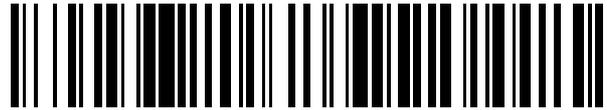


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 564 400**

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: **26.06.2009 E 09776858 (4)**

97 Fecha y número de publicación de la concesión europea: **06.01.2016 EP 2311032**

54 Título: **Codificador y descodificador de audio para codificar y descodificar muestras de audio**

30 Prioridad:

11.07.2008 US 79856

08.10.2008 US 103825

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
22.03.2016

73 Titular/es:

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

72 Inventor/es:

**LECOMTE, JÉRÉMIE;
GOURNAY, PHILIPPE;
BAYER, STEFAN;
MULTRUS, MARKUS;
BESSETTE, BRUNO y
GRILL, BERNHARD**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 564 400 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

Codificador y decodificador de audio para codificar y decodificar muestras de audio.

- 5 La presente invención se encuentra dentro del campo de la codificación de audio en diferentes dominios de codificación, por ejemplo, en el dominio del tiempo y en un dominio de la transformada.

10 En el contexto de la tecnología de codificación del habla y de audio de baja tasa de bits, se han empleado tradicionalmente diferentes técnicas de codificación con el fin de lograr una codificación de baja tasa de bits de tales señales con la mejor calidad subjetiva posible a una tasa de bits dada. Los codificadores para señales de música general / sonido buscan optimizar la calidad subjetiva mediante la conformación de una forma espectral (y temporal) del error de cuantificación de acuerdo con una curva de umbral de enmascaramiento que se estima a partir de la señal de entrada por medio de un modelo perceptual (“codificación de audio perceptual”). Por otro lado, se ha mostrado que la codificación del habla a muy bajas tasas de bits funciona de manera muy eficiente cuando se basa en un modelo de producción del habla humana, es decir, que emplea codificación predictiva lineal (LPC, *Linear Predictive Coding*) para los modelar los efectos resonantes del tracto vocal humano junto con una codificación eficiente de la señal de excitación residual.

20 Como consecuencia de estos dos enfoques diferentes, los codificadores de audio generales, como MPEG-1 Capa 3 (MPEG = *Moving Pictures Expert Group*, Grupo de Expertos en Imágenes en Movimiento), o MPEG-2 / 4 Codificación de Audio Avanzada (AAC, *Advanced Audio Coding*), en general, no funcionan tan bien para las señales del habla a muy bajas tasas de datos como los codificadores de habla basados en LPC dedicados, debido a la falta de aprovechamiento de un modelo fuente del habla. A la inversa, los codificadores de habla basados en LPC, en general, no alcanzan unos resultados convincentes cuando se aplican a señales de música generales debido a su incapacidad para dar forma de manera flexible a la envolvente espectral de la distorsión de codificación de acuerdo con una curva de umbral de enmascaramiento. En lo sucesivo, se describen conceptos que combinan las ventajas tanto de la codificación basada en LPC como de la codificación de audio perceptual en un mismo marco y por lo tanto describen una codificación de audio unificada que es eficiente para las señales tanto de audio general como de habla.

30 Tradicionalmente, los codificadores de audio perceptuales usan un enfoque basado en banco de filtros para codificar de manera eficiente señales de audio y dar forma a la distorsión de cuantificación de acuerdo con una estimación de la curva de enmascaramiento.

35 La figura 16a muestra un diagrama de bloques básico de un sistema de codificación perceptual monofónico. Se usa un banco de filtros de análisis 1600 para establecer una correspondencia de las muestras en el dominio del tiempo con componentes espectrales submuestreados. Dependiente del número de componentes espectrales, también se hace referencia al sistema como codificador de sub-bandas (número pequeño de sub-bandas, por ejemplo, 32) o codificador de transformada (número grande de líneas de frecuencia, por ejemplo, 512). Se usa un modelo perceptual (“psicoacústico”) 1602 para estimar el umbral real de enmascaramiento dependiente de tiempo. Las componentes espectrales (de “sub-banda” o de “dominio de la frecuencia”) se cuantifican y se codifican 1604 de manera que el ruido de cuantificación se esconde bajo la señal transmitida real y no es perceptible después de la decodificación. Esto se logra mediante la variación de la granularidad de cuantificación de los valores espectrales sobre el tiempo y la frecuencia.

45 Los valores de sub-banda o coeficientes espectrales cuantificados y codificados por entropía se introducen junto con una información complementaria en un formateador de tren de bits 1606, el cual proporciona una señal de audio codificada que es adecuada para transmitirse o almacenarse. El tren de bits de salida del bloque 1606 se puede transmitir por medio de Internet o se puede almacenar en cualquier soporte de datos legible por máquina.

50 En el lado de decodificador, una interfaz de entrada de decodificador 1610 recibe el tren de bits codificado. El bloque 1610 separa los valores espectrales / de sub-banda codificados por entropía y cuantificados de la información complementaria. Los valores espectrales codificados se introducen en un decodificador de entropía como un decodificador de Huffman, que se coloca entre 1610 y 1620. Las salidas de este decodificador de entropía son valores espectrales cuantificados. Estos valores espectrales cuantificados se introducen en un recuantificador, el cual realiza una cuantificación “inversa” como se indica en 1620 en la figura 16a. La salida del bloque 1620 se introduce en un banco de filtros de síntesis 1622, el cual realiza una filtración de síntesis incluyendo una transformada de frecuencia / tiempo y, por lo general, una operación de cancelación de distorsión por repliegue del espectro en el dominio del tiempo como una superposición o agregación y / o una operación de aplicación de función ventana de lado de síntesis para obtener finalmente la señal de audio de salida.

60 Tradicionalmente, la codificación del habla eficiente se ha basado en la codificación predictiva lineal (LPC) para modelar los efectos resonantes del tracto vocal humano junto con una codificación eficiente de la señal de excitación residual. Los parámetros tanto de LPC como de excitación se transmiten del codificador al decodificador. Este principio se ilustra en las figuras 17a y 17b.

La figura 17a indica el lado de codificador de un sistema de codificación / descodificación basado en la codificación predictiva lineal. La entrada de habla se introduce en un analizador de LPC 1701, el cual proporciona, en su salida, unos coeficientes de filtro de LPC. Sobre la base de estos coeficientes de filtro de LPC, se ajusta un filtro de LPC 1703. El filtro de LPC emite una señal de audio espectralmente blanqueada, que también se denomina "señal de error de predicción". Esta señal de audio espectralmente blanqueada se introduce en un codificador residual / de excitación 1705, el cual genera unos parámetros de excitación. Por lo tanto, la entrada de habla se codifica en parámetros de excitación, por un lado, y coeficientes de LPC, por otro lado.

En el lado de descodificador ilustrado en la figura 17b, los parámetros de excitación se introducen en un descodificador de excitación 1707, el cual genera una señal de excitación, la cual puede introducirse en un filtro de síntesis de LPC. El filtro de síntesis de LPC se ajusta usando los coeficientes de filtro de LPC transmitidos. Por lo tanto, el filtro de síntesis de LPC 1709 genera una señal de salida de habla reconstruida o sintetizada.

Con el tiempo, se han propuesto muchos métodos con respecto a una representación eficiente y perceptualmente convincente de la señal residual (de excitación), como Excitación de Multi-Pulso (MPE, *Multi-Pulse Excitation*), Excitación de Pulso Regular (RPE, *Regular Pulse Excitation*) y Predicción Lineal Excitada por Código (CELP, *Code-Excited Linear Prediction*).

La codificación predictiva lineal intenta producir una estimación del valor de muestra actual de una secuencia basándose en la observación de un cierto número de valores pasados como una combinación lineal de las observaciones pasadas. Con el fin de reducir la redundancia en la señal de entrada, el filtro de LPC de codificador "blanquea" la señal de entrada en su envolvente espectral, es decir, esta es un modelo de la inversa de la envolvente espectral de la señal. A la inversa, el filtro de síntesis de LPC de descodificador es un modelo de la envolvente espectral de la señal. En concreto, se sabe que el bien conocido análisis predictivo lineal autorregresivo (AR) modela la envolvente espectral de la señal por medio de una aproximación a todos los polos.

Por lo general, los codificadores de habla de banda angosta (es decir, codificadores de habla con una tasa de muestreo de 8 kHz) emplean un filtro de LPC con un orden de entre 8 y 12. Debido a la naturaleza del filtro de LPC, una resolución de frecuencia uniforme es efectiva a través de la totalidad del rango de frecuencias. Esto no se corresponde con una escala de frecuencias perceptual.

Con el fin de combinar las fortalezas de la codificación basada en LPC / CELP tradicional (la mejor calidad para señales de habla) y el enfoque de codificación de audio perceptual basado en banco de filtros tradicional (el mejor para música), se ha propuesto una codificación combinada entre estas arquitecturas. En el codificador de AMR-WB+ (AMR-WB = *Adaptive Multi-Rate WideBand*, banda ancha de multivelocidad adaptativo) B. Bessette, R. Lefebvre, R. Salami, "UNIVERSAL SPEECH / AUDIO CODING USING HYBRID ACELP / TCX TECHNIQUES," Proc. IEEE ICASSP 2005, páginas 301 - 304, 2005 dos núcleos de codificación alternativos operan sobre una señal residual de LPC. Uno está basado en ACELP (ACELP = *Algebraic Code Excited Linear Prediction*, Predicción Lineal por Excitación con Código Algebraico) y, por lo tanto, es extremadamente eficiente para la codificación de señales del habla. El otro núcleo de codificación está basado en TCX (TCX = *Transform Coded Excitation, excitación codificada de transformada*), es decir, un enfoque de codificación basado en banco de filtros que se asemeja a las técnicas de codificación de audio tradicionales con el fin de lograr una buena calidad para las señales musicales. Dependiendo de las características de las señales de entrada, uno de los dos modos de codificación se selecciona durante un periodo corto de tiempo para transmitir la señal residual de LPC. De esta forma, tramas de 80 ms de duración pueden dividirse en unas subtramas de 40 ms o de 20 ms en las cuales se toma una decisión entre los dos modos de codificación.

La AMR-WB+ (AMR-WB+ = *extended Adaptive Multi-Rate WideBand*, códec de Multivelocidad Adaptativo extendido), véase 3GPP (3GPP = *Third Generation Partnership Project*, Proyecto de Asociación de Tercera Generación) especificación técnica número 26.290, versión 6.3.0, junio de 2005, puede conmutar entre los dos modos esencialmente diferentes de ACELP y de TCX. En el modo de ACELP, una señal en el dominio del tiempo se codifica mediante excitación de código algebraico. En el modo de TCX se usa una transformada rápida de Fourier (FFT = *fast Fourier transform*) y los valores espectrales de la señal ponderada de LPC (de la que se puede obtener la excitación de LPC) se codifican sobre la base de una cuantificación vectorial.

La decisión de qué modos usar se puede tomar al probar y descodificar las dos opciones y comparar las relaciones de señal / ruido (SNR = *Signal-to-Noise Ratio*) segmentarias resultantes.

Este caso también se denomina decisión de lazo cerrado, ya que hay un lazo de control cerrado, que evalúa tanto las eficiencias como los rendimientos de codificación, respectivamente, y que elige entonces la que tenga la mejor SNR.

Se sabe bien que, para las aplicaciones de codificación de audio y habla, no es posible una transformada de bloque sin aplicación de una función ventana. Por lo tanto, para el modo de TCX se aplica una función ventana a la señal con una ventana de superposición baja con una superposición de 1 / 8. Esta región de superposición es necesaria con el fin del desvanecimiento gradual de un bloque o trama anterior al tiempo que aparece gradualmente el

próximo, por ejemplo, para suprimir artefactos debido a un ruido de cuantificación no correlacionado en tramas de audio consecutivas. De esta manera, la tara en comparación con un muestreo no crítico se mantiene razonablemente baja y la descodificación necesaria para la decisión de lazo cerrado reconstruye por lo menos 7 / 8 de las muestras de la trama actual.

5 La AMR-WB+ introduce 1 / 8 de tara en un modo de TCX, es decir, el número de valores espectrales que se va a codificar es 1 / 8 mayor que el número de muestras de entrada. Esto proporciona la desventaja de una tara de datos mayor. Asimismo, la respuesta de frecuencia de los filtros de paso de banda correspondientes no es ventajosa debido a la abrupta región de superposición de 1 / 8 de las tramas consecutivas.

10 Con el fin de profundizar en el detalle de la tara de código y en la superposición de tramas consecutivas, la figura 18 ilustra una definición de parámetros de ventana. La ventana que se muestra en la figura 18 tiene una parte de flanco de subida en el lado izquierdo, la cual se indica con "L" y también se denomina región de superposición izquierda, una región central que se indica con "1", la cual también se denomina región de 1 o parte de derivación, y una parte de flanco de bajada, que se indica con "R" y también se denomina la región de superposición derecha. Además, la figura 18 muestra una flecha que indica la región "PR" de reconstrucción perfecta dentro de una trama. Además, la figura 18 muestra una flecha que indica la longitud del núcleo de transformada, que se indica con "T".

15 La figura 19 muestra una vista gráfica de una secuencia de ventanas de AMR-WB+ y, en la parte inferior, una tabla de parámetros de ventana de acuerdo con la figura 18. La secuencia de ventanas que se muestra en la parte superior de la figura 19 es ACELP, TCX20 (para una trama de 20 ms de duración), TCX20, TCX40 (para una trama de 40 ms de duración), TCX80 (para una trama de 80 ms de duración), TCX20, TCX20, ACELP, ACELP.

20 A partir de la secuencia de ventanas se pueden ver las regiones de superposición variadas, las cuales se superponen 1 / 8 exacto de la parte central M. La tabla en la parte inferior de la figura 19 también muestra que la longitud de transformada "T" es siempre 1 / 8 más grande que la región de muestras nuevas perfectamente reconstruidas "PR". Asimismo, se ha de hacer notar que no es solo el caso de las transiciones de ACELP a TCX, sino también de las transiciones de TCXx a TCXx (en donde "x" indica tramas de TCX de una longitud arbitraria). Por lo tanto, en cada bloque se introduce una tara de 1 / 8, es decir, nunca se alcanza el muestreo crítico.

25 Cuando se conmuta de TCX a ACELP, las muestras de ventana se descartan de la trama de FFT-TCX en la región de superposición, como se indica, por ejemplo, en la parte superior de la figura 19 mediante la región marcada con 1900. Cuando se conmuta de ACELP a TCX la respuesta de entrada cero (*ZIR = Zero Input Response*), la cual también se indica mediante la línea de puntos 1910 en la parte superior de la figura 19, se elimina en el codificador antes de la aplicación de una función ventana y se agrega en el descodificador para la recuperación. Cuando se conmuta de tramas de TCX a TCX las muestras a las que se ha aplicado una función ventana se usan para un desvanecimiento cruzado. Dado que las tramas de TCX se pueden cuantificar de manera diferente, el error de cuantificación o ruido de cuantificación entre tramas consecutivas puede ser diferente y / o independiente. Con lo anterior, cuando se conmuta de una trama a la siguiente sin desvanecimiento cruzado, pueden ocurrir artefactos perceptibles y, en consecuencia, el desvanecimiento cruzado es necesario con el fin de lograr una cierta calidad.

30 Cuando se conmuta de TCX a ACELP, las muestras de ventana se descartan de la trama de FFT-TCX en la región de superposición, como se indica, por ejemplo, en la parte superior de la figura 19 mediante la región marcada con 1900. Cuando se conmuta de ACELP a TCX la respuesta de entrada cero (*ZIR = Zero Input Response*), la cual también se indica mediante la línea de puntos 1910 en la parte superior de la figura 19, se elimina en el codificador antes de la aplicación de una función ventana y se agrega en el descodificador para la recuperación. Cuando se conmuta de tramas de TCX a TCX las muestras a las que se ha aplicado una función ventana se usan para un desvanecimiento cruzado. Dado que las tramas de TCX se pueden cuantificar de manera diferente, el error de cuantificación o ruido de cuantificación entre tramas consecutivas puede ser diferente y / o independiente. Con lo anterior, cuando se conmuta de una trama a la siguiente sin desvanecimiento cruzado, pueden ocurrir artefactos perceptibles y, en consecuencia, el desvanecimiento cruzado es necesario con el fin de lograr una cierta calidad.

35 A partir de la tabla en la parte inferior de la figura 19, se puede ver que la región de desvanecimiento cruzado crece con una longitud creciente de la trama. La figura 20 proporciona otra tabla con ilustraciones de las diferentes ventanas para las posibles transiciones en AMR-WB+. Cuando se realiza una transición de TCX a ACELP se pueden descartar las muestras de superposición. Cuando se realiza una transición de ACELP a TCX, se puede eliminar en el codificador la respuesta de entrada cero a partir de la ACELP y se agrega el descodificador para la recuperación.

40 A continuación se arrojará luz sobre la codificación de audio, que usa una codificación en el dominio del tiempo (*TD = Time Domain*) y en el dominio de la frecuencia (*FD = Frequency Domain*). Asimismo, entre los dos dominios de codificación, se puede usar la conmutación. En la figura 21, se muestra una línea de tiempo durante la cual una primera trama 2101 es codificada por un codificador de FD, seguida por otra trama 2103, que es codificada por un codificador de TD y que se superpone en la región 2102 con la primera trama 2101. La trama codificada en el dominio del tiempo 2103 es seguida por una trama 2105, la cual es codificada de nuevo en el dominio de la frecuencia y que se superpone en la región 2104 con la trama precedente 2103. Las regiones de superposición 2102 y 2104 ocurren siempre que se conmute el dominio de codificación.

45 El fin de estas regiones de superposición es suavizar las transiciones. Sin embargo, las regiones de superposición pueden aún ser propensas a una pérdida de la eficiencia de codificación y a artefactos. Por lo tanto, las regiones de superposición o transiciones se eligen, a menudo, como un compromiso entre una cierta tara de la información transmitida, es decir, la eficiencia de codificación, y la calidad de la transición, es decir, la calidad de audio de la señal descodificada. Con el fin de establecer este compromiso, se ha de tener cuidado cuando se manipulan las transiciones y se diseñan las ventanas de transición 2111, 2113 y 2115 como se indica en la figura 21.

65 Son conceptos convencionales relativos a la manipulación de transiciones entre los modos de codificación en el dominio de la frecuencia y en el dominio del tiempo, por ejemplo, el uso de ventanas de desvanecimiento cruzado,

es decir, la introducción de una tara tan grande como la región de superposición. Se usa una ventana de desvanecimiento cruzado, desvaneciéndose gradualmente la trama precedente y apareciendo gradualmente la siguiente trama de forma simultánea. Este enfoque, debido a su tara, introduce deficiencias en una eficiencia de descodificación, dado que siempre que tenga lugar una transición, la señal no vuelve más a muestrearse críticamente. Se divulgan transformadas solapadas muestreadas críticamente, por ejemplo, en J. Princen, A. Bradley, "Analysis / Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation", IEEE Trans. ASSP, ASSP-34 (5): 1153 - 1161, 1986, y se usan, por ejemplo, en AAC (AAC = *Advanced Audio Coding*, Codificación de Audio Avanzada), véase *Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding*, Norma Internacional 13818-7, ISO/IEC JTC1/SC29/WG11 Grupo de Expertos en Imágenes en Movimiento, 1997.

Asimismo, las transiciones de desvanecimiento cruzado sin distorsión por repliegue del espectro se describen en Fielder, Louis D., Todd, Craig C., "The Design of a Video Friendly Audio Coding System for Distribution Applications", Escrito número 17-008, 17ª Conferencia Internacional de AES: *High-Quality Audio Coding* (agosto de 1999) y en Fielder, Louis D., Davidson, Grant A., "Audio Coding Tools for Digital Television Distribution", Número de preimpresión 5104, 108ª Convención de la AES (enero de 2000).

El documento WO 2008/071353 divulga un concepto para conmutar entre un codificador en el dominio del tiempo y en el dominio de la frecuencia. El concepto se podría aplicar a cualquier códec basado en una conmutación en el dominio del tiempo / dominio de la frecuencia. Por ejemplo, el concepto se podría aplicar a la codificación en el dominio del tiempo de acuerdo con el modo de ACELP del códec de AMR-WB+ y la AAC como un ejemplo de un códec en el dominio de la frecuencia. La figura 22 muestra un diagrama de bloques de un codificador convencional usando un descodificador en el dominio de la frecuencia en la rama superior y un descodificador en el dominio del tiempo en la rama inferior. La parte de descodificación de frecuencia se ejemplifica mediante un descodificador de AAC, que comprende un bloque de recuantificación 2202 y un bloque de transformada de coseno discreta modificada inversa 2204. En la AAC, la transformada de coseno discreta modificada (MDCT = *Modified Discrete Cosine Transform*) se usa como una transformación entre el dominio del tiempo y el dominio de la frecuencia. En la figura 22 la trayectoria de descodificación en el dominio del tiempo se ejemplifica como un descodificador de AMR-WB+ 2206 seguido por un bloque de MDCT 2208, con el fin de combinar el resultado del descodificador 2206 con el resultado del recuantificador 2202 en el dominio de la frecuencia.

Esto posibilita una combinación en el dominio de la frecuencia, al tiempo que una etapa de superposición y agregación, que no se muestra en la figura 22, se puede usar después de la MDCT inversa 2204, con el fin de combinar y de aplicar un desvanecimiento cruzado a bloques adyacentes, sin tener que considerar si se han codificado en el dominio del tiempo o el dominio de la frecuencia.

En otro enfoque convencional que se divulga en el documento WO2008/071353 es evitar la MDCT 2208 en la figura 22, es decir, DCT-IV e IDCT-IV para el caso de la descodificación en el dominio del tiempo, se puede usar otro enfoque para la así denominada cancelación de distorsión por repliegue del espectro en el dominio del tiempo (TDAC = *Time-Domain Aliasing Cancellation*). Esto se muestra en la figura 23. La figura 23 muestra otro descodificador que tiene el descodificador en el dominio de la frecuencia ejemplificado como un descodificador de AAC que comprende un bloque de recuantificación 2302 y un bloque de IMDCT 2304. La trayectoria en el dominio del tiempo se ejemplifica de nuevo mediante un descodificador de AMR-WB+ 2306 y el bloque TDAC 2308. El descodificador que se muestra en la figura 23 permite una combinación de los bloques descodificados en el dominio del tiempo, es decir, después de la IMDCT 2304, dado que la TDAC 2308 introduce la distorsión por repliegue del espectro de tiempo necesaria para una combinación apropiada, es decir, para una cancelación de distorsión por repliegue del espectro de tiempo, directamente en el dominio del tiempo. Para ahorrar algo de cálculo, y en lugar de usar MDCT en cada primera y última supertrama, es decir, en cada 1024 muestras, de cada segmento de AMR-WB+, se puede usar TDAC solo en las regiones o zonas de superposición en 128 muestras. Se puede mantener la distorsión por repliegue del espectro en el dominio del tiempo normal introducida por el procesamiento de AAC, al tiempo que se introduce la distorsión por repliegue del espectro en el dominio del tiempo inverso correspondiente en las partes de AMR-WB+.

Las ventanas de desvanecimiento cruzado sin distorsión por repliegue del espectro tienen la desventaja de que su codificación no es eficiente, debido a que generan unos coeficientes codificados no muestreados críticamente, y agregan una tara de la información que codificar. La introducción de TDA (TDA = *Time-Domain Aliasing*, distorsión por repliegue del espectro en el dominio del tiempo) en el descodificador en el dominio del tiempo, como por ejemplo en el documento WO 2008/071353, reduce esta tara, pero se podría aplicar solo a medida que las alineaciones temporales de trama de los dos codificadores coinciden las unas con las otras. De otro modo, la eficiencia de codificación se reduce de nuevo. Además, el TDA en el lado de descodificador podría ser problemático, en especial en el punto de partida de un codificador en el dominio del tiempo. Después de un restablecimiento potencial, un codificador o descodificador en el dominio del tiempo producirá, por lo general, una ráfaga de ruido de cuantificación debido a la vacuidad de las memorias del codificador o descodificador en el dominio del tiempo usando, por ejemplo, LPC (LPC = codificación predictiva lineal). Llevará entonces un cierto tiempo al descodificador antes de encontrarse en un estado permanente o estable y proporcionar un ruido de cuantificación más uniforme con el tiempo. Este error de ráfaga no es ventajoso, ya que por lo general es audible.

Por lo tanto, el objeto de la presente invención es la provisión de un concepto mejorado para conmutar en la codificación de audio en dominios múltiples.

5 Este objeto se logra mediante un codificador de acuerdo con la reivindicación 1, un método de codificación de audio de acuerdo con la reivindicación 3, un descodificador de audio de acuerdo con la reivindicación 4, un método de descodificación de audio de acuerdo con la reivindicación 5 y un programa informático de acuerdo con la reivindicación 6.

10 Es un hallazgo de la presente invención que se puede lograr una conmutación mejorada en un concepto de codificación de audio que usa una codificación en el dominio del tiempo y en el dominio de la frecuencia, cuando se adapta el entramado de los dominios de codificación correspondientes o se usan ventanas de desvanecimiento cruzado modificadas. En una realización, por ejemplo, se puede usar AMR-WB+ como códec en el dominio del tiempo y se puede usar AAC como un ejemplo de códec en el dominio de la frecuencia, se puede lograr una conmutación más eficiente entre los dos códecs mediante realizaciones, o bien mediante la adaptación del entramado de la parte de AMR-WB+ o bien mediante el uso de ventanas de inicio o de detención modificadas para la parte de codificación de AAC respectiva.

20 Es un hallazgo adicional de la invención que se puede aplicar TDAC en el descodificador y se pueden usar ventanas de desvanecimiento cruzado sin distorsión por repliegue del espectro.

Unas realizaciones de la presente invención pueden proporcionar la ventaja de que se puede reducir la información de tara, introducirse en una transición de superposición, al tiempo que se mantienen unas regiones de desvanecimiento cruzado moderadas lo cual asegura la calidad del desvanecimiento cruzado. Unas realizaciones de la presente invención se detallarán usando las figuras adjuntas, en las cuales

- 25 la figura 1a muestra una realización de un codificador de audio;
 la figura 1b muestra una realización de un descodificador de audio;
 las figuras 2a - 2j muestran ecuaciones para la MDCT / IMDCT;
 la figura 3 muestra una realización que usa un entramado modificado;
 30 la figura 4a muestra una señal cuasi periódica en el dominio del tiempo;
 la figura 4b muestra una señal vocal en el dominio de la frecuencia;
 la figura 5a muestra una señal de tipo ruido en un dominio del tiempo;
 la figura 5b muestra una señal no vocal en el dominio de la frecuencia;
 la figura 6 muestra un CELP de análisis por síntesis;
 35 la figura 7 ilustra un ejemplo de una fase de análisis de LPC en una realización;
 la figura 8a muestra una realización con una ventana de detención modificada;
 la figura 8b muestra una realización con una ventana de inicio - detención modificada;
 la figura 9 muestra una ventana de principio;
 la figura 10 muestra una ventana más avanzada;
 40 la figura 11 muestra un ejemplo de una ventana de detención modificada;
 la figura 12 ilustra una realización con diferentes zonas o regiones de superposición;
 la figura 13 ilustra una realización de una ventana de inicio modificada;
 la figura 14 muestra una realización de una ventana de detención modificada libre de distorsión por repliegue del espectro aplicada en un codificador;
 45 la figura 15 muestra una ventana de detención modificada libre de distorsión por repliegue del espectro aplicada en el descodificador;
 la figura 16 ilustra ejemplos de codificador y de descodificador convencionales;
 las figuras 17a, 17b ilustran LPC para señales vocales y no vocales;
 la figura 18 ilustra una ventana de desvanecimiento cruzado de la técnica anterior;
 50 la figura 19 ilustra una secuencia de la técnica anterior de ventanas de AMR-WB+;
 la figura 20 ilustra unas ventanas usadas para transmitir en AMR-WB+ entre ACELP y TCX;
 la figura 21 muestra una secuencia a modo de ejemplo de tramas de audio consecutivas en diferentes dominios de codificación;
 la figura 22 ilustra el enfoque convencional para la descodificación de audio en diferentes dominios; y
 55 la figura 23 ilustra un ejemplo de cancelación de distorsión por repliegue del espectro en el dominio del tiempo.

La figura 1a muestra un codificador de audio 100 para codificar muestras de audio. El codificador de audio 100 comprende un primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 para codificar muestras de audio en un primer dominio de codificación, teniendo el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 una primera regla de entramado, una ventana de inicio y una ventana de detención. Asimismo, el codificador de audio 100 comprende un segundo codificador 120 para codificar muestras de audio en el segundo dominio de codificación. El segundo codificador 120 tiene un número de muestras de audio de tamaño de trama predeterminado y un número de muestras de audio de periodo de puesta a punto de codificación. El periodo de puesta a punto de codificación puede ser uno cierto o predeterminado, este puede ser dependiente de las muestras de audio, una trama de muestras de audio o una

secuencia de señales de audio. El segundo codificador 120 tiene una segunda regla de entramado diferente. Una trama del segundo codificador 120 es una representación codificada de un número de muestras de audio oportunamente posteriores, siendo el número igual al número de muestras de audio de tamaño de trama predeterminado.

5 El codificador de audio 100 comprende adicionalmente un controlador 130 para conmutar del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 al segundo codificador 120 en respuesta a una característica de las muestras de audio y para modificar la segunda regla de entramado en respuesta a una conmutación del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 al segundo codificador 120 o para modificar la ventana de inicio o la ventana de detención del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110, en donde la segunda regla de entramado permanece sin modificaciones.

15 En unas realizaciones, el controlador 130 se puede adaptar para determinar la característica de las muestras de audio sobre la base de las muestras de audio de entrada o sobre la base de la salida del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 o el segundo codificador 120. Esto se indica mediante la línea de puntos en la figura 1a, a través de lo cual las muestras de audio de entrada se pueden proporcionar al controlador 130. Se proporcionarán a continuación detalles adicionales acerca de la decisión de conmutación.

20 En unas realizaciones, el controlador 130 puede controlar el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 y el segundo codificador 120 de una manera tal que ambos codifican las muestras de audio en paralelo, y el controlador 130 toma la decisión acerca de la conmutación sobre la base del resultado respectivo, y lleva a cabo las modificaciones antes de la conmutación. En otras realizaciones, el controlador 130 puede analizar las características de las muestras de audio y decidir qué rama de codificación usar, pero desconectando la otra rama. En una realización de este tipo, el periodo de puesta a punto de codificación del segundo codificador 120 se vuelve relevante, ya que antes de la conmutación, se ha de tener en cuenta el periodo de puesta a punto de codificación, lo cual se detallará a continuación.

30 En unas realizaciones, el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 puede comprender un transformador en el dominio de la frecuencia para transformar la primera trama de muestras de audio posteriores al dominio de la frecuencia. El primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 se puede adaptar para ponderar la primera trama codificada con la ventana de inicio, cuando la trama posterior es codificada por el segundo codificador 120 y se puede adaptar adicionalmente para ponderar la primera trama codificada con la ventana de detención cuando una trama precedente va a ser codificada por el segundo codificador 120.

40 Se ha de hacer notar que se pueden usar diferentes notaciones, el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 aplica una ventana de inicio o una ventana de detención. En el presente caso, y para el resto se supone que una ventana de inicio se aplica antes de la conmutación al segundo codificador 120 y cuando se conmuta de vuelta del segundo codificador 120 al primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 120, la ventana de detención se aplica en el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110. Sin pérdida de generalidad, la expresión se podría usar de forma recíproca con referencia al segundo codificador 120. Con el fin de evitar confusión, en el presente caso las expresiones "inicio" y "detención" se refieren a ventanas aplicadas en el primer codificador 110, cuando se inicia el segundo codificador 120 o después de que se detuviera.

50 En unas realizaciones, el transformador en el dominio de la frecuencia como se usa en el primer codificador de distorsión por repliegue del espectro en el dominio del tiempo 110 se puede adaptar para transformar la primera trama al dominio de la frecuencia sobre la base de una MDCT y el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 se puede adaptar para adaptar un tamaño de MDCT a las ventanas de inicio y de detención o de inicio y de detención modificadas. Los detalles para la MDCT y su tamaño se expondrán a continuación.

55 En unas realizaciones, el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 se puede adaptar, en consecuencia, para usar una ventana de inicio y / o de detención que tiene una parte libre de distorsión por repliegue del espectro, es decir, dentro de la ventana hay una parte sin distorsión por repliegue del espectro en el dominio del tiempo. Asimismo, el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 se puede adaptar para usar una ventana de inicio y / o una ventana de detención que tiene una parte libre de distorsión por repliegue del espectro en una parte de flanco de subida de la ventana, cuando la trama precedente es codificada por el segundo codificador 120, es decir, el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 usa una ventana de detención, que tiene una parte de flanco de subida que está libre de distorsión por repliegue del espectro. En consecuencia, se puede adaptar el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 para usar una ventana que tiene una parte de flanco de bajada que está libre de distorsión por repliegue del espectro, cuando una trama posterior es codificada por el segundo codificador 120, es decir,

usando una ventana de detención con una parte de flanco de bajada, la cual está libre de distorsión por repliegue del espectro.

5 En unas realizaciones, el controlador 130 se puede adaptar para iniciar el segundo codificador 120 de manera que una primera trama de una secuencia de tramas del segundo codificador 120 comprende una representación
 10 codificada de las muestras procesadas en la parte libre de distorsión por repliegue del espectro precedente del primer codificador de distorsión por repliegue del espectro en el dominio del tiempo 110. Dicho de otra forma, la salida del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 y el segundo codificador 120 pueden estar coordinadas por el controlador 130 de manera que la parte libre de
 15 distorsión por repliegue del espectro de las muestras de audio codificadas a partir del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 se superpone con las muestras de audio codificadas emitidas por el segundo codificador 120. El controlador 130 se puede adaptar adicionalmente para el desvanecimiento cruzado, es decir, el desvanecimiento gradual de un codificador al tiempo que aparece
 20 gradualmente el otro codificador.

El controlador 130 se puede adaptar para iniciar el segundo codificador 120 de manera que el número de muestras del periodo de puesta a punto de codificación se superpone con la parte libre de distorsión por repliegue del espectro de la ventana de inicio del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 y una trama posterior del segundo codificador 120 se superpone con la parte de distorsión por
 25 repliegue del espectro de la ventana de detención. Dicho de otra forma, el controlador 130 puede coordinar el segundo codificador 120 de manera que, para el periodo de puesta a punto de codificación, estén disponibles unas muestras de audio sin distorsión por repliegue del espectro a partir del primer codificador 110 y, cuando solo están disponibles unas muestras de audio de distorsión por repliegue del espectro a partir del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110, el periodo de puesta a punto del
 30 segundo codificador 120 ha terminado y están disponibles unas muestras de audio codificadas en la salida del segundo codificador 120 de una manera regular.

El controlador 130 se puede adaptar adicionalmente para iniciar el segundo codificador 120 de manera que el periodo de puesta a punto de codificación se superpone con la parte de distorsión por repliegue del espectro de la ventana de inicio. En la presente realización, durante la parte de superposición, las muestras de audio de distorsión por repliegue del espectro están disponibles a partir de la salida del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110, y en la salida del segundo codificador 120 pueden estar
 35 disponibles unas muestras de audio codificadas del periodo de puesta a punto, las cuales pueden experimentar un ruido de cuantificación aumentado. El controlador 130 aún se puede adaptar para un desvanecimiento cruzado entre las dos secuencias de audio codificadas de manera subóptima durante un periodo de superposición.

En otras realizaciones, el controlador 130 se puede adaptar adicionalmente para la conmutación a partir del primer codificador 110 en respuesta a una característica diferente de las muestras de audio y para modificar la segunda regla de entramado en respuesta a la conmutación del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 al segundo codificador 120 o para modificar la ventana de inicio o la
 40 ventana de detención del primer codificador, en donde la segunda regla de entramado permanece sin modificaciones. Dicho de otra forma, el controlador 130 se puede adaptar para conmutar hacia adelante y hacia atrás entre los dos codificadores de audio.

En otras realizaciones, el controlador 130 se puede adaptar para iniciar el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 de manera que la parte libre de distorsión por repliegue del espectro de la ventana de detención se superpone con la trama del segundo codificador 120. Dicho de otra forma, en unas realizaciones, el controlador se puede adaptar para el desvanecimiento cruzado entre las salidas de los dos codificadores. En algunas realizaciones, la salida del segundo codificador se desvanece gradualmente, al
 45 tiempo que aparecen gradualmente solo las muestras de audio codificadas de manera subóptima, es decir, con distorsión por repliegue del espectro del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110. En otras realizaciones, el controlador 130 se puede adaptar para el desvanecimiento cruzado entre una trama del segundo codificador 120 y tramas sin distorsión por repliegue del espectro del primer
 50 codificador 110.

En unas realizaciones, el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 puede comprender un codificador de AAC de acuerdo con *Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding*, Norma Internacional 13818-7, ISO/IEC JTC1/SC29/WG11 Grupo de
 55 Expertos en Imágenes en Movimiento, 1997.

En unas realizaciones, el segundo codificador 120 puede comprender un codificador de AMR-WB+ de acuerdo con 3GPP (3GPP = Proyecto de Asociación de Tercera Generación), Especificación Técnica 26.290, Versión 6.3.0 a partir de junio de 2005 "*Audio Codec Processing Function; Extended Adaptive Multi-Rate-Wide Band Codec; Transcoding Functions*", edición 6.

65 El controlador 130 se puede adaptar para modificar la regla de entramado de AMR o de AMR-WB+ de manera que

una primera supertrama de AMR comprende cinco tramas de AMR, en donde de acuerdo con la especificación técnica anteriormente mencionada, una supertrama comprende cuatro tramas de AMR regulares, compárense la figura 4, tabla 10 en la página 18 y la figura 5 en la página 20 de la Especificación Técnica anteriormente mencionada. Como se detallará adicionalmente más adelante, el controlador 130 se puede adaptar para agregar una trama extra a una supertrama de AMR. Se ha de hacer notar que, en unas realizaciones, la supertrama se puede modificar mediante una trama adjunta al comienzo o al final de cualquier supertrama, es decir, las reglas de entramado también se pueden hacer coincidir al final de una supertrama.

La figura 1b muestra una realización de un descodificador de audio 150 para descodificar tramas codificadas de muestras de audio. El descodificador de audio 150 comprende un primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 para descodificar muestras de audio en un primer dominio de descodificación. El primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 tiene una primera regla de entramado, una ventana de inicio y una ventana de detención. El descodificador de audio 150, comprende adicionalmente un segundo descodificador 170 para descodificar muestras de audio en un segundo dominio de descodificación. El segundo descodificador 170 tiene un número de muestras de audio de tamaño de trama predeterminado y un número de muestras de audio de periodo de puesta a punto de codificación. Asimismo, el segundo descodificador 170 tiene una segunda regla de entramado diferente. Una trama del segundo descodificador 170 se puede corresponder con una representación descodificada de un número de muestras de audio oportunamente posteriores, en donde el número es igual al número de muestras de audio de tamaño de trama predeterminado.

El descodificador de audio 150 comprende adicionalmente un controlador 180 para conmutar del primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 al segundo descodificador 170 sobre la base de una indicación en la trama codificada de muestras de audio, en donde se adapta el controlador 180 para modificar la segunda regla de entramado en respuesta a la conmutación del primer descodificador de introducción en el dominio del tiempo 160 al segundo descodificador 170 o para modificar la ventana de inicio o la ventana de detención del primer descodificador 160, en donde la segunda regla de entramado permanece sin modificaciones.

De acuerdo con la descripción anterior, por ejemplo, en el codificador y el descodificador de AAC, las ventanas de inicio y de detención se aplican en el codificador así como en el descodificador. De acuerdo con la descripción anterior del codificador de audio 100, el descodificador de audio 150 proporciona las componentes de descodificación correspondientes. La indicación de conmutación para el controlador 180 se puede proporcionar en términos de un bit, una bandera o cualquier información complementaria junto con las tramas codificadas.

En ciertas realizaciones, el primer descodificador 160 puede comprender un transformador en el dominio del tiempo para la transformación de una primera trama de muestras de audio descodificadas al dominio del tiempo. Se puede adaptar el primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 para ponderar la primera trama descodificada con la ventana de inicio cuando una trama posterior es descodificada mediante el segundo descodificador 170 y / o para ponderar la primera trama descodificada con la ventana de detención cuando una trama precedente se ha de descodificar mediante el segundo descodificador 170. El transformador en el dominio del tiempo se puede adaptar para transformar la primera trama al dominio del tiempo sobre la base de una MDCT inversa ($IMDCT = inverse\ MDCT$, MDCT inversa) y / o el primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 se puede adaptar para adaptar un tamaño de IMDCT a las ventanas de inicio y / o de detención o de inicio y / o de detención modificadas. Los tamaños de IMDCT se detallarán más adelante.

En unas realizaciones, el primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 se puede adaptar para usar una ventana de inicio y / o una ventana de detención que tienen una parte libre de distorsión por repliegue del espectro o libre de distorsión por repliegue del espectro. El primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 se puede adaptar adicionalmente para usar una ventana de detención que tiene una parte libre de distorsión por repliegue del espectro en una parte de subida de la ventana cuando la trama precedente ha sido descodificada por el segundo descodificador 170 y / o el primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 puede tener una ventana de inicio que tiene una parte libre de distorsión por repliegue del espectro en el flanco de bajada cuando la trama posterior es descodificada por el segundo descodificador 170.

En correspondencia con las realizaciones anteriormente descritas del codificador de audio 100, se puede adaptar el controlador 180 para iniciar el segundo descodificador 170 de tal modo que la primera trama de una secuencia de tramas del segundo descodificador 170 comprende una representación descodificada de una muestra procesada en la parte libre de distorsión por repliegue del espectro precedente del primer descodificador 160. El controlador 180 se puede adaptar para iniciar el segundo descodificador 170 de tal modo que el número de muestras de audio de periodo de puesta a punto de codificación se superpone con la parte libre de distorsión por repliegue del espectro de la ventana de inicio del primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 y una trama posterior del segundo descodificador 170 se superpone con la parte de distorsión por repliegue del espectro de la ventana de detención.

En otras realizaciones, el controlador 180 se puede adaptar para iniciar el segundo descodificador 170 de tal modo que el período de puesta a punto de codificación se superpone con la parte de distorsión por repliegue del espectro de la ventana de inicio.

5 En otras realizaciones, el controlador 180 se puede adaptar adicionalmente para conmutar del segundo descodificador 170 al primer descodificador 160 en respuesta a una indicación de las muestras de audio codificadas y para modificar la segunda regla de entramado en respuesta a la conmutación del segundo descodificador 170 al primer descodificador 160 o para modificar la ventana de inicio o la ventana de detención del primer descodificador 160, en donde la segunda regla de entramado permanece sin modificaciones. Se puede proporcionar la indicación
10 en términos de una bandera, un bit o cualquier información complementaria junto con las tramas codificadas.

En ciertas realizaciones, el controlador 180 se puede adaptar para iniciar el primer descodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 160 de tal modo que la parte de distorsión por repliegue del espectro de la ventana de detención se superpone con una trama del segundo descodificador 170.
15

El controlador 180 se puede adaptar para aplicar un desvanecimiento cruzado entre tramas consecutivas de muestras de audio descodificadas de los diferentes descodificadores. Asimismo, el controlador 180 se puede adaptar para determinar una distorsión por repliegue del espectro en una parte de distorsión por repliegue del espectro de la ventana de inicio o la ventana de detención a partir de una trama descodificada del segundo descodificador 170 y el controlador 180 se puede adaptar para reducir la distorsión por repliegue del espectro en la parte de distorsión por repliegue del espectro sobre la base de la distorsión por repliegue del espectro determinada.
20

En ciertas realizaciones, el controlador 180 se puede adaptar adicionalmente para descartar el período de puesta a punto de codificación de las muestras de audio a partir del segundo descodificador 170.
25

En lo sucesivo, se describirán la transformada de coseno discreta modificada (MDCT = Transformada de Coseno Discreta Modificada) y la IMDCT. La MDCT se explicará con más detalle con la ayuda de las ecuaciones ilustradas en las figuras 2a - 2j. La transformada de coseno discreta modificada es una transformada relacionada con Fourier basada en la transformada de coseno discreta de tipo IV (DCT-IV = Transformada de Coseno Discreta de Tipo IV), con la propiedad adicional de ser solapada, es decir, está diseñada para llevarse a cabo en bloques consecutivos de un conjunto de datos más grande, en donde bloques posteriores se solapan de manera que, por ejemplo, la última mitad de un bloque coincide con la primera mitad del bloque siguiente. Esta superposición, además de las cualidades de compactación de energía de la DCT, hace a la MDCT especialmente atractiva para las aplicaciones de compresión de señal, dado que esta ayuda a evitar que salgan artefactos de las fronteras de bloque. Por lo tanto, se emplea una MDCT en MP3 (MP3 = MPEG2 / 4 capa 3), AC-3 (AC-3 = Códec de Audio 3 de Dolby), Ogg Vorbis y AAC (AAC = Codificación de Audio Avanzada) para la compresión de audio, por ejemplo.
30
35

La MDCT fue propuesta por Princen, Johnson y Bradley en 1987, con posterioridad al trabajo anterior (1986) de Princen y Bradley para desarrollar el principio subyacente de la MDCT de la cancelación de distorsión por repliegue del espectro en el dominio del tiempo (TDAC), que se describe a continuación. También existe una transformada análoga, la MDST (MDST = *Modified DST*, DST Modificada, DST = *Discrete Sine Transform*, Transformada de Seno Discreta), basada en la transformada de seno discreta, así como otras formas, raramente usadas, de la MDCT basadas en tipos diferentes de combinaciones de DCT o de combinaciones de DCT / DST, que también se pueden usar en unas realizaciones por la transformada de introducción de distorsión por repliegue del espectro en el dominio del tiempo.
40
45

En MP3, la MDCT no se aplica a la señal de audio directamente, sino a una salida de un banco de filtros de cuadratura de polifase de 32 bandas (PQF = Filtro de Cuadratura de Polifase). La salida de esta MDCT se postprocesa mediante una fórmula de reducción de distorsión por repliegue del espectro para reducir la distorsión por repliegue del espectro típica del banco de filtros de PQF. Una combinación de este tipo de un banco de filtros con una MDCT se denomina banco de filtros híbrido o MDCT de sub-banda. AAC, por otro lado, usa normalmente una MDCT pura; solo la variante de MPEG-4 AAC-SSR (raramente usada) (por Sony) usa un banco de PQF de cuatro bandas seguido por una MDCT. ATRAC (ATRAC = *Adaptive TTransform Audio Coding*, Codificación de Audio de Transformada Adaptativa) usa filtros de espejo en cuadratura (QMF, *Quadrature Mirror Filter*) apilados seguido por una MDCT.
50
55

Como una transformada solapada, la MDCT es un poco inusual en comparación con las otras transformadas relacionadas con Fourier en que tiene la mitad de salidas que entradas (en lugar del mismo número). En particular, es una función lineal $F: \mathbf{R}^{2N} \rightarrow \mathbf{R}^N$, en donde \mathbf{R} denota el conjunto de los números reales. Los $2N$ números reales x_0, \dots, x_{2N-1} se transforman en los N números reales X_0, \dots, X_{N-1} de acuerdo con la fórmula en la figura 2a.
60

El coeficiente de normalización delante de esta transformada, en el presente caso la unidad, es una convención arbitraria y difiere entre tratamientos. Solo está restringido el producto de las normalizaciones de la MDCT y la IMDCT, a continuación.
65

La MDCT inversa se conoce como IMDCT. Dado que hay diferentes números de entradas y de salidas, en principio

puede parecer que la MDCT no debería ser invertible. Sin embargo, se alcanza una invertibilidad perfecta mediante la agregación de las IMDCT superpuestas de bloques superpuestos posteriores, lo cual causa que los errores se cancelen y que se recuperen los datos originales; esta técnica se conoce como cancelación de distorsión por repliegue del espectro en el dominio del tiempo (TDAC, *time-domain aliasing cancellation*).

5 La IMDCT transforma N números reales X_0, \dots, X_{N-1} en $2N$ números reales y_0, \dots, y_{2N-1} de acuerdo con la fórmula en la figura 2b. Al igual que para la DCT-IV, una transformada ortogonal, la inversa tiene la misma forma que la transformada directa.

10 En el caso de una MDCT a la que se ha aplicado una función ventana con la normalización de ventana usual (véase a continuación), el coeficiente de normalización delante de la IMDCT se ha de multiplicar por 2, es decir, se torna $2/N$.

15 A pesar de la aplicación directa de la fórmula de la MDCT requeriría operaciones de $O(N^2)$, es posible computar la misma cosa con una complejidad de solo $O(N \log N)$ mediante la factorización recursiva de la computación, como en la transformada rápida de Fourier (FFT, *Fast Fourier Transform*). También se pueden computar las MDCT a través de otras transformadas, por lo general una DFT (FFT) o una DCT, combinada con etapas de $O(N)$ de procesamiento previas y posteriores. Asimismo, como se describe a continuación, cualquier algoritmo para la DCT-IV proporciona inmediatamente un método para computar la MDCT y la IMDCT de tamaño par.

20 En aplicaciones de compresión de señal típicas, las propiedades de transformada se mejoran adicionalmente mediante el uso de una función ventana w_n ($n = 0, \dots, 2N-1$) que se multiplica por x_n e y_n en las fórmulas de MDCT y de IMDCT en lo que antecede con el fin de evitar discontinuidades en las fronteras $n = 0$ y $2N$ al hacer que la función vaya suavemente de cero a aquellos puntos. Es decir, se aplica una función ventana a la información antes de la MDCT y después de la IMDCT. En principio, x e y podrían tener diferentes funciones de ventana, y la función ventana también podría cambiar de un bloque al próximo, en especial para el caso en donde se combinan unos bloques de datos de diferentes tamaños, si bien por simplicidad se considera en primer lugar el caso común de las funciones de ventana idénticas para bloques de igual tamaño.

30 La transformada permanece invertible, es decir, la TDAC funciona para una ventana simétrica $w_n = w_{2N-1-n}$, siempre que w cumpla la condición de Princen - Bradley de acuerdo con la figura 2c.

35 Varias funciones de ventana diferentes son comunes, se da un ejemplo en la figura 2d para MP3 y MPEG-2 AAC, y en la figura 2e para Vorbis. AC-3 usa una ventana derivada de Kaiser - Bessel (KBD = *Kaiser - Bessel Derived*), y MPEG-4 AAC también puede usar una ventana de KBD.

40 Nótese que las ventanas aplicadas a la MDCT son diferentes de las ventanas usadas para otros tipos de análisis de señal, dado que estas han de cumplir la condición de Princen - Bradley. Una de las razones para esta diferencia es que las ventanas de MDCT se aplican dos veces, para la MDCT (filtro de análisis) y la IMDCT (filtro de síntesis).

45 Como se puede ver mediante la inspección de las definiciones, para N par la MDCT es esencialmente equivalente a una DCT-IV, en donde la entrada se desplaza $N/2$ y dos bloques N de datos se transforman al mismo tiempo. Mediante el examen de esta equivalencia de forma más cuidadosa, se pueden obtener fácilmente propiedades importantes como TDAC.

50 Con el fin de definir la relación precisa para la DCT-IV, se ha de observar que la DCT-IV se corresponde con alternar condiciones de frontera pares / impares, es par en su frontera izquierda (en torno a $n = -1/2$), impar en su frontera derecha (en torno a $n = N - 1/2$), y así sucesivamente (en lugar de límites periódicos como para una DFT). Esto se deduce de las identidades dadas en la figura 2f. Por lo tanto, si sus entradas con una serie x de longitud N , imagínese extender esta serie a $(x, -x_R, -x, x_R, \dots)$ y así sucesivamente se puede imaginar, en donde x_R denota x en un orden inverso.

55 Considérese una MDCT con $2N$ entradas y N salidas, en donde las entradas pueden dividirse en cuatro bloques (a, b, c, d) cada uno de tamaño $N/2$. Si estas se desplazan $N/2$ (desde el término $+N/2$ en la definición de MDCT), entonces (b, c, d) se extienden más allá del extremo de las N entradas de DCT-IV, de tal modo que estas se han de "plegar" de nuevo de acuerdo con las condiciones de frontera anteriormente descritas.

60 Por lo tanto, la MDCT de $2N$ entradas (a, b, c, d) es exactamente equivalente a una DCT-IV de las N entradas: $(-c_R - d, a - b_R)$, en donde R denota inversión como antes. De esta manera, cualquier algoritmo para computar la DCT-IV se puede aplicar trivialmente a la MDCT.

65 De forma similar, la fórmula de IMDCT como se mencionó anteriormente, es precisamente $1/2$ de la DCT-IV (que es su propia inversa), en donde la salida se desplaza $N/2$ y se extiende (por medio de las condiciones de frontera) a una longitud de $2N$. La DCT-IV inversa simplemente devolvería las entradas $(-c_R - d, a - b_R)$ a partir de lo anterior. Cuando esto se desplaza y se extiende por medio de las condiciones de frontera, se obtiene el resultado que se muestra en la figura 2g. La mitad de las salidas de IMDCT son por lo tanto redundantes.

Se puede entender ahora cómo funciona la TDAC. Supóngase que se computa la MDCT del bloque 2N (c, d, e, f) superpuesto al 50 % posterior. La IMDCT entonces dará, de forma análoga a lo anterior: $(c - d_R, d - c_R, e + f_R, e_R + f) / 2$. Cuando esto se agrega con el resultado de IMDCT anterior en la mitad superpuesta, los términos invertidos se cancelan y se obtiene simplemente (c, d), recuperando los datos originales.

Ahora está claro el origen de la expresión “cancelación de distorsión por repliegue del espectro en el dominio del tiempo”. El uso de datos de entrada que se extienden más allá de las fronteras de la DCT-IV lógica da lugar a que los datos estén sujetos a distorsión por repliegue del espectro exactamente de la misma manera que las frecuencias más allá de la frecuencia de Nyquist están sujetas a distorsión por repliegue del espectro a frecuencias más bajas, excepto por que esta distorsión por repliegue del espectro ocurre en el dominio del tiempo en lugar de en el dominio de la frecuencia. De aquí las combinaciones $c - d_R$ y así sucesivamente, que tienen precisamente los signos correctos para que las combinaciones se cancelen cuando se agregan.

Para N impar (que raramente se usa en la práctica), $N / 2$ no es un número entero de manera que la MDCT no es simplemente una permutación de desplazamiento de una DCT-IV. En este caso, el desplazamiento adicional por media muestra significa que la MDCT / IMDCT se torna equivalente a la DCT-III / II, y el análisis es análogo a lo anterior.

En lo que antecede la propiedad de TDAC se probó para la MDCT común lo cual muestra que agregar las IMDCT de bloques posteriores en su mitad de superposición recupera los datos originales. La derivación de esta propiedad inversa para la MDCT a la que se ha aplicado una función ventana solo es ligeramente más complicada.

Recuérdese de lo anterior que, cuando (a, b, c, d) y (c, d, e, f) se someten a MDCT, se someten a IMDCT y se agregan en su mitad superpuesta, se obtiene $(c + d_R, c_R + d) / 2 + (c - d_R, d - c_R) / 2 = (c, d)$, los datos originales.

A continuación, se supone la multiplicación de las entradas de MDCT y las salidas de IMDCT por una función ventana de longitud 2N. Como antes, se supone una función ventana simétrica, que es, por lo tanto, de la forma (w, z, z_R, w_R), en donde w y z son vectores de longitud N / 2 y R denota inversa como antes. Entonces la condición de Princen - Bradley se puede escribir

$$w^2 + z_R^2 = (1, 1, \dots),$$

con las multiplicaciones y sumas realizadas elemento a elemento, o de forma equivalente

$$w_R^2 + z^2 = (1, 1, \dots)$$

invirtiendo w y z.

Por lo tanto, en lugar de someter a MDCT (a, b, c, d), MDCT (w_a, z_b, z_{Rc}, w_{Rd}) se somete a MDCT con todas las multiplicaciones realizadas elemento a elemento. Cuando esto se somete a IMDCT y se multiplica de nuevo (elemento a elemento) por la función ventana, los resultados de la última mitad de N se muestran en la figura 2h.

Nótese que la multiplicación por 1/2 ha dejado de estar presente, debido a que la normalización de IMDCT difiere por un factor de 2 en el caso con la aplicación de una función ventana. De forma similar, la MDCT y la IMDCT a la que se ha aplicado una función ventana de (c, d, e, f) da, en su mitad primera N de acuerdo con la figura 2i. Cuando estas dos mitades se suman una a otra, se obtienen los resultados de la figura 2j, recuperando los datos originales.

En lo sucesivo, se detallará una realización en la cual el controlador 130 del lado de codificador y el controlador 180 del lado de descodificador, respectivamente, modifican la segunda regla de entramado en respuesta a la conmutación del primer dominio de codificación al segundo dominio de codificación. En la realización, se logra una transición suave en un codificador conmutado, es decir, conmutando entre codificación de AMR-WB+ y de AAC. Con el fin de tener una transición suave, se usa una cierta superposición, es decir, un segmento corto de una señal o una cantidad de muestras de audio, a las que se aplican ambos modos de codificación. Dicho de otra forma, en la siguiente descripción, se proporcionará una realización, en donde el primer codificador de distorsión por repliegue del espectro en el dominio del tiempo 110 y el primer descodificador de distorsión por repliegue del espectro en el dominio del tiempo 160 se corresponden con la codificación y la descodificación de AAC. El segundo codificador 120 y el descodificador 170 se corresponden con AMR-WB+ en el modo de ACELP. La realización se corresponde con una opción de los controladores respectivos 130 y 180 en los cuales se modifica el entramado de la AMR-WB+, es decir, la segunda regla de entramado.

La figura 3 muestra una línea de tiempo en la cual se muestra un número de ventanas y tramas. En la figura 3, una ventana regular de AAC 301 es seguida por una ventana de inicio de AAC 302. En la AAC, la ventana de inicio de AAC 302 se usa entre tramas largas y tramas cortas. Con el fin de ilustrar el entramado de AAC heredado, es decir, la primera regla de entramado del primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 y el descodificador 160, en la figura 3 se muestra también una secuencia de ventanas de

AAC cortas 303. La secuencia de ventanas cortas de AAC 303 se termina con una ventana de detención de AAC 304, la cual inicia una secuencia de ventanas largas de AAC. De acuerdo con la descripción que se ha mostrado en lo que antecede, se supone en la presente realización que el segundo codificador 120, el descodificador 170, respectivamente, usan el modo de ACELP de la AMR-WB+. La AMR-WB+ usa tramas de igual tamaño de las cuales se muestra una secuencia 320 en la figura 3. La figura 3 muestra una secuencia de tramas de prefiltro de diferentes tipos de acuerdo con ACELP en AMR-WB+. Con anterioridad a conmutar AAC por ACELP, el controlador 130 o 180 modifica el entramado de la ACELP de tal modo que la primera supertrama 320 está compuesta por cinco tramas en lugar de cuatro. Por lo tanto, los datos de ACE 314 se encuentran disponibles en el descodificador, al tiempo que también se encuentran disponibles los datos de AAC descodificados. Por lo tanto, la primera parte se puede descartar en el descodificador, dado que esta se refiere al período de puesta a punto de codificación del segundo codificador 120, el segundo descodificador 170, respectivamente. En general, en otras realizaciones, la supertrama de AMR-WB+ se puede extender mediante la anexión de tramas también al final de una supertrama.

La figura 3 muestra dos transiciones de modo, es decir, de AAC a AMR-WB+ y de AMR-WB+ a AAC. En una realización, se usan las típicas ventanas de inicio / detención 302 y 304 del códec de AAC y la longitud de trama del códec de AMR-WB+ se incrementa para superponerse con la parte de desvanecimiento de la ventana de inicio / detención del códec de AAC, es decir, se modifica la segunda regla de entramado. De acuerdo con la figura 3, las transiciones de AAC a AMR-WB+, es decir, del primer codificador de introducción de distorsión por repliegue del espectro de tiempo 110 al segundo codificador 120 o el primer descodificador de introducción de distorsión por repliegue del espectro de tiempo 160 al segundo descodificador 170, respectivamente, se maneja al mantener el entramado de AAC y la extensión de la trama en el dominio del tiempo en la transición con el fin de cubrir la superposición. La supertrama de AMR-WB+ en la transición, es decir, la primera supertrama 320 en la figura 3, usa cinco tramas en lugar de cuatro; cubriendo la quinta trama la superposición. Esto introduce una tara de datos, sin embargo, la realización proporciona la ventaja de que se asegura una transición suave entre los modos de AAC y de AMR-WB+.

Tal como se ha mencionado anteriormente, el controlador 130 se puede adaptar para conmutar entre los dos dominios de codificación sobre la base de la característica de las muestras de audio en donde resultan concebibles diferentes análisis o diferentes opciones. Por ejemplo, el controlador 130 puede conmutar el modo de codificación sobre la base de una fracción estacionaria o una fracción transitoria de la señal. Otra opción sería que se llevara a cabo la conmutación sobre la base de si las muestras de audio se corresponden con una señal de habla más vocal o no vocal. Con el fin de proporcionar una realización detallada para determinar las características de las muestras de audio, en lo sucesivo, una realización del controlador 130, la cual conmuta sobre la base de la similitud de voz de la señal.

A modo de ejemplo, se hace referencia a las figuras 4a y 4b, 5a y 5b, respectivamente. Los segmentos de señal o porciones de señal de tipo impulso cuasi-periódico y los segmentos de señal o porciones de señal de tipo ruido se analizan a modo de ejemplo. En general, los controladores 130, 180 se pueden adaptar para decidir sobre la base de criterios diferentes, tales como blancura espectral, transitoriedad, estacionario, etc. En lo sucesivo, se da un criterio a modo de ejemplo como parte de una realización. En concreto, se ilustra un habla vocal en la figura 4a en el dominio del tiempo y en la figura 4b en el dominio de la frecuencia y se analiza como un ejemplo para una porción de señal de tipo impulso cuasi-periódico y un segmento de habla no vocal como un ejemplo de una porción de señal de tipo ruido se analiza en conexión con las figuras 5a y 5b.

En general, el habla se puede clasificar como vocal, no vocal o mixta. El habla vocal es cuasi periódica en el dominio del tiempo y armónicamente estructurada en el dominio de la frecuencia, mientras que el habla no vocal es de tipo aleatorio y de banda ancha. Además, la energía de los segmentos vocales es, en general, mayor que la energía de los segmentos no vocales. El espectro a corto plazo del habla vocal se caracteriza por su estructura fina y formante. La estructura armónica fina es una consecuencia de la cuasi periodicidad del habla y se puede atribuir a las cuerdas vocales vibrantes.

La estructura formante, que también se denomina envolvente espectral, se debe a la interacción de la fuente y los tractos vocales. Los tractos vocales consisten en la faringe y la cavidad bucal. La forma de la envolvente espectral que "encaja" con el espectro a corto plazo del habla vocal está asociada con las características de transferencia del tracto vocal y la inclinación espectral (6 dB / octava) debido al pulso de la glotis.

La envolvente espectral está caracterizada por un conjunto de picos, que se denominan formantes. Las formantes son los modos resonantes del tracto vocal. Para el tracto vocal promedio hay de 3 a 5 formantes por debajo de 5 kHz. Las amplitudes y las ubicaciones de las tres primeras formantes, las cuales ocurren, por lo general, por debajo de 3 kHz, son bastante importantes, ambas, en la percepción y la síntesis del habla. Las formantes más altas son también importantes para las representaciones de habla no vocal y de banda ancha. Las propiedades del habla están relacionadas con los sistemas de producción del habla físicos tal como sigue. La excitación del tracto vocal con pulsos de aire de la glotis cuasi periódicos generados por la vibración de las cuerdas vocales produce el habla vocal. Se hace referencia a la frecuencia de los pulsos periódicos como frecuencia fundamental o tono. Forzar aire a través de una constricción en el tracto vocal produce un habla no vocal. Los sonidos nasales se deben al acoplamiento acústico del tracto nasal con el tracto vocal, y los sonidos oclusivos se reducen mediante la reducción

abrupta de la presión de aire, el cual se acumuló detrás del cierre del tracto.

Por lo tanto, una porción de tipo ruido de la señal de audio puede ser una porción estacionaria en el dominio del tiempo como se ilustra en la figura 5a o una porción estacionaria en el dominio de la frecuencia, la cual es diferente de la porción de tipo impulso cuasi-periódico como se ilustra en el ejemplo en la figura 4a, debido al hecho de que la porción estacionaria en el dominio del tiempo no muestra pulsos de repetición permanente. Como se bosquejará más adelante, sin embargo, la diferenciación entre las porciones de tipo ruido y las porciones de tipo impulso cuasi-periódico también se pueden observar después de una LPC para la señal de excitación. La LPC es un método que modela el tracto vocal y la excitación de los tractos vocales. Cuando se considera el dominio de la frecuencia de la señal, las señales de tipo impulso muestran la apariencia prominente de las formantes individuales, es decir, picos prominentes en la figura 4b, mientras que el espectro estacionario tiene un espectro bastante ancho como se ilustra en la figura 5b o, en el caso de las señales armónicas, un piso de ruido bastante continuo que tiene algunos picos prominentes que representan tonos específicos que ocurren, por ejemplo, en una señal de música, pero que no tienen una distancia tan regular del uno al otro como la señal de tipo impulso en la figura 4b.

Además, las porciones de tipo impulso cuasi-periódico y las porciones de tipo ruido pueden ocurrir de una forma oportuna, es decir, esto significa que una porción de la señal de audio de tiempo es ruidosa y otra porción de la señal de audio en el tiempo es cuasi-periódica, es decir, tonal. Como alternativa o adicionalmente, la característica de una señal puede ser diferente en diferentes bandas de frecuencia. Por lo tanto, la determinación de si la señal de audio es ruidosa o tonal, se puede llevar a cabo de una forma selectiva en frecuencia de manera que se considera que una cierta banda de frecuencia o varias ciertas bandas de frecuencia son ruidosas y se considera que otras bandas de frecuencia son tonales. En este caso, una cierta porción de tiempo de la señal de audio podría incluir componentes tonales y componentes ruidosos.

En lo sucesivo, se analizará un codificador de CELP de análisis por síntesis con respecto a la figura 6. También se pueden encontrar detalles de un codificador de CELP en el documento "Speech Coding: A tutorial review", Andreas Spanias, *Proceedings of IEEE*, Vol. 84, n.º 10, octubre de 1994, páginas 1541 - 1582. El codificador de CELP como se ilustra en la figura 6 incluye un componente de predicción a largo plazo 60 y un componente de predicción a corto plazo 62. Además, se usa un libro de códigos lo cual se indica en 64. Un filtro de ponderación perceptual $W(z)$ se implementa en 66, y un controlador de minimización de error se proporciona en 68. $s(n)$ es la señal de audio de entrada en el dominio del tiempo. Después de haber sido perceptualmente ponderada, la señal ponderada se introduce en un sustractor 69, el cual calcula el error entre la señal de síntesis ponderada en salida del bloque 66 y la señal ponderada real $s_w(n)$.

En general, la predicción a corto plazo $A(z)$ se calcula por una fase de análisis de LPC que se analizará más adelante. Dependiendo de esta información, la predicción a largo plazo $A_L(z)$ incluye la ganancia de predicción a largo plazo b y el retardo T (que también se conoce como ganancia de tono y retardo de tono). El algoritmo de CELP codifica a continuación la señal residual obtenida después de las predicciones a corto y a largo plazo usando un libro de códigos de, por ejemplo, secuencias gaussianas. El algoritmo de ACELP, en donde "A" significa "algebraico" tiene un libro de códigos específico diseñado algebraicamente.

El libro de códigos puede contener más o menos vectores, en donde cada vector tiene una longitud de acuerdo con un número de muestras. Un factor de ganancia g adapta a escala el vector de código y las muestras codificadas multiplicadas por la ganancia son filtrados por el filtro de síntesis a largo plazo y el filtro de síntesis de predicción a corto plazo. El vector de código "óptimo" se selecciona de tal manera que se minimiza el error cuadrático medio perceptualmente ponderado. El proceso de búsqueda en CELP es evidente a partir del esquema de análisis por síntesis ilustrado en la figura 6. Se ha de hacer notar que la figura 6 solo ilustra un ejemplo de un CELP de análisis por síntesis y que las realizaciones no se deben limitar a la estructura que se muestra en la figura 6.

En CELP, el predictor a largo plazo se implementa, a menudo, como un libro de códigos adaptativo que contiene la señal de excitación previa. La ganancia y el retardo de predicción a largo plazo se representan mediante una ganancia y un índice de libro de códigos adaptativo, los cuales se seleccionan también mediante la minimización del error cuadrático medio ponderado. En este caso la señal de excitación consiste en la adición de dos vectores adaptados a escala mediante la ganancia, uno a partir de un libro de códigos adaptativo y uno a partir de un libro de códigos fijo. El filtro de ponderación perceptual en AMR-WB+ está basado en el filtro de LPC, por lo tanto la señal perceptualmente ponderada es una forma de una señal en el dominio de LPC. En el codificador en el dominio de la transformada usado en AMR-WB+, la transformada se aplica a la señal ponderada. En el descodificador, la señal de excitación se puede obtener mediante la filtración de la señal ponderada descodificada a través de un filtro que consiste en la inversa de filtros de síntesis y de ponderación.

La funcionalidad de una realización de la fase de análisis de codificación predictiva 12 se analizará posteriormente de acuerdo con la realización que se muestra en la figura 7, usando análisis de LPC y síntesis de LPC en los controladores 130,180 en las realizaciones correspondientes.

La figura 7 ilustra una implementación más detallada de una realización de un bloque de análisis de LPC. La señal de audio se introduce en un bloque de determinación de filtro, el cual determina la información de filtro $A(z)$, es decir,

la información acerca de coeficientes para el filtro de síntesis. La información es cuantificada y se emite como la información de predicción a corto plazo requerida por el descodificador. En un sustractor 786, se introduce una muestra actual de la señal y un valor predicho para la muestra actual se sustrae de tal manera que, para esta muestra, la señal de error de predicción se genera en la línea 784. Nótese que la señal de error de predicción también se puede denominar señal de excitación o trama de excitación (por lo general, después de codificarse).

La figura 8a muestra otra secuencia de tiempo de ventanas que se logra con otra realización. En la realización considerada en lo sucesivo, el códec de AMR-WB+ se corresponde con el segundo codificador 120 y el códec de AAC se corresponde con el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110. La siguiente realización mantiene el entramado de códec de AMR-WB+, es decir, la segunda regla de entramado permanece sin modificaciones, pero se modifica la aplicación de una función ventana en la transición del códec de AMR-WB+ al códec de AAC, se manipulan las ventanas de inicio / detención del códec de AAC. Dicho de otra forma, la aplicación de una función ventana al códec de AAC será más larga en la transición.

Las figuras 8a y 8b ilustran la presente realización. Ambas figuras muestran una secuencia de ventanas de AAC convencionales 801 en las cuales, en la figura 8a, se introduce una nueva ventana de detención modificada 802 y, en la figura 8b, una nueva ventana de detención / inicio 803. Con respecto a la ACELP, se usa un entramado similar al que se muestra tal como ya se ha descrito con respecto a la realización en la figura 3. En la realización que da como resultado la secuencia de ventana tal como se muestra en las figuras 8a y 8b, se supone que no se mantiene el entramado de códec de AAC normal, es decir, se usan las ventanas de inicio, de detención o de inicio / detención modificadas. La primera ventana que se muestra en las figuras 8a es para la transición de AMR-WB+ a AAC, en donde el códec de AAC usará una ventana de detención larga 802. Se describirá otra ventana con la ayuda de la figura 8b, la cual muestra la transición de AMR-WB+ a AAC cuando el códec de AAC use una ventana corta, usando una ventana de AAC larga para esta transición, según se indica en la figura 8b. La figura 8a muestra que la primera supertrama 820 de la ACELP comprende cuatro tramas, es decir, es acorde al entramado de ACELP convencional, es decir, la segunda regla de entramado. Con el fin de mantener la regla de entramado de ACELP, es decir, la segunda regla de entramado se mantiene sin modificaciones, se usan unas ventanas modificadas 802 y 803 según se indica en las figuras 8a y 8b.

Por lo tanto, en lo sucesivo, se introducirán algunos detalles con respecto a la aplicación de función ventana, en general.

La figura 9 muestra una ventana rectangular general, en la cual la información de la secuencia de ventana puede comprender una primera parte cero, en la cual la ventana enmascara muestras, una segunda parte de derivación, en la cual las muestras de una trama, es decir, una trama en el dominio del tiempo de entrada o una trama en el dominio del tiempo superpuesta puede pasar a través sin modificaciones, y una tercera parte cero, la cual enmascara de nuevo muestras al final de una trama. Dicho de otra forma, se pueden aplicar unas funciones de ventana, lo cual suprime un número de muestras de una trama en una primera parte cero, pasa a través de unas muestras en una segunda parte de derivación, y suprime entonces unas muestras al final de una trama en una tercera parte cero. En este contexto, supresión también se puede referir a la aneación de una secuencia de ceros al comienzo y / o al final de la parte de derivación de la ventana. La segunda parte de derivación puede ser tal que la función ventana simplemente tiene un valor de 1, es decir, las muestras pasan a través sin modificaciones, es decir, la función ventana conmuta a través de las muestras de la trama.

La figura 10 muestra otra realización de una secuencia de ventana o función ventana, en donde la secuencia de ventana comprende adicionalmente una parte de flanco de subida entre la primera parte cero y la segunda parte de derivación y una parte de flanco de bajada entre la segunda parte de derivación y la tercera parte cero. La parte de flanco de subida también se puede considerar como una parte de aparición gradual y la parte de flanco de bajada se puede considerar como una parte de desvanecimiento gradual. En unas realizaciones, la segunda parte de derivación puede comprender una secuencia de unos para no modificar en absoluto las muestras de la trama de excitación.

Volviendo a la realización que se muestra en la figura 8a, la ventana de detención modificada, tal como se usa en un ejemplo que realiza una transición entre la AMR-WB+ y la AAC, cuando se realiza una transición de AMR-WB+ a AAC, se muestra con más detalle en la figura 11. La figura 11 muestra las tramas de ACELP 1101, 1102, 1103 y 1104. La ventana de detención modificada 802 se usa entonces para realizar una transición a AAC, es decir, el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110, el descodificador 160, respectivamente. De acuerdo con los detalles anteriores de la MDCT, la ventana se inicia ya en la mitad de la trama 1102, teniendo una primera parte cero de 512 muestras. A esta parte la sigue la parte de flanco de subida de la ventana, la cual se extiende a través de 128 muestras seguidas por la segunda parte de derivación que, en este ejemplo, se extiende a 576, es decir, 512 muestras después de la parte de flanco de subida en la cual se pliega la primera parte cero, seguidas por 64 muestras más de la segunda parte de derivación, las cuales resultan de la tercera parte cero al final de la ventana extendida a través de 64 muestras. La parte de flanco de bajada de la ventana con la misma da como resultado 1024 muestras, que van a superponerse con la siguiente ventana.

El ejemplo se puede describir también usando un pseudocódigo, que se ejemplifica mediante:

```

/* Conmutación de bloques basada en ataques */
Si (hay un ataque) {
nextwindowSequence = SHORT_WINDOW;
}
5 de lo contrario {
nextwindowSequence = LONG_WINDOW;
}
/* Conmutación de bloque basada en decisión de conmutación de ACELP */
si (la próxima trama es AMR) {
10 nextwindowSequence = SHORT_WINDOW;
}
/* Conmutación de bloque basada en decisión de conmutación de ACELP para STOP_WINDOW_1152 */
si (la trama actual es AMR y la próxima trama no es AMR) {
15 nextwindowSequence = STOP_WINDOW_1152;
}
/*Conmutación de bloque para STOPSTART_WINDOW_1152*/
si (nextwindowSequence == SHORT_WINDOW) {
si (windowSequence == STOP_WINDOW_1152) {
20 windowSequence = STOPSTART_WINDOW_1152;
}
}
}

```

Volviendo al ejemplo que se muestra en la figura 11, existe una sección de plegado de distorsión por repliegue del espectro de tiempo dentro de la parte de flanco de subida de la ventana, la cual se extiende a través de 128 muestras. Dado que esta sección se superpone con la última trama de ACELP 1104, la salida de la trama de ACELP 1104 se puede usar para la cancelación de distorsión por repliegue del espectro de tiempo en la parte de flanco de subida. La cancelación de distorsión por repliegue del espectro se puede llevar a cabo en el dominio del tiempo o en el dominio de la frecuencia, en línea con los ejemplos anteriormente descritos. Dicho de otra forma, la salida de la última trama de ACELP se puede transformar al dominio de la frecuencia y se puede superponer entonces con la parte de flanco de subida de la ventana de detención modificada 802. Como alternativa, se pueden aplicar TDA o TDAC a la última trama de ACELP antes de superponer esta con la parte de flanco de subida de la ventana de detención modificada 802.

El ejemplo anteriormente descrito reduce la tara generada en las transiciones. También elimina la necesidad de que se lleven a cabo modificaciones al entramado de la codificación en el dominio del tiempo, es decir, la segunda regla de entramado. Asimismo, también adapta el codificador en el dominio de la frecuencia, es decir, el codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 (AAC), que en general es más flexible en términos de la atribución de bits y la cantidad de coeficientes a transmitir que un codificador en el dominio del tiempo, es decir, el segundo codificador 120.

En lo sucesivo se describirá otra realización, la cual proporciona un desvanecimiento cruzado libre de distorsión por repliegue del espectro cuando se produce la conmutación entre el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo 110 y el segundo codificador 120, los descodificadores 160 y 170, respectivamente. La presente realización proporciona la ventaja de que se evite el ruido debido a TDAC, en especial a bajas tasas de bits, en el caso de procedimientos de inicio o de reinicio. La ventaja se logra mediante una realización que tiene una ventana de inicio de AAC modificada sin distorsión alguna por repliegue del espectro de tiempo en la parte derecha o en la parte de flanco de bajada de la ventana. La ventana de inicio modificada es una ventana asimétrica, es decir, la parte derecha o la parte de flanco de bajada de la ventana termina antes del punto de plegado de la MDCT. En consecuencia, la ventana está libre de distorsión por repliegue del espectro de tiempo. Al mismo tiempo, la región de superposición se puede reducir mediante unas realizaciones hasta 64 muestras en lugar de 128 muestras.

En unas realizaciones, puede llevar un cierto tiempo al codificador de audio 100 o al descodificador de audio 150 antes de entrar en un estado permanente o estable. Dicho de otra forma, durante el período de inicio del codificador en el dominio del tiempo, es decir, el segundo codificador 120 y también el descodificador 170, se necesita un cierto tiempo con el fin de iniciar, por ejemplo, los coeficientes de una LPC. Con el fin de suavizar el error en el caso de un restablecimiento, en unas realizaciones, se puede aplicar una función ventana a la parte izquierda de una señal de entrada de AMR-WB+ con una ventana de seno corta en el codificador 120, por ejemplo, que tiene una longitud de 64 muestras. Asimismo, se puede aplicar una función ventana a la parte izquierda de la señal de síntesis con la misma señal en el segundo descodificador 170. De este modo, la ventana de seno al cuadrado se puede aplicar de un modo similar a AAC, aplicando el seno al cuadrado a la parte derecha de su ventana de inicio.

Mediante el uso de esta aplicación de una función ventana, en una realización, la transición de AAC a AMR-WB+ se puede llevar a cabo sin distorsión por repliegue del espectro de tiempo y se puede llevar a cabo mediante una ventana de seno de desvanecimiento cruzado corto tal como, por ejemplo, 64 muestras. La figura 12 muestra una línea de tiempo que ejemplifica una transición de AAC a AMR-WB+ y de vuelta a AAC. La figura 12 muestra una

ventana de inicio de AAC 1201 seguida de la parte de AMR-WB+ 1203 que se superpone con la ventana de AAC 1201 y que se superpone con la región 1202, la cual se extiende a través de 64 muestras. La parte de AMR-WB+ es seguida por una ventana de detención de AAC 1205, que se superpone con 128 muestras.

- 5 De acuerdo con la figura 12, la realización aplica la ventana libre de distorsión por repliegue del espectro respectiva en la transición de AAC a AMR-WB+.

La figura 13 muestra la ventana de inicio modificada, tal como se aplica cuando se realiza una transición de AAC a AMR-WB+ en ambos lados en el codificador 100 y el descodificador 150, el codificador 110 y el descodificador 160, respectivamente.

La ventana que se ilustra en la figura 13 muestra que la primera parte cero no se encuentra presente. La ventana se inicia directamente con la parte de flanco de subida, la cual se extiende a través de 1024 muestras, es decir, el eje de plegado se encuentra en la mitad del intervalo de 1024 que se muestra en la figura 13. El eje de simetría se encuentra entonces en el lado derecho del intervalo de 1024. Tal como se puede ver a partir de la figura 13, la tercera parte cero se extiende a 512 muestras, es decir, no hay distorsión alguna por repliegue del espectro en la parte al lado derecho de la totalidad de la ventana, es decir, la parte de derivación se extiende desde el centro hasta el comienzo del intervalo de 64 muestras. También se puede observar que la parte de flanco de bajada se extiende a través de 64 muestras, lo cual proporciona la ventaja de que la sección de cruce sea angosta. El intervalo de 64 muestras se usa para un desvanecimiento cruzado, sin embargo, no se encuentra presente distorsión alguna por repliegue del espectro en este intervalo. Por lo tanto, solo se introduce una tara baja.

Unas realizaciones con las ventanas modificadas anteriormente descritas son capaces de evitar la codificación de demasiada información de tara, es decir, la codificación de algunas de las muestras dos veces. De acuerdo con la descripción que se ha mostrado anteriormente, se pueden aplicar opcionalmente unas ventanas diseñadas de manera similar para la transición de AMR-WB+ a AAC de acuerdo con una realización en la que se modifica de nuevo la ventana de AAC, reduciendo también la superposición a 64 muestras.

Por lo tanto, la ventana de detención modificada se alarga a 2304 muestras en una realización y se usa en una MDCT de 1152 puntos. La parte izquierda de la ventana se puede hacer libre de distorsión por repliegue del espectro de tiempo mediante el comienzo de la aparición gradual después del eje de plegado de MDCT. Dicho de otra forma, al hacer la primera parte cero más grande que un cuarto de la totalidad del tamaño de MDCT. La ventana de seno al cuadrado complementaria se aplica entonces a las 64 últimas muestras descodificadas del segmento de AMR-WB+. Estas dos ventanas de desvanecimiento cruzado permiten obtener una transición suave de AMR-WB+ a AAC mediante la limitación de la información transmitida de la tara.

La figura 14 ilustra una ventana para la transición de AMR-WB+ a AAC tal como se puede aplicar en el lado de codificador 100 en una realización. Se puede ver que el eje de plegado es después de 576 muestras, es decir, la primera parte cero se extiende a través de de 576 muestras. Esto tiene como consecuencia que la parte al lado izquierdo de la totalidad de la ventana se encuentre libre de distorsión por repliegue del espectro. El desvanecimiento cruzado comienza en el segundo cuarto de la ventana, es decir, después de 576 muestras o, dicho de otra forma, justo más allá del eje de plegado. La sección de desvanecimiento cruzado, es decir, la parte de flanco de subida de la ventana, puede estrecharse hasta 64 muestras de acuerdo con la figura 14.

La figura 15 muestra la ventana para la transición de AMR-WB+ a ACC aplicada en el lado de descodificador 150 en una realización. La ventana es similar a la ventana descrita en la figura 14, de tal modo que la aplicación de ambas ventanas a través de muestras que se están codificando y, entonces, descodificando de nuevo da como resultado una ventana de seno al cuadrado.

El siguiente pseudocódigo describe una realización de un procedimiento de selección de ventana de inicio, cuando ocurre la conmutación de AAC a AMR-WB+.

Estás realizaciones también se pueden describir mediante el uso de un pseudocódigo tal como, por ejemplo:

```

55 /* Ajustar a una secuencia de ventana permitida */
   si (nextwindowSequence == SHORT_WINDOW) {
     si (windowSequence == LONG_WINDOW){
       si (la trama actual no es AMR y la próxima trama es AMR) {
60   windowSequence = START_WINDOW_AMR;
       }
     de lo contrario {
       windowSequence = START_WINDOW;
     }
65 }

```

Unas realizaciones tal como se ha descrito anteriormente reducen la tara generada de la información mediante el

5 uso de pequeñas regiones de superposición en ventanas consecutivas durante la transición. Además, estas realizaciones proporcionan la ventaja de que estas pequeñas regiones de superposición siguen siendo suficientes para suavizar los artefactos de bloqueo, es decir, para tener un desvanecimiento cruzado suave. Asimismo, reduce el impacto de la ráfaga de error debido al inicio del codificador en el dominio del tiempo, es decir, el segundo codificador 120, el descodificador 170, respectivamente, mediante la inicialización del mismo con una entrada a la que se ha aplicado desvanecimiento.

10 Resumiendo, unas realizaciones de la presente invención proporcionan la ventaja de que unas regiones de cruce suavizadas se pueden llevar a cabo en un concepto de codificación de audio de modo múltiple con una eficiencia de codificación alta, es decir, las ventanas de transición introducen solo una tara baja en términos de la información adicional a ser transmitida. Además, unas realizaciones posibilitan el uso de codificadores de modo múltiple, al tiempo que se adapta el entramado o la aplicación de función ventana de un modo al otro.

15 A pesar de que algunos aspectos se han descrito en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del método correspondiente, en donde un bloque o dispositivo se corresponde con una etapa de método o con una característica de una etapa de método. De forma análoga, unos aspectos descritos en el contexto de una etapa de método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.

20 La señal de audio codificada se puede almacenar en un medio de almacenamiento digital o se puede transmitir por un medio de transmisión tal como un medio de transmisión inalámbrico o cableado tal como Internet.

25 Dependiendo de ciertos requisitos de implementación, unas realizaciones de la presente invención se pueden implementar en soporte físico o en soporte lógico. La implementación se puede realizar mediante el uso de un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga unas señales de control electrónicamente legibles almacenadas en el mismo, las cuales cooperan (o son capaces de cooperar) con un sistema informático programable de tal modo que se realice el método respectivo.

30 Algunas realizaciones de acuerdo con la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, capaces de cooperar con un sistema informático programable, de tal modo que se realice uno de los métodos descritos en el presente documento.

35 En general, unas realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, estando el código de programa operativo para realizar uno de los métodos cuando el producto del programa de informático se ejecuta en un ordenador. El código del programa se puede almacenar, por ejemplo, en un soporte legible por máquina.

40 Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenados en un soporte legible por máquina.

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

Una realización adicional de los métodos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en el mismo, el programa informático para realizar uno de los métodos descritos en el presente documento.

50 Una realización adicional de los métodos de la invención es, por lo tanto, un tren de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en el presente documento. El tren de datos o la secuencia de señales se pueden configurar, por ejemplo, para ser transferidos por medio de una conexión de comunicación de datos, por ejemplo, por medio de Internet.

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

Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los métodos descritos en el presente documento.

60 En algunas realizaciones, un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en campo) se puede usar para realizar algunas o la totalidad de las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programable en campo puede cooperar con un microprocesador con el fin de realizar uno de los métodos descritos en el presente documento. En general, los métodos son realizados, preferiblemente, por cualquier aparato de soporte físico.

65

Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento resultarán evidentes a los expertos en la materia. Por lo tanto, se tiene por objeto estar limitado solo por el alcance de las inminentes reivindicaciones de patente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

5

REIVINDICACIONES

1. Un codificador de audio (100) para codificar muestras de audio, que comprende:

5 un primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (110) para codificar, usando una codificación de AAC, muestras de audio en un primer dominio de codificación, teniendo el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (110) una primera regla de entramado, una ventana de inicio y una ventana de detención y comprendiendo un transformador en el dominio de la frecuencia para transformar una primera trama de muestras de audio
 10 posteriores al dominio de la frecuencia sobre la base de una transformación de coseno discreta modificada, MDCT, estando configurado el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (110) para adaptar un tamaño de MDCT a las ventanas de inicio y de detención;
 un segundo codificador (120) para codificar, usando una codificación de AMR-WB+, muestras en un segundo dominio de codificación, teniendo el segundo codificador (120) un número de muestras de audio de tamaño de trama predeterminado, y un número de muestras de audio de periodo de puesta a punto de codificación,
 15 teniendo el segundo codificador (120) una segunda regla de entramado diferente, siendo una trama del segundo codificador (120) una representación codificada de un número de muestras de audio oportunamente posteriores, siendo el número igual al número de muestras de audio de tamaño de trama predeterminado; y
 un controlador (130) para conmutar del primer codificador (110) al segundo codificador (120) en respuesta a una característica de las muestras de audio y para, en la conmutación del primer codificador (110) al segundo codificador (120), modificar la ventana de inicio del primer codificador (110) hasta el punto de que la ventana de inicio tiene una longitud de 2048 muestras y se usa en una MDCT de 1024 puntos,
 la ventana de inicio se inicia directamente con una parte de flanco de subida que tiene un primer eje de plegado de MDCT en la mitad de la misma, la cual se extiende a lo largo de un primer y un segundo cuartos de la ventana de inicio hasta un centro de la ventana de inicio,
 25 una parte de derivación se extiende desde el centro hasta una parte de flanco de bajada, proporcionando la parte de flanco de bajada una sección de cruce con una ventana de seno tiene una longitud de 64 muestras y se extiende hasta un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana de inicio, y
 30 una parte cero se extiende a través desde el segundo eje de plegado de MDCT hasta un extremo de la ventana de inicio, y en donde se aplica una función ventana a la parte izquierda de las muestras de audio en el segundo dominio de codificación con una ventana de seno de desvanecimiento cruzado de una longitud de 64 muestras, o conmutar del segundo codificador (120) al primer codificador (110) en respuesta a una característica diferente de las muestras de audio y para, en la conmutación del segundo codificador (120) al primer codificador (110),
 35 modificar la ventana de detención del primer codificador (110) hasta el punto de que la ventana de detención tiene una longitud de 2304 muestras y se usa en una MDCT de 1152 puntos, una parte cero de la ventana de detención se extiende a través de un primer cuarto de la ventana de detención, una parte de flanco de subida de la ventana de detención, que es una ventana de seno de una longitud de 64 muestras, se inicia en un segundo cuarto de la ventana de detención de manera que un desvanecimiento cruzado comienza justo más allá de un primer eje de plegado de MDCT colocado entre la parte cero y la parte de flanco de subida,
 40 una parte de derivación de la ventana de detención se extiende desde la parte de flanco de subida hasta el centro de la ventana de detención, y
 una parte de flanco de bajada de la ventana de detención se extiende desde el centro de la ventana de detención a lo largo de un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana de detención hasta un extremo de la ventana de detención,
 45 en donde la segunda regla de entramado permanece sin modificaciones.

2. El codificador de audio (100) de la reivindicación 1, en donde el primer codificador de distorsión por repliegue del espectro en el dominio del tiempo (110) comprende un codificador de AAC de acuerdo con *Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding*, norma internacional 13818-7, ISO/IEC JTC1/SC29/WG11 Grupo de Expertos en Imágenes en Movimiento, 1997.

3. Un método para codificar tramas de audio, que comprende las etapas de:

55 codificar, con un primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (110), usando una codificación de AAC, muestras de audio en un primer dominio de codificación usando una primera regla de entramado, una ventana de inicio y una ventana de detención y mediante la transformación de una primera trama de muestras de audio posteriores al dominio de la frecuencia sobre la base de una transformación de coseno discreta modificada, MDCT, estando configurado el primer codificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (110) para adaptar un tamaño de MDCT a las ventanas de inicio y de detención;
 60 codificar, usando una codificación de AMR-WB+, muestras de audio en un segundo dominio de codificación usando un número de muestras de audio de tamaño de trama predeterminado y un número de muestras de audio de periodo de puesta a punto de codificación y usando una segunda regla de entramado diferente, siendo la trama del segundo dominio de codificación una representación codificada de un número de muestras de audio

oportunamente posteriores, siendo el número igual al número de muestras de audio de tamaño de trama predeterminado; y
 conmutar del primer dominio de codificación (110) al segundo dominio de codificación (120) en respuesta a una característica de las muestras de audio y para, en la conmutación del primer al segundo dominio de codificación,
 5 modificar la ventana de inicio del primer dominio de codificación (110) hasta el punto de que la ventana de inicio tiene una longitud de 2048 muestras y se usa en una MDCT de 1024 puntos,
 la ventana de inicio se inicia directamente con una parte de flanco de subida que tiene un primer eje de plegado de MDCT en la mitad de la misma, la cual se extiende a lo largo de un primer y un segundo cuartos de la ventana de inicio hasta un centro de la ventana de inicio,
 10 una parte de derivación se extiende desde el centro hasta una parte de flanco de bajada,
 proporcionando la parte de flanco de bajada una sección de cruce con una ventana de seno tiene una longitud de 64 muestras y se extiende hasta un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana de inicio, y
 15 una parte cero se extiende a través desde el segundo eje de plegado de MDCT hasta un extremo de la ventana de inicio, y en donde se aplica una función ventana a la parte izquierda de las muestras de audio en el segundo dominio de codificación con una ventana de seno de desvanecimiento cruzado de una longitud de 64 muestras, o conmutar del segundo dominio de codificación (120) al primer dominio de codificación (110) en respuesta a una característica diferente de las muestras de audio y para, en la conmutación del segundo al primer dominio de codificación, modificar la ventana de detención del primer dominio de codificación (110) hasta el punto de que la
 20 ventana de detención tiene una longitud de 2304 muestras y se usa en una MDCT de 1152 puntos,
 una parte cero de la ventana de detención se extiende a través de un primer cuarto de la ventana de detención, una parte de flanco de subida de la ventana de detención, que es una ventana de seno de una longitud de 64 muestras, se inicia en un segundo cuarto de la ventana de detención de manera que un desvanecimiento cruzado comienza justo más allá de un primer eje de plegado de MDCT colocado entre la parte cero y la parte de
 25 flanco de subida,
 una parte de derivación de la ventana de detención se extiende desde la parte de flanco de subida hasta el centro de la ventana de detención, y
 una parte de flanco de bajada de la ventana de detención se extiende desde el centro de la ventana de detención a lo largo de un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana de detención
 30 hasta un extremo de la ventana de detención, en donde la segunda regla de entramado permanece sin modificaciones.

4. Un decodificador de audio (150) para decodificar tramas codificadas de muestras de audio, que comprende:

35 un primer decodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (160) para decodificar, usando una decodificación de AAC, muestras de audio en un primer dominio de decodificación, teniendo el primer decodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (160) una primera regla de entramado, una ventana de inicio y una ventana de detención,
 40 comprendiendo el primer decodificador (160) un transformador en el dominio del tiempo para transformar una primera trama de muestras de audio decodificadas al dominio del tiempo sobre la base de una transformación de coseno discreta modificada inversa, IMDCT, estando configurado el primer decodificador de introducción de distorsión por repliegue del espectro en el dominio del tiempo (160) para adaptar un tamaño de IMDCT a las ventanas de inicio y de detención;
 45 un segundo decodificador (170) para decodificar, usando una decodificación de AMR-WB+, muestras de audio en un segundo dominio de decodificación y teniendo el segundo decodificador (170) un número de muestras de audio de tamaño de trama predeterminado y un número de muestras de audio de periodo de puesta a punto de codificación, teniendo el segundo decodificador (170) una segunda regla de entramado diferente, siendo una trama del segundo decodificador (170) una representación codificada de un número de muestras de audio oportunamente posteriores, siendo el número igual al número de muestras de audio de tamaño de trama
 50 predeterminado; y
 un controlador (180) para conmutar del primer decodificador al segundo decodificador sobre la base de una indicación a partir de la trama codificada de muestras de audio o conmutar del segundo decodificador (170) al primer decodificador (160) sobre la base de una indicación a partir de la trama codificada de muestras de audio y, en la conmutación del primer decodificador al segundo decodificador, modificar la ventana de inicio del
 55 primer decodificador hasta el punto de que la ventana de inicio tiene una longitud de 2048 muestras y se usa en una IMDCT de 1024 puntos,
 la ventana de inicio se inicia directamente con una parte de flanco de subida que tiene un primer eje de plegado de MDCT en la mitad de la misma, la cual se extiende a lo largo de un primer y un segundo cuartos de la ventana de inicio hasta un centro de la ventana de inicio,
 60 una parte de derivación se extiende desde el centro hasta una parte de flanco de bajada,
 proporcionando la parte de flanco de bajada una sección de cruce con una ventana de seno tiene una longitud de 64 muestras y se extiende hasta un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana de inicio, y
 una parte cero se extiende a través desde el segundo eje de plegado de MDCT hasta un extremo de la ventana
 65 de inicio, y en donde se aplica una función ventana a la parte izquierda de las muestras de audio en el segundo dominio de decodificación con una ventana de seno de desvanecimiento cruzado de una longitud de 64

muestras; o

conmutar del segundo descodificador al primer descodificador en respuesta a una característica diferente de las muestras de audio y para, en la conmutación del segundo descodificador al primer descodificador modificar la ventana de detención del primer descodificador hasta el punto de que la ventana de detención tiene una longitud

5 de 2304 muestras y se usa en una IMDCT de 1152 puntos,
 una parte cero de la ventana de detención se extiende a través de un primer cuarto de la ventana de detención,
 una parte de flanco de subida de la ventana de detención, que es una ventana de seno de una longitud de 64
 10 muestras, se inicia en un segundo cuarto de la ventana de detención de manera que un desvanecimiento
 cruzado comienza justo más allá de un primer eje de plegado de MDCT colocado entre la parte cero y la parte de
 flanco de subida,
 una parte de derivación de la ventana de detención se extiende desde la parte de flanco de subida hasta el
 centro de la ventana de detención, y
 una parte de flanco de bajada de la ventana de detención se extiende desde el centro de la ventana de detención
 a lo largo de un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana de detención
 15 hasta un extremo de la ventana de detención, y en donde se aplica una función ventana a las últimas 64
 muestras descodificadas en el primer dominio de descodificación con una ventana de seno al cuadrado de una
 longitud de 64 muestras,
 en donde la segunda regla de entramado permanece sin modificaciones.

20 5. Un método para descodificar tramas codificadas de muestras de audio, que comprende las etapas de
 descodificar, usando una descodificación de AAC, muestras de audio en un primer dominio de descodificación,
 introduciendo el primer dominio de descodificación una distorsión por repliegue del espectro de tiempo, teniendo una
 primera regla de entramado, una ventana de inicio y una ventana de detención, y mediante el uso de la
 transformación de una primera trama de muestras de audio descodificadas al dominio del tiempo sobre la base de
 25 una transformación de coseno discreta modificada inversa, IMDCT, estando configurado el primer descodificador de
 introducción de distorsión por repliegue del espectro en el dominio del tiempo (160) para adaptar un tamaño de
 IMDCT a las ventanas de inicio y de detención;
 descodificar, usando una descodificación de AMR-WB+, muestras de audio en un segundo dominio de
 descodificación, teniendo el segundo dominio de descodificación un número de muestras de audio de tamaño de
 30 trama predeterminado y un número de muestras de audio de periodo de puesta a punto de codificación, teniendo el
 segundo dominio de descodificación una segunda regla de entramado diferente, siendo una trama del segundo
 dominio de descodificación una representación descodificada de un número de muestras de audio oportunamente
 posteriores, siendo el número igual al número de muestras de audio de tamaño de trama predeterminado; y
 conmutar del primer dominio de descodificación al segundo dominio de descodificación sobre la base de una
 35 indicación a partir de la trama codificada de muestras de audio o conmutar del segundo dominio de descodificación
 (170) al primer dominio de descodificación (160) sobre la base de una indicación a partir de la trama codificada de
 muestras de audio y, en la conmutación del primer dominio de descodificación al segundo dominio de
 descodificación, modificar la ventana de inicio del primer dominio de descodificación hasta el punto de que la
 ventana de inicio tiene una longitud de 2048 muestras y se usa en una IMDCT de 1024 puntos,
 40 la ventana de inicio se inicia directamente con una parte de flanco de subida que tiene un primer eje de plegado de
 MDCT en la mitad de la misma, la cual se extiende a lo largo de un primer y un segundo cuartos de la ventana de
 inicio hasta un centro de la ventana de inicio,
 una parte de derivación se extiende desde el centro hasta una parte de flanco de bajada,
 proporcionando la parte de flanco de bajada una sección de cruce con una ventana de seno tiene una longitud de 64
 45 muestras y se extiende hasta un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana
 de inicio, y
 una parte cero se extiende a través desde el segundo eje de plegado de MDCT hasta un extremo de la ventana de
 inicio y en donde se aplica una función ventana a la parte izquierda de las muestras de audio en el segundo dominio
 de descodificación con una ventana de seno de desvanecimiento cruzado de una longitud de 64 muestras; o
 50 conmutar del segundo dominio de descodificación al primer dominio de descodificación en respuesta a una
 característica diferente de las muestras de audio y para, en la conmutación del segundo dominio de descodificación
 al primer dominio de descodificación, modificar la ventana de detención del primer dominio de descodificación hasta
 el punto de que la ventana de detención tiene una longitud de 2304 muestras y se usa en una IMDCT de 1152
 puntos,
 55 una parte cero de la ventana de detención se extiende a través de un primer cuarto de la ventana de detención,
 una parte de flanco de subida de la ventana de detención, que es una ventana de seno de una longitud de 64
 muestras, se inicia en un segundo cuarto de la ventana de detención de manera que un desvanecimiento cruzado
 comienza justo más allá de un primer eje de plegado de MDCT colocado entre la parte cero y la parte de flanco de
 subida,
 60 una parte de derivación de la ventana de detención se extiende desde la parte de flanco de subida hasta el centro de
 la ventana de detención, y
 una parte de flanco de bajada de la ventana de detención se extiende desde el centro de la ventana de detención a
 lo largo de un segundo eje de plegado de MDCT entre un tercer y un cuarto cuarto de la ventana de detención hasta
 un extremo de la ventana de detención, y en donde se aplica una función ventana a las últimas 64 muestras de
 65 audio descodificadas en el primer dominio de descodificación con una ventana de seno al cuadrado de una longitud
 de 64 muestras,

en donde la segunda regla de entramado permanece sin modificaciones.

6. Un programa informático que tiene un código de programa adaptado para realizar el método de la reivindicación 3 o 5, cuando el código de programa se ejecuta en un ordenador o procesador.

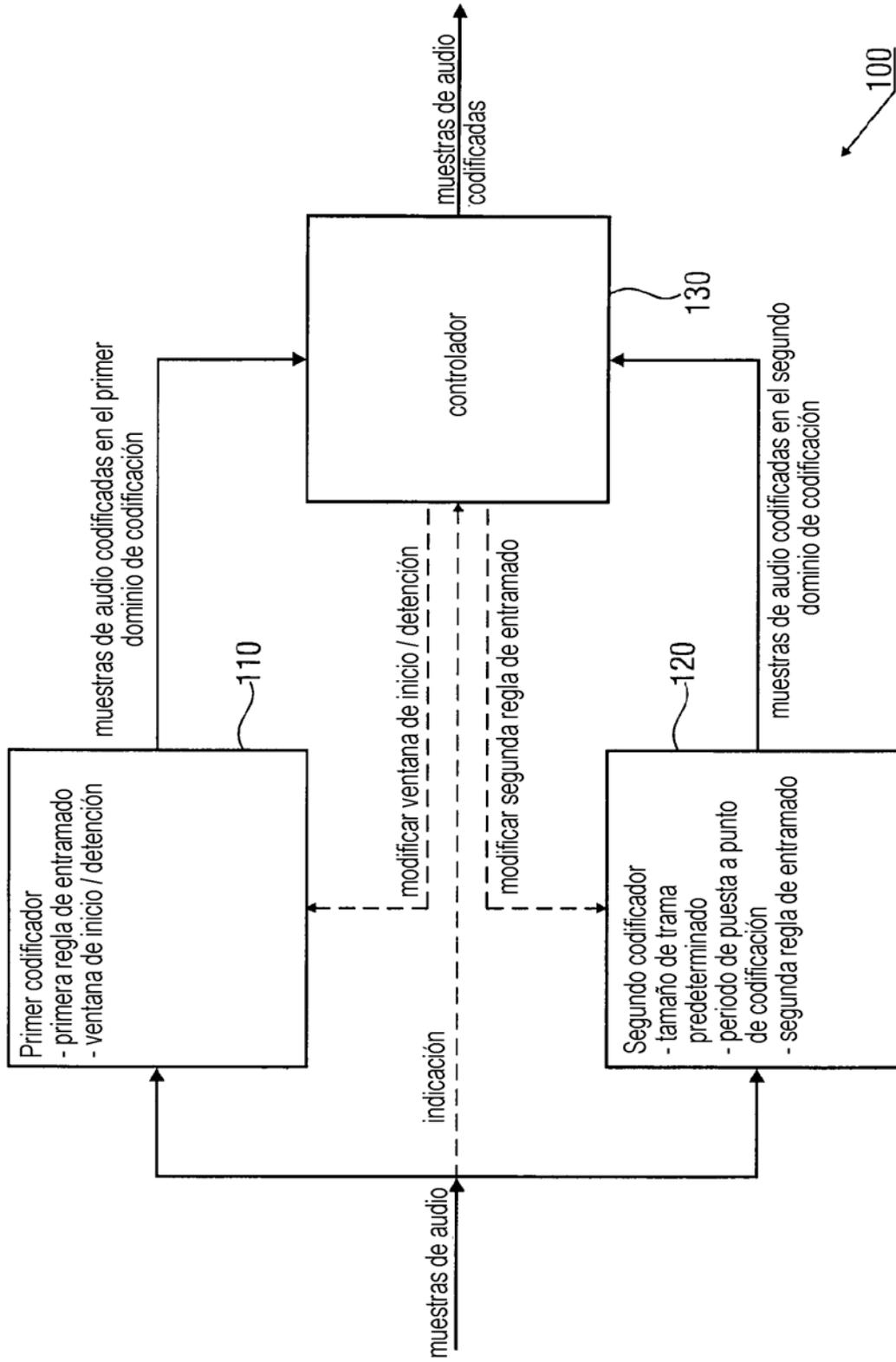


FIG 1A

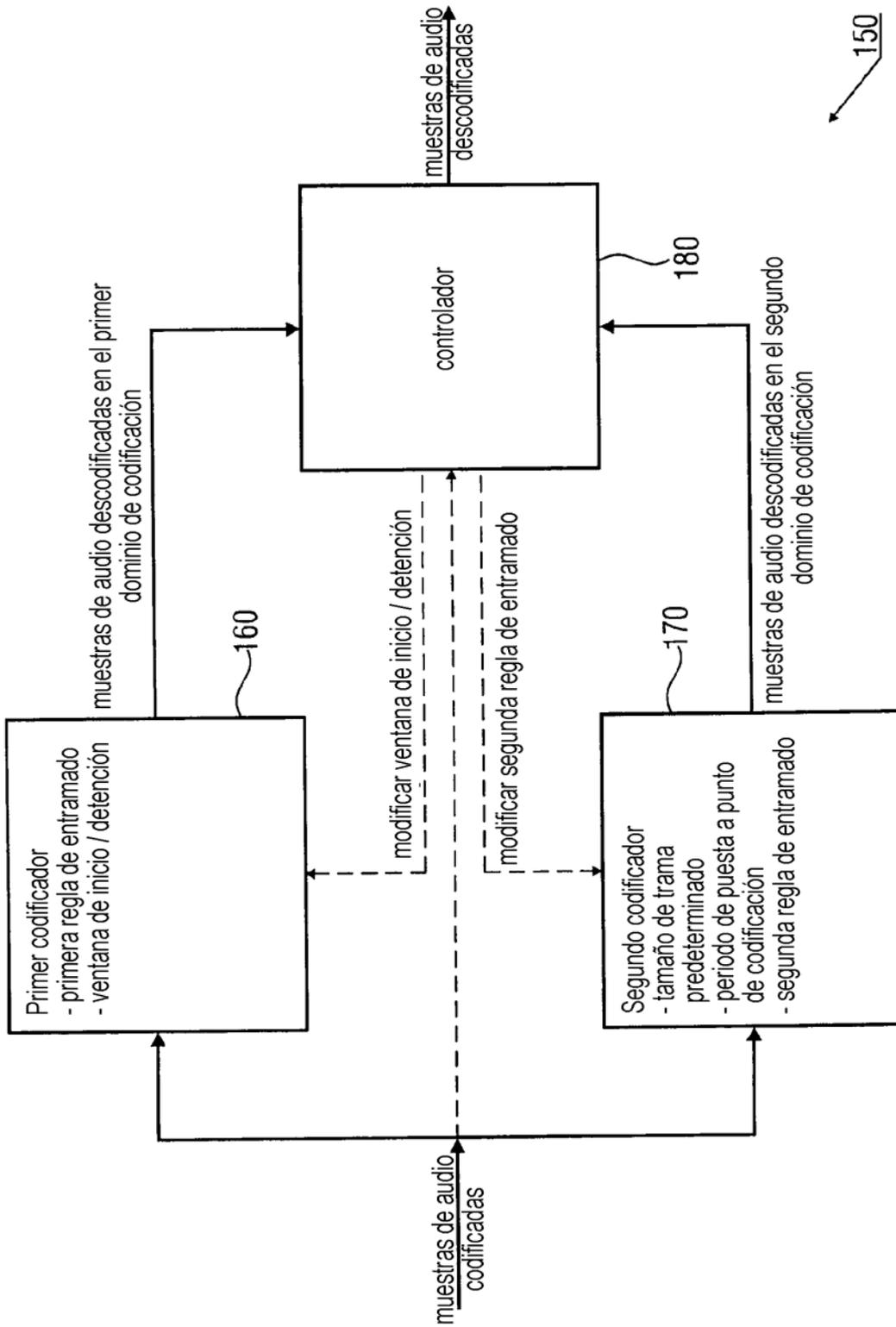


FIG 1B

- a)
$$X_k = \sum_{n=0}^{2N-1} X_n \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right]$$
- b)
$$y_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right]$$
- c)
$$w_n^2 + w_{n+N}^2 = 1$$
- d)
$$w_n = \sin \left[\frac{\pi}{2N} \left(n + \frac{1}{2} \right) \right]$$
- e)
$$w_n = \sin \left(\frac{\pi}{2} \sin^2 \left[\frac{\pi}{2N} \left(n + \frac{1}{2} \right) \right] \right)$$
- f)
$$\cos \left[\frac{\pi}{N} \left(-n-1 + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right] = \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right] y$$

$$\cos \left[\frac{\pi}{N} \left(2N-n-1 + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right] = -\cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right]$$
- g)
$$\text{IMDCT}(\text{MDCT}(a, b, c, d)) = (a-b_R, b-a_R, c+d_R, c_R+d)/2$$
- h)
$$(z_R, w_R) \cdot (z_R c + w d_R, z c_R + w_R d) = (z_R^2 c + w z_R d_R, w_R z c_R + w_R^2 d)$$
- i)
$$(w, z) \cdot (w c - z_R d_R, z d - w_R c_R) = (w^2 c + w z_R d_R, z^2 d - w_R z c_R)$$
- j)
$$(z_R^2 c + w z_R d_R, w_R z c_R + w_R^2 d) + (w^2 c - w z_R d_R, z^2 d - w_R z c_R)$$

$$= ([z_R^2 + w^2] c + [w z_R - w z_R] d_R, [w_R^2 + z^2] d + [w_R z - w_R z] c_R) = (c, d)$$

FIG 2

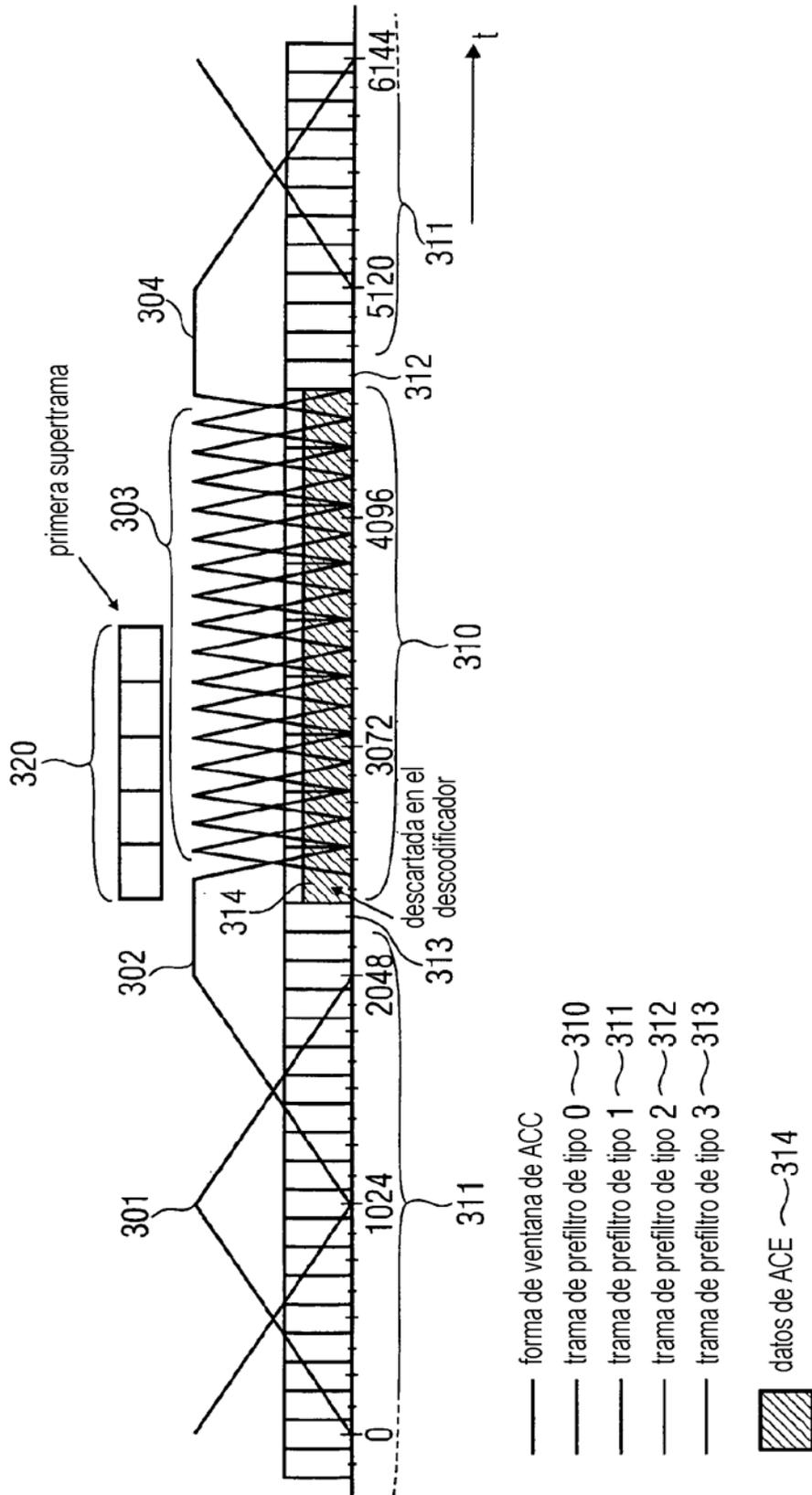


FIG 3

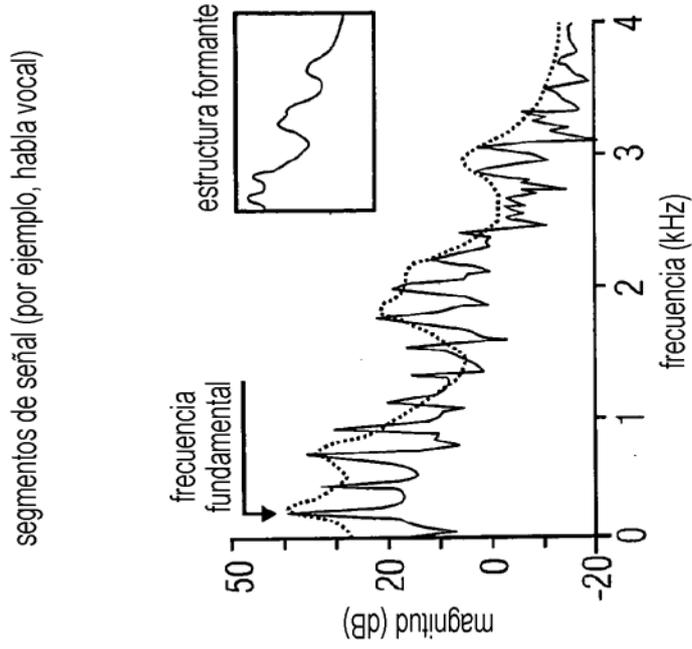


FIG 4B

FIG 4A

segmento / señal de tipo ruido (por ejemplo, habla no vocal)

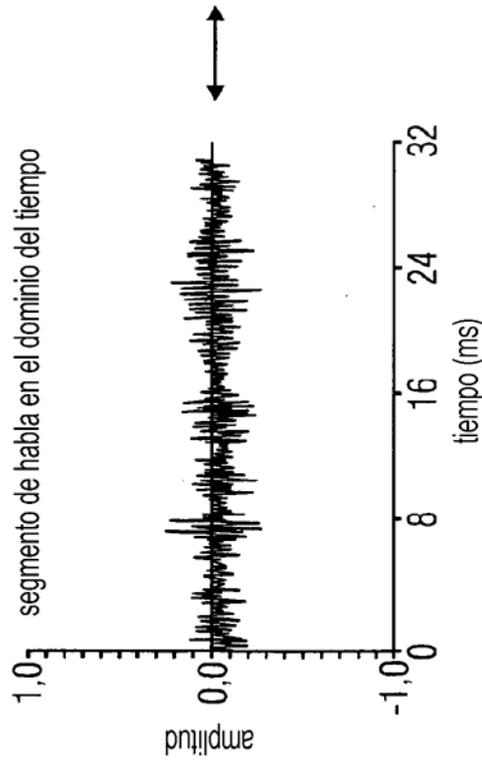


FIG 5A

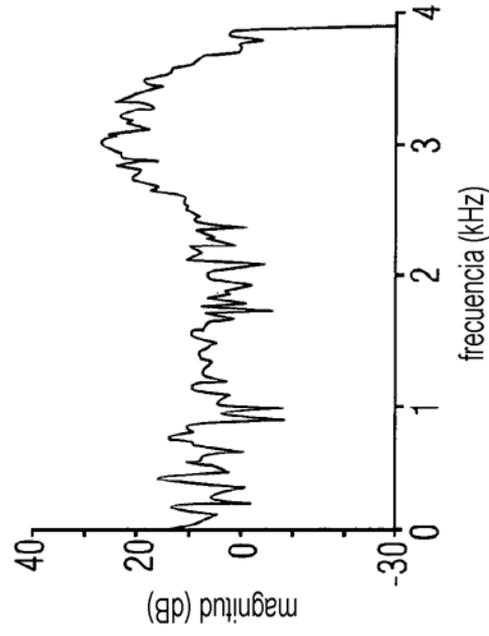
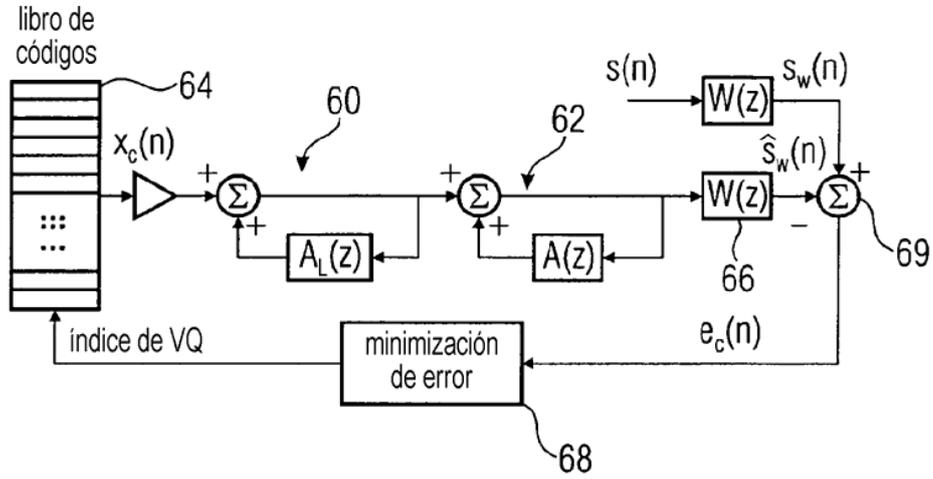


FIG 5B

CELP de análisis por síntesis



$A_L(z)$: predicción a largo plazo
 \cong estructura (fina) de paso

$A(z)$: predicción a corto plazo
 \cong estructura formante / envolvente espectral

FIG 6

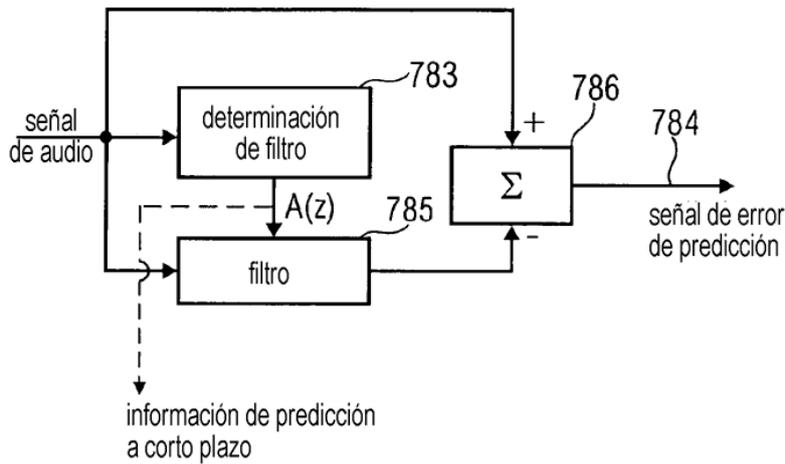


FIG 7

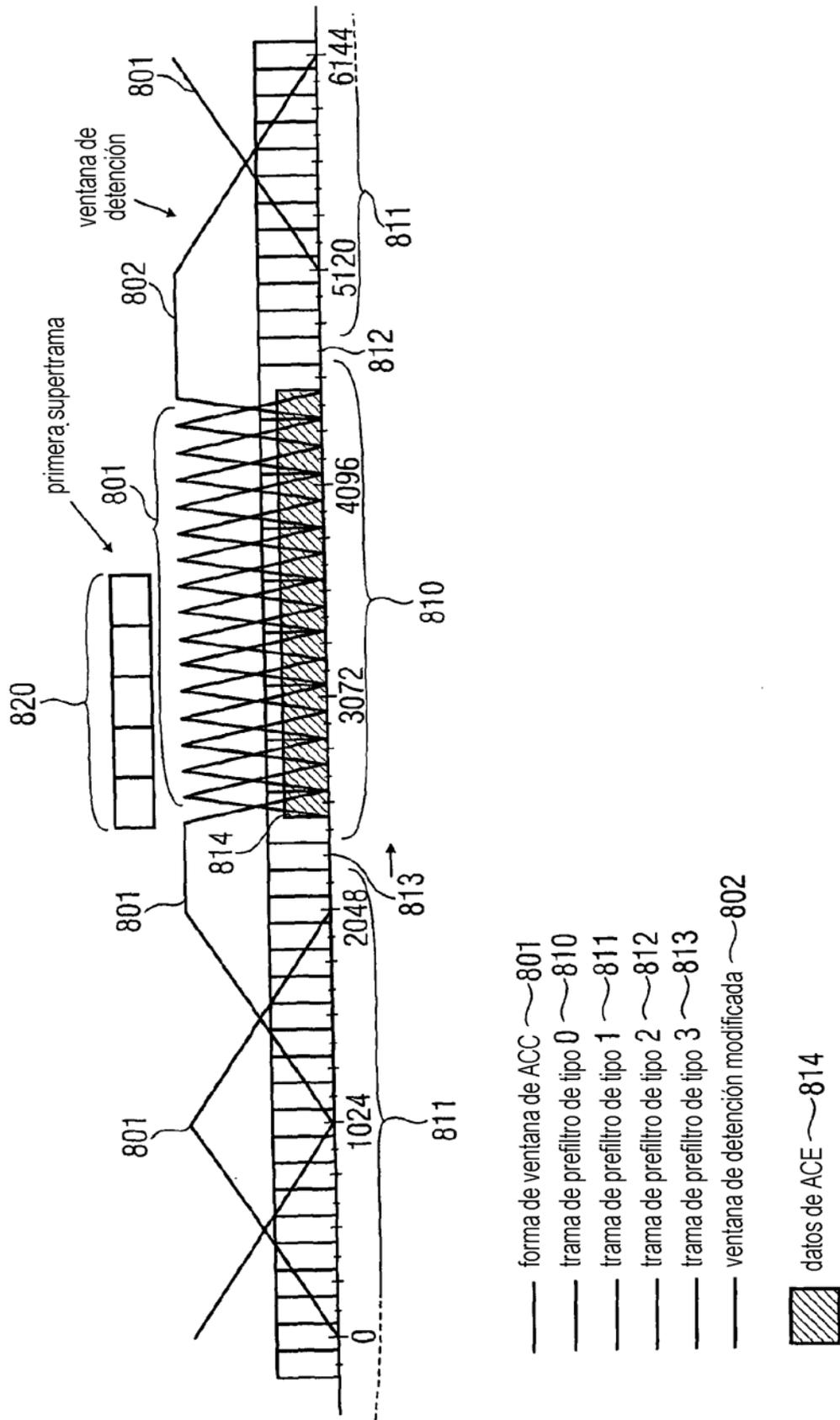
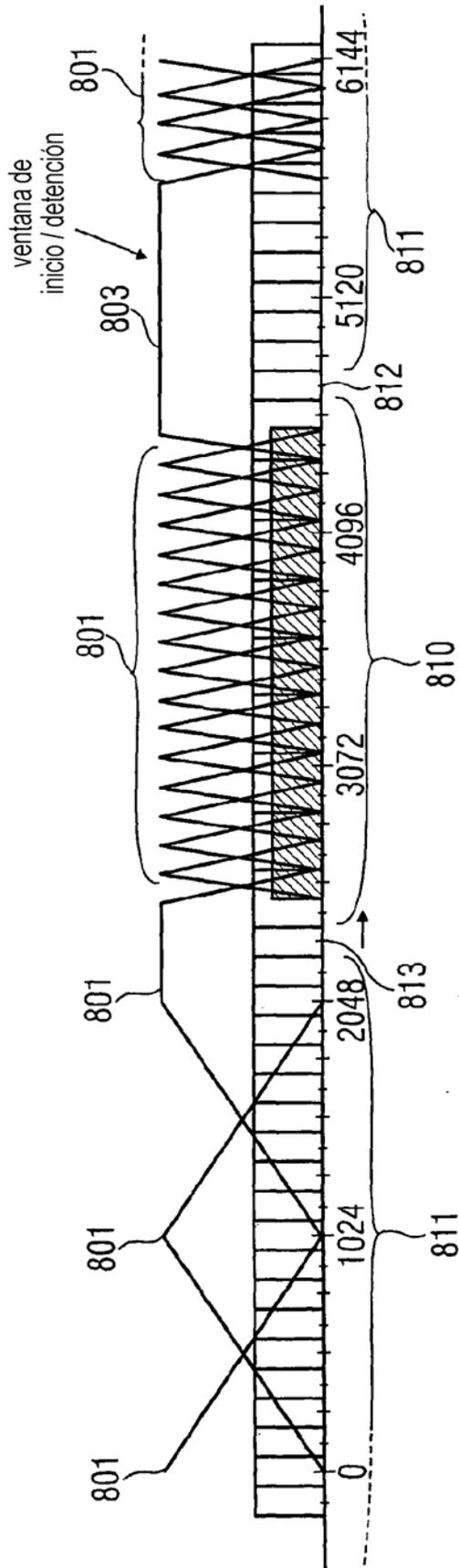


FIG 8A



- forma de ventana de ACC 801
- trama de prefiltro de tipo 0 810
- trama de prefiltro de tipo 1 811
- trama de prefiltro de tipo 2 812
- trama de prefiltro de tipo 3 813
- ventana de inicio / detención modificada 803

▨ datos de ACE

FIG 8B

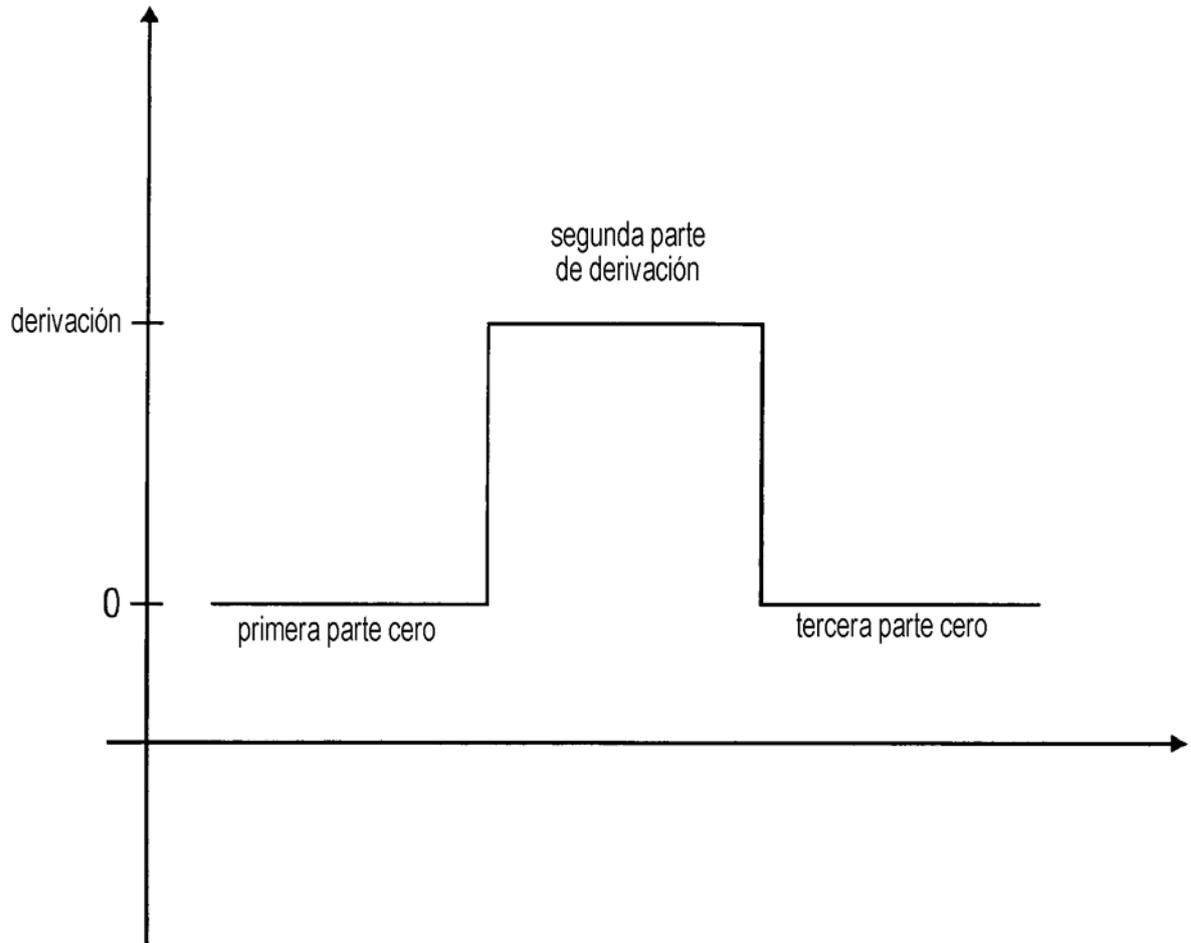


FIG 9

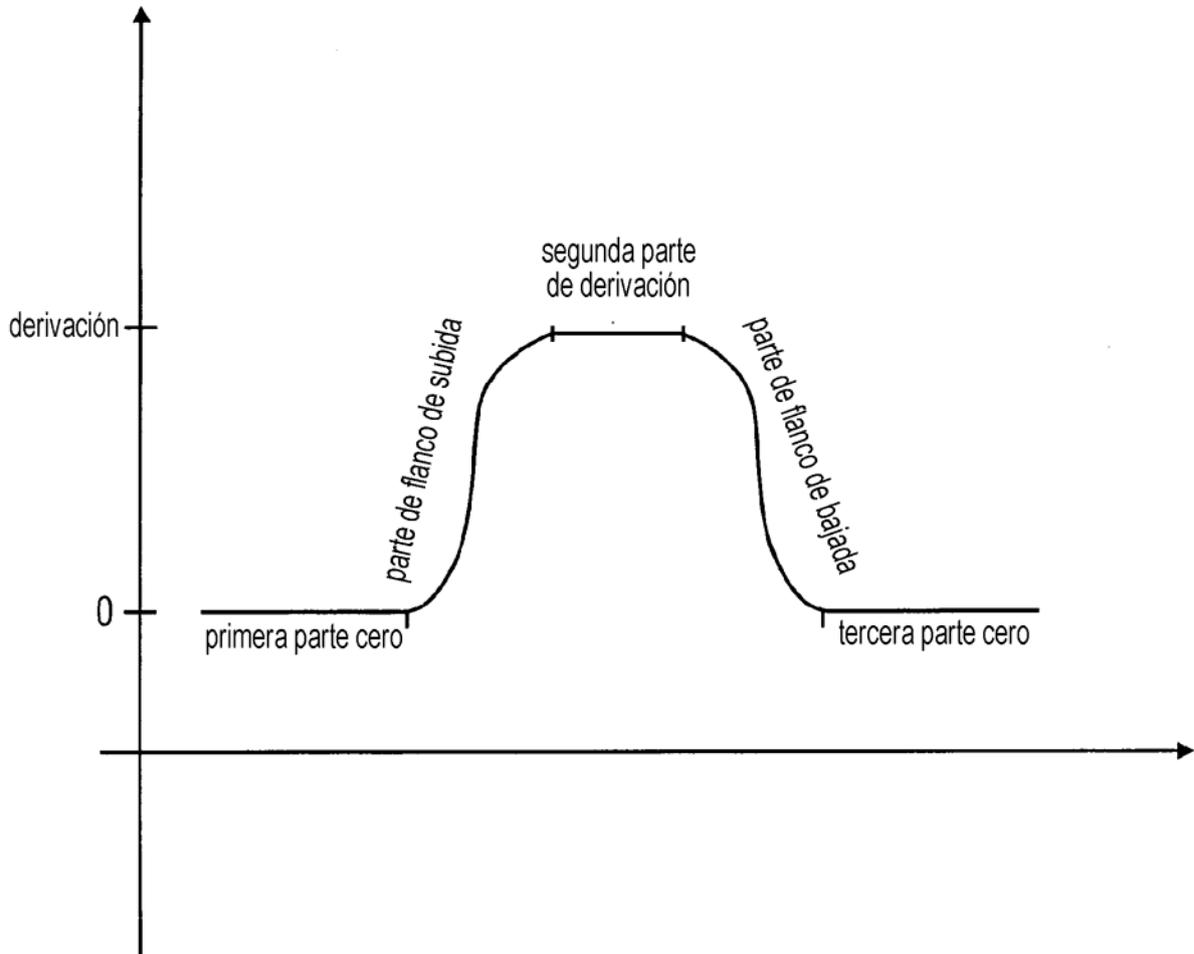


FIG 10

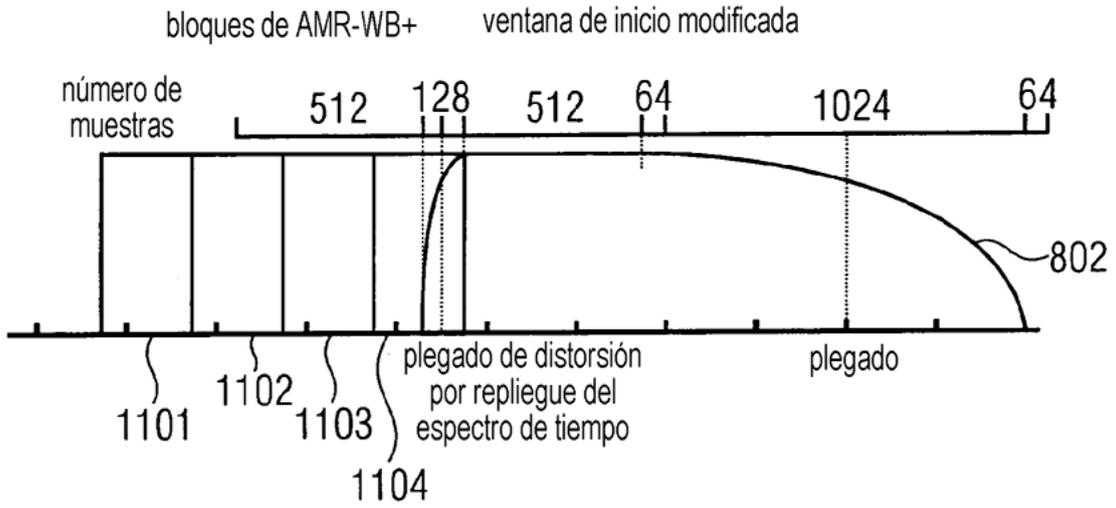


FIG 11

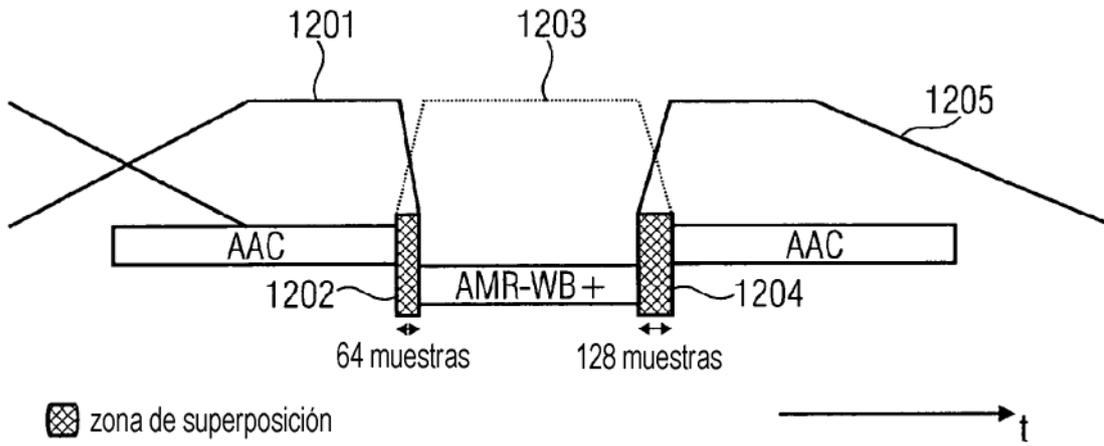


FIG 12

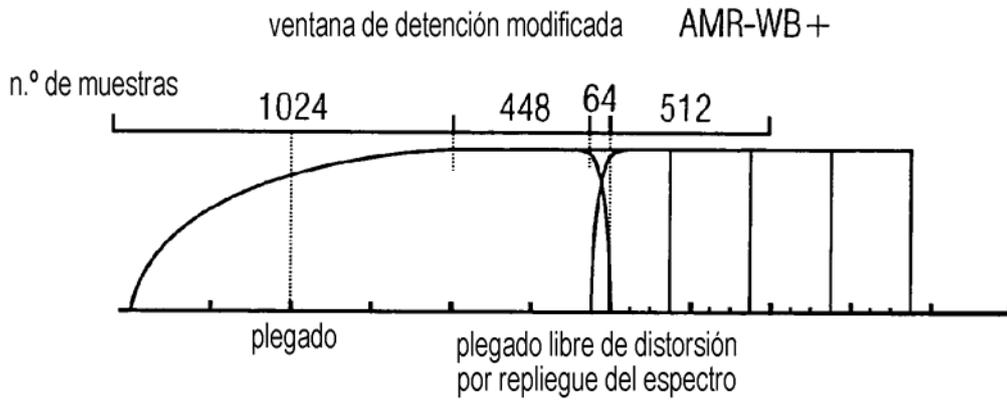


FIG 13

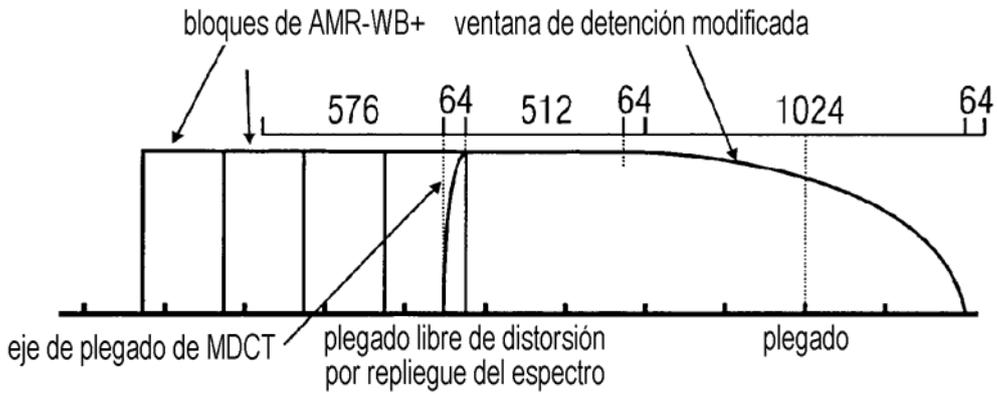


FIG 14

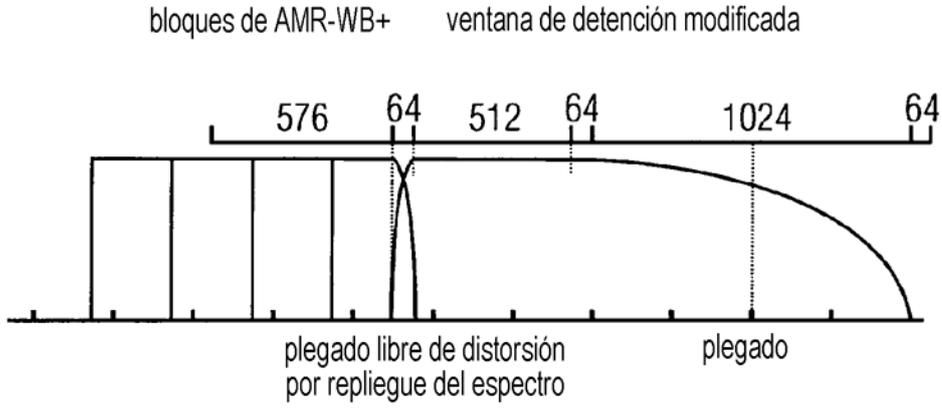


FIG 15

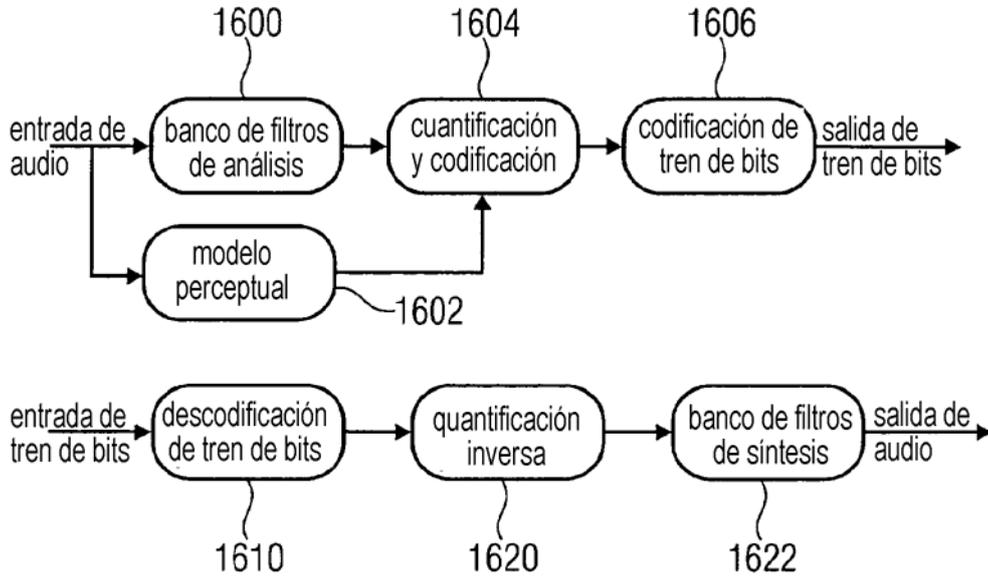


FIG 16
(TÉCNICA ANTERIOR)

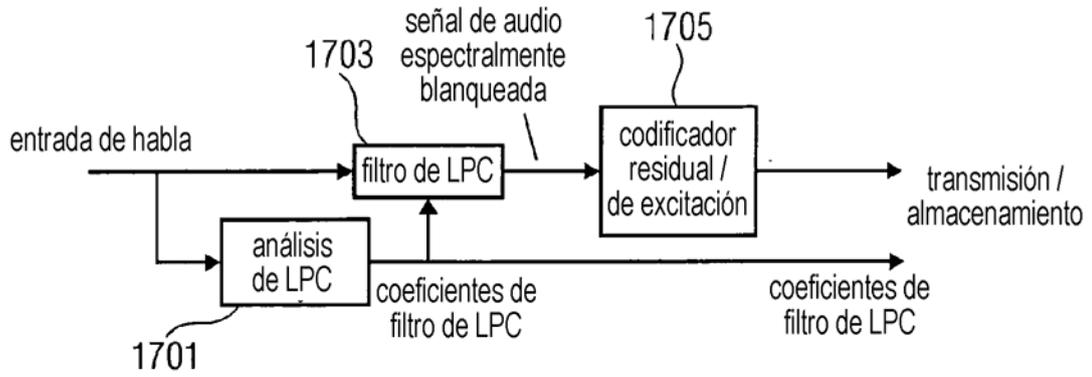


FIG 17A
(TÉCNICA ANTERIOR)

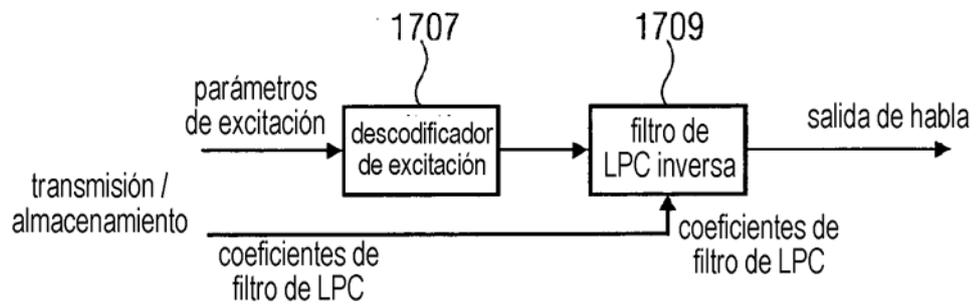


FIG 17B
(TÉCNICA ANTERIOR)

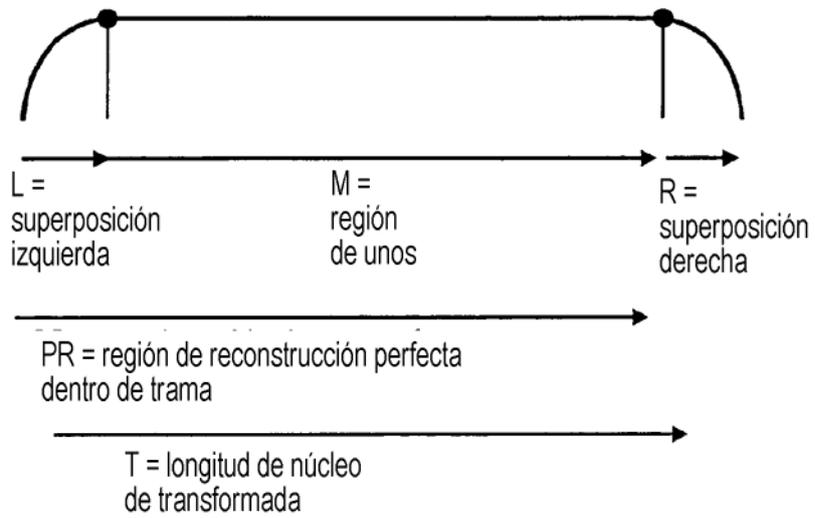
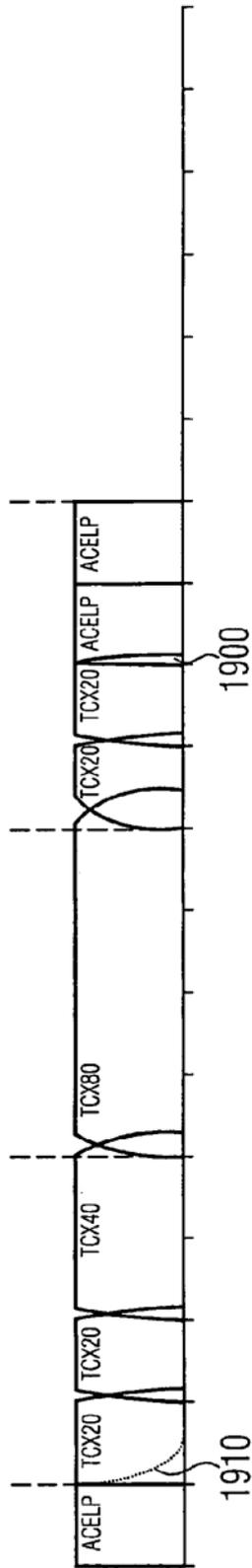


FIG 18
(TÉCNICA ANTERIOR)



parámetros de ventana para AMR-WB+

| transición a → de ↓ | TCX80 | TCX40 | TCX20 |
|------------------------|--------------------------------------|-----------------------------------|-----------------------------------|
| ACELP | L=0, M=1024, R=128, PR=1024, T=1152 | L=0, M=512, R=64, PR=512, T=576 | L=0, M=256, R=32, PR=256, T=288 |
| TCX20 | L=32, M=992, R=128, PR=1024, T=1152 | L=32, M=480, R=64, PR=512, T=576 | L=32, M=224, R=32, PR=256, T=288 |
| TCX40 | L=64, M=960, R=128, PR=1024, T=1152 | L=64, M=448, R=64, PR=512, T=576 | L=64, M=192, R=32, PR=256, T=288 |
| TCX80 | L=128, M=896, R=128, PR=1024, T=1152 | L=128, M=384, R=64, PR=512, T=576 | L=128, M=128, R=32, PR=256, T=288 |

FIG 19
(TÉCNICA ANTERIOR)

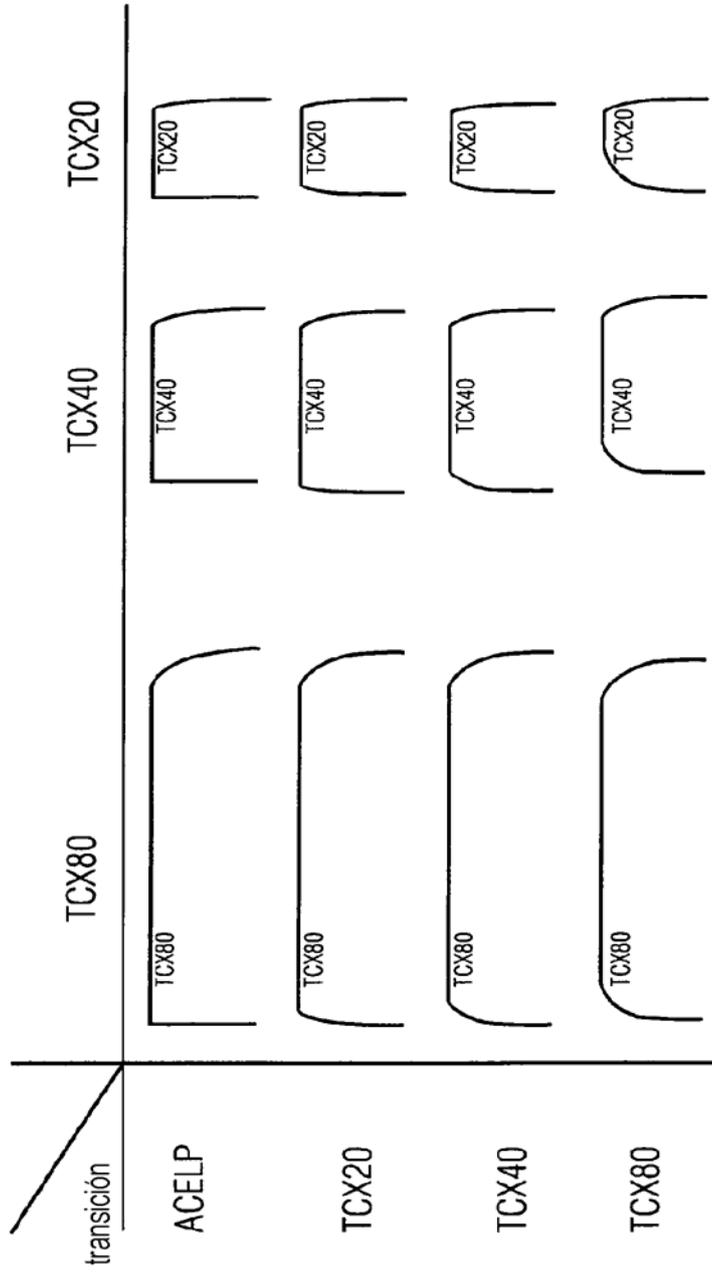


FIG 20
(TÉCNICA ANTERIOR)

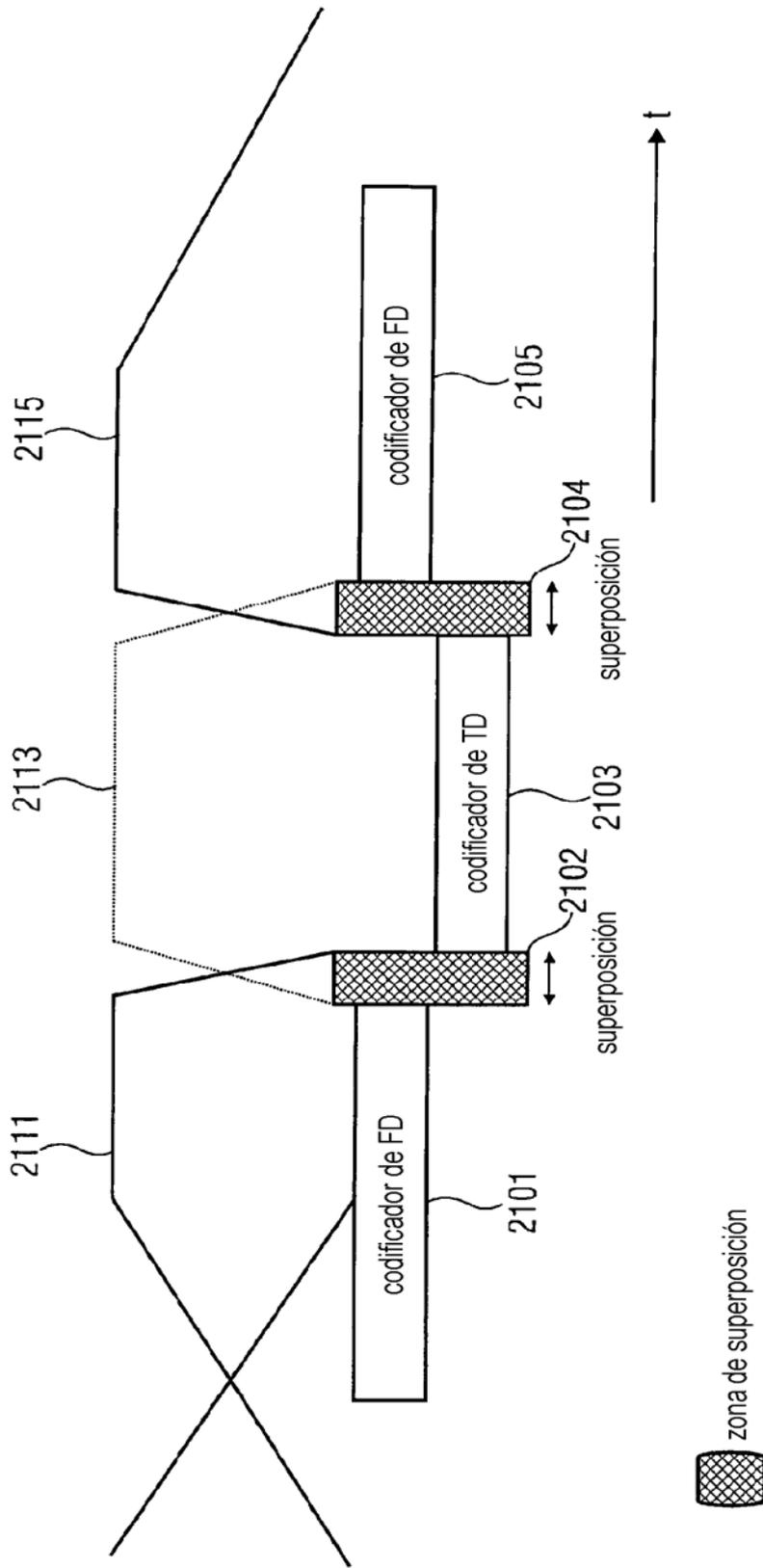


FIG 21
(TÉCNICA ANTERIOR)

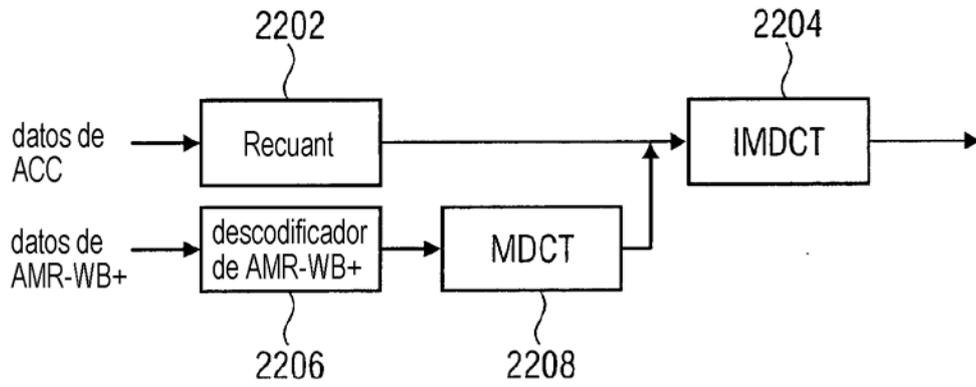


FIG 22
(TÉCNICA ANTERIOR)

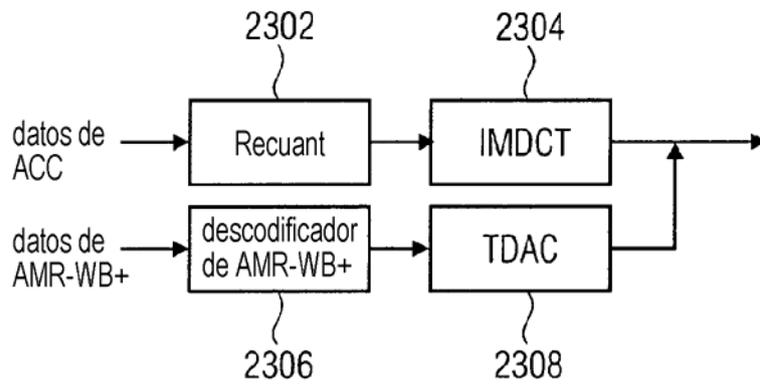


FIG 23
(TÉCNICA ANTERIOR)