

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 963 367**

51 Int. Cl.:

G10L 19/022 (2013.01)

G10L 19/04 (2013.01)

G10L 19/02 (2013.01)

G10L 19/18 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **14.02.2012** **E 19157006 (8)**

97 Fecha y número de publicación de la concesión europea: **30.08.2023** **EP 3503098**

54 Título: **Aparato y procedimiento de decodificación de una señal de audio usando una parte de anticipación alineada**

30 Prioridad:

14.02.2011 US 201161442632 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.03.2024

73 Titular/es:

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

72 Inventor/es:

**RAVELLI, EMMANUEL;
GEIGER, RALF;
SCHNELL, MARKUS;
FUCHS, GUILLAUME;
RUOPPILA, VESA;
BÄCKSTRÖM, TOM;
GRILL, BERNHARD y
HELMRICH, CHRISTIAN**

74 Agente/Representante:

PONTI & PARTNERS, S.L.P.

ES 2 963 367 T3

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

DESCRIPCIÓN

Aparato y procedimiento de decodificación de una señal de audio usando una parte de anticipación alineada

5 **[0001]** La presente invención se refiere a codificación de audio y, en particular, a codificación de audio que se basa en codificadores de audio conmutados, y de forma correspondiente en decodificadores de audio controlados, en particular adecuados para aplicaciones de bajo retardo.

[0002] Se conocen varios conceptos sobre codificadores de audio que dependen de códec conmutados. Un
10 concepto conocido sobre codificación de audio es el denominado códec de banda ancha de múltiple velocidad extendido adaptativo o AMR-WB+ (Extended Adaptive Multi-Rate-Wideband), tal como se describe en 3GPP TS 26.290 B10.0.0 (2011-03). El códec de audio AMR-WB+ contiene todos los modos del códec de voz AMR-WB 1 a 9 y AMR-WB VAD y DTX. AMR-WB+ extiende el códec AMR-WB añadiendo TCX, extensión de banda ancha y estéreo.

15 **[0003]** Los procesos del códec de audio AMR-WB+ introducen tramas iguales a 2.048 muestras a una frecuencia de muestreo interna F_s . La frecuencia de muestreo interna se limita al intervalo de 12.800 a 38.400 Hz. Las tramas de 2.048 muestras se dividen en dos bandas de frecuencia iguales sometidas a muestreo crítico. Se obtienen así dos supertramas de 1.024 muestras correspondientes a bandas de baja frecuencia (LF, low frequency)
20 y alta frecuencia (HF, high frequency). Cada supertrama se divide en cuatro tramas de 256 muestras. El muestreo a la velocidad de muestreo interna se obtiene usando un esquema de conversión de muestreo variable que remuestrea la señal de entrada.

[0004] A continuación, las señales de LF y HF se codifican usando dos enfoques diferentes: la señal LF se
25 codifica y decodifica usando el codificador/decodificador "central", basado en el modo ACELP y excitación codificada por transformada (TCX) conmutado. En el modo ACELP, se usa el códec AMR-WB estándar. La señal HF se codifica con relativamente pocos bits (16 bits/trama) usando el procedimiento de extensión de ancho de banda (BWE, Bandwidth Extension). Los parámetros transmitidos desde el codificador al decodificador son los bits de selección de modo, los parámetros LF y los parámetros HF. Los parámetros para cada supertrama de 1.024
30 muestras se descomponen en cuatro paquetes de idéntico tamaño. Cuando la señal de entrada es estéreo, los canales izquierdo y derecho se combinan en señales mono para la codificación ACELP-TCX, mientras que la codificación estéreo recibe los dos canales de entrada. En el lado del decodificador, las bandas LF y HF se decodifican por separado, después de lo cual las bandas se combinan en un banco de filtro de síntesis. Si la salida se restringe a mono sólo, los parámetros estéreo se omiten y el decodificador actúa en modo mono. El códec AMR-
35 WB+ aplica análisis LP para los modos ACELP y TCX cuando se codifica la señal LF. Los coeficientes LP se interpolan linealmente en cada subtrama de 64 muestras. La ventana de análisis LP es un medio coseno de 384 muestras de longitud. Para codificar la señal mono central se usa una codificación ACELP o TCX para cada trama. El modo de codificación se selecciona basado en un procedimiento de bucle cerrado de análisis-síntesis. Sólo se consideran tramas de 256 muestras para las tramas ACELP, mientras que las tramas de 256, 512 o 1.024 muestras
40 son posibles en el modo TCX. La ventana usada para el análisis LPC en AMR-WB+ se ilustra en la fig. 5b. Se usa una ventana de análisis LPC simétrica con anticipación de 20 ms. Anticipación significa que, como se ilustra en la fig. 5b, la ventana de análisis LPC para la trama actual ilustrada en 500 no sólo se extiende dentro de la trama actual indicada entre 0 y 20 ms en la fig. 5b ilustrada por 502, sino que se extiende en la trama futura entre 20 y 40 ms. Esto significa que al usar esta ventana de análisis LPC, se necesita un retardo adicional de 20 ms, es decir, una
45 trama futura completa. Por tanto, la parte de anticipación indicada en 504 en la fig. 5b contribuye al retardo sistemático asociado con el codificador AMR-WB+. Dicho de otro modo, deberá estar totalmente disponible una trama futura para que puedan calcularse los coeficientes del análisis LPC para la trama actual 502.

[0005] La fig. 5a ilustra un codificador adicional, denominado codificador AMR-WB y, en particular, la ventana
50 de análisis LPC usada para calcular los coeficientes de análisis para la trama actual. Una vez más, la trama actual se extiende entre 0 y 20 ms y la trama futura se extiende entre 20 y 40 ms. A diferencia de la fig. 5b, la ventana de análisis LPC de AMR-WB indicada en 506 tiene una parte de anticipación 508 de 5 ms sólo, es decir, la separación de tiempo entre 20 ms y 25 ms. De este modo, el retardo introducido por el análisis LPC se reduce sustancialmente con respecto a la fig. 5a. Por otro lado, sin embargo, se ha descubierto que una parte mayor de anticipación para
55 determinar los coeficientes LPC, es decir, una parte mayor de anticipación para la ventana de análisis LPC produce mejores coeficientes LPC y, por tanto, una menor energía en la señal residual y, por tanto, una velocidad binaria menor, ya que la predicción LPC se adapta mejor a la señal original.

[0006] Mientras las fig. 5a y 5b se refieren a codificadores que tienen sólo una única ventana de análisis para
60 determinar los coeficientes LPC para una trama, la fig. 5c ilustra la situación para el codificador de voz G.718. La especificación G.718 (06-2008) se refiere a sistemas de transmisión y sistemas y redes digitales de medios y, en particular, describe equipos de terminal digital y, en particular, una codificación de señales de voz y audio para dichos equipos. En particular, esta norma se refiere a codificación de voz y audio con velocidad binaria variable que incluye banda estrecha y banda ancha de 8-32 kbit/s robusta tal como se define en la recomendación ITU-T G.718.
65 La señal de entrada se procesa usando tramas de 20 ms. El retardo del códec depende de la velocidad de muestreo

de entrada y de salida. Para una entrada de banda ancha y una salida de banda ancha, el retardo algorítmico total de esta codificación es 42,875 ms. Consiste en una trama de 20 ms, un retardo de 1,875 ms de filtros de remuestreo de entrada y salida, 10 ms para la anticipación del codificador, 1 ms de retardo de posfiltrado y 10 ms en el decodificador para permitir la operación de superposición y suma de la codificación de transformada de capa superior. Para una entrada de banda estrecha y salida de banda estrecha no se usan capas superiores, sino que se usa el retardo del decodificador de 10 ms para mejorar el rendimiento de la codificación en presencia de borrados de trama y para señales de música. Si la salida se limita a la capa 2, el retardo de códec puede reducirse en 10 ms. La descripción del codificador es la siguiente. Las dos capas inferiores se aplican a una señal preacentuada sometida a muestreo a 12,8 kHz, y las tres capas superiores actúan en el dominio de la señal de entrada sometida a muestreo a 16 kHz. La capa de núcleo se basa en la tecnología de predicción lineal excitada por código (CELP, Code-Excited Linear Prediction), donde la señal de voz es modelada por una señal de excitación que hace pasar por un filtro de síntesis de predicción lineal (LP, linear prediction) que representa la envolvente espectral. El filtro LP se cuantifica en el dominio de frecuencia espectral de inmitancia (ISF, Immitance Spectral Frequency) usando un enfoque de predicción-conmutación y cuantificación de vector de etapas múltiples. El análisis de altura de tono en bucle abierto se realiza mediante un algoritmo de rastreo de altura de tono para asegurar un contorno de altura de tono suave. Se comparan dos contornos de evolución de altura de tono concurrente y se selecciona el rastro que produce el contorno más suave que produce la estimación de altura de tono más robusta. El preprocesamiento de nivel de trama comprende un filtrado de paso alto, una conversión de muestreo a 12.800 muestras por segundo, una preacentuación, un análisis espectral, una detección de entradas de banda estrecha, una detección de actividad de voz, una estimación de ruido, reducción de ruido, análisis de predicción lineal, conversión de LP a ISF y una interpolación, una computación de señal de voz ponderada, un análisis de altura de tono en bucle abierto, una actualización de ruido de fondo, una clasificación de señal para una selección de modo de codificación y un ocultamiento de borrado de trama. La codificación de capa 1 que usa el tipo de codificador seleccionado comprende un modo de codificación sin voz, un modo de codificación con voz, un modo de codificación de transición, un modo de codificación genérica y una transmisión discontinua y generación de ruido de confort (DTX/CNG, Discontinuous Transmission and Comfort Noise Generation).

[0007] Un análisis de predicción a largo plazo o predicción lineal (LP) que usa el enfoque de autocorrelación determina los coeficientes del filtro de síntesis del modelo CELP. Sin embargo, en CELP la predicción a largo plazo es normalmente el "libro de códigos adaptativo" y por tanto es diferente de la predicción lineal. La predicción lineal puede, por tanto, considerarse más una predicción a corto plazo. La autocorrelación de la voz dividida en ventanas se convierte en coeficientes LP usando el algoritmo de Levinson-Durbin. A continuación, los coeficientes LPC se transforman en pares espectrales de inmitancia (ISP, Immitance Spectral Pairs) y en consecuencia en frecuencias espectrales de inmitancia (ISF, Immitance Spectral Frequencies) con fines de cuantificación e interpolación. Los coeficientes interpolados cuantificados y no cuantificados se convierten nuevamente al dominio LP para construir filtros de síntesis y ponderación para cada subtrama. En caso de codificación de una trama de señal activa, se estiman dos grupos de coeficientes LP en cada trama usando las dos ventanas de análisis LPC indicadas en 510 y 512 en la fig. 5c. La ventana 512 se denomina "ventana LPC de mitad de trama", y la ventana 510 se denomina "ventana LPC de final de trama". Se usa una parte de anticipación 514 de 10 ms para calcular la autocorrelación de final de trama. La estructura de la trama se ilustra en la fig. 5c. La trama se divide en cuatro subtramas, teniendo cada subtrama una longitud de 5 ms correspondiente a 64 muestras a velocidad de muestreo de 12,8 kHz. Las ventanas para el análisis de final de trama y para el análisis de mitad de trama se centran en la cuarta subtrama y la segunda subtrama, respectivamente tal como se ilustra en la fig. 5c. Para la división en ventanas se usa una ventana Hamming con longitud de 320 muestras. Los coeficientes se definen en G.718, Sección 6.4.1. El cálculo de autocorrelación se describe en la Sección 6.4.2. El algoritmo de Levinson-Durbin se describe en la Sección 6.4.3, la conversión de LP a ISP se describe en la Sección 6.4.4 y la conversión de ISP a LP se describe en la Sección 6.4.5.

[0008] Se investigan los parámetros de codificación de voz como el retardo de libro de códigos adaptativo y la ganancia, el índice de libro de código algebraico y la ganancia reduciendo al mínimo el error entre la señal de entrada y la señal sintetizada en el dominio ponderado perceptualmente. La ponderación perceptual se realiza filtrando la señal mediante un filtro de ponderación perceptual obtenido de los coeficientes del filtro LP. La señal ponderada en forma perceptual se usa también en el análisis de altura de tono en bucle abierto.

[0009] El codificador G.718 es un codificador de voz puro que tiene únicamente modo de codificación de solo voz. Por tanto, el codificador G.718 no es un codificador conmutado y, de este modo, este codificador no es ventajoso porque tiene únicamente un modo de codificación de solo voz dentro de la capa de núcleo. De este modo, se producen los problemas de calidad cuando este codificador se aplica a señales que no sean señales de voz, es decir, a señales de audio generales, para las cuales el modelo subyacente a la codificación CELP no es apropiado.

[0010] Un códec conmutado adicional es el denominado códec USAC, es decir, el códec de audio de voz unificada tal como se define en ISO/IEC CD 23003-3 con fecha del 24 de septiembre de 2010. La ventana de análisis LPC usada para este códec conmutado se indica en la fig. 5d en 516. Nuevamente, se supone una trama actual que se extiende entre 0 y 20 ms y, por tanto, se muestra que la parte de anticipación 618 de este códec es 20 ms, es decir, es significativamente mayor que la parte de anticipación de G.718. De este modo, aunque el codificador USAC proporciona una buena calidad de audio debido a su naturaleza conmutada, el retardo es

considerable debido a la parte LPC de la ventana de análisis de anticipación 518 en la fig. 5d. La estructura general de USAC es la siguiente. Primero, existe un pre/posprocesamiento común que consiste en una unidad funcional de sonido envolvente MPEG (MPEGS) para gestionar procesamientos estéreo o de multicanales y una unidad mejorada SBR (eSBR) que gestiona la representación paramétrica de las frecuencias de audio mayores en la señal de entrada. A continuación, existen dos ramas, una consistente en una trayectoria de herramienta de codificación de audio avanzada (AAC, advanced audio coding) y la otra consistente en una trayectoria basada en codificación de predicción lineal (dominio LP o LPC), que a su vez presenta una representación en el dominio de frecuencias o una representación en el dominio de tiempo de la LPC residual. Todos los espectros transmitidos para las dos, AAC y LPC, se representan en el dominio MDCT que sigue al esquema de cuantificación y codificación aritmética. La representación en el dominio de tiempo usa un esquema de codificación por excitación ACELP. La herramienta ACELP proporciona una manera de representar de manera eficiente una señal de excitación en el dominio de tiempo combinando un predictor a largo plazo (palabra código adaptativa) con una secuencia de tipo pulso (palabra código de innovación). La excitación reconstruida se envía a través de un filtro de síntesis LP para formar una señal en el dominio de tiempo. La entrada a la herramienta ACELP comprende índices del libro de código adaptativo y de innovación, valores de ganancia de códigos adaptativos y de innovación, otros datos de control y coeficientes de filtro LPC inversamente cuantificados e interpolados. La salida de la herramienta ACELP es la señal de audio reconstruida en el dominio de tiempo.

[0011] La herramienta de decodificación TCX basada en MDCT se usa para cambiar la representación residual LP ponderada de un dominio MDCT de nuevo en la señal en el dominio de tiempo y envía la señal en el dominio de tiempo ponderada incluyendo el filtro de síntesis LP ponderado. La IMDCT puede configurarse de manera que soporte 256, 512 o 1.024 coeficientes espectrales. La entrada a la herramienta TCX comprende los espectros MDCT (inversamente cuantificados), y los coeficientes de filtro LPC inversamente cuantificados e interpolados. El envío de la herramienta TCX se encuentra en la señal de audio reconstruida en el dominio de tiempo.

[0012] La fig. 6 ilustra una situación en USAC, donde se observan las ventanas de análisis LPC 516 para la trama actual y 520 para la trama pasada o última, y donde, además, se ilustra una ventana TCX 522. La ventana TCX 522 está centrada en el centro de la trama actual que se extiende entre 0 y 20 ms y se extiende 10 ms en la trama pasada y 10 ms en la trama futura que se extiende entre 20 y 40 ms. De este modo, la ventana de análisis LPC 516 requiere una parte LPC de anticipación de entre 20 y 40 ms, es decir, 20 ms, mientras que la ventana de análisis TCX tiene además una parte de anticipación que se extiende entre 20 y 30 ms en la trama futura. Esto significa que el retardo introducido por la ventana de análisis USAC 516 es 20 ms, mientras que el retardo introducido en el codificador por la ventana TCX es 10 ms. De este modo, resulta claro que las partes de anticipación de los dos tipos de ventanas no están alineadas entre sí. Por tanto, aunque la ventana TCX 522 sólo introduce un retardo de 10 ms, el retardo completo del codificador es sin embargo de 20 ms debido a la ventana de análisis LPC 516. Por tanto, aunque existe una pequeña parte de anticipación para la ventana TCX, no reduce el retardo algorítmico total del codificador, ya que el retardo total se determina por la contribución más alta, es decir, es igual a 20 ms debido a la ventana de análisis LPC 516 que se extiende 20 ms en la trama futura, es decir, no sólo cubre la trama actual sino que además cubre la trama futura.

[0013] La publicación de la técnica anterior "Codificación de voz/audio universal que usa técnicas ACELP/TCX híbridas" B. Bessette y col., ICASSP 2005, páginas III-301 a III-304 describe un algoritmo de codificación de audio híbrido que integra una técnica de codificación basada en LP y una técnica de codificación por transformada más general. Los modos ACELP y TCX están integrados en el sentido de que los dos se basan en análisis LP y codificación de excitación. En ACELP, la excitación se codifica usando un libro de códigos disperso en el dominio de excitación, mientras que en TCX el libro de códigos está en el dominio objeto, o de señal ponderada. El análisis LP se realiza cada 20 ms, usando una ventana de medio seno situada en la parte central de la primera subtrama de 5 ms en la siguiente trama. Son posibles una trama TCX con 20 ms, una trama TCX con 40 ms o una TCX con 80 ms de longitud, en las que la duración de superposición en la parte derecha de la ventana que corresponde a una anticipación en la siguiente trama de la trama de 80 ms es igual a 128 muestras que corresponden a una duración de 10 ms a la vista de la velocidad de muestreo interna de 12,8 kHz en AMR-WB. Además, el documento WO 2012/004349 A1 se refiere a un códec que soporta un modo de codificación de transformada de cancelación de solapamiento en el dominio de tiempo y un modo de codificación en el dominio de tiempo así como cancelación de solapamiento directo para conmutación entre los dos modos.

[0014] Un objetivo de la presente invención es proporcionar un concepto de codificación mejorado para la codificación o decodificación de audio que, por un lado, proporcione una buena calidad de audio y, por otro lado, produzca un retardo reducido. Este objetivo se logra con un decodificador de audio según la reivindicación 1, un procedimiento de decodificación de audio según la reivindicación 8 o un programa informático según la reivindicación 9. Según la presente invención, se aplica un esquema de códec de audio conmutado que tiene una rama de codificación por transformada y una rama de codificación por predicción. De manera importante, los dos tipos de ventanas, es decir, la ventana de análisis de codificación por predicción por un lado y la ventana de análisis de codificación por transformada por otro lado están alineados con respecto a su parte de anticipación para que la parte de anticipación de codificación por transformada y la parte de anticipación de codificación por predicción sean

idénticas o sean diferentes entre sí en menos del 20 % de la parte de anticipación de codificación por predicción o en menos del 20 % de la parte de anticipación de codificación por transformada. Debe observarse que la ventana de análisis por predicción se usa no sólo en la rama de codificación por predicción, sino que en realidad se usa en las dos ramas. El análisis LPC se usa además para modelar el ruido en el dominio de transformada. Por tanto, dicho de otro modo, las partes de anticipación son idénticas o muy cercanas entre sí. Se asegura así que se obtenga un compromiso óptimo y que las características de calidad de audio o de retardo no se establezcan de manera subóptima. De este modo, para la codificación por predicción en la ventana de análisis se ha descubierto que el análisis LPC es mejor cuanto mayor es la parte de anticipación, pero, por otro lado, con una parte mayor de anticipación el retardo aumenta. Por otro lado, sucede lo mismo para la ventana TCX. Cuanto mayor es la parte de anticipación de la ventana TCX, mejor se reduce la velocidad binaria de TCX, ya que las ventanas TCX más largas producen en general menores velocidades binarias. Por tanto, a diferencia de la presente invención, las partes de anticipación son idénticas o muy cercanas entre sí y, en particular, con una diferencia entre sí de menos del 20 %. Por tanto, la parte de anticipación, que no se desea por motivos de retardo, por otro lado es usada de manera óptima por las dos ramas de codificación/decodificación.

A la vista de ello, la presente invención proporciona un concepto de codificación mejorado con, por un lado, un bajo retardo cuando a la parte de anticipación para las dos ventanas de análisis se le asigna un valor bajo y, por otro lado, proporciona un concepto de codificación/decodificación con buenas características debido al hecho de que el retardo que debe introducirse por motivos de calidad de audio o por motivos de velocidad binaria es utilizado de manera óptima por las dos ramas de codificación y no solo por una única rama de codificación.

[0015] Las realizaciones preferidas de la presente invención se describen a continuación con respecto a los dibujos adjuntos, en los que:

La fig. 1a ilustra un diagrama de bloques de un codificador de audio conmutado preferido;
La fig. 1b ilustra un diagrama de bloques de un decodificador conmutado correspondiente;
La fig. 1c ilustra más detalles en el decodificador de parámetros por transformado ilustrado en la fig. 1b;
La fig. 1d ilustra más detalles en el modo de codificación por transformada del decodificador de la fig. 1a;
La fig. 2a ilustra una realización preferida para el dispositivo de división en ventanas aplicado en el codificador para análisis LPC por un lado y análisis de codificación por transformada por otro lado, y es una representación de la ventana de síntesis usada en el decodificador de codificación por transformada de la fig. 1b;
La fig. 2b ilustra una secuencia de ventana de ventanas de análisis LPC y ventanas TCX alineadas para un intervalo de tiempo de más de dos tramas;
La fig. 2c ilustra una situación para una transición de TCX a ACELP y una ventana de transición para una transición de ACELP a TCX;
La fig. 3a ilustra más detalles del codificador de la fig. 1a;
La fig. 3b ilustra un procedimiento de análisis-síntesis para decidir un modo de codificación para una trama;
La fig. 3c ilustra otra realización para decidir entre los modos para cada trama;
La fig. 4a ilustra el cálculo y el uso de datos LPC obtenidos usando dos ventanas de análisis LPC diferentes para una trama actual;
La fig. 4b ilustra el uso de datos LPC obtenidos por división en ventanas usando una ventana de análisis LPC para la rama TCX del codificador;
La fig. 5a ilustra ventanas de análisis LPC para AMR-WB;
La fig. 5d ilustra ventanas simétricas para AMR-WB+ con fines de análisis LPC;
La fig. 5c ilustra ventanas de análisis LPC para un codificador G.718;
La fig. 5d ilustra ventanas de análisis LPC tal como se usan en USAC; y
La fig. 6 ilustra una ventana TCX para una trama actual con respecto a una ventana de análisis LPC para la trama actual.

[0016] La fig. 1a ilustra un aparato para codificar una señal de audio con un flujo de muestras de audio. Las muestras de audio o datos de audio entran en el codificador en 100. Los datos de audio se introducen en un dispositivo de división en ventanas 102 para aplicar una ventana de análisis de codificación por predicción al flujo de muestras de audio y obtener datos de división en ventanas para un análisis por predicción. El dispositivo de división en ventanas 102 se configura además para aplicar una ventana de análisis de codificación por transformada al flujo de muestras de audio con el fin de obtener datos divididos en ventanas para un análisis por transformada. Dependiendo de la implementación, la ventana LPC no se aplica de forma directa a la señal original sino a una señal "preacentuada" (como en AMR-WB, AMR-WB+, G718 y USAC). Por otro lado la ventana TCX se aplica en la señal original directamente (como en USAC). Sin embargo, las dos ventanas pueden aplicarse a la misma señal o la ventana TCX puede aplicarse también a una señal de audio procesada obtenida de la señal original como, por ejemplo, por preacentuación u otra ponderación usada para mejorar la calidad o eficiencia de compresión.

[0017] La ventana de análisis de codificación por transformada se asocia con muestras de audio en una trama actual de muestras de audio y con muestras de audio de una parte predefinida de la trama futura de muestras de audio que es una parte de anticipación de codificación por transformada.

[0018] Además, la ventana de análisis de codificación por predicción se asocia con al menos una parte de muestras de audio de la trama actual y con muestras de audio de una parte predefinida de la trama futura que es una parte de anticipación de codificación por predicción.

5 **[0019]** Tal como se representa en el bloque 102, la parte de anticipación de codificación por transformada y la parte de anticipación de codificación por predicción están alineadas entre sí, lo que significa que estas partes son idénticas o muy cercanas entre sí, por ejemplo diferentes entre sí en menos del 20 % de la parte de anticipación de codificación por predicción o en menos del 20 % de la parte de anticipación de codificación por transformada. Preferentemente, las partes de anticipación son idénticas o diferentes entre sí en menos del 5 % de la parte de
10 anticipación de codificación por predicción o en menos del 5 % de la parte de anticipación de codificación por transformada.

[0020] El codificador comprende además un procesador de codificación 104 para generar datos codificados por predicción para la trama actual usando los datos de división en ventanas para el análisis por predicción o para
15 generar datos codificados por transformada para la trama actual usando los datos de división en ventanas para el análisis por transformada.

[0021] Además, el codificador comprende preferentemente una interfaz de salida 106 para recibir, para una trama actual y, en realidad, para cada trama, datos LPC 108a y datos codificados por transformada (tales como
20 datos TCX) o datos codificados por predicción (datos ACELP) en la línea 108b. El procesador de codificación 104 proporciona estas dos clases de datos y recibe, como entrada, datos divididos en ventanas para un análisis por predicción indicado en 110a y datos divididos en ventanas para un análisis por transformada indicado en 110b. Además, el aparato para codificar comprende un selector de modo de codificación o controlador 112 que recibe, como entrada, los datos de audio 100 y proporciona, como salida, datos de control para el procesador de
25 codificación 104 por las líneas de control 114a, o datos de control para la interfaz de salida 106 por la línea de control 114b.

[0022] La fig. 3a proporciona otros detalles en el procesador de codificación 104 y el dispositivo de división en ventanas 102. El dispositivo de división en ventanas 102 comprende preferentemente, como primer módulo, el LPC
30 o dispositivo de división en ventanas de análisis de codificación por predicción 102a y, como segundo componente o módulo, el dispositivo de división en ventanas de codificación por transformada (tal como el dispositivo de división en ventanas TCX) 102b. Tal como indica la flecha 300, la ventana de análisis LPC y la ventana TCX están alineadas entre sí de manera que las partes de anticipación de las dos ventanas sean idénticas entre sí, lo que significa que las dos partes de anticipación se extienden hasta el mismo instante de tiempo en una trama futura. La rama superior
35 en la fig. 3a del dispositivo de división en ventanas LPC 102a hacia delante a la derecha es una rama de codificación por predicción con un analizador LPC e interpolador 302, un filtro de ponderación perceptual o bloque de ponderación 304 y un calculador de parámetro de codificación por predicción 306 tal como un calculador de parámetros ACELP. Los datos de audio 100 se proporcionan al aparato LPC para la división en ventanas 102a y el bloque de ponderación perceptual 304. Además, los datos de audio se proporcionan al dispositivo de división en
40 ventanas TCX, y la rama inferior de la salida del dispositivo de división en ventanas TCX a la derecha constituye una rama de codificación por transformada. Esta rama de codificación por transformada comprende un bloque de conversión en el dominio de tiempo-frecuencia 310, un bloque de ponderación espectral 312 y un bloque de codificación de procesamiento/cuantificación 314. El bloque de conversión en el dominio de tiempo-frecuencia 310 se implementa preferentemente como transformada de introducción de solapamiento tal como MDCT, MDST o
45 cualquier otra transformada que tenga un número de valores de entrada mayor que el número de valores de salida. La conversión de tiempo-frecuencia tiene, como entrada, los datos divididos en ventanas enviados por TCX o, dicho de forma general, un dispositivo de división en ventanas de codificación por transformada 102b.

[0023] Aunque la fig. 3a indica, para la rama de codificación por predicción, un procesamiento LPC con un
50 algoritmo de codificación ACELP, pueden aplicarse también otros codificadores de predicción como CELP o cualquier otro codificador en el dominio de tiempo conocidos en la técnica, si bien se prefiere el algoritmo ACELP debido por un lado a su calidad y por otro lado a su eficiencia.

[0024] Además, para la rama de codificación por transformada, se prefiere un procesamiento MDCT en
55 particular en el bloque de conversión de tiempo-frecuencia 310, aunque se pueden aplicar asimismo otras transformadas en el dominio espectral.

[0025] Además, la fig. 3a ilustra una ponderación espectral 312 para transformar los valores espectrales del
60 bloque 310 en un dominio LPC. Esta ponderación espectral 312 se realiza con datos de ponderación obtenidos de los datos de análisis LPC generados por el bloque 302 en la rama de codificación por predicción. Alternativamente, sin embargo, la transformada desde el dominio de tiempo al dominio LPC podría realizarse también en el dominio de tiempo. En este caso, se coloca un filtro de análisis LPC antes del dispositivo de división en ventanas TCX 102b para calcular los datos en el dominio de tiempo residuales por predicción. Sin embargo, se ha descubierto que la transformada desde el dominio de tiempo al dominio LPC se realiza preferentemente en el dominio espectral por
65 ponderación espectral de datos codificados por transformada usando datos de análisis LPC transformados de datos

LPC en factores de ponderación correspondientes en el dominio espectral como dominio MDCT.

[0026] La fig. 3b ilustra la visión general para ilustrar una determinación de análisis-síntesis o de “bucle cerrado” del modo de codificación para cada trama. Para tal fin, el codificador ilustrado en la fig. 3c comprende un codificador de codificación de transformada completa y un decodificador correspondiente indicado en 104b y, además, comprende un codificador de codificación de predicción completa y un decodificador correspondiente indicado en 104a en la fig. 3c. Los dos bloques 104a, 104b reciben, como entrada, los datos de audio y realizan una operación completa de codificación/decodificación. A continuación se comparan los resultados de la operación de codificación/decodificación para las dos ramas de codificación 104a, 104b con la señal original y se determina la medición de calidad para determinar qué modo de codificación tiene mejor calidad. La medición de calidad puede ser un valor de SNR segmentado o de SNR segmentado promedio como se describe, por ejemplo, en la Sección 5.2.3 de 3GPP TS 26.290. Sin embargo, pueden aplicarse otras mediciones de calidad que tengan en cuenta la comparación del resultado de codificación/decodificación con la señal original.

[0027] Basándose en la medición de calidad de cada rama 104a, 104b para el elemento de decisión 112, el elemento de decisión decide si la trama actual examinada debe ser codificada usando ACELP o TCX. Con posterioridad a la decisión, existen varias formas de realizar la selección del modo de codificación. Una forma consiste en que el elemento de decisión 112 controla los bloques de codificación/decodificación 104a, 104b correspondientes, para simplemente enviar el resultado de codificación para la trama actual a la interfaz de salida 106, de manera que se asegura que para cierta trama, sólo se transmite un único resultado de codificación en la señal codificada de salida en 107.

[0028] Alternativamente, los dos dispositivos 104a, 104b podrían remitir su resultado de codificación a la interfaz de salida 106, y los dos resultados se almacenan en la interfaz de salida 106 hasta que el elemento de decisión controle la interfaz de salida a través de la línea 105 para enviar el resultado desde el bloque 104b o desde el bloque 104a.

[0029] La fig. 3b ilustra más detalles sobre el concepto de la fig. 3c. En particular, el bloque 104a comprende un codificador ACELP completo y un decodificador ACELP completo y un comparador 112a. El comparador 112a proporciona una medición de calidad al comparador 112c. Lo mismo se aplica al comparador 112b, que tiene una medición de calidad debida a la comparación de la señal codificada TCX y decodificada de nuevo con la señal de audio original. Posteriormente, los dos comparadores 112a, 112b proporcionan sus mediciones de calidad al comparador final 112c. Dependiendo de qué medición de calidad sea mejor, el comparador se decide por una decisión CELP o TCX. La decisión puede refinarse introduciendo factores adicionales en la decisión.

[0030] Alternativamente, se puede llevar a cabo un modo en bucle abierto para determinar el modo de codificación para una trama actual basándose en el análisis de señal de los datos de audio para la trama actual. En este caso, el elemento de decisión 112 de la fig. 3c realiza un análisis de señal de datos de audio para la trama actual y controla un codificador ACELP o codificador TCX para codificar realmente la trama de audio actual. En esta situación, el codificador no necesita una decodificación completa, sino que sería suficiente una implementación de las etapas de codificación dentro del codificador. La clasificación de señal en bucle abierto y las decisiones de señal se describen, por ejemplo, en AMR-WB+ (3GPP TS 26.290).

[0031] La fig. 2a ilustra una implementación preferida del dispositivo de división en ventanas 102 y, en particular, las ventanas suministradas por el dispositivo de división en ventanas.

[0032] Preferentemente, la ventana de análisis de codificación por predicción para la trama actual está centrada en el centro de una cuarta subtrama y esta ventana se indica en 200. Además, se prefiere usar otra ventana de análisis LPC, es decir, una ventana de análisis LPC de mitad de trama indicada en 202 y centrada en el centro de la segunda subtrama de la trama actual. Además, la ventana de codificación por transformada como, por ejemplo, la ventana MDCT 204 se coloca con respecto a las dos ventanas de análisis LPC 200, 202 tal como se ilustra. En particular, la parte de anticipación 206 de la ventana de análisis tiene la misma longitud en tiempo de la parte de anticipación 208 de la ventana de análisis de codificación por predicción. Las dos partes de anticipación se extienden 10 ms en la trama futura. Además, se prefiere que la ventana de análisis de codificación por transformada no sólo tenga la parte de superposición 206, sino que tenga una parte de no superposición entre 10 y 20 ms 208 y la primera parte de superposición 210. Las partes de superposición 206 y 210 son tales que un dispositivo para la superposición-suma en un decodificador realiza un proceso de superposición-suma en la parte de superposición, pero no se necesita un procedimiento de superposición-suma para la parte de no superposición.

[0033] Preferentemente, la primera parte de superposición 210 comienza al principio de la trama, es decir, a cero ms y se extiende hasta el centro de la trama, es decir, 10 ms. Además, la parte de no superposición se extiende desde el final de la primera parte de la trama 210 hasta el final de la trama en 20 ms de manera que la segunda parte de superposición 206 coincida totalmente con la parte de anticipación. Esto tiene ventajas debido a la conmutación de un modo a otro. Desde un punto de vista del rendimiento de TCX, sería mejor usar una ventana de seno con superposición total (20 ms de superposición, como en USAC). Sin embargo, se necesitaría entonces una

tecnología de cancelación de solapamiento directo para las transiciones entre TCX y ACELP. La cancelación de solapamiento directo se usa en USAC para cancelar el solapamiento introducido por las tramas TCX siguientes perdidas (reemplazadas por ACELP). La cancelación de solapamiento directo requiere una cantidad importante de bits y por tanto no es adecuada para una velocidad binaria constante y, en particular, un códec de baja velocidad binaria como realización preferida del códec que se describe. Por tanto, según las realizaciones de la invención, en vez de usar FAC, la superposición de ventana TCX se reduce y la ventana cambia hacia el futuro de manera que toda la parte de superposición 206 se coloca en la trama futura. Además, la ventana ilustrada en la fig. 2a para la codificación por transformada tiene sin embargo una máxima superposición para recibir una perfecta reconstrucción en la trama actual, cuando la trama siguiente es ACELP y sin usar cancelación de solapamiento directo. Esta superposición máxima se fija preferentemente en 10 ms que es la parte de anticipación disponible en el tiempo, es decir, 10 ms como se observa claramente en la fig. 2a.

[0034] Aunque la fig. 2a se ha descrito con respecto a un codificador, donde la ventana 204 para codificación por transformada es una ventana de análisis, se observa que la ventana 204 representa además una ventana de síntesis para la decodificación por transformada. En una realización preferida, la ventana de análisis es idéntica a la ventana de síntesis, y las dos ventanas son simétricas en sí. Esto significa que las dos ventanas son simétricas en una línea central (horizontal). En otras aplicaciones, sin embargo, pueden usarse ventanas no simétricas, en las que la ventana de análisis tiene una forma diferente que la ventana de síntesis.

[0035] La fig. 2b ilustra una secuencia de ventanas sobre una parte de una trama pasada, una trama actual siguiente posteriormente, una trama futura que sigue a la trama actual y la siguiente trama futura que sigue a la trama futura.

[0036] Está claro que la parte de superposición-suma procesada por un procesador de superposición-suma ilustrado en 250 se extiende desde el inicio de cada trama hasta la mitad de cada trama, es decir, entre 20 y 30 ms para calcular los datos de la trama futura y entre 40 y 50 ms para calcular datos TCX para la trama futura siguiente o entre cero y 10 ms para calcular datos para la trama actual. Sin embargo, para calcular los datos en la segunda mitad de cada trama no se necesitan técnicas de superposición-suma y por tanto tampoco de cancelación de solapamiento directo. Este hecho se debe a que la ventana de síntesis tiene una parte de no superposición en la segunda mitad de cada trama.

[0037] Normalmente, la longitud de una ventana MDCT es dos veces la longitud de una trama. Así sucede también en la presente invención. Cuando, nuevamente, se considera la fig. 2a, sin embargo, está claro que la ventana de análisis/síntesis sólo se extiende de cero a 30 ms, pero la longitud completa de la ventana es 40 ms. Esta longitud completa es importante para proporcionar datos de entrada para la operación de plegado o no plegado correspondiente del cálculo MDCT. Con el fin de extender la ventana a una longitud total de 14 ms, se añaden 5 ms de valores cero entre -5 y 0 ms y se añaden 5 segundos de valores cero MDCT al final de la trama entre 30 y 35 ms. Sin embargo, estas partes adicionales que tienen sólo ceros no desempeñan ningún papel en cuanto a consideraciones de retardo, ya que para el codificador o decodificador se sabe que los últimos cinco ms de la ventana y los primeros cinco ms de la ventana son ceros, para que estos datos se encuentren ya presentes sin ningún retardo.

[0038] La fig. 2c ilustra las dos posibles transiciones. Sin embargo, para una transición de TCX a ACELP no se necesita un cuidado especial ya que cuando se supone con respecto a la fig. 2a que la trama futura es una trama ACELP, los datos del decodificador TCX de la última trama para la parte de anticipación 206 pueden simplemente eliminarse ya que la trama ACELP comienza inmediatamente al principio de la trama futura y, por tanto, no existe un hueco de datos. Los datos de ACELP son autoconsistentes y por tanto, un decodificador, cuando tiene una conmutación de TCX a ACELP, usa los datos calculados de TCX para la trama actual, descarta los datos obtenidos por el procesamiento de TCX para la trama futura y, en su lugar, usa los datos de la trama futura de la trama ACELP.

[0039] Sin embargo, cuando se realiza una transición de ACELP a TCX, se usa una ventana especial de transición ilustrada en la fig. 2c. Esta ventana comienza al principio de la trama de cero a 1, y tiene una parte de no superposición 220 y una parte de superposición al final indicada en 222 idéntica a la parte de superposición 206 de una ventana MDCT directa.

[0040] Además, esta ventana se rellena con ceros entre -12,5 ms y cero al principio de la ventana y entre 30 y 35,5 ms al final, es decir, después de la parte de anticipación 222. Se obtiene así una longitud de transformada aumentada. La longitud es 50 ms, pero la longitud de la ventana de análisis/síntesis directa tiene sólo 40 ms. Sin embargo, esto no disminuye la eficiencia ni aumenta la velocidad binaria, y esta transformada más larga es necesaria cuando se realiza una conmutación de ACELP a TCX. La ventana de transición usada en el decodificador correspondiente es idéntica a la ventana ilustrada en la fig. 2c.

[0041] A continuación se analiza con mayor detalle el decodificador. La fig. 1b ilustra un decodificador de audio para decodificar una señal de audio codificada. El decodificador de audio comprende un decodificador de parámetros por predicción 180, donde el decodificador de parámetros por predicción se configura para realizar la

decodificación de datos para una trama codificada por predicción desde la señal de audio codificada recibida en 181 e introducida en la interfaz 182. El decodificador comprende además un decodificador de parámetros por transformada 183 para realizar una decodificación de datos para una trama codificada por transformada a partir de la señal de audio codificada en la línea 181. El decodificador de parámetros por transformada se configura para 5 realizar, preferentemente, una transformada en tiempo espectral afectada por solapamiento y para aplicar una ventana de síntesis a los datos transformados con el fin de obtener datos para la trama actual y la trama futura. La ventana de síntesis tiene una primera parte de superposición, una segunda parte adyacente de no superposición y una tercera parte adyacente de superposición tal como se ilustra en la fig. 2a, en las que la tercera parte adyacente de superposición sólo se asocia a muestras de audio para la trama futura y la parte de no superposición sólo se 10 asocia a datos de la trama actual. Además, se proporciona un aparato para la superposición-suma 184 para superponer y agregar muestras de ventana de síntesis asociadas con la tercera parte de superposición de una ventana de síntesis para la trama actual y ventana de síntesis en las muestras asociadas con la primera parte de superposición de una ventana de síntesis para la trama futura y obtener una primera parte de muestras de audio para la trama futura. El resto de las muestras de audio para la trama futura son muestras sometidas a división en 15 ventana de síntesis asociadas con la segunda parte de no superposición de la ventana de síntesis para la trama futura obtenidas sin aparato para la superposición-suma cuando la trama actual y la trama futura comprenden datos codificados por transformada. Sin embargo, cuando se realiza un cambio de una trama a la trama siguiente, resulta útil un combinador 185 para realizar una buena conmutación de un modo de codificación al otro modo de codificación para obtener finalmente los datos de audio decodificados en la salida del combinador 185.

20 **[0042]** La fig. 1c ilustra más detalles de la construcción del decodificador de parámetros por transformada 183.

[0043] El decodificador comprende una fase de procesamiento de decodificación 183a configurada para 25 realizar todos los procesos necesarios para decodificar datos espectrales codificados tales como decodificación aritmética o decodificación de Huffman o generalmente, decodificación por entropía y una posterior descuantificación, relleno de ruido, etc., con el fin de obtener valores espectrales decodificados a la salida del bloque 183. Estos valores espectrales se introducen en un ponderador espectral 183b. El ponderador espectral 183b recibe los datos ponderados espectrales del calculador de datos ponderados LPC 183c, que son suministrados por datos 30 LPC generados a partir del bloque de análisis por predicción del lado del codificador y recibidos, en el decodificador, por la interfaz de entrada 182. A continuación se realiza una transformada espectral inversa que comprende, como primera fase, preferentemente una transformada inversa DCT-IV 183d y un procesamiento posterior de no plegado y ventana de síntesis 183e, antes de que los datos para la trama futura, por ejemplo, sean sometidos al dispositivo de superposición-suma 184. El dispositivo de superposición-suma realiza la operación de superposición-suma cuando 35 los datos para la trama futura siguiente están disponibles. Los bloques 183d y 183e en conjunto constituyen la transformada de tiempo espectral o, en la realización de la fig. 1c, una transformada inversa MDCT (MDCT^{-1}) preferida.

[0044] En particular, el bloque 183d recibe datos para una trama de 20 ms, y aumenta el volumen de datos 40 en la etapa de no plegado del bloque 183e en datos para 40 ms, es decir, el doble de datos de antes y, posteriormente, se aplica la ventana de síntesis que tiene una longitud de 40 ms (cuando las partes de ceros al inicio y al final de la ventana se suman conjuntamente) a estos 40 ms de datos. A continuación, a la salida del bloque 183e, los datos para el bloque actual y los datos dentro de la parte de anticipación para el bloque futuro se encuentran disponibles.

45 **[0045]** La fig. 1d ilustra el procesamiento correspondiente en el lado del codificador. Las características que se analizan en el contexto de la fig. 1d se implementan en el procesador de codificación 104 o por medio de bloques correspondientes en la fig. 3a. La conversión tiempo-frecuencia 310 en la fig. 3a preferentemente se implementa como MDCT y comprende una fase de división en ventanas y plegado 310a, donde la operación de división en 50 ventanas en el bloque 310a se implementa por medio del dispositivo de división en ventanas TCX 103d. De este modo, la primera operación real en el bloque 310 en la fig. 3a es la operación de plegado para llevar de nuevo 40 ms de datos de entrada a 20 ms de datos de la trama. A continuación, con los datos plegados que han recibido contribuciones de solapamiento, se realiza un DCT-IV tal como se ilustra en el bloque 310d. El bloque 302 (análisis LPC) proporciona datos LPC obtenidos del análisis usando la ventana LPC de final de trama a un bloque 302b (LPC a MDCT), y el bloque 302d genera factores de ponderación para realizar ponderaciones espectrales mediante un 55 ponderador espectral 312. Preferentemente, 16 coeficientes LPC para una trama de 20 ms en el modo de codificación TCX se transforman en 16 factores de ponderación en el dominio MDCT, usando preferentemente oDFT (odd Discrete Fourier Transform, transformada de Fourier discreta impar). Para otros modos, como los modos NB con velocidad de muestreo de 8 kHz, el número de coeficientes LPC puede ser inferior a 10. Para otros modos con 60 mayor velocidad de muestreo puede haber más de 16 coeficientes LPC. El resultado de esta oDFT son 16 valores de ponderación, y cada valor de ponderación se asocia a una banda de datos espectrales obtenidos por el bloque 310b. La ponderación espectral tiene lugar al dividir todos los valores espectrales MDCT para una banda por el mismo valor de ponderación asociado con esta banda de manera que se realice de forma muy eficiente esta operación de ponderación espectral en el bloque 312. De este modo, se dividen 16 bandas de valores MDCT cada 65 una por el factor de ponderación correspondiente para enviar los valores espectrales ponderados en forma espectral

que se procesan en el bloque 314 tal como se conoce en la técnica, es decir, por ejemplo, por cuantificación y codificación por entropía.

[0046] Por otra parte, en el lado del decodificador, la ponderación espectral correspondiente al bloque 312 en la fig. 1d será una multiplicación realizada por el ponderador espectral 183b ilustrado en la fig. 1c.

[0047] A continuación se analizan la fig. 4a y la fig. 4b para delinear el modo en que se usan los datos LPC generados por la ventana de análisis LPC o generados por las dos ventanas de análisis LPC ilustradas en la fig. 2 en el modo ACELP o en el modo TCX/MDCT.

10

[0048] Posteriormente a la aplicación de la ventana de análisis LPC, la computación de autocorrelación se realiza con los datos LPC divididos en ventanas. A continuación se aplica el algoritmo de Levinson-Durbin en la función de autocorrelación. Seguidamente, los 16 coeficientes LP para cada análisis LP, es decir, 16 coeficientes para la ventana de mitad de trama y 16 coeficientes para la ventana de final de trama se convierten en valores ISP. De este modo, las etapas del cálculo de autocorrelación a la conversión ISP se realizan, por ejemplo, en el bloque 400 de la fig. 4a. A continuación, el cálculo continúa en el lado del codificador mediante cuantificación de los coeficientes ISP. Seguidamente los coeficientes ISP son nuevamente descuantificados y convertidos al dominio de coeficientes LP. De este modo, se obtienen los datos LPC, o dicho de otro modo, los 16 coeficientes LPC apenas diferentes a los coeficientes LPC obtenidos en el bloque 400 (debido a la cuantificación y recuantificación) que pueden ser usados directamente para la cuarta subtrama tal como se indica en la etapa 401. Sin embargo, para las demás subtramas se prefiere realizar varias interpolaciones tal como se establece, por ejemplo, en la sección 6.8.3 de Rec. ITU-T G.718 (06/2008). Los datos LPC para la tercera subtrama se calculan interpolando datos LPC de final de trama y mitad de trama ilustrados en el bloque 402. La interpolación preferida consiste en que cada dato correspondiente se divide por dos y se suman ambos, es decir, un promedio de los datos LPC de final de trama y mitad de trama. Para calcular los datos LPC para la segunda subtrama tal como se ilustra en el bloque 403 se realiza además una interpolación. En particular, se usa el 10 % de los valores de los datos LPC de final de trama de la última trama, el 80 % de los datos LPC de mitad de trama para la trama actual y el 10 % de los valores de datos LPC para el final de trama de la trama actual para calcular finalmente los datos LPC para la segunda subtrama.

[0049] Finalmente, los datos LPC para la primera subtrama se calculan, tal como se indica en el bloque 404, obteniendo un promedio entre los datos LPC de final de trama de la última trama y los datos LPC de mitad de trama de la trama actual.

[0050] Para realizar una codificación ACELP, los dos grupos de parámetros cuantificados LPC, es decir, del análisis de mitad de trama y el análisis de final de trama, se transmiten a un decodificador.

[0051] Basándose en los resultados para las subtramas individuales calculadas por los bloques 401 a 404, los cálculos ACELP se realizan tal como se indica en el bloque 405 con el fin de obtener los datos ACELP que se transmitirán al decodificador.

40

[0052] A continuación se describe la fig. 4b. De nuevo, en el bloque 400, se calculan datos LPC de mitad de trama y final de trama. Sin embargo, como existe el modo de codificación TCX, sólo se transmiten datos LPC de final de trama al decodificador y los datos LPC de mitad de trama no se transmiten al decodificador. En particular, no se transmiten los coeficientes LPC en sí al decodificador, sino que se transmiten los valores obtenidos después de la transformada ISP y la cuantificación. De este modo se prefiere que, como datos LPC, los valores cuantificados ISP obtenidos de datos de coeficientes LPC de final de trama se transmitan al decodificador.

[0053] Sin embargo, en el codificador se realizan los procedimientos de las etapas 406 a 408 para obtener factores de ponderación con el fin de ponderar datos espectrales MDCT de la trama actual. Para tal objeto, se interpolan los datos LPC de final de trama de la trama actual y los datos LPC de final de trama de la trama pasada. Sin embargo, se prefiere no interpolar datos de coeficientes LPC en sí porque proceden directamente del análisis LPC.

[0054] En su lugar, se prefiere interpolar los valores ISP cuantificados y de nuevo descuantificados obtenidos de los coeficientes LPC correspondientes. De este modo, los datos LPC usados en el bloque 406 así como los datos LPC usados para otros cálculos en el bloque 401 a 404 siempre son datos ISP preferentemente cuantificados y nuevamente descuantificados obtenidos de los 16 coeficientes LPC originales por ventana de análisis LPC.

[0055] La interpolación en el bloque 406 es preferentemente un promediado puro, es decir, se suman los valores correspondientes y se dividen por dos. A continuación, en el bloque 407, los datos espectrales MDCT de la trama actual se ponderan usando los datos LPC interpolados y, en el bloque 408, se realiza un procesamiento adicional de los datos espectrales ponderados para obtener finalmente los datos espectrales codificados que se transmitirán desde el codificador a un decodificador. De este modo, los procedimientos realizados en el paso 407 corresponden al bloque 312, y el procedimiento realizado en el bloque 408 en la fig. 4d corresponde al bloque 314 en la fig. 4d. Las operaciones correspondientes se realizan en realidad en el lado del decodificador. Por tanto, son

65

necesarias las mismas interpolaciones en el lado del decodificador para, por una parte, calcular los factores de ponderación espectrales o, por otra parte, calcular los coeficientes LPC para las subtramas individuales por interpolación. De este modo, la fig. 4a y fig. 4b son aplicables igualmente al lado del decodificador en lo que respecta a los procedimientos de los bloques 401 a 404 o 406 de la fig. 4b.

5

[0056] La presente invención resulta especialmente útil para implementaciones de códec de bajo retardo. Esto significa que estos códec se diseñan para tener un retardo algorítmico o sistemático inferior preferentemente a 45 ms y, en algunos casos, igual o menor incluso que 35 ms. No obstante, la parte de anticipación para el análisis LPC y TCX es necesaria para obtener una buena calidad de audio. Por tanto, se necesita un buen compromiso entre estos dos requisitos contradictorios. Se ha descubierto que el buen compromiso entre el retardo, por un lado, y la calidad, por otro, puede obtenerse mediante un codificador o decodificador de audio conmutado con una trama de 20 ms de longitud, pero se ha descubierto que los valores para tramas con longitudes entre 15 y 30 ms proporcionan también resultados aceptables. Por otro lado, se ha descubierto que una parte de anticipación de 10 ms es aceptable en términos de retardo, pero los valores entre 5 ms y 20 ms son útiles dependiendo de la aplicación correspondiente. Además, se ha descubierto que la relación entre la parte de anticipación y la longitud de trama es útil cuando el valor es 0,5, aunque también son de utilidad otros valores comprendidos entre 0,4 y 0,6. Además, aunque la invención se ha descrito con ACELP por un lado y MDCT-TCX por otro lado, también son útiles otros algoritmos que operan en el dominio de tiempo tales como CELP o cualquier otra predicción así como algoritmos para formas de onda. En lo que respecta a TCX/MDCT, se pueden aplicar también otros algoritmos de codificación en el dominio de transformada tales como MDST o cualquier otro algoritmo basado en transformada.

10

15

20

[0057] Lo mismo sucede para la implementación específica de análisis LPC y cálculo LPC. Se prefiere basarse en los procedimientos que se han descrito anteriormente, si bien pueden usarse otros procedimientos para el cálculo/interpolación y análisis siempre que se basen en una ventana de análisis LPC.

25

[0058] Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que estos aspectos representan también una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa de procedimiento o a una característica de una etapa de procedimiento. De forma análoga, los aspectos descritos en el contexto de una etapa de procedimiento representan también una descripción de un bloque correspondiente o elemento o característica de un aparato correspondiente.

30

[0059] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando 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 tienen señales de control legibles electrónicamente almacenadas en los mismos, que cooperan (o son capaces de cooperar) con un sistema informático programable de manera que se lleve a cabo el procedimiento respectivo.

35

[0060] Algunas realizaciones según la invención comprenden un soporte de datos no transitorio que tiene señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable de manera que se lleve a cabo uno de los procedimientos descritos en la presente memoria.

40

[0061] Generalmente, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse por ejemplo en un soporte legible por máquina.

45

[0062] Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en la presente memoria, almacenado en un soporte legible por máquina.

50

[0063] Dicho de otro modo, una realización del procedimiento de invención es, por tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en la presente memoria, cuando el programa informático se ejecuta en un ordenador.

55

[0064] Una realización adicional de los procedimientos de la invención es, por tanto, un soporte de datos (o un medio de almacenamiento digital o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los procedimientos descritos en la presente memoria.

[0065] Una realización adicional del procedimiento de la invención es, por tanto, un tren de datos o una secuencia de señales que representa el programa informático para realizar uno de los procedimientos descritos en la presente memoria. El tren de datos o la secuencia de señales pueden configurarse por ejemplo de manera que sean transferidos por medio de una conexión de comunicación de datos, por ejemplo, por Internet.

60

[0066] Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado o adaptado para realizar uno de los procedimientos descritos en la

65

presente memoria.

[0067] Una realización adicional comprende un ordenador con un programa informático instalado en el mismo para realizar uno de los procedimientos descritos en la presente memoria.

5

[0068] En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo, una matriz de puertas programables de campo) para realizar algunas o todas las funcionalidades de los procedimientos descritos en la presente memoria. En algunas realizaciones, una matriz de puertas programables de campo puede cooperar con un microprocesador para realizar uno de los procedimientos descritos en la presente memoria. Generalmente, los procedimientos se realizan preferentemente mediante cualquier aparato de hardware.

10

[0069] Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Debe entenderse que para los expertos en la materia serán evidentes las modificaciones y variaciones de las disposiciones y los detalles descritos en la presente memoria. Se pretende, por tanto, estar limitado únicamente por el alcance de las reivindicaciones de patente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de la presente memoria.

15

REIVINDICACIONES

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

5 un decodificador de parámetros por predicción (180) para realizar una decodificación de datos para una trama codificada por predicción a partir de la señal de audio codificada;
 un decodificador de parámetros por transformada (183) para realizar una decodificación de datos para una trama codificada por transformada a partir de la señal de audio codificada,
 10 donde el decodificador de parámetros por transformada (183) está configurado para realizar una transformada de tiempo espectral y para aplicar una ventana de síntesis (204) a los datos transformados con el fin de obtener datos para la trama actual y una trama futura, de manera que la ventana de síntesis (204) tiene una primera parte de superposición (210), una segunda parte adyacente de no superposición (208) y una tercera parte adyacente de superposición (206), estando la tercera parte de superposición (206) asociada con muestras de audio para la trama futura y estando la parte de no superposición (208) asociada con datos de la trama actual; y
 15 un dispositivo de superposición-suma (184) para superponer y agregar muestras de síntesis divididas en ventanas asociadas con la tercera parte de superposición (206) de una ventana de síntesis (204) para la trama actual y muestras de síntesis divididas en ventanas asociadas con la primera parte de superposición (210) de una ventana de síntesis (204) para la trama futura con el fin de obtener una primera parte de muestras de audio para la trama futura, donde el resto de las muestras de audio para la trama futura son muestras de síntesis divididas en ventanas asociadas con la segunda parte de no superposición (206) de la ventana de síntesis (204) para la trama futura obtenida sin superposición-suma, cuando la trama actual y la trama futura comprenden datos codificados por transformada,
 20 donde el decodificador de parámetros por transformada (183) comprende: un ponderador espectral (183b) para ponderar datos espectrales por transformada decodificados para la trama actual usando datos codificados por predicción; y un calculador de datos ponderados de codificación por predicción (183c) para calcular los datos codificados por predicción calculando una suma ponderada de datos codificados por predicción obtenidos de una trama pasada y datos codificados por predicción obtenidos de la trama actual para obtener datos codificados por predicción interpolados.

30 2. Decodificador de audio según la reivindicación 1, donde la trama actual de la señal de audio codificada comprende datos codificados por transformada y la trama futura comprende datos codificados por predicción, donde el decodificador de parámetros por transformada (183) está configurado para realizar

una división en ventanas de síntesis usando la ventana de síntesis (204) para la trama actual con el fin de obtener muestras de audio divididas en ventanas asociadas con la segunda parte de no superposición (208) de la ventana de síntesis (204), donde las muestras de audio divididas en ventanas de síntesis asociadas con la tercera parte de superposición (206) de la ventana de síntesis (204) para la trama actual se descartan, y donde se proporcionan muestras de audio para la trama futura por parte del decodificador de parámetros por predicción (180) sin datos provenientes del decodificador de parámetros por transformada (183).

3. Decodificador de audio según la reivindicación 1 o 2,

40 donde la trama actual comprende datos codificados por predicción y la trama futura comprende datos codificados por transformada,
 45 donde el decodificador de parámetros por transformada (183) está configurado para usar una ventana de transición (220, 222) que es diferente de la ventana de síntesis (204),
 donde la ventana de transición (220, 222) comprende una primera parte de no superposición (220) al principio de la trama futura y una segunda parte de superposición (222) que empieza al final de la trama futura y que se extiende en la trama que sigue en el tiempo a la trama futura, y
 50 donde las muestras de audio para la trama futura se generan sin superposición y los datos de audio asociados con la segunda parte de superposición (222) de la ventana de transición (220, 222) para la trama futura son calculados por el dispositivo de superposición-suma (184) usando la primera parte de superposición (210) de la ventana de síntesis (204) para la trama que sigue a la trama futura.

55 4. Decodificador de audio según la reivindicación 1,

donde el calculador de datos ponderados de codificación por predicción (183c) está configurado para convertir los datos codificados por predicción en una representación espectral que tiene un valor de ponderación para cada banda de frecuencia, y
 60 donde el ponderador espectral (183b) está configurado para ponderar todos los valores espectrales en una banda mediante el mismo valor de ponderación para esta banda.

5. Decodificador de audio según cualquiera de las reivindicaciones 1 a 4, donde la ventana de síntesis (204) está configurada para tener una longitud de tiempo total menor que 50 ms y mayor que 25 ms, donde las partes de superposición primera y tercera (210, 206) tienen la misma longitud y donde la tercera parte de

superposición (206) tiene una longitud menor que 15 ms.

6. Decodificador de audio según cualquiera de las reivindicaciones 1 a 5, donde la ventana de síntesis (204) tiene una longitud de 30 ms sin partes rellenas con ceros, las partes de superposición primera y tercera (210, 206) tienen, cada una, una longitud de 10 ms y la parte de no superposición (208) tiene una longitud de 10 ms.

7. Decodificador de audio según la reivindicación 1, donde el decodificador de parámetros por transformada (183) está configurado para aplicar, para la transformada de tiempo espectral, una transformada DCT (183d) que tiene un número de muestras correspondiente a una longitud de trama, y una operación de no plegado (183e) para generar un número de valores de tiempo que es el doble del número de valores de tiempo antes de la DCT, y para aplicar (183e) la ventana de síntesis (204) a un resultado de la operación de no plegado, donde la ventana de síntesis (204) comprende, antes de la primera parte de superposición (210) y después de la tercera parte de superposición (206), partes de ceros que tienen una longitud que es la mitad de la longitud de las partes de superposición primera y tercera (210, 206)

8. Procedimiento de decodificación de una señal de audio codificada, que comprende:

la realización (180) de una decodificación de datos para una trama codificada por predicción a partir de la señal de audio codificada;
la realización (183) de una decodificación de datos para una trama codificada por transformada a partir de la señal de audio codificada,
donde la etapa de realización (183) de una decodificación de datos para una trama codificada por transformada comprende la realización de una transformada de tiempo espectral y la aplicación de una ventana de síntesis (204) a los datos transformados con el fin de obtener datos para la trama actual y una trama futura, teniendo la ventana de síntesis (204) una primera parte de superposición (210), una segunda parte adyacente de no superposición (208) y una tercera parte adyacente de superposición (206), de manera que la tercera parte de superposición (206) está asociada con muestras de audio para la trama futura y la segunda parte de no superposición (208) está asociada con datos de la trama actual; y
la superposición y suma (184) de muestras de síntesis divididas en ventanas asociadas con la tercera parte de superposición (206) de una ventana de síntesis (204) para la trama actual y muestras de síntesis divididas en ventanas asociadas con la primera parte de superposición (210) de una ventana de síntesis (204) para la trama futura con el fin de obtener una primera parte de muestras de audio para la trama futura, donde el resto de las muestras de audio para la trama futura son muestras de síntesis divididas en ventanas asociadas con la segunda parte de no superposición (208) de la ventana de síntesis (204) para la trama futura obtenida sin superposición-suma, cuando la trama actual y la trama futura comprenden datos codificados por transformada, donde la realización (183) de una decodificación de datos para una trama codificada por transformada comprende: la ponderación de datos espectrales por transformada decodificados para la trama actual usando datos codificados por predicción; y el cálculo de los datos codificados por predicción calculando una suma ponderada de datos codificados por predicción obtenidos de una trama pasada y datos codificados por predicción obtenidos de la trama actual con el fin de obtener datos codificados por predicción interpolados.

9. Programa informático que tiene un código de programa para realizar, cuando se ejecuta en un ordenador, el procedimiento de decodificación de una señal de audio codificada según la reivindicación 8.

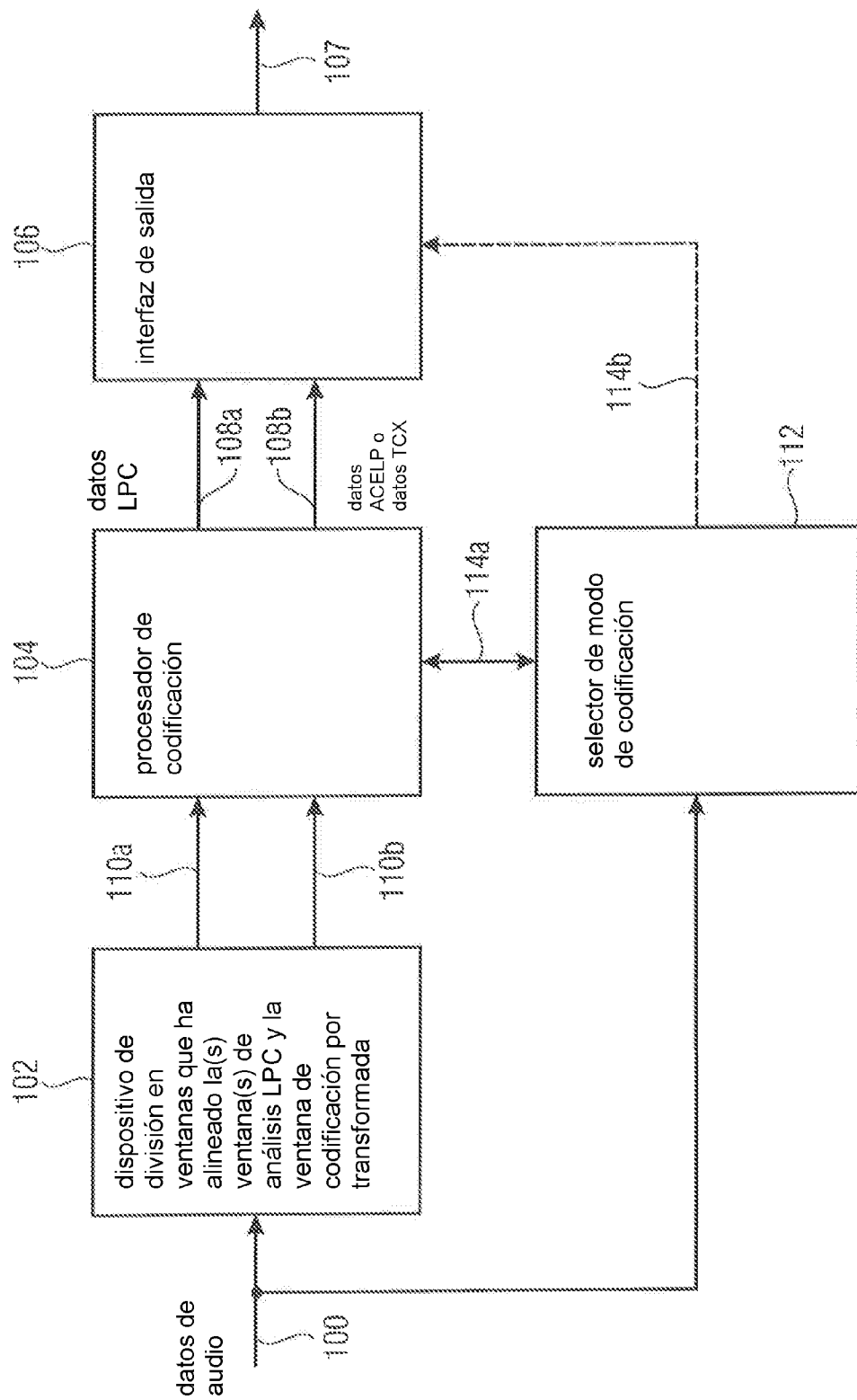


FIG 1A

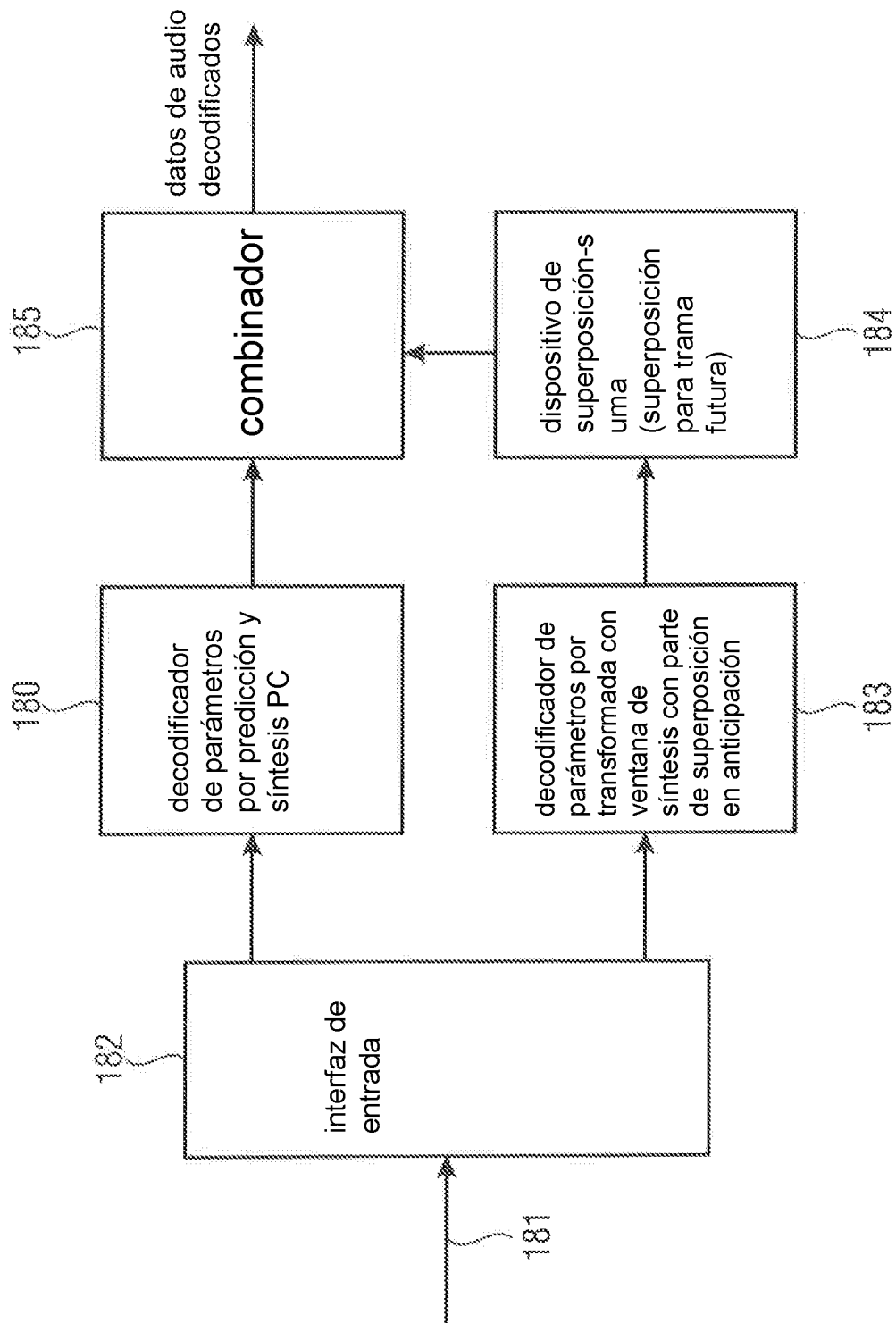


FIG 1B

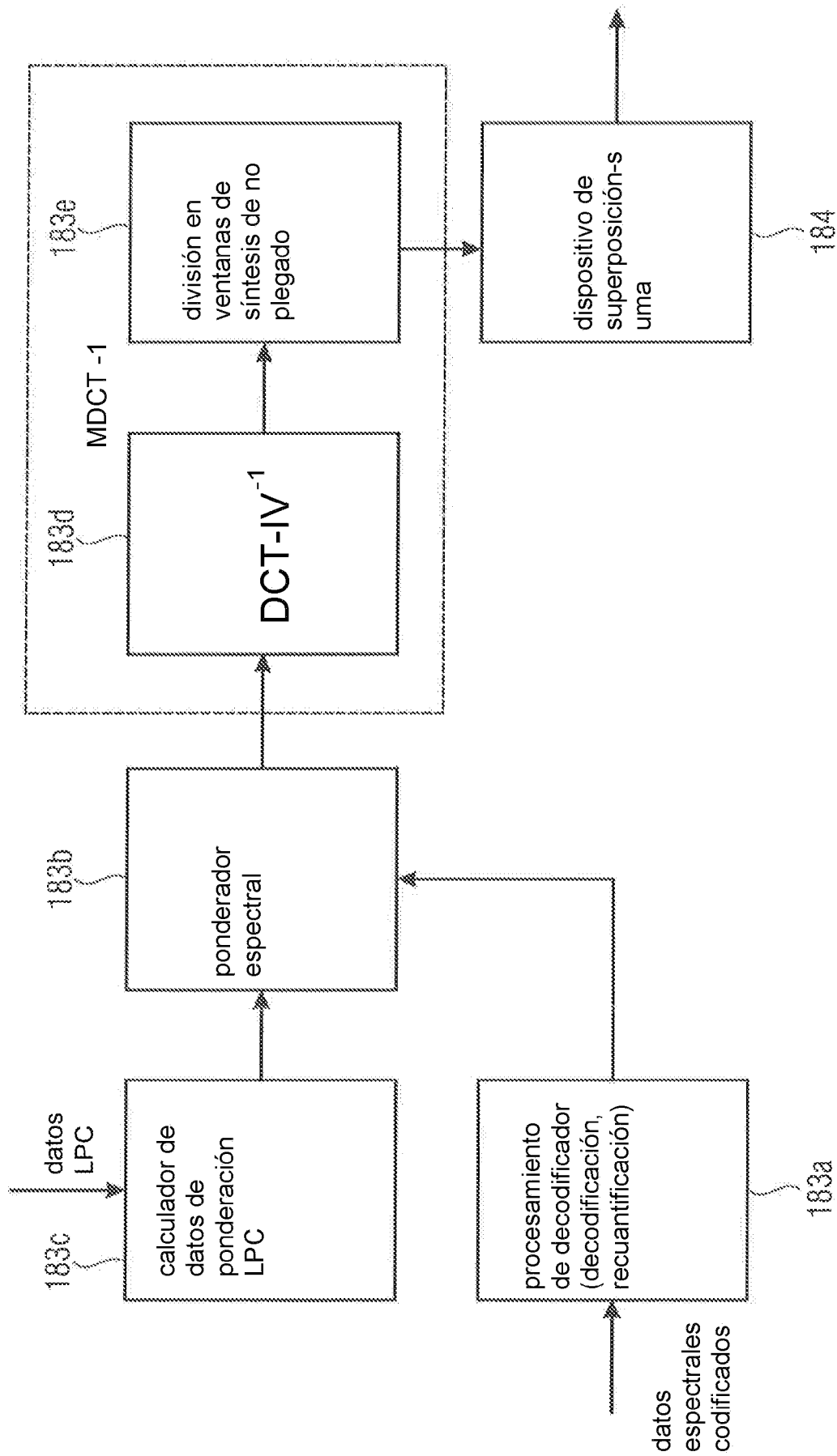


FIG 1C

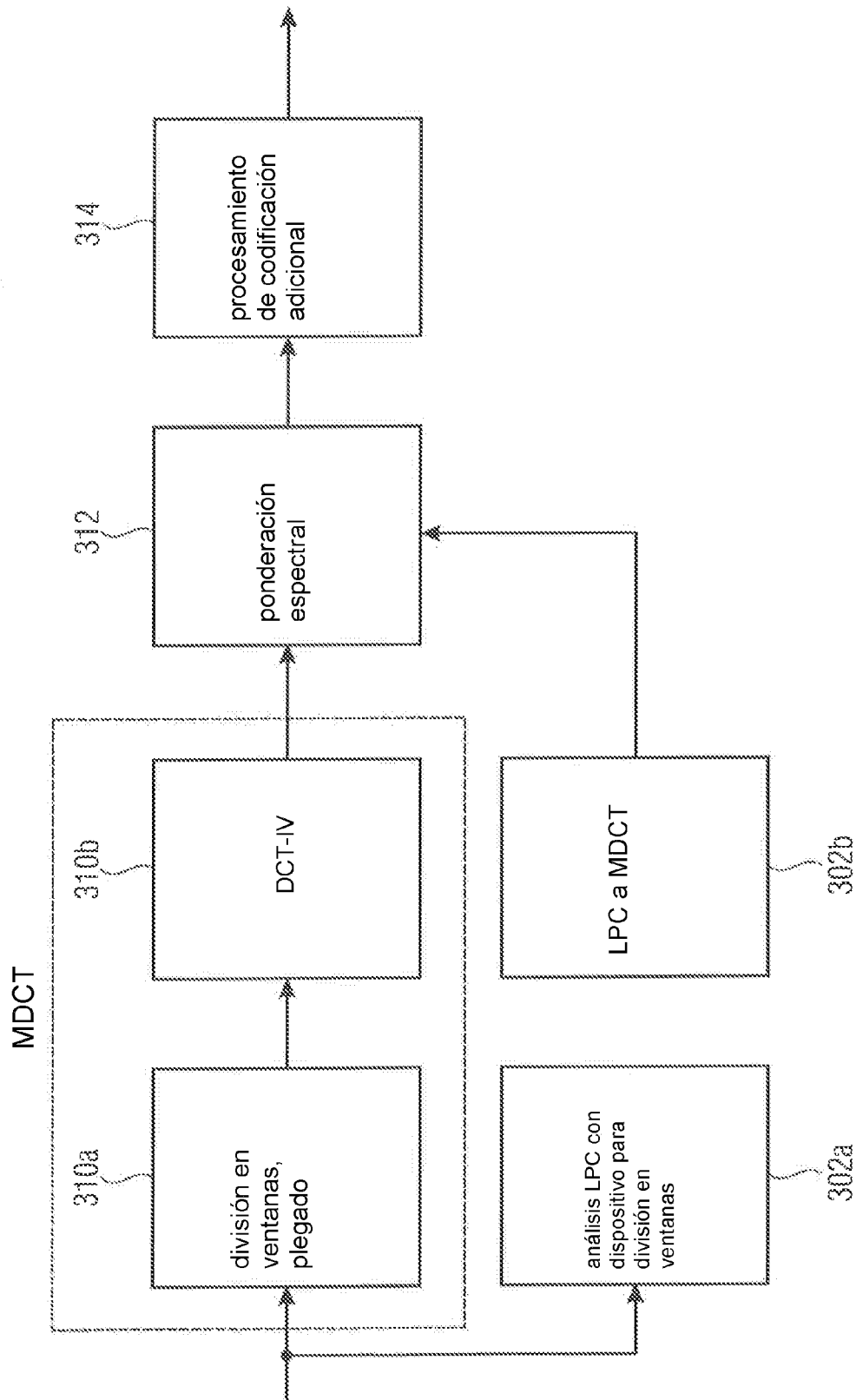
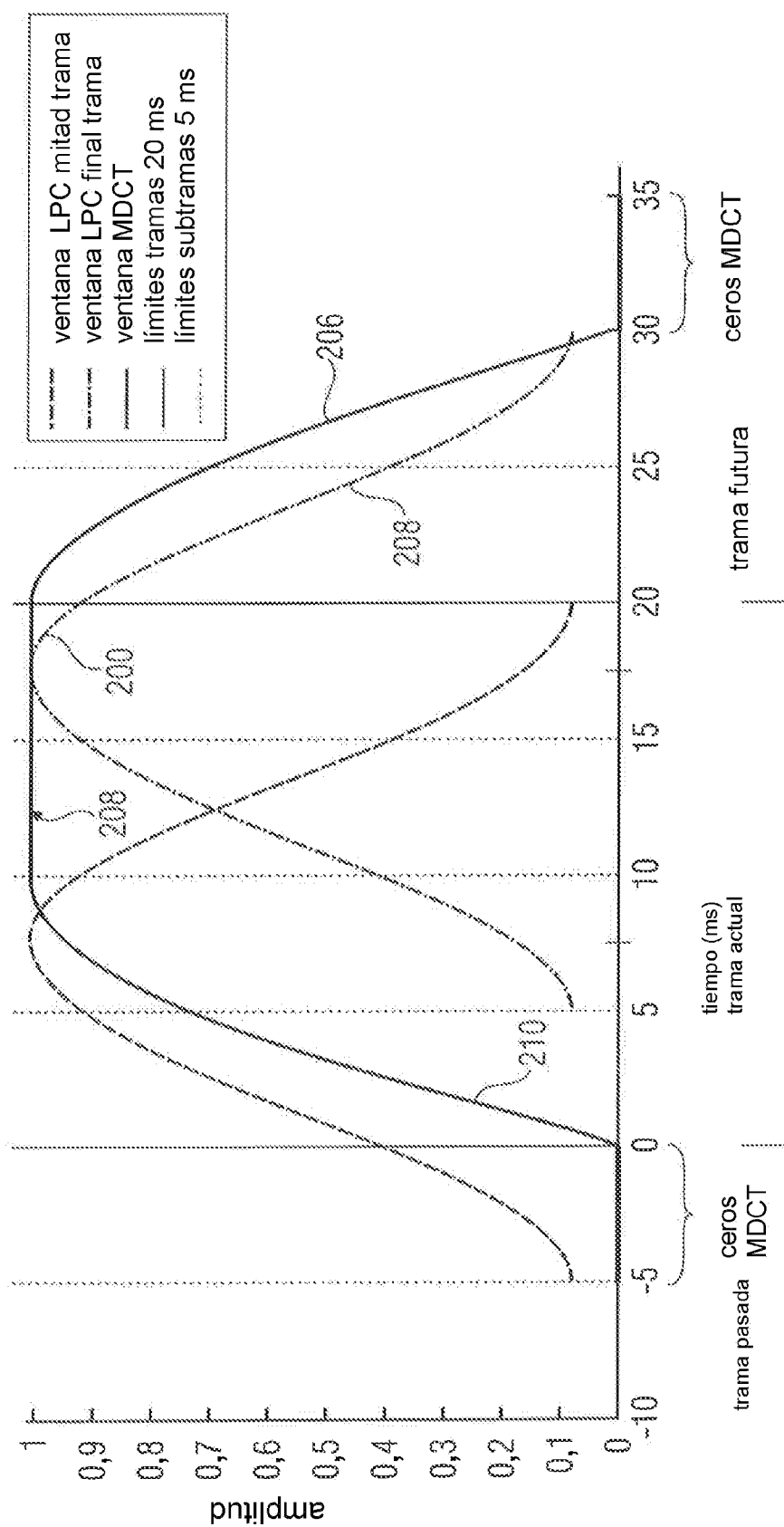


FIG 1D

superposición de tiempo ventanas MDCT; 10 ms
longitud total de ventana MDCT: 40 ms
longitud total de ventana LPC: 25 ms
las dos ventanas tienen una anticipación de 10 ms



ASGL

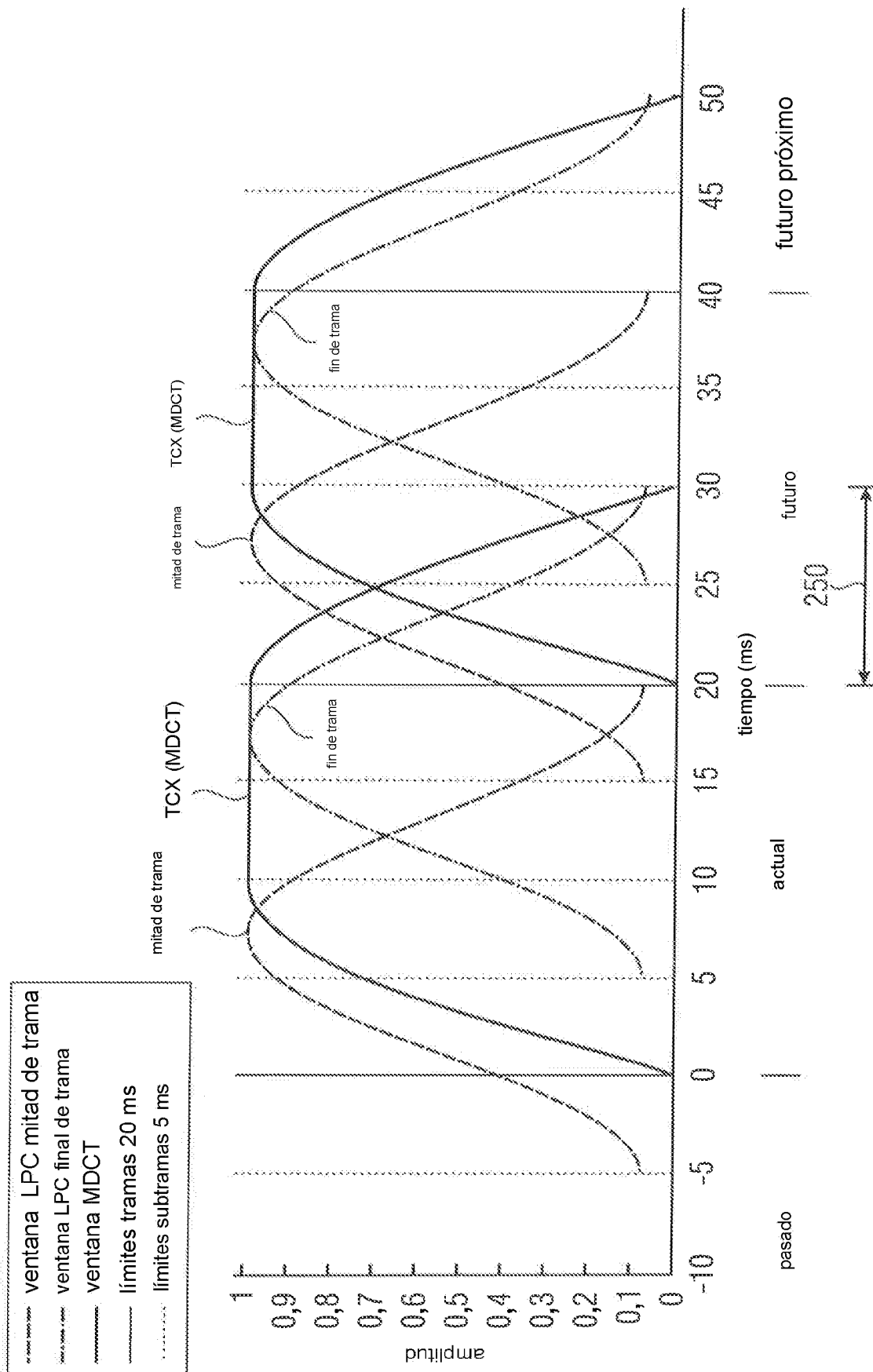
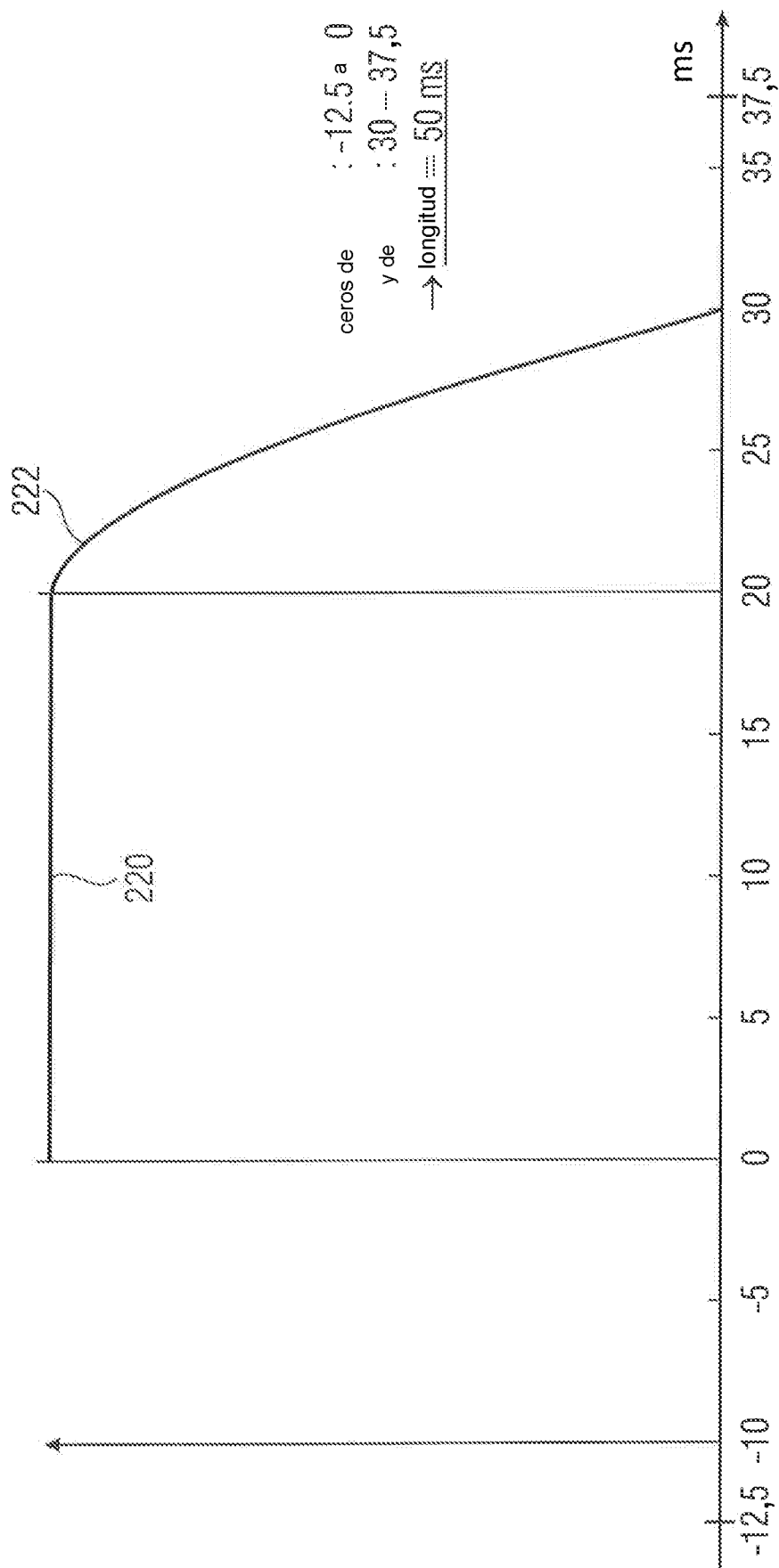


FIG 2B

※ transición de TCx a ACELP: sin precauciones especiales, ya que toda la superposición MDCT se encuentra en la región de anticipación



※ transición de ACELP a TCX: ventana de transición especial

FIG 2C

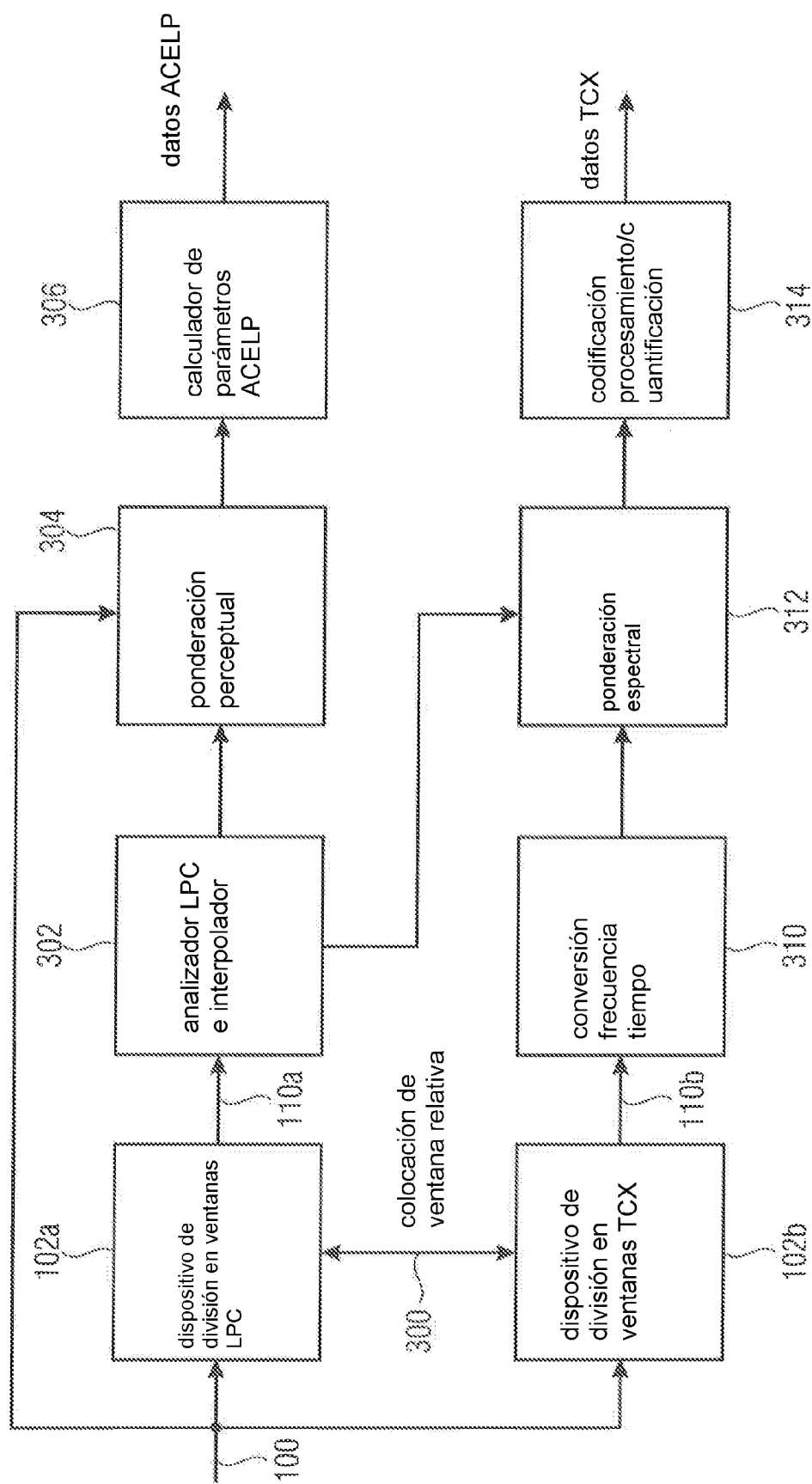


FIG 3A

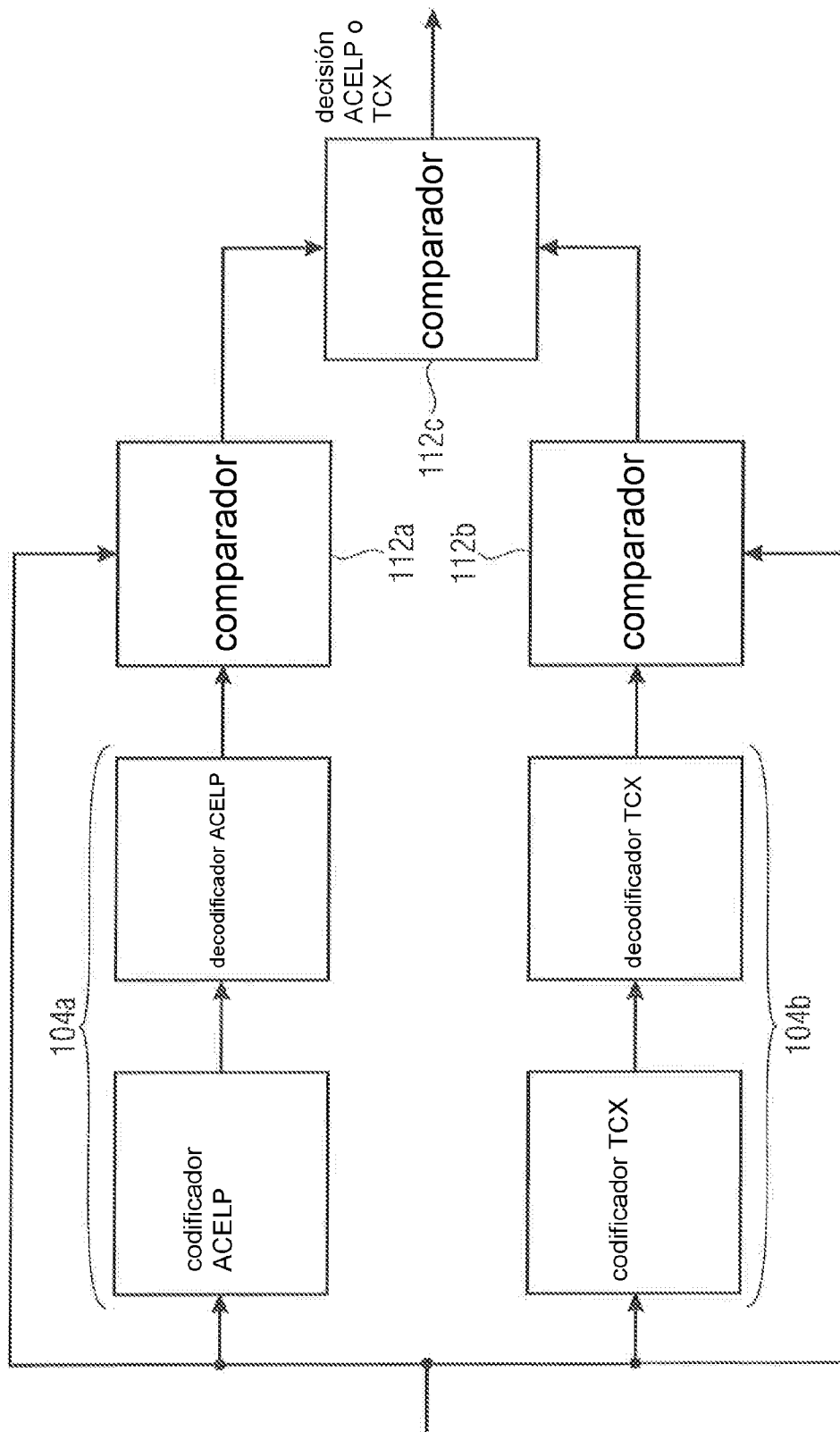


FIG 3B

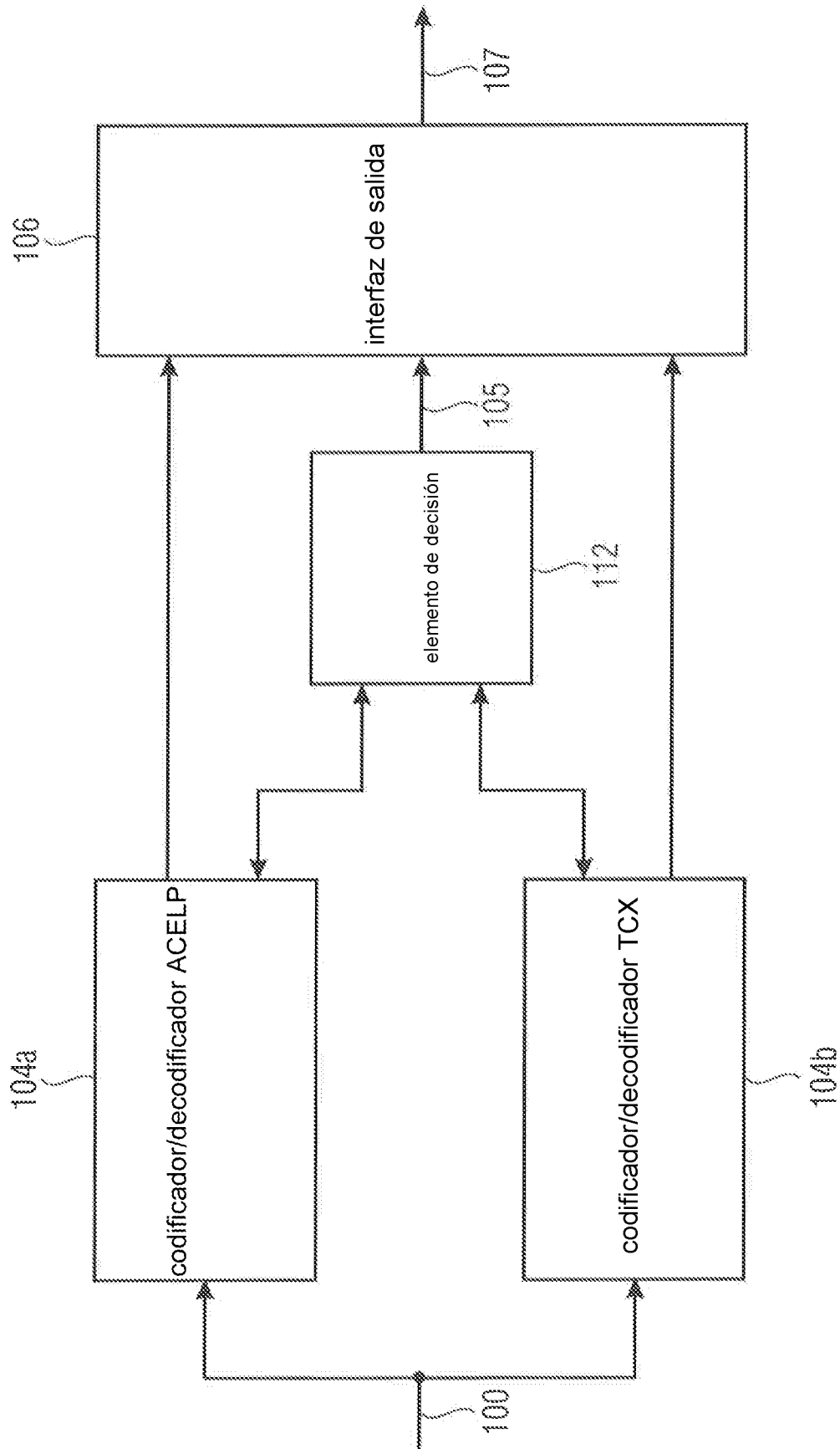


FIG 3C

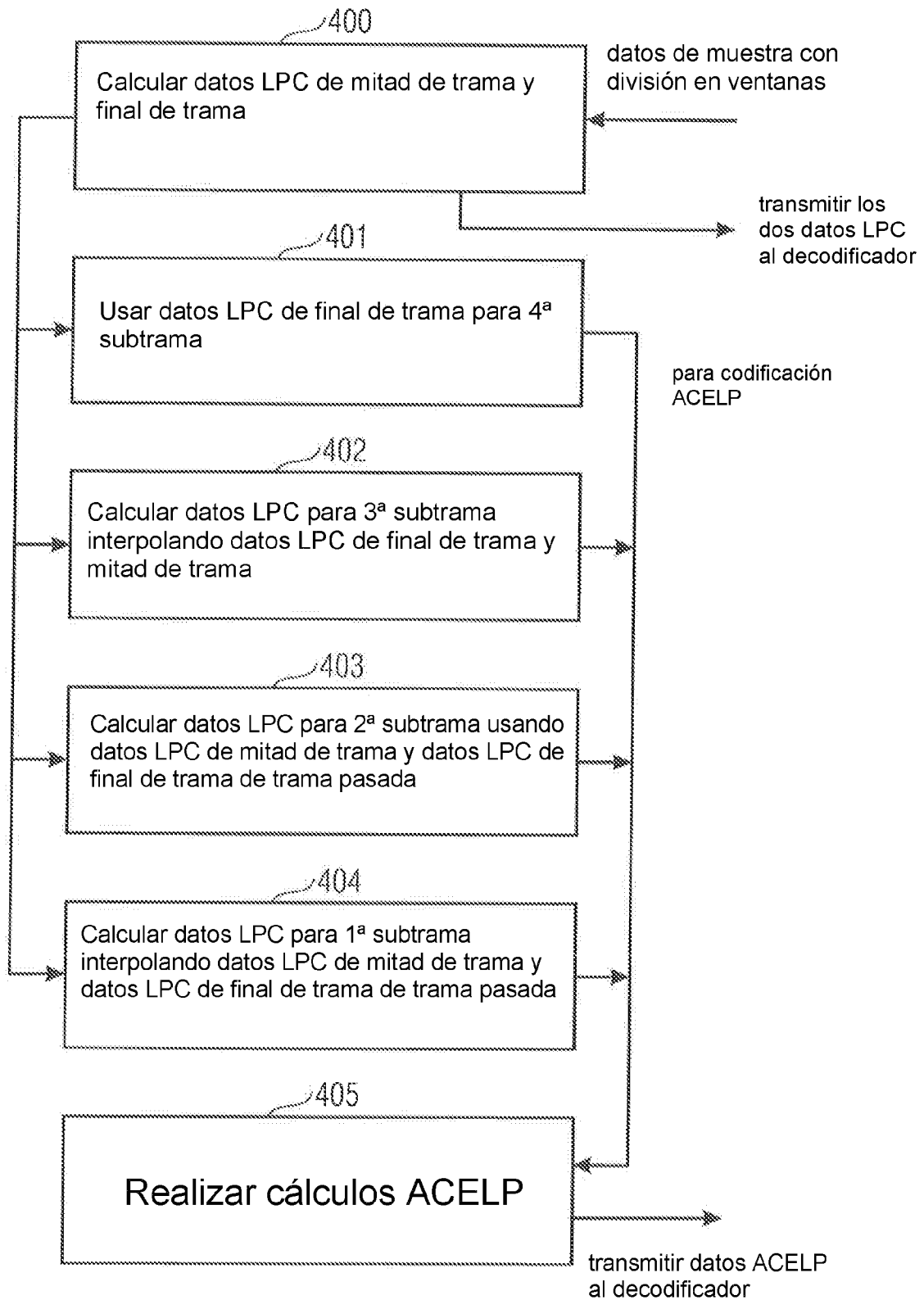


FIG 4A

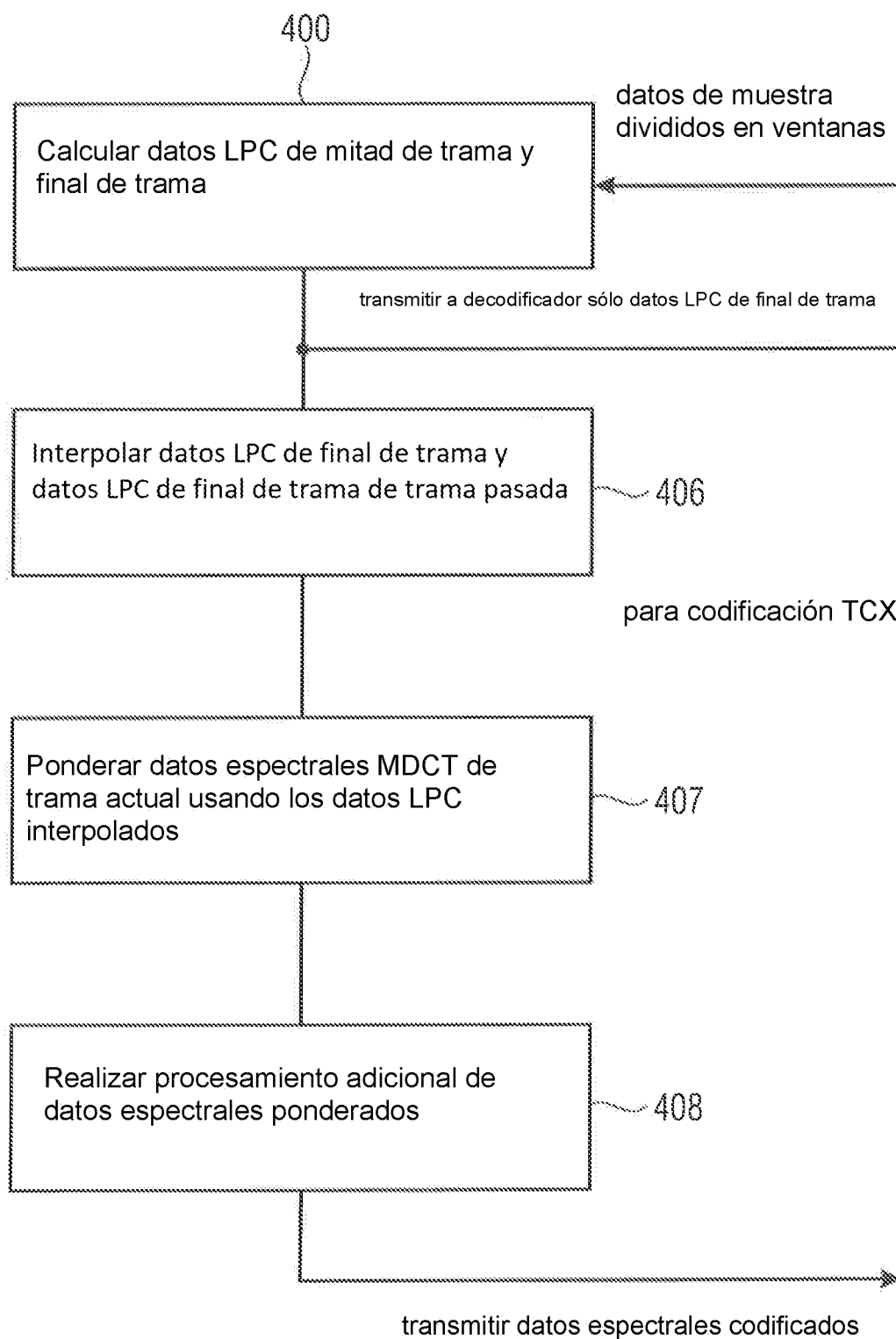


FIG 4B

* AMR-WB: ventana asimétrica con anticipación de 5 ms

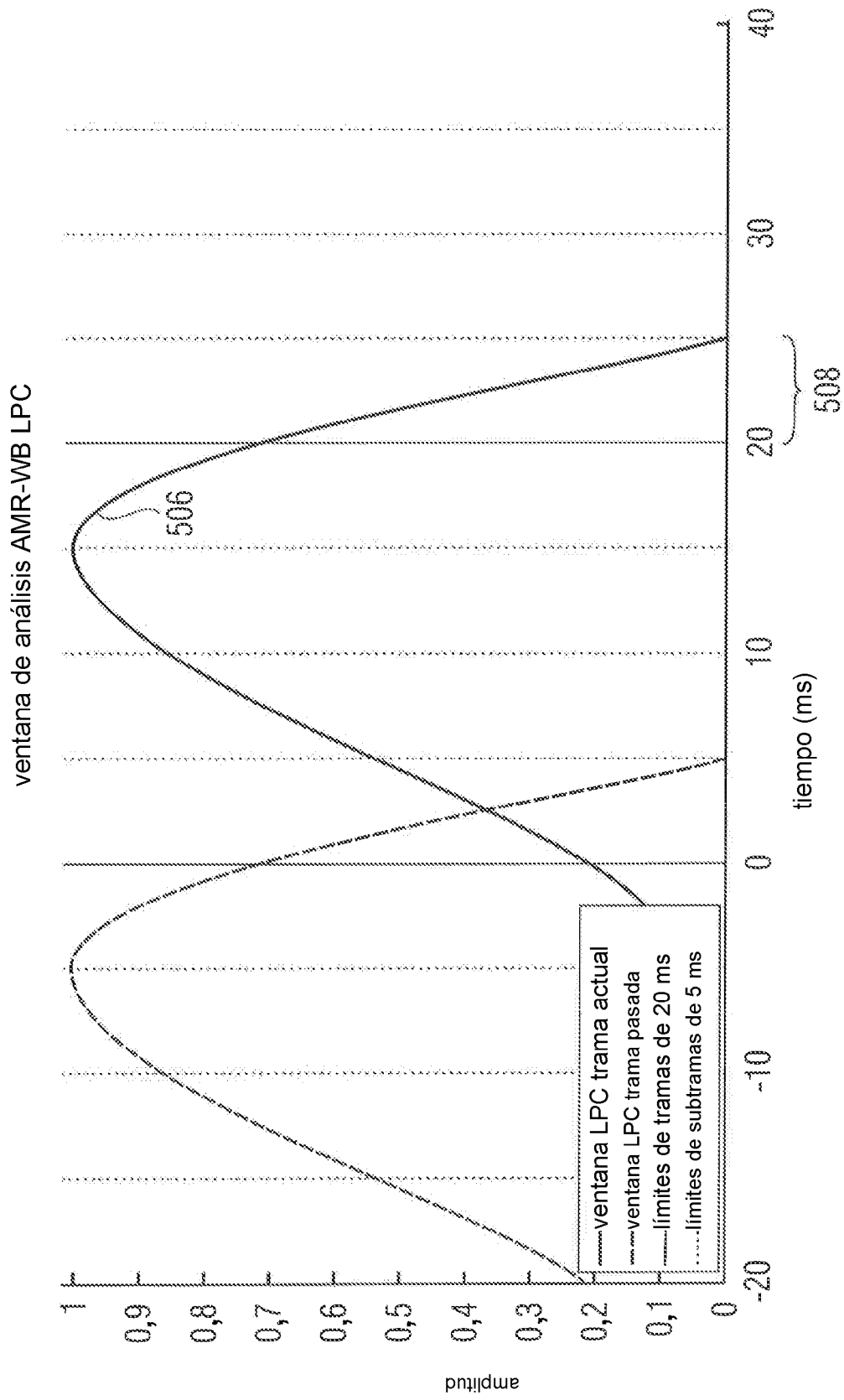


FIG 5A

* AMR-WB+:

ventana simétrica con anticipación de 20 ms

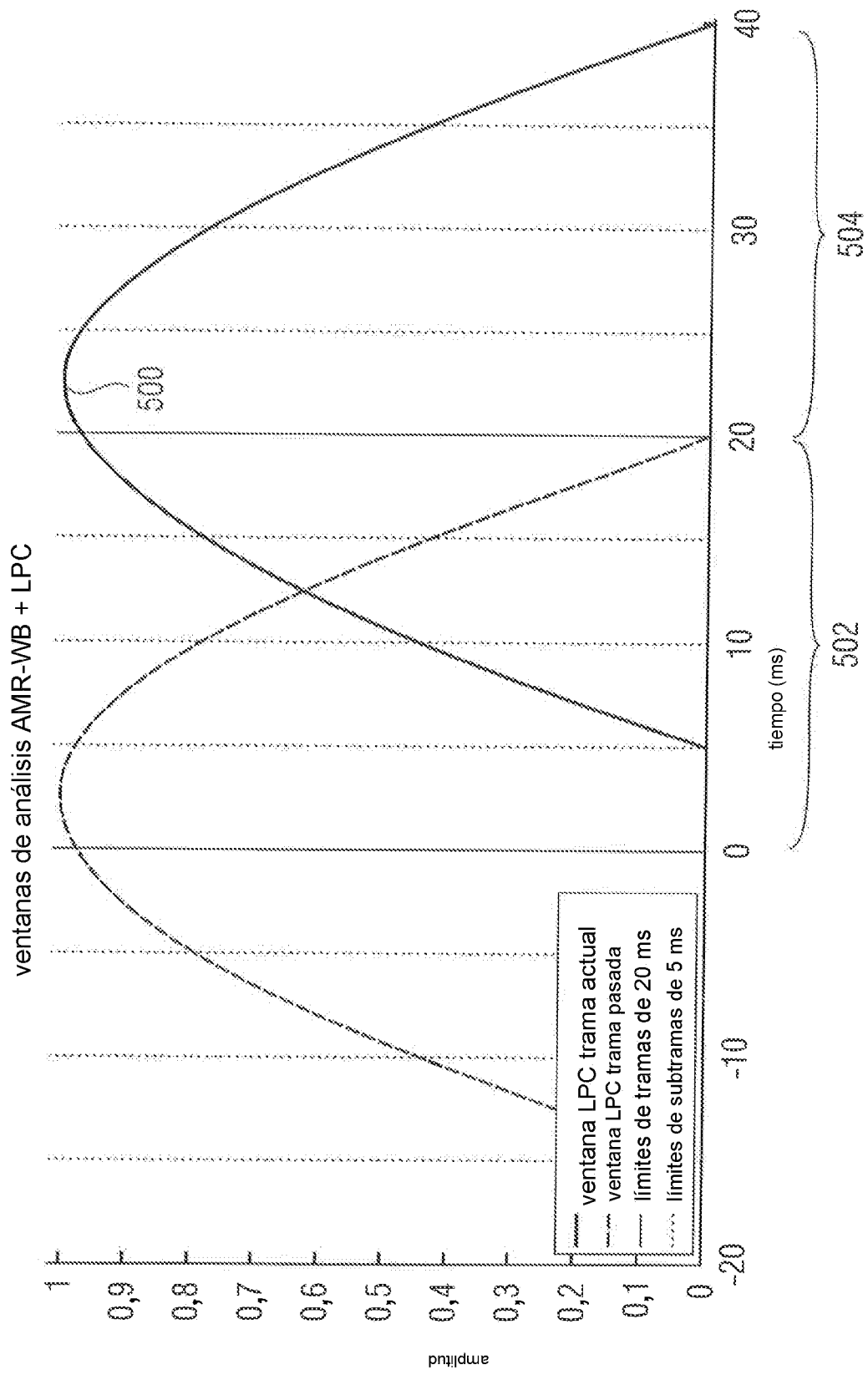


FIG 5B

- G.718: ventana simétrica con anticipación de 10 ms. Ventana adicional de mitad de trama.

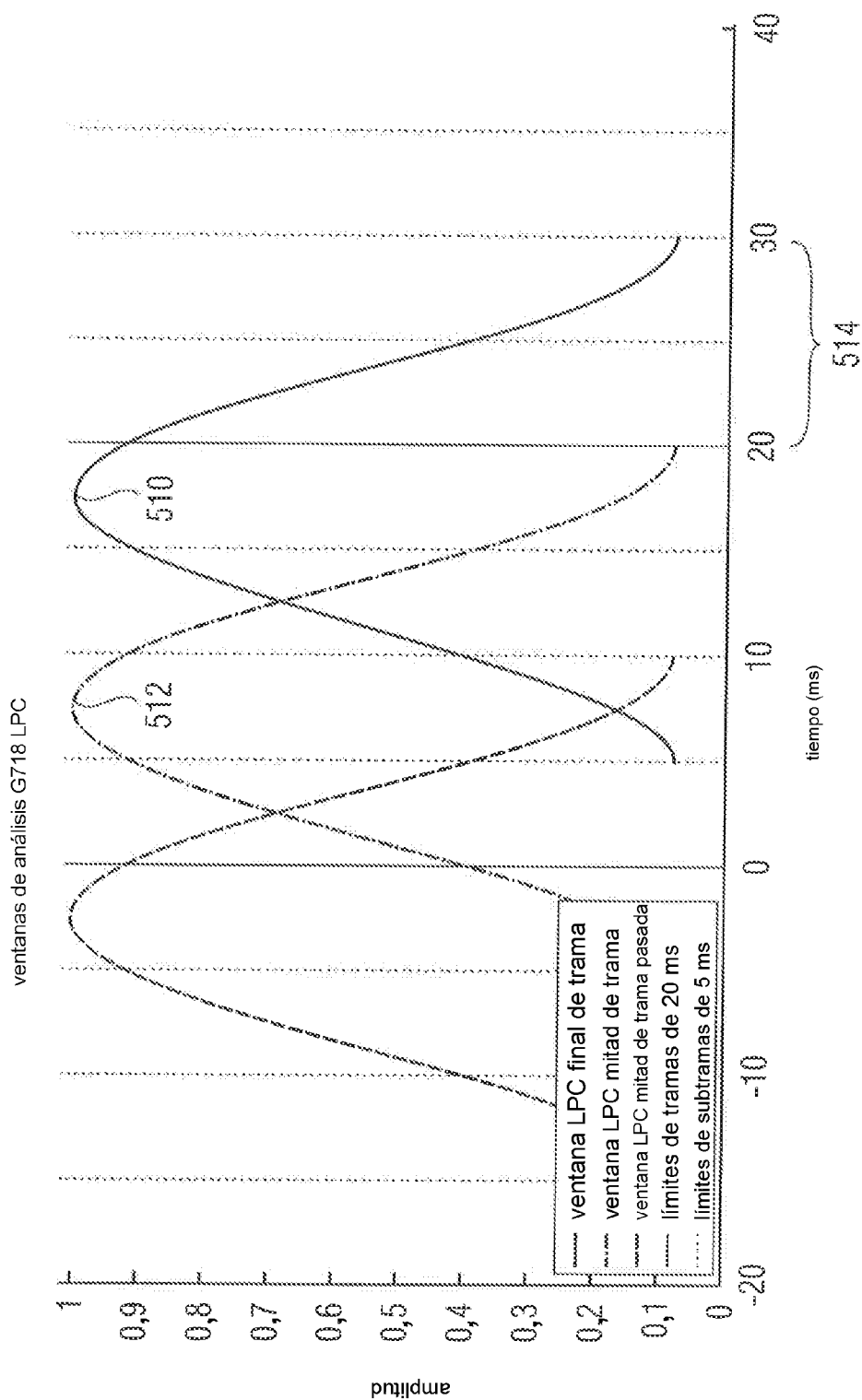


FIG 5C

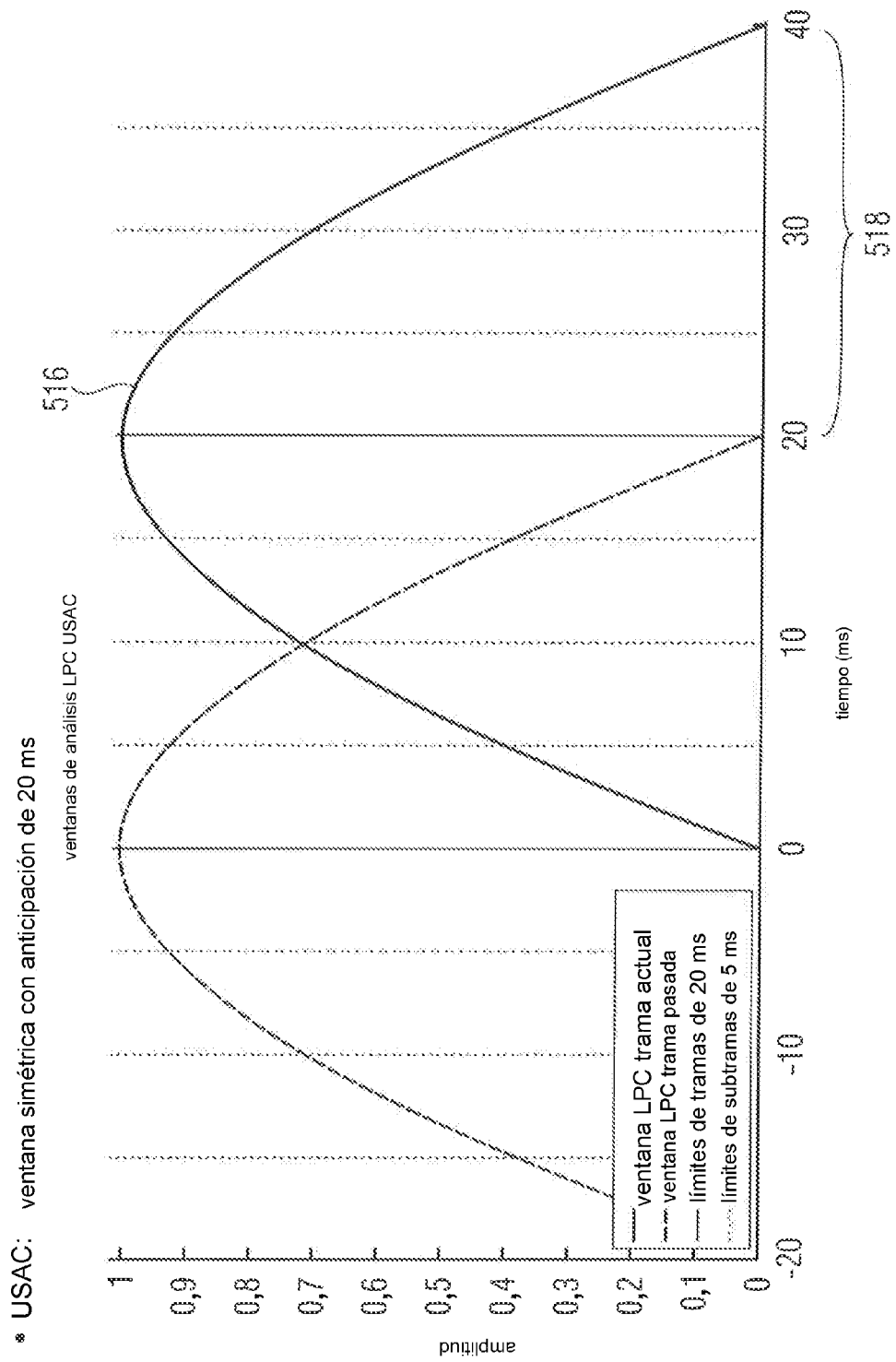


FIG 5D

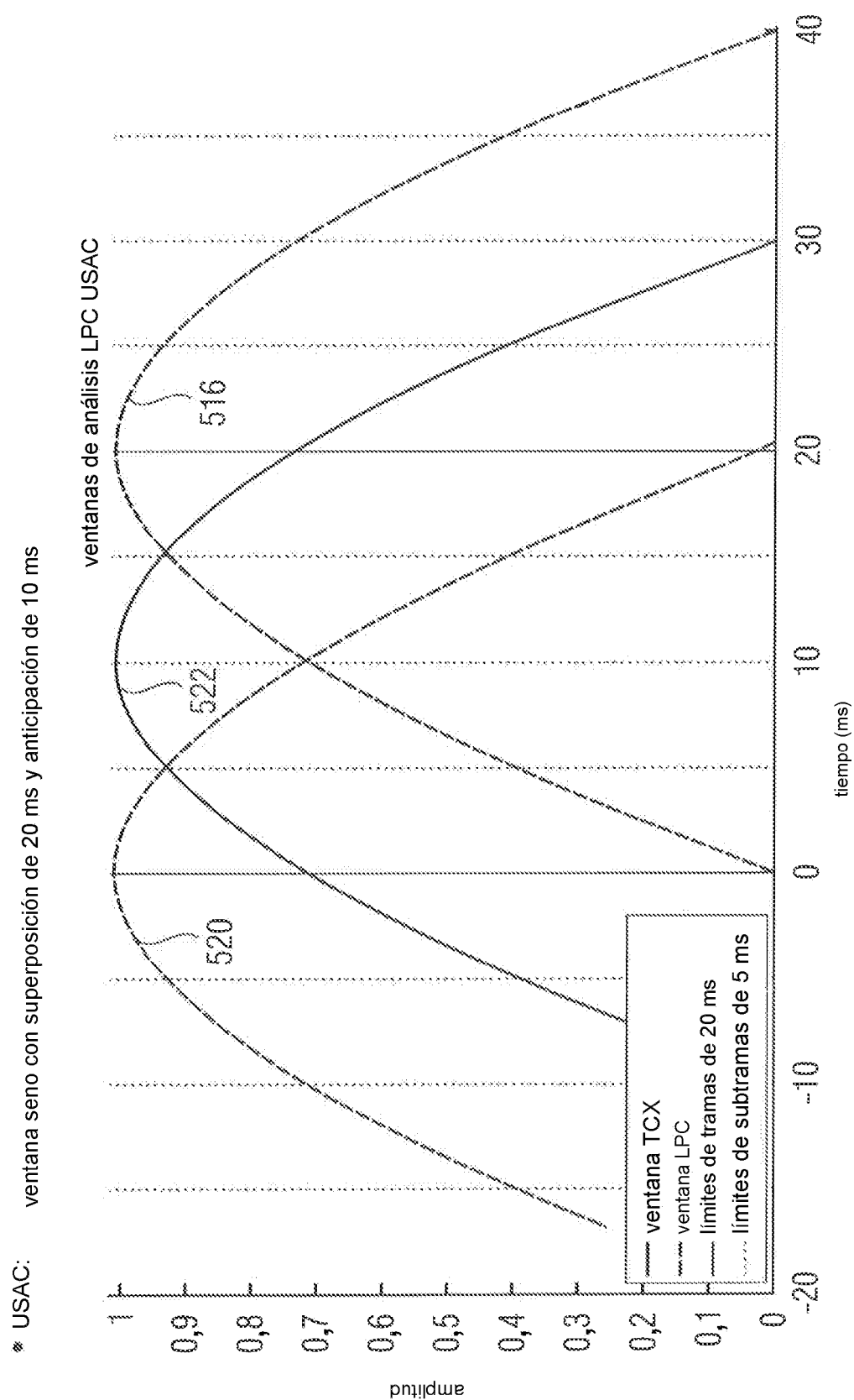


FIG 6