

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 706 061**

51 Int. Cl.:

**G10L 19/02** (2013.01)

**G10L 19/18** (2013.01)

**G10L 19/06** (2013.01)

**G10L 19/022** (2013.01)

**G10L 19/12** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.01.2011 PCT/CA2011/000040**

87 Fecha y número de publicación internacional: **21.07.2011 WO11085483**

96 Fecha de presentación y número de la solicitud europea: **13.01.2011 E 11732606 (6)**

97 Fecha y número de publicación de la concesión europea: **31.10.2018 EP 2524374**

54 Título: **Decodificación de audio con cancelación directa de distorsión por repliegue espectral en el dominio del tiempo usando filtrado predictivo lineal**

30 Prioridad:

**13.01.2010 US 294688 P**

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

**27.03.2019**

73 Titular/es:

**VOICEAGE CORPORATION (100.0%)  
Suite 250 750 Lucerne Road City of Mount Royal  
Quebec H3R 2H6, CA**

72 Inventor/es:

**BESSETTE, BRUNO**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 706 061 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

Decodificación de audio con cancelación directa de distorsión por repliegue espectral en el dominio del tiempo usando filtrado predictivo lineal

**Campo técnico**

- 5 La presente descripción se refiere al campo de la codificación y decodificación de señales de audio. Más específicamente, la presente descripción se refiere a la cancelación de distorsión por repliegue espectral en el dominio del tiempo en una señal de audio codificada.

**Antecedentes**

- 10 La codificación de audio de la tecnología de vanguardia usa la descomposición tiempo-frecuencia para representar la señal de una forma significativa para la reducción de datos. Más específicamente, los codificadores de audio usan transformadas para realizar una correlación de las muestras en el dominio del tiempo con coeficientes en el dominio de la frecuencia. Las transformadas de tiempo discreto usadas para esta correlación de tiempo a frecuencia se basan típicamente en núcleos de funciones sinusoidales, tales como la Transformada de Fourier Discreta (DFT) y la Transformada de Coseno Discreta (DCT). Se puede demostrar que tales transformadas logran compactación de energía de la señal de audio. Compactación de energía significa que, en el dominio de la transformada (o de la frecuencia), la distribución de energía se sitúa en menos coeficientes significativos en el dominio de la frecuencia que en las muestras en el dominio del tiempo. Se pueden lograr entonces ganancias de codificación aplicando asignación de bits adaptativa y cuantificación adecuada a los coeficientes en el dominio de la frecuencia. En el receptor, los bits que representan los parámetros cuantificados y codificados (incluyendo los coeficientes en el dominio de la frecuencia) se usan para recuperar los coeficientes cuantificados en el dominio de la frecuencia (u otros datos cuantificados tales como ganancias), y la transformada inversa genera la señal de audio en el dominio del tiempo. Se hace referencia generalmente a tales esquemas de codificación como codificación por transformada.

- 25 Por definición, la codificación por transformada opera en bloques consecutivos (normalmente llamados "tramas") de muestras de la señal de audio de entrada. Dado que la cuantificación introduce alguna distorsión en cada bloque sintetizado de la señal de audio, el uso de bloques no superpuestos puede introducir discontinuidades en los límites del bloque que pueden degradar la calidad de la señal de audio. Por lo tanto, en la codificación por transformada, para evitar discontinuidades, los bloques codificados de la señal de audio se superponen antes de aplicar la transformada, y se muestran en pequeñas ventanas apropiadamente en el segmento superpuesto para permitir una transición suave de un bloque de muestras decodificado al siguiente. Usar una transformada tal como la DFT (o su equivalente rápida, la Transformada Rápida de Fourier (FFT)) o la DCT y aplicarla a bloques superpuestos de muestras desafortunadamente da como resultado lo que se llama "muestreo no crítico". Por ejemplo, tomar una condición de superposición del 50% típica, codificando un bloque de  $N$  muestras consecutivas en el dominio del tiempo en realidad requiere tomar una transformada en  $2N$  muestras consecutivas, incluyendo  $N$  muestras del bloque presente y  $N$  muestras de las partes superpuestas del bloque anterior y siguiente. Por lo tanto, para cada bloque de  $N$  muestras en el dominio del tiempo, se codifican  $2N$  coeficientes en el dominio de la frecuencia. El muestreo crítico en el dominio de la frecuencia implica que  $N$  muestras de entrada en el dominio del tiempo producen solamente  $N$  coeficientes en el dominio de la frecuencia a ser cuantificados y codificados.

- 40 Se han diseñado transformadas especializadas para permitir el uso de ventanas superpuestas y mantener aún el muestreo crítico en el dominio de la transformada. Con tales transformadas especializadas, las  $2N$  muestras en el dominio del tiempo en la entrada de la transformada dan como resultado  $N$  coeficientes en el dominio de la frecuencia en la salida de la transformada. Para lograr esto, el bloque de  $2N$  muestras en el dominio del tiempo primero se reduce a un bloque de  $N$  muestras en el dominio del tiempo a través de una inversión de tiempo especial, la suma de partes específicas de la señal mostrada en pequeñas ventanas de longitud de  $2N$  muestras en un extremo de la ventana y la resta de partes específicas de la señal mostrada en pequeñas ventanas de longitud de  $2N$  muestras entre sí en el otro extremo de la ventana. Esta inversión de tiempo especial, suma y resta introducen lo que se denomina "distorsión por repliegue espectral en el dominio del tiempo" (TDA). Una vez que se introduce la TDA en el bloque de muestras de la señal de audio, no se puede eliminar usando solamente ese bloque. Es esta señal de distorsión por repliegue espectral en el dominio del tiempo la que es la entrada de una transformada de tamaño  $N$  (y no  $2N$ ), que produce los  $N$  coeficientes en el dominio de la frecuencia de la transformada. Para recuperar las  $N$  muestras en el dominio del tiempo, la transformada inversa usa los coeficientes de transformada de dos tramas o bloques consecutivos y superpuestos para cancelar la TDA, en un proceso llamado cancelación de distorsión por repliegue espectral en el dominio del tiempo (TDAC).

- 55 Un ejemplo de una transformada tal que aplica la TDAC, que se usa ampliamente en codificación de audio, es la Transformada de Coseno Discreta Modificada (MDCT). En realidad, la MDCT introduce TDA sin plegado explícito en el dominio del tiempo. Más bien, la distorsión por repliegue espectral en el dominio del tiempo se introduce cuando se considera tanto la MDCT directa como la MDCT inversa (IMDCT) de un único bloque de muestras. Esto proviene de la construcción matemática de la MDCT y es bien sabido por los expertos en la técnica. Pero también se sabe que esta distorsión por repliegue espectral en el dominio del tiempo implícita se puede ver como equivalente a las

primeras partes de inversión de las muestras en el dominio del tiempo y sumar (o restar) estas partes invertidas a otras partes de la señal. Esto se conoce como “plegado”.

5 Surge un problema cuando un codificador de audio conmuta entre dos modos de codificación, uno que usa la TDAC y el otro no. Supongamos, por ejemplo, que un códec conmuta de un modo de codificación de TDAC a un modo de codificación de no TDAC. El lado del bloque de muestras codificadas que usa el modo de codificación de TDAC, y que es común al bloque codificado sin usar la TDAC, contiene TDA que no se puede cancelar usando el bloque de muestras codificadas que usa el modo de codificación de no TDAC.

Una primera solución es descartar las muestras que contienen distorsión por repliegue espectral que no se pueden cancelar.

10 Esta primera solución da como resultado un uso ineficiente del ancho de banda de transmisión debido a que el bloque de muestras para el que no se puede cancelar la TDA se codifica dos veces, una vez por el códec basado en TDAC y una segunda vez por el códec no basado en TDAC.

15 Una segunda solución es usar ventanas especialmente diseñadas que no introducen TDA en al menos una parte de la ventana cuando se aplica el proceso de inversión de tiempo y de suma/resta. La Figura 1 es un diagrama de un ejemplo de una ventana de  $2N$  muestras que introduce TDA en su lado izquierdo pero no en su lado derecho. La ventana 100 de la Figura 1 es útil para transiciones de un códec basado en TDAC a un códec no basado en TDAC. La primera mitad de la ventana 100 está conformada de modo que introduce la TDA 110, que se puede cancelar si la ventana anterior también usa TDA con superposición. No obstante, el lado derecho de la ventana 100 en la Figura 1 tiene una región con valor cero 120 después del punto de plegado en la posición  $3N/2$ . Esta región 120 de la ventana 100 por lo tanto no introduce ninguna TDA cuando el proceso de inversión de tiempo y suma/resta (o plegado) se realiza alrededor del punto de plegado en la posición  $3N/2$ .

20 Como se ilustra en la Figura 1, la ventana 100 contiene una región plana 130 precedida por una región estrechada 140 en el lado izquierdo. El propósito de la región estrechada 140 es proporcionar una buena resolución espectral cuando se calcula la transformada y suavizar la transición durante las operaciones de superposición y suma entre bloques adyacentes. Aumentar la duración de la región plana 130 de la ventana 100 reduce la sobrecarga de información. No obstante, la región 120 disminuye el rendimiento espectral de la ventana 100 dado que la información de muestra de valor cero solamente se transporta en la región 120.

25 Por lo tanto, hay una necesidad de una técnica de TDAC mejorada que se pueda utilizar, por ejemplo, en el Códec de Habla y Audio Unificado (USAC) del Grupo de Expertos de Imagen en Movimiento (MPEG), para gestionar las diferentes transiciones entre tramas que usan ventanas rectangulares y no superpuestas y tramas que usan ventanas no rectangulares y superpuestas, al tiempo que aseguran una resolución espectral adecuada, reducción de sobrecarga de datos y suavidad de transición entre estos diferentes tipos de tramas.

30 La solicitud de patente PCT N° WO 2011/048117 A1 describe un codificador de señal de audio, un decodificador de señal de audio, y un método para codificar o decodificar una señal de audio que usan una cancelación de distorsión por repliegue espacial. Un decodificador de señal de audio para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada del contenido de audio comprende un camino en el dominio de la transformada configurado para obtener una representación en el dominio del tiempo de una parte del contenido de audio codificado en un modo en el dominio de la transformada sobre la base de un primer conjunto de coeficientes espectrales, una representación de una señal de estímulo de cancelación de distorsión por repliegue espectral y una pluralidad de parámetros en el dominio de la predicción lineal. El camino en el dominio de la transformada comprende un procesador de espectro configurado para aplicar una conformación de espectro al primer conjunto de coeficientes espectrales en dependencia de al menos un subconjunto de parámetros en el dominio de la predicción lineal, para obtener una versión conformada espectralmente del primer conjunto de coeficientes espectrales. El camino en el dominio de la transformada comprende un primer convertidor en el dominio de la frecuencia a dominio del tiempo configurado para obtener una representación en el dominio del tiempo del contenido de audio sobre la base de la versión conformada espectralmente del primer conjunto de coeficientes espectrales. El camino en el dominio de la transformada comprende un filtro de estímulo de cancelación de distorsión por repliegue espectral configurado para filtrar la señal de estímulo de cancelación de distorsión por repliegue espectral en dependencia de al menos un subconjunto de los parámetros en el dominio de la predicción lineal, para derivar una señal de síntesis de cancelación de distorsión por repliegue espectral de la señal de estímulo de cancelación de distorsión por repliegue espectral. El camino en el dominio de la transformada también comprende un combinador configurado para combinar la representación en el dominio del tiempo del contenido de audio con la señal de síntesis de cancelación de distorsión por repliegue espectral, o una versión procesada posteriormente del mismo, para obtener una señal en el dominio del tiempo de distorsión por repliegue espectral reducida.

55 El artículo técnico de MAX NEUENDORF ET AL “Completion of Core Experiment on unification of USAC Windowing and Frame Transitions”, 91. REUNIÓN DEL MPEG; 18-1-2010 – 22-1-2010; KYOTO (GRUPO DE EXPERTOS DE IMAGEN EN MOVIMIENTO O ISO/IEC JTC1/SC29/WG11) se conoce en la técnica anterior. Un experimento central descrito en este documento de la técnica anterior aspira a mejorar la calidad de audio y el diseño estructural del sistema USAC simplificando y distorsionando por repliegue espectral la estructura de trama y las diversas

transiciones de trama, unificando la conformación de ruido de cuantificación y reduciendo el alto número de diferentes longitudes de transformada para las herramientas de codificación por transformada. Además, se eliminaron las partes de señal codificadas de manera redundante y se amplió el intervalo de transiciones permitidas, aumentando la flexibilidad de la estructura.

- 5 La solicitud de patente PCT N° WO 2012/004349 describe un codificador que usa cancelación directa de distorsión por repliegue espectral. Un códec que soporta conmutación entre el modo de codificación por transformada de cancelación de distorsión por repliegue espectral en el dominio del tiempo y el modo de codificación en el dominio del tiempo se hace menos propenso a la pérdida de tramas añadiendo una parte de sintaxis adicional a las tramas, dependiendo de lo que el analizador sintáctico del decodificador puede seleccionar entre una primera acción de esperar que la trama actual comprenda, y de este modo leer datos de cancelación directa de distorsión por repliegue espectral de la trama actual y una segunda acción de no esperar que la trama actual comprenda, y de este modo no leer datos de cancelación directa de distorsión por repliegue espectral de la trama actual. En otras palabras, aunque se pierde un poco de eficiencia de codificación debido a la provisión de la nueva parte de sintaxis, es meramente la nueva parte de sintaxis la que proporciona la capacidad de usar el códec en caso de un canal de comunicación con pérdida de tramas. Sin la nueva parte de sintaxis, el decodificador no sería capaz de decodificar ninguna parte de flujo de datos después de una pérdida y se estrellaría al intentar reanudar el análisis sintáctico. De este modo, en un entorno propenso a errores, la eficiencia de codificación se evita que desaparezca por la introducción de la nueva parte de sintaxis.

- 20 El artículo técnico de BRUNO BESSETTE ET AL: "Alternatives for windowing in USAC", 89. REUNIÓN DEL MPEG; 29-6-2009 – 3-7-2009; LONDRES (GRUPO DE EXPERTOS DE IMAGEN EN MOVIMIENTO O ISO/IEC JTC1/SC29/WG11) es conocido en la técnica anterior. Este documento de la técnica anterior propone alternativas a la selección de ventanas aplicada en el modo de TCX del Códec de Habla y Audio Unificado (USAC). La selección de ventanas y la conmutación de modo es una parte esencial de USAC, con diferentes posibilidades para opciones y compromisos de diseño. Se propone modificar, y en realidad armonizar, las formas de las ventanas en los modos de TCX de USAC, con el fin de aliviar algunas limitaciones. Se describen los cambios requeridos en el USAC para permitir ventanas más coherentes en TCX, incluyendo cómo cancelar el efecto de selección de ventanas y la distorsión por repliegue espectral en el dominio del tiempo en la transición entre las tramas de ACELP y de TCX. También se proporciona evidencia de las ventajas de rendimiento y otros beneficios de los cambios propuestos.

### Compendio

- 30 Por lo tanto, hay una necesidad de una técnica de cancelación de distorsión por repliegue espectral para soportar la conmutación entre modos de codificación, en donde la técnica compensa los efectos de distorsión por repliegue espectral en un punto de conmutación entre estos modos.

La invención se expone en el conjunto de reivindicaciones adjuntas.

- 35 La anterior y otras características llegarán a ser más evidentes tras la lectura de la siguiente descripción no restrictiva de realizaciones ilustrativas del dispositivo y del método para cancelación directa de distorsión por repliegue espectral en el dominio del tiempo, dada a modo de ejemplo solamente con referencia a los dibujos que se acompañan.

### Breve descripción de los dibujos

En los dibujos adjuntos:

- 40 la Figura 1 es un diagrama esquemático de un ejemplo de ventana que introduce la TDA en su lado izquierdo pero no en su lado derecho;
- la Figura 2 es un diagrama esquemático de un ejemplo de transición de una trama que usa una ventana rectangular no superpuesta a una trama que usa una ventana superpuesta;
- la Figura 3 es un diagrama esquemático que muestra el plegado y la TDA aplicada al diagrama de la Figura 2;
- 45 la Figura 4 es un diagrama esquemático de una secuencia de operaciones de un método ejemplar de cálculo de un objetivo de FAC;
- la Figura 5 es un diagrama de bloques esquemático que muestra la cuantificación del objetivo de FAC de la Figura 4;
- 50 la Figura 6 es un diagrama esquemático de una secuencia de operaciones de un método ejemplar de cálculo de una síntesis de una señal de audio, que usa parámetros de FAC representativos del objetivo de FAC de la Figura 4;
- la Figura 7 es un diagrama de bloques esquemático de un ejemplo no limitativo de dispositivo para cancelación directa de distorsión por repliegue espectral en el dominio del tiempo en una señal de audio codificada recibida en un flujo de bits; y

la Figura 8 es un diagrama de bloques de un ejemplo no limitativo de dispositivo para cancelación directa de distorsión por repliegue espectral en el dominio del tiempo en una señal de audio codificada para transmisión a un decodificador.

**Descripción detallada**

5 La siguiente descripción aborda el problema de cancelación de los efectos de la distorsión por repliegue espectral en el dominio del tiempo y la selección de ventanas no rectangulares cuando una señal de audio se codifica usando tanto ventanas superpuestas como no superpuestas en tramas contiguas. Usando la tecnología descrita en la presente memoria, el uso de ventanas no óptimas y especiales se puede evitar al tiempo que aún se permite una gestión correcta de transiciones de tramas entre modos de codificación que usan tanto ventanas rectangulares y no  
10 superpuestas como ventanas no rectangulares y superpuestas.

La codificación Predictiva Lineal (LP), por ejemplo, la codificación de ACELP (Predicción Lineal Excitada por Código Algebraico), es un ejemplo de modo de codificación en el que una trama se codifica usando selección de ventanas rectangulares y no superpuestas. Alternativamente, un ejemplo de modo de codificación que usa selección de  
15 ventanas no rectangulares y superpuestas es la codificación de Excitación Codificada por Transformada (TCX), como se aplica en el Códec de Habla y Audio Unificado (USAC) MPEG. Otro ejemplo de modo de codificación que usa selección de ventanas no rectangulares y superpuestas es la codificación por transformada perceptiva como en el modo de FD de USAC, donde también se usa una MDCT como transformada y se usa un modelo perceptual para asignar dinámicamente los bits a los coeficientes de transformada. En USAC, las tramas de TCX usan tanto  
20 ventanas superpuestas como Transformada de Coseno Discreta Modificada (MDCT), que introduce Distorsión por Repliegue Espectral en el Dominio del tiempo (TDA). USAC también es un ejemplo típico donde las tramas contiguas se pueden codificar usando o bien ventanas rectangulares y no superpuestas, tales como en tramas de ACELP, o bien ventanas no rectangulares y superpuestas, tales como en tramas de TCX. Sin pérdida de generalidad, la presente descripción considera, de este modo, el ejemplo específico de USAC para ilustrar los beneficios del dispositivo y del método para cancelación directa de distorsión por repliegue espectral en el dominio  
25 del tiempo.

Se abordan dos casos distintos en la presente descripción. El primer caso se refiere a una transición de una trama que usa una ventana rectangular y no superpuesta a una trama que usa una ventana no rectangular y superpuesta. El segundo caso se refiere a una transición de una trama que usa una ventana no rectangular y superpuesta a una  
30 trama que usa una ventana rectangular y no superpuesta. Con el propósito de ilustración y sin sugerir limitación, las tramas que usan una ventana rectangular y no superpuesta se pueden codificar usando el modo de codificación de ACELP, y las tramas que usan una ventana no rectangular y superpuesta se pueden codificar usando el modo de codificación de TCX. Además, se pueden usar duraciones específicas para algunas tramas, por ejemplo, 20 milisegundos para una trama de TCX, señalada TCX20. No obstante, se debería tener en cuenta que estos ejemplos se usan solamente con propósitos de ilustración, y que se pueden contemplar otras longitudes de tramas y modos  
35 de codificación distintos de ACELP y TCX.

El caso de una transición de una trama con ventana rectangular y no superpuesta a una trama con ventana no rectangular y superpuesta se abordará ahora en relación con la siguiente descripción tomada junto con la Figura 2, que es un diagrama esquemático de un ejemplo de transición de una trama que usa una ventana rectangular y no superpuesta a una trama que usa una ventana superpuesta.

Más específicamente, la Figura 2 ilustra un ejemplo de trama de ACELP 201 que usa una ventana rectangular y no superpuesta 202 y un ejemplo de trama de TCX20 203 que usa una ventana no rectangular y superpuesta 204. TCX20 se refiere a las tramas de TCX cortas en USAC, que nominalmente tienen una duración de 20 ms, como lo hacen las tramas de ACELP en muchas aplicaciones. La Figura 2 muestra qué muestras se usan en cada trama, y cómo se muestran en pequeñas ventanas en un codificador. La misma ventana 204 se aplica en un decodificador,  
45 de manera que el efecto combinado visto en el decodificador es el cuadrado de la forma de la ventana mostrada en la Figura 2. Por supuesto, esta selección de ventana doble, una vez en el codificador y una segunda vez en el decodificador, es típica en codificación por transformada. La ventana no rectangular 204 para la trama de TCX20 203 mostrada en la Figura 2 se elige de manera que, si las tramas anterior y siguiente también usan ventanas no rectangulares y superpuestas, entonces las partes de superposición 204a y 204d de la ventana 204 son, después de la segunda selección de ventana en el decodificador, complementarias y permiten recuperar la señal “sin mostrar pequeñas ventanas” en la región de superposición de las ventanas.  
50

Para codificar la trama de TCX20 203 de la Figura 2 de una manera eficiente, la distorsión por repliegue espectral en el dominio del tiempo (TDA) se aplica típicamente a las muestras mostradas en pequeñas ventanas para esa trama de TCX20 203. Más específicamente, las partes izquierda 204a y derecha 204d de la ventana 204 se pliegan y combinan. La Figura 3 es un diagrama esquemático que muestra el plegado y la TDA aplicados al diagrama de la  
55 Figura 2. En la Figura 3, la ventana no rectangular 204 de la Figura 2 se muestra en cuatro cuartos. El 1º y 4º cuartos, 204a y 204d de la ventana 204 se muestran en línea discontinuas, ya que se combinan con el 2º y 3º cuartos, 204b, 204c, mostrados en línea continua. La combinación del 1º y 4º cuartos 204a, 204d, con el 2º y 3º cuartos 204b, 204c, usa un proceso similar al usado en codificación por MDCT, de la siguiente manera. El 1º cuarto 204a se invierte en el tiempo, luego se alinea, muestra por muestra, con el 2º cuarto 204b de la ventana, y  
60

finalmente el 1º cuarto 204e invertido en el tiempo y desplazado se resta del 2º cuarto 204b de la ventana 203. De manera similar, el 4º cuarto 204d de la ventana se invierte en tiempo y se desplaza para formar el 4º cuarto 204f invertido en el tiempo y desplazado alineado con el 3º cuarto 204c de la ventana 204, y finalmente se suma al 3º cuarto 204c en la ventana 204. Si la ventana de TCX20 204 mostrada en la Figura 2 tiene  $2N$  muestras, luego al final de este proceso se obtienen  $N$  muestras que se extienden exactamente desde el principio hasta el final de la trama de TCX20 206 de la Figura 3. Entonces estas  $N$  muestras forman la entrada de una transformada apropiada para codificación eficiente en el dominio de la transformada. Usando la distorsión específica por repliegue espectral en el dominio del tiempo descrita en la Figura 3, la MDCT puede ser la transformada usada con este propósito.

Después de la combinación de partes desplazadas e invertidas en el tiempo de la ventana descrita en la Figura 3, ya no es posible recuperar las muestras originales en el dominio del tiempo en la trama de TCX20 debido a que están mezcladas con versiones invertidas en el tiempo de muestras fuera de la trama de TCX20. En un codificador de audio basado en MDCT tal como el AAC de MPEG, donde todas las tramas se codifican usando la misma transformada y ventanas superpuestas, se puede cancelar esta distorsión por repliegue espectral en el dominio del tiempo, y se pueden recuperar las muestras de audio usando dos tramas superpuestas consecutivas. No obstante, cuando las tramas contiguas no usan el mismo proceso de selección de ventana y superposición, como en la Figura 2, donde la trama de TCX20 (ventana no rectangular y superpuesta) está precedida por una trama de ACELP (ventana rectangular y no superpuesta), el efecto de la distorsión por repliegue espectral en el dominio del tiempo y de la ventana no rectangular no se puede eliminar usando solamente la información de la trama de ACELP anterior y la siguiente trama de TCX20.

Se presentaron en lo que antecede técnicas para gestionar este tipo de transición. La presente descripción propone un planteamiento alternativo para gestionar estas transiciones. Este planteamiento no usa ventanas no óptimas y asimétricas en las tramas donde se usa codificación en el dominio de la transformada basada en MDCT. En su lugar, el dispositivo y el método introducidos en la presente memoria permiten el uso de ventanas simétricas, centradas en el medio de la trama codificada, tal como por ejemplo la trama de TCX20 de la Figura 3, y con un 50% de superposición con tramas codificadas por MDCT usando también ventanas no rectangulares. El dispositivo y el método introducidos en la presente memoria proponen, de este modo, enviar desde el codificador al decodificador, como información adicional en el flujo de bits, información de corrección para cancelar el efecto de selección de ventana y la distorsión por repliegue espectral en el dominio del tiempo cuando se conmuta de tramas codificadas con una ventana rectangular y no superpuesta a tramas codificadas con una ventana no rectangular y superpuesta, y viceversa.

En la Figura 2, se muestra una selección de ventana rectangular y no superpuesta para una trama de ACELP, mientras que se muestra una ventana no rectangular y superpuesta para una trama de TCX20. Usando la TDA introducida en la Figura 3, un decodificador que recibe al principio los bits de la trama de ACELP tiene información suficiente para decodificar completamente esta trama de ACELP hasta su última muestra. Pero luego, recibiendo los bits de la trama de TCX20, decodificar correctamente todas las muestras en la trama de TCX20 se ve afectado por el efecto de distorsión por repliegue espectral en el tiempo causado por la presencia de la trama de ACELP anterior. Si la siguiente trama también usa una ventana superpuesta, entonces la selección de ventana no rectangular y la TDA introducida en el codificador se pueden cancelar en la segunda mitad de la trama de TCX20 mostrada y las muestras se pueden decodificar correctamente. Es, de este modo, en la primera mitad de la trama de TCX20 de la Figura 3, donde el 1º cuarto 204e invertido en el tiempo y desplazado se resta del 2º cuarto 204b, por lo que el efecto de la ventana no rectangular y la TDA introducida en el codificador no se puede cancelar dado que la trama de ACELP previa usa una ventana rectangular y no superpuesta.

El dispositivo y el método introducidos en la presente memoria proponen transmitir información adicional en forma de parámetros de Cancelación Directa de Distorsión por Repliegue Espectral (FAC), para cancelar estos efectos y para recuperar correctamente tramas de TCX

Una realización de particular interés usa Conformación de Ruido en el Dominio de la Frecuencia (FDNS), por ejemplo, como en la solicitud PCT N° PCT/CA2010/001649 presentada el 15 de octubre de 2010 y titulada "SIMULTANEOUS TIME-DOMAIN AND FREQUENCY-DOMAIN NOISE SHAPING FOR TDAC TRANSFORMERS" para conformar el ruido de cuantificación en tramas codificadas por transformada, tales como las tramas de TCX. En esta realización, la corrección de FAC se puede aplicar directamente en el dominio de señal original, tal como una señal de audio que no tiene ponderación aplicada a la misma. En un códec conmutado multimodo tal como USAC, esto implica que la conformación de ruido de cuantificación se realiza en el dominio de la transformada, por ejemplo, usando MDCT, en todos los modos de codificación que implican una transformada. Específicamente, en tramas de TCX que usan FDNS, la transformada (MDCT) se aplica directamente a la señal original (como en el modo de codificación por transformada perceptual) en lugar del residuo ponderado. La FDNS opera de tal forma que obtiene una conformación de ruido en tramas de TCX lo cual es esencialmente equivalente a usar el filtro de ponderación perceptual en el dominio del tiempo pero solamente operando en los coeficientes de transformada (MDCT). La corrección de FAC se puede aplicar entonces con el procedimiento descrito en lo sucesivo.

El códec de audio de USAC se usa en la presente memoria como ejemplo no limitante de un códec. Se han propuesto tres modos de codificación para el códec de USAC, de la siguiente manera:

Modo de codificación 1: Codificación por transformada perceptual de la señal de audio original;

Modo de codificación 2: Codificación por transformada del residuo ponderado de un filtro de LPC;

Modo de codificación 3: Codificación de ACELP.

5 En el modo de codificación 1, la conformación de ruido de cuantificación ya se logra en el dominio de la transformada a través de la aplicación de factores de escala derivados de un modelo perceptual, como es bien sabido por los expertos en la técnica de codificación de audio. No obstante, en el modo de codificación 2, la conformación de ruido de cuantificación se aplica típicamente en el dominio del tiempo usando un filtro perceptual, o ponderando,  $W(z)$  derivado de un filtro de codificación predictiva lineal (LPC) calculado para la trama actual. Una transformada, por ejemplo, una transformada DCT, se aplica después de este filtrado en el dominio del tiempo para  
10 obtener un objetivo de FAC sea cuantificado y codificado como parámetro de FAC. Esto evita unir tramas sucesivas codificadas en los modos 1 y 2 usando directamente las propiedades de Cancelación de Distorsión por Repliegue Espectral en el Dominio del Tiempo (TDAC) de la MDCT, dado que la MDCT no se aplica en el mismo dominio para los modos de codificación 1 y 2.

15 En consecuencia, en una realización del dispositivo y del método para cancelación directa de distorsión por repliegue espectral en el dominio del tiempo, la conformación de ruido de cuantificación para el modo de codificación 2 se hace a través del filtrado en el dominio de la frecuencia usando el proceso de FDNS de la solicitud PCT N° PCT/CA2010/001649, en lugar de filtrado en el dominio del tiempo. Por lo tanto, la transformada, que es por ejemplo MDCT en el caso de USAC, se aplica a la señal de audio original en lugar de a una versión ponderada de esa señal de audio a la salida del filtro  $W(z)$ . Esto asegura la uniformidad entre el modo de codificación 1 y el modo de  
20 codificación 2 y permite unir tramas sucesivas codificadas en los modos 1 y 2 usando la propiedad de TDAC de la MDCT.

No obstante, la aplicación de la conformación de ruido de cuantificación en el dominio de la transformada para el modo de codificación 2 usa un procesamiento especial cuando se manejan transiciones desde y hasta el modo de ACELP.

25 La Figura 4 es un diagrama esquemático de una secuencia de operaciones de un método ejemplar de cálculo de un objetivo de FAC. Se muestra el procesamiento en el codificador cuando una trama 402 codificada en el modo 2 está precedida por una trama 404 codificada en el modo 3 y seguida por una trama 406 codificada en el modo 3, en donde la ACELP se usa como ejemplo del modo 3 con propósitos de ilustración solamente. La Figura 4 muestra marcadores en el dominio del tiempo tales como 408 y límites de tramas. Los límites de trama identificados específicamente con los marcadores de línea de discontinua vertical LPC1 y LPC2 muestran el principio y el final de la trama 402, que está codificada en el modo 2. Los marcadores LPC1 y LPC2 indican además el centro de la ventana de análisis para calcular dos filtros LPC: un primer filtro de LPC se calcula al principio de la trama 402 (que también corresponde al punto de plegado izquierdo de la ventana) y un segundo filtro de LPC se calcula al final de la misma trama 402 (que también corresponde al punto de plegado derecho de la ventana).

35 Hay cuatro líneas (línea 1 a línea 4) en la Figura 4. Cada línea representa una operación en el procesamiento en el codificador. Como se ilustra, las líneas 1-4 de la Figura 4 están alineadas en el tiempo entre sí.

La línea 1 de la Figura 4 representa una señal de audio original 410, segmentada en tramas que están delimitadas por los marcadores LPC1 y LPC2. Por lo tanto, a la izquierda del marcador LPC1, la señal de audio original se codifica en el modo 3. Entre los marcadores LPC1 y LPC2, la señal de audio original se codifica en el modo 2, con conformación de ruido de cuantificación aplicada directamente en el dominio de la transformada usando el proceso de FDNS, por ejemplo, como en la solicitud PCT N° PCT/CA2010/001649 en lugar de en el dominio del tiempo. A la derecha del marcador LPC2, la señal de audio original se codifica de nuevo en el modo de codificación 3. Esta secuencia de modos de codificación, que implica a la ACELP en el modo 3, luego a la TCX en el modo 2, y luego de nuevo a la ACELP en el modo 3, se elige para ilustrar el procesamiento relacionado con ambas transiciones desde  
40 el modo 3 al modo 2 y desde el modo 2 al modo 3. En un códec multimodo, son posibles, por supuesto, otras secuencias de modo. Obviamente, la presente descripción no se limita a la secuencia de modo específica elegida en el ejemplo de la Figura 4.

La línea 2 de la Figura 4 corresponde a señales de síntesis decodificadas 412, 414, 416 en cada trama. A la izquierda, el marcador LPC1 está una señal de síntesis 414 de la trama 404 se ha codificado en modo 3. Por lo tanto, la señal de síntesis 414 se identifica como una señal de síntesis de ACELP. En principio, hay una alta similitud entre la señal de síntesis de ACELP 414 y la señal original en la trama 404, dado que el modo de codificación de ACELP intenta codificar y sintetizar la señal de audio tan precisamente como sea posible. Luego, la trama 402 entre los marcadores LPC1 y LPC2 en la línea 2 de la Figura 4 representa una señal de síntesis 412 obtenida como salida de una MDCT inversa (IMDCT) aplicada a la trama correspondiente. La Figura 4 describe una realización en la que  
55 la conformación de ruido de cuantificación en la trama de Codificación por Transformada (TC) 402 se logra en el dominio de la transformada. Esto se puede lograr, por ejemplo, filtrando los coeficientes de MDCT usando la información espectral del primer y segundo filtros LPC calculados mencionados anteriormente, como se ha explicado anteriormente, en los límites de trama o marcadores LPC1 y LPC2. También, la señal de síntesis 412 contiene un

efecto de selección de ventana y efecto de distorsión por repliegue espectral en el dominio del tiempo, o de plegado, al principio y al final de la trama 402. Este efecto de plegado está formado por partes de síntesis de ACELP mostrada en pequeñas ventanas y plegada 418 y 420 de las tramas 404 y 406, respectivamente. Las partes de síntesis de ACELP mostrada en pequeñas ventanas y plegada 418 y 420 forman dos partes de una señal de error de codificación por transformada. La curva superior de la señal de síntesis 412, que se extiende desde el principio hasta el final de la trama 402, muestra el efecto de selección de ventana en la señal de síntesis 412, que es relativamente plana en la parte intermedia, pero no al principio y al final de la trama 402. El efecto de plegado se muestra por las partes de síntesis de ACELP mostrada en pequeñas ventanas y plegada 418 y 420 al principio y al final de la trama 402, respectivamente. El signo “-” asociado con la parte de síntesis de ACELP mostrada en pequeñas ventanas y plegada 418 al principio de la trama 402, indica una resta de esa parte de síntesis de ACELP mostrada en pequeñas ventanas y plegada 418 de la señal de síntesis 412, mientras que el signo “+” asociado con la parte de síntesis de ACELP mostrada en pequeñas ventanas y plegada 420 al final de la trama 402 indica una suma de esa parte de síntesis de ACELP mostrada en pequeñas ventanas y plegada 420 a la señal de síntesis 412. Este efecto de selección de ventana y efecto de distorsión por repliegue espectral en el dominio del tiempo, o plegado, son inherentes a la MDCT. Esta señal de error de codificación por transformada se puede cancelar cuando se codifican tramas consecutivas usando la MDCT, como se ha explicado anteriormente. No obstante, en el caso donde una trama codificada por MDCT no esté precedida y/o seguida por otras tramas codificadas por MDCT, este efecto de selección de ventana y efecto de distorsión por repliegue espectral en el dominio del tiempo, o de plegado, no se cancelan y permanece en la señal en el dominio del tiempo después de la IMDCT. Se puede usar entonces FAC para corregir estos efectos. Finalmente, la trama 406 después del marcador LPC2 en la Figura 4 también se codifica en el modo 3, usando por ejemplo la ACELP. Para obtener la señal de síntesis 416 en esa trama 406, los estados de filtro en la memoria de los predictores a largo plazo y corto plazo al principio de la trama 406 se establecen de la manera descrita a continuación, lo que implica que los efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo, o de plegado al final de trama 402 anterior, entre los marcadores LPC1 y LPC2, se cancelan por la aplicación de FAC. Para resumir, la línea 2 en la Figura 4 contiene las señales de síntesis 414, 412, 416 de las tramas 404, 402, 406 consecutivas, incluyendo las partes de la señal de error de codificación por transformada 418, 420 causadas por la selección de ventana y la distorsión por repliegue espectral en el dominio del tiempo a la salida del IMDCT en la trama 402 entre los marcadores LPC1 y LPC2.

Entonces, los detalles de la codificación de ACELP ejemplar se pueden usar para aliviar al menos en parte la señal de error de codificación por transformada inducida al principio de la señal de síntesis 412. Una predicción para el uso en la reducción de una energía de la señal de error de codificación por transformada se muestra en la línea 3 de la Figura 4. La predicción se basa en una estimación de que en una salida de síntesis de ACELP final, se ha usado la ACELP al principio de la trama 402. La predicción se basa en una similitud automática esperada de la señal de audio original 410 inmediatamente antes y después del marcador LPC1 y se puede obtener de la siguiente manera:

Al principio de la trama 402 entre los marcadores LPC1 y LPC2 de la línea 3, se pueden colocar dos contribuciones de los estados del filtro de síntesis de ACELP inmediatamente a la izquierda del marcador LPC1. Una primera contribución 422 comprende una versión mostrada en pequeñas ventanas e invertida en el tiempo, o plegada, de las últimas muestras de síntesis de ACELP de la trama 404. La longitud y la forma de ventana para esta señal invertida en el tiempo 422 es la misma que la parte de síntesis de ACELP mostrada en pequeñas ventanas y plegada 418 en el lado izquierdo de la trama de Codificación por Transformada (TC) decodificada 402 en la línea 2. Esta contribución 422 representa una buena aproximación de la distorsión por repliegue espectral en el dominio del tiempo presente en la trama de TC de la línea 2. Una segunda contribución 424 comprende una respuesta de entrada cero (ZIR) mostrada en pequeñas ventanas del filtro de síntesis de ACELP, con estados iniciales tomados como los estados finales de este filtro al final de la trama de síntesis de ACELP 404, inmediatamente a la izquierda del marcador LPC1. La longitud y la forma de ventana de esta segunda contribución 424 se toma como el complemento del cuadrado de la ventana de transformada usada en la trama codificada por transformada que, en el caso ejemplar de USAC, es la MDCT.

Entonces, habiendo colocado opcionalmente estas dos contribuciones de predicción (síntesis de ACELP mostrada en pequeñas ventanas y plegada 422 y ZIR de ACELP 424 mostrada en pequeñas ventanas) en la línea 3, la línea 4 se obtiene restando la línea 2 y la línea 3 de la línea 1, usando los sumadores 426 y 427. Se debería observar que la diferencia calculada durante esta operación se detiene en el marcador LPC2. Una vista aproximada de la envolvente en el dominio del tiempo esperada de la señal de error de codificación por transformada se muestra en la línea 4. Se espera que la envolvente en el dominio del tiempo de un error de codificación de ACELP 430 en la trama de ACELP 404 sea aproximadamente plana en amplitud, a condición de que la señal codificada sea estacionaria para esta duración. Entonces, se espera que la envolvente en el dominio del tiempo del error de codificación por transformada en la trama de TC 402, entre los marcadores LPC1 y LPC2, presente la forma general que se muestra en esta trama en la línea 4. Esta forma esperada de la envolvente en el dominio del tiempo del error de codificación por transformada solamente se muestra aquí con propósitos de ilustración y puede variar dependiendo de la señal codificada en la trama de TC entre los marcadores LPC1 y LPC2. Esta ilustración de la envolvente en el dominio del tiempo del error de codificación por transformada expresa que se espera que sea relativamente grande cerca del principio y del final de la trama de TC 402, entre los marcadores LPC1 y LPC2. Al principio de la trama 402, donde se muestra una primera parte de objetivo de FAC 432, el error de codificación por transformada se reduce usando las dos contribuciones de predicción de ACELP 422, 424, mostradas en la línea 3. Esta reducción no está presente

al final de la trama de TC 402, donde se muestra una segunda parte de objetivo de FAC 434. En la segunda parte de objetivo de FAC 434, los efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo no se pueden reducir usando la síntesis de la siguiente trama, que comienza después del marcador LPC2, dado que la trama de TC 402 necesita ser decodificada antes de que se pueda decodificar la siguiente trama.

5 El ruido de cuantificación puede ser típicamente como la envolvente esperada de la señal de error mostrada en la línea 4 de la Figura 4 cuando el decodificador usa solamente las señales de síntesis 414, 412, 416 de la línea 2 para producir la señal de audio decodificada. Este error proviene de los efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo inherentes a un par MDCT/IMDCT. Los efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo se han reducido al principio de la trama de TC 402, sumando las dos contribuciones de la trama de ACELP anterior 404 expuesta anteriormente, pero no se pueden cancelar completamente como en la operación de TDAC real de la MDCT, cuando se usa TC como el único modo de codificación. Además, a la derecha de la trama de TC en la línea 4 de la Figura 4, justo antes del marcador LPC2, todos los efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo permanecen del par MDCT/IMDCT. Las partes de alta amplitud 432 y 434 de la señal de error de codificación de la línea 4, al principio y al final de la trama de TC 402, constituyen ambas partes de objetivo de FAC, que es el objeto de la corrección de FAC.

De este modo, se entiende que los parámetros para la corrección de FAC han de ser enviados al decodificador para compensar esta señal de error de codificación, lo cual afecta al principio y al final de la trama de TC 402. Los efectos de selección de ventana y de distorsión por repliegue espectral se cancelan de una manera que mantiene el ruido de cuantificación en un nivel correcto, similar al de la trama de ACELP, y eso evita discontinuidades en los límites entre la trama de TC 402 y las tramas codificadas en otros modos tales como 404 y 406. Estos efectos se pueden cancelar usando FAC en el dominio de la frecuencia. Esto se logra filtrando los coeficientes de MDCT usando información derivada del primer y segundo filtros LPC calculados en los límites LPC1 y LPC2, aunque también se puede usar otra Conformación de Ruido en el Dominio de la Frecuencia (FDNS).

25 Para compensar de manera eficiente los efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo al principio y al final de la trama de TC 402 en la línea 4 de la Figura 4, se aplica la FAC siguiendo el procesamiento descrito en la Figura 4. La Figura 5 es un diagrama de bloques que muestra la cuantificación del objetivo de FAC de la Figura 4. La cuantificación como se muestra en la Figura 5 es de particular interés en el caso del proceso de FDNS, por ejemplo, como en la solicitud PCT N° PCT/CA2010/001649. La FAC cuantifica el error de codificación por transformada en el dominio ponderado usando LPC en el límite de trama. Una discontinuidad potencial debida a la cuantificación se enmascara entonces por filtrado inverso. Este procesamiento se describe tanto para la parte izquierda de la trama de TC 402, alrededor del marcador LPC1, como para la parte derecha de la trama de TC 402, alrededor del marcador LPC2. Como se mencionó anteriormente, la trama de TC 402 de la Figura 4 está precedida por una trama de ACELP 404, en el límite del marcador LPC1, y seguida por una trama de ACELP 406, en el límite del marcador LPC2.

Para compensar los efectos de selección de ventanas y de distorsión por repliegue espectral en el dominio del tiempo alrededor del marcador LPC1, el procesamiento puede ser como se describe en la parte superior de la Figura 5. Primero, en el caso de FDNS, un filtro de ponderación  $W_1(z)$  501 se puede calcular a partir del primer filtro de LPC calculado en el límite de trama de LPC1, o a partir de un filtro de LPC interpolado usando tanto el primer filtro de LPC calculado en el límite de trama de LPC1 como el segundo filtro de LPC calculado en el límite de trama de LPC2. La primera parte de objetivo de FAC 432, desde el principio de la trama de TC 402 en la línea 4 de la Figura 4, se filtra a través del filtro de ponderación  $W_1(z)$  501. El filtro de ponderación  $W_1(z)$  tiene como estado inicial, o memoria de filtro, constituido por el error de ACELP 430 mostrado en la línea 4 de la Figura 4. La salida del filtro  $W_1(z)$  501 de la Figura 5 entonces forma la entrada de una transformada, por ejemplo, una DCT 502. Los coeficientes de transformada de la DCT 502 se cuantifican entonces en el cuantificador Q 503 y se pueden codificar además en el cuantificador Q 503. Estos coeficientes codificados se transmiten entonces a un decodificador como parámetros de FAC. Los parámetros de FAC comprenden el coeficiente de DCT cuantificado, que entonces llega a ser, en el decodificador, la entrada de una transformada inversa, por ejemplo, una IDCT 504, usada para formar una señal en el dominio del tiempo. Esta señal en el dominio del tiempo se puede filtrar entonces a través del filtro inverso  $1/W_1(z)$  505 que tiene un estado inicial cero. El filtrado a través del filtro inverso  $1/W_1(z)$  505 se extiende pasada la longitud de la primera parte de objetivo de FAC 432 usando una entrada cero para las muestras que se extienden después de la primera parte de objetivo de FAC 432. La salida del filtro inverso  $1/W_1(z)$  es una primera parte de síntesis de FAC 506, que es una señal de corrección que ahora se puede aplicar al principio de la trama de TC 402 para compensar los efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo.

55 Ahora, volviendo al procesamiento para corrección de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo al final de la trama de TC 402, antes del marcador LPC2, se considera la parte inferior de la Figura 5. La segunda parte de objetivo de FAC 434, al final de la trama de TC 402 en la línea 4 de la Figura 4, se puede filtrar a través de un filtro de ponderación  $W_2(z)$  calculado a partir del segundo filtro de LPC calculado en el límite de trama de LPC2 o un filtro de LPC interpolado usando tanto el primer filtro de LPC calculado en el límite de trama de LPC1 como el segundo filtro de LPC calculado en el límite de filtro de LPC2. El segundo filtro de LPC calculado en el límite de trama de LPC2 tiene como estado inicial, o memoria de filtro, formado por el error de codificación por transformada en la trama de TC en la línea 4 de la Figura 4. Entonces, todas las operaciones de

procesamiento adicionales son las mismas que para la parte superior de la Figura 5 (véase la DCT 508, el cuantificador Q 509, la IDCT 510 y el filtro de ponderación inverso  $1/W_2(z)$  511), que trató con el procesamiento del objetivo de FAC al principio de la trama de TC 402, excepto por el uso del filtro de ponderación  $W_2(z)$  en lugar del filtro de ponderación  $W_1(z)$ ), que proporciona una segunda parte de síntesis de FAC 512.

- 5 Todo el proceso de la Figura 5 se realiza cuando se aplica en el codificador, con el fin de obtener la síntesis de FAC local. En el decodificador, el procesamiento de la Figura 5 solamente se aplica desde un punto donde los parámetros de FAC, recibidos del cuantificador Q 503 o 509 del codificador, se usan como entrada en la IDCT. Esto también produce una síntesis de FAC en el decodificador.

- 10 La Figura 6 es un diagrama esquemático de una secuencia de operaciones de un método ejemplar de cálculo de una síntesis de una señal de audio original, usando parámetros de FAC representativos del objetivo de FAC de la Figura 4. El cálculo de la síntesis se hace en el dominio original usando la FAC. El uso de LPC permite que la FAC se use en el contexto de FDNS, por ejemplo, como se describe en la solicitud PCT N° PCT/CA2010/001649 presentada el 15 de octubre de 2010 y titulada "SIMULTANEOUS TIME-DOMAIN AND FREQUENCY-DOMAIN NOISE SHAPING FOR TDAC TRANSFORMS". Las discontinuidades potenciales se enmascaran por el filtrado
- 15 inverso, como se hace en el contexto de TCX usando LPC. La Figura 6 muestra cómo se puede obtener una señal de síntesis completa 604, 602, 606 usando la síntesis de FAC como se muestra en la Figura 5 y aplicando una inversa de las operaciones de la Figura 4. En la Figura 6, la trama de ACELP 404 a la izquierda del marcador LPC1 se sintetiza ya hasta el marcador LPC1, mostrado como la síntesis de ACELP 604 en la línea B. La trama 406 después del marcador LPC2 también es una trama de ACELP. Entonces, para producir una señal de síntesis 602 en
- 20 la trama de TC 402, entre los marcadores LPC1 y LPC2, se realizan las siguientes etapas:

La trama de TC codificada por MDCT 402 recibida se decodifica por IMDCT y la señal en el dominio del tiempo 608 resultante se produce entre los marcadores LPC1 y LPC2 como se muestra en la línea B de la Figura 6. Esta trama de TC 402 decodificada contiene efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo 610, 612.

- 25 La señal de síntesis de FAC 506, 512 como en la Figura 5 se coloca al principio y al final de la trama de TC 402. Más específicamente, los parámetros de FAC recibidos se decodifican, si es aplicable, por transformada inversa, por ejemplo, usando IDCT (504, 510), y se filtran usando el filtro  $1/W_1(z)$  505 para la primera parte 506 y el filtro  $1/W_2(z)$  511 para la segunda parte 512. Esto produce dos partes de síntesis de FAC 506, 512 como se ilustra en la Figura 5. La primera parte de síntesis de FAC 506 está colocada al principio de la trama de TC 402 en la línea A, y la segunda
- 30 parte de síntesis de FAC 512 está colocada al final de la trama de TC 402 en la línea A.

La síntesis de ACELP mostrada en pequeñas ventanas y plegada (invertida en el tiempo) 618 de la trama de ACELP 404 que precede a la trama de TC 402 y la ZIR 620 del filtro de síntesis de ACELP se colocan al principio de la trama de TC 402. Esto se muestra en la línea C.

- 35 Las líneas A, B y C se suman a través de los sumadores 622 y 624 para formar la señal de síntesis 602 para la trama de TC en el dominio original en la línea D. Este procesamiento ha producido, en la trama de TC 402, la señal de síntesis 602 donde los efectos de distorsión por repliegue espectral en el dominio del tiempo y de selección de ventana se han cancelado al principio y al final de la trama 402, y donde la discontinuidad potencial en el límite de trama alrededor del marcador LPC1 además puede haber sido suavizada y enmascarada perceptualmente por los filtros  $1/W_1(z)$  505 y  $1/W_2(z)$  511 de la Figura 5.

- 40 Por supuesto, la suma de las señales de las líneas A a C se puede realizar en cualquier orden sin cambiar el resultado del procesamiento descrito.

La FAC también se puede aplicar directamente a la salida de síntesis de la trama de TC sin ninguna selección de ventana en el decodificador. En este caso, la forma de la FAC está adaptada para tener en cuenta la diferente selección de ventana (o la falta de selección de ventana) de la trama de TC 402 decodificada.

- 45 La longitud de la trama de FAC se puede cambiar durante la codificación. Por ejemplo, longitudes de trama ejemplares pueden ser 64 o 128 muestras, dependiendo de la naturaleza de la señal. Por ejemplo, se puede usar una trama de FAC más corta en el caso de señales sin voz. Información acerca de la longitud de la trama de FAC se puede señalar al decodificador, usando por ejemplo un indicador de 1 bit, o una bandera, para indicar tramas de 64 o 128 muestras. Un ejemplo de secuencia de transmisión con longitud de FAC de señalización puede comprender el
- 50 siguiente juego:

- TC con superposición (256 bits)
- FAC + longitud de FAC de señalización (128 bits)
- ACELP
- FAC + longitud de FAC de señalización (64 bits)

- TC con superposición (128 bits)

Se puede transmitir información de señalización adicional para indicar ciertas funciones de procesamiento a ser realizadas por el decodificador. Un ejemplo es la señalización de la activación del procesamiento posterior, específica de las tramas ACELP. El procesamiento posterior se puede encender o apagar durante un cierto período que consiste en varias tramas de ACELP consecutivas. En una transición de TC a ACELP, se puede incluir una bandera de 1 bit dentro de la información de FAC para señalar la activación del procesamiento posterior. En una realización, esta bandera solamente se transmite en una primera trama en una secuencia de varias tramas de ACELP. De este modo, la bandera se puede añadir a la información de FAC, que también se envía para la primera trama de ACELP.

La Figura 7 es un diagrama de bloques de un ejemplo no limitativo de dispositivo para cancelación directa de distorsión por repliegue espectral en el dominio del tiempo en una señal de audio codificada recibida en un flujo de bits. Un dispositivo 700 se da, con el propósito de ilustración, con referencia al objetivo de FAC de las Figuras 5 y 6, usando información del modo de ACELP. Los expertos en la técnica apreciarán que un dispositivo 700 correspondiente se puede implementar en relación con todos los demás ejemplos de modos de codificación y corrección de FAC dados en la presente descripción.

El dispositivo 700 comprende un receptor 710 para recibir un flujo de bits 701 representativo de una señal de audio codificada incluyendo los parámetros de FAC representativos del objetivo de FAC.

Los parámetros (prm) para las tramas de ACELP del flujo de bits 701 se suministran desde el receptor 710 a un decodificador de ACELP 711 incluyendo un filtro de síntesis de ACELP. El decodificador de ACELP 711 produce una respuesta de entrada cero (ZIR) 704 del filtro de síntesis de ACELP. También, el decodificador de síntesis de ACELP 711 produce una señal de síntesis de ACELP 702. La señal de síntesis de ACELP 702 y la ZIR 704 se concatenan para formar una señal de síntesis de ACELP seguida por la ZIR. Una ventana de FAC 703, que tiene características que coinciden con la selección de ventana aplicadas en la Figura 6, línea C, se aplica entonces a las señales 707 y 704 concatenadas. La señal de síntesis de ACELP 707 se muestra en pequeñas ventanas y se pliega para producir la síntesis de ACELP 618 de la línea C de la Figura 6 mientras que la ZIR 704 se muestra en pequeñas ventanas para producir la ZIR de ACELP 620 de la Figura 6. Ambas se suman en el procesador 705, y luego se aplican a una entrada positiva de un sumador 720 para proporcionar una primera parte (opcional) de la señal de audio en las tramas de TCX.

Los parámetros (prm) para las tramas de TCX 20 del flujo de bits 701 se suministran a un decodificador de TCX 706, seguido por una transformada IMDCT 713 y una ventana 714 para la IMDCT, para producir una señal de síntesis de TCX 702 (véase 608, 610 y 612 de la línea B de la Figura 6) aplicada a una entrada positiva de un sumador 716 para proporcionar una segunda parte de la señal de audio en las tramas de TCX 20.

No obstante, en una transición entre los modos de codificación (por ejemplo, desde una trama de ACELP hasta una trama de TCX 20), una parte de la señal de audio no se decodificaría correctamente sin el uso de un procesador de FAC 715. En el ejemplo de la Figura 7, el procesador de FAC 715 comprende un decodificador de FAC 717 para decodificar a partir del flujo de bits 701 recibido los parámetros de FAC (salida de la DCT 502 y 508 de la Figura 5), que corresponde al objetivo de FAC después del filtrado (véanse los filtros 501 y 507 de la Figura 5) y la transformada DCT (véase la DCT 502 y 508 de la Figura 5), como se produce por el cuantificador Q (503, 509) de la Figura 5. Una IDCT 718 (correspondiente a la IDCT 504 y 505 en la Figura 5) aplica una DCT inversa a los parámetros de FAC decodificados del decodificador 717, y la salida de la IDCT 718 se suministra a una entrada positiva del sumador 720. La salida del sumador 720 se suministra a un filtro 719, que aplica características del filtro de ponderación inverso  $1/W_1(z)$  (505 a la Figura 5) a una primera parte (correspondiente a 432 en la Figura 5) del objetivo de FAC y los del filtro de ponderación inverso  $1/W_2(z)$  (511 en la Figura 5) a una segunda parte (correspondiente a 434 en la Figura 5) del objetivo de FAC. La salida del filtro 719 se suministra a una entrada positiva del sumador 716.

La salida global del sumador 716 representa la señal de síntesis cancelada por FAC (602 en la Figura 6) para una trama de TCX que sigue a una trama de ACELP.

La Figura 8 es un diagrama de bloques esquemático de un ejemplo no limitativo del dispositivo 800 para cancelación directa de distorsión por repliegue espectral en el dominio del tiempo en una señal codificada para su transmisión a un decodificador. El dispositivo 800 se da, con el propósito de ilustración, con referencia al objetivo de FAC de las Figuras 4 y 5, usando información del modo de ACELP. Los expertos en la técnica apreciarán que un dispositivo 800 correspondiente se puede implementar en relación con todos los demás ejemplos de modos de codificación y corrección de FAC dados en la presente descripción.

Una señal de audio 801 a ser codificada se aplica al dispositivo 800. Una lógica (no mostrada) aplica tramas de ACELP de la señal de audio 801 a un codificador de ACELP 810. Una salida del codificador de ACELP 810, los parámetros codificados por ACELP 802, se aplica a una primera entrada de un multiplexor (MUX) 811 para su transmisión a un receptor (no mostrado). Otra salida del codificador de ACELP es una señal de síntesis de ACELP 860 seguida de la respuesta de entrada cero (ZIR) 861 de un filtro de síntesis de ACELP que forma parte del

codificador de ACELP 810. Una ventana de FAC 805 que tiene características que coinciden con la selección de ventana aplicada en la Figura 4, línea 3, se aplica por un procesador de ventana de FAC 805 a la concatenación de señales 860 y 861. La salida (correspondiente a la Figura 4, línea 3) del procesador de ventana de FAC 805 se aplica a una entrada negativa de un sumador 851 (correspondiente al sumador 427 de la Figura 4).

5 La lógica (no mostrada) también aplica tramas de TCX 20 (véase la línea 402 de la Figura 4) de la señal de audio 801 a un módulo de codificación de MDCT 812 para producir los parámetros codificados de TCX 20 803 aplicados a una segunda entrada del multiplexor 811 para su transmisión a un receptor (no mostrado). El módulo de codificación de MDCT 812 comprende una ventana de MDCT 831, una transformada MDCT 832, y un cuantificador 833. La señal de audio 801 se muestra en pequeñas ventanas mediante la ventana de MDCT 831 y la señal mostrada en  
10 pequeñas ventanas de MDCT se suministra desde la ventana de MDCT 831 a una entrada positiva de un sumador 850 (correspondiente al sumador 426 de la Figura 4). La señal mostrada con pequeñas ventanas de MDCT de la ventana de MDCT 831 también se suministra a una MDCT para producir coeficientes MDCT suministrados a un cuantificador 833 para producir el parámetro de TCX 803 y los coeficientes de MDCT cuantificados 804 aplicados a una MDCT inversa (IMDCT) 833. La salida de la IMDCT 833 es una señal de síntesis (correspondiente a la señal de  
15 síntesis 412 de la Figura 4) suministrada a una entrada negativa del sumador 850 (correspondiente al sumador 426 de la Figura 4). La salida del sumador 850 forma un error de cuantificación de TCX, que se muestra en pequeñas ventanas en el procesador 836. La salida del procesador 836 se suministra a una entrada positiva del sumador 851.

Tras una transición entre modos de codificación (por ejemplo, de una trama de ACELP a una trama de TCX 20), algunas de las tramas de audio codificadas por el módulo de MDCT 812 pueden no ser decodificadas correctamente sin información adicional. Un calculador 813 proporciona esta información adicional, más específicamente un  
20 objetivo de FAC codificado y cuantificado. Todos los componentes del calculador 813 se pueden ver como un productor de los parámetros de FAC 806. La salida del sumador 851 es el objetivo de FAC (correspondiente a la línea 4 de la Figura 4). El objetivo de FAC se introduce en un filtro 808, que aplica las características del filtro de ponderación  $W_1(z)$  501 (Figura 5) a la primera parte 432 del objetivo de FAC y las del filtro de ponderación  $W_2(z)$  507 (Figura 5) a la segunda parte 434 del objetivo de FAC. La salida del filtro 804 se aplica entonces a la DCT 834  
25 (correspondiente a la DCT 502 y 508 de la Figura 5), seguido por la cuantificación de la salida de la DCT 834 en el cuantificador 837 (correspondiente a los cuantificadores 503 y 509 de la Figura 5) para producir los parámetros de FAC 806 que se aplican a una entrada del multiplexor 811 para su transmisión a un receptor (no mostrado).

La señal a la salida del multiplexor 811 representa la señal de audio codificada 855 a ser transmitida a un receptor  
30 (no mostrado) a través de un transmisor 856 en un flujo de bits codificado 857.

Los expertos en la técnica se darán cuenta de que la descripción del dispositivo y del método para cancelación directa de distorsión por repliegue espectral en el dominio del tiempo en una señal codificada es solamente  
35 ilustrativa y no se pretende que sea limitante de ninguna forma. Otras realizaciones se sugerirán fácilmente por sí mismas a tales personas expertas en la técnica que tienen el beneficio de esta descripción. Además, el dispositivo y el método descritos se pueden personalizar para ofrecer soluciones valiosas a las necesidades y los problemas existentes de cancelación de distorsión por repliegue espectral en el dominio del tiempo en una señal codificada.

Los expertos en la técnica también apreciarán que numerosos tipos de terminales u otros aparatos pueden incorporar tanto aspectos de codificación para transmisión de audio codificado, como aspectos de decodificación que sigue a la recepción de audio codificado, en un mismo dispositivo.

40 En interés de la claridad, no se muestran y describen todas las características rutinarias de las implementaciones de cancelación directa de distorsión por repliegue espectral en el dominio del tiempo en una señal codificada. Se apreciará, por supuesto, que en el desarrollo de cualquier implementación real de la codificación de audio, se deben tomar numerosas decisiones específicas de la implementación con el fin de lograr las metas específicas del desarrollador, tales como el cumplimiento de las restricciones relacionadas con la aplicación, el sistema, la red y el  
45 negocio, y que estas metas específicas variarán de una implementación a otra y de un desarrollador a otro. Además, se apreciará que un esfuerzo de desarrollo podría ser complejo y requerir mucho tiempo, pero, sin embargo, sería una tarea rutinaria de ingeniería para los expertos en el campo de los sistemas de codificación de audio que tienen el beneficio de esta descripción.

Según esta descripción, los componentes, las etapas de proceso y/o las estructuras de datos descritos en la  
50 presente memoria se pueden implementar usando diversos tipos de sistemas operativos, plataformas informáticas, dispositivos de red, programas de ordenador y/o máquinas de propósito general. Además, los expertos en la técnica reconocerán que también se pueden usar dispositivos de una naturaleza de propósito menos general, tales como dispositivos cableados, agrupaciones de puertas programables en campo (FPGA), circuitos integrados de aplicaciones específicas (ASIC) o similares. Cuando un método que comprende una serie de etapas de proceso se  
55 implementa mediante un ordenador o una máquina y esas etapas de proceso se pueden almacenar como una serie de instrucciones legibles por la máquina, se pueden almacenar en un medio tangible.

Los sistemas y módulos descritos en la presente memoria pueden comprender software, microprograma, hardware o cualquier combinación o combinaciones de software, microprograma o hardware adecuado para los propósitos descritos en la presente memoria. El software y otros módulos pueden residir en servidores, estaciones de trabajo,

5 ordenadores personales, tabletas computarizadas, PDA y otros dispositivos adecuados para los propósitos descritos en la presente memoria. El software y otros módulos pueden ser accesibles a través de la memoria local, a través de una red, a través de un navegador u otra aplicación en un contexto de ASP o a través de otros medios adecuados para los propósitos descritos en la presente memoria. Las estructuras de datos descritas en la presente memoria pueden comprender archivos de ordenador, variables, agrupaciones de programación, estructuras de programación o cualquier esquema o método de almacenamiento electrónico de información, o cualquier combinación de los mismos, adecuados para los propósitos descritos en la presente memoria.

## REIVINDICACIONES

1. Un método de producción de una señal de síntesis en una primera trama codificada por transformada de coseno discreta modificada, MDCT, que se extiende desde un límite de trama de LPC1 hasta un límite de trama de LPC2, que comprende:

5 (a) recibir la trama codificada por MDCT que se extiende desde el límite de trama de LPC1 hasta el límite de trama de LPC2, y (b) decodificar la primera trama aplicando una transformada de coseno discreta inversa modificada, IMDCT, a la primera trama para producir entre los límites de trama de LPC1 y LPC2 una señal en el dominio del tiempo que contiene efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo;

10 (a) recibir parámetros de cancelación directa de distorsión por repliegue espectral directa, FAC, (b) decodificar los parámetros de FAC para producir una primera parte de una señal de síntesis de FAC al principio de la primera trama y una segunda parte de la señal de síntesis de FAC al final de la primera trama, mediante (i) transformada inversa de los parámetros de FAC, y (ii) filtrado de los parámetros de FAC de transformada inversa usando un filtro inverso  $1/W_1(z)$  con estado inicial cero para la primera parte de la señal de síntesis de FAC y un filtro inverso  $1/W_2(z)$  con estado inicial cero para la segunda parte de la señal de síntesis de FAC, en donde  $W_1(z)$  es un filtro de ponderación calculado a partir de un primer filtro de LPC calculado en el límite de trama de LPC1 o a partir de un filtro de LPC interpolado usando tanto el primer filtro de LPC calculado en el límite de trama de LPC1 y un segundo filtro de LPC calculado en el límite de trama de LPC2, y en donde  $W_2(z)$  es un filtro de ponderación calculado a partir del segundo filtro de LPC calculado en el límite de trama de LPC2 o a partir de un filtro de LPC interpolado usando tanto el primer filtro de LPC calculado en el límite de trama de LPC1 y el segundo filtro de LPC calculado en el límite de trama de LPC2;

colocar al principio de la primera trama una síntesis de ACELP mostrada en pequeñas ventanas e invertida en el tiempo a partir de la trama de ACELP que precede a la primera trama y una respuesta de entrada cero mostrada en pequeñas ventanas del filtro de síntesis de ACELP calculada en el límite de trama de LPC1;

25 añadir (a) la señal en el dominio del tiempo que contiene efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo, (b) la primera y segunda partes de la señal de síntesis de FAC, (c) la síntesis de ACELP mostrada en pequeñas ventanas e invertida en el tiempo, y (d) la respuesta de entrada cero mostrada en pequeñas ventanas del filtro de síntesis de ACELP para formar la señal de audio de síntesis en la primera trama en el dominio del tiempo donde los efectos de distorsión por repliegue espectral en el dominio del tiempo y de selección de ventana se han cancelado al principio y al final de la primera trama, y en donde la discontinuidad potencial en el límite de trama de LPC1 se ha suavizado y enmascarado perceptualmente por los filtros inversos  $1/W_1(z)$  y  $1/W_2(z)$ .

2. Un dispositivo para producir una señal de síntesis en una primera trama codificada por transformada de coseno discreta modificada, MDCT, que se extiende desde un límite de trama de LPC1 hasta un límite de trama de LPC2, que comprende:

35 medios para (a) recibir la trama codificada por MDCT que se extiende desde el límite de trama de LPC1 hasta el límite de trama de LPC2, y (b) decodificar la primera trama aplicando una transformada de coseno discreta inversa modificada, IMDCT, a la primera trama para producir entre los límites de trama de LPC1 y LPC2 una señal en el dominio del tiempo que contiene efectos de selección de ventanas y de distorsión por repliegue espectral en el dominio del tiempo;

40 medios para (a) recibir parámetros de cancelación directa de distorsión por repliegue espectral, FAC, (b) decodificar los parámetros de FAC para producir una primera parte de una señal de síntesis de FAC al principio de la primera trama y una segunda parte de la señal de síntesis de FAC al final de la primera trama, mediante (i) transformada inversa de los parámetros de FAC, y (ii) filtrado de los parámetros de FAC de transformada inversa usando un filtro inverso  $1/W_1(z)$  con estado inicial cero para la primera parte de la señal de síntesis de FAC, y un filtro inverso  $1/W_2(z)$  con estado inicial cero para la segunda parte de la señal de síntesis de FAC, en donde  $W_1(z)$  es un filtro de ponderación calculado a partir de un primer filtro de LPC calculado en el límite de trama de LPC1 o a partir de un filtro de LPC interpolado usando tanto el primer filtro de LPC calculado en el límite de trama de LPC1 como un segundo filtro de LPC calculado en el límite de trama de LPC2, y en donde  $W_2(z)$  es un filtro de ponderación calculado a partir del segundo filtro de LPC calculado en el límite de trama de LPC2 o a partir de un filtro de LPC interpolado usando tanto el primer filtro de LPC calculado en el límite de trama de LPC1 como el segundo filtro de LPC calculado en el límite de trama de LPC2;

50 medios para colocar al principio de la primera trama una síntesis de ACELP mostrada con pequeñas ventanas e invertida en el tiempo a partir de la trama de ACELP que precede a la primera trama y una respuesta de entrada cero mostrada con pequeñas ventanas del filtro de síntesis de ACELP calculada en el límite de trama de LPC1;

55 medios para añadir (a) la señal en el dominio del tiempo que contiene efectos de selección de ventana y de distorsión por repliegue espectral en el dominio del tiempo, (b) la primera y segunda partes de la señal de síntesis de FAC, (c) la síntesis de ACELP mostrada con pequeñas ventanas e invertida en el tiempo, y (d) la

respuesta de entrada cero mostrada con pequeñas ventanas del filtro de síntesis de ACELP para formar la señal de síntesis de audio en la primera trama en el dominio del tiempo donde se han cancelado los efectos de distorsión por repliegue espectral en el dominio del tiempo y selección de ventana al principio y al final de la primera trama, y en donde la discontinuidad potencial en el límite de trama de LPC1 se ha suavizado y enmascarado perceptualmente por los filtros inversos  $1/W_1(z)$  y  $1/W_2(z)$ .

5

Figura 1

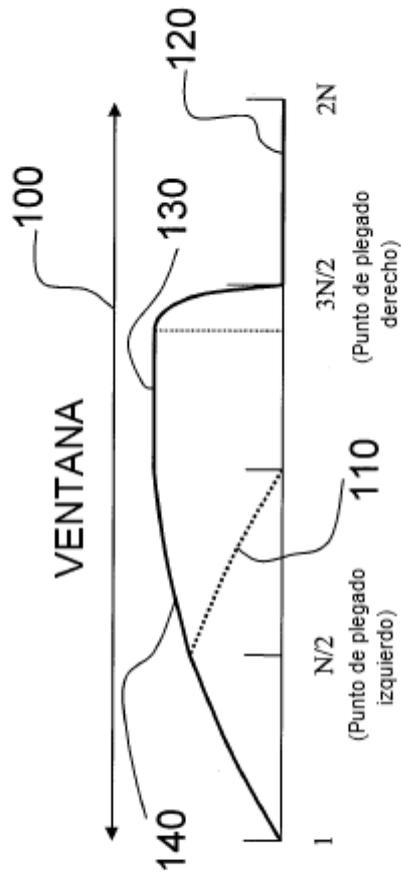


Figura 2

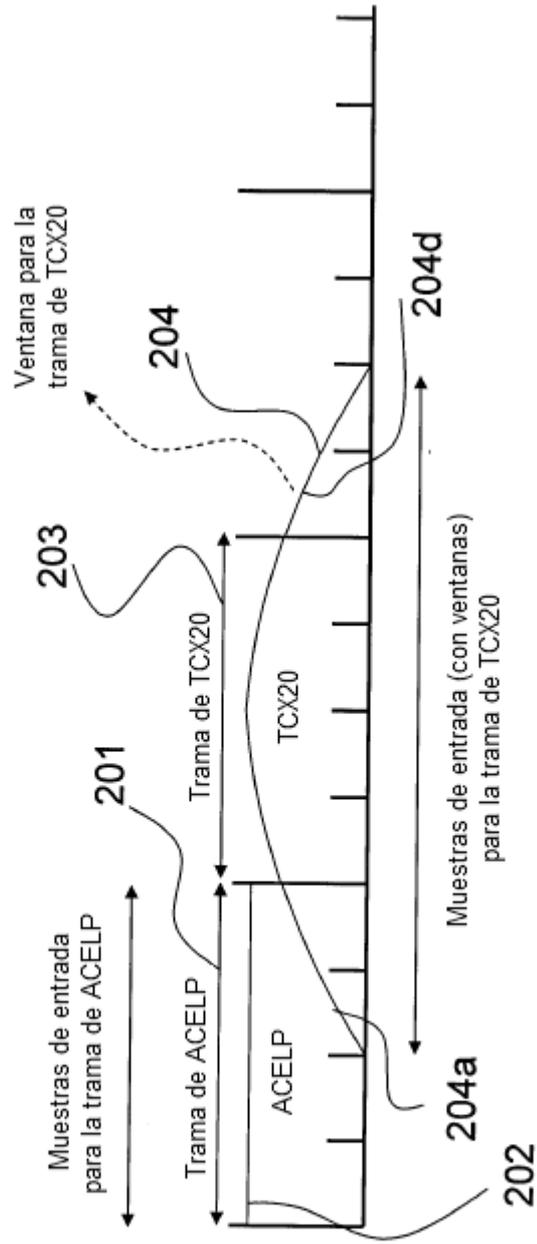


Figura 3

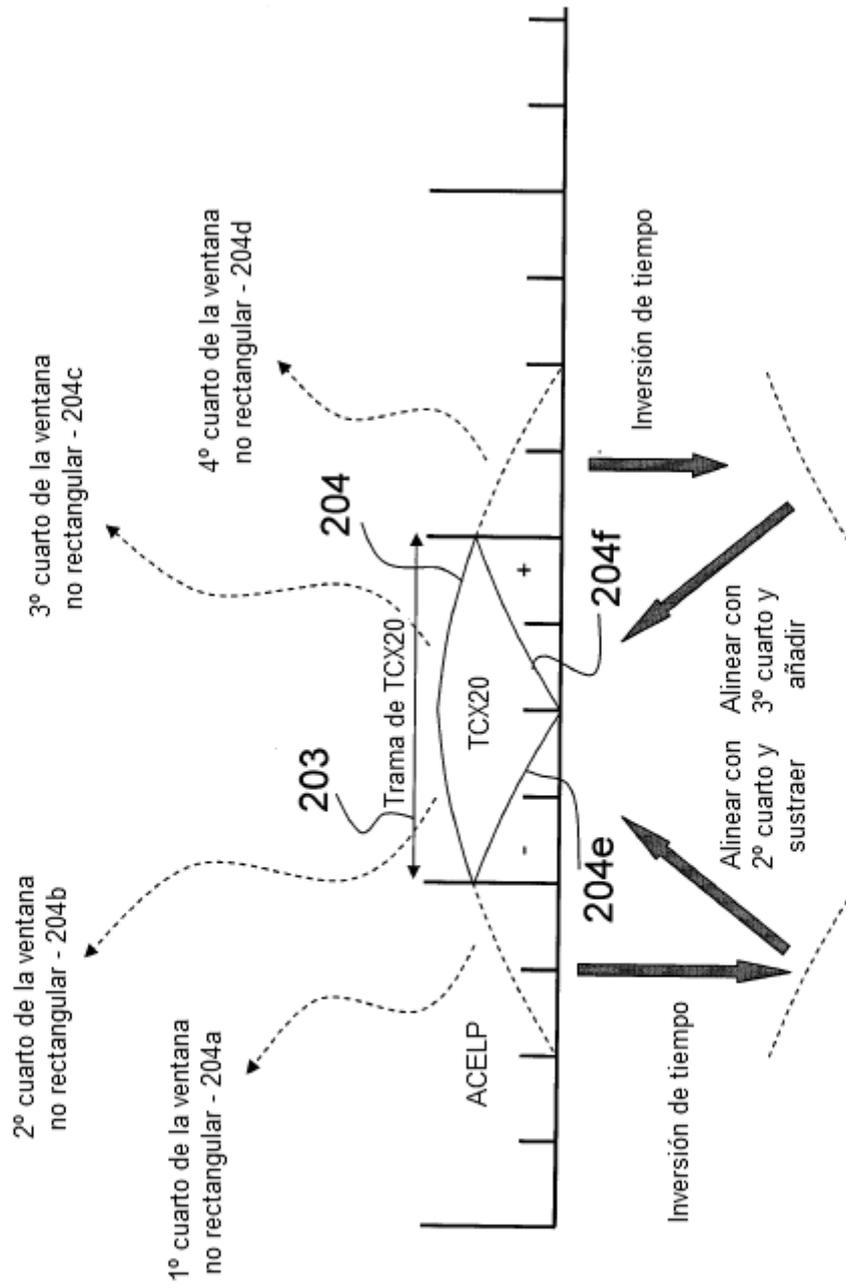
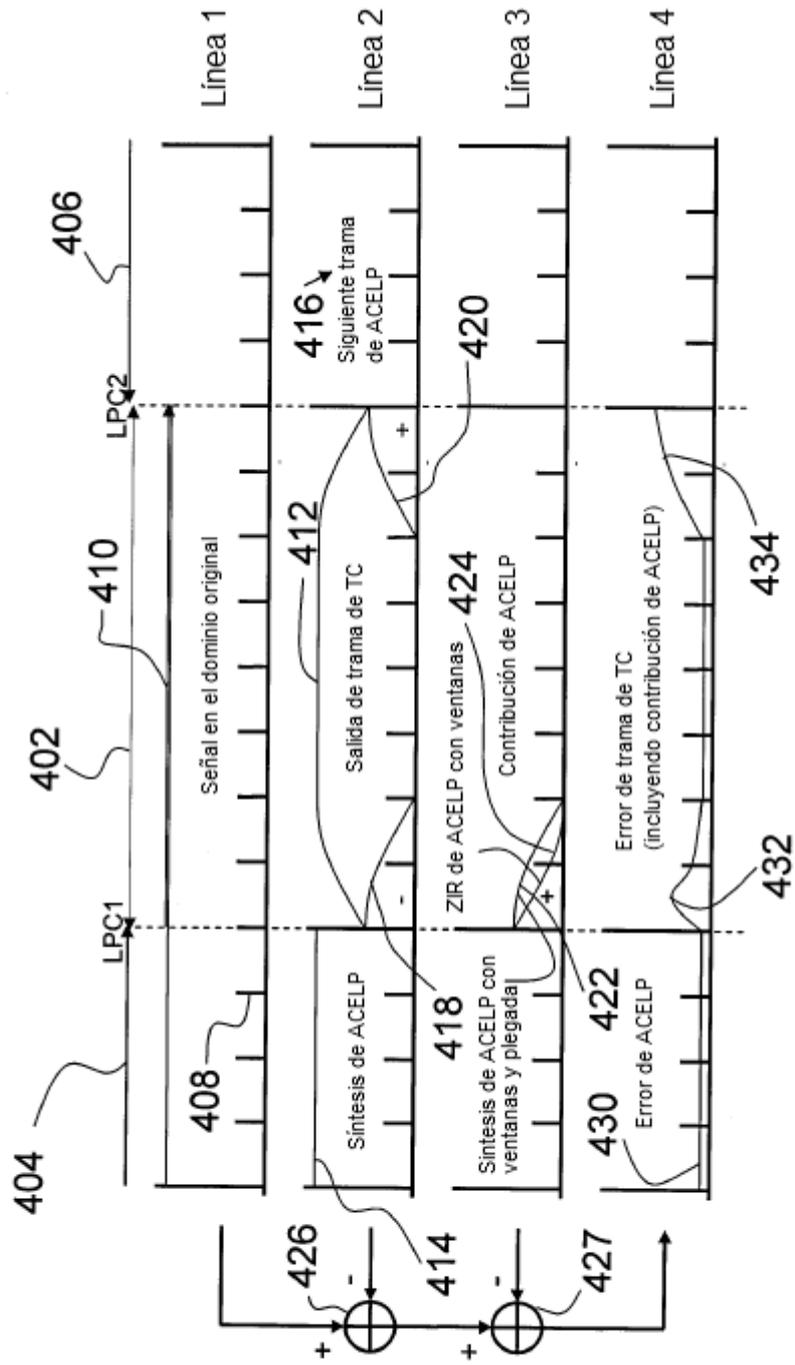


Figura 4





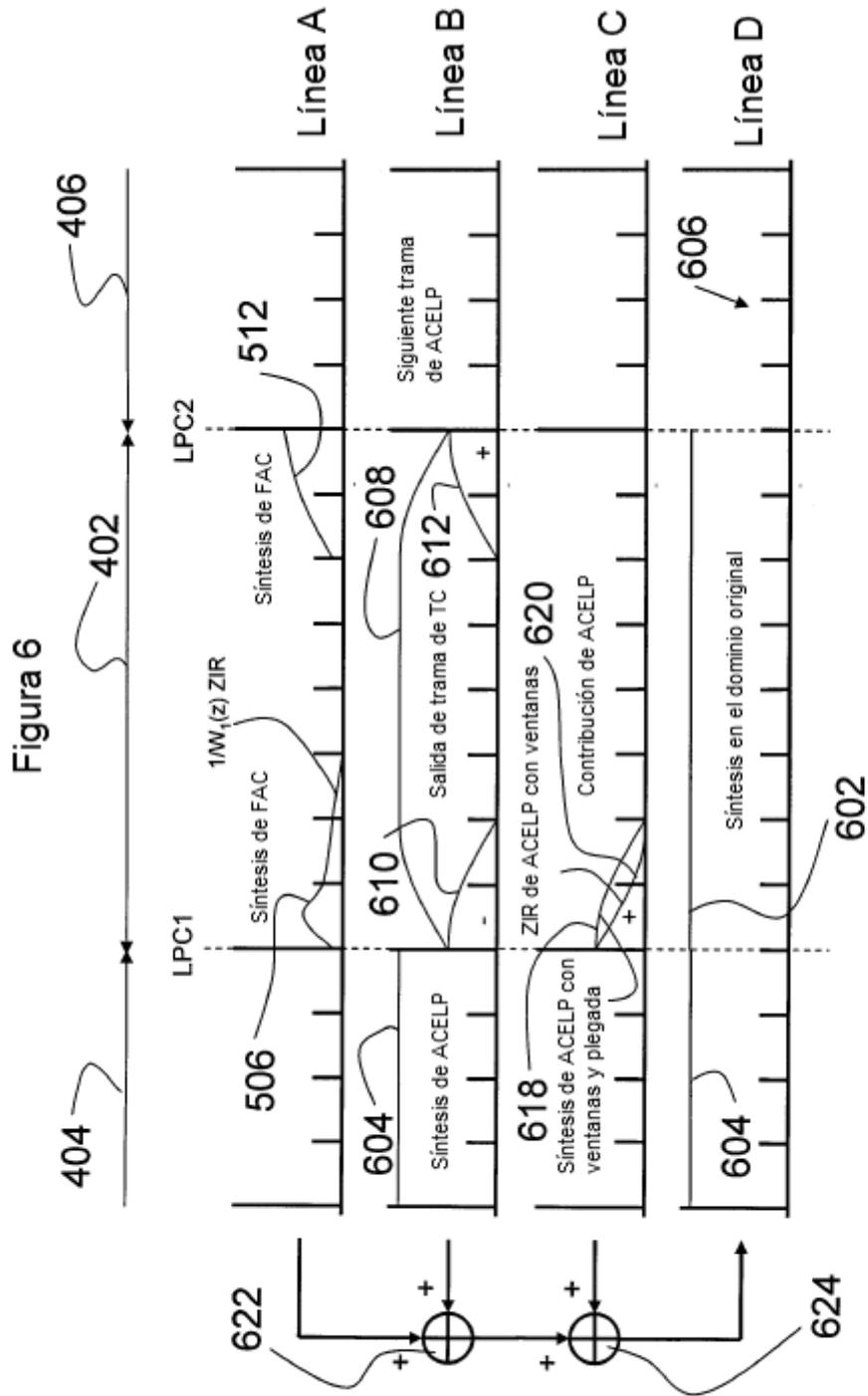




Figura 8

