

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 028 558**

51 Int. Cl.:

**G10L 19/18** (2013.01)  
**G10L 19/02** (2013.01)  
**G10L 19/04** (2013.01)  
**G10L 19/03** (2013.01)  
**G10L 19/12** (2013.01)  
**G10L 19/20** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.10.2010** **E 24160719 (1)**

97 Fecha y número de publicación de la concesión europea: **23.04.2025** **EP 4362014**

54 Título: **Decodificador de señal de audio, procedimiento correspondiente y programa informático**

30 Prioridad:

**20.10.2009 US 25346809 P**

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

**19.06.2025**

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR  
FÖRDERUNG DER ANGEWANDTEN  
FORSCHUNG E.V. (25.00%)  
Hansastraße 27c  
80686 München, DE;  
VOICEAGE CORPORATION (25.00%);  
KONINKLIJKE PHILIPS N.V. (25.00%) y  
DOLBY INTERNATIONAL AB (25.00%)**

72 Inventor/es:

**BESSETTE, BRUNO;  
NEUENDORF, MAX;  
GEIGER, RALF;  
GOURNAY, PHILIPPE;  
LEFEBVRE, ROCH;  
GRILL, BERNHARD;  
LECOMTE, JÉRÉMIE;  
BAYER, STEFAN;  
RETTELBACH, NIKOLAUS;  
VILLEMOES, LARS;  
SALAMI, REDWAN y  
DEN BRINKER, ALBERTUS C.**

74 Agente/Representante:

**PONTI & PARTNERS, S.L.P.**

ES 3 028 558 T3

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

## DESCRIPCIÓN

Decodificador de señal de audio, procedimiento correspondiente y programa informático

### 5 Campo técnico

**[0001]** Las realizaciones según la invención crean un decodificador de señal de audio para proporcionar una representación decodificada de un contenido de audio en base a una representación codificada del contenido de audio.

10 **[0002]** Las realizaciones según la invención crean un procedimiento para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada del contenido de audio.

**[0003]** Las realizaciones según la invención crean un programa informático para llevar a cabo uno de dichos procedimientos.

15 **[0004]** Las realizaciones según la invención crean un concepto para la unificación de la formación de ventanas de codificación unificada de voz y audio (también designada brevemente como USAC, *Unification Of Unified-Speech-And-Audio-Coding*) y de las transiciones de tramas.

### 20 Antecedentes de la invención

**[0005]** A continuación se explicarán algunos de los antecedentes de la invención para facilitar la comprensión de la invención y las ventajas de la misma.

25 **[0006]** Durante la última década, se hicieron grandes esfuerzos en crear la posibilidad de almacenar y distribuir de manera digital el contenido de audio. Un logro importante en este camino es la definición de la Norma Internacional ISO/IEC 14496-3. La parte 3 de esta norma está relacionada con la codificación y la decodificación de los contenidos de audio, y la subparte 4 de la parte 3 está relacionada con la codificación de audio general. ISO/IEC 14496, parte 3, subparte 4 define un concepto para codificar y decodificar el contenido de audio general. Además, se han propuesto mejoras adicionales para mejorar la calidad y/o reducir la velocidad de transferencia requerida. Además, se ha descubierto que el rendimiento de los codificadores de audio basados en el dominio frecuencial no es óptimo para los contenidos de audio comprendiendo la voz. Recientemente, se ha propuesto un códec unificado de voz y audio que combina de manera eficiente las técnicas de ambas palabras, a saber codificación de audio y codificación de voz. 30 Para algunos detalles, se hace referencia a la publicación "A Novel Scheme for Low Bitrate Unified Speech and Audio Coding – MPEG-RM0 " de M. Neuendorf y col. (presentada en la 126ª Convención de la Sociedad de Ingenieros del Sonido, del 7 al 10 de mayo de 2009, Múnich, Alemania).

**[0007]** En tal codificador de audio, algunas tramas de audio se codifican en el dominio frecuencial y algunos 40 tramas de audio se codifican en el dominio de predicción lineal.

**[0008]** El artículo "Universal Speech/Audio Coding Using Hybrid ACELP/TCX Techniques" de B. Bessette y col. (publicado en la Conferencia internacional sobre procesamiento de señal, acústica y discurso IEE del 2005, Nueva Jersey, EE. UU., vol. 3, 18 de marzo del 2005, pp. 301-304) describe un algoritmo de codificación de audio híbrido que 45 integra una técnica de codificación basada en LP y una técnica de codificación de transformada más general. ACELP se utiliza en el modo de codificación basado en LP, mientras que TCX algebraico se utiliza en el modo de codificación de transformada. Una longitud de trama se incrementa a 80 ms, se utilizan subtramas de multi-longitud adaptativa con la formación de ventanas de superposición, un VQ algebraico multi-tasa extendido se aplica al espectro de TCX para evitar la saturación del cuantificador y se mejora la formación de sonido.

50 **[0009]** El documento "Finalización del experimento principal sobre la unificación de las transiciones de tramas y formación de ventana de USAC" de Max Neuendorf y col. (ISO/IEC JTC1/SC29/WG11, MPEG 2010/M17167, enero de 2010, Kioto, Japón) describe dos tecnologías que pueden utilizarse en combinación: la corrección de solapamiento hacia delante (FAC) y el modelado de ruido en el dominio de la frecuencia (FDNS), que sustituye al filtrado LPC ponderado en la rama de codificación de la transformada de predicción lineal ponderada (wLPT) de USAC. 55

**[00010]** El documento "Alternativas para la formación en ventana en USAC" de B. Bessette y col. (ISO/IEC JTC1/SC29/WG11, MPEG 2009/M16688, junio-julio de 2009, Londres, Reino Unido) describe alternativas a la formación en ventana aplicada en el modo TCX del códec unificado de voz y audio (USAC). Se describe que la 60 formación en ventana y el cambio de modo son parte esencial de USAC. En este documento, se propone modificar las formas de las ventanas en los modos USACTCX para mitigar algunas limitaciones.

**[00011]** Sin embargo, se ha descubierto que es difícil realizar una transición entre las tramas codificadas en dominios diferentes sin sacrificar una cantidad significativa de velocidad de transferencia.

65

**[00012]** En vista de esta situación, se desea crear un concepto para codificar y decodificar un contenido de audio comprendiendo tanto audio de voz y general, que permita una realización eficaz de transiciones entre porciones codificadas utilizando modos diferentes.

## 5 Resumen de la invención

**[00013]** Las realizaciones según la invención se definen por las reivindicaciones.

**[00014]** Un ejemplo crea un decodificador de señal de audio para proporcionar una representación decodificada de un contenido de audio sobre la base de una representación codificada de un contenido de audio. El decodificador de señal de audio comprende una ruta de dominio de la transformada (por ejemplo, una ruta de dominio de predicción lineal con excitación por código de la transformada) configurada para obtener una representación de dominio temporal del contenido de audio codificado en un modo de dominio de la transformada en la base de un primer conjunto de coeficientes espectrales, una representación de una señal de estímulo de cancelación de solapamiento, y una pluralidad de parámetros de dominios de predicción lineal (por ejemplo, coeficientes de filtrado de codificación de predicción lineal). La ruta de dominio de la transformada comprende un procesador de espectros configurado para aplicar un moldeado espectral al (primer) conjunto de coeficientes espectrales dependiendo de al menos un subconjunto de parámetros de dominio de predicción lineal para obtener una versión de moldeado espectral del primer conjunto de coeficientes espectrales. La ruta de dominio de la transformada comprende también un (primer) conversor de dominio frecuencial a dominio temporal configurado para obtener una representación de dominio temporal del contenido de audio en la base a una versión de moldeado espectral del primer conjunto de coeficientes espectrales. La ruta del dominio de la transformada comprende también un filtro del estímulo de cancelación de solapamiento configurado para filtrar la señal del estímulo de cancelación de solapamiento dependiendo de al menos un subconjunto de los parámetros de dominio de predicción lineal, para derivar una señal de síntesis de cancelación de solapamiento desde la señal del estímulo de cancelación de solapamiento. La ruta de dominio de la transformada comprende también un combinador configurado para combinar la representación del dominio temporal del contenido de audio con la señal de síntesis de cancelación de solapamiento, o una versión postprocesada del mismo, para obtener una señal de dominio temporal con solapamiento reducido.

**[00015]** Las realizaciones según la invención se basan en el descubrimiento de un decodificador de audio que realiza un moldeado espectral de los coeficientes espectrales del primer conjunto de coeficientes espectrales en el dominio frecuencial, y que calcula una señal de síntesis de cancelación de solapamiento mediante filtrado del dominio temporal de una señal de estímulo de cancelación de solapamiento, donde tanto el moldeado espectral de los coeficientes espectrales y el filtrado de dominio temporal de la señal del estímulo de cancelación de solapamiento se realizan dependiendo de los parámetros del dominio de predicción lineal, es adecuada para las transiciones desde y hasta las porciones (por ejemplo, tramas) de la señal de audio codificada con diferentes moldeados de sonido y también para transiciones desde y hasta tramas que se codifican en dominios diferentes. Por consiguiente, las transiciones (por ejemplo, entre tramas de superposición y de no superposición) de la señal de audio, que se codifican en modos diferentes de una codificación de señal de audio de múltiples modos, pueden procesarse mediante el decodificador de señal de audio con buena calidad auditiva y a un nivel moderado de sobrecarga.

**[00016]** Por ejemplo, la realización del moldeado espectral del primer conjunto de coeficientes en el dominio frecuencial permite codificar las transiciones entre las porciones (por ejemplo, tramas) del contenido de audio con diferentes conceptos de moldeado de sonido en el dominio de la transformada, donde una cancelación de solapamiento puede obtenerse con buena eficacia entre las diferentes porciones del contenido de audio codificado con diferentes procedimientos de moldeado de sonido (por ejemplo, moldeado de sonido basado en factor de escala y moldeado de sonido basado en parámetros de dominio de predicción lineal). Además, los conceptos descritos anteriormente también permiten una reducción eficaz de los artefactos de solapamiento entre las porciones (por ejemplo, tramas) del contenido de audio codificado en diferentes dominios (por ejemplo, uno en un dominio de la transformada y uno en el dominio de predicción lineal excitada por código algebraico). El uso de un filtrado de dominio de tiempo de la señal del estímulo de cancelación de solapamiento permite una cancelación de solapamiento en la transición desde y hasta una porción del contenido de audio codificado en el modo de predicción lineal excitada por código algebraico incluso si el moldeado de sonido de la porción actual del contenido de audio (que puede estar codificado, por ejemplo, en el modo de dominio de predicción lineal de excitación por código de la transformada) se lleva a cabo en el dominio frecuencial, en vez de mediante el filtrado de dominio de tiempo.

**[00017]** Para resumir lo anterior, las realizaciones según la presente invención permiten un buen intercambio entre la información secundaria requerida y una calidad porcentual de transiciones entre las porciones del contenido de audio codificado en tres modos diferentes (por ejemplo, modo de dominio frecuencial, modo de dominio de predicción lineal de excitación por código de la transformada, y el modo de predicción lineal excitada por código algebraico).

**[00018]** En un ejemplo, el decodificador de señal de audio es un decodificador de señal de audio de múltiples modos configurado para conmutar entre una pluralidad de modos de codificación. En este caso, el ramal del dominio de la transformada está configurado para obtener de manera selectiva una señal de síntesis de cancelación de

solapamiento para una porción del contenido de audio que sigue a una porción previa del contenido de audio que no permite una operación de superposición y suma de cancelación de solapamiento o que es seguida por una porción posterior del contenido de audio que no permite una operación de superposición y suma de cancelación de solapamiento. Se ha descubierto que la aplicación de un moldeado de sonido, que se lleva a cabo por el moldeado 5 espectral de los coeficientes espectrales del primer conjunto de coeficientes espectrales, permite una transición entre las porciones del contenido de audio codificado en el dominio de la transformada y usando diferentes conceptos de moldeado de sonido (por ejemplo, un concepto de moldeado de sonido basado en el factor de escala y un concepto de moldeado de sonido basado en parámetros de dominio de predicción lineal) sin utilizar señales de cancelación de solapamiento, debido a que el uso del primer conversor de dominio frecuencial a dominio temporal después del 10 moldeado espectral permite una cancelación de solapamiento eficaz entre las tramas posteriores codificadas en el dominio de la transformada, incluso si se utilizan diferentes enfoques de moldeado de sonido en las tramas de audio posteriores. De este modo, la eficacia de la velocidad de transferencia puede obtenerse de manera selectiva obteniendo la señal de síntesis de cancelación de solapamiento solo para las transmisiones desde o hasta una porción del contenido de audio codificado en un dominio que no es de transformada (por ejemplo, en un modo de predicción 15 lineal excitada por código algebraico).

**[00019]** En un ejemplo, el decodificador de la señal de audio está configurado para conmutar entre el modo de dominio de predicción lineal de excitación por código de la transformada, que utiliza información de excitación por código de la transformada y un modo de dominio frecuencial, que utiliza información del coeficiente espectral e 20 información del factor de escala. En este caso, la ruta del dominio de la transformada está configurada para obtener el primer conjunto de coeficientes espectrales en base a la información de excitación por código de la transformada y para obtener los parámetros de dominio de predicción lineal en base a la información de los parámetros de dominio de predicción lineal. El decodificador de señal de audio comprende una ruta de dominio de frecuencia configurada para obtener una representación de dominio temporal del contenido de audio codificado en el modo de dominio 25 frecuencial en base a un conjunto del modo de dominio frecuencial de coeficientes espectrales descritos por la información de coeficientes espectrales y dependiendo de un conjunto de factores de escala descritos por la información del factor de escala. La ruta del dominio frecuencial comprende un procesador espectral configurado para aplicar un moldeado espectral al conjunto del modo del dominio frecuencial de coeficientes espectrales, o una versión previamente procesada del mismo, dependiendo de los factores de escala para obtener un conjunto del modo del 30 dominio frecuencial con moldeado espectral de coeficientes espectrales. La ruta del dominio frecuencial comprende también un conversor del dominio frecuencial a dominio temporal configurado para obtener una representación del dominio temporal del contenido de audio en base al conjunto del modo del dominio frecuencial con moldeado espectral de coeficientes espectrales. El decodificador de la señal de audio está configurado de manera que las representaciones de dos porciones posteriores del contenido de audio, una de las cuales está codificada en el modo 35 del dominio de predicción lineal de excitación por código de la transformada, y otra de las cuales está codificada en el modo de dominio frecuencial, comprende una superposición temporal para cancelar un solapamiento del dominio temporal provocado por la conversión del dominio frecuencial a dominio temporal.

**[0020]** Como ya se trató, el concepto conforme a las realizaciones y ejemplos de la invención es adecuado 40 para las transiciones entre porciones de contenido de audio codificado en el modo del dominio de predicción lineal de excitación por código de la transformada y en el modo del dominio frecuencial. Se obtiene una muy buena calidad de cancelación de solapamiento debido al hecho de que se realiza el moldeado espectral en el dominio frecuencial en el modo del dominio de predicción con excitación por código de la transformada.

**[0021]** En un ejemplo, el decodificador de la señal de audio está configurado para conmutar entre el modo de dominio de predicción lineal con excitación por código de la transformada que utiliza información de excitación por código de la transformada e información del parámetro del dominio de predicción lineal, y un modo de predicción lineal 45 excitada por código algebraico, que utiliza información de excitación por código algebraico e información de parámetros de dominio de predicción lineal. En este caso, la ruta del dominio de la transformada está configurada para obtener un primer conjunto de coeficientes espectrales en base a la información de excitación por código de la transformada y para obtener los parámetros de dominio de predicción lineal en base a la información de parámetros de dominio de predicción lineal. El decodificador de la señal de audio comprende una ruta de predicción lineal excitada por código algebraico para obtener una representación del dominio temporal del contenido de audio codificado en el 50 modo de predicción lineal excitada por código algebraico (también designado brevemente con ACELP a continuación), en base a la información de excitación por código algebraico y la información de los parámetros de dominio de predicción lineal. En este caso, la ruta ACELP comprende un procesador de excitación ACELP configurado para proporcionar una señal de excitación de dominio temporal en base a la información con excitación por código algebraico y un filtro de síntesis configurado para llevar a cabo el filtrado del dominio temporal, para proporcionar una 55 señal reconstruida en base a la señal con excitación del dominio temporal y que depende de los coeficientes del filtro del dominio de predicción lineal obtenidos en base a la información de los parámetros del dominio de predicción lineal. La ruta del dominio de la transformada está configurada para proporcionar de manera selectiva una señal de síntesis de cancelación de solapamiento para una porción del contenido de audio codificado en el modo de dominio de predicción lineal con excitación por código de la transformada que sigue a una porción del contenido de audio en el 60 modo ACELP y para una porción del contenido codificado en el modo de dominio de predicción lineal de excitación por código de la transformada que precede a una porción del contenido de audio codificado en el modo ACELP. Se 65

ha descubierto que la señal de síntesis de cancelación de solapamiento es muy adecuada para las transiciones entre porciones (por ejemplo, tramas) codificadas en el modo de dominio de predicción lineal con excitación por código de la transformada (a continuación también designado brevemente como TCX-LPD) y el modo ACELP.

5 **[0022]** En un ejemplo, el filtro del estímulo de cancelación de solapamiento está configurado para filtrar las señales del estímulo de cancelación de solapamiento dependiendo de los parámetros del filtro del dominio de predicción lineal que corresponden a un punto de plegado de solapamiento izquierdo del primer conversor del dominio  
10 frecuencial al dominio temporal para una porción del contenido de audio codificada en el modo TCX-LPD que sigue una porción del contenido de audio codificado en el modo ACELP. El filtro del estímulo de cancelación de solapamiento está configurado para filtrar la señal del estímulo de cancelación de solapamiento que depende de los parámetros del  
15 filtro del dominio de predicción lineal que corresponden a un punto del plegado del solapamiento derecho del segundo conversor del dominio frecuencial al dominio temporal para una porción del contenido de audio codificada en el modo de predicción lineal con excitación codificada de la transformada que precede a una porción del contenido de audio codificado en el modo ACELP. Al aplicar los parámetros del filtro del dominio de predicción lineal, que corresponden  
20 a los puntos de plegado de solapamiento, se puede obtener una cancelación de solapamiento extremadamente eficaz. También, los parámetros del filtro del dominio de predicción lineal, que corresponden a los puntos de plegado de solapamiento, se pueden, por lo general, obtener fácilmente debido a que los puntos de plegado de solapamiento por lo general están en la transición de una trama a la siguiente, de manera que la transmisión de dichos parámetros del filtro del dominio de predicción lineal se requiere de todos modos. Por consiguiente, se mantienen al mínimo las sobrecargas.

**[0023]** En un ejemplo adicional, el decodificador de la señal de audio está configurado para inicializar los valores de memoria del filtro del estímulo de cancelación de solapamiento a cero para proporcionar la señal de síntesis de cancelación de solapamiento, y para alimentar muestras M de la señal del estímulo de cancelación de solapamiento  
25 en el filtro del estímulo de cancelación de solapamiento para obtener las muestras de respuesta de entrada no cero correspondientes de la señal de síntesis de cancelación de solapamiento, y para obtener además una pluralidad de muestras de respuesta de entradas cero de la señal de síntesis de cancelación de solapamiento. El combinador está preferentemente configurado para combinar la representación del dominio temporal del contenido de audio con las muestras de respuesta de entrada no cero y las muestras de respuesta de entradas cero posteriores, para obtener  
30 una señal del dominio temporal con solapamiento reducido en una transición desde una porción del contenido de audio codificado en el modo ACELP a una porción del contenido de audio codificado en el modo TCX-LPD que sigue a la porción del contenido de audio codificado en el modo ACELP. Al aprovechar ambas, las muestras de respuesta de entrada no cero y las muestras de respuesta de entradas cero, se puede hacer un muy buen uso del filtro del estímulo de cancelación de solapamiento. También, se puede obtener una señal de síntesis de cancelación de solapamiento  
35 muy leve mientras se mantiene lo más pequeña posible la cantidad de muestras requeridas de la señal del estímulo de cancelación de solapamiento. Además, se ha descubierto que una forma de la señal de síntesis de cancelación de solapamiento está muy bien adaptada a los artefactos de solapamiento típicos por el uso del concepto antes mencionado. De este modo, se puede obtener un buen intercambio entre la eficiencia de la codificación y la cancelación del solapamiento.

40 **[0024]** En un ejemplo, el decodificador de la señal de audio está configurado para combinar una versión plegada y formada en ventana de al menos una porción de una representación del dominio temporal con el modo ACELP con una representación del dominio temporal de una porción posterior del contenido de audio obtenido usando el modo TCX-LPD, para cancelar al menos parcialmente un solapamiento. Se ha descubierto que el uso de tales  
45 mecanismos de cancelación de solapamiento además de la generación de la señal de síntesis de cancelación de solapamiento, ofrece la posibilidad de obtener una cancelación de solapamiento en una velocidad de transferencia muy eficaz. En particular, la señal del estímulo de cancelación de solapamiento requerida puede codificarse con alta eficiencia si la señal de síntesis de cancelación de solapamiento es soportada, en la cancelación de solapamiento por la versión plegada y formada en ventana de al menos una porción de una representación del dominio temporal  
50 utilizando el modo ACELP.

**[0025]** En un ejemplo, el decodificador de la señal de audio está configurado para combinar una versión formada en ventana de una respuesta de impulso cero del filtro de síntesis del ramal ACELP con una representación del dominio temporal de una porción posterior del contenido de audio obtenido con el modo TCX-LPD, para cancelar  
55 al menos parcialmente un solapamiento. Se ha descubierto que el uso de dicha respuesta de impulso cero puede ayudar también a mejorar la eficacia de codificación de la señal del estímulo de cancelación de solapamiento, debido a que la respuesta de impulso cero del filtro de la síntesis del ramal ACELP generalmente cancela al menos una parte del solapamiento en la porción codificada TCX-LPD del contenido de audio. Por consiguiente, la energía de la señal de síntesis de cancelación de solapamiento se reduce, lo que, a su vez, resulta en una reducción de la energía de la  
60 señal del estímulo de cancelación de solapamiento. Sin embargo, codificar señales con una menor energía es posible por lo general con requisitos de velocidad de transferencia reducida.

**[0026]** En un ejemplo, el decodificador de la señal de audio está configurado para conmutar entre el modo TCX-LPD, donde se utiliza la transformada del dominio frecuencial al dominio temporal limitada, un modo de dominio  
65 frecuencial, donde se utiliza la transformada del dominio frecuencial al dominio temporal derivada, así como un modo

de predicción lineal excitada por código algebraico. En este caso, el decodificador de señal de audio está configurado para cancelar al menos parcialmente un solapamiento en una transición entre una porción del contenido de audio codificado en el modo TCX-LPD y una porción del contenido de audio codificado en el modo de dominio frecuencial llevando a cabo una operación de superposición y suma entre las muestras del dominio temporal de las porciones de superporción posteriores del contenido de audio. También, el decodificador de la señal de audio está configurado para cancelar al menos parcialmente un solapamiento en una transición entre una porción del contenido de audio codificado en el modo TCX-LPD y una porción del contenido de audio codificado en el modo ACELP usando la señal de síntesis de cancelación de solapamiento. Se ha descubierto que el decodificador de señal de audio también es adecuado para conmutar entre los diferentes modos de operación, donde el solapamiento se cancela de manera muy eficaz.

**[0027]** En un ejemplo, el decodificador de la señal de audio está configurado para aplicar un valor de ganancia común para la escala de ganancia de una representación del dominio temporal proporcionada por el primer conversor de dominio frecuencial a dominio temporal de la ruta de dominio de la transformada (por ejemplo, la ruta TCX-LPF) y para una escala de ganancia de la señal del estímulo de cancelación de solapamiento o de la señal de síntesis de cancelación de solapamiento. Se ha descubierto que reutilizar este valor de ganancia común tanto para la escala de la representación de dominio temporal proporcionada por el primer conversor de dominio frecuencial a dominio temporal y para la escala de la señal del estímulo de cancelación de solapamiento o la señal de síntesis de cancelación de solapamiento permite la reducción de la velocidad de transferencia requerida en una transición entre las porciones del contenido de audio codificado en modos diferentes. Esto es muy importante, debido a que el requisito de velocidad de transferencia aumenta por la codificación de la señal del estímulo de cancelación de solapamiento en el entorno de una transición entre porciones del contenido de audio codificado en modos diferentes.

**[0028]** En un ejemplo, el decodificador de la señal de audio está configurado para aplicar, además del moldeado espectral realizado dependiendo de, al menos, un subconjunto de parámetros del dominio de predicción lineal, un desmoldeado espectral a, al menos, un subconjunto del primer conjunto de coeficientes espectrales. En este caso, el decodificador de señal de audio está configurado para aplicar el desmoldeado espectral a, al menos, un subconjunto de un conjunto de coeficientes espectrales de cancelación de solapamiento del cual se deriva la señal del estímulo de cancelación de solapamiento. Aplicar un desmoldeado espectral tanto al primer conjunto de coeficientes espectrales y a los coeficientes espectrales de cancelación de solapamiento del cual se deriva la señal del estímulo de cancelación de solapamiento, asegura que la señal de síntesis de cancelación de solapamiento está bien adaptada a la señal del contenido de audio "principal" proporcionada por el primer conversor de dominio frecuencial a dominio temporal. De nuevo, se mejora la eficacia de la codificación para codificar la señal del estímulo de cancelación de solapamiento.

**[0029]** En un ejemplo, el decodificador de la señal de audio comprende un segundo conversor de dominio frecuencial a dominio temporal configurado para obtener una representación del dominio temporal de la señal del estímulo de cancelación de solapamiento dependiendo de un conjunto de coeficientes espectrales que representan la señal del estímulo de cancelación de solapamiento. En este caso, el primer conversor del dominio frecuencial a dominio temporal está configurado para llevar a cabo una transformada traslapada, que comprende un solapamiento del dominio temporal. El segundo conversor del dominio frecuencial a dominio temporal se configura para realizar una transformada no traslapada. Por consiguiente, se puede mantener una alta eficacia de codificación utilizando la transformada traslapada para la síntesis de la señal "principal". No obstante, la cancelación de solapamiento se logra utilizando una conversión de dominio frecuencial a dominio temporal, que no es traslapada. Sin embargo, se ha descubierto que la combinación de la conversión del dominio frecuencial al dominio temporal traslapada y la conversión del dominio frecuencial al dominio temporal no traslapada permite una codificación más eficaz de las transiciones en comparación con una única conversión del dominio frecuencial al dominio temporal no traslapado.

**[0030]** Un ejemplo crea un codificador de señal de audio para proporcionar una representación codificada de un contenido de audio comprendiendo un primer conjunto de coeficientes espectrales, una representación de una señal de estímulo de cancelación de solapamiento y una pluralidad de parámetros de dominio de predicción lineal en base a una representación de la entrada del contenido de audio. El codificador de la señal de audio comprende un conversor del dominio temporal al dominio frecuencial configurado para procesar la representación de entrada del contenido de audio, para obtener una representación del dominio frecuencial del contenido de audio. El codificador de la señal de audio comprende también un procesador espectral configurado para aplicar el moldeado espectral a un conjunto de coeficientes espectrales, o a una versión preprocesada del mismo, dependiendo de un conjunto de parámetros del dominio de predicción lineal para una porción del contenido de audio que se va a codificar en el dominio de predicción lineal, para obtener una representación del dominio frecuencial de moldeado espectral del contenido de audio. El codificador de señal de audio comprende también un proveedor de información de cancelación de solapamiento configurado para proporcionar una representación de una señal del estímulo de cancelación de solapamiento, de manera que una filtración de la señal del estímulo de cancelación de solapamiento que depende de al menos un subconjunto de parámetros de dominio de predicción lineal resulta en una señal de síntesis de cancelación de solapamiento para cancelar los artefactos de solapamiento en un decodificador de la señal de audio.

**[0031]** El codificador de la señal de audio que se trata en la presente es adecuado para cooperar con el codificador de señal de audio que se describió anteriormente. En particular, el codificador de la señal de audio está configurado para proporcionar una representación del contenido de audio donde se mantiene razonablemente

pequeña una sobrecarga de la velocidad de transferencia requerida para cancelar el solapamiento en las transiciones entre porciones (por ejemplo, tramas o subtramas) del contenido de audio codificado en modos diferentes.

**[0032]** Realizaciones y ejemplos adicionales según la invención crean un procedimiento para proporcionar una representación decodificada del contenido de audio y un procedimiento para proporcionar una representación codificada de un contenido de audio. Dichos procedimientos se basan en las mismas ideas que el aparato mencionado anteriormente.

**[0033]** Las realizaciones según la invención crean programas informáticos para realizar uno de dichos procedimientos. Los programas informáticos se basan también en las mismas consideraciones.

#### Breve descripción de las figuras

**[0034]** Posteriormente se describirán las realizaciones según la presente invención y ejemplos adicionales haciendo referencia a las figuras adjuntas, donde:

La Fig. 1 muestra un diagrama de bloque esquemático de un codificador de señal de audio, según un ejemplo;  
 La Fig. 2 muestra un diagrama de bloque esquemático de un decodificador de señal de audio, según un ejemplo;  
 La Fig. 3a muestra un diagrama de bloque esquemático del decodificador de señal de audio de referencia según el borrador de trabajo 4 de la norma en borrador de la codificación unificada de voz y audio (USAC);  
 La Fig. 3b muestra un diagrama de bloque esquemático de un decodificador de señal de audio, según otro ejemplo;  
 La Fig. 4 muestra una representación gráfica de una transición de la ventana de referencia según el borrador de trabajo 4 de la norma en borrador de la USAC;  
 La Fig. 5 muestra una representación esquemática de las transiciones de ventana que pueden utilizarse en una codificación de señal de audio, conforme a un ejemplo;  
 La Fig. 6 muestra una representación esquemática que proporciona un panorama de todos los tipos de ventana utilizadas en un codificador de señal de audio según un ejemplo o un decodificador de señal de audio según un ejemplo;  
 La Fig. 7 muestra una representación en tabla de una secuencia de ventana permitida, que puede utilizarse en un codificador de señal de audio según un ejemplo, o un decodificador de señal de audio según una realización de la invención;  
 La Fig. 8 muestra un diagrama de bloque esquemático detallado de un codificador de señal de audio, según un ejemplo;  
 La Fig. 9 muestra un diagrama de bloque esquemático detallado de un decodificador de señal de audio, según una realización de la invención;  
 La Fig. 10 muestra una representación esquemática de operaciones de decodificación de cancelación de solapamiento hacia delante (FAC, por sus siglas en inglés) para transiciones desde y a ACELP;  
 La Fig. 11 muestra una representación esquemática de un cálculo de una FAC objetivo en un codificador;  
 La Fig. 12 muestra una representación esquemática de una cuantificación de una FAC objetivo en el contexto de un moldeado de sonido del dominio frecuencial (FNDS, por sus siglas en inglés);  
 La tabla 1 muestra las condiciones para la presencia de un filtro LPC dado en una corriente de bits;  
 La Fig. 13 muestra una representación esquemática de un principio de un cuantificador inverso de LPC algebraica ponderada;  
 La tabla 2 muestra una representación de posibles modos absolutos y relativos de cuantificación y la señalización de la velocidad de transferencia correspondiente de "mode\_lpc";  
 La tabla 3 muestra una representación en tabla de los modos de codificación para los números del libro de códigos  $n_k$ ;  
 La tabla 4 muestra una representación en tabla de una normalización del vector  $W$  para la cuantificación AVQ;  
 La tabla 5 muestra una representación en tabla de mapeado para una energía de excitación promedio  $\bar{E}$ ;  
 La tabla 6 muestra una representación en tabla de una cantidad de coeficientes espectrales como una función de "mod[]";  
 La Fig. 14 muestra una representación de una sintaxis de una corriente del canal del dominio frecuencial "fd\_channel\_stream()";  
 La Fig. 15 muestra una representación de una sintaxis de una corriente del canal del dominio de predicción lineal "lpd\_channel\_stream()"; y  
 La Fig. 16 muestra una representación de una sintaxis de la información de cancelación de solapamiento hacia delante "fac\_data()".

#### Descripción detallada de los ejemplos y las realizaciones

##### 1. Decodificador de la señal de audio según la Fig. 1

**[0035]** La Fig. 1 muestra un diagrama en bloque esquemático de un codificador de señal de audio 100, según un ejemplo. El codificador de la señal de audio 100 está configurado para recibir una representación de entrada 110 de un contenido de audio y proporcionar, sobre esta base, una representación codificada 112 del contenido de audio. La representación codificada 112 del contenido de audio comprende un primer conjunto 112a de coeficientes

espectrales, una pluralidad de parámetros de dominio de predicción lineal 112b y una representación 112c de una señal del estímulo de cancelación de solapamiento.

**[0036]** El codificador de la señal de audio 100 comprende un conversor de dominio temporal a dominio  
5 frecuencial 120 que está configurado para procesar la representación de entrada 110 del contenido de audio (o, de manera equivalente, una versión preprocesada 110' del mismo), para obtener una representación del dominio frecuencial 122 del contenido de audio (que puede adquirir la forma de un conjunto de coeficientes espectrales).

**[0037]** El codificador de señal de audio 100 comprende también un procesador espectral 130 que está  
10 configurado para aplicar un moldeado espectral a la representación del dominio frecuencial 122 del contenido de audio, o a una versión preprocesada 122' del mismo, dependiendo de un conjunto 140 de parámetros de dominio de predicción lineal para una porción del contenido de audio que se va a codificar en el dominio de predicción lineal, para obtener una representación del dominio frecuencial de moldeado espectral 132 del contenido del audio. El primer conjunto 112a de coeficientes espectrales puede ser igual a la representación del dominio frecuencial de moldeado espectral 132 del contenido de audio, o puede derivar de la representación del dominio frecuencial de moldeado espectral 132 del contenido de audio.  
15

**[0038]** El codificador de señal de audio 100 comprende también un proveedor de información de solapamiento  
150, que está configurado para proporcionar una representación 112c de una señal del estímulo de cancelación de solapamiento, de manera que un filtrado de la señal del estímulo de cancelación de solapamiento que depende de, al menos, un subconjunto de los parámetros del dominio de predicción lineal 140 resulta en una señal de síntesis de cancelación de solapamiento para cancelar los artefactos de solapamiento en un decodificador de señal de audio.

**[0039]** También debería observarse que los parámetros del dominio de predicción lineal 112b pueden, por  
25 ejemplo, ser iguales a los parámetros del dominio de predicción lineal 140.

**[0040]** El codificador de la señal de audio 110 proporciona información que es adecuada para una  
reconstrucción del contenido de audio, incluso si se codifican diferentes porciones (por ejemplo, tramas o subtramas) del contenido de audio en modos diferentes. Para una porción del contenido de audio codificada en el dominio de predicción lineal, por ejemplo, en un modo de dominio de predicción lineal con excitación por código de la transformada, el moldeado espectral, que presenta un moldeado de sonido y por lo tanto permite una cuantificación del contenido de audio con una velocidad de transferencia comparativamente pequeña, se lleva a cabo después de la conversión del dominio temporal al dominio frecuencial. Esto permite una superposición y suma de la cancelación de solapamiento de una porción del contenido de audio codificado en el dominio de predicción lineal con una porción anterior o posterior del contenido de audio codificado en un modo de dominio frecuencial. Al utilizar los parámetros de dominio de predicción lineal 140 para el moldeado espectral, el moldeado espectral está bien adaptado a contenidos de audio tipo voz, de manera que se puede obtener una eficacia particularmente buena en la codificación para los contenidos de audio tipo voz. Además, la representación de la señal del estímulo de cancelación de solapamiento permite una cancelación de solapamiento eficaz en las transiciones desde o hacia una porción (por ejemplo, trama o subtrama) del contenido de audio codificado en el modo de predicción lineal excitada por código algebraico. Al proporcionar la representación de la señal del estímulo de cancelación de solapamiento dependiendo de los parámetros del dominio de predicción lineal, se obtiene una representación particularmente eficaz de la señal del estímulo de cancelación de solapamiento, que puede decodificarse en el lado del decodificador considerando los parámetros de dominio de predicción lineal, que de todos modos se conocen en el decodificador.  
40  
45

**[0041]** Para resumir, el codificador de la señal de audio 100 es adecuado para permitir transiciones entre  
porciones del contenido de audio codificado en modos de codificación diferentes y es capaz de proporcionar información de cancelación de solapamiento en una forma particularmente compacta.

## 50 2. Decodificador de la señal de audio según la Fig. 2

**[0042]** La Fig. 2 muestra un diagrama en bloque esquemático de un decodificador de señal de audio 200, según  
un ejemplo. El decodificador de la señal de audio 200 está configurado para recibir una representación codificada 210 del contenido de audio y para proporcionar, sobre esta base, la representación decodificada 212 del contenido de  
55 audio, por ejemplo, en la forma de una señal del dominio temporal con solapamiento reducido.

**[0043]** El decodificador de la señal de audio 200 comprende una ruta de dominio de la transformada (por  
ejemplo, una ruta del dominio de predicción lineal con excitación por código de la transformada) configurada para obtener una representación del dominio temporal 212 del contenido de audio codificado en un modo de dominio de la transformada en base a un (primer) conjunto 220 de coeficientes espectrales, una representación 224 de una señal del estímulo de cancelación de solapamiento y una pluralidad de parámetros de dominio de predicción lineal 222. La ruta del dominio de la transformada comprende un procesador espectral 230 configurado para aplicar un moldeado espectral del (primer) conjunto 220 de coeficientes espectrales dependiendo de, al menos, un subconjunto de parámetros de dominio de predicción lineal 222, para obtener una versión de moldeado espectral 232 del primer  
60 conjunto 220 de coeficientes espectrales. La ruta de dominio de la transformada comprende también un (primer)  
65



conversor de dominio frecuencial a dominio temporal 240 configurado para obtener una representación de dominio temporal 242 del contenido de audio en base a una versión de moldeado espectral 232 del (primer) conjunto 220 de coeficientes espectrales. La ruta del dominio de la transformada comprende también un filtro del estímulo de cancelación de solapamiento 250, que está configurado para filtrar la señal del estímulo de cancelación de solapamiento (que está representada por la representación 224) dependiendo de al menos un subconjunto de los parámetros de dominio de predicción lineal 222, para derivar una señal de síntesis de cancelación de solapamiento 252 desde la señal del estímulo de cancelación de solapamiento. La ruta de dominio de la transformada comprende también un combinador 260 configurado para combinar la representación del dominio temporal 242 del contenido de audio (o, de manera equivalente, una versión postprocesada 242' del mismo) con la señal de síntesis de cancelación de solapamiento 252 (o, de manera equivalente, una versión postprocesada 252' del mismo), para obtener una señal de dominio temporal con solapamiento reducido 212.

**[0044]** El decodificador de señal de audio 200 puede comprender un procesamiento opcional 270 para derivar la configuración del procesador espectral 230, que realiza, por ejemplo, un moldeado de sonido de escala y/o de dominio frecuencial, desde al menos un subconjunto de parámetros de dominio de predicción lineal.

**[0045]** El decodificador de la señal de audio 200 comprende también un procesamiento opcional 280, que está configurado para derivar la configuración del filtro del estímulo de cancelación de solapamiento 250, que puede, por ejemplo, realizar un filtrado de síntesis para sintetizar la señal de síntesis de cancelación de solapamiento 252, desde al menos un subconjunto de parámetros de dominio de predicción lineal 222.

**[0046]** El decodificador de la señal de audio 200 está configurado para proporcionar una señal de dominio temporal de solapamiento reducido 212, que es adecuada para una combinación tanto, con una señal de dominio temporal que represente un contenido de audio y que se obtenga en un modo de dominio frecuencial de operación, y para/en combinación con una señal de dominio temporal que represente un contenido de audio y que esté codificada en un modo ACELP de operación. Existen características particularmente buenas de superposición y suma entre las porciones (por ejemplo, tramas) del contenido de audio decodificadas utilizando un modo de dominio frecuencial de operación (usando una ruta del dominio frecuencial no ilustrada en la Fig. 2) y porciones (por ejemplo, una trama o subtrama) del contenido de audio decodificado usando la ruta de dominio de la transformada de la Fig. 2, debido a que el moldeado de sonido es realizado por el procesador espectral 230 en el dominio frecuencial, es decir, antes de la conversión del dominio frecuencial a dominio temporal 240. Además, se pueden obtener también cancelaciones de solapamiento particularmente buenas entre una porción (por ejemplo, una trama o una subtrama) del contenido de audio decodificado usando la ruta de dominio de la transformada de la Fig. 2 y una porción (por ejemplo, una trama o subtrama) del contenido de audio decodificado usando una ruta de decodificación de ACELP debido al hecho de que la señal de síntesis de cancelación de solapamiento 252 se proporciona en base a un filtrado de una señal del estímulo de cancelación de solapamiento dependiente de los parámetros de dominio de predicción lineal. Una señal de síntesis de cancelación de solapamiento 252, que se obtiene de esta manera, está por lo general bien adaptada a los artefactos de solapamiento que tienen lugar en la transición entre una porción del contenido de audio codificado en el modo TCX-LPD y una porción del contenido de audio codificado en el modo ACELP. Se describirán a continuación detalles adicionales y opcionales relacionados con la operación de la decodificación de la señal de audio.

### 3. Decodificadores de audio conmutados conforme a las Figs. 3a y 3b.

**[0047]** A continuación, el concepto de un decodificador de señal de audio de modos múltiples se tratará brevemente haciendo referencia a las Fig. 3a y 3b.

#### 3.1 Decodificador de señal de audio 300 conforme a la Fig. 3a

**[0048]** La Fig. 3a muestra un diagrama en bloque sistemático de un decodificador de señal de audio de modos múltiples, y la Fig. 3b muestra un diagrama en bloque esquemático de un decodificador de señal de audio de modos múltiples, según una realización de la invención. En otras palabras, la Fig. 3a muestra un flujo de señal del decodificador básico de un sistema de referencia (por ejemplo, según el borrador de trabajo 4 del borrador de la norma de la USAC), y la Fig. 3b muestra un flujo de señal del decodificador básico del sistema propuesto según un ejemplo.

**[0049]** El decodificador de la señal de audio 300 se describirá primero haciendo referencia a la Fig. 3a. El decodificador de la señal de audio 300 comprende un multiplexor de bits 310, que está configurado para recibir una corriente de bits de entrada y proporcionar la información incluida en la corriente de bits a las unidades de procesamiento apropiadas de los ramales de procesamiento.

**[0050]** El decodificador de señal de audio 300 comprende una ruta del modo de dominio frecuencial 320, que está configurada para recibir información del factor de escala 322 e información del coeficiente espectral codificado 324, y proporcionar, en esta base, una representación del dominio temporal 326 de una trama de audio codificada en el modo de dominio frecuencial. El decodificador de la señal de audio 300 comprende también una ruta del dominio de predicción lineal con excitación por código de la transformada 330, que está configurado para recibir información de excitación por código de la transformada codificada 332 e información del coeficiente de predicción lineal 334,

(también designada como información de codificación de predicción lineal o como información del dominio de predicción lineal o como información del filtro de codificación de predicción lineal) y para proporcionar, en esta base, una representación de dominio temporal de una trama de audio o subtrama de audio codificada en el modo de dominio de predicción lineal con excitación por código de la transformada (TCX-LPD). El decodificador de la señal de audio 5 300 comprende también una ruta de predicción lineal excitada por código algebraico (ACELP) 340, que está configurada para recibir información de excitación codificada 342 e información de codificación de predicción lineal 344, (designada también como información del coeficiente de predicción lineal o como información del dominio de predicción lineal o como información del filtro de codificación de predicción lineal) y para proporcionar, en esta base, una información de codificación de predicción lineal del dominio temporal a una representación de una trama de audio 10 o subtrama de audio codificada en el modo ACELP. El decodificador de la señal de audio 300 comprende también una formación en ventanas de transición, que está configurada para recibir las representaciones del dominio temporal 326, 336, 346 de las tramas o subtramas del contenido de audio codificadas en modos diferentes y para combinar la representación del dominio temporal utilizando una formación en ventanas de transición.

15 **[0051]** La ruta del dominio frecuencial 320 comprende un decodificador aritmético 320a configurado para decodificar la representación espectral codificada 324, para obtener una representación espectral decodificada 320b, un cuantificador inverso 320d configurado para proporcionar una representación espectral cuantificada de manera inversa 320e en base a la representación espectral decodificada 320b, una escala 320e configurada para presentar en escala la representación espectral cuantificada de manera inversa 320d dependiendo de los factores de escala, 20 para obtener una representación espectral en escala 320f y una transformada de coseno discreta modificada (inversa) 320g para proporcionar una representación del dominio temporal 326 en base a la representación espectral en escala 320f.

**[0052]** El ramal TCX-LPD 330 comprende un decodificador aritmético 330a configurado para proporcionar una 25 representación espectral decodificada 330b en base a la representación espectral codificada 332, un cuantificador inverso 330c configurado para proporcionar una representación espectral cuantificada de manera inversa 330d en base a la representación espectral 330b, una transformada de coseno discreta modificada (inversa) 330e para proporcionar una señal de excitación 330f en base a la representación espectral cuantificada de manera inversa 330d, y un filtro de síntesis de codificación de predicción lineal 330g para proporcionar la representación del dominio temporal 30 336 en base a la señal de excitación 330f y a los coeficientes del filtro de codificación de predicción lineal 334 (también designados por lo general como coeficientes del filtro del dominio de predicción lineal).

**[0053]** El ramal ACELP 340 comprende un procesador de excitación ACELP 340a configurado para proporcionar una señal de excitación ACELP 340b en base a la señal de excitación codificada 342 y un filtro de la 35 síntesis de codificación de predicción lineal 340c para proporcionar la representación del dominio temporal 346 en base a la señal de excitación ACELP 340b y los coeficientes del filtro de codificación de predicción lineal 344.

### 3.2 Formación en ventana de transición conforme a la Fig. 4

40 **[0054]** Tomando como referencia ahora la Fig. 4, la formación en ventana de transición 350 se describirá en más detalle. Primero, se describirá la estructura de trama general de un decodificador de señal de audio 300. Sin embargo, debería observarse que una estructura de trama muy similar con solo diferencias menores, o incluso una estructura de trama general idéntica, se utilizará en otros codificadores o decodificadores de señal de audio descritos en esta invención. También debería observarse que las tramas de audio generalmente comprenden una longitud de 45 muestras N, donde N puede ser igual a 2048. Las tramas posteriores del contenido de audio pueden estar superpuestas por aproximadamente el 50 %, por ejemplo, por N/2 muestras de audio. Una trama de audio se puede codificar en el dominio frecuencial, de manera que las muestras del dominio temporal N de una trama de audio estén representadas por un conjunto de, por ejemplo, coeficientes espectrales N/2. De manera alternativa, las muestras del dominio temporal N de una trama de audio pueden estar representadas también por una pluralidad de, por ejemplo, ocho 50 conjuntos de, por ejemplo, 128 coeficientes espectrales. Por consiguiente, se puede obtener una resolución temporal mayor.

**[0055]** Si las muestras del dominio temporal N de una trama de audio están codificadas en el modo de dominio frecuencial usando un único conjunto de coeficientes espectrales, una única ventana tal como, por ejemplo, una 55 ventana denominada "STOP\_START", una ventana denominada "AAC Long", una ventana denominada "AAC 35 Start", o una ventana denominada "AAC Stop" se puede aplicar para formar en ventana las muestras del dominio temporal 326 proporcionadas por la transformada de coseno discreta modificada inversa 320g. En contraste, una pluralidad de ventanas más cortas, por ejemplo del tipo "AAC Short" pueden aplicarse para formar en ventana representaciones de dominio temporal que se obtuvieron usando diferentes conjuntos de coeficientes espectrales, si 60 las muestras del dominio temporal N de una trama de audio se codifican utilizando una pluralidad de conjuntos de coeficientes espectrales. Por ejemplo, se pueden aplicar ventanas cortas separadas a las representaciones del dominio temporal obtenidas en base a conjuntos individuales de coeficientes espectrales asociados con una única trama de audio.

65 **[0056]** Una trama de audio codificada en el modo de dominio de predicción lineal puede estar subdividida en

una pluralidad de subtramas, que están a veces designadas como "tramas". Cada una de las subtramas puede estar codificada ya sea en el modo TCX-LPD o en el modo ACELP. De manera acorde, sin embargo, en el modo TCX-LPD, se pueden codificar dos o incluso cuatro de las subtramas juntas utilizando un único conjunto de coeficientes espectrales que describan la excitación codificada de la transformada.

**[0057]** Una subtrama (o un conjunto de dos o cuatro subtramas) codificadas en el modo TCX-LPD puede estar representada por un conjunto de coeficientes espectrales y uno o más conjuntos de coeficientes de filtrado de codificación de predicción lineal. Una subtrama del contenido de audio codificado en el dominio ACELP puede estar representada por una señal de excitación ACELP codificada y uno o más conjuntos de coeficientes de filtro de codificación de predicción lineal.

**[0058]** Haciendo referencia ahora a la Fig. 4, se describirá la implementación de las transiciones entre tramas o subtramas. En la representación esquemática de la Fig. 4, las abscisas 402a a 402i describen un tiempo en términos de muestras de audio, y las ordenadas 404a a 404i describen ventanas y/o regiones temporales para las que se proporcionan muestras del dominio temporal.

**[0059]** En el numeral de referencia 410, se representa una transición entre dos tramas de superposición codificadas en el dominio frecuencial. En el numeral de referencia 420, se ilustra una transición desde una subtrama codificada en el modo ACELP a una trama codificada en el modo de dominio frecuencial. En el numeral de referencia 430, una transición de una trama (o una subtrama) codificada en el modo TCX-LPD (también designado como modo "wLPT") a una trama codificada en el modo de dominio frecuencial según se ilustra. En el numeral de referencia 440, se ilustra una transición entre una trama codificada en el modo de dominio frecuencial y una subtrama codificada en el modo ACELP. En el numeral de referencia 450, se ilustra una transición entre subtramas codificadas en el modo ACELP. En el numeral de referencia 460, se ilustra una transición desde una subtrama codificada en el modo TCX-LPD a una subtrama codificada en el modo ACELP. En el numeral de referencia 470, se ilustra una transición desde una trama codificada en el modo de dominio frecuencial y una subtrama codificada en el modo TCX-LPD. En el numeral de referencia 480, se ilustra una transición entre una subtrama codificada en el modo ACELP a una subtrama codificada en el modo TCX-LPD. En el numeral de referencia 490, se ilustra una transición entre subtramas codificadas en el modo.

**[0060]** De manera interesante, la transición desde el modo TCX-LPD al modo de dominio frecuencial, que se ilustra en el numeral de referencia 430, es algo ineficaz o incluso TCX-LPD es muy ineficaz debido al hecho de que parte de la información transmitida al decodificador se deja de lado. De manera similar, las transiciones entre el modo ACELP y el modo TCX-LPD, que se ilustran en los numerales de referencia 460 y 480, están implementadas de manera ineficaz debido al hecho de que una parte de la información transmitida al decodificador se deja de lado.

### 3.3 Decodificador de señal de audio 360 conforme a la Fig. 3b

**[0061]** A continuación, se describirá el decodificador de señal de audio 360, conforme a un ejemplo.

**[0062]** La señal de audio 360 comprende un multiplexor de bits o analizador sintáctico de corriente de bits 362, que está configurado para recibir representaciones de la corriente de bits 361 de un contenido de audio y para proporcionar, en esta base, elementos de información a ramales diferentes del decodificador de la señal de audio 360.

**[0063]** El decodificador de la señal de audio 360 comprende un ramal de dominio frecuencial 370 que recibe información del factor de escala 372 codificado e información espectral codificada 374 desde el multiplexor de la corriente de bits 362 y para proporcionar, en esta base, una presentación del dominio temporal 376 de una trama codificada en el modo del dominio frecuencial. El decodificador de la señal de audio 360 comprende también una ruta TCX-LPD 380 que está configurada para recibir una representación espectral codificada 382 y coeficientes de filtrado de codificación de predicción lineal codificados 384 para proporcionar, en esta base, una representación de dominio temporal 386 de una trama de audio o subtrama de audio codificada en el modo TCX-LPD.

**[0064]** El decodificador de la señal de audio 360 comprende una ruta ACELP 390 que está configurada para recibir una excitación ACELP codificada 392 y coeficientes de filtrado de codificación de predicción lineal codificados 394 para proporcionar, en esta base, una representación de dominio temporal 396 de una subtrama de audio codificada en el modo ACELP.

**[0065]** El decodificador de la señal de audio 360 comprende también una formación en ventana de transición 398, que está configurada para aplicar una formación en ventana de transición apropiada a las representaciones del dominio temporal 376, 386, 396 de las tramas y las subtramas codificadas en los modos diferentes, para derivar una señal de audio contigua.

**[0066]** Debería observarse aquí que el ramal del dominio frecuencial 370 puede ser idéntico en su estructura general y funcionalidad respecto al ramal de dominio frecuencial 320, a pesar de que hay mecanismos diferentes o adicionales de cancelación de solapamiento en el ramal del dominio frecuencial 370. Además, el ramal ACELP 390

puede ser idéntico al ramal ACELP 340 en su estructura general y funcionalidad, de manera que la descripción anterior también aplica.

**[0067]** Sin embargo, el ramal TCX-LPD 380 difiere del ramal TCX-LPD 330 en lo que respecta a que el moldeado de sonido se lleva a cabo antes de la transformada de coseno discreta modificada inversa en el ramal TCX-LPD 380. También el ramal TCX-LPD 380 comprende funcionalidades adicionales de cancelación de solapamiento.

**[0068]** El ramal TCX-LPD 380 comprende un decodificador aritmético 380a que está configurado para recibir una representación espectral codificada 382 y para proporcionar, en esta base, una representación espectral decodificada 380b. El ramal TCX-LPD 380 comprende también un cuantificador inverso 380c configurado para recibir la representación espectral decodificada 380b y para proporcionar, en esta base, una representación espectral cuantificada de manera inversa 380d. El ramal TCX-LPD 380 comprende también una escala y/o moldeado de sonido de dominio frecuencial 380e que está configurado para recibir la representación espectral inversamente cuantificada 380d y una información de moldeado espectral 380f y para proporcionar, en esta base, una representación espectral con moldeado espectral 380g a una transformada de coseno discreta modificada inversa 380h, que proporciona la representación de dominio temporal 386 en la base de la representación espectral con moldeado espectral 380g. El ramal TCX-LPD 380 comprende también una transformada de coeficiente de predicción lineal a dominio frecuencial 380i que está configurada para proporcionar la información de escala espectral 380f en base a los coeficientes de filtrado de codificación de predicción lineal 384.

**[0069]** En relación a la funcionalidad del decodificador de señal de audio 360 se puede decir que el ramal del dominio frecuencial 370 y el ramal TCX-LPD 380 son muy similares en que cada uno de ellos comprende una cadena de procesamiento que tiene una decodificación aritmética, una cuantificación inversa, una escala espectral y una transformada de coseno discreta modificada en el mismo orden de procesamiento. Por consiguiente, las señales de salida 376, 386 del ramal del dominio frecuencial 370 y del ramal TCX-LPD 380 son muy similares en que ambas pueden ser señales de salida no filtradas (con la excepción de la formación en ventana de transición) de las transformadas de coseno discreta modificada. Por consiguiente, las señales de dominio temporal 376, 386 son muy adecuadas para una operación de superposición y suma, donde se logra una cancelación de solapamiento de dominio temporal mediante operación de superposición y suma. De este modo, las transiciones entre la trama de audio codificadas en el modo de dominio frecuencial y una trama de audio o subtrama de audio codificada en el modo TCX-LPD pueden llevarse a cabo de manera eficaz por una simple operación de superposición y suma sin requerir ninguna información adicional de cancelación de solapamiento y sin dejar de lado ninguna información. De este modo, es suficiente una cantidad mínima de información secundaria.

**[0070]** Además, debería observarse que la escala de la representación espectral cuantificada de manera inversa, que se lleva a cabo en la ruta del dominio frecuencial 370 dependiendo de la información del factor de escala, trae de manera eficaz un moldeado de sonido del sonido de cuantificación introducido por la cuantificación realizada por el codificador y la cuantificación inversa realizada por el decodificador 320c, dicho moldeado de sonido está bien adaptado a las señales de audio general tal como, por ejemplo, señales de música. En contraste, el moldeado de sonido de escala y/o de dominio frecuencial 380e, que se realiza dependiendo de los coeficientes de filtrado de codificación de predicción lineal, trae de manera eficaz un moldeado de sonido de un sonido de cuantificación provocado por una cuantificación realizada por el codificador y la cuantificación inversa realizada por el decodificador 380c, que está bien adaptada a las señales de audio tipo voz. Por consiguiente, la funcionalidad del ramal de dominio frecuencial 370 y del ramal TCX-LPD 380 meramente difiere en que se aplica un moldeado de sonido diferente en el dominio frecuencial, de tal manera que una codificación eficaz (o calidad de audio) es particularmente buena para las señales de audio generales cuando se utiliza el ramal de dominio frecuencial 370, y de tal manera que es particularmente alta una eficacia de codificación o calidad de audio para las señales de audio tipo voz cuando se utiliza el ramal TCX-LPD 380.

**[0071]** Debería observarse que el ramal TCX-LPD 380 comprende preferentemente mecanismos de cancelación de solapamiento adicional para las transiciones entre las tramas de audio o subtramas de audio codificadas en el modo TCX-LPD y en el modo ACELP. Se describirán los detalles a continuación.

### 3.4 Formación en ventana de transición conforme a la Fig. 5

**[0072]** La Fig. 5 muestra una representación gráfica de un ejemplo de un esquema de formación en ventana previsto, que puede aplicarse en el decodificador de la señal de audio 360 o en cualquier otro codificador y decodificador de señal de audio conforme con la presente invención. La Fig. 5 representa una formación en ventana en posibles transiciones entre tramas o subtramas codificados en modos diferentes. Las abscisas 502a a 502i describen un tiempo con respecto a las muestras de audio y las ordenadas 504a a 504i describen las ventanas o subtramas para proporcionar una representación del dominio temporal de un contenido de audio.

**[0073]** Una representación gráfica en el numeral de referencia 510 muestra una transición entre las tramas posteriores codificadas en el modo de dominio frecuencial. Como puede verse, las muestras de dominio temporal proporcionadas para una primera mitad derecha de una trama (por ejemplo, mediante una transformada de coseno

discreta modificada de manera inversa (MDCT) 320g) se forman en ventanas por una mitad derecha 512 de una ventana, que puede ser, por ejemplo, del tipo de ventana "AAC Long " o del tipo de ventana "AAC Stop ".

De manera similar, las muestras de dominio temporal proporcionadas por una mitad izquierda de una segunda trama posterior (por ejemplo, mediante la MDCT 320g) pueden visualizarse utilizando una mitad izquierda 514 de una ventana, que puede, por ejemplo ser del tipo de ventana "AAC Long " o "AAC Start ". La mitad derecha 512 puede, por ejemplo, comprender una pendiente de transición derecha comparativamente larga y la mitad izquierda 514 de la ventana posterior puede comprender una pendiente de transición izquierda comparativamente larga. Una versión formada en ventana de la representación del dominio temporal de la primera trama de audio (formada en ventana utilizando la mitad de ventana derecha 512) y una versión formada en ventana de la representación del dominio temporal de la segunda trama de audio posterior (formada en ventana utilizando la mitad de la ventana izquierda 514) pueden estar superpuestas y sumadas. Por consiguiente, el solapamiento que surge de la MDCT, se puede cancelar eficazmente.

- 15 **[0074]** Una representación gráfica en el numeral de referencia 520, muestra una transición desde una subtrama codificada en el modo ACELP a una trama codificada en el modo de dominio frecuencial. Una cancelación de solapamiento hacia delante puede aplicarse para reducir los artefactos de solapamiento en tal transición.

- 20 **[0075]** Una representación gráfica en el numeral de referencia 530, muestra una transición desde una subtrama codificada en el modo TCX-LPD a una trama codificada en el modo de dominio frecuencial. Como puede verse, una ventana 532 se aplica a las muestras del dominio temporal proporcionadas por la MDCT inversa 380h de la ruta TCX-LPD, dicha ventana 532 puede, por ejemplo, ser del tipo de ventana "TCX256 ", "TCX512 ", o "TCX1024 ". La ventana 532 puede comprender una pendiente de transición derecha 533 de 128 muestras de longitud de dominio temporal . Una ventana 534 se aplica a las muestras de dominio temporal proporcionadas por la MDCT de la ruta de dominio frecuencial 370 para la trama de audio posterior codificada en el modo de dominio frecuencial. La ventana 534 puede, por ejemplo, ser una ventana tipo "Stop Start " o "AAC Stop ", y puede comprender una pendiente de transición izquierda 535 que tenga una longitud de, por ejemplo, 128 muestras de dominio temporal. Las muestras del dominio temporal de la subtrama del modo TCX-LPD que se forman en ventana por la pendiente de transición derecha 533 están superpuestas y sumadas con las muestras del dominio temporal de la trama de audio posterior codificada en el modo de dominio frecuencial que se visualizan en cascada por la pendiente de transición izquierda 535. Las pendientes de transición 533 y 535 coinciden, de manera que se obtiene una cancelación de solapamiento en la transición desde la subtrama codificada en modo TCX-LPD y la subtrama codificada en modo de dominio frecuencial posterior. La cancelación de solapamiento es posible mediante la ejecución del moldeado de sonido de dominio frecuencial/de escala 380e antes de la ejecución de la MDCT inversa 380h. En otras palabras, la cancelación de solapamiento es provocada por el hecho de que tanto, la MDCT inversa 320g de la ruta de dominio frecuencial 370 y la MDCT inversa 380h de la ruta TCX-LPD 380 se alimentan con coeficientes espectrales a los que ya se ha aplicado el moldeado de sonido (por ejemplo, en la forma de escala dependiente de factor de escala y de escala dependiente de coeficiente de filtro LPC).

- 40 **[0076]** Una representación gráfica en el numeral de referencia 540, ilustra una transición desde una trama de audio codificada en el modo de dominio frecuencial a una subtrama codificada en el modo ACELP . Como puede verse, una cancelación de solapamiento hacia delante (FAC) se aplica para reducir o incluso eliminar los artefactos de solapamiento en dicha transición.

- 45 **[0077]** Una representación gráfica en el numeral de referencia 550, ilustra una transición desde una subtrama de audio codificada en el modo ACELP a otra subtrama de audio codificada en el modo ACELP. No se requiere un procesamiento de cancelación de solapamiento específico en la presente en algunos ejemplos.

- 50 **[0078]** Una representación gráfica en el numeral de referencia 560, muestra una transición desde una subtrama codificada en el modo TCX-LPD (también designado como modo wLPT) a una subtrama codificada en el modo ACELP. Como puede verse, las muestras de dominio temporal proporcionadas por el MDCT 380h del ramal TCX-LPD 380 se forman en ventana utilizando una ventana 562, que puede, por ejemplo, ser del tipo de ventana "TCX256 ", "TCX512 " o "TCX1024 ". La ventana 562 comprende una pendiente de transición derecha comparativamente corta 563. Las muestras de dominio temporal proporcionadas para la subtrama de audio posterior codificada en el modo ACELP comprenden una superposición parcial temporal con las muestras de audio proporcionadas para la subtrama de audio codificada en el modo TCX-LPD precedente que se forman en ventana por la pendiente de transición derecha 563 de la ventana 562. Las muestras de audio de dominio temporal proporcionadas para la subtrama de audio codificada en el modo ACELP se ilustran mediante un bloque en el numeral de referencia 564.

- 60 **[0079]** Como puede verse, una señal de cancelación de solapamiento hacia delante 566 se agrega a la transición desde una trama de audio codificada en el modo TCX-LPD a la trama de audio codificada en el modo ACELP para reducir o incluso eliminar los artefactos de alineación. Se describirán a continuación los detalles relacionados con la provisión de la señal de cancelación de solapamiento 566.

- 65 **[0080]** Una representación gráfica en el numeral de referencia 570, ilustra una transición desde una trama

codificada en el modo del dominio frecuencial a una trama posterior codificada en el modo TCX-LPD. Las muestras de dominio temporal proporcionadas por la MDCT 320g inversa del ramal del dominio frecuencial 370 pueden formarse en ventana por una ventana 572 que tiene una pendiente de transición derecha comparativamente corta 573, por ejemplo, por una ventana del tipo "Stop Start" o una ventana del tipo "AAC Start". Una representación del dominio temporal proporcionada por la MDCT inversa 380h del ramal TCX-LPD 380 para una subtrama de audio posterior codificada en el modo TCX-LPD puede formarse en ventana por una ventana 574 que comprende una pendiente de transición del lado izquierdo comparativamente corta 575, dicha ventana 574 puede ser, por ejemplo, una ventana tipo "TCX256", "TCX512", o "TCX1024". Las muestras del dominio temporal formadas en ventana por la pendiente de transición del lado derecho 573 y las muestras del dominio temporal formadas en ventana por la pendiente de transición del lado izquierdo 575 están superpuestas y sumadas por la formación en ventana de transición 398, de manera que los artefactos de solapamiento están reducidos o incluso eliminados. Por consiguiente, no se requiere información secundaria adicional para realizar una transición desde una trama de audio codificada en el modo de dominio frecuencial a una subtrama de audio codificada en el modo TCX-LPD.

- 15 **[0081]** Una representación gráfica en el numeral de referencia 580 muestra una transición desde una trama de audio codificada en el modo ACELP a una trama de audio codificada en el modo TCX-LPD (también designado como modo wLPT). Una región temporal para la cual se proporcionan muestras de dominio temporal por el ramal ACELP se designa con 582. Una ventana 584 se aplica a las muestras del dominio temporal proporcionadas por la MDCT inversa 380h del ramal TCX-LPD 380. La ventana 584, que puede ser del tipo "TCX256", "TCX512", o "TCX1024", puede comprender una pendiente de transición izquierda comparativamente corta 585. La pendiente de transición izquierda 585 de la ventana 584 se superpone parcialmente con las muestras del dominio temporal proporcionadas por el ramal ACELP, que están representadas por el bloque 582. Además, se proporciona una señal de cancelación de solapamiento 586 para reducir, o incluso eliminar, los artefactos de solapamiento que ocurren en la transición desde la subtrama de audio codificada en el modo ACELP a la subtrama de audio codificada en el modo TCX-LPD. Se tratarán a continuación los detalles relacionados con la provisión de la señal de cancelación de solapamiento 586.

- [0082]** Una representación esquemática en el numeral de referencia 590, muestra una transición desde una subtrama de audio codificada en el modo TCX-LPD a otra subtrama de audio codificada en el modo TCX-LPD. Las muestras del dominio temporal de una primera subtrama de audio codificada en el modo TCX-LPD se forman en ventana con una ventana 592, que puede, por ejemplo, ser del tipo "TCX256", "TCX512", o "TCX1024", y puede comprender una pendiente de transición derecha comparativamente corta 593. Las muestras de audio del dominio temporal de una segunda subtrama de audio codificada en el modo TCX-LPD, que se proporciona por la MDCT inversa 380h del ramal TCX-LPD 380 se visualizan, por ejemplo, con una ventana 594 que puede ser del tipo "TCX256", "TCX512", o "TCX1024" y puede comprender una pendiente de transición izquierda comparativamente corta 595. Las muestras del dominio temporal visualizadas utilizando una pendiente de transición derecha 593 y las muestras del dominio temporal visualizadas utilizando la pendiente de transición izquierda 595 están superpuestas y sumadas por la formación de ventana transicional 398. Por consiguiente, lo que se provoca por la MDCT 380h (inversa) se reduce o incluso elimina.

#### 40 4. Panorama sobre todos los tipos de ventana

- [0083]** A continuación, se proporcionará un panorama de todos los tipos de ventana. Para este fin, se hace referencia a la Fig. 6, que muestra una representación gráfica de los diferentes tipos de ventana y sus características. En la tabla de la Fig. 6, una columna 610 describe una longitud de superposición izquierda, que puede ser igual a la longitud de la pendiente de transición izquierda. La columna 612 describe una longitud de la transformada, es decir, una cantidad de coeficientes espectrales utilizados para generar la representación del dominio temporal que se forma en ventana por la ventana respectiva. La columna 614 describe una longitud de superposición derecha, que puede ser igual a la longitud de una pendiente de transición derecha. Una columna 616 describe un nombre del tipo de ventana. La columna 618 muestra una representación gráfica de la ventana respectiva.

- 50 **[0084]** Una primera fila 630 muestra las características de una ventana del tipo "AAC Short". Una segunda fila 632 muestra las características de una ventana del tipo "TCX256". Una tercera fila 634 muestra las características de una ventana del tipo "TCX512". Una cuarta fila 636 muestra las características de las ventanas de los tipos "TCX1024" y "Stop Start". Una quinta fila 638 muestra las características de una ventana del tipo "AAC Long". Una sexta fila 640 muestra las características de una ventana de tipo "AAC Start", y una séptima fila 642 muestra las características de una ventana del tipo "AAC Stop".

- [0085]** Notablemente, las pendientes de transición de las ventanas del tipo "TCX256", "TCX512", y "TCX1024" están adaptadas a la pendiente de transición derecha de la ventana del tipo "AAC Start" y a la pendiente de transición izquierda de la ventana del tipo "AAC Stop", para permitir una cancelación de solapamiento del dominio temporal superponiendo y sumando representaciones del dominio temporal visualizadas utilizando diferentes tipos de ventanas. En un ejemplo, las pendientes de ventanas de lado izquierdo (pendientes de transición) de todos los tipos de ventana que tienen longitudes de superposición de lado izquierdo idénticas pueden ser idénticas y las pendientes de transición de lado derecho de todos los tipos de ventana que tienen longitudes de superposición de lado derecho idénticas pueden ser idénticas. También, las pendientes de transición de lado izquierdo y las pendientes de transición de lado

derecho que tienen longitudes de superposición idénticas pueden adaptarse para permitir una cancelación de solapamiento, cumpliendo las condiciones de la cancelación de solapamiento MDCT.

#### 5. Secuencias de ventana permitidas

5

**[0086]** A continuación, las secuencias de ventana permitidas se describirán, haciendo referencia a la Fig. 7, lo que muestra una representación en tabla de dichas secuencias de ventana permitidas. Como se puede ver en la tabla de la Fig. 7, una trama de audio codificada en el modo de dominio frecuencial, cuyas muestras de dominio temporal se visualizan con una ventana de tipo "AAC Stop", pueden ser continuadas por una trama de audio codificada en el modo de dominio frecuencial, cuyas muestras de dominio temporal se visualizan utilizando una ventana de tipo "AAC Long" o una ventana de tipo "AAC Start".

10

**[0087]** Una trama de audio codificada en el modo de dominio frecuencial, cuyas muestras de dominio temporal se visualizan con una ventana de tipo "AAC Long" pueden ser seguidas por una trama de audio codificada en el modo de dominio frecuencial, cuyas muestras de dominio temporal se visualizan utilizando una ventana de tipo "AAC Long" o "AAC Start".

15

**[0088]** Las tramas de audio codificadas en el modo de predicción lineal, cuyas muestras de dominio temporal se visualizan utilizando una ventana de tipo "AAC Start", usando ocho ventanas del tipo "AAC Short" o utilizando una ventana de tipo "AAC Stop-Start", pueden estar seguidas por una trama de audio codificada en el modo de dominio frecuencial, cuyas tramas de tiempo se visualizan con ocho ventanas de tipo "AAC Short", usando una ventana de tipo "AAC Short" o usando una ventana de tipo "AAC Stop-Start". Alternativamente, las tramas de audio codificadas en el modo de dominio frecuencial, cuyas muestras del dominio temporal se visualizan utilizando una ventana de tipo "AAC Start", con ocho ventanas de tipo "AAC Short" o con una ventana de tipo "AAC Stop-Start" pueden estar seguidas por una trama o una subtrama de audio codificada en el modo TCX-LPD (también designado como LPD-TCX) o por una trama de audio o subtrama de audio codificada en el modo ACELP (también designado como LPD-ACELP).

20

25

**[0089]** Una trama de audio o una subtrama de audio codificada en el modo TCX-LPD puede estar seguida por tramas de audio codificadas en el modo de dominio frecuencial, cuyas muestras de dominio temporal se visualizan utilizando ocho ventanas "AAC Short" y utilizando la ventana "AAC Stop" o utilizando una ventana "AAC Stop-Start", o por una trama de audio o una subtrama de audio codificada en el modo TCX-LPD o por una trama de audio o una subtrama de audio codificada en el modo ACELP.

30

**[0090]** Una trama de audio codificada en el modo ACELP puede estar seguida por tramas de audio codificadas en el modo de dominio frecuencial, cuyas muestras de dominio temporal se visualizan con ocho ventanas "AAC Short", utilizando una ventana "AAC Stop", con una ventana "AAC Stop-Start" mediante una trama de audio codificada en el modo TCX-LPD o mediante una trama de audio codificada en el modo ACELP.

35

**[0091]** Para las transiciones desde una trama de audio codificada en el modo ACELP hacia una trama de audio codificada en el modo de dominio frecuencial o hacia una trama de audio codificada en el modo TXC-LPD, se realiza la bien conocida cancelación de solapamiento hacia delante (FAC). Por consiguiente, una señal de síntesis de cancelación de solapamiento se suma a la representación del dominio temporal en dicha transición de trama, donde los artefactos de solapamiento se reducen o incluso se eliminan. De manera similar, se realiza también una FAC cuando se cambia desde una trama o una subtrama codificada en el modo de dominio frecuencial, o desde una trama o una subtrama codificada en el modo TCX-LPD, a una trama o subtrama codificada en el modo ACELP.

40

45

**[0092]** Los detalles relacionados con la FAC se tratarán a continuación.

#### 50 6. Codificador de la señal de audio según la Fig. 8

**[0093]** A continuación, se describirá un codificador de señal de audio de modos múltiples 800 tomando como referencia la Fig. 8.

55

**[0094]** El codificador de la señal de audio 800 está configurado para recibir una representación de entrada 810 de un contenido de audio y para proporcionar, sobre esta base, una corriente de bits 812 que representa el contenido de audio. El codificador de la señal de audio 800 está configurado para operar en modos múltiples de operación, a saber un modo de dominio frecuencial, un modo de dominio de predicción lineal con excitación por código de la transformada y un modo de dominio de predicción lineal excitada por código algebraico. El codificador de señal de audio 800 comprende un controlador de codificación 814 que está configurado para seleccionar uno de los modos para codificar una porción del contenido de audio dependiendo de las características de la representación de entrada 810 del contenido de audio y/o dependiendo de la eficacia o calidad de la codificación alcanzable.

60

**[0095]** El codificador de la señal de audio 800 comprende un ramal de dominio frecuencial 820 que está configurado para proporcionar coeficientes espectrales codificados 822, factores de escala codificados 824, y

65

opcionalmente, coeficientes de cancelación de solapamiento codificados 826, en base a la representación de entrada 810 del contenido de audio. El codificador de la señal de audio 800 comprende un ramal TCX-LPD 850 que está configurado para proporcionar coeficientes espectrales codificados 852, parámetros de dominio de predicción lineal codificados 854, y coeficientes de cancelación de solapamiento codificados 856, dependiendo de la representación de entrada 810 del contenido de audio. El decodificador de la señal de audio 800 comprende también un ramal ACELP 880 que está configurado para proporcionar una excitación ACELP codificada 882 y parámetros de dominio de predicción lineal codificados 884 dependiendo de la representación de entrada 810 del contenido de audio.

**[0096]**

El ramal del dominio frecuencial 820 comprende una conversión del dominio temporal al dominio frecuencial 830 que está configurado para recibir la representación de entrada 810 del contenido de audio, o una versión preprocesada de la misma, y para proporcionar, en esta base, una representación del dominio frecuencial 832 del contenido de audio. El ramal del dominio frecuencial 820 comprende también un análisis psicoacústico 834, que está configurado para evaluar los efectos de ocultamiento de frecuencia y/o los efectos de ocultamiento de tiempo del contenido de audio, y para proporcionar, en esta base, una información del factor de escala 836 que describe los factores de escala. El ramal del dominio frecuencial 820 comprende también un procesador espectral 838 configurado para recibir la representación del dominio frecuencial 832 del contenido de audio y la información del factor de escala 836 y para aplicar una escala dependiente de frecuencia y de tiempo a los coeficientes espectrales de la representación del dominio frecuencial 832 dependiendo de la información del factor de escala 836, para obtener una representación del dominio frecuencial en escala 840 del contenido de audio. El ramal de dominio frecuencial comprende también una cuantificación/codificación 842 configurada para recibir la representación del dominio frecuencial en escala 840 y para realizar una cuantificación y una codificación para obtener los coeficientes espectrales codificados 822 en base a una representación del dominio frecuencial en escala 840. El ramal del dominio frecuencial comprende también una cuantificación/codificación 844 configurada para recibir la información del factor de escala 836 y para proporcionar, en esta base, una información del factor de escala codificado 824. Opcionalmente, el ramal del dominio frecuencial 820 comprende también un cálculo del coeficiente de cancelación de solapamiento 846 que puede estar configurado para proporcionar los coeficientes de la cancelación de solapamiento 826.

**[0097]**

El ramal TCX-LPD 850 comprende una conversión del dominio temporal al dominio frecuencial 860, que puede estar configurado para recibir la representación de entrada 810 del contenido de audio, y para proporcionar en esta base, una representación del dominio frecuencial 861 del contenido de audio. El ramal TCX-LPD 850 comprende también un cálculo del parámetro del dominio de predicción lineal 862 que está configurado para recibir la representación de entrada 810 del contenido de audio, o una versión preprocesada del mismo, y para derivar uno o más parámetros de dominio de predicción lineal (por ejemplo, los coeficientes del filtro de codificación de predicción lineal) 863 desde la representación de entrada 810 del contenido de audio. El ramal TCX-LPD 850 comprende también una conversión del dominio de predicción lineal al dominio espectral 864, que está configurada para recibir los parámetros del dominio de predicción lineal (por ejemplo, los coeficientes del filtro de codificación de predicción lineal) y para proporcionar una representación del dominio espectral o una representación del dominio frecuencial 865 en esta base. La representación del dominio espectral o la representación del dominio frecuencial de los parámetros del dominio de predicción lineal puede, por ejemplo, representar una respuesta de filtro de un filtro definido por los parámetros de dominio de predicción lineal en el dominio frecuencial o en el dominio espectral. El ramal TCX-LPD 850 comprende también un procesador espectral 866, que está configurado para recibir la representación del dominio frecuencial 861, o una versión preprocesada 861' del mismo, y la representación del dominio frecuencial o la representación del dominio espectral de los parámetros del dominio de predicción lineal 863. El procesador espectral 866 está configurado para realizar un moldeado espectral de la representación del dominio frecuencial 861, o de la versión preprocesada 861' del mismo, donde la representación del dominio frecuencial o la representación del dominio espectral 865 de los parámetros del dominio de predicción lineal 863 sirven para ajustar la escala de los diferentes coeficientes espectrales de la representación del dominio frecuencial 861 o de la versión preprocesada 861' del mismo. Por consiguiente, el procesador espectral 866 proporciona una versión de moldeado espectral 867 de la representación del dominio frecuencial 861 o de la versión preprocesada 861' del mismo, dependiendo de los parámetros de dominio de predicción lineal 863. El ramal TCX-LPD 850 comprende también una cuantificación/codificación 868 que está configurada para recibir la representación del dominio frecuencial de moldeado espectral 867 y para proporcionar, en esta base, coeficientes espectrales codificados 852. El ramal TCX-LPD 850 comprende también otra cuantificación/codificación 869, que está configurada para recibir los parámetros de dominio de predicción lineal 863 y para proporcionar, en esta base, los parámetros del dominio de predicción lineal codificada 854.

**[0098]**

El ramal TCX-LPD 850 comprende además una provisión del coeficiente de cancelación de solapamiento que está configurado para proporcionar los coeficientes de cancelación de solapamiento codificados 856. La provisión del coeficiente de cancelación de solapamiento comprende un cálculo de error 870 que está configurado para calcular una información de error de solapamiento 871 dependiendo de los coeficientes espectrales codificados, así como dependiendo de la representación de entrada 810 del contenido de audio. El cálculo de error 870 puede opcionalmente tomar en consideración una información 872 en relación a los componentes de cancelación de solapamiento adicionales, que pueden proporcionarse por otros mecanismos. La provisión del coeficiente de cancelación de solapamiento comprende también un cálculo del filtro de análisis 873 que está configurado para proporcionar una información 873a que describa un filtrado de error dependiendo de los parámetros de dominio de



predicción lineal 863. La provisión del coeficiente de cancelación de solapamiento comprende también un filtrado de análisis de error 874, que está configurado para recibir la información de error de solapamiento 871 y la información de configuración del filtro de análisis 873a, y para aplicar un filtrado de análisis de error, que se ajusta dependiendo de la información del filtrado de análisis 873a, a la información de error de solapamiento 871, para obtener una información de error de solapamiento filtrado 874a. La provisión de cancelación de solapamiento comprende también una conversión de dominio temporal al dominio de frecuencia 875, que puede tomar la funcionalidad de una transformada de coseno discreta de tipo IV, y que está configurada para recibir la información de error de solapamiento filtrado 874a y para proporcionar en esta base, una representación del dominio frecuencial 875a de la información de error de solapamiento filtrado 874a. La provisión del coeficiente de cancelación de solapamiento comprende también una cuantificación/codificación 876 que está configurada para recibir la representación del dominio frecuencial 875a, y para proporcionar en esta base, coeficientes de cancelación de solapamiento 856, de manera que los coeficientes de cancelación de solapamiento codificados 856 codifiquen la representación del dominio frecuencial 875a.

**[0099]** La provisión del coeficiente de cancelación de solapamiento comprende también un cálculo opcional 877 de la contribución de ACELP a una cancelación de solapamiento. El cálculo 877 puede estar configurado para computar o estimar una contribución a una cancelación de solapamiento que puede derivarse de una subtrama de audio codificada en el modo ACELP que precede a una trama de audio codificada en el modo TCX-LPD. El cálculo de la contribución de ACELP respecto a la cancelación de solapamiento puede comprender un cálculo de una síntesis post-ACELP, una formación en ventana de la síntesis post-ACELP y un solapamiento de síntesis post-ACELP formada en ventana, para obtener la información 872 en relación a los componentes de cancelación de solapamiento adicionales, que pueden derivarse de una subtrama de audio anterior en el modo ACELP. Además, o alternativamente, el cálculo 877 puede comprender un cálculo de una respuesta de entrada cero de un filtro inicializado por una decodificación de una subtrama de audio anterior codificada en el modo ACELP y una ventana de dicha respuesta de entrada cero, para obtener la información 872 sobre los componentes de cancelación de solapamiento adicionales.

**[0100]** A continuación, el ramal ACELP 880 se tratará brevemente. El ramal ACELP 880 comprende un cálculo de parámetro de dominio de predicción lineal 890 que está configurado para computar los parámetros del dominio de predicción lineal 890a en base a la representación de entrada 810 del contenido de audio. El ramal ACELP 880 comprende también un cálculo de excitación ACELP 892 configurado para computar una información de excitación ACELP 892 dependiendo de la representación de entrada 810 del contenido de audio y de los parámetros del dominio de predicción lineal 890a. El ramal ACELP 880 comprende también una codificación 894 configurada para codificar la información de excitación ACELP 892, para obtener la excitación ACELP codificada 882. Además, el ramal ACELP 880 comprende también una cuantificación/codificación 896 configurada para recibir los parámetros del dominio de predicción lineal 890a y para proporcionar, en esta base, los parámetros del dominio de predicción lineal 884.

**[0101]** El decodificador de la señal de audio 800 comprende también un formateador de la corriente de bits 898 que está configurado para proporcionar corriente de bits 812 en la base de los coeficientes espectrales codificados 822, la información del factor de escala codificado 824, los coeficientes de cancelación de solapamiento 826, los coeficientes espectrales codificados 852, los parámetros del dominio de predicción lineal codificados 852, los coeficientes de cancelación de solapamiento codificados, la excitación ACELP codificada 882, y los parámetros del dominio de predicción lineal codificados 884.

**[0102]** Se describirán a continuación los detalles relacionados con la provisión de los coeficientes de cancelación de solapamiento 852 codificados.

## 7. Decodificador de la señal de audio según la Fig. 9

**[0103]** A continuación, se describirá un decodificador de la señal de audio 900 conforme a la Fig. 9.

**[0104]** El decodificador de la señal de audio 900 conforme a la Fig. 9 es similar al decodificador de la señal de audio 200 conforme a la Fig. 2 y también al decodificador de la señal de audio 360 conforme a la Fig. 3b, de manera que se mantienen las explicaciones anteriores también.

**[0105]** El decodificador de la señal de audio 900 comprende un multiplexor de bits 902 que está configurado para recibir una corriente de bits y para proporcionar información extraída desde la corriente de bits a las rutas de procesamiento correspondientes.

**[0106]** El decodificador de señal de audio 900 comprende un ramal de dominio frecuencial 910 que está configurado para recibir coeficientes espectrales codificados 912 e información de factor de escala codificada 914. El ramal del dominio frecuencial 910 está opcionalmente configurado para recibir también coeficientes de cancelación de solapamiento codificados, que permiten la bien llamada cancelación de solapamiento hacia delante, por ejemplo, en una transición entre una trama de audio codificada en el modo de dominio frecuencial y una trama de audio codificada en el modo ACELP. La ruta de dominio frecuencial 910 proporciona una representación de dominio temporal 918 del contenido de audio de la trama de audio codificada en el modo de dominio frecuencial.

**[0107]** El decodificador de señal de audio 900 comprende un ramal TCX-LPD 930, que está configurado para recibir coeficientes espectrales codificados 932, parámetros del dominio de predicción lineal codificados 934 y coeficientes de cancelación de solapamiento codificados 936, y para proporcionar, en esta base, una representación del dominio temporal de la trama de audio o una subtrama codificada en el modo TCX-LPD. El decodificador de señal de audio 900 comprende también un ramal ACELP 980, que está configurado para recibir una excitación ACELP codificada 982 y parámetros del dominio de predicción lineal codificados 984 y para proporcionar, en esta base, una representación del dominio temporal 986 de la trama de audio o una subtrama de audio codificada en el modo ACELP.

#### 7.1 Ruta del dominio frecuencial

**[0108]** A continuación, se describirán detalles relacionados con la ruta del dominio frecuencial 910. Debería observarse que la ruta del dominio frecuencial es similar a la ruta del dominio frecuencial 320 del decodificador de audio 300, de manera que se hace referencia a la descripción anterior. El ramal de dominio frecuencial 910 comprende una decodificación aritmética 920, que recibe los coeficientes espectrales codificados 912 y proporciona, en esa base, los coeficientes espectrales codificados 920a, y una cuantificación inversa 921 que recibe los coeficientes espectrales decodificados 920a, y proporciona, en esa base, coeficientes espectrales cuantificados inversamente 921a. El ramal del dominio frecuencial 910 comprende también una decodificación del factor de escala 922, que recibe la información del factor de escala codificado y proporciona, en esa base, una información del factor de escala decodificado 922a. El ramal del dominio frecuencial comprende una escala 923 que recibe los coeficientes espectrales inversamente cuantificados 921a y escala los coeficientes espectrales inversamente cuantificados conforme a los factores de escala 922a, para obtener coeficientes espectrales escalados 923a. Por ejemplo, los factores de escala 922a pueden proporcionarse para una pluralidad de bandas de frecuencia, donde una pluralidad de intervalos de frecuencia de los coeficientes espectrales 921a están asociadas a cada banda de frecuencia. Por consiguiente, se puede realizar la escala de banda de frecuencia de los coeficientes espectrales 921a. De este modo, una cantidad de factores de escala asociados con una trama de audio es típicamente más pequeña en comparación con una cantidad de coeficientes espectrales 921a asociada con la trama de audio. El ramal del dominio frecuencial 910 comprende también una MDCT inversa 924, que está configurada para recibir los coeficientes espectrales en escala 923a y proporcionar, en esa base, una representación del dominio temporal 924a del contenido de audio de la trama de audio actual. El ramal del dominio frecuencial 912 comprende también, opcionalmente, una combinación 925, que está configurada para combinar la representación del dominio temporal 924a con una señal de síntesis de cancelación de solapamiento 929a, para obtener la representación del dominio temporal 918. Sin embargo, en algunas otras realizaciones la combinación 925 puede omitirse, de manera que la representación del dominio temporal 924a se proporcione como una representación del dominio temporal 918 del contenido de audio.

**[0109]** Para proporcionar la señal de síntesis de cancelación de solapamiento 929a, la ruta del dominio frecuencial comprende una decodificación 926a, que proporciona coeficientes de cancelación de solapamiento decodificados 926b, en base a los coeficientes de cancelación de solapamiento codificados 916, y una escala 926c de coeficientes de cancelación de solapamiento, que proporciona coeficientes de cancelación de solapamiento en escala 926d en base a coeficientes de cancelación de solapamiento 926b. La ruta del dominio frecuencial comprende también una transformada de coseno discreta inversa del tipo IV 927, que está configurada para recibir los coeficientes de cancelación de solapamiento en escala 926d, y para proporcionar, en esta base, una señal de estímulo de cancelación de solapamiento 927a, que se introduce en un filtrado de síntesis 927b. El filtrado de síntesis 927b está configurado para realizar una operación de filtrado de síntesis en base a la señal del estímulo de cancelación de solapamiento 927a y en dependencia de los coeficientes de filtrado de síntesis 927c, que se proporcionan por un cálculo del filtro de síntesis 927d, para obtener, como resultado del filtrado de síntesis, la señal de cancelación de solapamiento 929a. El cálculo del filtro de síntesis 927d proporciona los coeficientes del filtro de síntesis 927c dependiendo de los parámetros del dominio de predicción lineal, que pueden derivarse, por ejemplo, de los parámetros del dominio de predicción lineal proporcionados en la corriente de bits para una trama codificada en el modo TCX-LPD, o para una trama proporcionada en el modo ACELP (o puede ser igual a tales parámetros del dominio de predicción lineal).

**[0110]** Por consiguiente, el filtrado de síntesis 927b es capaz de proporcionar la señal de síntesis de cancelación de solapamiento 929a, que puede ser equivalente a la señal de síntesis de cancelación de solapamiento 522 ilustrada en la Fig. 5, o a la señal de síntesis de cancelación de solapamiento 542 ilustrada en la Fig. 5.

#### 7.2 Ruta TCX-LPD

**[0111]** A continuación, la ruta TCX-LPD del decodificador de señal de audio 900 se tratará resumidamente. Se proporcionarán a continuación detalles adicionales.

**[0112]** La ruta TCX-LPD 930 comprende una síntesis de señal principal 940 que está configurada para proporcionar una representación del dominio temporal 940a del contenido de audio de la trama de audio o subtrama de audio en base a los coeficientes espectrales codificados 932 y a los parámetros del dominio de predicción lineal codificados 934. El ramal TCX-LPD 930 comprende también un procesamiento de cancelación de solapamiento que se describirá a continuación.

**[0113]** La síntesis de señal principal 940 comprende una decodificación aritmética 941 de coeficientes espectrales, donde los coeficientes espectrales decodificados 941a se obtienen en base a los coeficientes espectrales codificados 932. La síntesis de señal principal 940 comprende también una cuantificación inversa 942, que está configurada para proporcionar coeficientes espectrales inversamente cuantificados 942a en base a coeficientes espectrales decodificados 941a. Un relleno de sonido opcional 943 puede aplicarse a los coeficientes espectrales inversamente cuantificados 942a para obtener coeficientes espectrales rellenos con sonido. El coeficiente espectral inversamente cuantificado y relleno con sonido 943a puede estar designado también con  $r[i]$ . Los coeficientes espectrales inversamente cuantificados y con ruido 943a,  $r[i]$  pueden ser procesados por un desmoldeado espectral 944, para obtener coeficientes espectrales con desmoldeado espectral 944a, que también a veces son designados con  $r[i]$ . Una escala 945 puede configurarse como un moldeado de sonido del dominio temporal 945. En el moldeado de sonido del dominio frecuencial 945, se obtiene un conjunto con moldeado espectral de coeficientes espectrales 945a, que también se designa con  $rr[i]$ . En el moldeado de sonido de dominio frecuencial 945, los aportes de los coeficientes espectrales con desmoldeado espectral 944a en los coeficientes espectrales con moldeado espectral 945a están determinados por los parámetros de moldeado de sonido del dominio frecuencial 945b, que se proporcionan por una provisión de parámetros de moldeado de sonido del dominio frecuencial que se tratará a continuación. Mediante el moldeado de sonido del dominio frecuencial 945, se les da a los coeficientes espectrales del conjunto con desmoldeado espectral de coeficientes espectrales 944a un peso comparativamente grande, si una respuesta de dominio frecuencial de un filtro de predicción lineal descrito por los parámetros del dominio de predicción lineal 934 toma un valor comparativamente pequeño para la frecuencia asociada con el coeficiente espectral respectivo (de un conjunto 944a de coeficientes espectrales) en consideración. Por otro lado, se le proporciona a un coeficiente espectral del conjunto 944a del coeficiente espectral un peso comparativamente más grande cuando se obtienen los coeficientes espectrales correspondientes del conjunto 945a de coeficientes espectrales con moldeado espectral, si la respuesta del dominio frecuencial del filtro de predicción lineal descrito por los parámetros del dominio de predicción lineal 934 toma un valor comparativamente pequeño para la frecuencia asociada con el coeficiente espectral (del conjunto 944a) en consideración. Por consiguiente, un moldeado espectral, que se define por los parámetros de predicción lineal 934, se aplica en el dominio frecuencial cuando deriva el coeficiente espectral con moldeado espectral 945a del coeficiente espectral con desmoldeado espectral 944a.

**[0114]** La síntesis de señal principal 940 comprende también una MDCT inversa 946, que está configurada para recibir los coeficientes espectrales con moldeado espectral 945a, y para proporcionar, en esta base, una representación del dominio temporal 946a. Una escala de ganancia 947 se aplica a la representación del dominio temporal 946a, para derivar la representación del dominio temporal 940a del contenido de audio de la señal del dominio temporal 946a. Un factor de ganancia  $g$  se aplica en la escala de ganancia 947, que es preferentemente una operación independiente de la frecuencia (selectiva de no frecuencia).

**[0115]** La síntesis de señal principal comprende también un procesamiento de los parámetros del moldeado de sonido del dominio frecuencial 945b, que se describirá a continuación. Para los fines de proporcionar los parámetros del moldeado de sonido del dominio frecuencial 945b, la síntesis de la señal principal 940 comprende una decodificación 950, que proporciona parámetros del dominio de predicción lineal decodificados 950a en base a los parámetros del dominio de predicción lineal codificados 934. Los parámetros del dominio de predicción lineal decodificados pueden, por ejemplo, tomar la forma de un primer conjunto LPC1 de parámetros de dominio de predicción lineal decodificados y un segundo conjunto LPC2 de los parámetros del dominio de predicción lineal. El primer conjunto LPC1 de los parámetros de dominio de predicción lineal pueden, por ejemplo, estar asociados con una transición izquierda de una trama o subtrama codificada en el modo TCX-LPD, y el segundo conjunto LPC2 de los parámetros de dominio de predicción lineal puede estar asociado con una transición derecha de la trama de audio o subtrama de audio codificada TCX-LPD. Los parámetros del dominio de predicción lineal decodificados son alimentados en un cálculo espectral 951, que proporciona una representación del dominio frecuencial de una respuesta de impulso definida por los parámetros del dominio de predicción lineal 950a. Por ejemplo, los conjuntos separados de coeficientes de dominio frecuencial  $X_0[k]$  pueden proporcionarse para el primer conjunto LPC1 y para el segundo conjunto LPC2 de los parámetros del dominio de predicción lineal decodificada 950.

**[0116]** Un cálculo de ganancia 952 mapea los valores espectrales  $X_0[k]$  en valores de ganancia, donde un primer conjunto de valores de ganancia  $g_1[k]$  está asociado con el primer conjunto LPC1 de coeficientes espectrales y donde un segundo conjunto de valores de ganancia  $g_2[k]$  está asociado con el segundo conjunto LPC2 de coeficientes espectrales. Por ejemplo, los valores de ganancia pueden ser inversamente proporcionales a una magnitud de los coeficientes espectrales correspondientes. Un cálculo de parámetros de filtro 953 puede recibir los valores de ganancia 952a y proporcionar, en esta base, los parámetros de filtro 945b para el moldeado del dominio frecuencial. Por ejemplo, se pueden proporcionar los parámetros de filtro  $a[i]$  y  $b[i]$ . Los parámetros de filtro 945d determinan la contribución de los coeficientes espectrales con desmoldeado espectral 944a en los coeficientes espectrales con moldeado espectral 945a. Los detalles relacionados con un posible cálculo de los parámetros de filtro se proporcionarán a continuación.

**[0117]** El ramal TCX-LPD 930 comprende un cálculo de señal de síntesis de cancelación de solapamiento hacia delante, que comprende dos ramales. Un primer ramal de la generación de señal de síntesis de cancelación de solapamiento (hacia delante) comprende una decodificación 960, que está configurada para recibir coeficientes de

cancelación de solapamiento codificado 936, y para proporcionar en esta base, coeficientes de cancelación de solapamiento decodificados 960a, que se presentan en escala mediante una escala 961 dependiendo del valor de ganancia  $g$  para obtener coeficientes de cancelación de solapamiento en escala 961a. El mismo valor de ganancia  $g$  puede utilizarse para la escala 961 de los coeficientes de cancelación de solapamiento 960a y para la escala de ganancia 947 de la señal de dominio temporal 946a proporcionada por la MDCT inversa 946 en algunas realizaciones. La generación de señal de síntesis de cancelación de solapamiento comprende también un desmoldeado espectral 962, que puede estar configurado para aplicar un desmoldeado espectral a los coeficientes de cancelación de solapamiento en escala 961a, para obtener una ganancia en escala y coeficientes de cancelación de solapamiento con desmoldeados espectrales 962a. El desmoldeado espectral 962 puede realizarse de manera similar al desmoldeado espectral 944, que se describirá en más detalle a continuación. Los coeficientes de cancelación de solapamiento desmoldeados en escala de ganancia y espectrales 962a se introducen en una transformada de coseno discreta inversa de tipo IV, que está designada con un numeral de referencia 963, y que proporciona una señal de estímulo de cancelación de solapamiento 963a como resultado de la transformada de coseno discreta inversa que se realiza en base a los coeficientes de cancelación de solapamiento con desmoldeado espectral de ganancia en escala 962a. Un filtrado de síntesis 964 recibe la señal del estímulo de cancelación de solapamiento 963a y proporciona una primera señal de síntesis de cancelación de solapamiento hacia delante 964a mediante el filtrado de síntesis de la señal del estímulo de cancelación de solapamiento 963a utilizando un filtro de síntesis configurado según los coeficientes del filtro de síntesis 965a, que son proporcionados por el cálculo del filtro de síntesis 965 según los parámetros del dominio de predicción lineal LPC1, LPC2. Se describirán a continuación los detalles relacionados con el filtrado de síntesis 964 y el cálculo de los coeficientes del filtro de síntesis 965a.

**[0118]** La primera señal de síntesis de cancelación de solapamiento 964a se basa consecuentemente en los coeficientes de cancelación de solapamiento 936 así como en los parámetros del dominio de predicción lineal. Una buena consistencia entre la señal de síntesis de cancelación de solapamiento 964a y la representación del dominio temporal 940a del contenido de audio se alcanza aplicando el mismo factor de escala  $g$  tanto en la provisión de la representación del dominio temporal 940a del contenido de audio y en la provisión de la señal de síntesis de cancelación de solapamiento 964, y aplicando un desmoldeado espectral 944, 962 similar, o incluso idéntico, en la provisión de la representación del dominio temporal 940a del contenido de audio y en la provisión de la señal de síntesis de cancelación de solapamiento 964.

**[0119]** El ramal TCX-LPD 930 comprende además una provisión de las señales de síntesis de cancelación de solapamiento 973a, 976a según una trama o subtrama ACELP anterior. Este cálculo 970 de una contribución ACELP a la cancelación de solapamiento está configurado para recibir información ACELP tal como, por ejemplo una representación del dominio temporal 986 proporcionada por el ramal ACELP 980 y/o un contenido del filtro de síntesis ACELP. El cálculo 970 de la contribución ACELP a la cancelación de solapamiento comprende un cálculo 971 de una síntesis post-ACELP 971a, una formación en ventana 972 de la síntesis post-ACELP 971a y un plegado 973 de la síntesis post-ACELP 972a. Por consiguiente, se obtiene una síntesis plegada y formada en ventana post-ACELP 973a mediante el plegado de una síntesis formada en ventana post-ACELP 972a. Además, el cálculo 970 de una contribución ACELP para la cancelación de solapamiento comprende también un cálculo 975 de una respuesta de entrada cero, que puede computarse para un filtro de síntesis utilizado para sintetizar una representación de dominio temporal de una subtrama ACELP anterior, donde el estado inicial de dicho filtro de síntesis puede ser igual al estado de un filtro de síntesis ACELP en el extremo de la subtrama ACELP anterior. Por consiguiente, se obtiene una respuesta de entrada cero 975a, a la cual se aplica una formación en ventana 976 para obtener una respuesta de entrada cero formada en ventana 976a. Se describirán a continuación los detalles adicionales relacionados con la provisión de una respuesta de entrada cero formada en ventana 976a.

**[0120]** Finalmente, se realiza una combinación 978 para combinar la representación del dominio temporal 940a del contenido de audio, la primera señal de síntesis de cancelación de solapamiento hacia delante 964a, la segunda señal de síntesis de cancelación de solapamiento hacia delante 973a y la tercera señal de síntesis de cancelación de solapamiento hacia delante 976a. Por consiguiente, la representación del dominio temporal 938 de la trama de audio o de la subtrama de audio codificada en el modo TCX-LPD se proporciona como un resultado de la combinación 978, como se describirá en más detalle a continuación.

### 7.3 Ruta ACELP

**[0121]** A continuación, se describirá resumidamente el ramal ACELP 980 del decodificador de señal de audio 900. El ramal ACELP 980 comprende una decodificación 988 de la excitación ACELP codificada 982, para obtener una excitación ACELP decodificada 988a. Posteriormente, se lleva a cabo un cálculo y postprocesamiento de la señal de excitación 989 de la excitación para obtener una señal de excitación postprocesada 989a. El ramal ACELP 980 comprende una decodificación 990 de los parámetros de dominio de predicción lineal 984, para obtener parámetros de dominio de predicción lineal decodificados 990a. La señal de excitación postprocesada 991a se filtra, y el filtrado de síntesis 991 se realiza, según los parámetros de dominio de predicción lineal 990a para obtener una señal ACELP sintetizada 991a. La señal ACELP sintetizada 991a se procesa a continuación con un postprocesamiento 992 para obtener la representación del dominio temporal 986 de una subtrama de audio codificada en la carga ACELP.

#### 7.4 Combinación

**[0122]** Finalmente, se realiza una combinación 996 para obtener la representación del dominio temporal 918 de una trama de audio codificada en el modo de dominio frecuencial, la representación del dominio temporal 938 de una trama de audio codificada en el modo TCX-LPD, y la representación del dominio temporal 986 de una trama de audio codificada en el modo ACELP, para obtener una representación del dominio temporal 998 del contenido de audio.

**[0123]** Se describirán detalles adicionales a continuación.

#### 8. Detalles del codificador y decodificador

##### 8.1 Filtro LPC

##### 8.1.1 Descripción de la herramienta

**[0124]** A continuación, se describirán detalles relacionados con la codificación y decodificación utilizando coeficientes de filtro de codificación de predicción lineal.

**[0125]** En el modo ACELP, los parámetros transmitidos incluyen filtros LPC 984, índices de libros de códigos adaptativos y fijos 982, ganancias de libros de códigos adaptativos y fijos 982.

**[0126]** En el modo TCX, los parámetros transmitidos incluyen filtros LPC 934, parámetros de energía e índices de cuantificación 932 de los coeficientes de MDCT. Esta sección describe la decodificación de los filtros LPC, por ejemplo de los coeficientes del filtro LPC  $a_1$  a  $a_{16}$ , 950a, 990a.

##### 8.1.2 Definiciones

**[0127]** A continuación, se proporcionarán algunas definiciones.

**[0128]** El parámetro "nb\_lpc" describe un número total de conjuntos de parámetros LPC que están decodificados en la corriente de bits.

**[0129]** El parámetro de la corriente de bits "mode\_lpc" describe un modo de codificación del conjunto de parámetros LPC posterior.

**[0130]** El parámetro de la corriente de bits "lpc[k][x]" describe un número de parámetros LPC x del conjunto k.

**[0131]** El parámetro de la corriente de bits "qn k" describe un código binario asociado con los números  $n_k$  del libro de códigos correspondiente.

##### 8.1.3 Número de filtros LPC

**[0132]** El número real de filtros LPC "nb\_lpc" que están codificados dentro de la corriente de bits depende de la combinación del modo ACELP/TCX de la supertrama, donde la supertrama puede ser idéntica a una trama comprendiendo una pluralidad de subtramas. La combinación del modo ACELP/TCX se extrae del campo "lpc\_mode" que a su vez determina los modos de codificación "mod[k]" para  $k=0$  to 3, para cada una de las 4 tramas (también designadas como subtramas) que forman la supertrama. El valor del modo es 0 para ACELP, 1 para TCX corto (256 muestras), 2 TCX de tamaño medio (512 muestras), 3 para TCX largo (1024 muestras). Debería observarse aquí que el parámetro de la corriente de bits "lpc\_mode" que puede considerarse como un "modo" de campo de bits define los modos de codificación para cada una de las cuatro tramas dentro de una supertrama de la corriente del canal del dominio de predicción lineal (que corresponde a una trama de audio del modo de dominio frecuencial tal como, por ejemplo, una trama de codificación de audio avanzado o una trama AAC). Los modos de codificación se almacenan en una matriz "mod[]" y toma valores de 0 a 3. El mapeado desde el parámetro de la corriente de bits "LPD\_mode" a la matriz "mod[]" se puede determinar a partir de la tabla 7.

**[0133]** En relación con la matriz "mod[0...3]" se puede decir que la matriz "mod[]" indica los modos de codificación respectivos en cada trama. Para obtener detalles, se hace referencia en la tabla 8, que describe los modos de codificación indicados por la matriz "mod[]".

**[0134]** Además de los 1 a 4 filtros LPC (Codificación por predicción lineal, por sus siglas en inglés) de la supertrama, se transmite un filtro LPC opcional, LPC0 para la primera supertrama de cada segmento codificado utilizando el códec de núcleo LPD. Esto se le indica al procedimiento de decodificación LPC por medio de un indicador "first\_lpd\_flag" configurado en 1.

**[0135]** El orden en el cual los filtros LPC normalmente se encuentran en la corriente de bits es el siguiente: LPC4, y LPC0, LPC2, LPC1 y LPC3 opcionales. La condición para la presencia de un filtro LPC dado dentro de la corriente de bits se resume en la Tabla 1.

- 5 **[0136]** La corriente de bits se analiza para extraer los índices de cuantificación correspondientes a cada uno de los filtros LPC requeridos por la combinación de los modos ACELP/TCX. A continuación se describen las operaciones necesarias para decodificar uno de los filtros LPC.

#### 8.1.4 Principio general del cuantificador inverso

10

**[0137]** La cuantificación inversa de un filtro LPC, que puede llevarse a cabo en la decodificación 950 o la decodificación 990, se realiza según se describe en la Fig. 13. Los filtros LPC se cuantifican utilizando la representación de la Frecuencia espectral lineal (LSF, por sus siglas en inglés). Se calcula en primer lugar una aproximación de primera etapa según se describe en la sección 8.1.6. A continuación, se calcula un refinamiento del vector de cuantificación algebraica (AVQ, por sus siglas en inglés) 1330 según se describe en la sección 8.1.7. El vector LSF cuantificado se reconstruye mediante la suma 1350 de la aproximación de primera etapa y la contribución del AVQ ponderado inverso 1342. La presencia de un refinamiento del AVQ depende del modo de cuantificación real del filtro LPC, como se explica en la sección 8.1.5. El vector LSF de cuantificación inversa se convierte posteriormente en un vector de parámetros LSP (Pares espectrales lineales, por sus siglas en inglés), a continuación se interpola y se vuelve a convertir en parámetros LPC.

20

#### 8.1.5 Decodificación del modo de cuantificación con LPC

- 25 **[0138]** A continuación se describirá la decodificación del modo de cuantificación con LPC, que puede ser parte de la decodificación 950 de o de la decodificación 990.

**[0139]** LPC4 siempre se cuantifica utilizando una estrategia de cuantificación absoluta. Los otros filtros LPC se pueden cuantificar utilizando una estrategia de cuantificación absoluta, o uno de varios enfoques de cuantificación relativa. Para estos filtros LPC, la primera información extraída de la corriente de bits es el modo de cuantificación. Esta información se denota como "**mode\_lpc**" y se señala en la corriente de bits utilizando un código binario de longitud variable, según se indica en la última columna de la Tabla 2.

30

#### 8.1.6 Aproximación de primera etapa

- 35 **[0140]** Para cada filtro LPC, el modo cuantificación determina cómo se calcula la aproximación de primera etapa de la Fig. 13.

**[0141]** Para el modo de cuantificación absoluta (**mode\_lpc**=0), se extrae un índice de 8 bits que corresponde a la aproximación estocástica de primera etapa cuantificada por VQ de la corriente de bits. La aproximación de primera etapa 1320 se calcula a continuación por una simple tabla de consulta.

40

**[0142]** Para los modos de cuantificación relativa, se calcula la aproximación de primera etapa utilizando filtros LPC ya cuantificados en forma inversa, según se indica en la segunda columna de la Tabla 2. Por ejemplo, para LPC0 hay sólo un modo de cuantificación relativa para el cual el filtro LPC4 cuantificado en forma inversa constituye la aproximación de primera etapa. Para LPC1, hay dos modos posibles de cuantificación relativa, uno donde el LPC2 cuantificado en forma inversa constituye la aproximación de primera etapa y el otro para el cual el promedio entre los filtros LPC0 y LPC2 constituye la aproximación de primera etapa. Al igual que con todas las otras operaciones relacionadas con la cuantificación con LPC, el cálculo de la aproximación de primera etapa se realiza en el dominio de la Frecuencia espectral lineal (LSF).

50

#### 8.1.7 Refinamiento del AVQ

##### 8.1.7.1 General

- 55 **[0143]** La siguiente información extraída de la corriente de bits está relacionada con el refinamiento del AVQ necesario para construir el vector LSF cuantificado en forma inversa. La única excepción es con LPC1: la corriente de bits no contiene refinamiento del AVQ cuando este filtro se codifica relativamente a (LPC0+LPC2)/2.

**[0144]** El AVQ se basa en el cuantificador de vectores de la red RE8 de 8 dimensiones utilizado para cuantificar el espectro en modos TCX en AMR-WB+. La decodificación de los filtros LPC incluye la decodificación de los dos subvectores de 8 dimensiones  $\hat{B}_k$ ,  $k=1$  y  $2$ , del vector LSF residual ponderado.

60

**[0145]** La información del AVQ para estos dos subvectores se extrae de la corriente de bits. Comprende dos números codificados del libro de códigos "**qn1**" y "**qn2**", y los correspondientes índices de AVQ. Estos parámetros se

decodifican de la siguiente manera.

### 8.1.7.2 Decodificación de números del libro de códigos

5 **[0146]** Los primeros parámetros extraídos de la corriente de bits para decodificar el refinamiento del AVQ son los dos números del libro de códigos  $n_k$ ,  $k=1$  y  $2$ , para cada uno de los dos subvectores mencionados anteriormente. La forma en que se codifican los números del libro de códigos depende del filtro LPC (LPC0 a LPC4) y de su modo de cuantificación (absoluta o relativa). Como se muestra en la Tabla 3, hay cuatro formas diferentes de codificar  $n_k$ . Los detalles en los códigos utilizados para  $n_k$  se proporcionan a continuación:

10 modos 0 y 3 de  $n_k$ :

**[0147]** El número  $n_k$  del libro de códigos se codifica como un código **qnk** de longitud variable, de la siguiente manera:

15  $Q_2 \rightarrow$  el código para  $n_k$  es 00  
 $Q_3 \rightarrow$  el código para  $n_k$  es 01  
 $Q_4 \rightarrow$  el código para  $n_k$  es 10  
 Otros: el código para  $n_k$  es 11 seguido por:  
 20  $Q_5 \rightarrow 0$   
 $Q_6 \rightarrow 10$   
 $Q_0 \rightarrow 110$   
 $Q_7 \rightarrow 1110$   
 $Q_8 \rightarrow 11110$   
 25 etc.

modo 1 de  $n_k$ :

**[0148]** El número  $n_k$  del libro de códigos se codifica como un código **qnk** unario de la siguiente manera:

30  $Q_0 \rightarrow$  el código unario para  $n_k$  es 0  
 $Q_2 \rightarrow$  el código unario para  $n_k$  es 10  
 $Q_3 \rightarrow$  el código unario para  $n_k$  es 110  
 $Q_4 \rightarrow$  el código unario para  $n_k$  es 1110  
 35 etc.

modo 2 de  $n_k$ :

**[0149]** El número  $n_k$  del libro de códigos se codifica como un código **qnk** de longitud variable, de la siguiente manera:

40  $Q_2 \rightarrow$  el código para  $n_k$  es 00  
 $Q_3 \rightarrow$  el código para  $n_k$  es 01  
 $Q_4 \rightarrow$  el código para  $n_k$  es 10  
 45 Otros: el código para  $n_k$  es 11 seguido por:  
 $Q_0 \rightarrow 0$   
 $Q_5 \rightarrow 10$   
 $Q_6 \rightarrow 110$   
 50 etc.

### 8.1.7.3 Decodificación de índices AVQ

**[0150]** La decodificación de los filtros LPC implica decodificar los parámetros de VQ algebraica que describen

55 cada subvector  $\hat{B}_k$  cuantificado de los vectores LSF residuales ponderados. Recuerde que cada bloque  $B_k$  tiene una dimensión 8. Para cada bloque  $\hat{B}_k$  tres conjuntos de índices binarios son recibidos por el decodificador:

a) el número  $n_k$  del libro de códigos, transmitido utilizando un código de entropía “**qnk**” según se describió anteriormente:

60 b) la categoría  $l_k$  de un punto de la red seleccionado  $z$  en un denominado libro de códigos de base, que indica qué permutación se debe aplicar a un líder específico para obtener un punto entramado  $z$ ,

c) y, si el bloque  $\hat{B}_k$  cuantificado (un punto de la red) no está en el libro de códigos, los 8 índices del vector  $\mathbf{k}$  de índice de extensión Voronoi; de los índices de extensión Voronoi, se puede calcular un vector  $\mathbf{v}$  de extensión. El número de bits en cada componente del vector  $\mathbf{k}$  del índice es proporcionado por el orden de extensión  $r$ , que se puede obtener del valor del código  $n_k$  del índice. El factor de escala  $M$  de la extensión Voronoi es proporcionado

5 por  $M = 2^r$ .

[0151] A continuación, del factor de escala  $M$ , el vector  $\mathbf{v}$  de extensión Voronoi (un punto de la red en  $PE_8$ ) y el punto entramado  $\mathbf{z}$  en el libro de códigos de base (también un punto de la red en  $RE_8$ ), cada bloque  $\hat{B}_k$  en escala cuantificado se puede calcular de la siguiente manera:

10

$$\hat{B}_k = M \mathbf{z} + \mathbf{v}$$

[0152] Cuando no hay una extensión Voronoi (es decir,  $n_k < 5$ ,  $M=1$  y  $\mathbf{z}=0$ ), el libro de códigos de base es el libro de códigos  $Q_0$ ,  $Q_2$ ,  $Q_3$  o  $Q_4$  de M. Xie y J.-P. Adoul, "Embedded algebraic vector quantization (EAVQ) with application to wideband audio coding", "Conferencia internacional de IEEE sobre acústica, habla y procesamiento de señal (ICASSP), Atlanta, GA, EE. UU., vol. 1, pág. 240-243, 1996. En ese caso no se requieren bits para transmitir el vector

15

$\mathbf{k}$ . De lo contrario, cuando se utiliza la extensión Voronoi porque  $\hat{B}_k$  es lo suficientemente grande, sólo se utiliza  $Q_3$  o  $Q_4$  de la referencia antes mencionada como libro de códigos de base. La selección de  $Q_3$  o  $Q_4$  está implícita en el valor  $n_k$  del número del libro de códigos.

20

#### 8.1.7.4 Cálculo de los pesos de LSF

[0153] En el codificador, los pesos aplicados a los componentes del vector LSF residual antes de la cuantificación de AVQ son los siguientes:

25

$$w(i) = \frac{1}{W} * \frac{400}{\sqrt{d_i \cdot d_{i+1}}}, \quad i = 0..15$$

con:

$$\begin{aligned} d_0 &= LSF1ra[0] \\ d_{16} &= SF/2 - LSF1ra[15] \\ d_i &= LSF1ra[i] - LSF1ra[i-1], \quad i = 1..15 \end{aligned}$$

30

donde  $LSF1ra$  es la aproximación LSF de primera etapa y  $W$  es un factor de escala que depende del modo de cuantificación (Tabla 4).

[0154] El peso inverso correspondiente 1340 se aplica al decodificador para recuperar el vector LSF residual cuantificado.

35

#### 8.1.7.5 Reconstrucción del vector LSF cuantificado en forma inversa

[0155] El vector LSF cuantificado en forma inversa se obtiene en primer lugar mediante la concatenación de dos subvectores de refinamiento de AVQ  $\hat{B}_1$  y  $\hat{B}_2$  decodificados según se explica en las secciones 8.1.7.2 y 8.1.7.3 para formar un único vector LSF residual ponderado, a continuación, la aplicación a este vector LSF residual de peso específico del valor inverso de los pesos calculados según se explica en la sección 8.1.7.4 para formar el vector LSF residual y, a continuación, de nuevo agregar este vector LSF residual a la aproximación de primera etapa calculada como en la sección 8.1.6.

45

#### 8.1.8 Reorganización de las LSF cuantificadas

[0156] Las LSF cuantizadas de forma inversa se reorganizan y se introduce una distancia mínima entre las LSF adyacentes de 50 Hz antes de ser utilizadas.

50

#### 8.1.9 Conversión en parámetros LSP



**[0157]** El procedimiento de cuantificación inversa que se describió hasta el momento da como resultado el conjunto de parámetros LPC en el dominio LSF. Las LSF se convierten entonces al dominio coseno (LSP) utilizando la relación  $q_i = \cos(\omega_i)$ ,  $i=1, \dots, 16$  donde  $\omega_i$  representa a las frecuencias espectrales lineales (LSF).

#### 8.1.10 Interpolación de parámetros LSP

**[0158]** Para cada trama ACELP (o subtrama), aunque se transmite sólo un filtro LPC correspondiente al extremo de la trama, se utiliza la interpolación lineal para obtener un filtro diferente en cada subtrama (o parte de una subtrama) (4 filtros por trama o subtrama ACELP). La interpolación se realiza entre el filtro LPC que corresponde al extremo de la trama (o subtrama) anterior y el filtro LPC que corresponde al extremo de la trama ACELP (actual). Si  $LSP^{(nuevo)}$  es el nuevo vector LSP disponible y  $LSP^{(antiguo)}$  es el vector LSP previamente disponible. Los vectores LSP interpolados para las subtramas  $N_{sfr}=4$  se proporcionan mediante:

$$LSP_i = (0,875 - \frac{i}{N_{sfr}})LSP^{(vieja)} + (0,125 + \frac{i}{N_{sfr}})LSP^{(nuevo)} \quad \text{para } i = 0, \dots, N_{sfr} - 1$$

**[0159]** Los vectores LSP interpolados se utilizan para calcular un filtro LP diferente en cada subtrama utilizando el procedimiento de conversión de LSP a LP que se describe a continuación.

#### 8.1.11 Conversión de LSP a LP

**[0160]** Para cada subtrama, los coeficientes LSP interpolados se convierten en coeficientes de filtro LP  $a_k$ , 950a, 990a que se utilizan para sintetizar la señal reconstruida en la subtrama. Por definición, los LSP de un filtro LP de orden 16 son las raíces de los dos polinomios

$$F_1'(z) = A(z) + z^{-17} A(z^{-1})$$

y

$$F_2'(z) = A(z) - z^{-17} A(z^{-1})$$

que se puede expresar como

$$F_1'(z) = (1 + z^{-1})F_1(z)$$

y

$$F_2'(z) = (1 - z^{-1})F_2(z)$$

con

$$F_1(z) = \prod_{i=1,3,\dots,15} (1 - 2q_i z^{-1} + z^{-2})$$

y

$$F_2(z) = \prod_{i=2,4,\dots,16} (1 - 2q_i z^{-1} + z^{-2})$$

donde  $q_i$ ,  $i = 1, \dots, 16$  son las LSF en el dominio coseno, también denominados LSP. La conversión al dominio LP se realiza de la siguiente manera. Los coeficientes de  $F_1(z)$  y  $F_2(z)$  se encuentran mediante la expansión de las ecuaciones anteriores conociendo los LSP cuantificados e interpolados. Se utiliza la siguiente relación recursiva para calcular  $F_i(z)$ :

para  $i=1$  a 8

$$f_1(i) = -2q_{2i-1}f_1(i-1) + 2f_1(i-2)$$

para  $j=i-1$  hasta 1

$$f_1(j) = f_1(j) - 2q_{2j-1}f_1(j-1) + f_1(j-2)$$

final

final

con valores iniciales  $f_1(0) = 1$  y  $f_1(-1) = 0$ . Los coeficientes de  $F_2(z)$  se calculan de manera similar mediante el reemplazo de  $q_{2i-1}$  por  $q_{2i}$ .

5

[0161] Una vez que se han encontrado los coeficientes de  $F_1(z)$  y  $F_2(z)$ ,  $F_1(z)$  y  $F_2(z)$  se multiplican por  $1+z^{-1}$  y  $1-z^{-1}$ , respectivamente, para obtener  $F'_1(z)$  y  $F'_2(z)$ ; es decir

$$f'_1(i) = f_1(i) + f_1(i-1),$$

$$i = 1, \dots, 8$$

10

$$f'_2(i) = f_2(i) - f_2(i-1),$$

$$i = 1, \dots, 8$$

[0162] Finalmente, los coeficientes LP se calculan de  $f'_1(i)$  y  $f'_2(i)$  por

15

$$a_i = \begin{cases} 0,5f'_1(i) + 0,5f'_2(i), & i = 1, \dots, 8 \\ 0,5f'_1(17-i) - 0,5f'_2(17-i), & i = 9, \dots, 16 \end{cases}$$

[0163] Esto deriva directamente de la ecuación  $A(z) = (F'_1(z) + F'_2(z)) / 2$ , y considerando el hecho de que  $F'_1(z)$  y  $F'_2(z)$  son polinomios simétricos y asimétricos, respectivamente.

20

## 8.2. ACELP

[0164] A continuación, se explicarán algunos detalles acerca del procesamiento realizado por el ramal 980 de ACELP del decodificador de señales de audio 900 para facilitar la comprensión de los mecanismos de cancelación de solapamiento, lo que se describirá posteriormente.

25

### 8.2.1 Definiciones

[0165] A continuación, se proporcionarán algunas definiciones.

30

[0166] El elemento "mean\_energy" de la corriente de bits describe la energía de excitación cuantificada promedio por trama. El elemento "acb\_index[sfr]" de la corriente de bits indica el índice adaptativo del libro de códigos para cada subtrama.

35

[0167] El elemento "ltp\_filtering\_flag[sfr]" de la corriente de bits es un indicador de filtro de excitación del libro de códigos adaptativo. El elemento "lcb\_index[sfr]" de la corriente de bits indica el índice del libro de códigos de innovación para cada subtrama. El elemento "gains[sfr]" de la corriente de bits describe las ganancias cuantificadas del libro de códigos adaptativo y la contribución del libro de códigos de innovación a la excitación.

40

[0168] Además, para obtener detalles acerca de la codificación del elemento "mean\_energy" de la corriente de bits, se hace referencia al mismo en la tabla 5.

### 8.2.2 Configuración del tampón de excitación de ACELP utilizando la anterior síntesis FD y LPC0

**[0169]** A continuación se describirá una inicialización opcional del tampón de excitación de ACELP, lo que se puede realizar mediante un bloque 990b.

**[0170]** En caso de una transición de FD a ACELP, el anterior tampón de excitación  $u(n)$  y el tampón que contiene la síntesis anterior previamente enfatizada  $\hat{s}(n)$  se actualizan utilizando la anterior síntesis FD (incluyendo FAC) y LPC0 (es decir, los coeficientes del filtro LPC del conjunto de coeficientes del filtro LPC0) antes de la decodificación de la excitación ACELP. Para ello, la síntesis FD se enfatiza previamente mediante la aplicación del filtro de énfasis previo  $(1-0,68z^{-1})$  y el resultado se copia a  $\hat{s}(n)$ . La síntesis previamente enfatizada resultante se filtra entonces mediante el filtro de análisis  $A(z)$  utilizando LPC0 para obtener la señal de excitación  $u(n)$ .

### 8.2.3 Decodificación de excitación CELP

**[0171]** Si el modo de una trama es el modo CELP, la excitación está formada por los vectores del libro de códigos adaptativo en escala y del libro de códigos fijo. En cada subtrama, la excitación se construye mediante la repetición de las siguientes etapas:

La información requerida para decodificar la información de CELP se puede considerar como la excitación ACELP 982 codificada. También se debe observar que la decodificación de la excitación CELP se puede llevar a cabo mediante los bloques 988, 989 del ramal ACELP 980.

#### 8.2.3.1 Decodificación de la excitación del libro de códigos adaptativo, dependiendo del elemento "acb\_index[]" de la corriente de bits

**[0172]** El índice de tono recibido (índice del libro de códigos adaptativo) se utiliza para encontrar la parte entera y fraccionaria del retardo del tono.

**[0173]** El vector de excitación del libro de códigos adaptativo inicial  $v'(n)$  se encuentra mediante la interpolación de la excitación anterior  $u(n)$  y el retraso de tono y la fase (fracción) utilizando un filtro de interpolación FIR (Respuesta finita al impulso, por sus siglas en inglés).

**[0174]** La excitación del libro de códigos adaptativo se calcula para el tamaño de la subtrama de 64 muestras. El índice de filtro adaptativo recibido (**ltp\_filtering\_flag[]**) se utiliza a continuación para decidir si el libro de códigos adaptativo filtrado es  $v(n) = v'(n)$  o  $v(n) = 0,18v'(n) + 0,64v'(n-1) + 0,18v'(n-2)$ .

#### 8.2.3.2 Decodificación de la excitación del libro de códigos de innovación utilizando el elemento "icb\_index[]" de la corriente de bits

**[0175]** El índice del libro de códigos algebraico recibido se utiliza para extraer las posiciones y amplitudes (signos) de los pulsos de excitación y para encontrar el vector de código algebraico  $c(n)$ . Es decir

$$c(n) = \sum_{i=0}^{M-1} s_i \delta(n - m_i)$$

donde  $m_i$  y  $s_i$  son las posiciones del pulso y los signos y  $M$  es el número de pulsos.

**[0176]** Una vez que el vector de código  $c(n)$  algebraico se decodifica, se realiza un procedimiento de nitidez de tono. Primero, el  $c(n)$  se filtra mediante un filtro de énfasis previo definido de la siguiente manera:

$$F_{enf}(z) = 1 - 0,3 z^{-1}$$

**[0177]** El filtro de énfasis previo tiene el rol de reducir la energía de la excitación a bajas frecuencias. A continuación, se realiza una mejora de periodicidad por medio de un prefiltro adaptativo con una función de transferencia definida como:

$$F_p(z) = \begin{cases} 1 & \text{si } n < \min(T, 64) \\ (1 + 0,85 \cdot z^{-T}) & \text{si } T < 64 \text{ y } T \leq n < \min(2T, 64) \\ 1/(1 - 0,85 \cdot z^{-T}) & \text{si } 2T < 64 \text{ y } 2T \leq n < 64 \end{cases}$$

donde  $n$  es el índice de la subtrama ( $n=0, \dots, 63$ ), y donde  $T$  es una versión redondeada de la parte entera  $T_0$  y la parte fraccionaria  $T_{0, \text{frac}}$  del retardo del tono y se proporciona mediante:

$$T = \begin{cases} T_0 + 1 & \text{si } T_{0, \text{frac}} > 2 \\ T_0 & \text{de otro modo} \end{cases}$$

**[0178]** El prefiltro adaptativo  $F_p(z)$  da color al espectro mediante la reducción de frecuencias interarmónicas, que son molestas para el oído humano en caso de señales de voz.

**8.2.3.3 Decodificación de ganancias del libro de códigos adaptativo e innovador, descrita por el elemento "gains[]"** de la corriente de bits

**[0179]** El índice de 7 bits recibido por subtrama proporciona directamente la ganancia del libro de códigos  $\hat{g}_p$  y el factor de corrección  $\hat{\gamma}$  de la ganancia del libro de códigos fijo. La ganancia del libro de códigos fijo se calcula entonces mediante la multiplicación del factor de corrección de la ganancia por una ganancia estimada del libro de códigos fijo. La ganancia estimada del libro de códigos fijo  $g'_c$  se encuentra de la siguiente manera. Primero, se encuentra la energía de innovación promedio por:

$$E_i = 10 \log \left( \frac{1}{N} \sum_{i=0}^{N-1} c^2(i) \right)$$

**[0180]** A continuación se encuentra la ganancia  $G'_c$  estimada en dB por:

$$G'_c = \bar{E} - E_i$$

donde  $\bar{E}$  es la energía de excitación promedio decodificada por trama. La energía de excitación innovadora promedio en una trama,  $\bar{E}$ , se codifica con 2 bits por trama (18, 30, 42 ó 54 dB) como "mean\_energy".

**[0181]** La ganancia de predicción en el dominio lineal se proporciona por

$$g'_c = 10^{0,05 G'_c} = 10^{0,05 (\bar{E} - E_i)}$$

**[0182]** La ganancia del libro de códigos fijo cuantificada se proporciona por

$$\hat{g}_c = \hat{\gamma} \cdot g'_c$$

**8.2.3.4 Cálculo de la excitación reconstruida**

**[0183]** Las siguientes etapas son para  $n = 0, \dots, 63$ . La excitación total se construye por:

$$u'(n) = \hat{g}_p v(n) + \hat{g}_c c(n)$$

donde  $c(n)$  es el vector de código del libro de códigos fijo después de filtrarlo a través del prefiltro adaptativo  $F(z)$ . La señal de excitación  $u'(n)$  se utiliza para actualizar el contenido del libro de códigos adaptativo. La señal de excitación  $u'(n)$  se postprocesa a continuación según se describe en la siguiente sección para obtener la señal de excitación postprocesada  $u(n)$  que se utiliza en la entrada del filtro de síntesis  $1/\hat{A}(z)$ .

### 8.3 Postprocesamiento de la excitación

#### 8.3.1 General

**[0184]** A continuación, se describirá el postprocesamiento de señales de excitación, que se puede realizar en el bloque 989. En otras palabras, para la síntesis de señales se puede realizar un postprocesamiento de elementos de excitación de la siguiente manera.

#### 8.3.2 Suavizamiento de ganancia para aumento de sonido

**[0185]** Se aplica una técnica de suavizamiento de ganancia no lineal a la ganancia  $\hat{g}_c$  del libro de códigos fijo para aumentar la excitación en el sonido. En base a la estabilidad y la sonorización del segmento de voz, la ganancia del vector del libro de códigos fijo se suaviza para poder reducir la fluctuación en la energía de la excitación en el caso de señales estacionarias. Esto mejora el rendimiento en el caso de un sonido de fondo estacionario. El factor de sonorización se proporciona mediante

$$\lambda = 0.5(1 - r_v)$$

con

$$r_v = (E_v - E_c) / (E_v + E_c)$$

donde  $E_v$  y  $E_c$  son las energías del vector de código del tono en escala y el código vector de innovación en escala, respectivamente ( $r_v$  proporciona una medida de la periodicidad de la señal). Observe que dado que el valor de  $r_v$  está entre  $-1$  y  $1$ , el valor de  $\lambda$  está entre  $0$  y  $1$ . Observe que el factor  $\lambda$  está relacionado con la cantidad de no sonorización con un valor de  $0$  de segmentos puramente sonorizados y un valor de  $1$  para segmentos puramente no sonorizados.

**[0186]** Se calcula un factor de estabilidad  $\theta$  en base a una medida de distancia entre los filtros LP adyacentes. Aquí, el factor  $\theta$  está relacionado con la medida de la distancia del ISF (filtro infrasónico, por sus siglas en inglés). La distancia del ISF se proporciona mediante

$$ISF_{dist} = \sum_{i=0}^{14} (f_i - f_i^{(p)})^2$$

donde  $f_i$  representa los ISF de la trama actual, y  $f_i^{(p)}$  representa los ISF de la trama anterior. El factor de estabilidad  $\theta$  se proporciona por

$$\theta = 1.25 - ISF_{dist} / 400000$$

Limitado por  $0 \leq \theta \leq 1$

**[0187]** La medida de la distancia del ISF es menor en el caso de las señales estables. Dado que el valor de  $\theta$  está inversamente relacionado con la medida de distancia del ISF, los valores mayores a  $\theta$  corresponden a señales más estables. El factor  $S_m$  de suavizamiento de ganancia se proporciona por

$$S_m = \lambda \theta$$

**[0188]** El valor de  $S_m$  se acerca a  $1$  en el caso de las señales no sonorizadas y estables, que es el caso de las

señales de sonido de fondo estacionario. En el caso de las señales puramente sonorizadas, o de señales inestables, el valor de  $S_m$  se acerca a 0. Se calcula una ganancia inicial  $g_0$  modificada mediante la comparación de la ganancia  $\hat{g}_c$  del libro de códigos fijo con un valor umbral aportado por la ganancia inicial modificada de la subtrama anterior,  $g_{-1}$ . Si  $\hat{g}_c$  es mayor o igual a  $g_{-1}$ , entonces  $g_0$  se calcula mediante el decremento de  $\hat{g}_c$  por 1,5 dB unido por  $g_0 \geq g_{-1}$ . Si  $\hat{g}_c$  es menor que  $g_{-1}$ , entonces  $g_0$  se calcula mediante el incremento de  $\hat{g}_c$  por 1,5 dB limitado por  $g_0 \leq g_{-1}$ .

[0189] Finalmente, la ganancia se actualiza con el valor de la ganancia suavizada de la siguiente manera

$$\hat{g}_{sc} = S_m g_0 + (1 - S_m) \hat{g}_c$$

10

### 8.3.3 Potenciador de tono

[0190] Un esquema potenciador de tono modifica la excitación total  $u'(n)$  mediante el filtrado de la excitación del libro de códigos fijo a través de un filtro de innovación cuya respuesta a la frecuencia enfatiza las frecuencias más altas y reduce la energía de la porción de baja frecuencia del vector de código innovador, y cuyos coeficientes están relacionados con la periodicidad en la señal. Un filtro de la forma

$$F_{in}(z) = -c_{pe} z + 1 - c_{pe} z^{-1}$$

20 se utiliza donde  $c_{pe} = 0,125(1 + r_v)$ , donde  $r_v$  es el factor de periodicidad dado por  $r_v = (E_v - E_c)/(E_v + E_c)$ , según se describió anteriormente. El vector de código del libro de códigos fijo filtrado es proporcionado por

$$c'(n) = c(n) - c_{pe} (c(n+1) + c(n-1))$$

25 y la excitación postprocesada actualizada se proporciona por

$$u(n) = \hat{g}_p v(n) + \hat{g}_x c'(n)$$

[0191] El procedimiento anterior se puede llevar a cabo en una etapa mediante la actualización de la excitación 30 989a,  $u(n)$  de la siguiente manera

$$u(n) = \hat{g}_p v(n) + \hat{g}_x c(n) - \hat{g}_x c_{pe} (c(n+1) + c(n-1))$$

### 8.4 Síntesis y postprocesamiento

35

[0192] A continuación se describirá el filtrado de síntesis 991 y el postprocesamiento 992.

#### 8.4.1 General

40 [0193] La síntesis LP se realiza mediante el filtrado de la señal de excitación postprocesada 989a  $u(n)$  a través del filtro de síntesis LP  $1/\hat{A}(z)$ . El filtro LP interpolado por subtrama se utiliza en la síntesis LP mediante el filtrado de la señal reconstruida en una subtrama proporcionada por

$$\hat{s}(n) = u(n) - \sum_{i=1}^{16} \hat{a}_i \hat{s}(n-i),$$

$$n = 0, \dots, 63$$

45

[0194] La señal sintetizada se desenfatiza entonces mediante filtrado a través del filtro  $1/(1-0,68z^{-1})$  (la inversa del filtro de preénfasis aplicada en la entrada del codificador).

#### 50 8.4.2 Postprocesamiento de la señal de síntesis

[0195] Después de la síntesis LP, la señal reconstruida se postprocesa utilizando una potenciación de tono de

baja frecuencia. Se utiliza una descomposición de dos bandas y se aplica el filtrado adaptativo sólo en la banda más baja. Esto da como resultado un postprocesamiento total, que principalmente se dirige a frecuencias cercanas a la primera armonía de la señal de voz sintetizada.

- 5 **[0196]** La señal se procesa en dos ramales. En el ramal más alto, la señal decodificada se filtra mediante un filtro de paso alto para producir la señal de banda más alta  $s_H$ . En el ramal más bajo, la señal decodificada se procesa en primer lugar a través de un potenciador de tono adaptativo y, a continuación, se filtra a través de un filtro de paso bajo para obtener la señal  $s_{LEF}$  postprocesada de banda más baja. La señal decodificada postprocesada se obtiene mediante la adición de la señal postprocesada de banda más baja y la señal de banda más alta. El objetivo del
- 10 potenciador de tono es reducir el sonido interarmónico en la señal decodificada, lo cual se logra en este caso con un filtro lineal de variación de tiempo con una función de transferencia

$$H_E(z) = (1 - \alpha) + \frac{\alpha}{2} z^T + \frac{\alpha}{2} z^{-T}$$

- 15 y se describe mediante la siguiente ecuación:

$$s_{LE}(n) = (1 - \alpha)\hat{s}(n) + \frac{\alpha}{2}\hat{s}(n - T) + \frac{\alpha}{2}\hat{s}(n + T)$$

- donde  $\alpha$  es un coeficiente que controla la atenuación interarmónica,  $T$  es el período de tono de la señal de entrada
- 20  $\hat{s}(n)$ , y  $s_{LE}(n)$  es la señal de salida del potenciador de tono. Los parámetros  $T$  y  $\alpha$  varían con el tiempo y son proporcionados por el módulo de rastreo de tono. Con un valor de  $\alpha = 0,5$ , la ganancia del filtro es exactamente 0 en las frecuencias  $1/(2T), 3/(2T), 5/(2T)$ , etc.; es decir, en el punto medio entre las frecuencias armónicas  $1/T, 3/T, 5/T$ , etc. Cuando  $\alpha$  se acerca a 0, la atenuación entre la armonía producida por el filtro se reduce.

- 25 **[0197]** Para confinar el postprocesamiento a la región de baja frecuencia, la señal potenciada  $s_{LE}$  se filtra a través del filtro de paso bajo para producir la señal  $s_{LEF}$  que se agrega a la señal filtrada a través del filtro de paso alto  $s_H$  para obtener la señal de síntesis postprocesada,  $s_E$ .

- [0198]** Un procedimiento alternativo equivalente al que se describió anteriormente se utiliza para eliminar la
- 30 necesidad de realizar un filtrado a través del filtro de paso alto. Esto se logra mediante la representación de la señal postprocesada  $s_E(n)$  en el dominio  $Z$  como

$$S_E(z) = \hat{S}(z) - \alpha \hat{S}(z) P_{LT}(z) H_{LP}(z)$$

- 35 donde  $P_{LT}(z)$  es la función de transferencia del filtro de predicción de largo plazo proporcionada por

$$P_{LT}(z) = 1 - 0,5z^T - 0,5z^{-T}$$

- 40 y  $H_{LP}(z)$  es la función de transferencia del filtro de paso bajo.

**[0199]** De este modo, el postprocesamiento es equivalente a la sustracción de la señal de error a largo plazo filtrada a través de un filtro de paso bajo en escala de la señal de síntesis  $\hat{s}(n)$ .

- [0200]** El valor  $T$  lo proporciona el retardo del tono en bucle cerrado recibido en cada subtrama (el retardo de
- 45 tono fraccionario redondeado al número entero más cercano). Se realiza un rastreo simple para verificar el duplicado del tono. Si la correlación del tono normalizada en el retardo  $T/2$  es mayor que 0,95, se utiliza el valor  $T/2$  como el nuevo retardo del tono para el postprocesamiento.

- [0201]** El factor  $\alpha$  es proporcionado por
- 50

$$\alpha = 0,5 \hat{g}_p$$

limitado a  $0 \leq \alpha \leq 0,5$  donde  $\hat{g}_p$  es la ganancia del tono decodificado.

**[0202]** Observe que en el modo TCX y durante la codificación del dominio frecuencial, el valor de  $\alpha$  se configura en cero. Se utiliza un filtro de paso bajo de fase lineal FIR con 25 coeficientes, con una frecuencia de corte a 25 5Fs/256 kHz (el retardo del filtro es de 12 muestras).

#### 8.5 TCX basada en MDCT (Transformada de coseno discreta modificada, por sus siglas en inglés)

**[0203]** A continuación, se describirá la TCX basada en MDCT en detalle, que se realiza por la síntesis de la señal principal 940 del ramal TXC-LPD 930.

##### 8.5.1 Descripción de las herramientas

**[0204]** Cuando la variable "core\_mode" de la corriente de bits es igual a 1, lo que indica que la codificación se realiza utilizando parámetros de dominio de predicción lineal, y cuando uno o más de los tres modos TCX se selecciona como la codificación del "dominio de predicción lineal", es decir, una de las 4 entradas de matriz de mod[] es mayor que 0, se utiliza la herramienta TCX basada en MDCT. La herramienta TCX basada en MDCT recibe los coeficientes espectrales cuantificados 941a del decodificador aritmético 941. Los coeficientes cuantificados 941a (o una versión inversamente cuantificada 942a de los mismos) se completan en primer lugar mediante un sonido confortable (relleno de sonido 943). El moldeado del sonido de dominio frecuencial basado en LPC 945 se aplica a continuación a los coeficientes espectrales resultantes 943a (o una versión de desmoldeado espectral 944a de los mismos) y se realiza una transformación de MDCT inversa 946 para obtener la señal de síntesis de dominio temporal 946a.

##### 8.5.2 Definiciones

**[0205]** A continuación, se proporcionarán algunas definiciones. La variable "lg" describe un número de coeficientes espectrales cuantificados generados por el decodificador aritmético. El elemento "noise\_factor" de la corriente de bits describe un índice de cuantificación de nivel de sonido. La variable "noise\_level" describe un nivel de sonido inyectado en un espectro reconstruido. La variable "noise[]" describe un vector de sonido generado. El elemento "global\_gain" de la corriente de bits describe un índice de cuantificación de ganancias de puesta a cero. La variable "g" describe una ganancia de puesta a cero. La variable "rms" describe una media cuadrática de la señal sintetizada de dominio temporal, x[]. La variable "x[]" describe una señal sintetizada de dominio temporal.

##### 8.5.3 Proceso de decodificación

**[0206]** La TCX basada en MDCT solicita del decodificador aritmético 941 un número de coeficientes espectrales cuantificados, lg, que se determina por el valor mod[]. Este valor (lg) también define la longitud y forma de la ventana que se va a aplicar en la MDCT inversa. La ventana, que se puede aplicar durante o después de la MDCT inversa 946, está compuesta por tres partes, una superposición del lado izquierdo de muestras L, una parte media de algunas de las muestras M y una superposición derecha de muestras R. Para obtener una ventana de MDCT con una longitud de 2\*lg, se agregan ceros ZL del lado izquierdo y ceros ZR del lado derecho. En caso de una transición de o a SHORT\_WINDOW, puede ser necesario que la correspondiente región de superposición L o R se reduzca a 128 con el fin de adaptarse a la pendiente de la ventana más corta de SHORT\_WINDOW. Por consiguiente, puede que sea necesario que la región M y la correspondiente región cero ZL o ZR se expandan a 64 muestras cada una.

**[0207]** La ventana MDCT, que se puede aplicar durante la MDCT inversa 946 o después de la MDCT inversa 946, es proporcionada por

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

**[0208]** La tabla 6 muestra un número de coeficientes espectrales como función de mod[].

**[0209]** Los coeficientes espectrales cuantificados, quant[] 941a, proporcionados por el decodificador aritmético 941, o los coeficientes espectrales cuantificados de manera inversa 942a, se completan opcionalmente con un sonido confortable (relleno de sonido 943). El nivel del sonido inyectado se determina por la variable decodificada noise\_factor de la siguiente manera:

$$\text{noise\_level} = 0,0625 * (8 - \text{noise\_factor})$$



**[0210]** A continuación, se calcula un vector de sonido, noise[], utilizando una función aleatoria, random\_sign(), que proporciona de manera aleatoria el valor -1 o +1.

5 noise[i] = random\_sign()\*noise\_level;

**[0211]** Los vectores quant[] y noise[] se combinan para formar el vector de coeficientes espectrales reconstruidos, r[] 942a, de tal forma que la partida de 8 ceros consecutivos en quant[] sea reemplazada por los componentes de noise[]. Una partida de 8 no-ceros se detecta según la fórmula:

10

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

**[0212]** Se obtiene el espectro reconstruido 943a de la siguiente manera:

$$r[i] = \begin{cases} noise[i] & \text{if } rl[i] = 0 \\ quant[i] & \text{de otro modo} \end{cases}$$

15

**[0213]** El desmoldeado de un espectro 944 se aplica opcionalmente al espectro reconstruido 943a según las siguientes etapas:

- 20 1. Calcular la energía  $E_m$  del bloque de 8 dimensiones en el índice m para cada bloque de 8 dimensiones del primer cuarto del espectro.  
2. Calcular la relación  $R_m = \sqrt{E_m/E_l}$ , donde l es el índice de bloques con el valor máximo de todos los  $E_m$   
3. Si  $R_m < 0,1$ , entonces configurar  $R_m = 0,1$   
25 4. Si  $R_m < R_{m-1}$ , entonces configurar  $R_m = R_{m-1}$

25

**[0214]** Cada bloque de 8 dimensiones correspondiente al primer cuarto del espectro se multiplica entonces por el factor  $R_m$ . De manera acorde, se obtienen los coeficientes espectrales con desmoldeado espectral 944a.

**[0215]** Antes de aplicar la MDCT inversa 946, los dos filtros LPC cuantificados, LPC1, LPC2 (cada uno de los cuales puede describirse por los coeficientes de filtros del 1 al 10) que corresponden a ambas extremidades del bloque MDCT (es decir, los puntos de plegado izquierdo y derecho) se recuperan (bloque 950), se calculan sus 15 versiones ponderadas y se calculan los correspondientes espectros diezmos (64 puntos, cualquiera que sea la longitud de la transformada) 951a (bloque 951). Estos espectros LPC ponderados 951a se calculan mediante la aplicación de una ODFT (Transformada de Fourier discreta impar, por sus siglas en inglés) a los coeficientes del filtro LPC 950a. Se aplica una modulación compleja a los coeficientes LPC antes de calcular la ODFT de manera que los intervalos de frecuencia de la ODFT (utilizados en el cálculo de espectros 951) se alineen perfectamente con los intervalos de frecuencia de la MDCT (de la MDCT inversa 946). Por ejemplo, el espectro de síntesis LPC ponderado 951a de un filtro LPC dado  $\hat{A}(z)$  (definido, por ejemplo, por coeficientes de filtro de dominio temporal  $a_1$  a  $a_{16}$ ) se calcula de la siguiente manera:

40

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

con

$$x_t[n] = \begin{cases} \hat{w}[n] e^{-j \frac{\pi}{M} n} & \text{si } 0 \leq n < lpc\_order + 1 \\ 0 & \text{si } lpc\_order + 1 \leq n < M \end{cases}$$

donde  $\hat{w}[n]$ ,  $n=0 \dots lpc\_order + 1$ , son los coeficientes (de dominio temporal) del filtro LPC ponderado proporcionados por:

45

$$\hat{W}(z) = \hat{A}(z) / \gamma_f$$

con  $\gamma_f = 0,92$

**[0216]** Las ganancias  $g[k]$  952a se pueden calcular a partir de la representación espectral  $X_0[k]$ , 951a de los 5 coeficientes LPC según:

$$g[k] = \sqrt{\frac{1}{X_0[k] X_0^*[k]}} \quad \forall k \in \{0, \dots, M-1\}$$

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

10

**[0217]** En el caso que  $g_1[k]$  y  $g_2[k]$ ,  $k=0\dots 63$ , sean los espectros LPC diezmados que corresponden respectivamente a los puntos de plegado izquierdo y derecho calculados según se explicó anteriormente. La operación de FDNS inverso 945 consiste en el filtrado del espectro reconstruido  $r[i]$ , 944a utilizando el filtro recursivo:

$$r[i] = a[i] \cdot r[i] + b[i] \cdot r[i-1], \quad i=0 \dots lg,$$

15

donde  $a[i]$  y  $b[i]$ , 945b derivan de las ganancias izquierda y derecha  $g_1[k]$ ,  $g_2[k]$ , 952a, utilizando las siguientes fórmulas:

$$a[i] = 2 \cdot g_1[k] \cdot g_2[k] / (g_1[k] + g_2[k]),$$

20

$$b[i] = (g_2[k] - g_1[k]) / (g_1[k] + g_2[k]).$$

**[0218]** En la fórmula anterior, la variable  $k$  es igual a  $i/(lg/64)$  para tener en cuenta el hecho de que los espectros LPC están diezmados.

25

**[0219]** El espectro reconstruido  $rr[i]$ , 945a es alimentado en una MDCT inversa 946. La señal de salida que no está formada en ventana,  $x[i]$ , 946a, es puesta a cero por la ganancia,  $g$ , obtenida por una cuantificación inversa del índice "global\_gain" decodificado:

$$g = \frac{10^{global\_gain/28}}{2 \cdot rms},$$

30

donde  $rms$  se calcula como:

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

35

**[0220]** Entonces, la señal de dominio temporal sintetizada puesta a cero 940a es igual a:

$$x_{w,0}[i] = x[i] \cdot g$$

40

**[0221]** Después de la puesta a cero, se aplica la formación en ventana y la suma de superposición, por ejemplo, en el bloque 978.

**[0222]** La síntesis de TCX  $x(n)$  938 reconstruida a continuación se filtra opcionalmente a través del filtro de 45 preénfasis  $(1 - 0,68z^{-1})$ . La síntesis previamente enfatizada resultante entonces se filtra mediante el filtro de análisis  $\hat{A}(z)$  para obtener la señal de excitación. La excitación calculada actualiza el libro de códigos adaptativo ACELP y

permite la conmutación de TCX a ACELP en una trama posterior. La señal finalmente se reconstruye desenfatizando la síntesis preenfatzada, mediante la aplicación del filtro  $1/(1-0,68z^{-1})$ . Observe que los coeficientes del filtro de análisis están interpolados en subtramas.

- 5 **[0223]** Observe también que la longitud de la síntesis de TCX es proporcionada por la longitud de trama de TCX (sin la superposición): 256, 512 ó 1024 muestras para el mod[] de 1, 2 ó 3 respectivamente.

#### 8.6 Herramienta de cancelación de solapamiento hacia delante (FAC, por sus siglas en inglés)

##### 10 8.6.1 Descripción de la herramienta de cancelación de solapamiento hacia delante

**[0224]** A continuación se describen operaciones de cancelación de solapamiento hacia delante (FAC) que se llevan a cabo durante las transiciones entre ACELP y la codificación de la transformada (TC, por sus siglas en inglés) (por ejemplo, en el modo de dominio frecuencial o en el modo TCX-LPD) con el fin de obtener la señal final de síntesis.

- 15 El objetivo de la FAC consiste en cancelar el solapamiento de dominio temporal introducido por la TC y que no puede cancelarse por medio de la trama ACELP anterior o posterior. Aquí, la noción de TC incluye la MDCT en bloques largos y cortos (modo de dominio frecuencial) así como la TCX basada en MDCT (modo TCX-LPD).

- [0225]** La Fig. 10 representa las distintas señales intermedias que se calculan para obtener la señal de síntesis final de la trama TC. En el ejemplo ilustrado, la trama TC (por ejemplo, una trama 1020 codificada en el modo de dominio frecuencial o en el modo TCX-LPD) está precedido por y seguido por una trama ACELP (tramas 1010 y 1030). En los otros casos (una trama ACELP seguida por más de una trama TC, o más de una trama TC seguida por una trama ACELP), sólo se calculan las señales requeridas.

- 25 **[0226]** Con respecto a la Fig. 10, se proporcionará un panorama general de la cancelación de solapamiento hacia delante, donde se debe observar que la cancelación de solapamiento hacia delante será realizada por los bloques 960, 961, 962, 963, 964, 965 y 970.

- [0227]** En la representación gráfica de las operaciones de decodificación de cancelación de solapamiento hacia delante, que se muestran en la Fig. 10, las abscisas 1040a, 1040b, 1040c, 1040d describen un tiempo en términos de muestras de audio. Una ordenada 1042a describe una señal de síntesis de cancelación de solapamiento hacia delante, por ejemplo, en términos de una amplitud. Una ordenada 1042b describe señales que representan un contenido de audio codificado, por ejemplo una señal de síntesis ACELP y una señal de salida de trama de codificación de la transformada. Una ordenada 1042c describe las contribuciones ACELP a una cancelación de solapamiento, como por ejemplo, una respuesta impulso cero ACELP formada en ventana y una síntesis ACELP plegada y formada en ventana. Una ordenada 1042d describe una señal de síntesis en un dominio original.

- [0228]** Como se puede observar, se proporciona una señal de síntesis de cancelación de solapamiento hacia delante 1050 en una transición de la trama de audio 1010 codificada en el modo ACELP a la trama de audio 1020 codificada en el modo TCX-LPD. La señal de síntesis de cancelación de solapamiento hacia delante 1050 se proporciona mediante la aplicación del filtrado de síntesis 964 y una señal de estímulo de cancelación de solapamiento 963a, que se proporciona por la DCT (Transformada de coseno discreta, por sus siglas en inglés) inversa de tipo IV 963. El filtrado de síntesis 964 se basa en los coeficientes del filtro de síntesis 965a, que derivan de un conjunto LPC1 de parámetros de dominio de predicción lineal o coeficientes de filtro LPC. Como se puede observar en la Fig. 10, una primera porción 1050a de la (primera) señal de síntesis de cancelación de solapamiento hacia delante 1050 puede ser una respuesta de entrada no cero proporcionada por el filtrado de síntesis 964 para una señal de estímulo de cancelación de solapamiento no cero 963a. Sin embargo, la señal de síntesis de cancelación de solapamiento hacia delante 1050 comprende también una porción de respuesta de entrada cero 1050b, que puede ser proporcionada por el filtrado de síntesis 964 para una porción cero de la señal de estímulo de cancelación de solapamiento 963a. Por consiguiente, la señal de síntesis de cancelación de solapamiento hacia delante 1050 puede comprender una porción de respuesta de entrada no cero 1050a y una porción de respuesta de entrada cero 1050b. Se debe observar que la señal de síntesis de cancelación de solapamiento hacia delante 1050 puede proporcionarse preferiblemente sobre la base del conjunto LPC1 de parámetros de dominio de predicción lineal, que está relacionado con la transición entre la trama o subtrama 1010 y la trama o subtrama 1020. Además, otra señal de síntesis de cancelación de solapamiento hacia delante 1054 se proporciona en una transición de la trama o subtrama 1020 al trama o subtrama 1030. La señal de síntesis de cancelación de solapamiento hacia delante 1054 puede ser proporcionada por el filtrado de síntesis 964 de una señal de estímulo de cancelación de solapamiento 963a, que es proporcionada por una DCT IV inversa, 963 sobre la base de coeficientes de cancelación de solapamiento. Se debe observar que la provisión de la señal de síntesis de cancelación de solapamiento hacia delante 1054 puede basarse en un conjunto de parámetros de predicción lineal, LPC2, que estén asociados a la transición entre la trama o subtrama 1020 y la posterior trama o subtrama 1030.

- [0229]** Además, las señales de síntesis de cancelación de solapamiento adicionales 1060, 1062 se proporcionarán en una transición de una trama o subtrama ACELP 1010 a una trama o subtrama TXC-LPD 1020. Por ejemplo, una versión plegada y formada en ventana 973a, 1060 de una señal de síntesis ACELP 986, 1056 se puede

proporcionar, por ejemplo, mediante los bloques 971, 972, 973. Además, se proporcionará una respuesta ACELP formada en ventana de entrada cero 976a, 1062, por ejemplo, mediante los bloques 975, 976. Por ejemplo, la señal de síntesis ACELP plegada y formada en ventana 973a, 1060 se puede obtener mediante la formación en ventana de la señal de síntesis ACELP 986, 1056 y mediante la aplicación de un plegado temporal 973 del resultado de la formación en ventana, tal como se describirá en mayor detalle a continuación. La respuesta ACELP formada en ventana de entrada cero 976a, 1062 se puede obtener mediante la provisión de una entrada cero a un filtro de síntesis 975, lo cual es igual al filtro de síntesis 991, que se utiliza para proporcionar la señal de síntesis ACELP 986, 1056, donde un estado inicial del filtro de síntesis 975 es igual a un estado del filtro de síntesis 981 al final de la provisión de la señal de síntesis ACELP 986, 1056 de la trama o subtrama 1010. De este modo, la señal de síntesis ACELP plegada y formada en ventana 1060 puede ser equivalente a la señal de síntesis de cancelación de solapamiento hacia delante 973a, y la respuesta ACELP formada en ventana de entrada cero 1062 puede ser equivalente a la señal de síntesis de cancelación de solapamiento hacia delante 976a.

**[0230]** Finalmente, la trama de codificación de la transformada genera la señal 1050a, que puede ser igual a una versión formada en ventana de la representación de dominio temporal 940a, según se combina con las señales de síntesis de cancelación de solapamiento hacia delante 1052, 1054, y las contribuciones ACELP adicionales 1060, 1062 a la cancelación de solapamiento.

## 8.6.2 Definiciones

**[0231]** A continuación, se proporcionarán algunas definiciones. El elemento "fac\_gain" de la corriente de bits describe un índice con ganancia de 7 bits. El elemento "nq[i]" de la corriente de bits describe un número del libro de códigos. El elemento "FAC[i]" de sintaxis describe datos de cancelación de solapamiento hacia delante. La variable "fac\_length" describe una longitud de una transformada de cancelación de solapamiento hacia delante, que puede ser igual a 64 para transiciones desde y hasta una ventana del tipo "EIGHT\_SHORT\_SEQUENCES" y que puede ser 128 de otro modo. La variable "use\_gain" indica el uso de información de ganancias explícita.

## 8.6.3 Proceso de decodificación

**[0232]** A continuación, se describirá el proceso de decodificación. Con este fin, se resumirán brevemente las distintas etapas.

### 1. Decodificación de parámetros AVQ (bloque 960)

- La información de la FAC se codifica utilizando la misma herramienta de cuantificación de vector algebraica (AVQ) que la que se utiliza para la codificación de los filtros LPC (véase la sección 8.1).
- Para  $i=0 \dots \text{longitud de la transformada FAC}$ :

- o Se codifica un número del libro de códigos  $nq[i]$  utilizando un código unario modificado
- o Se codifican los datos de FAC correspondientes  $FAC[i]$  con  $4 \cdot nq[i]$  bits

- Por lo tanto, se extrae un vector  $FAC[i]$  para  $i=0, \dots, \text{fac\_length}$  de la corriente de bits

### 2. Aplicación de un factor de ganancia $g$ a los datos de FAC (bloque 961)

- Para las transiciones con TCX basada en MDCT (wLPT), se utiliza la ganancia del elemento "tcx\_coding" correspondiente
- Para otras transiciones, una información de ganancia "fac\_gain" se ha recuperado de la corriente de bits (codificada utilizando un cuantificador escalar de 7 bits). La ganancia  $g$  se calcula como  $g=10^{\text{fac\_gain}/28}$  utilizando esa información de ganancia.

3. En el caso de las transiciones entre TCX basada en MDCT y ACELP, se aplica desmoldeado espectral 962 al primer cuarto de los datos espectrales FAC 961a. Las ganancias del desmoldeado son aquellas calculadas para la correspondiente TCX basada en MDCT (para su uso por el desmoldeado espectral 944) según se explicó en la sección 8.5.3 de manera que el sonido de cuantificación de FAC y TCX basada en MDCT tengan la misma forma.

### 4. Cálculo de la DCT-IV inversa de los datos FAC en escala de ganancia (bloque 963).

- La longitud de la transformada FAC,  $\text{fac\_length}$ , es por defecto, igual a 128
- Para transiciones con bloques cortos, esta longitud se reduce a 64.

5. Aplicación (bloque 064) del filtro de síntesis ponderado  $1/\hat{W}(z)$  (descrito, por ejemplo por los coeficientes de filtro de síntesis 965a) para obtener la señal de síntesis FAC 964a. La señal resultante se representa en la línea (a) de la Fig. 10.

- El filtro de síntesis ponderado se basa en el filtro LPC que corresponde al punto de plegado (en la Fig. 10 se identifica como LPC1 para transiciones de ACELP a TCX-LPD y como LPC2 para transiciones de wLPD TC (TCX-LPD) a ACELP o LPC0 para transiciones de FD TC (codificación de la transformada del código frecuencial) a ACELP)
- Se utiliza el mismo factor de peso LPC para las operaciones ACELP:

$$\hat{W}(z) = A(z/\gamma_1),$$

donde  $\gamma_1 = 0,92$

- Para calcular la señal de síntesis FAC 964a, la memoria inicial del filtro de síntesis ponderado 964 se establece en 0
  - Para las transiciones de ACELP, la señal de síntesis FAC 1050 se extiende aún más anexando la respuesta de entrada cero (ZIR) 1050b del filtro de síntesis ponderado (128 muestras).
6. En el caso de las transiciones de ACELP, calcular la síntesis ACELP anterior formada en ventana 972a, plegarla (por ejemplo, para obtener la señal 973a o a la señal 1060) y agregarle la señal ZIR formada en ventana (por ejemplo, la señal 976a o la señal 1062). La respuesta ZIR se calcula utilizando LPC1. La ventana aplicada a las muestras de síntesis ACELP anterior fac\_length es:

$$\text{sine}[n + \text{fac\_length}] * \text{sine}[\text{fac\_length} - 1 - n],$$

$n = -\text{fac\_length} \dots -1$ ,  
y la ventana aplicada a ZIR es:

$$1 - \text{sine}[n + \text{fac\_length}]^2,$$

$n = 0 \dots \text{fac\_length} - 1$ ,  
donde  $\text{sine}[n]$  es un cuarto de un ciclo de senos:

$$\text{sine}[n] = \sin(n * \pi / (2 * \text{fac\_length})),$$

- $n = 0 \dots 2 * \text{fac\_length} - 1$ .  
La señal resultante se representa en la línea (c) en la Fig. 10 y se denota como la contribución ACELP (contribuciones de señales 1060, 1062).

- 7. Agregar la síntesis FAC 964a, 1050 (y la contribución ACELP 973a, 976a, 1060, 1062 en el caso de transiciones de ACELP) a la trama TC (que se representa como la línea (b) en la FIG. 10) (o a una versión formada en ventana de la representación de dominio temporal 940a) para obtener la señal de síntesis 998 (que se representa como la línea (d) en la Fig. 10).

#### 8.7 Proceso de codificación de la cancelación de solapamiento hacia delante (FAC)

- [0233]** A continuación se describirán algunos detalles acerca de la codificación de la información requerida para la cancelación de solapamiento hacia delante. En particular, se describirá el cálculo y la codificación de los coeficientes de cancelación de solapamiento 936.
- [0234]** La Fig. 11 muestra las etapas de procesamiento en el codificador cuando una trama 1120 codificada con Codificación de la transformada (TC) es precedida y seguida por una trama 1110, 1130 codificada con ACELP. Aquí, la noción de TC incluye la MDCT en bloques largos y cortos como en AAC, así como también la TCX basada en MDCT (TCX-LPD). La Figura 11 muestra marcadores de dominio temporal 1140 y límites de trama 1142, 1144. Las líneas punteadas verticales muestran el comienzo 1142 y el final 1144 de la trama 1120 codificada con TC. LPC1 y LPC2 indican el centro de la ventana de análisis para calcular dos filtros LPC: El LPC1 calculado al comienzo 1142 de la trama 1120 codificada con TC, y el LPC2 calculado al final 1144 de la misma trama 1120. Se asume que la trama 1110 a la izquierda del marcador "LPC1" se ha codificado con ACELP. También se asume que la trama 1130 a la derecha del marcador "LPC2" se ha codificado con ACELP.
- [0235]** Hay cuatro líneas 1150, 1160, 1170, 1180 en la Fig. 11. Cada línea representa una etapa en el cálculo de la FAC blanco en el codificador. Se debe comprender que cada línea está alineada en términos del tiempo con la línea de arriba.
- [0236]** La línea 1 (1150) de la Fig. 11 representa la señal de audio original, segmentada en las tramas 1110, 1120, 1130 como se especificó anteriormente. Se asume que la trama media 1120 está codificada en el dominio MDCT, utilizando FDNS y se denominará trama TC. Se asume que la señal en la trama anterior 1110 se ha codificado en el modo ACELP. Esta secuencia de modos de codificación (ACELP, a continuación TC, a continuación ACELP) se elige de manera para ilustrar todo el procesamiento en FAC ya que FAC tiene relación con ambas transiciones (ACELP a TC y TC a ACELP).

**[0237]** La línea 2 (1160) de la Fig. 11 corresponde a las señales decodificadas (síntesis) en cada trama (que pueden determinarse mediante el codificador utilizando el conocimiento del algoritmo de decodificación). La curva superior 1162, que se extiende desde el comienzo hasta el final de la trama TC, muestra el efecto de la formación en ventana (plano en el medio pero no en el comienzo y en el final). El efecto de plegado se muestra a través de las curvas inferiores 1164, 1166 al comienzo y al final del segmento (con el signo “-” al comienzo del segmento y “+” al final del segmento). Entonces se puede usar la FAC para corregir estos efectos.

**[0238]** La línea 3 (1170) de la Fig. 11 representa la contribución ACELP, utilizada al comienzo de la trama TC para reducir la carga de codificación de FAC. Esta contribución ACELP está formada por dos partes: 1) la síntesis ACELP plegada, formada en ventana 877f, 1170 del final de la trama anterior, y 2) la respuesta de entrada cero formada en ventana 877j, 1172 del filtro LPC1.

**[0239]** Se debe observar aquí que la síntesis ACELP plegada y formada en ventana 1110 puede ser equivalente a la síntesis ACELP plegada y formada en ventana 1060, y que la respuesta de entrada cero formada en ventana 1172 puede ser equivalente a la respuesta de entrada cero ACELP formada en ventana 1062. En otras palabras, el codificador de señal de audio puede estimar (o calcular) el resultado de la síntesis 1162, 1164, 1166, 1170, 1172, que se obtendrá en el lado de un decodificador de señal de audio (bloques 869a y 877).

**[0240]** El error ACELP que se muestra en la línea 4 (1180) se obtiene entonces simplemente mediante la sustracción de la línea 2 (1160) y la línea 3 (1170) de la línea 1 (1150) (bloque 870). Una vista aproximada de la envolvente esperada de la señal de error 871, 1182 en el dominio temporal se muestra en la línea 4 (1180) de la Fig. 11. Se espera que el error en la trama ACELP (1120) sea aproximadamente plano en amplitud en el dominio temporal. A continuación, se espera que el error en la trama TC (entre los marcadores LPC1 y LPC2) exhiba la forma general (envolvente de dominio temporal) según se muestra en este segmento 1182 de la línea 4 (1180) de la Fig. 11.

**[0241]** Para compensar de manera eficiente los efectos de solapamiento de dominio temporal y de la formación en ventana en el comienzo y en el final de la trama TC de la línea 4 de la Fig. 10, y asumiendo que la trama TC utiliza FDNS, se aplica FAC según la Fig. 11. Se debe observar que la Fig. 11 describe este procesamiento tanto para la parte izquierda (transición de ACELP a TC) como la parte derecha (transición de TC a ACELP) de la trama TC.

**[0242]** Para resumir, el error de la trama de codificación de la transformada 871, 1182, que está representado por los coeficientes de cancelación de solapamiento codificados 856, 936 se obtiene mediante la sustracción de la salida de la trama de codificación de la transformada 1162, 1164, 1166 (descrita, por ejemplo, por la señal 869b) y la contribución ACELP 1170, 1172 (descrita, por ejemplo, por la señal 872) de la señal 1152 en el dominio original (es decir, el dominio temporal). Por consiguiente, se obtiene la señal de error de la trama de la codificación de la transformada 1182.

**[0243]** A continuación, se describirá la codificación del error de la trama de codificación de la transformada 871, 1182.

**[0244]** En primer lugar, se calcula un filtro de ponderación 874, 1210  $W_1(z)$  a partir del filtro LPC1. La señal de error 871, 1182 al comienzo de la trama TC 1120 en la línea 4 (1180) de la Fig. 11 (que también se denomina FAC blanco en las Fig. 11 y 12) se filtra entonces a través de  $W_1(z)$ , que tiene un estado inicial, o memoria de filtro, el error ACELP 871, 1182 en la trama ACELP 1120 en la línea 4 de la Fig. 11. La salida del filtro 874, 1210  $W_1(z)$  en la parte superior de la Fig. 12 forma a continuación la entrada de una transformada DCT-IV 875, 1220. Los coeficientes de la transformada 875a, 1222 de la DCT-IV 875, 1220 se cuantifican entonces y se codifican utilizando la herramienta AVQ 876 (representada por Q, 1230). Esta herramienta AVQ es la misma que la que se utiliza para cuantificar los coeficientes LPC. Estos coeficientes codificados se transmiten al decodificador. La salida de AVQ 1230 es entonces la entrada de una DCT-IV inversa 963, 1240 para formar una señal de dominio temporal 963a, 1242. Esta señal de dominio temporal se filtra a continuación a través del filtro inverso 964, 1250,  $1/W_1(z)$  que tiene memoria cero (estado inicial cero). El filtrado a través de  $1/W_1(z)$  se extiende más allá de la longitud de la FAC blanco utilizando una entrada cero para las muestras que se extienden después de la FAC blanco. La salida 964a, 1252 del filtro 1250,  $1/W_1(z)$  es la síntesis FAC, que es la señal de corrección (por ejemplo, señal 964a) que ahora se puede aplicar al comienzo de la trama TC para compensar por los efectos de la formación en ventana y el solapamiento del dominio temporal.

**[0245]** Ahora, haciendo referencia al procesamiento para la corrección de la formación en ventana y el solapamiento del dominio de tiempo al final de la trama TC, consideramos la parte inferior de la Fig. 12. La señal de error 871, 1182b al final de la trama TC 1120 en la línea 4 de la Fig. 11 (FAC blanco) se filtra a través del filtro 874, 1210;  $W_2(z)$ , que tiene un estado inicial, o memoria del filtro, el error en la trama TC 1120 en la línea 4 de la Fig. 11. A continuación, todas las etapas de procesamiento adicionales son las mismas que para la parte superior de la Fig. 12 que hizo referencia al procesamiento del objetivo FAC en el comienzo de la trama TC, con la excepción de la extensión ZIR en la síntesis FAC.

**[0246]** Cabe observar que el procesamiento en la Fig. 12 se realiza completamente (de izquierda a derecha)

cuando se aplica al codificador (para obtener la síntesis FAC local), mientras que del lado del decodificador el procesamiento en la Fig. 12 sólo se aplica comenzando por los coeficientes decodificados DCT-IV recibidos.

## 9. Corriente de bits

5

**[0247]** A continuación, se describirán algunos detalles acerca de la corriente de bits para facilitar la comprensión de la presente invención. Se debe observar aquí que se puede incluir una cantidad significativa de información en la corriente de bits.

10 **[0248]** Sin embargo, un contenido de audio de una trama codificada en el modo de dominio frecuencial está representado principalmente por un elemento de la corriente de bits denominado "fd\_channel\_stream()". Este elemento "fd\_channel\_stream()" de la corriente de bits comprende información de una ganancia global "global\_gain", datos del factor de escala codificados "scale\_factor\_data()" y datos espectrales aritméticamente codificados "ac\_spectral\_data". Además, el elemento "fd\_channel\_stream()" de la corriente de bits comprende selectivamente datos de cancelación de solapamiento hacia delante que incluyen la información de una ganancia (también denominada "fac\_data(1)"), si (y solo si) se ha codificado una trama anterior (también denominada "supertrama" en algunas realizaciones y ejemplos) en el modo de dominio de predicción lineal y si la última subtrama de la trama anterior se codificó en el modo ACELP. En otras palabras, los datos de cancelación de solapamiento hacia delante que incluyen información de una ganancia se proporcionan de modo selectivo para una trama de audio de modo de dominio frecuencial, si la trama o subtrama anterior se codificó en el modo ACELP. Esto es ventajoso, ya que una cancelación de solapamiento se puede efectuar por una mera funcionalidad de superposición y suma entre una trama de audio anterior o una subtrama de audio codificada en el modo TCX-LPD y la trama de audio actual codificada en el modo de dominio frecuencial, como se ha explicado anteriormente.

25 **[0249]** Para obtener detalles, se hace referencia a la Fig. 14, que muestra una representación de la sintaxis del elemento "fd\_channel\_stream()" de la corriente de bits que comprende la información de la ganancia global "global\_gain", los datos del factor de escala "scale\_factor\_data()", los datos espectrales codificados aritméticamente "ac\_spectral\_data()". La variable "core\_mode\_last" describe un último modo de núcleo y toma el valor de cero para una codificación de dominio frecuencial basada en el factor de escala y toma el valor de uno para una codificación basada en parámetros de dominio de predicción lineal (TCXLPD o ACELP). La variable "last\_lpd\_mode" describe un modo LPD de una última trama o subtrama y toma el valor de cero para una trama o subtrama codificada en el modo ACELP.

**[0250]** Haciendo referencia ahora a la Fig. 15, se describirá la sintaxis para un elemento "lpd\_channel\_stream()" de la corriente de bits, que codifica la información de una trama de audio (también denominada "supertrama") codificada en el modo de dominio de predicción lineal. La trama de audio ("supertrama") codificada en el modo de dominio de predicción lineal puede comprender una pluralidad de subtramas (a veces también denominadas "tramas", por ejemplo, en combinación con la terminología "supertrama"). Las subtramas (o "tramas") pueden ser de diferentes tipos, de manera que algunas de las subtramas se puedan codificar en el modo TCX-LPD, mientras que otras subtramas se puedan codificar en el modo ACELP.

**[0251]** La variable "acelp\_core\_mode" de la corriente de bits describe el esquema de asignación de bits en caso de que se utilice ACELP. El elemento "lpd\_mode" de la corriente de bits se ha explicado anteriormente. La variable "first\_tcx\_flag" se establece como verdadera al comienzo de cada trama codificada en el modo LPD. La variable "first\_lpd\_flag" es un indicador que indica si la trama o supertrama actual es la primera de una secuencia de tramas o supertramas que se codifican en el dominio de codificación de predicción lineal. La variable "last\_lpd" se actualiza para describir el modo (ACELP; TCX256; TCX512; TCX1024) en el cual se codificó la última subtrama (o trama). Como se puede observar en el número de referencia 1510, los datos de cancelación de solapamiento hacia delante sin información de una ganancia ("fac\_data(0)") se incluyen en una subtrama que se codifica en el modo TCX-LPD ( $\text{mod}[k]>0$ ) si la última subtrama se codificó en el modo ACELP ( $\text{last\_lpd\_mode}==0$ ) y en una subtrama codificada en el modo ACELP ( $\text{mod}[k]==0$ ) si la subtrama anterior se codificó en el modo TCX-LPD ( $\text{last\_lpd\_mode}>0$ ).

**[0252]** Si, en contraste, la trama anterior se codificó en el modo de dominio frecuencial ( $\text{core\_mode\_last}=0$ ) y la primera subtrama de la trama actual se codifica en el modo ACELP ( $\text{mod}[0]==0$ ), hay datos de cancelación de solapamiento hacia delante que incluyen información de una ganancia ("fac\_data(1)") en el elemento "lpd\_channel\_stream" de la corriente de bits.

**[0253]** Para resumir, los datos de cancelación de solapamiento hacia delante que incluyen un valor de ganancia de cancelación de solapamiento hacia delante específico se incluyen en la corriente de bits, si hay una transición directa entre una trama codificada en el dominio frecuencial y una trama o subtrama codificada en el modo ACELP. En contraste, si hay una transición entre una trama o subtrama codificada en el modo TCX-LPD y una trama o subtrama codificada en el modo ACELP, se incluye información de cancelación de solapamiento hacia delante sin un valor de ganancia específico de cancelación de solapamiento hacia delante en la corriente de bits.

65 **[0254]** Haciendo referencia ahora a la Fig. 16, se describirá la sintaxis de los datos de cancelación de

solapamiento hacia delante, que se describe mediante el elemento "fac\_data()" de la corriente de bits. El parámetro "useGain" indica si hay un elemento "fac\_gain" específico de la corriente de bits con valor de ganancia de cancelación de solapamiento hacia delante, como se puede observar en el número de referencia 1610. Además, el elemento "fac\_data" de la corriente de bits comprende una pluralidad de elementos "nq[i]" de la corriente de bits de un número 5 del libro de códigos y un número de elementos "fac[i]" de la corriente de bits "fac\_data".

**[0255]** La decodificación de dicho número del libro de códigos y dichos datos de cancelación de solapamiento hacia delante se ha descrito anteriormente.

## 10. Alternativas de implementación

**[0256]** Aunque se han descrito algunos aspectos en el contexto de un aparato, es claro que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa de un procedimiento o una característica de una etapa de un procedimiento. De manera análoga, los 15 aspectos descritos en el contexto de una etapa de un procedimiento representan también una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente. Algunas o todas las etapas del procedimiento se pueden ejecutar mediante (o utilizando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunos ejemplos, alguna o muchas de las etapas del procedimiento más importantes se pueden ejecutar mediante dicho aparato.

**[0257]** La señal de audio codificada de la invención se puede almacenar en un medio de almacenamiento digital o se puede transmitir en un medio de transmisión tal como un medio de transmisión inalámbrica o un medio de transmisión con cable tal como Internet.

**[0258]** Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o software. La implementación se puede llevar a cabo utilizando un medio de almacenamiento digital, por ejemplo un disquete, un DVD, un Blue-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tengan señales de control que se puedan leer electrónicamente almacenadas allí, que cooperen con (o sean capaces de cooperar con) un sistema informático programable de tal 25 manera que se realice el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

**[0259]** Algunos ejemplos comprenden un soporte de datos que tiene señales de control que se pueden leer electrónicamente, que son capaces de cooperar con un sistema informático programable, tal como se realiza uno de 35 los procedimientos descritos en esta invención.

**[0260]** Generalmente, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, donde el código de programa es operativo para realizar uno de los procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa, por 40 ejemplo, se puede almacenar en un soporte legible por máquina.

**[0261]** Otros ejemplos comprenden el programa informático para realizar uno de los procedimientos descritos en esta invención, almacenados en un soporte legible por ordenador.

**[0262]** En otras palabras, un ejemplo del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en esta invención, cuando el programa informático se ejecuta en un ordenador.

**[0263]** Un ejemplo adicional es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) comprendiendo, grabado allí, el programa informático para realizar uno de los procedimientos descritos en esta invención. El soporte de datos, el medio de almacenamiento digital o el medio grabado son típicamente tangibles y/o no transicionales.

**[0264]** Un ejemplo adicional es, por lo tanto, una corriente de datos o una secuencia de señales que representa al programa informático para realizar uno de los procedimientos descritos en esta invención. La corriente de datos o la secuencia de señales, por ejemplo, se puede configurar para transferirse vía una conexión de comunicación de 55 datos, por ejemplo, vía Internet.

**[0265]** Un ejemplo adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado a o adaptado para realizar uno de los procedimientos descritos en esta invención. 60

**[0266]** Un ejemplo adicional comprende un ordenador que tiene instalado en sí el programa informático para realizar uno de los procedimientos descritos en esta invención.

**[0267]** Un ejemplo adicional comprende un aparato o un sistema configurado para transferir (por ejemplo, 65



electrónicamente u ópticamente) un programa informático para realizar uno de los procedimientos descritos en esta invención a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa informático al receptor.

**[0268]** En algunos ejemplos, un dispositivo lógico programable (por ejemplo una matriz de puerta programable del campo) se puede utilizar para realizar alguna o todas las funcionalidades de los procedimientos descritos en esta invención. En algunos ejemplos, una matriz de puerta programable del campo puede cooperar con un microprocesador para realizar uno de los procedimientos descritos en esta invención. Generalmente, los procedimientos son preferiblemente realizados por cualquier aparato de hardware.

**[0269]** Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Se comprende que las modificaciones y variaciones de las disposiciones y los detalles descritos en esta invención serán aparentes para otros expertos en la técnica. Es la intención, por lo tanto, tener limitación solo por el alcance de las reivindicaciones de la patente inminente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de esta invención.

## 11. Conclusión

**[0270]** A continuación se resumirá la presente propuesta para la unificación de la formación en ventana de codificación unificada de voz y audio (USAC) y las transiciones de trama.

**[0271]** En primer lugar, se proporcionará una introducción y algo de información de fondo. Un diseño actual (también denominado diseño de referencia) del modelo de referencia de USAC está compuesto por (o comprende) tres módulos de codificación diferentes. Para cada sección de señal de audio dada (por ejemplo, una trama o subtrama), se elige un módulo de codificación (o modo de codificación) para codificar/decodificar esa sección resultante en distintos modos de codificación. Dado que estos módulos se van alternando en actividad, se debe prestar especial atención a las transiciones de un modo al otro. En el pasado, diversas contribuciones han propuesto modificaciones que trataban estas transiciones entre modos de codificación.

**[0272]** Las realizaciones según la presente invención crean una formación en ventana general y un esquema de transición previstos. Se describirá el progreso que se ha logrado en el camino a que se complete este esquema, exhibiendo cada evidencia promisorio de calidad y las mejoras estructurales sistemáticas.

**[0273]** El presente documento resume los cambios propuestos al diseño de referencia (que también se denomina diseño en borrador de trabajo 4) con el fin de crear una estructura de codificación más flexible para USAC, para reducir la sobrecodificación y la complejidad de las secciones codificadas por transformada del códec.

**[0274]** Con el fin de alcanzar un esquema de formación en ventana que evite el costoso muestreo no crítico (sobrecodificación), se introducen dos componentes, que se pueden considerar esenciales en algunas realizaciones:

- 1) la ventana de cancelación de solapamiento hacia delante (FAC); y
- 2) el moldeado del sonido del dominio frecuencial (FDNS) para el ramal de codificación por transformada en el códec de núcleo LPD (TCX, también denominado TCX-LPD o wLPT).

**[0275]** La combinación de ambas tecnologías hace posible el uso de un esquema de formación en ventana que permite una conmutación altamente flexible de la longitud de la transformada a una demanda de bits mínima.

**[0276]** A continuación se describirán los desafíos de sistemas de referencia para facilitar la comprensión de las ventajas proporcionadas por las realizaciones según la invención. Un concepto de referencia según el borrador de trabajo 4 del borrador de norma USAC está compuesto por un códec de núcleo conmutado que trabaja junto con una etapa anterior/posterior de procesamiento formada por (o comprendiendo) MPEG envolvente y un módulo SBR (Replicación de banda espectral, por sus siglas en inglés) mejorado. El núcleo conmutado presenta un códec de dominio frecuencial (FD) y un códec de dominio de predicción lineal (LPD). El último utiliza un módulo ACELP y un codificador de transformada que trabaja en el dominio ponderado ("Transformada de predicción lineal ponderada" (wLPT), también conocido como excitación por código de la transformada, (TCX)). Se ha descubierto que debido a los principios de codificación fundamentalmente diferentes, es especialmente desafiante el manejo de las transiciones entre los modos. Se ha descubierto que se debe procurar que los modos se entremezclen eficientemente.

**[0277]** A continuación, se describirán los desafíos que surgen en las transiciones de dominio temporal a dominio frecuencial (ACELP↔wLPT, ACELP↔FD). Se ha descubierto que las transiciones de la codificación de dominio temporal a la codificación de dominio de la transformada son engañosas, en particular, ya que el codificador de la transformada se basa en la propiedad de cancelación de solapamiento de dominio de la transformada (TDAC) de los bloques vecinos en MDCT. Se ha descubierto que un bloque codificado de dominio frecuencial no se puede decodificar en su totalidad sin información adicional de sus bloques superpuestos adyacentes.

**[0278]** A continuación, se describirán los desafíos que surgen en las transiciones del dominio de señal al dominio de predicción lineal (FD $\leftrightarrow$ ACELP, FD $\leftrightarrow$ wLPT). Se ha descubierto que las transiciones a y desde el dominio de predicción lineal implican una transición de distintos paradigmas de moldeado de sonido de cuantificación. Se ha descubierto que los paradigmas utilizan una forma diferente de transmitir y aplicar información de moldeado de sonido psicoacústicamente motivada que puede generar discontinuidades en la calidad percibida en lugares donde el modo de codificación cambia.

**[0279]** A continuación, se describirán detalles acerca de una matriz de transición de trama de un concepto de referencia según el trabajo en borrador 4 del borrador de norma USAC. Debido a la naturaleza híbrida del modelo de referencia USAC, hay una multitud de transiciones de ventana concebibles. La tabla de 3 por 3 de la Fig. 4 muestra un panorama general de estas transiciones como se implementan según el concepto del trabajo en borrador 4 del borrador de norma USAC.

**[0280]** Cada una de las contribuciones enumeradas anteriormente trata una o más de las transiciones de la tabla de la Fig. 4. Cabe observar que cada una de las transiciones no homogéneas (las que no están en la diagonal principal) aplican distintas etapas de procesamiento específicas, que son el resultado de un compromiso entre intentar lograr un muestreo crítico, evitar bloquear artefactos, encontrar un esquema común de formación en ventana y permitir una decisión de modo del codificador de bucle cerrado. En algunos casos, este compromiso incluye el coste de eliminar muestras codificadas y transmitidas.

**[0281]** A continuación, se describirán algunos cambios del sistema propuesto. En otras palabras, se describirán las mejoras del concepto de referencia según el trabajo en borrador 4 de USAC. Para abordar las dificultades enumeradas en las transiciones de la ventana, las realizaciones según la invención introducen dos modificaciones al sistema existente, en comparación con los conceptos según el sistema de referencia según el trabajo en borrador 4 del borrador de norma USAC. La primera modificación apunta a mejorar de modo universal la transición del dominio temporal al dominio frecuencial mediante la adopción de una ventana suplementaria de cancelación de solapamiento hacia delante. La segunda modificación asimila el procesamiento de dominios de predicción lineal y de señal mediante la introducción de una etapa de transmutación para los coeficientes LPC, que se pueden aplicar a continuación en el dominio frecuencial.

**[0282]** A continuación se describirá el concepto de moldeado de sonido de dominio frecuencial (FDNS), el cual permite la aplicación de LPC en el dominio frecuencial. El objetivo de esta herramienta (FDNS) consiste en permitir el procesamiento TDAC de los codificadores MDCT que trabajan en distintos dominios. Mientras que la MDCT de la parte del dominio frecuencial de la USAC actúa en el dominio de señal, la wLPT (o TCX) del concepto de referencia opera en el dominio filtrado ponderado. Mediante el reemplazo del filtro de síntesis LPC ponderado, que se utiliza en el concepto de referencia, por una etapa de procesamiento equivalente en el dominio frecuencial, la MDCT de ambos codificadores de la transformada opera en el mismo dominio y se puede lograr la TDAC sin introducir discontinuidades en el moldeado de sonido de cuantificación.

**[0283]** En otras palabras, el filtro de síntesis LPC ponderado 330g se reemplaza por el moldeado de sonido de dominio frecuencial/escala 380e en combinación con la conversión de LPC a dominio frecuencial 380i. Por consiguiente, la MDCT 320g de la ruta de dominio frecuencial y la MDCT 380h del ramal TCX-LPD operan en el mismo dominio, de manera que se logra la cancelación de solapamiento de dominio de la transformada (TDAC).

**[0284]** A continuación se describirán algunos detalles acerca de la ventana de cancelación de solapamiento hacia delante (ventana FAC). La ventana de cancelación de solapamiento hacia delante (FAC) ya se ha introducido y descrito. Esta ventana suplementaria compensa la información de TDAC que falta que, en un código de transformada continuamente en funcionamiento, generalmente es contribuida por la ventana anterior o la siguiente. Dado que el codificador de dominio temporal ACELP no exhibe una superposición a tramas adyacentes, la FAC puede compensar la falta de esta superposición que falta.

**[0285]** Se ha descubierto que mediante la aplicación del filtro LPC en el dominio frecuencial, la ruta de codificación LPD pierde algo del impacto de suavizamiento del filtrado LPC interpolado entre segmentos codificados por ACELP y wLPT (TCX-LPD). Sin embargo, se ha descubierto que, dado que la FAC se diseñó para permitir una transición favorable exactamente en este lugar, también puede compensar este efecto.

**[0286]** Como consecuencia de la introducción de la ventana FAC y FDNS, todas las transiciones concebibles se pueden lograr sin una sobreecodificación inherente.

**[0287]** A continuación, se describirán algunos detalles acerca del esquema de formación en ventana.

**[0288]** Ya se ha descrito cómo la ventana FAC puede fusionar las transiciones entre ACELP y wLPT. Para obtener mayores detalles, se hace referencia al siguiente documento: ISO/IEC JTC1/SC29/WG11, MPEG2009/M16688, junio-julio 2009, Londres, Reino Unido, "Alternatives for windowing in USAC".

**[0289]** Dado que FDNS conmuta la wLPT al dominio de señal, la ventana FAC se puede aplicar ahora a ambas, las transiciones de/a ACELP a/de wLPT y también de/a ACELP a/de FD exactamente de la misma manera (o, al menos, en una manera similar).

5

**[0290]** De manera similar, las transiciones del codificador de la transformada basada en TDAC que previamente eran posibles exclusivamente entre ventanas FD o entre ventanas wLPT (es decir, de/a FD de/a FD; o de/a wLPT a/de wLPT) se pueden aplicar ahora al transgredir del dominio frecuencial a wLPT o viceversa. De este modo, ambas tecnologías combinadas permiten la conmutación de las 64 muestras de la red de tramas ACELP a la derecha (hacia

10

“más adelante” en el eje de tiempo). Al hacerlo, la ventana de 64 muestras de superposición-suma en un extremo y la ventana de la transformada de dominio frecuencial extra larga en el otro extremo ya no son necesarias. En ambos casos, se puede evitar una sobrecodificación de 64 muestras en realizaciones según la invención en comparación con los conceptos de referencia. Más importante aún, todas las otras transiciones permanecen como están y no se necesitan más modificaciones adicionales.

15

**[0291]** A continuación se discutirá brevemente la nueva matriz de transición de tramas. Se proporciona un ejemplo de una nueva matriz de transición en la Fig. 5. Las transiciones en la diagonal principal permanecen como están en el trabajo en borrador 4 del borrador de norma USAC. Todas las otras transiciones se pueden tratar con la ventana FAC o TDAC directa en el dominio de señal. En algunas realizaciones solo se necesitan dos longitudes de superposición entre ventanas de dominio de la transformada adyacentes para el esquema anterior, a saber 1024

20

muestras y 128 muestras, aunque otras longitudes de superposición también son concebibles.

## 12. Evaluación subjetiva

**[0292]** Se debe observar que se han llevado a cabo dos pruebas de audio para mostrar que en el estado actual de implementación la tecnología nueva propuesta no compromete la calidad. Eventualmente, se espera que las realizaciones según la invención proporcionen un aumento de la calidad debido a los ahorros de bits en los lugares donde las muestras se descartaron previamente. Como otro efecto lateral, el control clasificador en el codificador puede ser mucho más flexible ya que las transiciones de modo no están más afectadas con muestreo no crítico.

30

## 13. Comentarios adicionales

**[0293]** Para resumir lo anterior, la presente descripción describe un esquema de transición y formación en ventana previsto para la USAC que tiene varias virtudes, en comparación con el esquema existente, utilizado en el trabajo en borrador 4 del borrador de norma USAC. El esquema propuesto de formación en ventana y transición mantiene un muestreo crítico en todas las tramas de codificación de transformada, evita la necesidad de transformadas sin potencia de dos y alinea de manera adecuada todas las tramas de codificación de transformada. La propuesta se basa en dos herramientas nuevas. La primera herramienta, la cancelación de solapamiento hacia delante (FAC) se describe en la referencia [M16688]. La segunda herramienta, el moldeado de sonido de dominio frecuencial (FDNS) permite el procesamiento de tramas de dominio frecuencial y tramas wLPT en el mismo dominio sin introducir discontinuidades en el moldeado de sonido de cuantificación. De este modo, todas las transiciones de modos en USAC se pueden manejar con estas dos herramientas básicas, permitiendo una formación en ventana armonizada para todos los modos por codificación de transformada. Los resultados subjetivos de las pruebas también se proporcionaron en la presente descripción, mostrando que las herramientas propuestas proporcionan una calidad equivalente o mejor en comparación con el concepto de referencia según el trabajo en borrador 4 del borrador de norma USAC.

35

40

45

## Referencias

**[0294]** [M16688] ISO/IEC JTC1/SC29/WG11, MPEG2009/M16688, junio-julio 2009, Londres, Reino Unido, “Alternatives for windowing in USAC”

50

## REIVINDICACIONES

1. Un decodificador de señal de audio multi-modo (200; 360; 900) para proporcionar una representación decodificada (212; 399; 998) de un contenido de audio en base a una representación codificada (210; 361; 901) del contenido de audio,

donde el decodificador de señal de audio multi-modo está configurado para conmutar entre tres modos, un modo del dominio de la frecuencia, que utiliza una información de coeficiente espectral y una información de factor de escala, un modo de dominio de predicción lineal de excitación codificada por transformada, que utiliza una información de excitación codificada por transformada y una información de parámetro del dominio de predicción lineal, y un modo de predicción lineal excitado de código algebraico, que utiliza una información de excitación de código algebraico y una información de parámetro del dominio de predicción lineal, comprendiendo el decodificador de la señal de audio:

una ruta del dominio de la transformada (230, 240, 242, 250, 260; 270, 280; 380; 930) configurada para obtener una representación del dominio temporal (212; 386; 938), en la forma de una señal del dominio de tiempo reducida de solapamiento, de una porción del contenido de audio codificado en un modo de dominio de la transformada en base a un primer conjunto (220; 382; 944a) de coeficientes espectrales, sobre la base de una representación (224; 936) de una señal de estímulo de cancelación de solapamiento y sobre la base de una pluralidad de parámetros de dominio de predicción lineal (222; 384; 950a), donde la ruta de dominio de la transformada comprende un procesador espectral (230; 380e; 945) configurado para aplicar un moldeado espectral al primer conjunto (944a) de coeficientes espectrales dependiendo de al menos un subconjunto de los parámetros de dominio de predicción lineal, para obtener una versión de moldeado espectral (232; 380g; 945a) del primer conjunto de coeficientes espectrales, donde la ruta de dominio de la transformada comprende un primer conversor de dominio frecuencial a dominio temporal (240; 380h; 946) configurado para obtener una representación de dominio temporal (242; 940a) del contenido de audio en base a una versión de moldeado espectral del primer conjunto de coeficientes espectrales; donde la ruta de dominio de la transformada comprende un filtro de estímulo de cancelación de solapamiento (250; 964) configurado para filtrar la señal de estímulo de cancelación de solapamiento (224; 963a) dependiendo de al menos un subconjunto de los parámetros de dominio de predicción lineal (222; 384; 934), para derivar una señal de síntesis de cancelación de solapamiento (252; 964a) para cancelar los artefactos de solapamiento de la señal de estímulo de cancelación de solapamiento; y donde la ruta de dominio de la transformada comprende también un combinador (260; 978) configurado para combinar la representación de dominio temporal (242; 940a) del contenido de audio con la señal de síntesis de cancelación de solapamiento (252; 964); o una versión postprocesada del mismo, para obtener una señal de dominio temporal con solapamiento reducido como la representación decodificada (212) del contenido de audio; donde la ruta del dominio de transformada es una ruta del dominio de predicción lineal de excitación codificada de transformada,

donde el decodificador de señal de audio multimodo comprende una etapa de postprocesamiento comprendiendo un módulo envolvente MPEG.

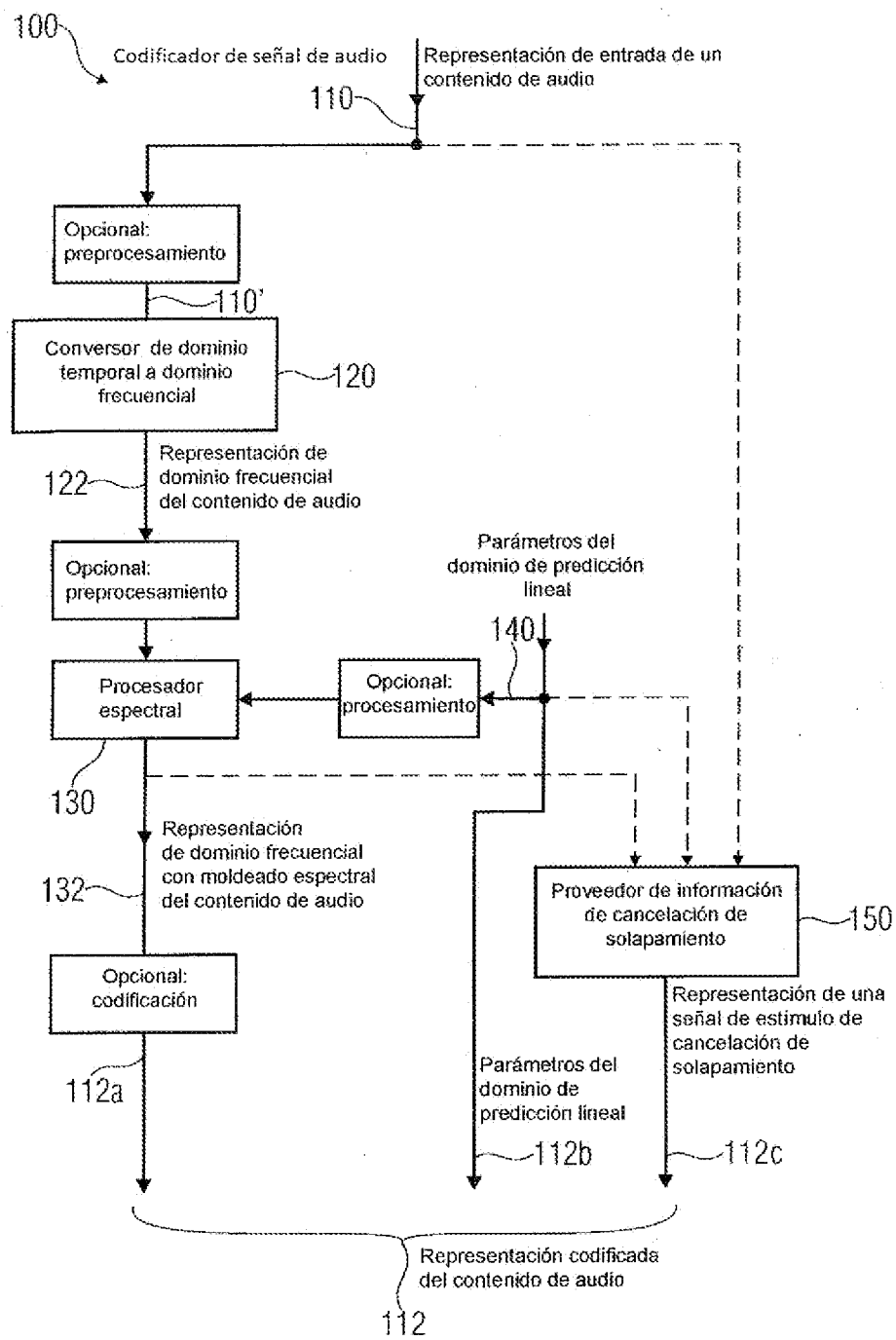
2. Un procedimiento para proporcionar una representación decodificada de un contenido de audio en base a una representación codificada del contenido de audio, comprendiendo el procedimiento:

obtener una representación de dominio temporal de una porción del contenido de audio codificada en un modo de dominio de predicción lineal de excitación codificada de transformada en base a un primer conjunto de coeficientes espectrales, sobre la base de una representación de una señal de estímulo de cancelación de solapamiento y sobre la base de la pluralidad de parámetros de dominio de predicción lineal, donde se aplica un modelado espectral al primer conjunto de coeficientes espectrales dependiendo de al menos un subconjunto de los parámetros del dominio de predicción lineal, para obtener una versión modelada espectralmente del primer conjunto de coeficientes espectrales, y donde una conversión del dominio de la frecuencia al dominio del tiempo se aplica para obtener una representación del dominio del tiempo del contenido de audio sobre la base de la versión moldeada espectralmente del primer conjunto de coeficientes espectrales, y donde la señal de estímulo de cancelación de solapamiento se filtra dependiendo de al menos un subconjunto de los parámetros del dominio de predicción lineal, para derivar una señal de síntesis de cancelación de solapamiento desde la señal de estímulo de cancelación de solapamiento, y donde la representación del dominio del tiempo del contenido de audio se combina con la señal de síntesis de cancelación de solapamiento, o una versión postprocesada de la misma, para obtener una señal del dominio del tiempo de solapamiento reducido, donde el procedimiento es un procedimiento de decodificación multi-modo, donde el procedimiento comprende la conmutación entre tres modos, un modo del dominio de la frecuencia, que utiliza una información del coeficiente espectral y una información del factor de escala, el modo del dominio de predicción lineal de excitación codificada de transformada, que utiliza una información de excitación codificada de

transformada y una información de parámetro del dominio de predicción lineal, y un modo de predicción lineal excitada de código algebraico, que utiliza una información de excitación de código algebraico y una información de parámetro del dominio de predicción lineal; y  
donde el procedimiento comprende realizar un postprocesamiento de sonido envolvente MPEG.

5

3. Un programa informático para realizar el procedimiento según la reivindicación 2, cuando el programa informático se ejecuta en un ordenador.



**FIG. 1**

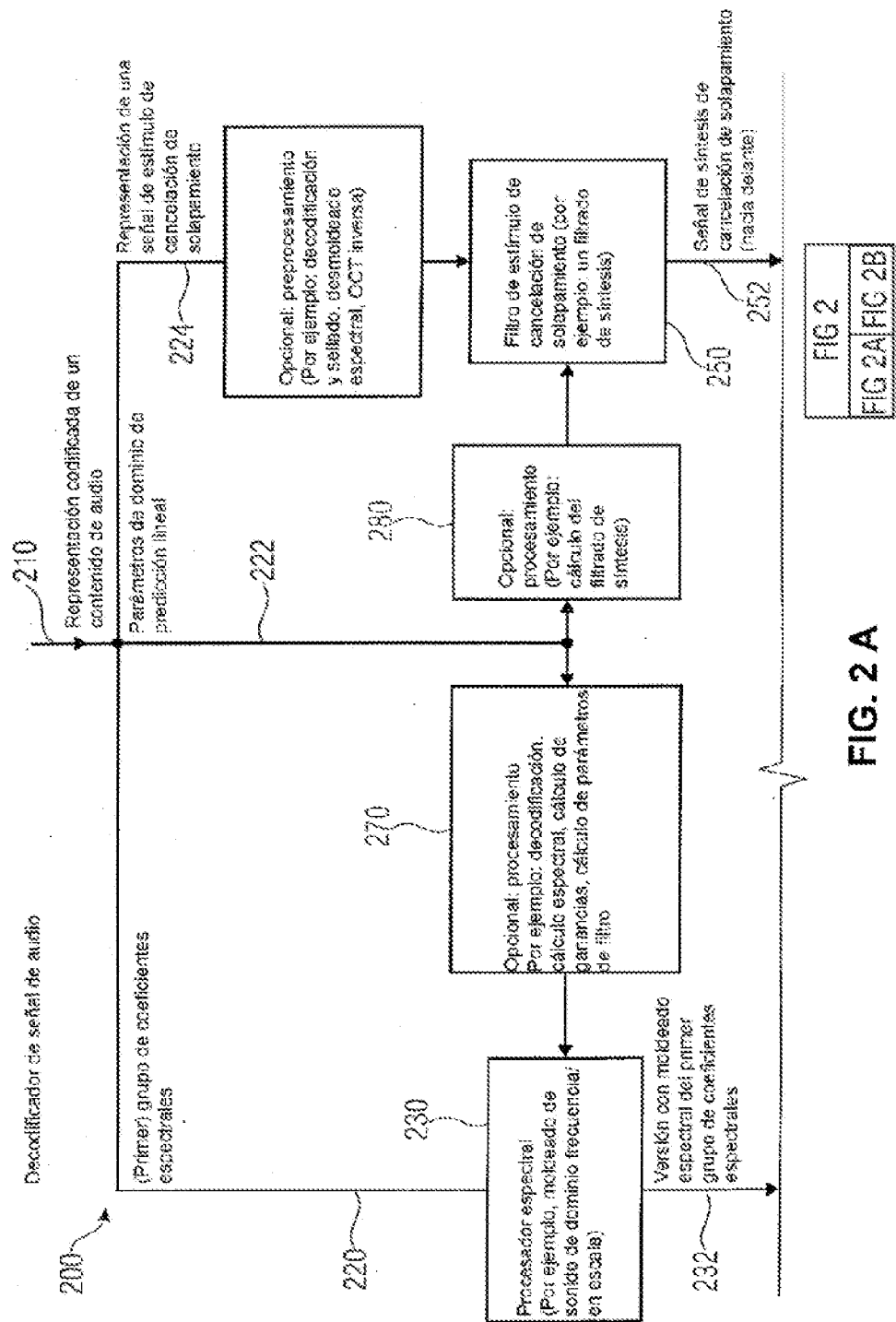


FIG. 2  
FIG. 2A FIG. 2B

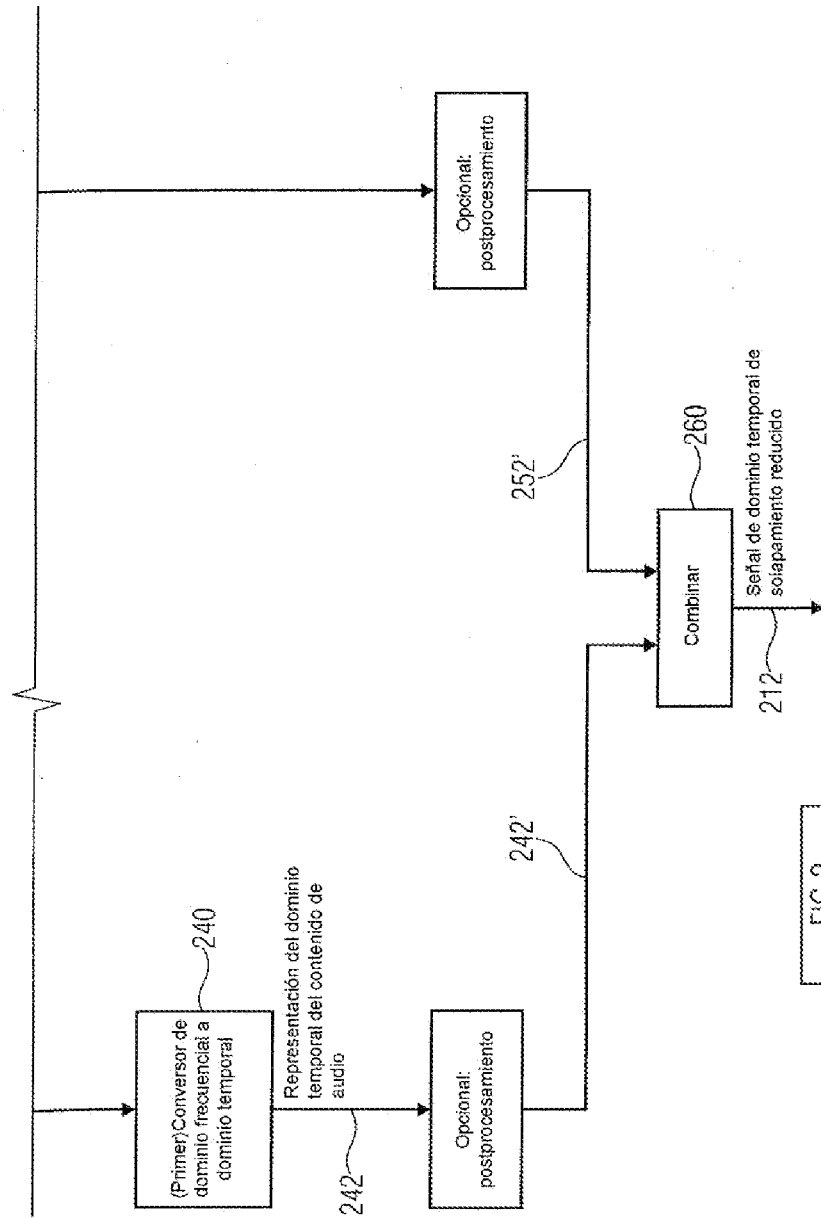


FIG. 2 B

FIG 2  
FIG 2A FIG 2B



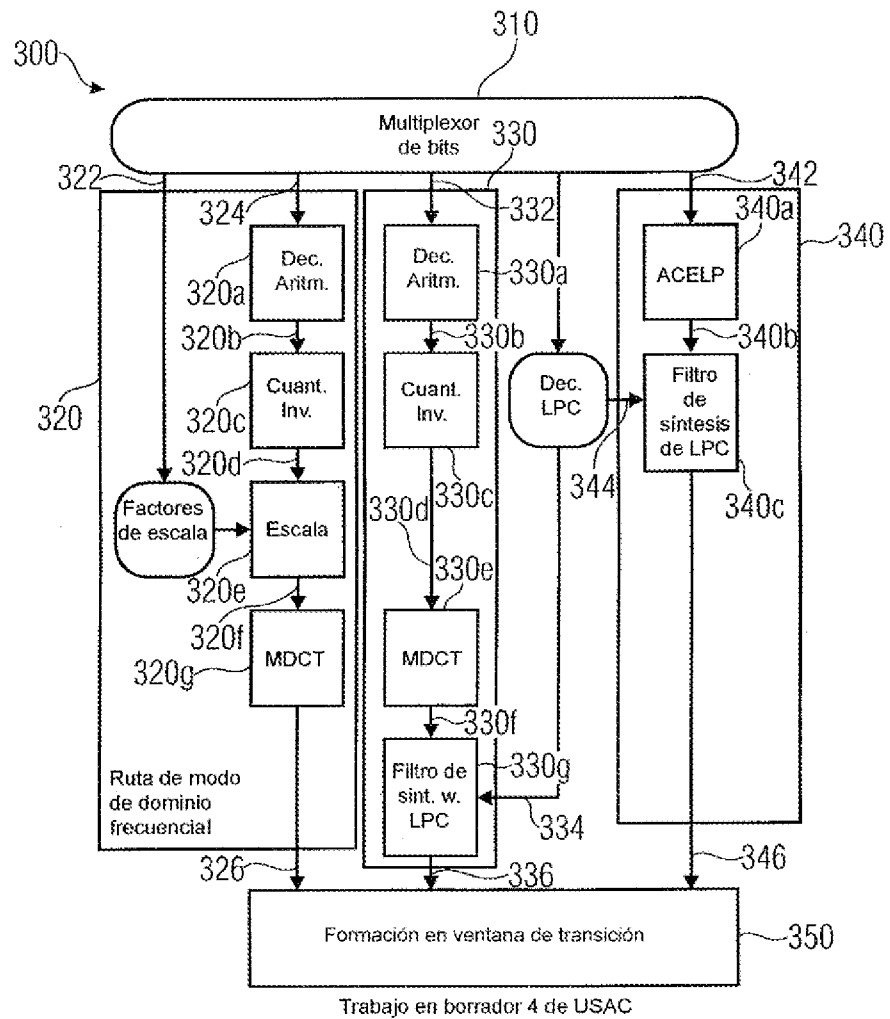


FIG. 3 A

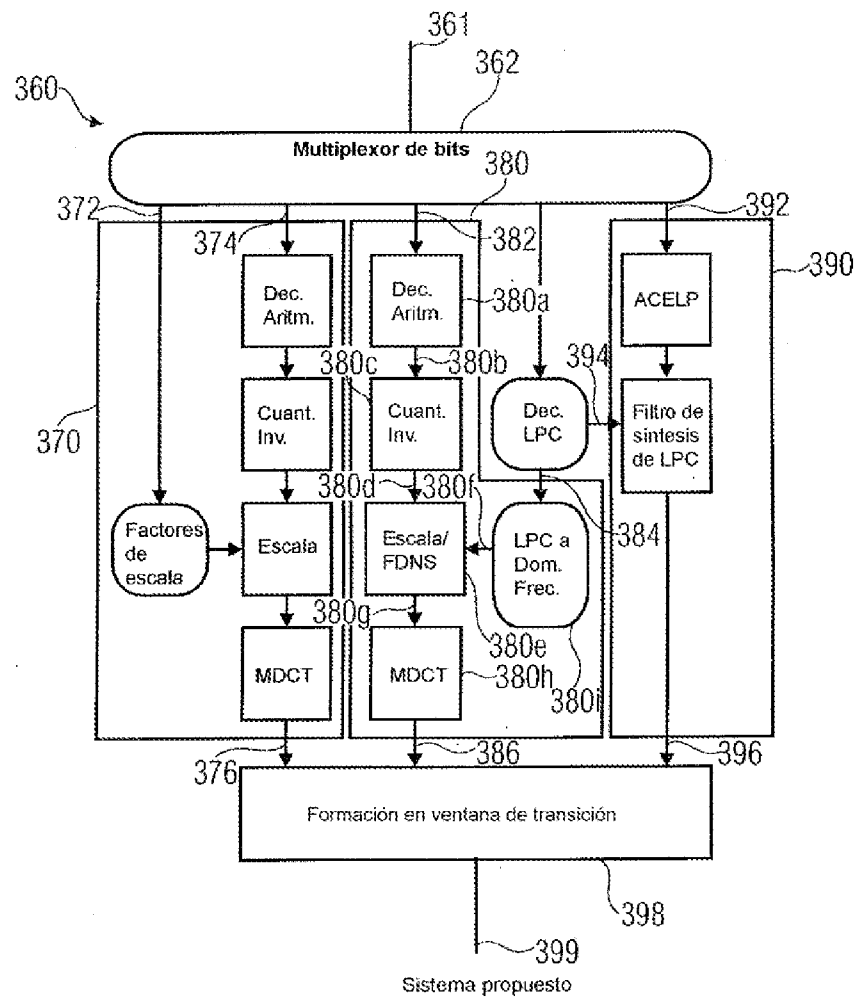


FIG. 3 B

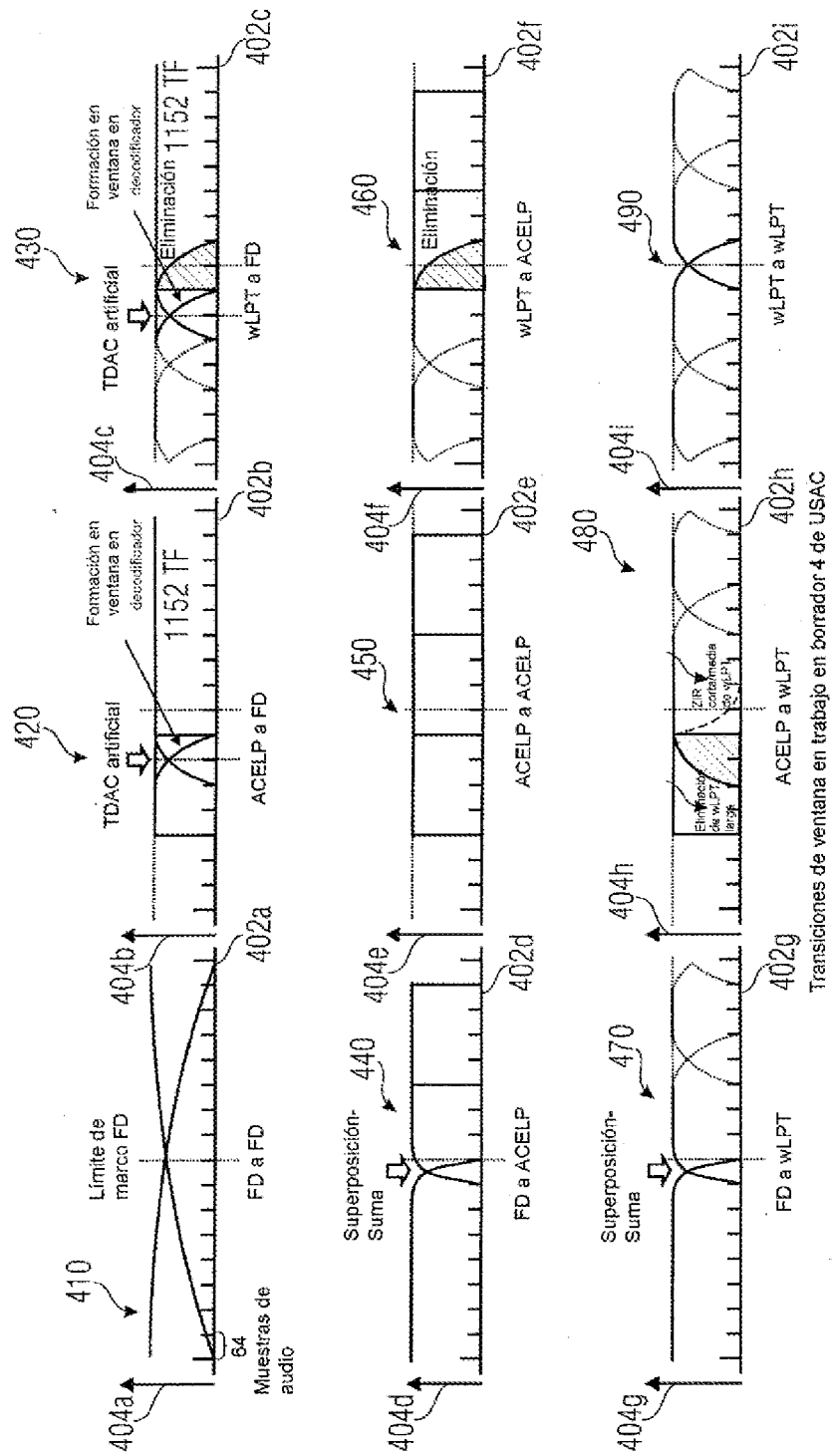
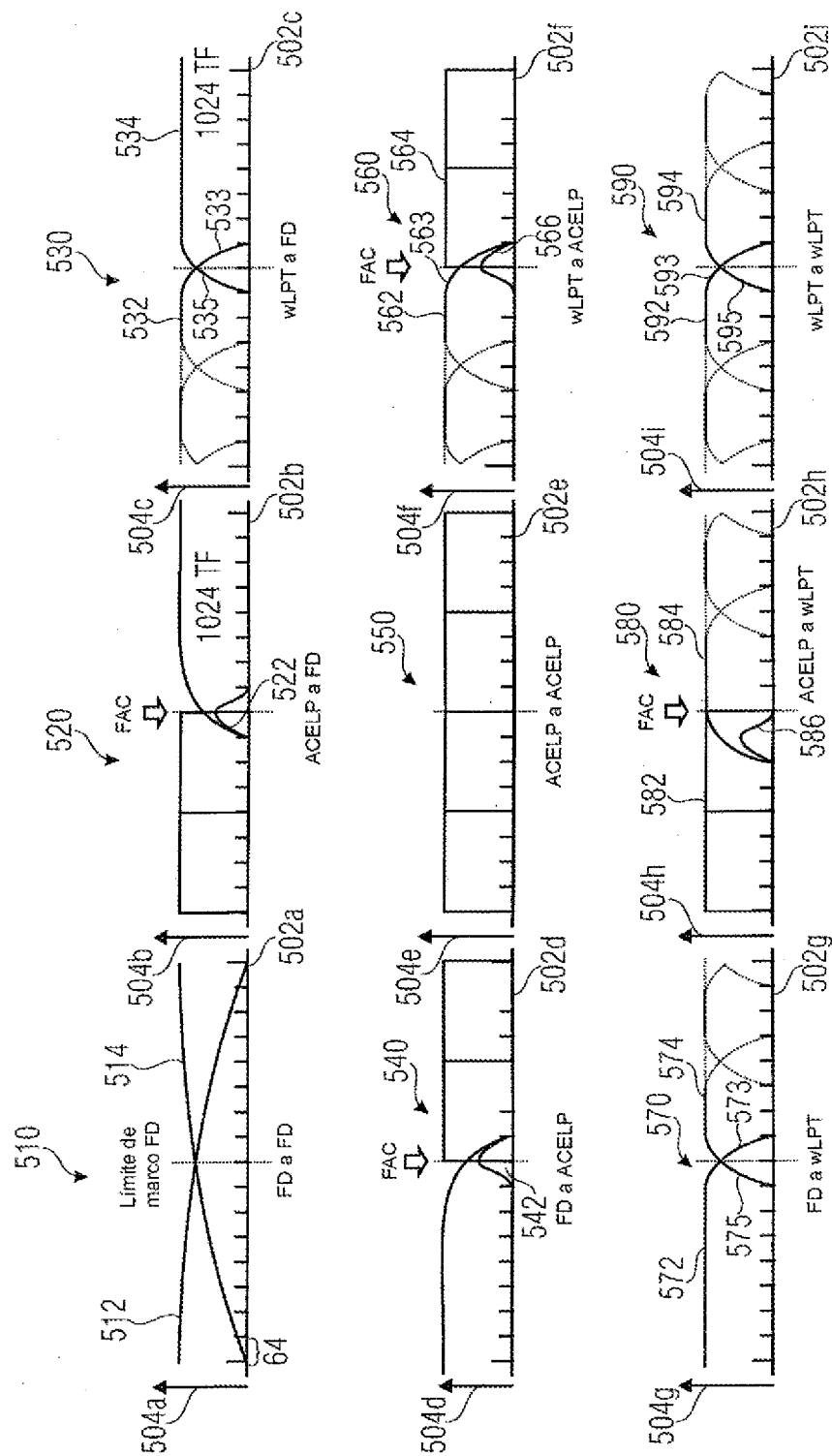


FIG. 4

Transiciones de ventana en trabajo en borrador 4 de USAC



**FIG. 5**

Panorama sobre todos los tipos de ventanas

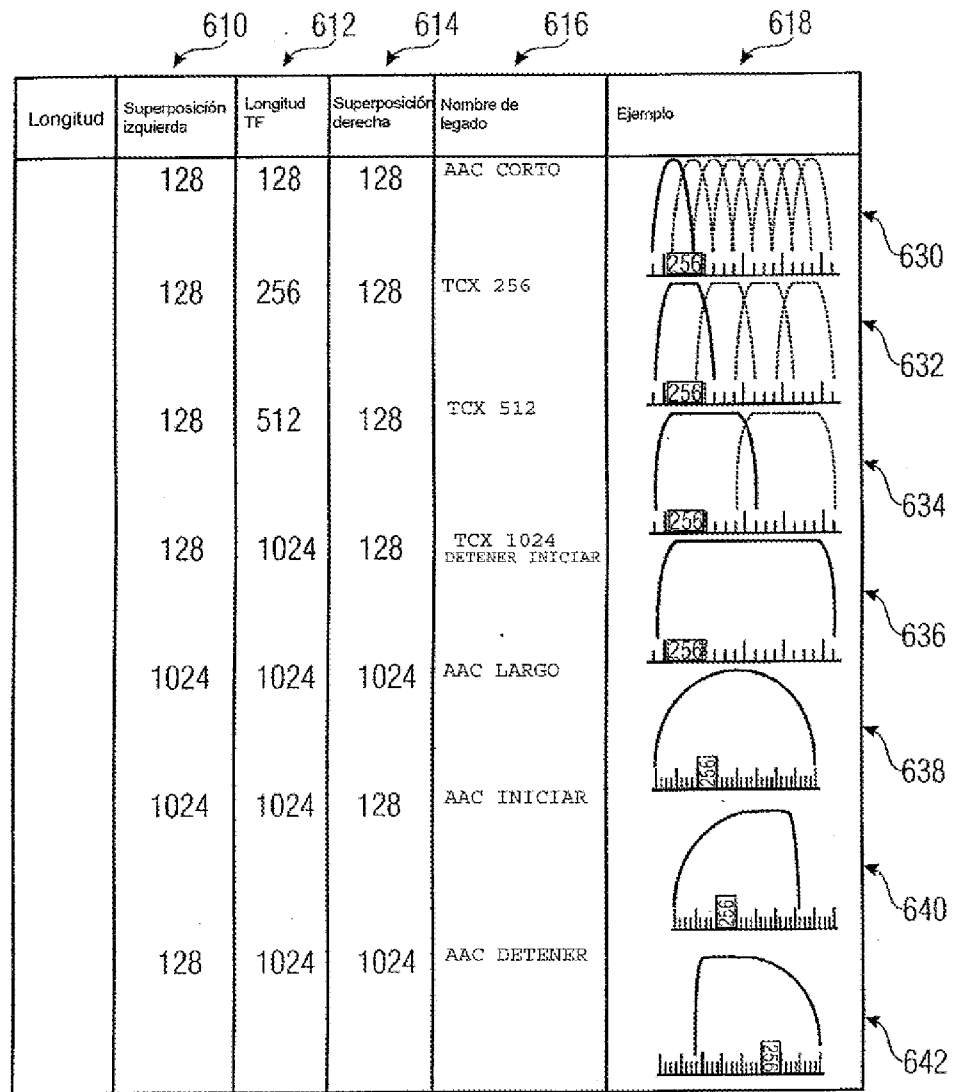


FIG. 6

SECUENCIAS DE VENTANA PERMITIDAS

Sequencia de ventana De ↓ hasta ↑	"AAC long"	"AAC start"	8 x "AAC short"	"AAC stop"	"AAC stopstart"	"LPD TCX"	"LPD ACELP"
"AAC stop"	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					
"AAC long"	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					
"AAC start"			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> (FAC)
8 x "AAC short"			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> (FAC)
"AAC stopstart"			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> (FAC)
"LPD TCX"			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> (FAC)
"LPD ACELP"			<input checked="" type="checkbox"/> (FAC)	<input checked="" type="checkbox"/> (FAC)	<input checked="" type="checkbox"/> (FAC)	<input checked="" type="checkbox"/> (FAC)	<input checked="" type="checkbox"/>

FIG 7

FIG 8A

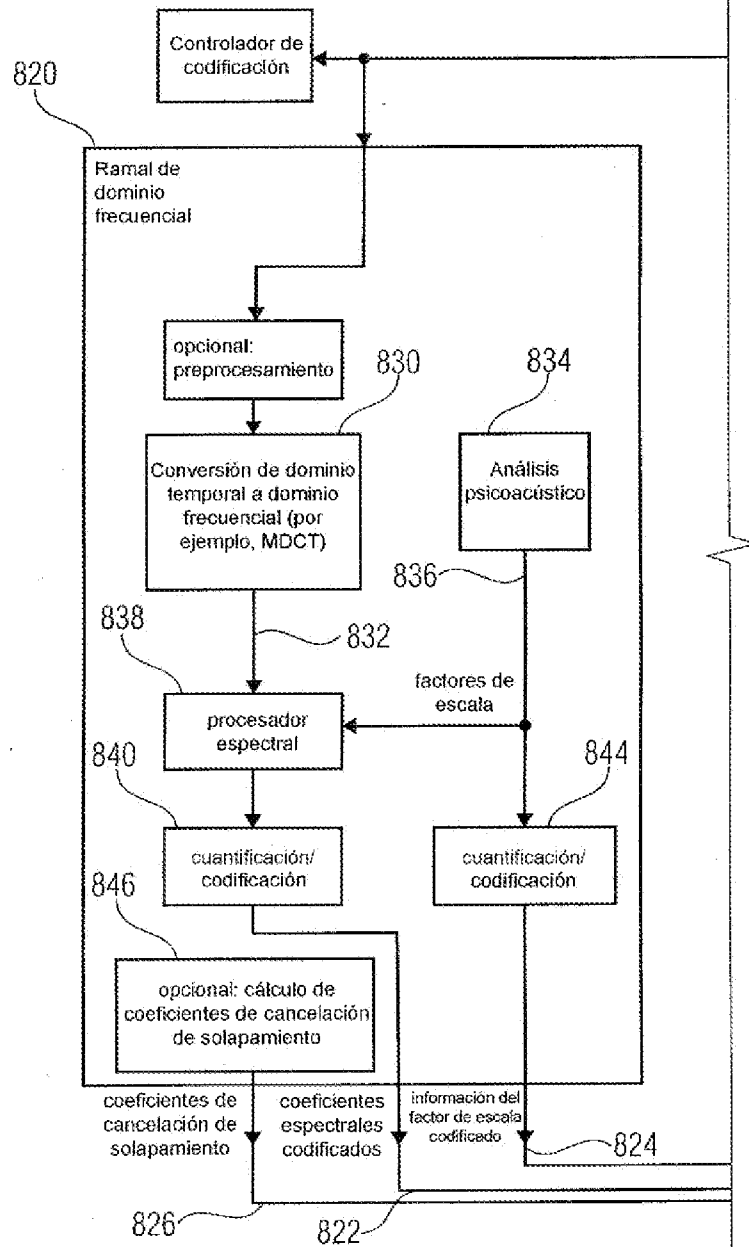


FIG 8  
FIG 8A FIG 8B FIG 8C FIG 8D

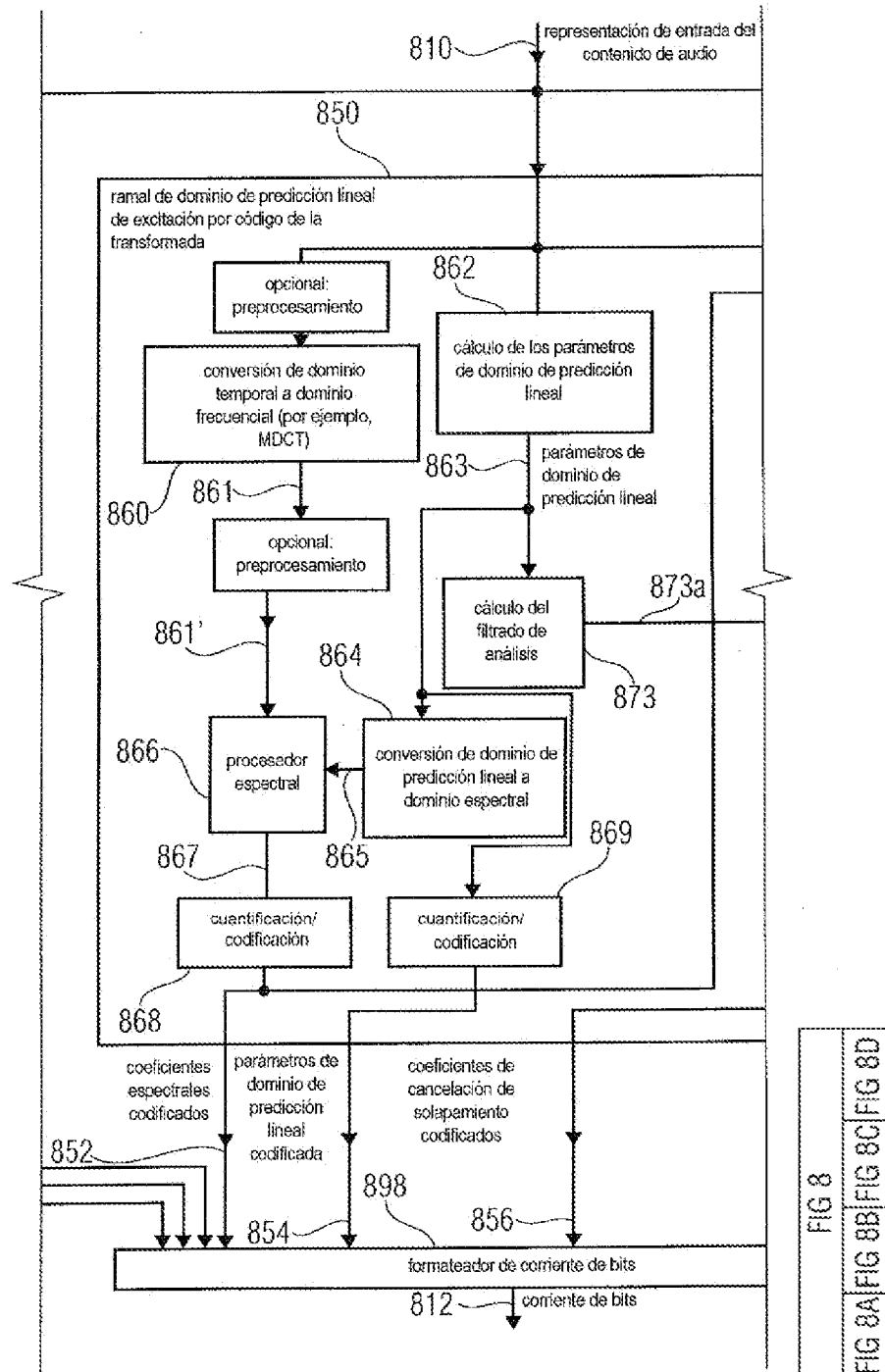


FIG. 8 B

FIG 8  
FIG 8A FIG 8B FIG 8C FIG 8D



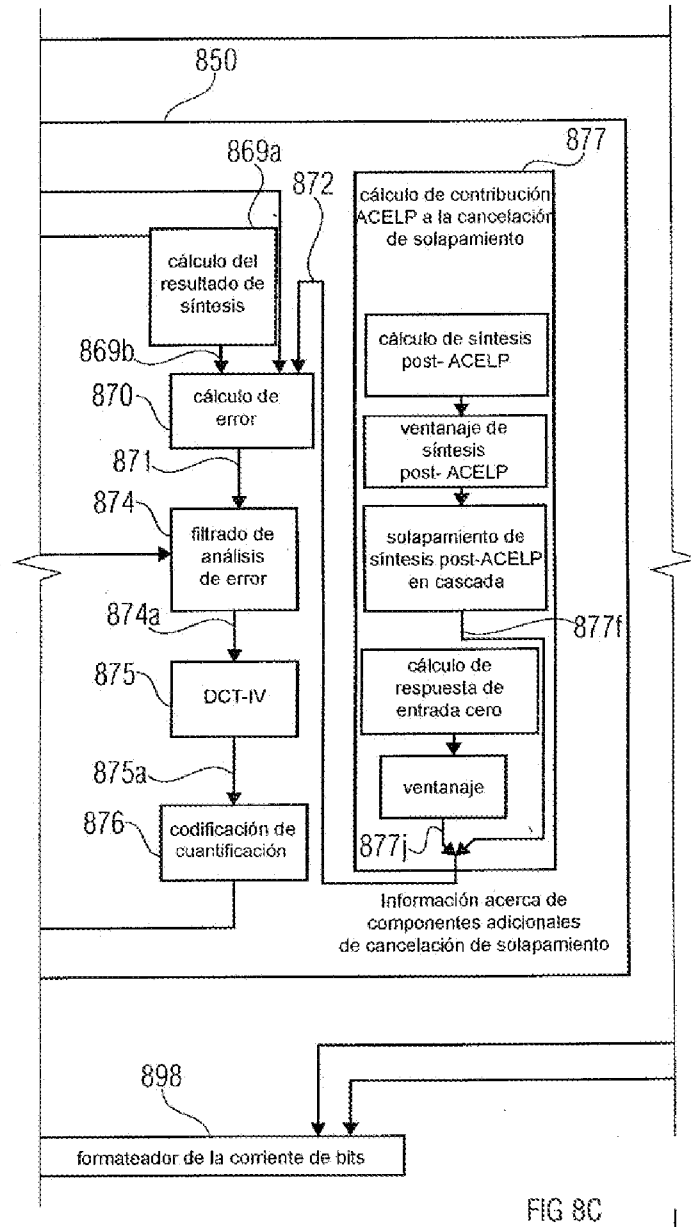


FIG 8C

FIG 8			
FIG 8A	FIG 8B	FIG 8C	FIG 8D

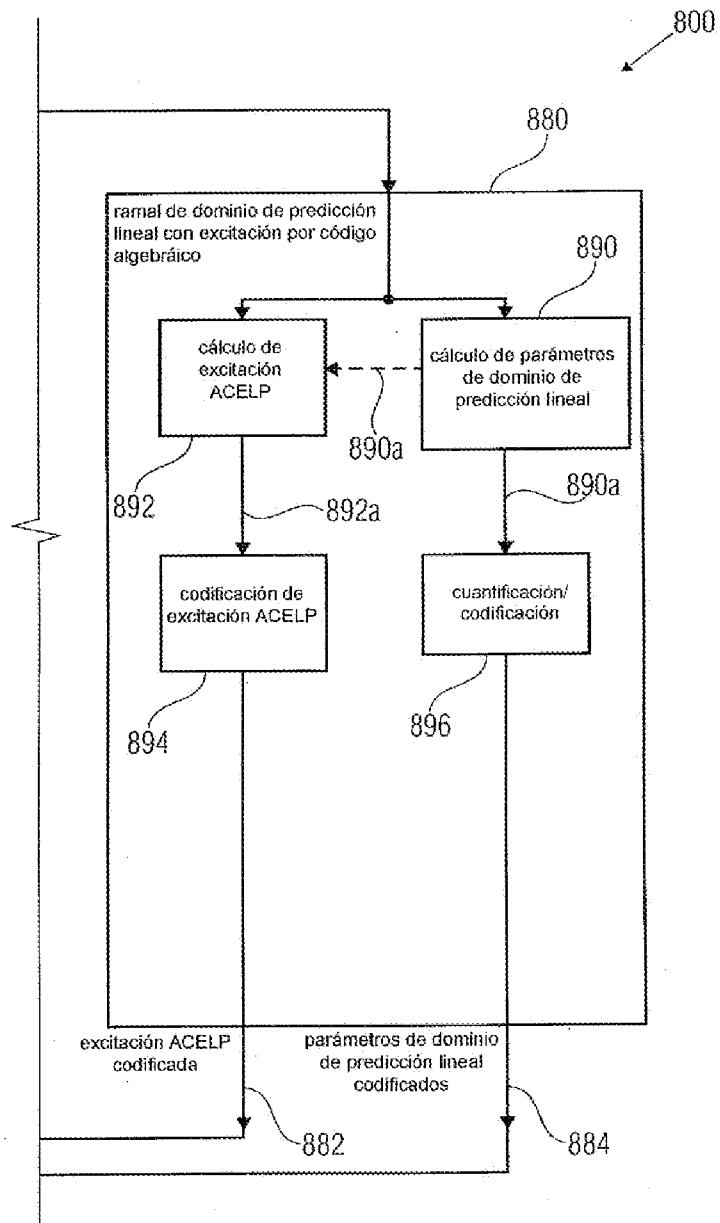
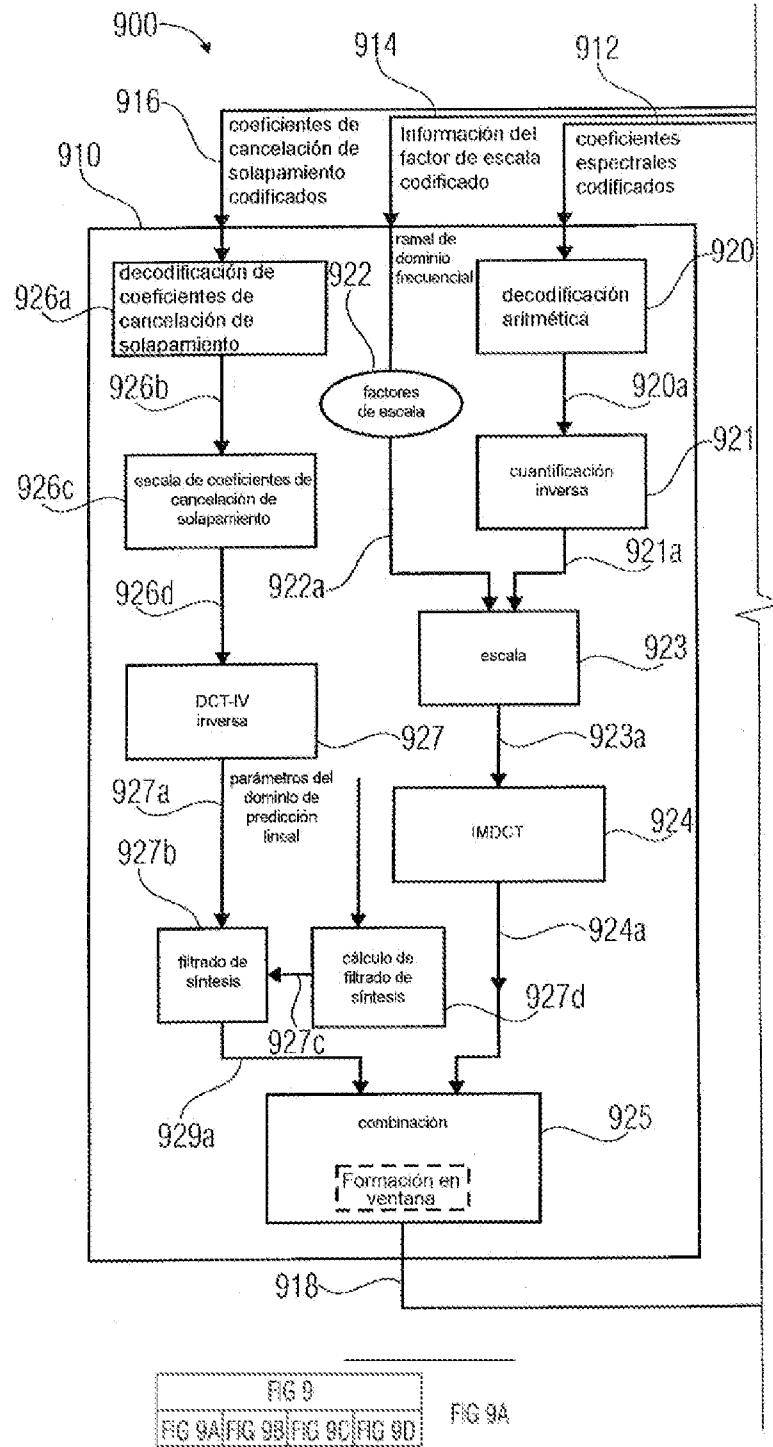


FIG 80

FIG 8				
FIG 8A	FIG 8B	FIG 8C	FIG 8D	FIG 8E



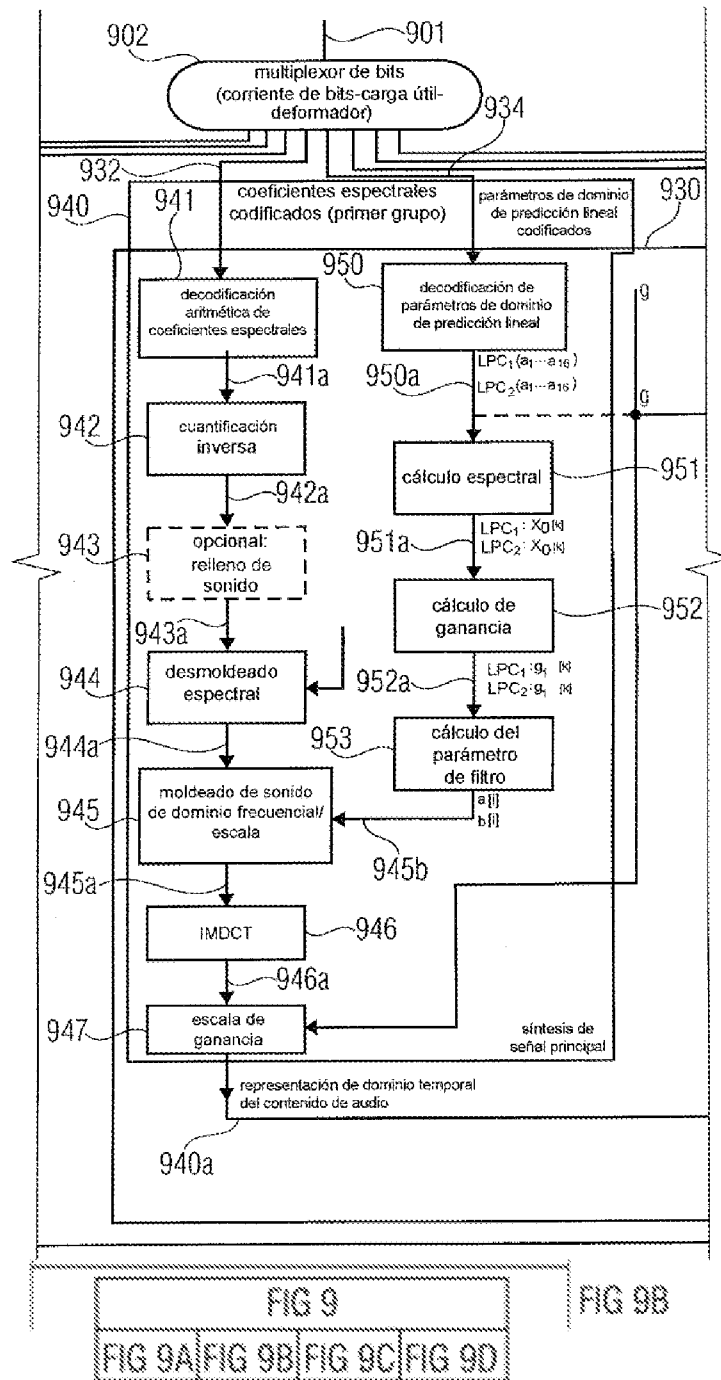


FIG 9

FIG 9B

FIG 9A | FIG 9B | FIG 9C | FIG 9D

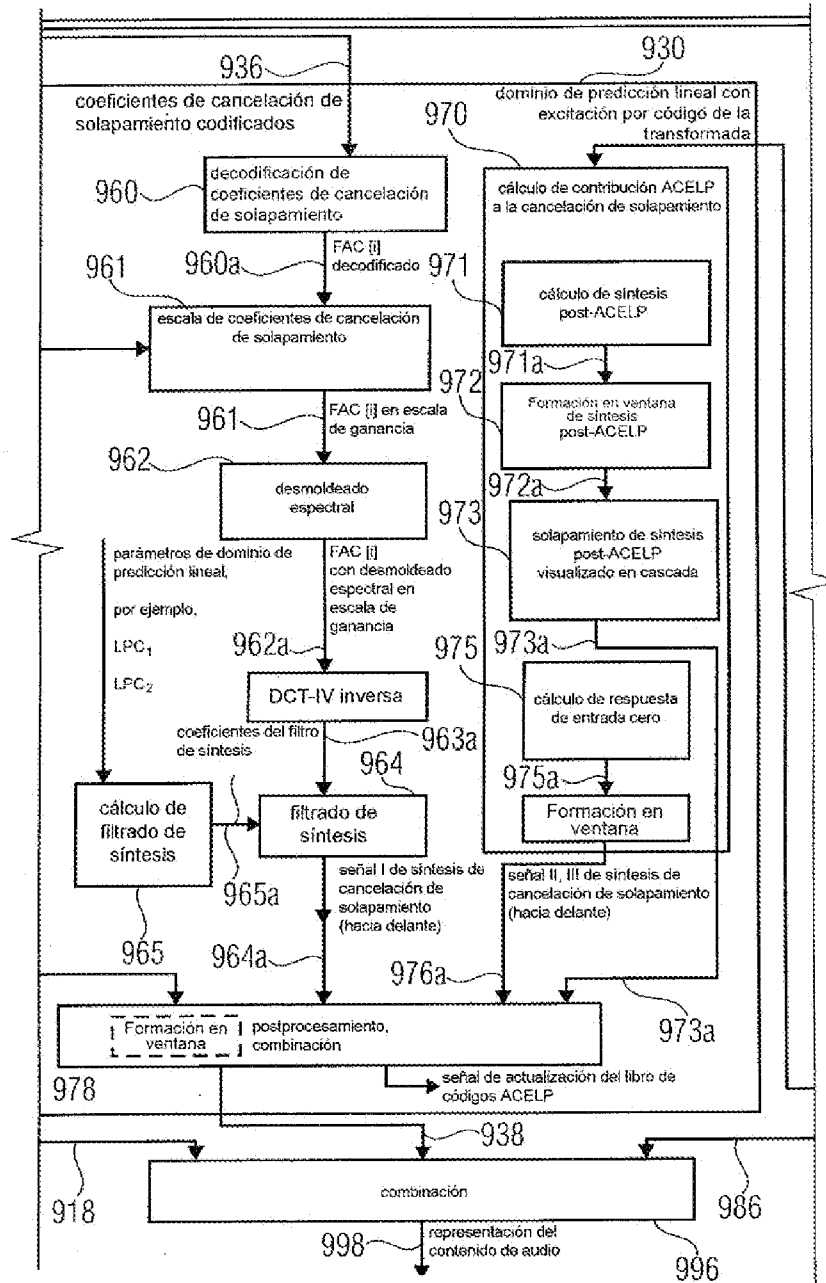


FIG 9  
FIG 9A FIG 9B FIG 9C FIG 9D

FIG 9C

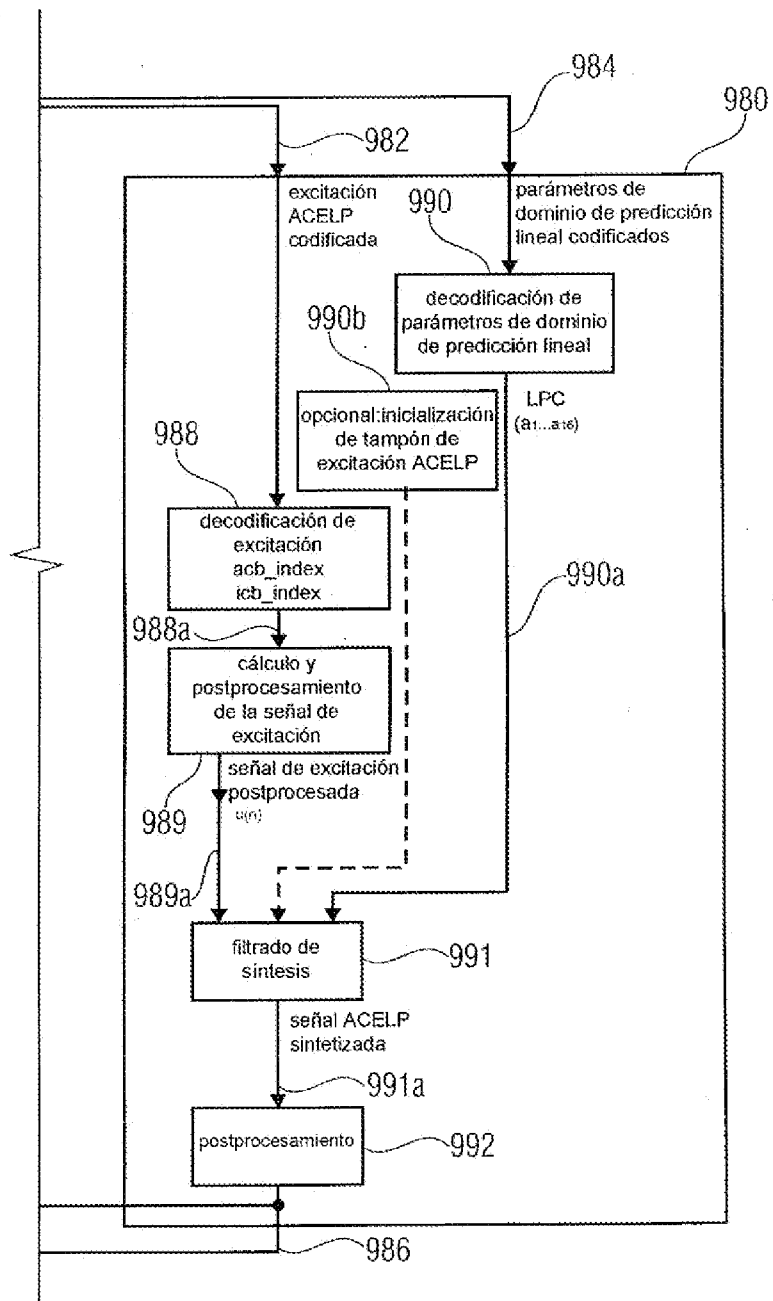


FIG 9  
FIG 9A | FIG 9B | FIG 9C | FIG 9D

FIG 9D

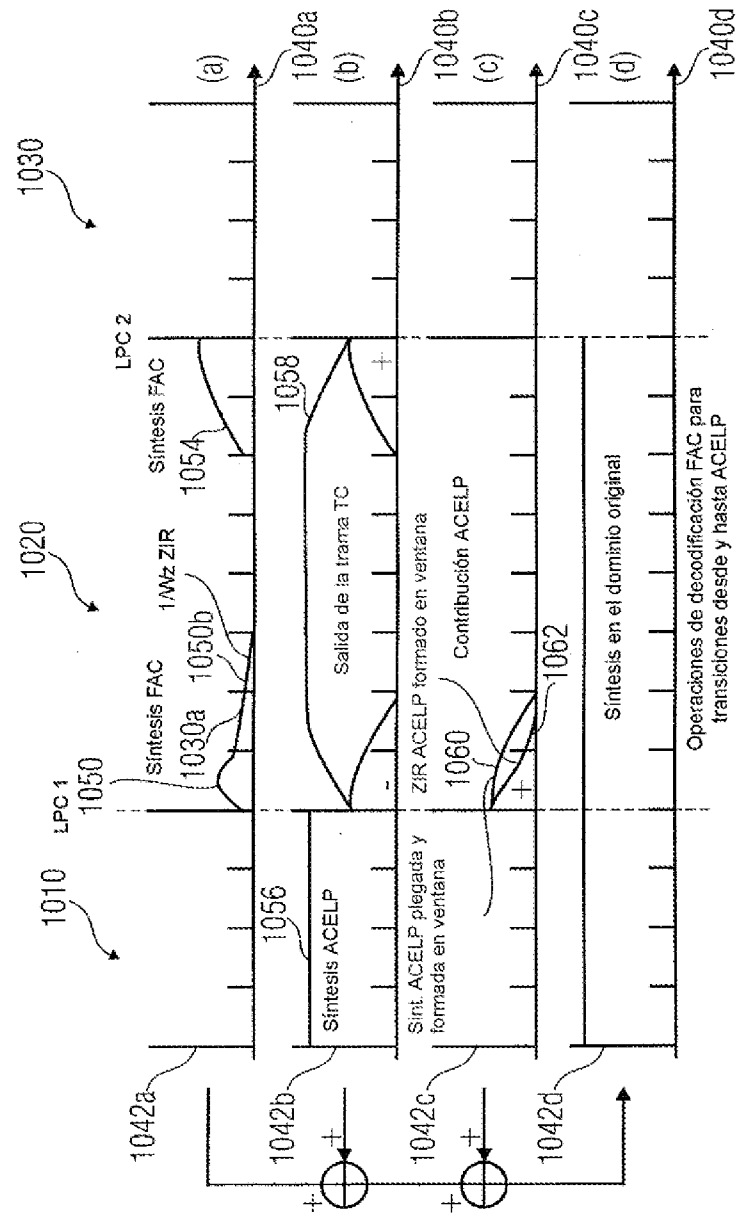
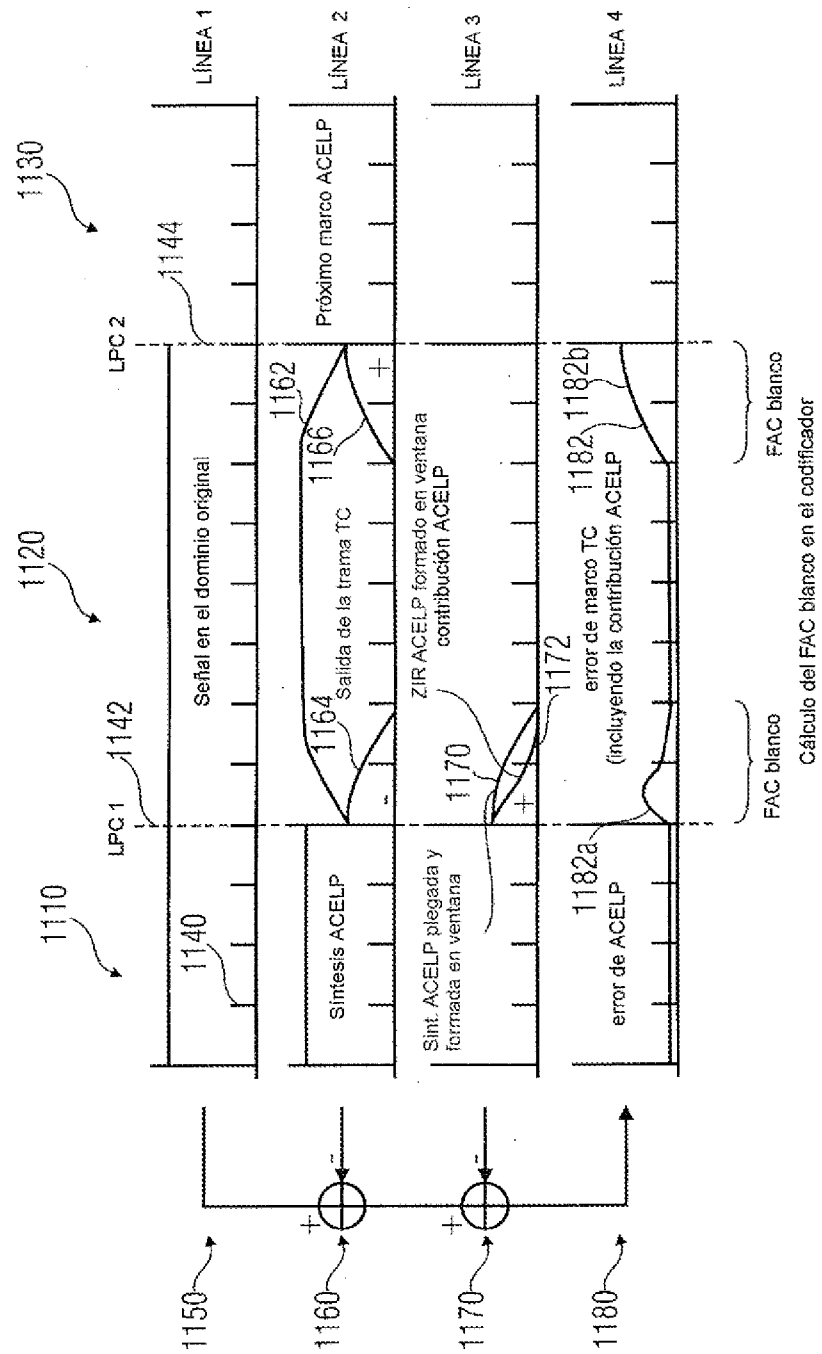
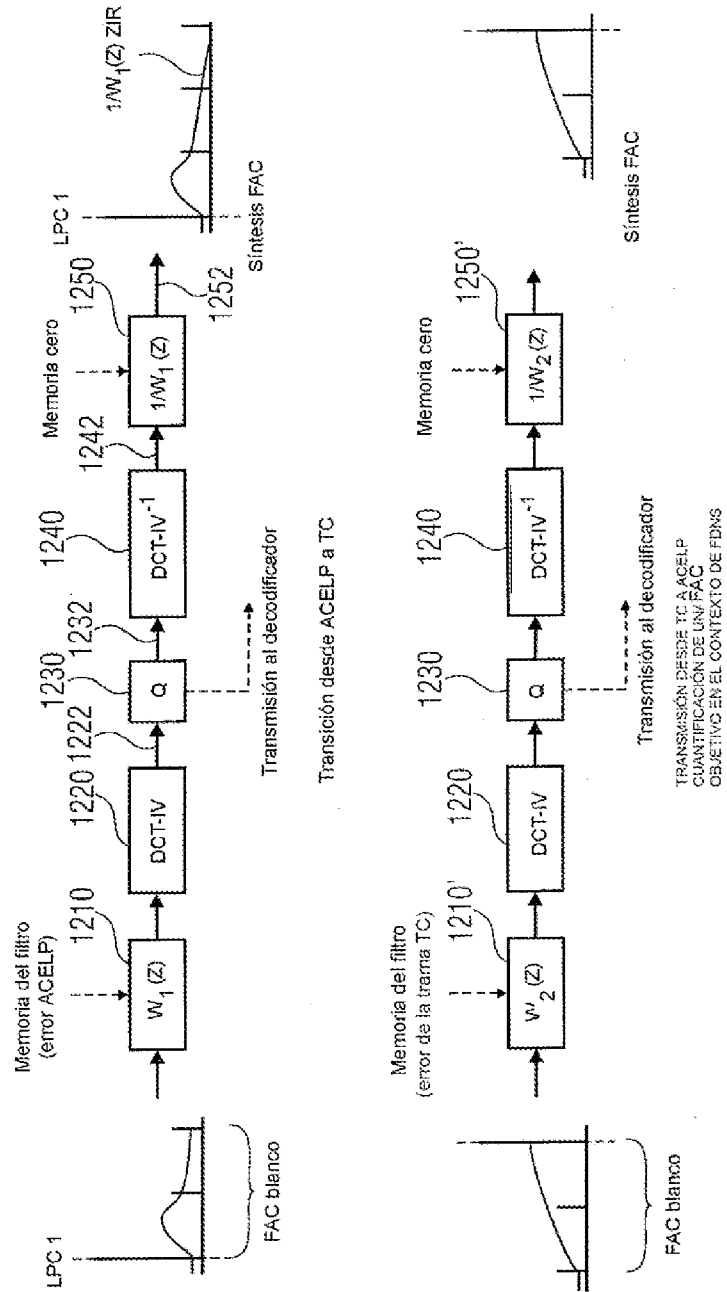


FIG. 10



**FIG. 11**

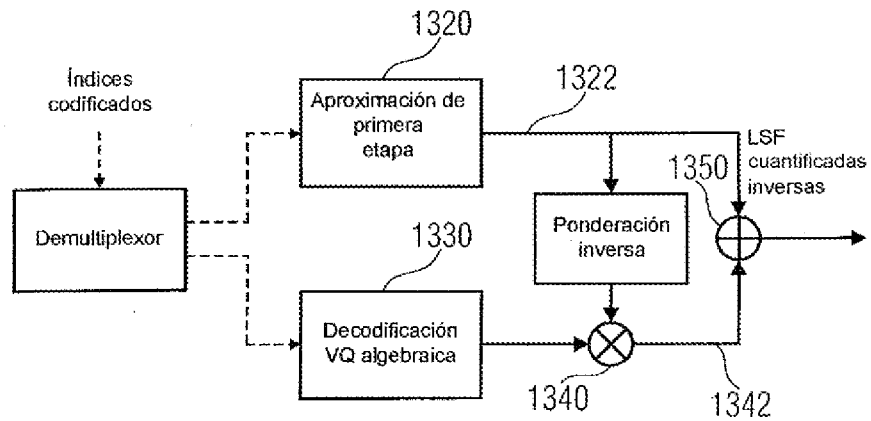




CONDICIONES PARA LA PRESENCIA DE UN FILTRO  
LPC DADO EN UNA CORRIENTE DE BITS

filtro de LPC	presente si
LPC 0	first_lpd_flag=1
LPC 1	$\text{mod}[0] < 2$
LPC 2	$\text{mod}[2] < 3$
LPC 3	$\text{mod}[2] < 2$
LPC 4	siempre

TABLA 1



Principio del cuantificador inverso de LPC algebraica ponderada

FIG. 13

POSIBLES MODOS ABSOLUTOS Y RELATIVOS  
DE CUANTIFICACIÓN Y SEÑALIZACIÓN DE LA CORRIENTE  
DE BITS CORRESPONDIENTE DE MODE\_LPC

filtro	modo de cuantificación	mode_lpc	cód. binario
LPC4	absoluto	0	(ninguno)
LPC0	absoluto	0	0
	relativo LPC4	1	1
LPC2	absoluto	0	0
	relativo LPC4	1	1
LPC1	absoluto	0	10
	relativo $(LPC0 + LPC2)/2$ (Nota 1)	1	11
	relativo LPC2	2	0
LPC3	absoluto	0	10
	relativo $(LPC2 + LPC4)/2$	1	0
	relativo LPC2	2	110
	relativo LPC4	3	111

Nota 1: en este modo, no existe cuantificación AVQ de segunda fase

**TABLA 2**

MODOS DE CODIFICACIÓN PARA LOS  
NÚMEROS DEL LIBRO DE CÓDIGOS  $n_k$

filtro	modo de cuantificación	modo $n_k$
LPC4	absoluto	0
LPC0	absoluto	0
	relativo LPC4	3
LPC2	absoluto	0
	relativo LPC4	3
LPC1	absoluto	0
	relativo $(LPC0 + LPC2)/2$	1
	relativo LPC2	2
LPC3	absoluto	0
	relativo $(LPC2 + LPC4)/2$	1
	relativo LPC2	2
	relativo LPC4	2

TABLA 3

FACTOR DE NORMALIZACIÓN DE W  
PARA LA CUANTIFICACIÓN AVQ

filtro	modo de cuantificación	W
LPC4	absoluto	60
LPC0	absoluto	60
	relativo LPC4	63
LPC2	absoluto	60
	relativo LPC4	63
LPC1	absoluto	60
	relativo $(LPC0 + LPC2)/2$	65
	relativo LPC2	64
LPC3	absoluto	60
	relativo $(LPC2 + LPC4)/2$	65
	relativo LPC2	64
	relativo LPC4	64

TABLA 4

ENERGÍA DE EXCITACIÓN PROMEDIO  $\bar{E}$ 

mean_energy	energía de excitación promedio decodificada, $\bar{E}$
0	18 dB
1	30 dB
2	42 dB
3	54 dB

TABLA 5

NÚMERO DE COEFICIENTES ESPECTRALES  
COMO FUNCION DE MOD[]

valor de mod[x]	número lg de coeficientes espectrales	ZL	L	M	R	ZR
1	256	0	256	0	256	0
2	512	128	256	256	256	128
3	1024	384	256	768	256	384

TABLA 6

MAPEADO DE MODOS DE CODIFICACIÓN PARA LPD\_CHANNEL\_STREAM()

modo de lpd	significado de los bits en el "modo de campo de bits"					resto de mod[] enties
	bit 4	bit 3	bit 2	bit 1	bit 0	
0..15	0	mod[3]	mod[2]	mod[1]	mod[0]	
16..19	1	0	0	mod[3]	mod[2]	mod[1] = 2 mod[0] = 2
20..23	1	0	1	mod[1]	mod[0]	mod[3] = 2 mod[2] = 2
24	1	1	0	0	0	mod[3] = 2 mod[2] = 2 mod[1] = 2 mod[0] = 2
25	1	1	0	0	1	mod[3] = 3 mod[2] = 3 mod[1] = 3 mod[0] = 3
26..31						reservado

TABLA 7

MODOS DE CODIFICACIÓN INDICADOS CON 'MOD[]'

valor de mod [X]	modo de codificación en trama	elemento de corriente de bit
0	ACELP	acelp_coding()
1	una trama de TCX	tcx_coding()
2	TCX que cubre la mitad de una supertrama	tcx_coding()
3	TCX que cubre toda una supertrama	tcx_coding()

TABLA 8

Syntaxis de fd\_channel\_stream()

fd_channel_stream(common_window,	core_mode_last)
{	
global_gain;	8 uimbsf
scale_factor_data();	
ac_spectral_data();	
if(core_mode_last == 1 && last_lpd_mode == 0) {	
lac_data(1)	1 uimbsf
}	
}	

FIG. 14



Sintaxis	Sintaxis de lpd_channel_stream()	Nº de bits	Mnemonic
<pre> lpd_channel_stream(core_mode_last) {     acelp_core_mode     lpd_mode      first_tox_flag=TRUE;     k=0;     si(first_lpd_flag){last_lpd_mode = -1}     mientras(k&lt;4){         si(last_lpd_mode==0 &amp;&amp; mod[k]&gt;0               (last_lpd_mode&gt;0 &amp;&amp; mod[k]==0)){             fac_data(0)         }         si(mod[k]==0){             acelp_coding(acelp_core_mode);             last_lpd_mode=0;             k+=1         }     } } </pre>	<pre> 3  uimbsf 5  uimbsf,     NOTA 1      NOTA 2 </pre>	<p>1510</p>	

FIG. 15 A

FIG 15
FIG 15A FIG 15B

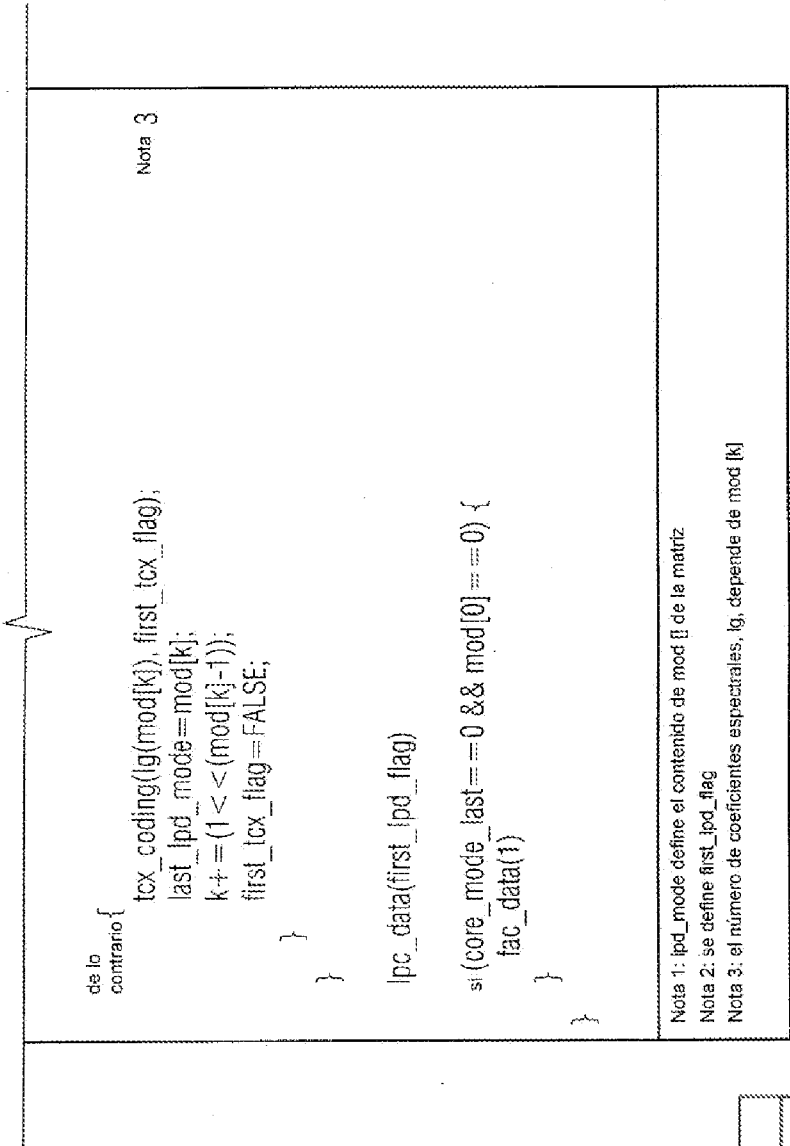


FIG 15

FIG 15A/FIG 15B

FIG. 15 B

Sintaxis de fac\_data ( )

Sintaxis	Nº de bits	Mnemonic
<pre> fac_data(useGain) {     si (useGain { } 1610         fac_gain     }     para (i=0; i&lt;fac_length/8;i++) {         nq[i]         FAC[i]     } } </pre>	<p>7</p> <p>1..n</p> <p>4*nq[i]</p>	<p>uimsbf</p> <p>vlclbf, Nota 1 uimsbf</p>
<p>Nota 1: este valor se codifica utilizando un código unario modificado, donde qn=0 está representado por un bit "0" y cualquier valor qn mayor o igual a 2 está representado por bits qn-1 "1", seguido por un bit de detención "0". Observe que qn=1 no se puede señalar, porque el libro de códigos Q1 no está definido.</p>		

FIG. 16