 15 [0234] La secuencia **Window_sequence** (secuencia de ventanas) es necesaria para conmutar desde una secuencia de ventanas "EIGHT_SHORT_SEQUENCE" (secuencia de ocho ventanas cortas) o un tipo de ventana "LPD_SEQUENCE" (secuencia LPD) nuevamente a un tipo de ventana "ONLY_LONG_SEQUENCE" (sólo secuencia larga).

- 20 [0235] En el caso de que la secuencia de ventanas anterior no sea una secuencia LPD_SEQUENCE (secuencia LPD):

la longitud de la ventana N_l y N_s se fija en 2048 (1920) y 256 (240), respectivamente.

[0236] En el caso de que la secuencia de ventana anterior sea una secuencia LPD_SEQUENCE (secuencia LPD) : la longitud de la ventana N_l y N_s se fija en 2048 (1920) y 512 (480), respectivamente.

5 [0237] Si **window_shape** (forma de ventana) == 1, la ventana para el tipo de ventana “LONG_STOP_SEQUENCE” (secuencia larga de parada) es la siguiente:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{LEFT, N_s}(n - \frac{N_l - N_s}{4}), & \text{para } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{para } \frac{N_l + N_s}{4} \leq n < N_l / 2 \\ W_{KBD_RIGHT, N_l}(n), & \text{para } N_l / 2 \leq n < N_l \end{cases}$$

(N. del T.: LEFT = izquierdo; KBD _ RIGHT = KBD _ derecho).

10 [0238] Si **window_shape** (forma de ventana) == 0, la ventana para LONG_START_SEQUENCE (secuencia larga de inicio) está determinada por:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l - N_s}{4} \\ W_{LEFT, N_s}(n - \frac{N_l - N_s}{4}), & \text{para } \frac{N_l - N_s}{4} \leq n < \frac{N_l + N_s}{4} \\ 1.0, & \text{para } \frac{N_l + N_s}{4} \leq n < N_l / 2 \\ W_{SIN_RIGHT, N_l}(n), & \text{para } N_l / 2 \leq n < N_l \end{cases}$$

(N. del T.: LEFT = izquierdo; SIN _ RIGHT = seno _ derecho).

15 [0239] Los valores de dominio de tiempo de ventana dividido en ventanas se pueden calcular de acuerdo con la fórmula explicada en a).

e) **STOP_START_SEQUENCE** (secuencia de inicio – de parada):

20 [0240] El tipo de ventana “STOP_START_SEQUENCE” (secuencia de inicio – de parada) se puede utilizar para obtener una operación correcta de solapar y agregar para una transición de bloque desde cualquier bloque con una mitad de ventana de bajo solapamiento (pendiente de ventana corta) de la derecha a cualquier bloque con una mitad de ventana de bajo solapamiento (pendiente de ventana corta) de la izquierda y si es necesaria una única transformación larga para el cuadro actual.

25 [0241] En el caso de que la secuencia de ventana siguiente no sea LPD_SEQUENCE (secuencia LPD):

la longitud de la ventana N_l y N_{sr} se fijan en 2048 (1920) y 256 (240), respectivamente.

30 [0242] En el caso de que la secuencia de ventana siguiente sea LPD_SEQUENCE (secuencia LPD):

la longitud de la ventana N_l y N_{sr} se fijan en 2048 (1920) y 512 (480), respectivamente.

35 [0243] En el caso de que la secuencia de ventana anterior no sea LPD_SEQUENCE (secuencia LPD):

la longitud de la ventana N_l y N_s se fijan en 2048 (1920) y 256 (240), respectivamente.

[0244] En el caso de que la secuencia de ventana anterior sea LPD_SEQUENCE (secuencia LPD):

40 la longitud de la ventana N_l y N_s se fijan en 2048 (1920) y 512 (480), respectivamente.

[0245] Si **window_shape** (forma de ventana) == 1, la ventana para el tipo de ventana “STOP_START_SEQUENCE” (secuencia de inicio – de parada) es la siguiente:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l - N_{sl}}{4} \\ W_{LEFT, N_{sl}}(n - \frac{N_l - N_{sl}}{4}), & \text{para } \frac{N_l - N_{sl}}{4} \leq n < \frac{N_l + N_{sl}}{4} \\ 1.0, & \text{para } \frac{N_l + N_{sl}}{4} \leq n < \frac{3N_l - N_{sr}}{4} \\ W_{KBD_RIGHT, N_{sr}}(n + \frac{N_{sr}}{2} - \frac{3N_l - N_{sr}}{4}), & \text{para } \frac{N_l - N_{sr}}{4} \leq n < \frac{3N_l + N_{sr}}{4} \\ 0.0, & \text{para } \frac{3N_l + N_{sr}}{4} \leq n < N_l \end{cases}$$

(N. del T.: LEFT = izquierdo; KBD _ RIGHT = KBD _ derecho).

5 [0246] Si **window_shape** (forma de ventana) == 0, la ventana para el tipo de ventana "STOP_START_SEQUENCE" (secuencia de inicio – de parada) es la siguiente:

$$W(n) = \begin{cases} 0.0, & \text{para } 0 \leq n < \frac{N_l - N_{sl}}{4} \\ W_{LEFT, N_{sl}}(n - \frac{N_l - N_{sl}}{4}), & \text{para } \frac{N_l - N_{sl}}{4} \leq n < \frac{N_l + N_{sl}}{4} \\ 1.0, & \text{para } \frac{N_l + N_{sl}}{4} \leq n < \frac{3N_l - N_{sr}}{4} \\ W_{SIN_RIGHT, N_{sr}}(n + \frac{N_{sr}}{2} - \frac{3N_l - N_{sr}}{4}), & \text{para } \frac{3N_l - N_{sr}}{4} \leq n < \frac{3N_l + N_{sr}}{4} \\ 0.0, & \text{para } \frac{3N_l + N_{sr}}{4} \leq n < N_l \end{cases}$$

(N. del T.: LEFT = izquierdo; SIN _ RIGHT = seno _ derecho).

10 [0247] Los valores de dominio de tiempo dividido en ventanas se pueden calcular con la fórmula explicada en a).

9.3.3 Banco de Filtros y Conmutación de Bloque – Solapar y Agregar con Secuencia de Ventana Previa

15 [0248] Además de solapar y agregar dentro de la secuencia de ventanas EIGHT_SHORT (secuencia de ocho ventanas cortas), la primera parte (izquierda) de cada **window sequence** (secuencia de ventanas) (o de cada cuadro o subcuadro) se solapa y agrega con la segunda parte (derecha) de **window sequence** previa (o el cuadro anterior o subcuadro) dando lugar a los valores finales de dominio de tiempo . La expresión matemática para esta operación se puede describir de la siguiente manera.

20 [0249] En caso de ONLY_LONG_SEQUENCE (sólo secuencia larga), LONG_START_SEQUENCE (secuencia larga de inicio), EIGHT_SHORT_SEQUENCE (secuencia de ocho ventanas cortas), LONG_STOP_SEQUENCE (secuencia larga de parada), STOP_START_SEQUENCE (secuencia de inicio – de parada):

$$out_{i,n} = z_{i,n} + z_{i-1, n + \frac{N}{2}};$$

$$\text{para } 0 \leq n < \frac{N}{2}, \quad N = 2048 (1920)$$

25 [0250] La ecuación anterior para la operación de solapar y agregar entre los cuadros de audio codificados en el modo de dominio de frecuencia también se puede utilizar para el solapamiento y agregado de representaciones de dominio de tiempo de los cuadros de audio codificado en diferentes modos.

[0251] Alternativamente, la operación de solapar y agregar se puede definir de la siguiente manera:

5 En caso de ONLY_LONG_SEQUENCE (sólo secuencia larga), LONG_START_SEQUENCE (secuencia larga de inicio), EIGHT_SHORT_SEQUENCE (secuencia de ocho ventanas cortas), LONG_STOP_SEQUENCE (secuencia larga de parada), STOP_START_SEQUENCE (secuencia de inicio – de parada):

$$out[i_{out} + n] = Z_{i,n} + Z_{i-1, n + \frac{N-L}{2}}; \forall 0 \leq n < \frac{N-L}{2}$$

[0252] N_L es el tamaño de la secuencia de ventanas. i_{out} indexa el registro intermedio de salida out y aumenta

$$\frac{N-L}{2}$$

10 por el número de muestras escritas.

En caso de LPD_SEQUENCE (secuencia LPD):

[0253] A continuación se describirá un primer procedimiento que se puede utilizar para reducir los artefactos de solapamiento. Cuando proviene de ACELP se puede utilizar una ventana específica para la siguiente TCX reduciendo R a 0, y luego eliminando la región de solapamiento entre los dos cuadros siguientes.

[0254] A continuación se describirá un segundo procedimiento que se puede utilizar para reducir artefactos de solapamiento (como se describe en USAC WD5 y anteriores). Cuando proviene de ACELP, la siguiente ventana TCX se amplía aumentando M (longitud media) en 128 muestras y aumentando también un número de coeficientes MDCT asociados a la ventana TCX. En el decodificador, la parte derecha de la ventana, es decir, las primeras muestras R no cero decodificadas son simplemente eliminadas y reemplazadas por las muestras ACELP decodificadas. En otras palabras, al proporcionar coeficientes MDCT adicionales (por ejemplo, 1152 en vez de 1024) se reducen los artefactos de solapamiento. Es decir, al proporcionar más coeficientes MDCT (de modo tal que el número de coeficientes MDCT sea mayor que la mitad del número de muestras de dominio de tiempo por cuadro de audio) se puede obtener una porción libre de solapamiento de la representación de dominio de tiempo se puede obtener, lo que elimina la necesidad de una cancelación de solapamiento especializada a costa de un muestreo no-crítico del espectro.

[0255] De otro modo, cuando la señal decodificada anterior dividida en ventanas $Z_{i-1,n}$ proviene de TCX basada en MDCT se lleva a cabo una operación de solapar y agregar convencional para obtener la señal de tiempo final out . La operación de solapar y agregar puede estar expresada mediante la siguiente fórmula cuando la secuencia de ventanas del modo FD ventana es LONG_START_SEQUENCE (secuencia larga de inicio) o EIGHT_SHORT_SEQUENCE (secuencia de ocho ventanas cortas):

$$out[i_{out} + n] = \begin{cases} Z_{i, \frac{N-L-N_s}{4} + n} + Z_{i-1, \frac{3N-L-N_s}{4} + n}; \forall 0 \leq n < \frac{N-s}{2} \\ Z_{i, \frac{N-L-N_s}{4} + n}; \forall \frac{N-s}{2} \leq n < \frac{N-L+N_s}{4} \end{cases}$$

[0256] N_{i-1} corresponde al tamaño $2lg$ de la ventana previa aplicada en TCX basada en MDCT. i_{out} indexa el registro intermedio de salida out y aumenta por el número $(N_L + N_s)/4$ de las muestras escritas. $N_s/2$ debería ser igual al valor de L de la TCX basada en MDCT previa definida en la tabla de la Figura 15.

[0257] Para STOP_START_SEQUENCE (secuencia de inicio – de parada), la operación de solapar y agregar entre el modo FD y la TCX basada en MDCT está expresada de la siguiente manera:

$$out[i_{out} + n] = \begin{cases} Z_{i, \frac{N-L-N_{sl}}{4} + n} + Z_{i-1, \frac{3N-L-2N_{sl}}{4} + n}; \forall 0 \leq n < \frac{N-sl}{2} \\ Z_{i, \frac{N-L-N_{sl}}{4} + n}; \forall \frac{N-sl}{2} \leq n < \frac{N-L+N_{sl}}{4} \end{cases}$$

[0258] N_{l-1} corresponde al tamaño $2lg$ de la ventana previa aplicada en TCX basada en MDCT. i_{out} indexa el registro intermedio out y aumenta por el número $(N_{l-1} + N_{sl})/4$ de las muestras escritas. $N_{sl}/2$ debería ser igual al valor de L de la TCX basada en MDCT previa definida en la tabla de la Figura 15.

5 10. Detalles de Información relativos al Cómputo de $\hat{w}[n]$

[0259] A continuación se describirán algunos detalles sobre el cómputo de los valores de ganancia de dominio de predicción lineal $g[k]$ a fin de facilitar la comprensión del presente documento. Por lo general, una corriente de bits que representan el contenido de audio codificado (codificado en el modo de predicción lineal) comprende coeficientes codificados de filtro LPC. Los coeficientes codificados de filtro LPC se pueden describir, por ejemplo, mediante palabras clave correspondientes y pueden describir un filtro de predicción lineal para recuperar el contenido de audio. Cabe señalar que el número de conjuntos de coeficientes de filtro LPC, transmitidos por cuadro de audio codificado por LPC, puede variar. De hecho, el número real de los conjuntos de coeficientes de filtro LPC que están codificados dentro de la corriente de bits para un cuadro de audio codificado en el modo de predicción lineal depende de la combinación del modo ACELP–TCX del cuadro de audio (que a veces también se denomina “supercuadro”). Esta combinación de modo ACELP–TCX puede estar determinada por una variable de corriente de bits. Sin embargo, también hay, naturalmente, casos en los que sólo hay un modo TCX disponible, y también hay casos en los que no hay un modo ACELP disponible.

20 **[0260]** La corriente de bits normalmente se analiza sintácticamente para extraer los índices de cuantificación correspondientes a cada uno de los conjuntos de coeficientes de filtro LPC requeridos por la combinación del modo ACELP TCX.

25 **[0261]** En un primer paso del procesamiento 1810 se lleva a cabo una cuantificación inversa del filtro LPC. Cabe señalar que los filtros LPC (es decir, los conjuntos de coeficientes de filtro LPC, por ejemplo, a_1 a a_{16}) se cuantifican utilizando la representación de frecuencia espectral lineal (LSF) (que es una representación de codificación de los coeficientes de filtro LPC). En el primer paso del procesamiento 1810, las frecuencias espectrales lineales de cuantificación inversa (LSF) provienen de los índices codificados.

30 **[0262]** A tal efecto se puede computar una primera aproximación de etapa y se puede calcular un refinamiento de cuantificación vectorial algebraica opcional (AVQ). Las frecuencias espectrales lineales de cuantificación inversa se pueden reconstruir agregando la primera aproximación de etapa y la contribución AVQ inversa ponderada. La presencia del refinamiento AVQ puede depender del modo de cuantificación real del filtro LPC.

35 **[0263]** El vector de frecuencias espectrales lineales de cuantificación inversa, que puede provenir de la representación codificada de los coeficientes de filtro LPC, posteriormente se convierte en un vector de parámetros de pares de líneas espectrales, luego se interpola y convierte nuevamente en parámetros de LPC. El procedimiento de cuantificación inversa, llevado a cabo en la etapa de procesamiento 1810, da lugar a un conjunto de parámetros LPC en el dominio de frecuencia espectral lineal. Las frecuencias espectrales lineales luego se convierten, en un paso de procesamiento 1820, en el dominio de coseno, que está descrito por pares de líneas espectrales. Por lo tanto se obtienen pares de líneas espectrales q_i . Para cada cuadro o subcuadro, los coeficientes de pares de líneas espectrales q_i (o una versión interpolada de los mismos) se convierten en coeficientes de filtro de predicción lineal a_k , que se utilizan para sintetizar la señal reconstruida en el cuadro o subcuadro. La conversión en el dominio de predicción lineal de dominio se lleva a cabo de la siguiente manera. Los coeficientes $f_{1(i)}$ y $f_{2(i)}$ se pueden obtener, utilizando la siguiente relación recursiva:

$$\begin{aligned}
 & \text{para } i = 1 \text{ a } 8 \\
 & \quad f_1(i) = -2q_{2i-1}f_1(i-1) + 2f_1(i-2) \\
 & \quad \text{para } j = i-1 \text{ hasta } 1 \\
 & \quad \quad f_1(j) = f_1(j) - 2q_{2i-1}f_1(j-1) + f_1(j-2) \\
 & \quad \text{fin} \\
 & \text{fin}
 \end{aligned}$$

con los valores iniciales $f_1(0) = 1$ y $f_1(-1) = 0$. Los coeficientes $f_2(i)$ se computan de manera similar reemplazando q_{2i-1} por q_{2i} .

50 **[0264]** Una vez encontrados los coeficientes de $f_1(i)$ y $f_2(i)$, los coeficientes $f_1(i)$ y $f_2(i)$ se computan de acuerdo con

$$f_1'(i) = f_1(i) + f_1(i-1), \quad i = 1, \dots, 8$$

$$f_2'(i) = f_2(i) - f_2(i-1), \quad i = 1, \dots, 8$$

[0265] Por último, los coeficientes LP a_i se computan a partir de $f_1(i)$ y $f_2(i)$ por

$$a_i = \begin{cases} 0.5f_1'(i) + 0.5f_2'(i), & i = 1, \dots, 8 \\ 0.5f_1'(17-i) - 0.5f_2'(17-i), & i = 9, \dots, 16 \end{cases}$$

5 [0266] En resumen, la derivación de los coeficientes LPC a_i de los coeficientes de pares de líneas espectrales q_i se lleva a cabo utilizando los pasos de procesamiento 1830, 1840, 1850, como se explicó anteriormente.

10 [0267] Los coeficientes $\hat{w}[n]$, $n=0 \dots \text{lpc_order}-1$, que son los coeficientes de un filtro LPC ponderado se obtienen en un paso de procesamiento 1860. Cuando se obtienen los coeficientes $\hat{w}[n]$ a partir de los coeficientes a_i , se considera que los coeficientes a_i son coeficientes de dominio de tiempo de un filtro que tiene características de un filtro $\hat{A}[z]$, y que los coeficientes $\hat{w}[n]$ son coeficientes de dominio de tiempo de un filtro que tiene respuesta de dominio de frecuencia $\hat{W}[z]$. Además, se considera que tiene la siguiente relación:

$$\hat{W}(z) = \hat{A}(z/\gamma_1) \quad \text{con } \gamma_1 = 0.92$$

15 [0268] En vista de lo anterior, se puede observar que los coeficientes $\hat{w}[n]$ pueden extraerse fácilmente de los coeficientes codificados de filtro LPC, que están representados, por ejemplo, por los índices respectivos en la corriente de bits.

20 [0269] Cabe señalar además que se ha analizado anteriormente la derivación de $x_i[n]$, que se lleva a cabo en el paso de procesamiento 1870. Del mismo modo, el cómputo de $X_0[k]$ también se ha analizado anteriormente. En forma similar, también se ha analizado anteriormente el cómputo de valores de ganancia de dominio de predicción lineal de los valores $g[k]$, que se lleva a cabo en el paso 1890.

11. Solución Alternativa para el Modelado Espectral

25 [0270] Cabe señalar que un concepto de modelado espectral se ha descrito anteriormente, el cual se aplica para los cuadros de audio codificados en el dominio de predicción lineal, y que está basado en una transformación de coeficientes de filtro LPC $\hat{w}_n[n]$ en una representación espectral $X_0[k]$ de la que provienen los valores de ganancia del dominio de predicción lineal. Como se mencionó anteriormente, los coeficientes de filtro LPC $\hat{w}[n]$ se transforman en una representación de dominio de frecuencia $X_0[k]$, utilizando una transformada discreta de Fourier impar que tiene 64 intervalos de frecuencia igualmente espaciados. Sin embargo, naturalmente no es necesario obtener los valores de dominio de frecuencia $x_0[k]$, que están igualmente espaciados en la frecuencia. En cambio, en algunas ocasiones puede ser conveniente utilizar los valores de dominio de frecuencia $x_0[k]$, que están espaciados de forma no lineal en la frecuencia. Por ejemplo, los valores de dominio de frecuencia $x_0[k]$ pueden estar espaciados logarítmicamente en la frecuencia o pueden estar espaciados en la frecuencia de acuerdo con una escala de Bark. Dicho espaciado no lineal de los valores de dominio de frecuencia $X_0[k]$ y de los valores de ganancia de dominio de predicción lineal $g[k]$ puede dar lugar a un equilibrio particularmente adecuado entre la impresión auditiva y la complejidad computacional. Sin embargo, no es necesario aplicar dicho concepto de espaciado de frecuencia no uniforme de los valores de ganancia de dominio de predicción lineal.

12. Concepto Mejorado de Transición

40 [0271] A continuación se describirá un concepto mejorado para la transición entre un cuadro de audio codificado en el dominio de frecuencia y un cuadro de audio codificado en el dominio de predicción lineal. Este concepto mejorado utiliza lo que se denomina ventana de inicio del modo de predicción lineal, que se explicará a continuación.

45 [0272] Tomando como referencia primero a las Figuras 17a y 17b, cabe señalar que normalmente las ventanas que tienen una pendiente de transición comparativamente corta en el lado derecho de transición se aplican a las muestras de dominio de tiempo de un cuadro de audio codificado en el modo de dominio de frecuencia cuando se lleva a cabo una transición para un cuadro de audio codificado en el modo de predicción lineal. Como se puede observar en la Figura 17a, una ventana de tipo "LONG_START_SEQUENCE" (secuencia larga de inicio), una

ventana de tipo "EIGHT_SHORT_SEQUENCE" (secuencia de ocho ventanas cortas), una ventana de tipo "STOP_START_SEQUENCE" (secuencia de inicio – de parada) se aplican convencionalmente antes de un cuadro de audio codificado en el dominio de predicción lineal. Por lo tanto, convencionalmente, no hay posibilidad de efectuar directamente una transición desde un cuadro de audio codificado en el dominio de frecuencia, al que se aplica una ventana que tiene una pendiente comparativamente larga en el lado derecho, a un cuadro de audio codificado en el modo de predicción lineal. Esto se debe al hecho de que existen, convencionalmente, graves problemas causados por la porción larga de solapamiento de dominio de tiempo de un cuadro de audio codificado en el dominio de frecuencia al que se aplica una ventana que tiene una pendiente de transición comparativamente larga en el lado derecho. Como se puede observar en la Figura 17a, no es convencionalmente posible efectuar una transición desde un cuadro de audio al que está asociado la ventana de tipo "only_long_sequence" (sólo secuencia larga), o desde un cuadro de audio al que está asociado la ventana de tipo "long_stop_sequence" (secuencia larga de parada), a un cuadro de audio posterior codificado en el modo de predicción lineal.

[0273] Sin embargo, en algunas formas de realización de la invención, se utiliza un nuevo tipo de cuadro de audio, es decir, un cuadro de audio al que está asociado una ventana de inicio del modo de predicción lineal.

[0274] Un nuevo tipo de cuadro de audio (también brevemente designado como cuadro de inicio del modo de predicción lineal) está codificado en el sub-modo TCX del modo de dominio de predicción lineal. El cuadro de inicio del modo de predicción lineal consta de un solo cuadro TCX (es decir, no está sub-dividido en subcuadros TCX). En consecuencia, tanto como 1024 coeficientes MDCT están incluidos en la corriente de bits, en una forma codificada, para el cuadro de inicio del modo de predicción lineal. En otras palabras, el número de coeficientes MDCT relacionado con un cuadro de inicio de predicción lineal es idéntico al número de coeficientes MDCT asociados al cuadro de audio codificado en el dominio de frecuencia al que está asociado una ventana de tipo "only_long_sequence" (sólo secuencia larga). Asimismo, la ventana asociada al cuadro de inicio de modo de predicción lineal puede ser del tipo de ventana "LONG_START_SEQUENCE" (secuencia larga de inicio). Por lo tanto, el cuadro de inicio del modo de predicción lineal puede ser muy similar al cuadro codificado en el dominio de frecuencia al que está asociado una ventana de tipo "long_start_sequence" (secuencia larga de inicio). Sin embargo, el cuadro de inicio del modo de predicción lineal difiere de dicho cuadro de audio codificado en el dominio de frecuencia debido a que el modelado espectral se lleva a cabo dependiendo de los valores de ganancia del dominio de predicción lineal, y no depende de los valores de factores de escala. Por lo tanto, los coeficientes de filtro de codificación de predicción lineal están incluidos en la corriente de bits para el cuadro de inicio del modo de predicción lineal.

[0275] Como la MDCT inversa 1354, 1382 se aplica en el mismo dominio (como se explicó anteriormente), tanto para un cuadro de audio codificado en el modo de dominio de frecuencia como para un cuadro de audio codificado en el modo de predicción lineal, una operación de solapar y agregar de cancelación de solapamiento del dominio de tiempo que tiene características adecuadas de cancelación de solapamiento de tiempo se puede llevar a cabo entre un cuadro de audio previo codificado en el modo de dominio de frecuencia y que tiene una pendiente de transición comparativamente larga en el lado derecho (por ejemplo, de 1024 muestras) y el cuadro de inicio del modo de predicción lineal que tiene una pendiente de transición comparativamente larga en el lado izquierdo (por ejemplo, de 1024 muestras), en donde las pendientes de transición coinciden para la cancelación de solapamiento de tiempo. Por lo tanto, el cuadro de inicio del modo de predicción lineal está codificado en el modo de predicción lineal (es decir, utilizando los coeficientes del filtro de codificación de predicción lineal) y comprende una pendiente de transición en el lado izquierdo significativamente más larga (por ejemplo, por lo menos por el factor de 2, o por lo menos por el factor de 4, o por lo menos por el factor de 8) que otros cuadros de audio codificados en el modo de predicción lineal para crear posibilidades adicionales de transición.

[0276] Como consecuencia, un cuadro de inicio del modo de predicción lineal puede reemplazar al cuadro de audio codificado en el dominio de frecuencia que tiene el tipo de ventana "long_sequence" (secuencia larga). El cuadro de inicio del modo de predicción lineal comprende la ventaja de que los coeficientes de filtro MDCT son transmitidos para el cuadro de inicio del modo de predicción lineal, que están disponibles para un cuadro de audio posterior codificado en el modo de predicción lineal. En consecuencia, no es necesario incluir información extra sobre coeficientes de filtro LPC en la corriente de bits a fin de obtener información de inicialización para una decodificación del cuadro de audio posterior codificado en el modo de predicción lineal.

[0277] La Figura 14 ilustra este concepto. La Figura 14 muestra una representación gráfica de una secuencia de cuatro cuadros de audio 1410, 1412, 1414, 1416, comprendiendo todas ellas una longitud de 2048 muestras de audio, y que se solapan en aproximadamente un 50%. El primer cuadro de audio 1410 está codificado en el modo de dominio de frecuencia utilizando una "only_long_sequence" (sólo secuencia larga) 1420, el segundo cuadro de audio 1412 está codificado en el modo de predicción lineal utilizando una ventana de inicio del modo de predicción lineal, que es igual a la ventana "long_start_sequence" (secuencia larga de inicio), el tercer cuadro de audio 1414 está codificado en el modo de predicción lineal utilizando, por ejemplo, una ventana $\hat{W}[n]$ como se definió anteriormente para un valor de $\text{mod}[x] = 3$, que está designado como 1424. Cabe señalar que la ventana de inicio del modo de predicción lineal 1422 comprende una pendiente de transición en el lado izquierdo cuya longitud es de 1024 muestras de audio y una pendiente de transición en el lado derecho cuya longitud es de 256 muestras. La ventana 1424 comprende una pendiente de la transición en el lado izquierdo cuya longitud es de 256 muestras y una

pendiente de transición en el lado derecho cuya longitud es de 256 muestras. El cuarto cuadro de audio 1416 está codificado en el modo de dominio de frecuencia utilizando una ventana “long_stop_sequence” (secuencia larga de parada) 1426, que comprende una pendiente de transición en el lado izquierdo cuya longitud es de 256 muestras y una pendiente de transición en el lado derecho cuya longitud es de 1024 muestras.

5 **[0278]** Como se puede observar en la Figura 14 muestras de dominio de tiempo para los cuadros de audio son proporcionadas por las transformadas inversas coseno discretas modificadas 1460, 1462, 1464, 1466. Para los cuadros de audio 1410, 1416 codificados en el modo de dominio de frecuencia, el modelado espectral se lleva a cabo dependiendo de los factores de escala y de los valores de factores de escala. Para los cuadros de audio 1412, 1414, que están codificados en el modo de predicción lineal, el modelado espectral se lleva a cabo dependiendo de los valores de ganancia del dominio de predicción lineal que provienen de los coeficientes codificados de filtro de codificación de predicción lineal. En cualquiera de los casos, los valores espectrales son proporcionados por una decodificación (y, opcionalmente, una cuantificación inversa).

15 13. Conclusión

[0279] En resumen, las formas de realización de acuerdo con la invención utilizan un modelado de ruido basado en LPC aplicado en el dominio de frecuencia para un codificador de audio conmutado.

20 **[0280]** Las formas de realización de acuerdo con la invención aplican un filtro basado en LPC en el dominio de frecuencia para facilitar la transición entre los diferentes codificadores en el contexto de un codec de audio conmutado.

25 **[0281]** Por consiguiente, algunas formas de realización resuelven los problemas para diseñar transiciones eficaces entre los tres modos de codificación, codificación de dominio de frecuencia, TCX (dominio de predicción lineal transformación con código de excitación) y ACELP (modo de predicción lineal excitada por código algebraico). Sin embargo, en algunas otras formas de realización, es suficiente tener sólo dos de dichos modos, por ejemplo, la codificación del dominio de frecuencia y el modo TCX.

30 **[0282]** Las formas de realización de acuerdo con la invención tienen un desempeño superior con respecto a las soluciones alternativas siguientes:

- Transiciones muestreadas en forma no-crítica entre el codificador de dominio de frecuencia y el codificador de dominio de predicción lineal (véase, por ejemplo, la referencia [4]):

- generan un muestreo no crítico, equilibrio entre el tamaño de solapamiento y la información de cabecera, no utilizan plenamente la capacidad (cancelación de solapamiento del dominio de tiempo TDAC) de las MDCT.

- necesitan enviar un conjunto extra de LPC de coeficientes cuando pasa desde el codificador de dominio de frecuencia al codificador LPD.

- Aplican una cancelación de solapamiento del dominio de tiempo (TDAC) en diferentes dominios (véase, por ejemplo, la referencia [5]). El filtrado LPC se lleva a cabo dentro de la MDCT entre el plegamiento y la DCT:

- la señal solapada de dominio de tiempo puede no ser adecuada para el filtrado; y

- es necesario enviar un conjunto extra de LPC de coeficientes cuando pasa desde el codificador de dominio de frecuencia al codificador LPD.

- Computa coeficientes LPC en el dominio MDCT para un codificador no conmutado (T_{winVQ}) (véase, por ejemplo, la referencia [6]);

- utiliza la LPC sólo como una presentación de envolvente espectral para aplanar el espectro. No aprovecha la LPC ni para modelar el ruido de cuantificación, ni para facilitar las transiciones cuando conmuta hacia otro codificador de audio.

60 **[0283]** Las formas de realización de acuerdo con la presente invención realizan la MDCT del codificador de dominio de frecuencia y del codificador LPC en el mismo dominio utilizando al mismo tiempo la LPC para modelar el error de cuantificación en el dominio MDCT. Esto trae aparejado un número de ventajas:

- La LPC todavía se puede utilizar para conmutar a un codificador de voz tipo ACELP.

- La cancelación de solapamiento del dominio de tiempo (TDAC) es posible durante la transición desde/a TCX a/desde el codificador de dominio de frecuencia, entonces el muestreo crítico se mantiene.

• La LPC se sigue utilizando como un modelador de ruido alrededor de ACELP, lo que hace posible utilizar la misma función del objetivo para maximizar tanto para TCX y ACELP, (por ejemplo, la SNR segmental ponderada basada en LPC en un proceso de decisión de circuito cerrado).

[0284] Resumiendo nuevamente, es un aspecto importante que

1. la transición entre la transformación con código de excitación (TCX) y el dominio de frecuencia (FD) se simplifica/unifica significativamente aplicando la codificación de predicción lineal en el dominio de frecuencia, y que

2. manteniendo la transmisión de los coeficientes LPC en el caso TCX, las transiciones entre TCX y ACELP se pueden llevar a cabo ventajosamente como en otras implementaciones (cuando se aplica el filtro LPC en el dominio de tiempo).

Alternativas de Aplicación

[0285] Si bien se han descrito algunos aspectos en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del procedimiento correspondiente, en donde un bloque o dispositivo corresponde a un paso del procedimiento o a una característica de un paso del procedimiento. En forma análoga, los aspectos descritos en el contexto de un paso del procedimiento también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente. Algunos o todos los pasos del procedimiento pueden ser ejecutados (o utilizados) por un aparato de hardware, como por ejemplo, un microprocesador, una computadora programable o un circuito electrónico. En algunas formas de realización, algunos pasos o la mayoría de los pasos más importantes del procedimiento pueden ser ejecutados por un aparato de esas características.

[0286] Dependiendo de determinados requisitos de aplicación, las formas de realización de la invención se pueden implementar en hardware o en software. La implementación se puede llevar a cabo utilizando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blue-Ray, un CD, una memoria ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tienen señales de control electrónicamente legibles que están almacenadas en ellos, que cooperan (o son capaces de cooperar) con un sistema de computadora programable de forma tal que el procedimiento respectivo se lleve a cabo. Por lo tanto, el medio de almacenamiento digital puede ser legible por computadora.

[0287] Las formas de realización descritas anteriormente 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 este documento serán evidentes para otros expertos en la materia. Por lo tanto, la presente invención tiene el propósito de limitarse sólo al alcance de las reivindicaciones de patente inminentes y no a los detalles específicos presentados a modo de descripción y explicación de las formas de realización del presente documento.

Referencias:

[0288]

[1] "Unified speech and audio coding scheme for high quality at low bitrates", Max Neuendorf et al., in IEEE Int. Conf. Acoustics, Speech and Signal Processing, ICASSP, 2009
 [2] Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding. International Standard 13818-7, ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group, 1997
 [3] "Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec", 3GPP TS 26.290 V6.3.0, 2005-06, Technical Specification
 [4] "Audio Encoder and Decoder for Encoding and Decoding Audio Samples", FH080703PUS, F49510, incorporado a modo de referencia.
 [5] "Apparatus and Method for Encoding/Decoding an Audio Signal Using an Aliasing Switch Scheme", FH080715PUS, F49522, incorporado a modo de referencia.
 [6] "High-quality audio-coding at less than 64 kbits/s "by using transform-domain weighted interleaved vector quantization (Twin VQ)", N. Iwakami y T. Moriya y S. Miki, IEEE ICASSP, 1995.

REIVINDICACIONES

- 5 1. Un decodificador multimodo para señal de audio (1100; 1200) para proporcionar una representación decodificada (1112; 1212) de un contenido de audio sobre la base de una representación codificada (1110; 1208) del contenido de audio, en donde el decodificador para señal de audio comprende:
- 10 un determinador de valor espectral (1130; 1230a; 1230c) configurado para obtener conjuntos (1132; 1230d) de coeficientes espectrales decodificados (1132; 1230d; $r[i]$) para una pluralidad de porciones (1410, 1412, 1414, 1416) del contenido de audio;
- 15 un procesador espectral (1230e; 1378) configurado para aplicar un modelado espectral de un conjunto (1132; 1230d; $r[i]$) de coeficientes espectrales decodificados, o a una versión pre-procesada (1132') del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal para una porción del contenido de audio codificado en el modo de predicción lineal, y para aplicar un modelado espectral a un conjunto (1132; 1230d; $r[i]$) de coeficientes espectrales decodificados, o una versión pre-procesada (1232') del mismo, dependiendo de un conjunto de parámetros de factores de escala (1152; 1260b) para una porción (1410; 1416) del contenido de audio codificado en el modo de dominio de frecuencia, y
- 20 un convertidor de dominio de frecuencia a dominio de tiempo (1160; 1230g) configurado para obtener una representación de dominio de tiempo (1162; 1232; x_i, n) del contenido de audio sobre la base de un conjunto modelado espectralmente (1158; 1230f) de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de predicción lineal, y para obtener una representación de dominio de tiempo (1162; 1232) del contenido de audio sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de dominio de frecuencia.
- 25 2. El decodificador multimodo para señal de audio de acuerdo con la reivindicación 1, en donde el decodificador multimodo para señal de audio comprende además un solapador (1233) configurado para solapar y agregar una representación de dominio de tiempo de una porción del contenido de audio codificado en el modo de predicción lineal con una porción del contenido de audio codificado en el modo de dominio de frecuencia.
- 30 3. El decodificador multimodo para señal de audio de acuerdo con la reivindicación 2, en donde el convertidor de dominio de frecuencia a dominio de tiempo (1160; 1230g) está configurado para obtener una representación de dominio de tiempo del contenido de audio para una porción (1412; 1414) del contenido de audio codificado en el modo de predicción lineal utilizando una transformación solapada, y para obtener una representación de dominio de tiempo del contenido de audio para una porción (1410; 1416) del contenido de audio codificado en el modo de dominio de frecuencia utilizando una transformación solapada, y
- 35 en donde el solapador está configurado para solapar las representaciones de dominio de tiempo de las porciones posteriores del contenido de audio codificado en diferentes modos.
- 40 4. El decodificador multimodo para señal de audio de acuerdo con la reivindicación 3, en donde el convertidor de dominio de frecuencia a dominio de tiempo (1160; 1230g) está configurado para aplicar transformaciones solapadas del mismo tipo de transformación para obtener representaciones de dominio de tiempo del contenido de audio para porciones del contenido de audio codificado en diferentes modos y
- 45 en donde el solapador está configurado para solapar y agregar las representaciones de dominio de tiempo de las porciones posteriores del contenido de audio codificado en diferentes modos de manera tal que se reduce o elimina un solapamiento de dominio de tiempo causado por la transformación solapada.
- 50 5. El decodificador multimodo para señal de audio de acuerdo con la reivindicación 4, en donde el solapador está configurado para solapar y agregar una representación de dominio de tiempo dividida en ventanas de una primera porción (1414) del contenido de audio codificado en un primero de los modos proporcionados por una transformación solapada asociada, o una versión ajustada a escala de amplitud pero espectralmente no distorsionada del mismo, y una representación de dominio de tiempo dividida en ventanas de una segunda porción posterior (1416) del contenido de audio codificado en un segundo de los modos, proporcionados por una transformación solapada asociada, o una versión ajustada a escala de amplitud pero espectralmente no distorsionada del mismo.
- 55 6. El decodificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 1 a 5, en donde el convertidor de dominio de frecuencia a dominio de tiempo (1160; 1230g) está configurado para proporcionar representaciones de dominio de tiempo de las porciones (1410, 1412, 1414, 1416) del contenido de audio codificado en diferentes modos de manera tal que las representaciones de dominio de tiempo proporcionadas se encuentran en un mismo dominio, ya que son linealmente combinables sin aplicar una operación de filtrado de modelado de señales, a excepción de una operación de transición dividida en ventanas, a una o ambas representaciones de dominio de tiempo proporcionadas.
- 60 7. El decodificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 1 a 6, en donde el convertidor de dominio de frecuencia a dominio de tiempo (1160; 1230g) está configurado para llevar a cabo una transformada inversa coseno discreta modificada, para obtener, como resultado de la transformada inversa coseno
- 65

discreta modificada, una representación de dominio de tiempo del contenido de audio en un dominio de señal de audio tanto para una porción del contenido de audio codificado en el modo de predicción lineal como para una porción del contenido de audio codificado en el modo de dominio de frecuencia.

5 **8.** El decodificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 1 a 7, que comprende:

un determinador de coeficientes de filtro de codificación de predicción lineal configurado para obtener coeficientes de filtro decodificados de codificación de predicción lineal (α_1 a α_{16}) sobre la base de una representación codificada de los coeficientes de filtro de codificación de predicción lineal para una porción del contenido de audio codificado en el modo de predicción lineal;

10 un transformador de coeficientes del filtro (1260e) configurado para transformar los coeficientes decodificados de codificación de predicción lineal (1260d; α_1 a α_{16}) en una representación espectral (1260f; $X_0[k]$), con el fin de obtener los valores de ganancia del modo de predicción lineal ($g[k]$) asociados con diferentes frecuencias;

15 un determinador de factores de escala (1260a) configurado para obtener valores decodificados de factores de escala (1260f) sobre la base de una representación codificada (1254) de los valores de factores de escala para una porción del contenido de audio codificado en un modo de dominio de frecuencia;

20 en donde el procesador espectral (1150; 1230e) comprende un modificador espectral configurado para combinar un conjunto (1132; 1230d; $r[i]$) de coeficientes espectrales decodificados asociados a una porción del contenido de audio codificado en el modo de predicción lineal, o una versión pre-procesada del mismo, con los valores de ganancia del modo de predicción lineal ($g[k]$), con el fin de obtener una versión procesada de ganancias (1158; 1230f; $rr[i]$) de los coeficientes espectrales decodificados, en los que las contribuciones de los coeficientes espectrales decodificados (1130; 1230d; $r[i]$), o de la versión pre-procesada del mismo, se ponderan dependiendo de los valores de ganancia del modo de predicción lineal ($g[k]$), y que también está configurado para combinar un conjunto (1132; 1230d; $x_{ac_invquant}$) de coeficientes espectrales decodificados asociados a una porción del contenido de audio codificado en el modo de dominio de frecuencia, o una versión pre-procesada del mismo, con los valores de factores de escala (1260b), con el fin de obtener una versión procesada a factores de escala (x_{rescal}) de los coeficientes espectrales decodificados ($x_{ac_invquant}$) en los que las contribuciones de los coeficientes espectrales decodificados, o de la versión pre-procesada del mismo, se ponderan dependiendo de los valores de factores de escala.

9. El decodificador multimodo para señal de audio de acuerdo con la reivindicación 8, en donde el transformador de coeficientes de filtro (1260e) está configurado para transformar los coeficientes de filtro decodificados de codificación de predicción lineal (1260d), que representan una respuesta al impulso de dominio de tiempo ($W[n]$) de un filtro de codificación de predicción lineal, en una representación espectral ($X_0[k]$) utilizando una transformada discreta impar de Fourier; y

40 en donde el transformador de coeficientes de filtro (1260e) está configurado para obtener los valores de ganancia del modo de predicción lineal ($g[k]$) a partir de la representación espectral ($X_0[k]$) de los coeficientes de filtro decodificados de codificación de predicción lineal (1260d; a_1 a a_{16}), de manera tal que los valores de ganancia dependen de las magnitudes de los coeficientes ($X_0[k]$) de la representación espectral ($X_0[k]$).

10. El decodificador multimodo para señal de audio de acuerdo con la reivindicación 8 o con la reivindicación 9, en donde el transformador de coeficientes de filtro (1260e) y el combinador (1230e) están configurados de manera tal que una contribución de un determinado coeficiente espectral decodificado ($r[i]$), o de una versión pre-procesada del mismo, a una versión de ganancia pre-procesada ($rr[i]$) del coeficiente espectral determinado está determinado por una magnitud de un valor de ganancia del modo de predicción lineal ($g[k]$) asociados con el coeficiente espectral decodificado determinado ($r[i]$).

11. El decodificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 1 a 9, en donde el procesador espectral (1230e) está configurado de manera tal que una ponderación de una contribución de un determinado coeficiente espectral decodificado ($r[i]$), o de una versión pre-procesada del mismo, a una versión procesada de ganancia ($rr[i]$) del coeficiente espectral determinado aumenta al aumentar la magnitud de un valor del ganancia del modo de predicción lineal ($g[k]$) asociado con el coeficiente espectral decodificado determinado ($r[i]$), o de manera tal que una ponderación de una contribución de un determinado coeficiente decodificado espectral ($r[i]$), o de una versión pre-procesada del mismo, a una versión procesada de ganancia ($rr[i]$) del coeficiente espectral asociado disminuye al aumentar la magnitud de un coeficiente espectral asociado ($X_0[k]$) de una representación espectral de los coeficientes de filtro decodificados de codificación de predicción lineal.

12. El decodificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 1 a 11, en donde el determinador de valor espectral (1130; 1230^a; 1230c) está configurado para aplicar una cuantificación inversa a coeficientes espectrales cuantificados decodificados, con el fin de obtener coeficientes espectrales decodificados inversamente cuantificados (1132; 1230d); y

65 en donde el procesador espectral (1230e) está configurado para llevar a cabo un modelado de ruido de cuantificación ajustando un paso de cuantificación eficaz para un coeficiente espectral decodificado determinado

(r[i]) dependiendo de una magnitud de un valor de ganancia de modo de predicción lineal (g[k]) asociado con el coeficiente espectral decodificado determinado (r[i]).

5 **13.** El decodificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 1 a 12, en donde el decodificador para señal de audio está configurado para utilizar un cuadro de inicio de modo de predicción lineal intermedio (1212) con el fin de realizar una transición desde un cuadro de modo de dominio de frecuencia (1410) a un cuadro combinado de modo de predicción lineal/modo de predicción lineal excitada por código algebraico, en donde el decodificador para señal de audio está configurado para obtener un conjunto de coeficientes espectrales decodificados para el cuadro de inicio del modo de predicción lineal,
10 para aplicar un modelado espectral al conjunto de coeficientes espectrales decodificados para el cuadro de inicio del modo de predicción lineal, o a una versión pre–procesada del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal asociados con los mismos,
para obtener una representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados, y
15 para aplicar una ventana de inicio que tiene una pendiente de transición comparativamente larga en el lado izquierdo y una pendiente de transición comparativamente corta en el lado derecho a la representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal.

20 **14.** El decodificador multimodo para señal de audio de acuerdo con la reivindicación 13, en donde el decodificador para señal de audio está configurado para solapar una porción del lado derecho de una representación de dominio de tiempo de un cuadro de modo de dominio de frecuencia (1410) que precede al cuadro de inicio del modo de predicción lineal (1412) con una porción del lado izquierdo de una representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal, para obtener una reducción o cancelación de solapamiento del dominio de tiempo.

25 **15.** El decodificador multimodo para señal de audio de acuerdo con la reivindicación 13 o con la reivindicación 14, en donde el decodificador para señal de audio está configurado para utilizar los parámetros de dominio de predicción lineal asociados al cuadro de inicio del modo de predicción lineal (1412) con el fin de inicializar un decodificador de modo de predicción lineal excitada por código algebraico para decodificar por lo menos una porción del cuadro combinado de modo de predicción lineal/modo de predicción lineal excitada por código algebraico a continuación del cuadro de inicio del modo de predicción lineal.

30 **16.** Un codificador multimodo para señal de audio (100; 300; 900; 1000) para proporcionar una representación codificada (112; 312; 1012) de un contenido de audio sobre la base de una representación de entrada (110; 310; 1010) del contenido de audio, en donde el codificador para señal de audio comprende:
un convertidor de dominio de tiempo a dominio de frecuencia (120; 330a; 350a; 1030a) configurado para procesar la representación de entrada (110; 310; 1010) del contenido de audio, para obtener una representación de dominio de frecuencia (122; 330b; 1030b) del contenido de audio;
un procesador espectral (130; 330e; 350d; 1030e) configurado para aplicar un modelado espectral a un conjunto de coeficientes espectrales, o una versión pre–procesada del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal (134; 340b) para una porción del contenido de audio a codificar en el modo de predicción lineal, y para aplicar un modelado espectral a un conjunto de coeficientes espectrales, o una versión pre–procesada del mismo, dependiendo de un conjunto de parámetros de factores de escala (136) para una porción del contenido de audio a codificar en el modo de dominio de frecuencia; y
40 un codificador de cuantificación (140; 330g; 330i; 350f; 350h; 1030g; 1030i) configurado para proporcionar una versión codificada (142; 322; 342; 1032) de un conjunto modelado espectralmente (132; 350e; 1030f) de coeficientes espectrales para la porción del contenido de audio a codificar en el modo de predicción lineal, y para proporcionar una versión codificada (142; 322; 342; 1032) de un conjunto modelado espectralmente (132; 330f; 1030f) de coeficientes espectrales para la porción del contenido de audio a codificar en el modo de dominio de frecuencia.

50 **17.** El codificador multimodo para señal de audio de acuerdo con la reivindicación 16, en donde el convertidor de dominio de tiempo a dominio de frecuencia (120; 330a; 350a; 1030a) está configurado para convertir una representación de dominio de tiempo (110; 310; 1010) de un contenido de audio en un dominio de señal de audio en una representación de dominio de frecuencia (122; 330b; 1030b) del contenido de audio tanto para una porción del contenido de audio a codificar en el modo de predicción lineal como para una porción del contenido de audio a codificar en el modo de dominio de frecuencia.

55 **18.** El codificador multimodo para señal de audio de acuerdo con la reivindicación 16 o con la reivindicación 17, en donde el convertidor de dominio de tiempo a dominio de frecuencia (120; 330a; 350a; 1030a) está configurado para aplicar transformaciones solapadas del mismo tipo de transformación para obtener representaciones de dominio de frecuencia para porciones del contenido de audio a codificar en diferentes modos.

60 **19.** El codificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 16 a 18, en donde el procesador espectral (130; 330e; 350b; 1030e) está configurado para aplicar selectivamente el modelado espectral al conjunto (122; 330b; 1030b) de coeficientes espectrales, o una versión pre–procesada del mismo, dependiendo de un conjunto (134; 340b) de parámetros de dominio de predicción lineal obtenidos utilizando un análisis basado en

65

correlación de una porción del contenido de audio a codificar en el modo de predicción lineal, o dependiendo de un conjunto (136; 330d; 1070b) de parámetros de factores de escala obtenidos utilizando un análisis de modelo psicoacústico (330c; 1070a) de una porción del contenido de audio a codificar en el modo de dominio de frecuencia.

- 5 **20.** El codificador multimodo para señal de audio de acuerdo con la reivindicación 19, en donde el codificador para señal de audio comprende un selector de modo configurado para analizar el contenido de audio con el fin de decidir si se codificará una porción del contenido de audio en el modo de predicción lineal o en el modo de dominio de frecuencia.
- 10 **21.** El codificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 16 a 20, en donde el codificador multicanal para señal de audio está configurado para codificar un cuadro de audio, que está entre un cuadro de modo de dominio de frecuencia y un cuadro combinado de modo de predicción lineal de transformación con código de excitación/modo de predicción lineal excitada por código algebraico como un cuadro de inicio de modo de predicción lineal, en donde el codificador multimodo para señal de audio está configurado para
- 15 aplicar una ventana de inicio que tiene una pendiente de transición comparativamente larga en el lado izquierdo y una pendiente de transición comparativamente corta en el lado derecho a la representación de dominio de tiempo del cuadro de inicio del modo de predicción lineal, para obtener una representación de dominio de tiempo dividida en ventanas,
- 20 obtener una representación de dominio de frecuencia de la representación de dominio de tiempo dividida en ventanas del cuadro de inicio del modo de predicción lineal,
- obtener un conjunto de parámetros de dominio de predicción lineal para el cuadro de inicio del modo de predicción lineal,
- aplicar un modelado espectral a la representación de dominio de frecuencia de la representación de dominio de tiempo dividida en ventanas del cuadro de inicio del modo de predicción lineal, o una versión pre-procesada del mismo, dependiendo del conjunto de parámetros de dominio de predicción lineal, y
- 25 codificar el conjunto de parámetros de dominio de predicción lineal y la representación de dominio de frecuencia modelada espectralmente de la representación de dominio de tiempo dividida en ventanas del cuadro de inicio del modo de predicción lineal.
- 30 **22.** El codificador multimodo para señal de audio de acuerdo con la reivindicación 21, en donde el codificador multimodo para señal de audio está configurado para utilizar los parámetros de dominio de predicción lineal asociados al cuadro de inicio del modo de predicción lineal para inicializar un codificador del modo de predicción lineal excitada por código algebraico para codificar por lo menos una porción del cuadro combinado del modo de predicción lineal de transformación excitada por código/modo de predicción lineal excitada por código algebraico a
- 35 continuación del cuadro de inicio del modo de predicción lineal.
- 23.** El codificador multimodo para señal de audio de acuerdo con una de las reivindicaciones 16 a 22, en donde el codificador para señal de audio comprende:
- 40 un determinador de coeficientes del filtro de codificación de predicción lineal (340a; 1070c) configurado para analizar una porción del contenido de audio a codificar en un modo de predicción lineal, o una versión pre-procesada del mismo, para determinar los coeficientes de filtro de codificación de predicción lineal asociados a la porción del contenido de audio a codificar en el modo de predicción lineal;
- 45 un transformador de coeficientes de filtro (350b; 1070d) configurado para transformar los coeficientes de filtro de codificación de predicción lineal en una representación espectral ($X_0[k]$), con el fin de obtener los valores de ganancia del modo de predicción lineal ($g[k]$, 350c) asociados a diferentes frecuencias;
- un determinador de factores de escala (330c; 1070a) configurado para analizar una porción del contenido de audio a codificar en el modo de dominio de frecuencia, o una versión pre-procesada del mismo, para determinar los factores de escala asociados a la porción del contenido de audio a codificar en el modo de dominio de frecuencia;
- 50 un dispositivo combinador (330e, 350d; 1030e) configurado para combinar una representación del dominio de frecuencia de una porción del contenido de audio a codificar en el modo de predicción lineal, o una versión pre-procesada del mismo, con los valores de ganancia del modo de predicción lineal ($g[k]$), para obtener componentes espectrales de ganancia procesados, en donde las contribuciones de los
- 55 componentes espectrales de la representación de dominio de frecuencia del contenido de audio se ponderan dependiendo de los valores de ganancia del modo de predicción lineal, y
- para combinar una representación de dominio de frecuencia de una porción del contenido de audio a codificar en el modo de dominio de frecuencia, o una versión pre-procesada del mismo, con los factores de escala, para obtener componentes espectrales de ganancia procesados, en donde las contribuciones de los
- 60 componentes espectrales de la representación de dominio de frecuencia del contenido de audio se ponderan dependiendo de los factores de escala,
- en donde los componentes espectrales procesados forman conjuntos modelados espectralmente de coeficientes espectrales.
- 65 **24.** Un procedimiento para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada del contenido de audio, cuyo procedimiento comprende:

obtener conjuntos de coeficientes espectrales decodificados para una pluralidad de porciones del contenido de audio;

5 aplicar un modelado espectral a un conjunto de coeficientes espectrales decodificados, o una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal para una porción del contenido de audio codificado en un modo de predicción lineal, y aplicar un modelado espectral a un conjunto de coeficientes espectrales decodificados, o una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de factores de escala para una porción del contenido de audio codificado en un modo de dominio de frecuencia, y

10 obtener una representación de dominio de tiempo del contenido de audio sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de predicción lineal, y obtener una representación de dominio de tiempo del contenido de audio sobre la base de un conjunto modelado espectralmente de coeficientes espectrales decodificados para una porción del contenido de audio codificado en el modo de dominio de frecuencia.

15 **25.** Un procedimiento para proporcionar una representación codificada de un contenido de audio sobre la base de una representación de entrada del contenido de audio, cuyo procedimiento comprende:

20 procesar la representación de entrada del contenido de audio, para obtener una representación de dominio de frecuencia del contenido de audio;

aplicar un modelado espectral a un conjunto de coeficientes espectrales, o una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de dominio de predicción lineal para una porción del contenido de audio a codificar en el modo de predicción lineal;

25 aplicar un modelado espectral a un conjunto de coeficientes espectrales, o una versión pre-procesada del mismo, dependiendo de un conjunto de parámetros de factores de escala para una porción del contenido de audio a codificar en el modo de dominio de frecuencia;

proporcionar una representación codificada de un conjunto modelado espectralmente de coeficientes espectrales para la porción del contenido de audio a codificar en el modo de predicción lineal utilizando una codificación de cuantificación, y

30 proporcionar una versión codificada de un conjunto modelado espectralmente de coeficientes espectrales para la porción del contenido de audio a codificar en el modo de dominio de frecuencia utilizando una codificación de cuantificación.

35 **26.** Un programa de computadora para llevar a cabo el procedimiento de acuerdo con la reivindicación 24 o 25 cuando el programa de computadora se ejecuta en una computadora.

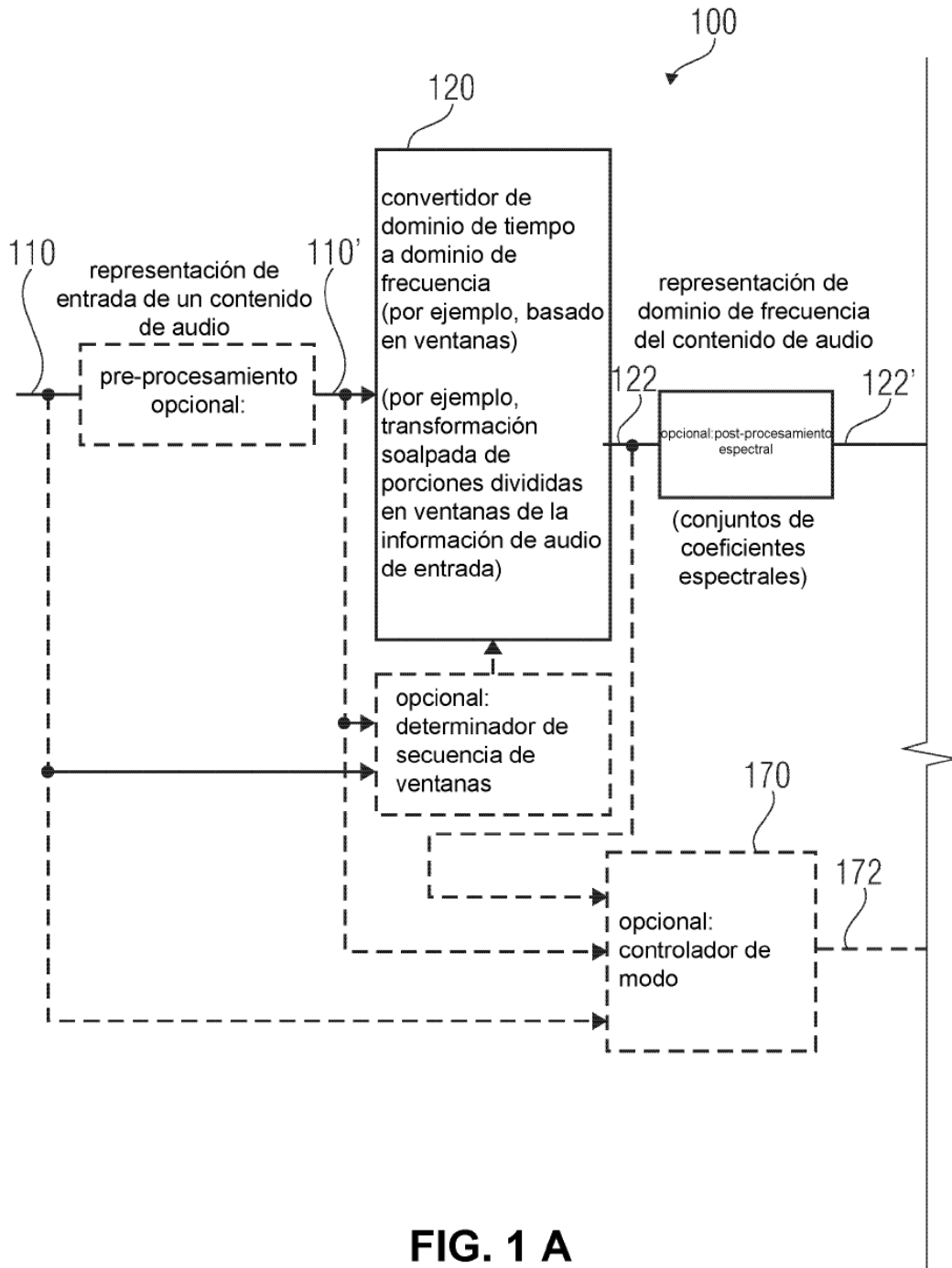


FIG. 1 A

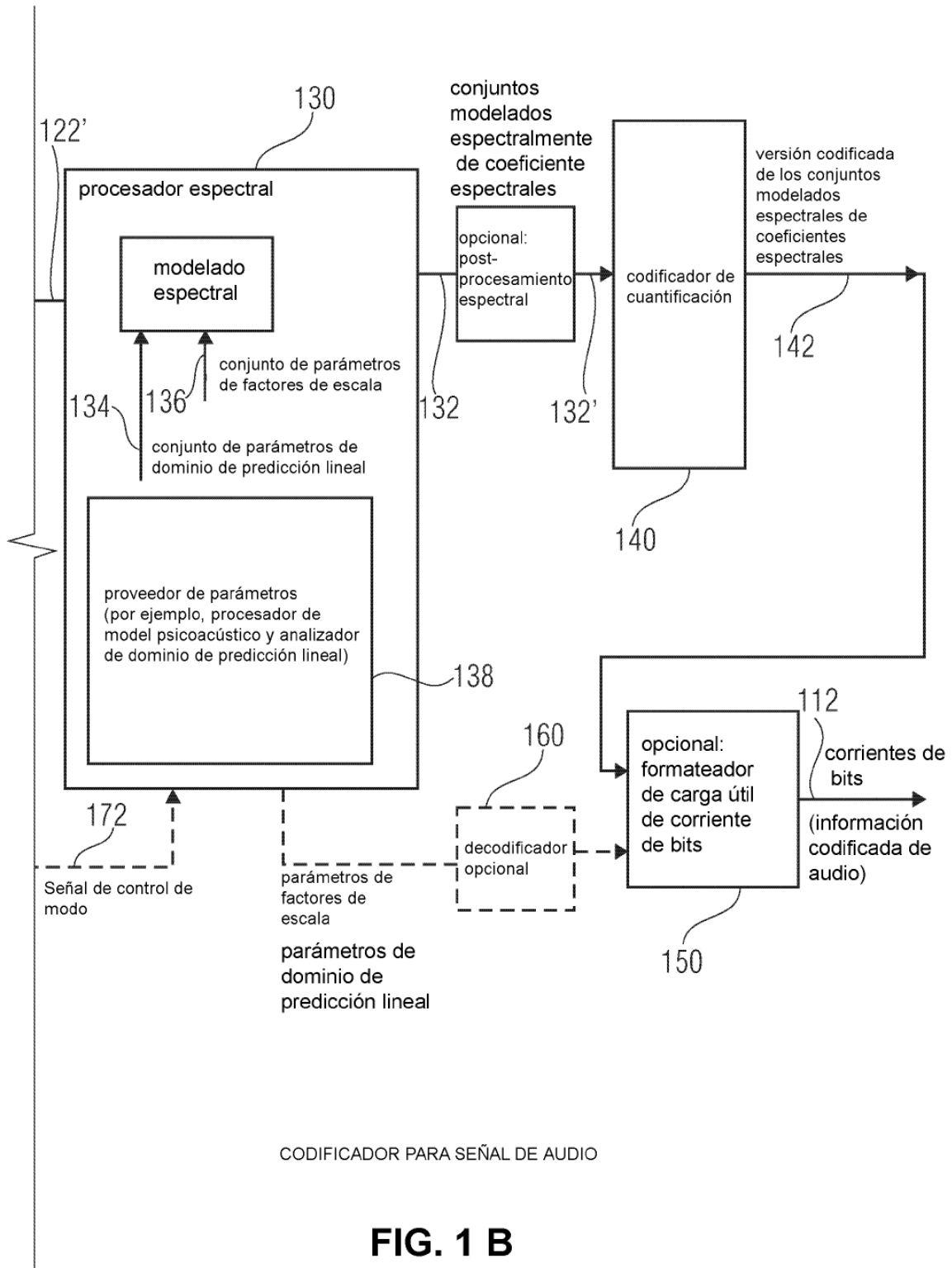


FIG. 1 B

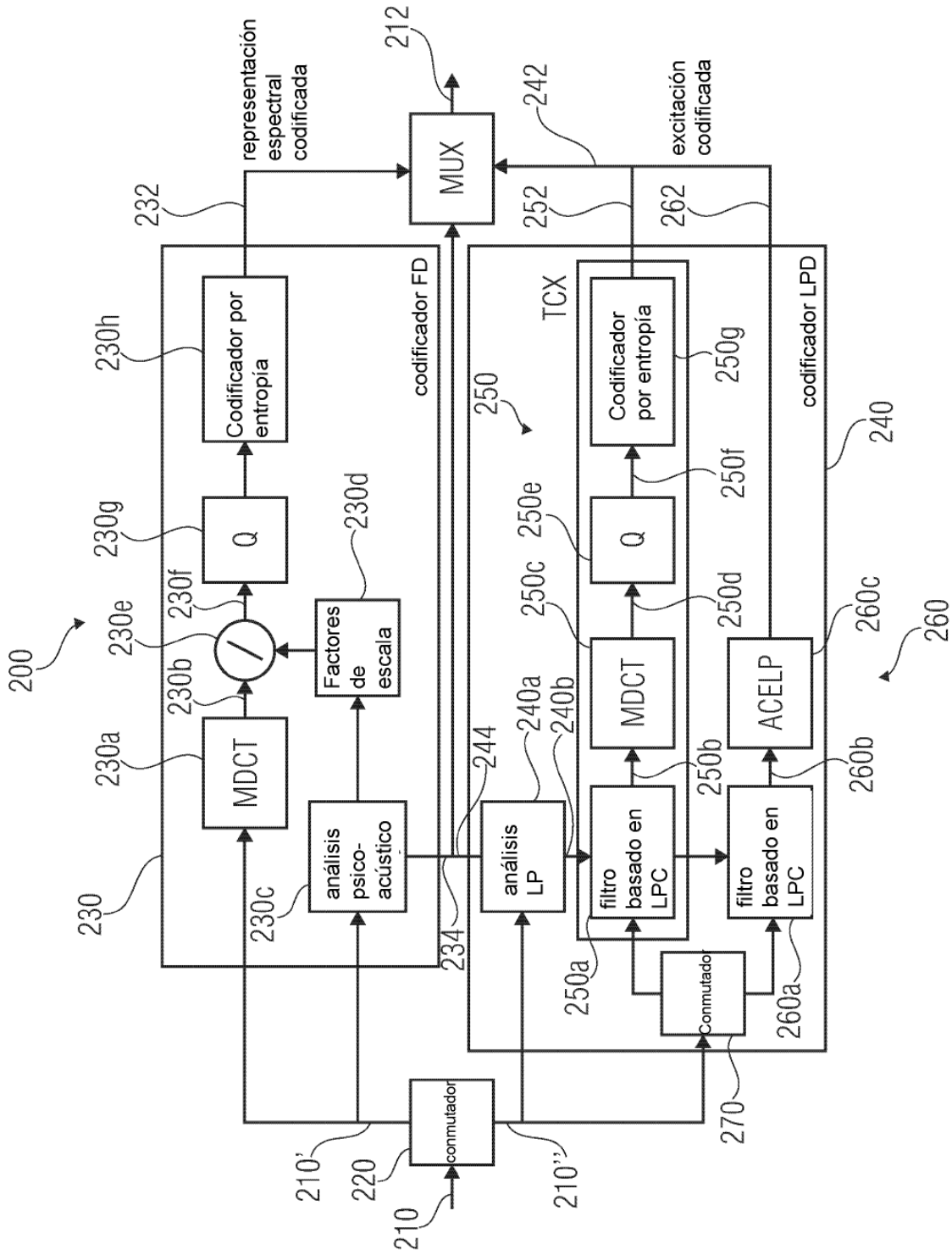


FIG. 2

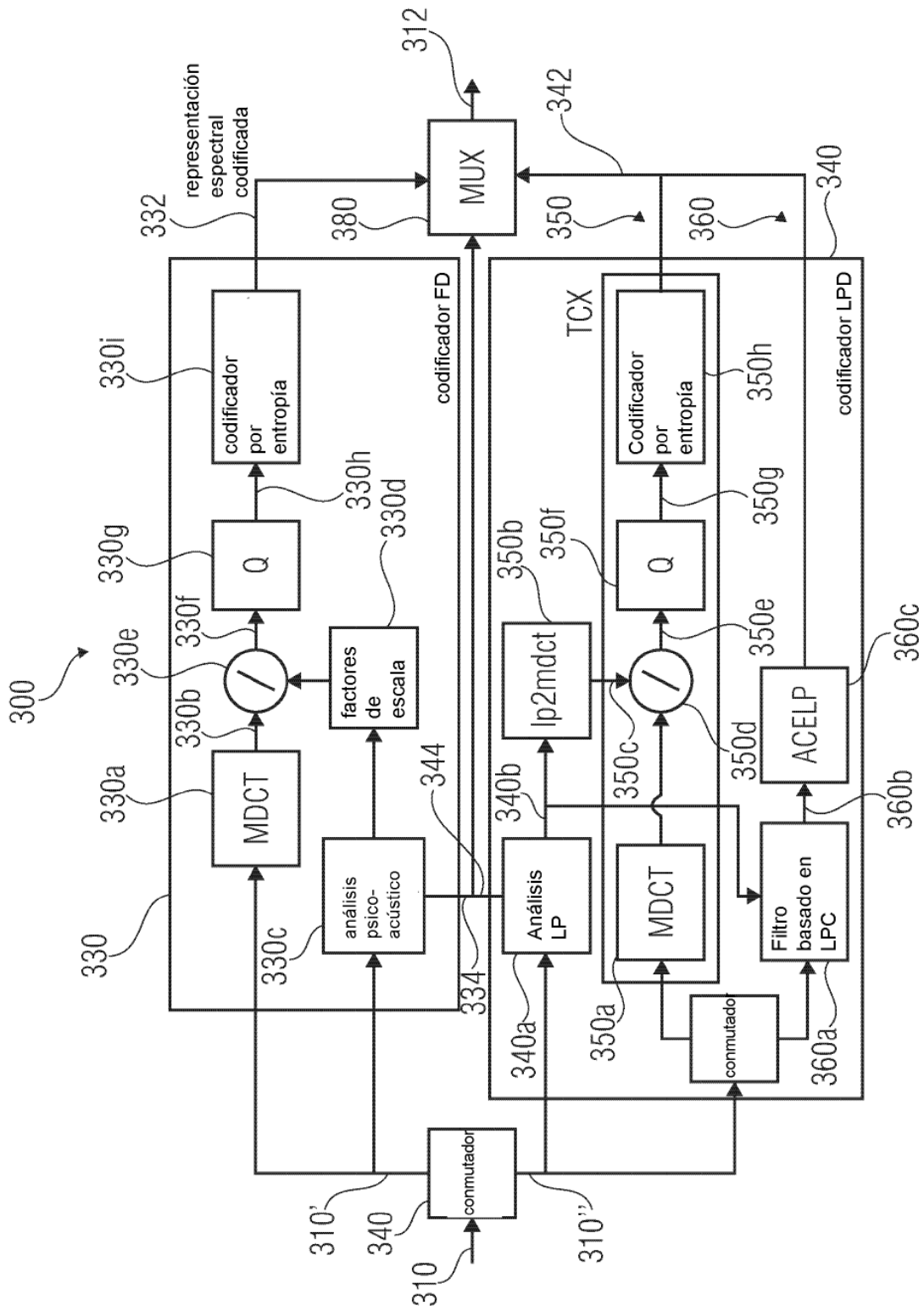


FIG. 3

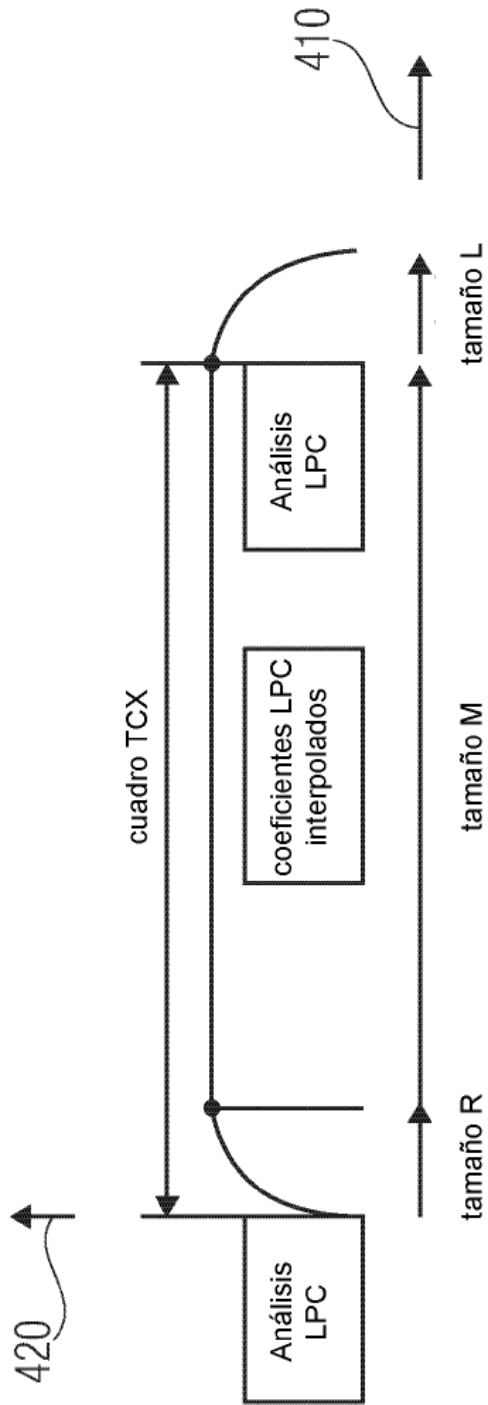


FIG. 4

```

Función lpc2mdct(input lpc_coefs, lpc_order,
    input sizeR, sizeM, sizeL,
    output mdct_scaleFactors) {
    int sizeN=2*(sizeL/2+sizeM+sizeR/2);
    float InRealData[sizeN];
    float OutRealData[sizeN];
    float InImagData[sizeN];
    float OutImagData[mdct_size];

    /*ODFT*/
    para(i=0; i<lpc_order; i++) {
        InRealData[i]=lpcCoefs[i]*cos(i*PI/(float)(sizeN));
        InImagData[i]=-lpcCoefs[i]*sin(i*PI/(float)(sizeN));
    }
    para(j<sizeN; j++) {
        InRealData[j]=0.f;
        InImagData[j]=0.f;
    }
    Complex_fft(InRealData, InImagData, OutRealData, OutImagData, sizeN);

    /*Get amplitude*/
    para(i=0; i<sizeN/2; i++) {
        mdct_scaleFactors[i]=sqrt(1/(OutRealData[i]*OutRealData[i]+OutImagData[i]*OutImagData[i]));
    }
}

```

FIG. 5

```

para (i=0; i < sizeL/2 + sizeM + sizeR/2; i++) {
    mdct[i] = mdct[i]/mdct_scaleFactors[i];
}
    
```

FIG. 6

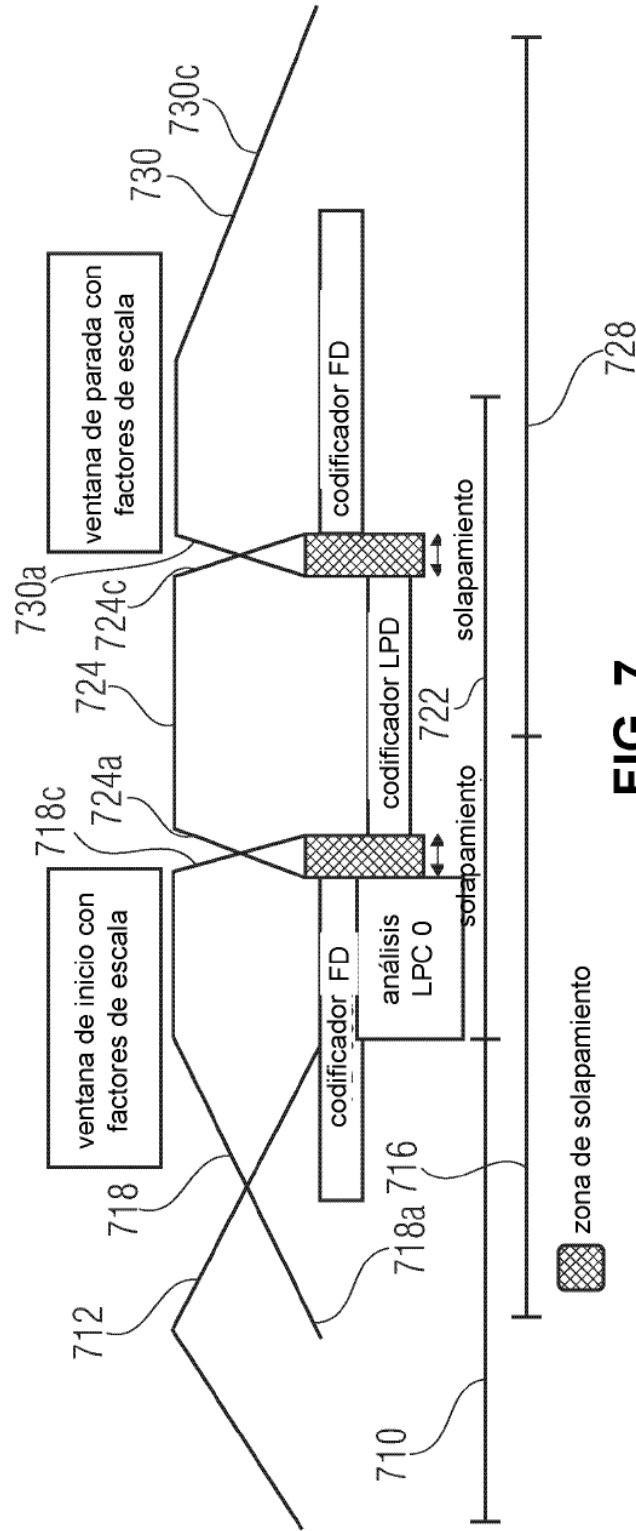


FIG. 7

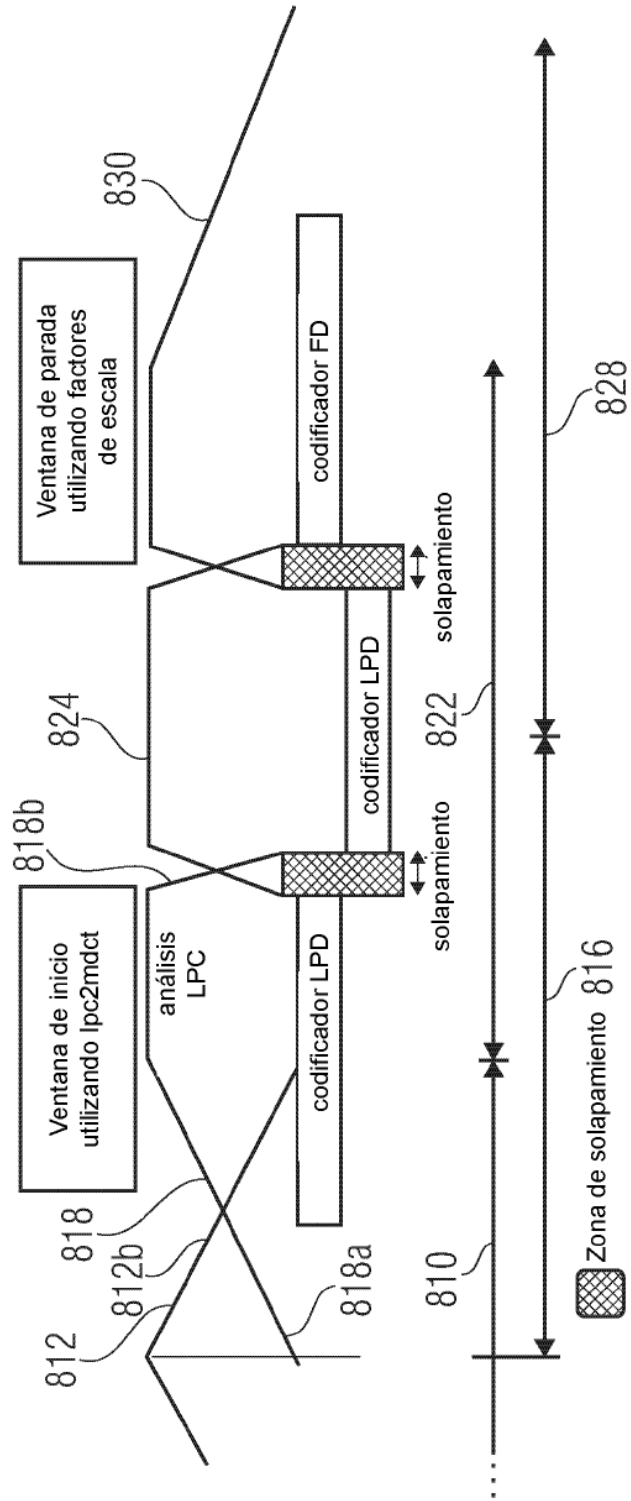


FIG. 8

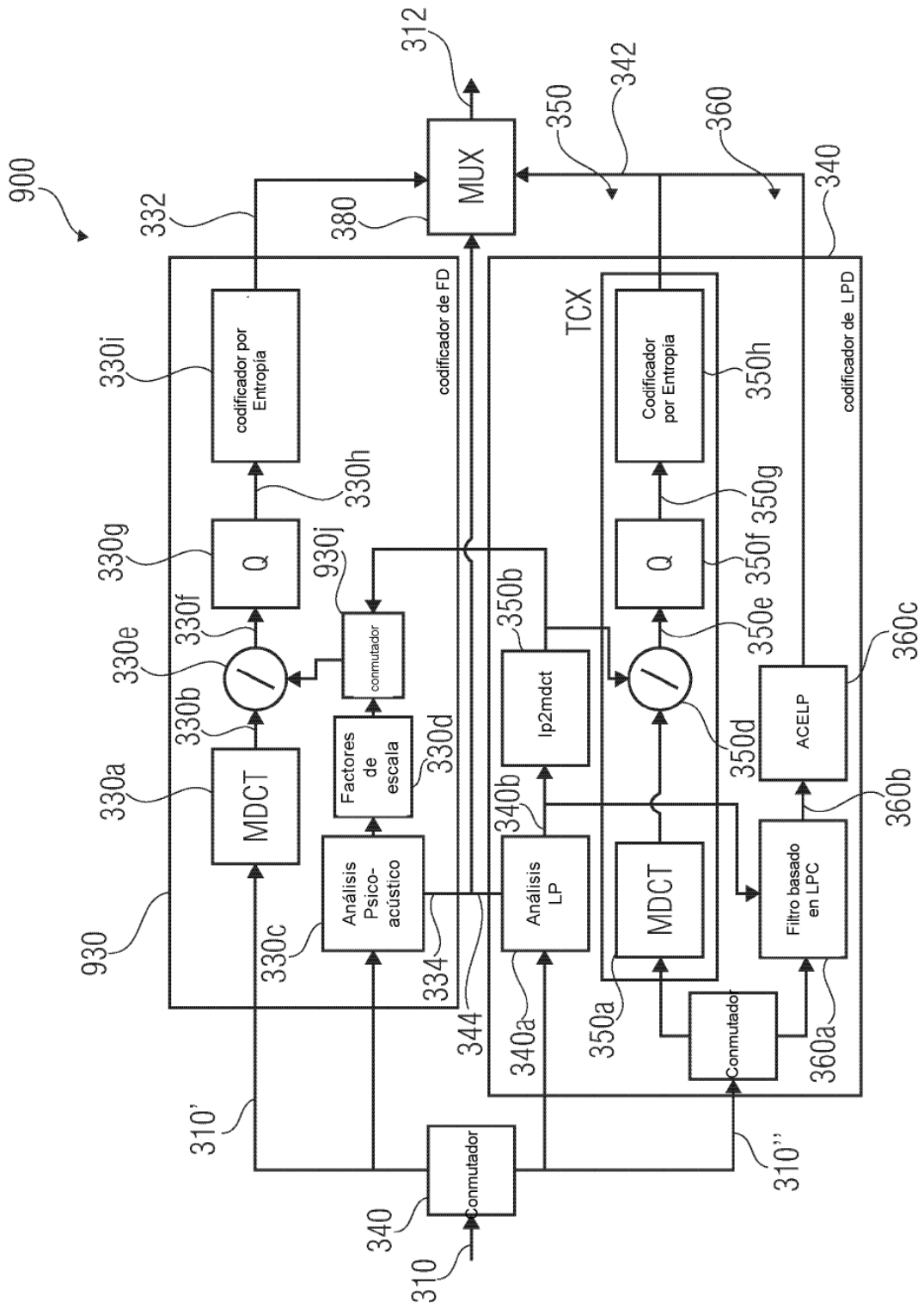


FIG. 9

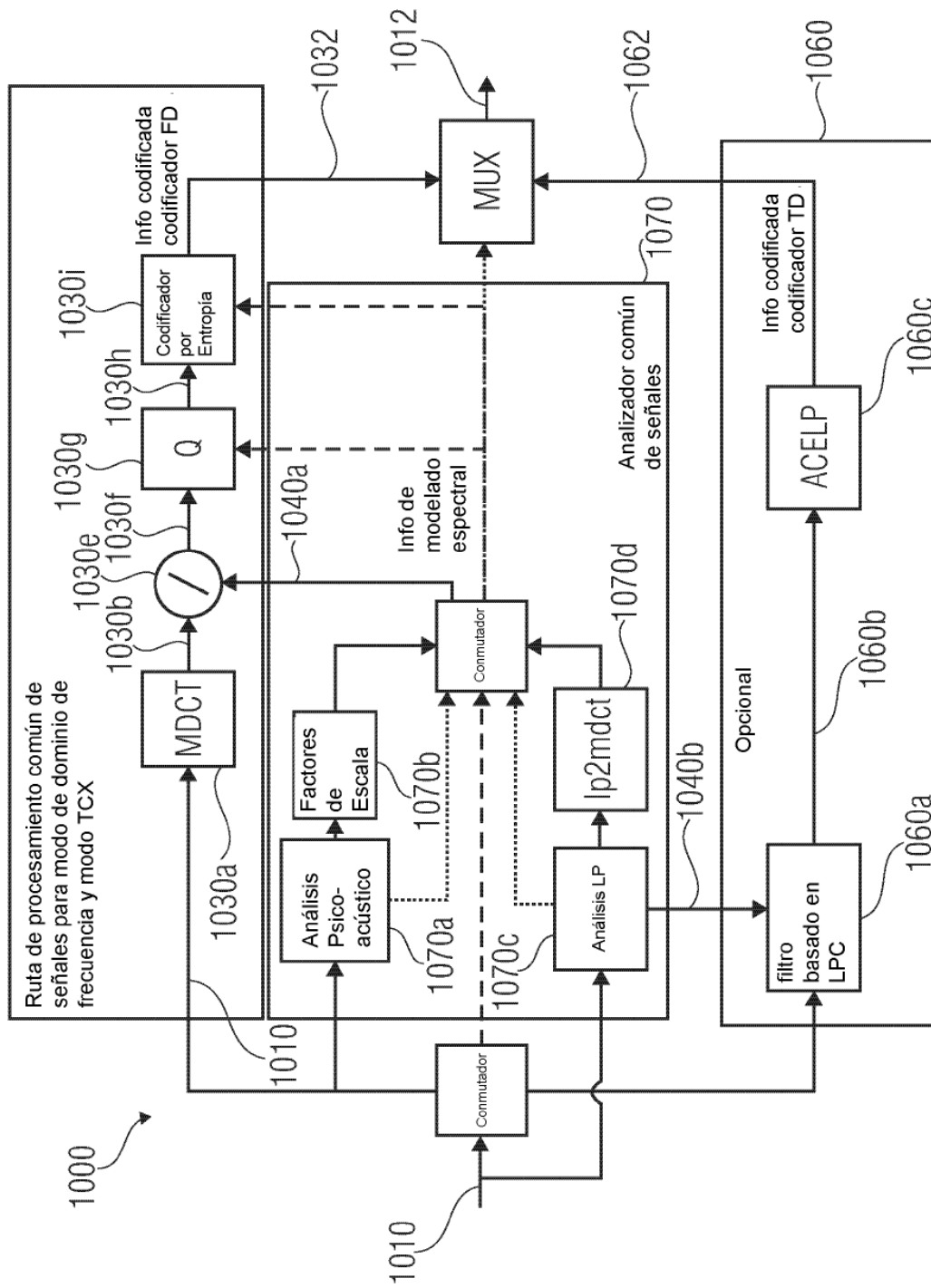
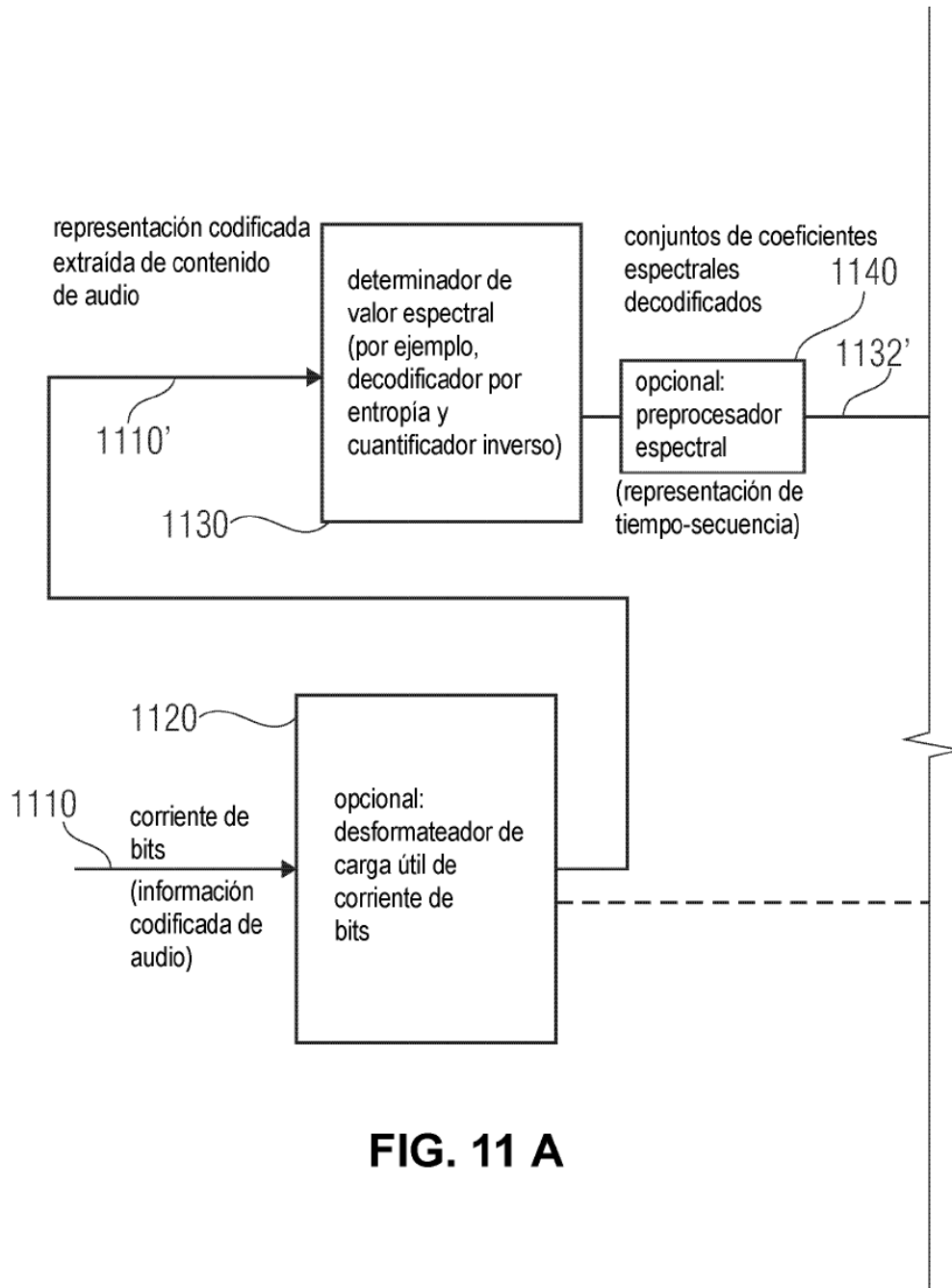


FIG. 10



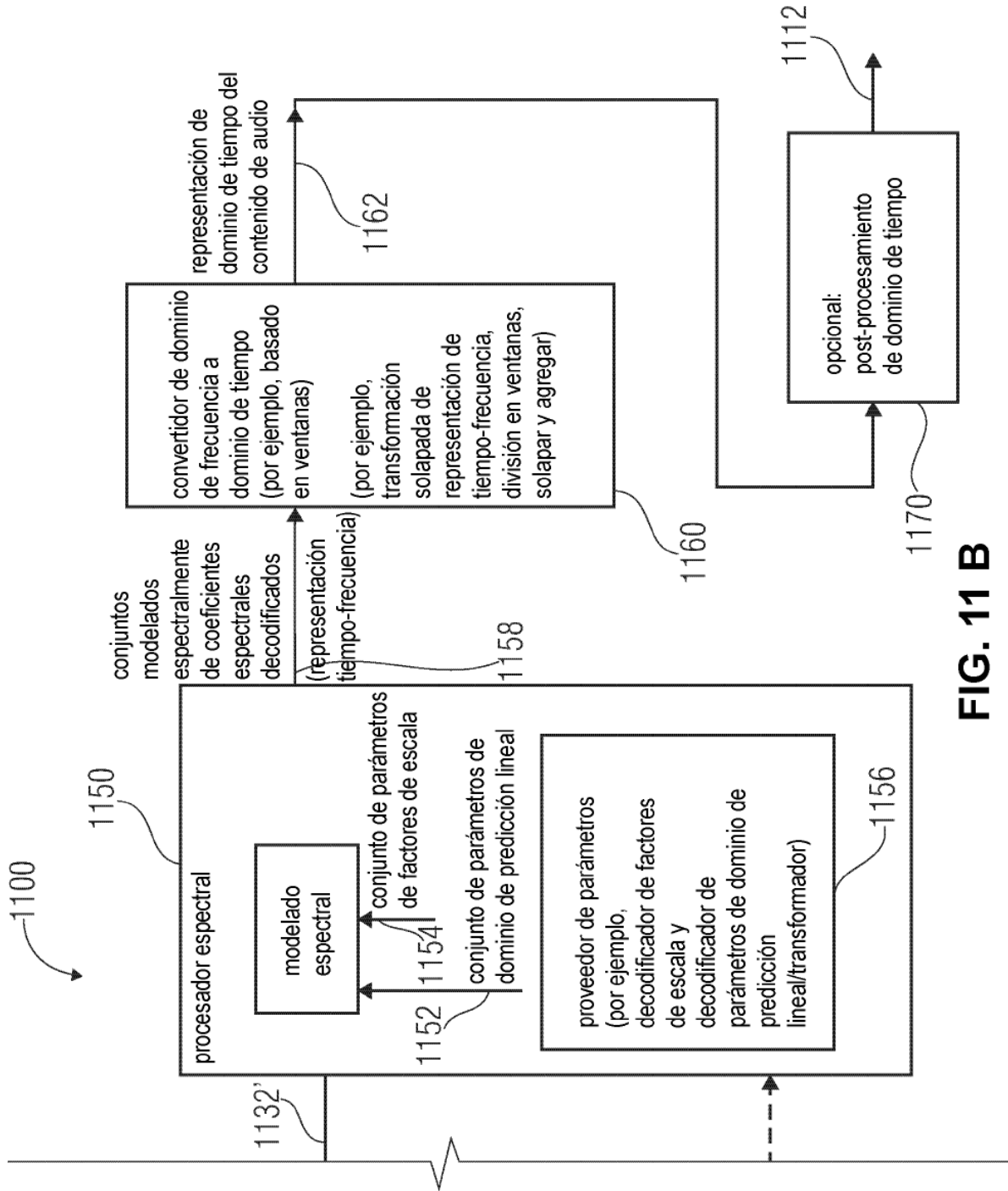


FIG. 11 B

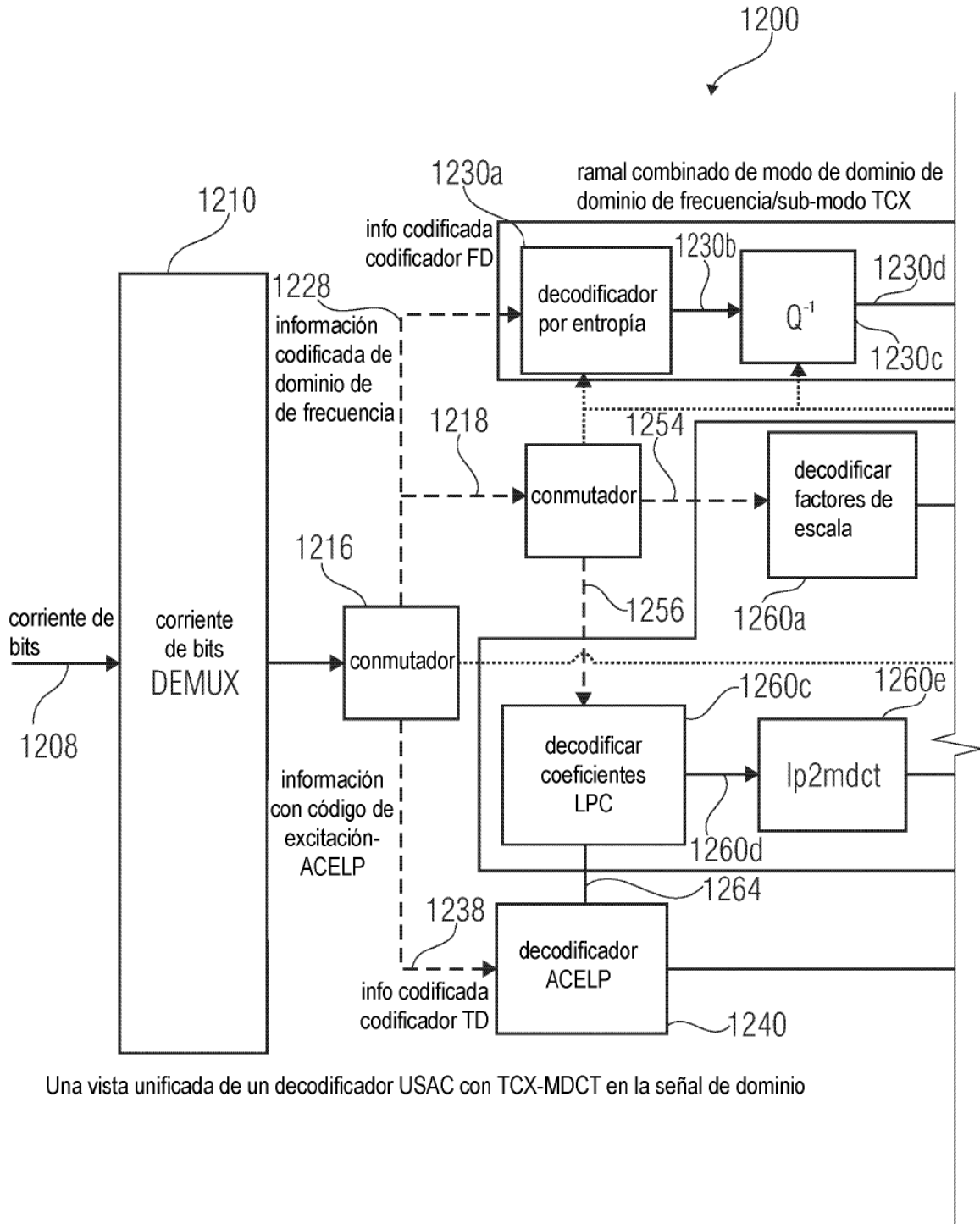
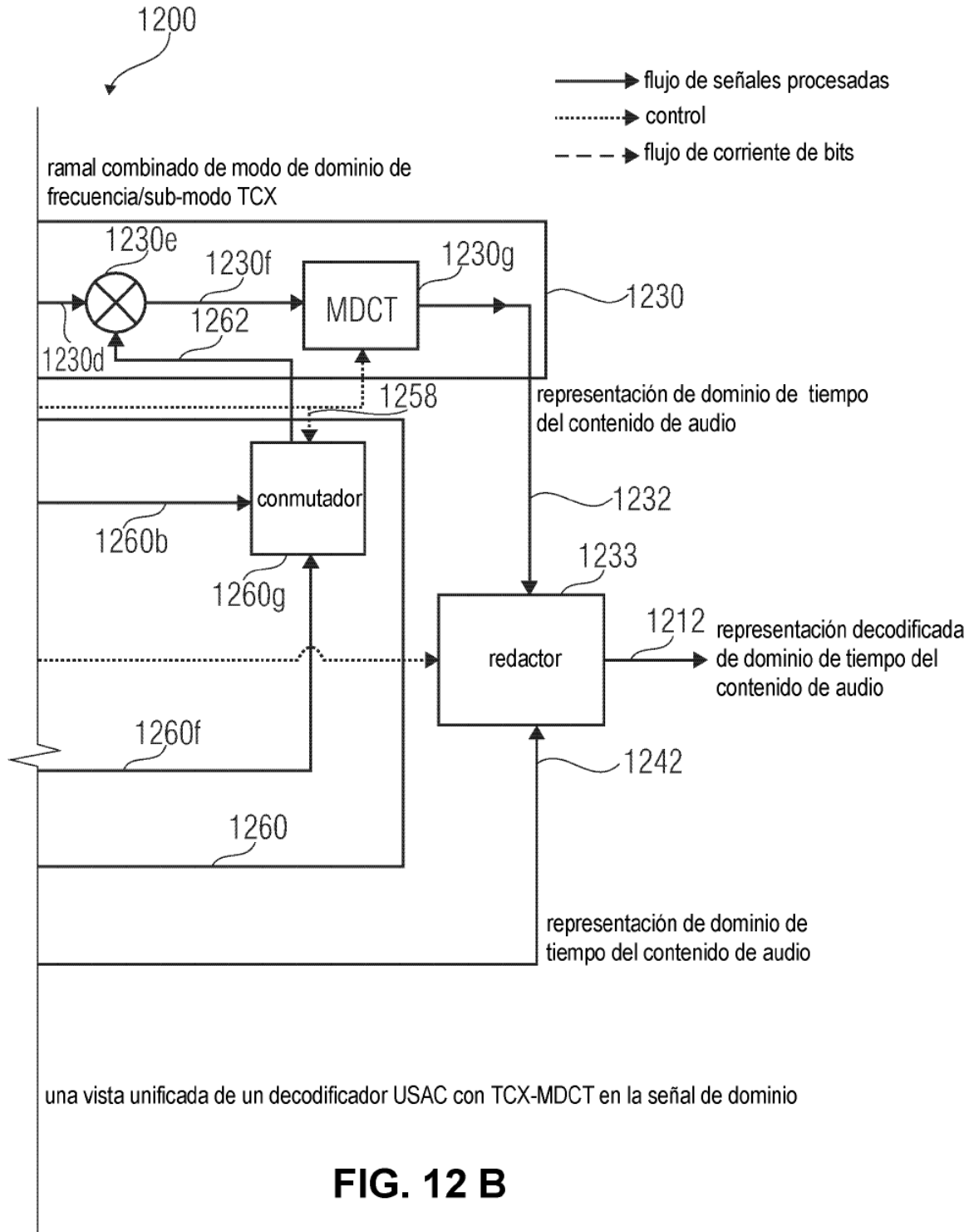
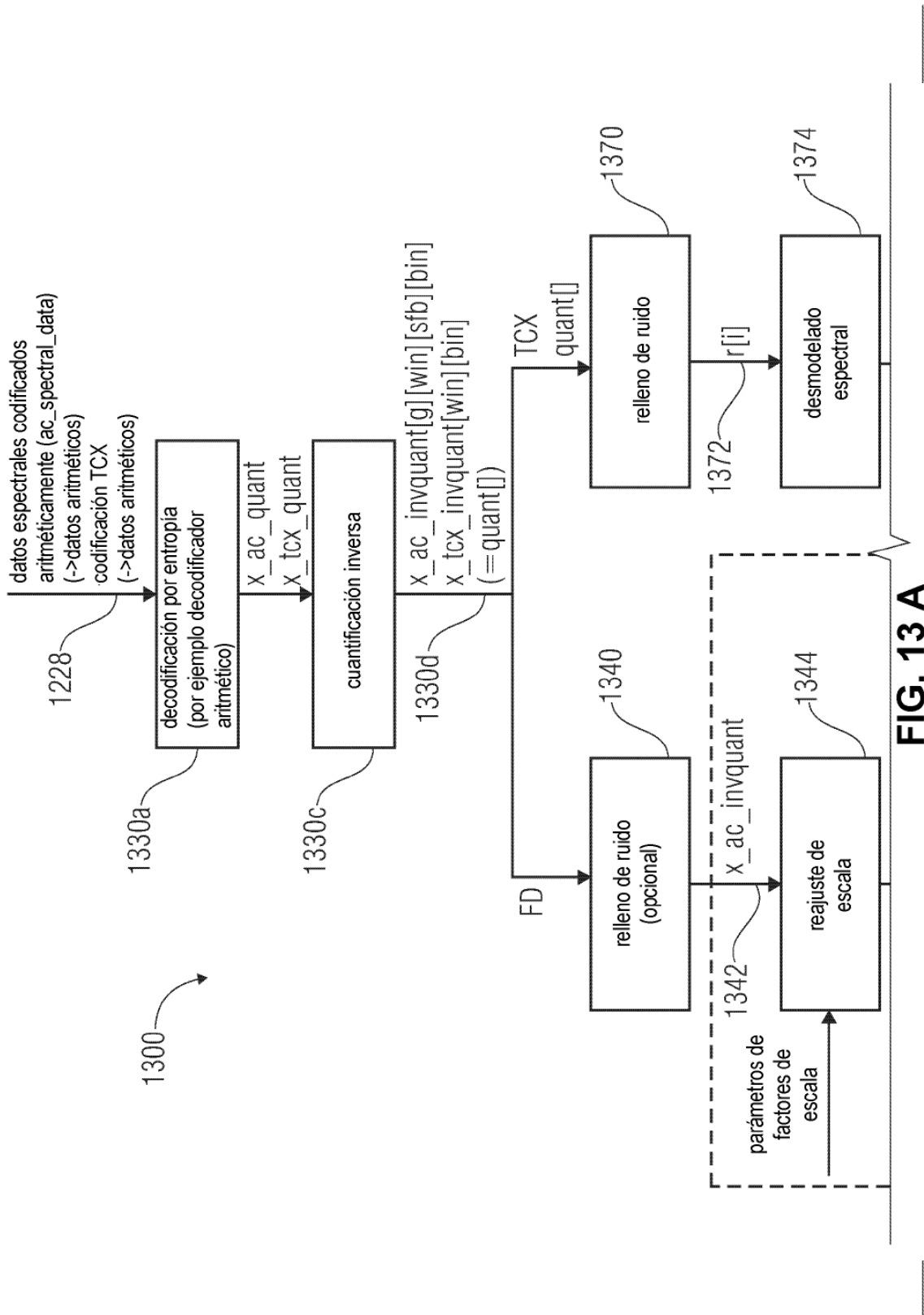


FIG. 12 A





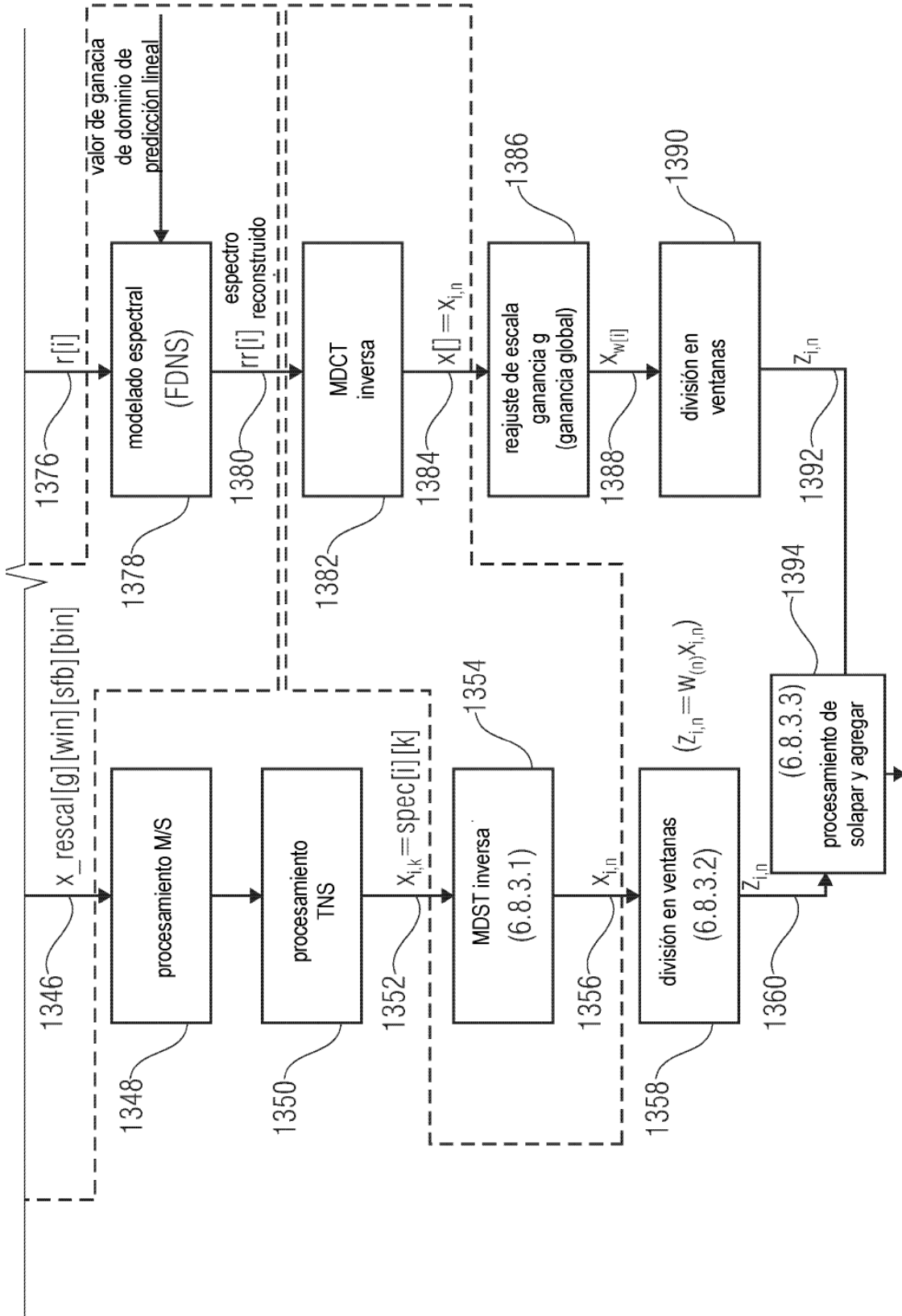


FIG. 13 B

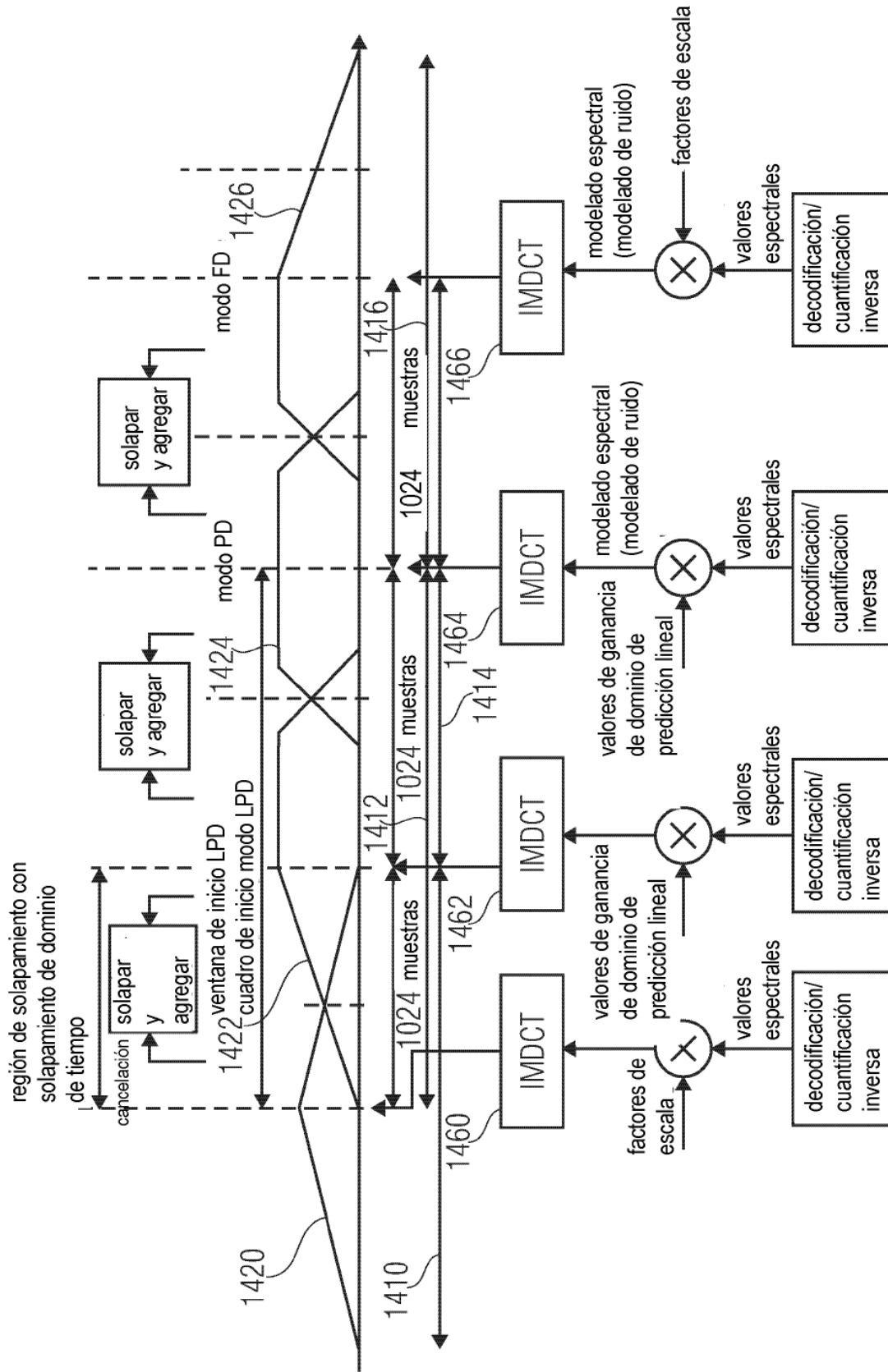


FIG. 14

Número de coeficientes espectrales en función de mod[]

valor de mod [x]	número lg de coeficientes espectrales	ZL	L	M	R	ZR	Número de muestras de dominio de tiempo por subcuadro o cuadro (incluyendo solapamiento)
1	256	0	256	0	256	0	512 (sub-cuadro)
2	512	128	256	256	256	128	1024 (sub-cuadro)
3	1024	384	256	768	256	384	2048 (cuadro completo)

FIG. 15

Secuencias de ventanas y ventanas de transformación





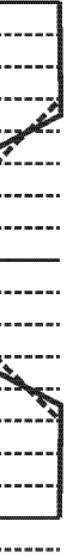
Valor	Ventana	# coef.	Forma de la Ventana
0	sólo secuencia larga ventana larga	1024/960	
1	secuencia larga de inicio ventana larga de inicio	1024/960	
2	secuencia de ocho ventanas cortas 8 ventana cortas	8*(128/120)	
3	secuencia larga de parada ventana larga de parada	1024/960	
1	secuencia de inicio-parada ventana de inicio parada	1024/960	

FIG. 16

Transiciones de ventana permitidas

Secuencia de ventanas Desde ↓ A →	sólo secuencia larga	secuencia larga de inicio	secuencia de ocho ventanas cortas	secuencia larga de parada	secuencia de inicio-parada	secuencia LPD
sólo secuencia larga	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
secuencia larga de inicio			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
secuencia de ocho ventanas cortas			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
secuencia larga de parada	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
secuencia de inicio-parada			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
secuencia LPD			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

FIG. 17 A

Transiciones de ventana permitidas

Secuencia de ventanas Desde ↓ A →	sólo secuencia larga	secuencia larga de inicio	secuencia de ocho ventanas cortas	secuencia larga de parada	secuencia de inicio-parada	secuencia LPD	secuencia de inicio LPD
sólo secuencia larga	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>
secuencia larga de inicio			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
secuencia de ocho ventanas cortas			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
secuencia larga de parada	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>
secuencia de inicio-parada			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
secuencia LPD			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
secuencia de inicio LPD						<input checked="" type="checkbox"/>	

FIG. 17 B

COEFICIENTES CODIFICADOS DE FILTRO LPC

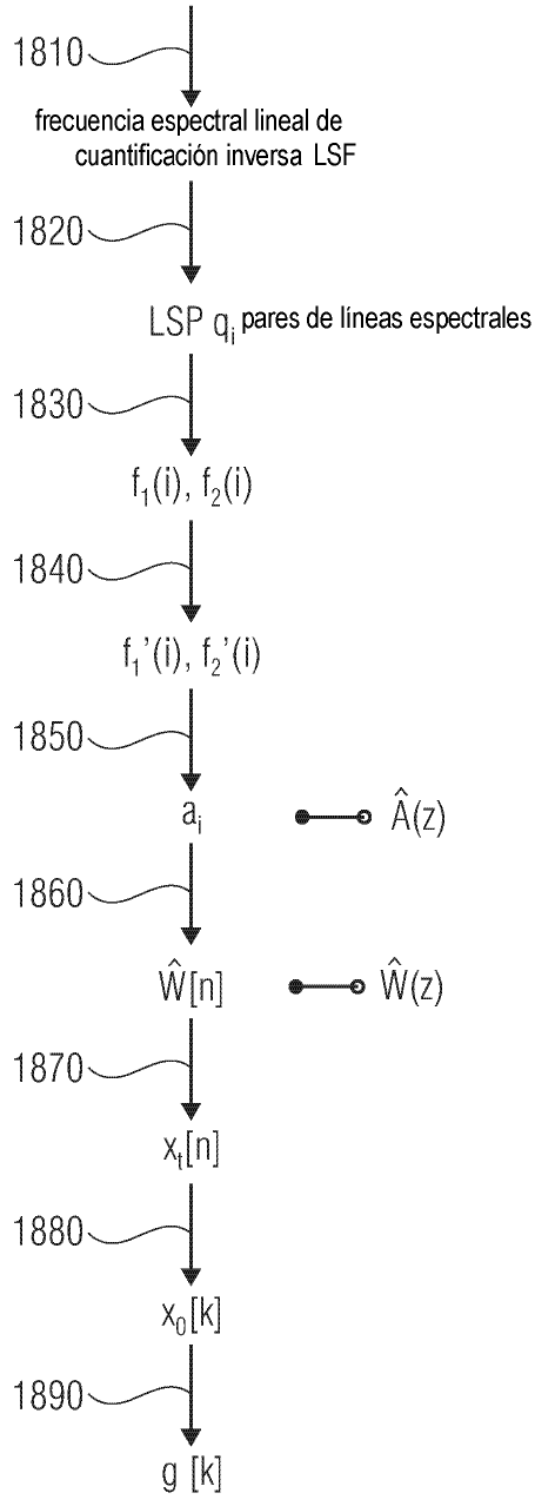


FIG. 18