

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 693 229**

51 Int. Cl.:

G10L 19/08 (2006.01)

G10L 19/20 (2006.01)

G10L 19/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.10.2011 E 11835383 (8)**

97 Fecha y número de publicación de la concesión europea: **01.08.2018 EP 2633521**

54 Título: **Codificación de señales de audio genérico a bajas tasas de bits y bajo retardo**

30 Prioridad:

25.10.2010 US 406379 P

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

10.12.2018

73 Titular/es:

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

72 Inventor/es:

**VAILLANCOURT, TOMMY y
JELINEK, MILAN**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 693 229 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

Codificación de señales de audio genérico a bajas tasas de bits y bajo retardo

Sector técnico

5 La presente invención se refiere a dispositivos y procedimientos de codificación mixta en el dominio de tiempo/dominio de frecuencia para codificar una señal de sonido de entrada, y a un codificador y un descodificador correspondientes que utilizan estos dispositivos y procedimientos de codificación mixta en el dominio de tiempo/dominio de frecuencia.

Antecedentes

10 Un códec conversacional del estado de la técnica puede representar con muy buena calidad una señal de voz limpia con una tasa de bits de aproximadamente 8 kbps y acercarse a la transparencia a una tasa de bits de 16 kbps. Sin embargo, a tasas de bits por debajo de 16 kbps, los códecs conversacionales con bajo retardo de procesamiento, que la mayor parte de las veces codifican la señal de voz de entrada en el dominio de tiempo, no son adecuados para señales de audio genérico, tales como música y voz reverberante. Para superar este inconveniente, se han introducido códecs conmutados, básicamente utilizando el enfoque del dominio de tiempo para codificar señales de
15 entrada dominadas por voz y un enfoque del dominio de frecuencia para codificar señales de audio genérico. Sin embargo, dichas soluciones conmutadas requieren habitualmente un mayor retardo de procesamiento, necesario tanto para la clasificación voz-música como para la transformada al dominio de frecuencia.

20 El documento "High quality coding of wideband audio signals using transform coded excitation (TCX)", Lefebvre et al., Procedimientos de ICASSP '94, propone un enfoque que utiliza una combinación de técnicas del dominio de tiempo y el dominio de frecuencia. El enfoque propuesto utiliza un modelo de síntesis similar al de los codificadores de predicción lineal tales como CELP. Sin embargo, en el codificador, el análisis por síntesis de alta complejidad se evita mediante cuantificación directa de la denominada señal objetivo (la señal ponderada perceptualmente con correlaciones de tono y sonido de filtro eliminadas) en el dominio de frecuencia. La excitación novedosa se obtiene en el descodificador mediante filtrado inverso de la señal objetivo cuantificada.

25 Para superar el inconveniente anterior, se propone un modelo de dominio de tiempo y dominio de frecuencia más unificado.

Compendio

30 La presente invención se refiere a un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia para codificar una señal de sonido de entrada, que comprende: un calculador de una contribución de excitación del dominio de tiempo, en respuesta a la señal de sonido de entrada; un calculador de una frecuencia de corte para la contribución de excitación del dominio de tiempo en respuesta a la señal de sonido de entrada; un filtro sensible a la frecuencia de corte para ajustar una extensión de frecuencia de la contribución de excitación del dominio de tiempo; un calculador de una contribución de excitación del dominio de frecuencia en respuesta a la señal de sonido de
35 entrada; y un sumador de la contribución de excitación del dominio de tiempo filtrada y la contribución de excitación del dominio de frecuencia en el dominio de frecuencia para formar una excitación mixta en el dominio de tiempo/dominio de frecuencia que constituye una versión codificada de la señal de sonido de entrada.

40 La presente invención se refiere asimismo a un codificador que utiliza un modelo de dominio de tiempo y dominio de frecuencia, que comprende: un clasificador de una señal de sonido de entrada como de voz o no de voz; un codificador solamente del dominio de tiempo; el dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia descrito anteriormente; y un selector de uno del codificador solamente del dominio de tiempo y el dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia, para codificar la señal de sonido de entrada en función de la clasificación de la señal de sonido de entrada.

45 La presente invención se refiere además a un descodificador para descodificar una señal de sonido codificada utilizando el dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia que se ha descrito anteriormente, que comprende: un convertidor de la excitación mixta en el dominio de tiempo/dominio de frecuencia al dominio de tiempo; y un filtro de síntesis para sintetizar la señal de sonido en respuesta a la excitación mixta en el dominio de tiempo/dominio de frecuencia convertida al dominio de tiempo.

50 La presente invención se refiere asimismo a un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia para codificar una señal de sonido de entrada, que comprende: calcular una contribución de excitación del dominio de tiempo en respuesta a la señal de sonido de entrada; calcular una frecuencia de corte para la contribución de excitación del dominio de tiempo en respuesta a la señal de sonido de entrada; en respuesta a la frecuencia de corte, ajustar una extensión de frecuencia de la contribución de excitación del dominio de tiempo; calcular una contribución de excitación del dominio de frecuencia en respuesta a la señal de sonido de entrada; y sumar la contribución de excitación del dominio de tiempo ajustada y la contribución de excitación del dominio de
55 frecuencia en el dominio de frecuencia para formar una excitación mixta en el dominio de tiempo/dominio de frecuencia que constituye una versión codificada de la señal de sonido de entrada.

En la presente invención, se describe además un procedimiento de codificación que utiliza un modelo de dominio de tiempo y dominio de frecuencia, que comprende: clasificar una señal de sonido de entrada como de voz o no de voz; proporcionar un procedimiento de codificación solamente del dominio de tiempo; proporcionar el procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia descrito anteriormente, y seleccionar uno del procedimiento de codificación solamente del dominio de tiempo y el procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia para codificar la señal de sonido de entrada en función de la clasificación de la señal de sonido de entrada.

En la presente invención, se describe además un procedimiento de descodificación de una señal de sonido codificada utilizando el procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia que se ha descrito anteriormente, que comprende: convertir la excitación mixta en el dominio de tiempo/dominio de frecuencia al dominio de tiempo; y sintetizar la señal de sonido mediante un filtro de síntesis en respuesta a la excitación mixta en el dominio de tiempo/dominio de frecuencia convertida al dominio de tiempo.

Las anteriores y otras características resultarán más evidentes tras la lectura de la siguiente descripción no restrictiva de una realización ilustrativa del modelo de dominio de tiempo y dominio de frecuencia propuesto, proporcionada solamente a modo de ejemplo, haciendo referencia a los dibujos adjuntos.

Breve descripción de los dibujos

En los dibujos adjuntos:

la figura 1 es un diagrama esquemático de bloques que muestra una visión general de un codificador CELP (Code-Excited Linear Prediction, predicción lineal con excitación por código) mejorado, por ejemplo un codificador ACELP (Algebraic Code-Excited Linear Prediction; predicción lineal con excitación por código algebraico);

la figura 2 es un diagrama esquemático de bloques de una estructura más detallada del codificador CELP mejorado de la figura 1;

la figura 3 es un diagrama esquemático de bloques de una visión general de un calculador de frecuencia de corte;

la figura 4 es un diagrama esquemático de bloques de una estructura más detallada del calculador de frecuencia de corte de la figura 3;

la figura 5 es un diagrama esquemático de bloques de una visión general de un cuantificador de frecuencia; y

la figura 6 es un diagrama esquemático de bloques de una estructura más detallada del cuantificador de frecuencia de la figura 5.

Descripción detallada

El modelo más unificado de dominio de tiempo y dominio de frecuencia propuesto puede mejorar la calidad de síntesis para señales de audio genérico tales como, por ejemplo, música y/o voz reverberante, sin aumentar el retardo de procesamiento y la tasa de bits. Este modelo actúa, por ejemplo, en un dominio de residuos de predicción lineal (LP, Linear Prediction), donde los bits disponibles son asignados dinámicamente entre un libro de códigos adaptativo, uno o varios libros de códigos fijos (por ejemplo, un libro de códigos algebraico, un libro de códigos gaussiano, etc.) y un modo de codificación en el dominio de frecuencia, dependiendo de las características de la señal de entrada.

Para conseguir un códec conversacional de bajo retardo de procesamiento y baja tasa de bits, que mejore la calidad de síntesis de las señales de audio genérico, tales como música y/o voz reverberante, se puede integrar lo mejor posible un modo de codificación en el dominio de frecuencia en el modo de codificación en el dominio de tiempo CELP (predicción lineal con excitación por código). Con este propósito, el modo de codificación en el dominio de frecuencia utiliza, por ejemplo, una transformada de frecuencia realizada en el dominio de residuos LP. Esto permite una conmutación casi sin artefactos de una trama, por ejemplo una trama de 20 ms, a otra. Asimismo, la integración de los dos (2) modos de codificación es lo suficientemente estrecha como para permitir una reasignación dinámica del presupuesto de bits a otro modo de codificación si se determina que el modo de codificación actual no es lo suficientemente eficiente.

Una característica del modelo más unificado de dominio de tiempo y dominio de frecuencia propuesto, es el soporte temporal variable del componente del dominio de tiempo, que varía desde un cuarto de trama hasta una trama completa en una base trama a trama, y se denominará subtrama. Como un ejemplo ilustrativo, una trama representa 20 ms de señal de entrada. Esto corresponde a 320 muestras si la frecuencia de muestreo interna del códec es 16 kHz, o a 256 muestras por trama si la frecuencia de muestreo interna del códec es de 12,8 kHz. Entonces un cuarto de trama (la subtrama) representa 64 u 80 muestras en función de la frecuencia de muestreo interna del códec. En la siguiente realización ilustrativa, la frecuencia de muestreo interna del códec es de 12,8 kHz proporcionando una longitud de trama de 256 muestras. El soporte temporal variable hace posible capturar eventos temporales principales con una tasa de bits mínima para crear una contribución básica de excitación del dominio de tiempo. A una tasa de bits muy baja, el soporte temporal es normalmente toda la trama. En ese caso, la contribución del

dominio de tiempo a la señal de excitación se compone solamente del libro de códigos adaptativo, y la correspondiente información de tono con la correspondiente ganancia se transmiten una vez por cada trama. Cuando está disponible más de una tasa de bits, es posible capturar más eventos temporales acortando el soporte temporal (y aumentando la tasa de bits asignada al modo de codificación en el dominio de tiempo). Al final, cuando el soporte temporal es lo suficientemente corto (de un cuarto de trama), y la tasa de bits disponible es lo suficientemente alta, la contribución del dominio de tiempo puede incluir la contribución del libro de códigos adaptativo, una contribución de libro de códigos fijo, o ambas, con las correspondientes ganancias. Los parámetros que describen los índices de libro de códigos y las ganancias se transmiten a continuación para cada subtrama.

A una baja tasa de bits, los códec conversacionales no pueden codificar adecuadamente las frecuencias superiores. Esto provoca una degradación importante de la calidad de la síntesis cuando la señal de entrada incluye música y/o voz reverberante. Para resolver este problema, se añade una característica para calcular la eficiencia de la contribución de excitación del dominio de tiempo. En algunos casos, la contribución de excitación del dominio de tiempo no es valiosa para cualesquiera tasa de bits de entrada y soporte temporal de las tramas. En estos casos, todos los bits son reasignados a la siguiente etapa de codificación en el dominio de frecuencia. Pero la mayor parte del tiempo, la contribución de excitación del dominio de tiempo es valiosa solamente hasta una determinada frecuencia (la frecuencia de corte). En estos casos, la contribución de excitación del dominio de tiempo se filtra por encima de la frecuencia de corte. La operación de filtrado permite mantener información valiosa codificada con la contribución de excitación del dominio de tiempo y eliminar la información no valiosa por encima de la frecuencia de corte. En una realización ilustrativa, el filtrado se lleva a cabo en el dominio de frecuencia ajustando a cero las cajas de frecuencia por encima de una determinada frecuencia.

El soporte temporal variable en combinación con la frecuencia de corte variable hace muy dinámica la asignación de bits dentro del modelo integrado de dominio de tiempo y dominio de frecuencia. La tasa de bits después de la cuantificación del filtro de LP se puede asignar íntegramente al dominio de tiempo o íntegramente al dominio de frecuencia, o a cualquier situación intermedia. La asignación de la tasa de bits entre los dominios de tiempo y de frecuencia se realiza como una función del número de subtramas utilizadas para la contribución del dominio de tiempo, del presupuesto de bits disponible y de la frecuencia de corte calculada.

Para crear una excitación total que se ajuste más eficientemente al residuo de entrada, se aplica el modo de codificación en el dominio de frecuencia. Una característica de la presente invención es que la codificación en el dominio de frecuencia se lleva a cabo sobre un vector que contiene la diferencia entre una representación en frecuencia (transformada de frecuencia) del residuo LP de entrada y una representación en frecuencia (transformada de frecuencia) de la contribución filtrada de excitación del dominio de tiempo hasta la frecuencia de corte, y que contiene la representación en frecuencia (transformada de frecuencia) del propio residuo LP de entrada por encima de dicha frecuencia de corte. Se introduce una transición espectral suave entre ambos segmentos inmediatamente por encima de la frecuencia de corte. En otras palabras, en primer lugar la parte de alta frecuencia de la representación en frecuencia de la contribución de excitación del dominio de tiempo se reduce a cero. Se introduce una zona de transición entre la parte no modificada del espectro y la parte reducida a cero del espectro, inmediatamente por encima de la frecuencia de corte, para garantizar una transición suave entre ambas partes del espectro. Este espectro modificado de la contribución de excitación del dominio de tiempo se resta a continuación de la representación en frecuencia del residuo LP de entrada. El espectro resultante corresponde por lo tanto a la diferencia de ambos espectros por debajo de la frecuencia de corte, y a la representación en frecuencia del residuo LP por encima de la misma, con una cierta zona de transición. La frecuencia de corte, tal como se ha mencionado anteriormente, puede variar de una trama a otra.

Cualquiera que sea el procedimiento de cuantificación de frecuencia (modo de codificación en el dominio de frecuencia) escogido, existe siempre una posibilidad de pre-eco, especialmente con ventanas largas. En esta técnica, las ventanas utilizadas son ventanas cuadradas, de tal modo que la longitud de ventana extra comparada con la señal codificada es cero (0), es decir, no se utiliza suma en solapamiento. Aunque esto corresponde a la mejor ventana para reducir cualquier pre-eco potencial, puede seguir siendo audible algún pre-eco en ataques temporales. Existen muchas técnicas para resolver dicho problema de pre-eco, pero la presente invención propone una característica simple para cancelar este problema de pre-eco. Esta característica se basa en un modo de codificación en el dominio de tiempo sin memoria que se obtiene a partir del "modo de transición" de la recomendación de ITU-T G.718; referencia [recomendación ITU-T G.718 "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s", junio de 2008, sección 6.8.1.4 y sección 6.8.4.2]. La idea tras esta característica es sacar partido del hecho de que el modelo más unificado de dominio de tiempo y dominio de frecuencia propuesto está integrado en el dominio de residuos LP, lo que permite casi todo el tiempo una conmutación sin artefactos. Cuando una señal se considera audio genérico (música y/o voz reverberante) y cuando se detecta un ataque temporal en una trama, entonces esta trama solamente se codifica con este modo de codificación en el dominio de tiempo sin memoria especial. Este modo se encargará del ataque temporal, evitando de ese modo el pre-eco que podría ser introducido con la codificación en el dominio de frecuencia de dicha trama.

Realización ilustrativa

En el modelo más unificado de dominio de tiempo y dominio de frecuencia propuesto, el libro de códigos adaptativo mencionado anteriormente, uno o varios libros de códigos fijos (por ejemplo, un libro de códigos algebraico, un libro de códigos gaussiano, etc.), es decir, los denominados libros de códigos del dominio de tiempo, y la cuantificación en el dominio de frecuencia (modo de codificación en el dominio de frecuencia se pueden ver como una biblioteca de libros de códigos, y los bits se pueden distribuir entre todos los libros de códigos disponibles, o un subconjunto de los mismos. Esto significa, por ejemplo, que si la señal de sonido de entrada es una voz limpia, todos los bits serán asignados al modo de codificación en el dominio de tiempo, básicamente reduciendo la codificación al esquema CELP heredado. Por otra parte, para algunos segmentos de música, en ocasiones todos los bits asignados para codificar el residuo LP de entrada están mejor empleados en el dominio de frecuencia, por ejemplo en un dominio de transformada.

Tal como se ha indicado en la descripción anterior, el soporte temporal para los modos de codificación en el dominio de tiempo y el dominio de frecuencia no tiene por qué ser el mismo. Mientras los bits invertidos en los diferentes procedimientos de cuantificación en el dominio de tiempo (búsquedas de libros de códigos adaptativos y algebraicos) se distribuyen normalmente en un esquema por subtrama (habitualmente un cuarto de trama, o 5 ms de soporte temporal), los bits asignados al modo de codificación en el dominio de frecuencia se distribuyen en un esquema por trama (habitualmente, 20 ms de soporte temporal) para mejorar la resolución de la frecuencia.

El presupuesto de bits asignado al modo de codificación CELP en el dominio de tiempo puede asimismo controlarse dinámicamente en función de la señal de sonido de entrada. En algunos casos, el presupuesto de bits asignado al modo de codificación CELP en el dominio de tiempo puede ser cero, lo que significa de hecho que todo el presupuesto de bits se atribuye al modo de codificación en el dominio de frecuencia. La elección de trabajar en el dominio de residuos LP para los enfoques tanto del dominio de tiempo como del dominio de frecuencia tiene dos (2) beneficios principales. En primer lugar, es compatible con el modo de codificación CELP, que ha mostrado ser eficiente en codificación de señales de voz. Por consiguiente, no se introduce ningún artefacto debido a la conmutación entre los dos tipos de modos de codificación. En segundo lugar, una menor dinámica del residuo LP con respecto a la señal de sonido de entrada original, y su relativa planitud, facilitan la utilización de una ventana cuadrada para las transformadas de frecuencia, permitiendo por lo tanto la utilización de una ventana sin solapamiento.

En un ejemplo no limitativo donde la frecuencia de muestreo interna del códec es de 12,8 kHz (lo que significa 256 muestras por trama), de manera similar a la recomendación G.718 de ITU-T, la longitud de las subtramas utilizadas en el modo de codificación CELP del dominio de tiempo puede variar desde un típico 1/4 de longitud de trama (5 ms) a media trama (10 ms), o una longitud de trama completa (20 ms). La decisión de la longitud de la subtrama se basa en la tasa de bits disponible y en un análisis de la señal de sonido de entrada, en particular de las dinámicas espectrales de la señal de sonido de entrada. La decisión de la longitud de la subtrama se puede llevar a cabo en un modo en bucle cerrado. Para ahorrar complejidad, es posible asimismo basar la decisión de la longitud de la subtrama en un modo en bucle abierto. La longitud de la subtrama se puede modificar de una trama a otra.

Una vez se ha elegido la longitud de las subtramas en una trama particular, se lleva a cabo un análisis estándar de tono en bucle cerrado y se selecciona la primera contribución a la señal de excitación desde el libro de códigos adaptativo. A continuación, dependiendo del presupuesto de bits disponible y de las características de la señal de sonido de entrada (por ejemplo, en el caso de una señal de voz de entrada), se puede añadir una segunda contribución desde uno o varios libros de códigos fijos antes de la codificación en el dominio de transformada. La excitación resultante se denominará la contribución de excitación del dominio de tiempo. Por otra parte, a tasas de bits muy bajas y en caso de audio genérico, a menudo es mejor omitir la etapa de libro de códigos fijo y utilizar todos los bits restantes para el modo de codificación en el dominio de transformada. El modo de codificación en el dominio de transformada puede ser, por ejemplo, un modo de codificación en el dominio de frecuencia. Tal como se ha descrito anteriormente, la longitud de la subtrama puede ser de la longitud de un cuarto de la trama, de la mitad de la trama o de una trama. La contribución de libro de códigos fijo se utiliza solamente si la longitud de la subtrama es igual a un cuarto de la longitud de trama. En caso de que se decida que la longitud de la subtrama tenga una longitud de media trama o de la trama entera, se utiliza entonces solamente la contribución de libro de códigos adaptativo para representar la excitación en el dominio de tiempo, y todos los bits restantes son asignados al modo de codificación en el dominio de frecuencia.

Una vez se ha completado el cálculo de la contribución de excitación del dominio de tiempo, es necesario evaluar y cuantificar su eficiencia. Si la ganancia de la codificación en el dominio de tiempo es muy baja, es más eficiente eliminar por completo la contribución de excitación del dominio de tiempo y utilizar en su lugar todos los bits para el modo de codificación en el dominio de frecuencia. Por otra parte, por ejemplo en el caso de una voz de entrada limpia, el modo de codificación en el dominio de frecuencia no es necesario y todos los bits son asignados al modo de codificación en el dominio de tiempo. Pero a menudo la codificación en el dominio de tiempo es eficiente solamente hasta una determinada frecuencia. Esta frecuencia se denominará la frecuencia de corte de la contribución de excitación del dominio de tiempo. La determinación de dicha frecuencia de corte garantiza que toda la codificación en el dominio de tiempo está contribuyendo a obtener una mejor síntesis final en lugar de trabajar contra la codificación en el dominio de frecuencia.

La frecuencia de corte se estima en el dominio de frecuencia. Para calcular la frecuencia de corte, los espectros tanto del residuo LP como de la contribución codificada del dominio de tiempo se dividen en primer lugar en un número predefinido de bandas de frecuencia. El número de bandas de frecuencia y el número de cajas de frecuencia cubiertas por cada banda de frecuencia pueden variar de una implementación a otra. Para cada una de las bandas de frecuencia, se calcula una correlación normalizada entre la representación en frecuencia de la contribución de excitación del dominio de tiempo y la representación en frecuencia del residuo LP, y la correlación es suavizada entre bandas de frecuencia adyacentes. Las correlaciones por banda se limitan por debajo a 0,5 y se normalizan entre 0 y 1. A continuación se calcula la correlación promedio, como el promedio de las correlaciones para todas las bandas de frecuencia. Con el objetivo de una primera estimación de la frecuencia de corte, a continuación la correlación promedio se escala entre 0 y la mitad de la velocidad de muestreo (correspondiendo la mitad de la velocidad de muestreo al valor de correlación normalizada de 1). A continuación, se encuentra la primera estimación de la frecuencia de corte como el límite superior de la banda de frecuencia que se acerca más a dicho valor. En un ejemplo de implementación, se definen dieciséis (16) bandas de frecuencia a 12,8 kHz para el cálculo de correlación.

Sacando partido de la propiedad psicoacústica del oído humano, la fiabilidad de la estimación de la frecuencia de corte se mejora comparando la posición estimada de la frecuencia del octavo armónico del tono con la frecuencia de corte estimada por el cálculo de correlación. Si esta posición está por encima de la frecuencia de corte estimada por el cálculo de correlación, la frecuencia de corte se modifica en correspondencia con la posición de la frecuencia del octavo armónico del tono. A continuación, el valor final de la frecuencia de corte es cuantificado y transmitido. En un ejemplo de implementación, se utilizan 3 o 4 bits para dicha cuantificación, que proporcionan 8 o 16 posibles frecuencias de corte en función de la tasa de bits.

Una vez se conoce la frecuencia de corte, se lleva a cabo la cuantificación de frecuencia de la contribución de excitación del dominio de frecuencia. En primer lugar, se determina la diferencia entre la representación en frecuencia (transformada de frecuencia) de la contribución de excitación del dominio de tiempo. A continuación se crea un nuevo vector, que consiste en esta diferencia hasta la frecuencia de corte, y una transición suave a la representación en frecuencia del residuo LP de entrada para el resto del espectro. A continuación se aplica una cuantificación de frecuencia a todo el nuevo vector. En un ejemplo de implementación, la cuantificación consiste en codificar el signo y la posición de los pulsos espectrales dominantes (más energéticos). El número de pulsos a cuantificar por banda de frecuencia está relacionado con la tasa de bits disponible para el modo de codificación en el dominio de frecuencia. Si no existen suficientes bits disponibles para cubrir todas las bandas de frecuencia, las bandas restantes se llenan solamente con ruido.

La cuantificación de frecuencia de una banda de frecuencia utilizando el procedimiento de cuantificación descrito en el párrafo anterior no garantiza que se cuantifiquen todas las cajas de frecuencia dentro de esta banda. Esto es especialmente cierto a tasas de bits bajas donde el número de pulsos cuantificados por banda de frecuencia es relativamente bajo. Para impedir la aparición de artefactos audibles debido a estas cajas no cuantificadas, se añade algún ruido para llenar estos huecos. Dado que a tasas de bits bajas los pulsos cuantificados deberían dominar el espectro en lugar de hacerlo el ruido introducido, la amplitud del espectro del ruido corresponde solamente una fracción de la amplitud de los pulsos. La amplitud del ruido añadido en el espectro es mayor cuando el presupuesto de bits disponibles es bajo (permitiendo más ruido) y menor cuando el presupuesto de bits disponible es alto.

En el modo de codificación en el dominio de frecuencia, se calculan las ganancias para cada banda de frecuencia con el fin de ajustar la energía de la señal no cuantificada a la señal cuantificada. Las ganancias son cuantificadas vectorialmente y aplicadas por cada banda a la señal cuantificada. Cuando el codificador cambia su asignación de bits desde el modo de codificación solamente del dominio de tiempo al modo de codificación mixta en el dominio de tiempo/dominio de frecuencia, la energía del espectro de excitación por banda del modo de codificación solamente del dominio de tiempo no se ajusta a la energía del espectro de excitación por banda del modo de codificación mixta en el dominio de tiempo/dominio de frecuencia. Este desajuste de energía puede crear algunos artefactos de conmutación, especialmente a baja tasa de bits. Para reducir cualquier degradación audible creada por esta reasignación de bits, se puede calcular una ganancia a largo plazo para cada banda y se puede aplicar con el fin de corregir la energía de cada banda de frecuencia para unas pocas tramas después de la conmutación del modo de codificación en el dominio de tiempo al modo de codificación mixta en el dominio de tiempo/dominio de frecuencia.

Después de completarse el modo de codificación en el dominio de frecuencia, la excitación total se obtiene sumando la contribución de excitación del dominio de frecuencia a la representación en frecuencia (transformada de frecuencia) de la contribución de excitación del dominio de tiempo, y a continuación la suma de las contribuciones de excitación se transforma de nuevo al dominio de tiempo para formar una excitación total. Finalmente, la señal sintetizada se calcula filtrando la excitación total a través del filtro de síntesis LP. En una realización, si bien las memorias de codificación CELP se actualizan en un esquema por subtramas utilizando solamente la contribución de excitación del dominio de tiempo, se utiliza la excitación total para actualizar estas memorias en los límites de trama. En otra posible implementación, las memorias de codificación CELP se actualizan en un esquema por subtrama y asimismo en los límites de trama utilizando solamente la contribución de excitación del dominio de tiempo. Esto tiene como resultado una estructura integrada donde la señal cuantificada en el dominio de frecuencia constituye una capa de cuantificación superior independiente de la capa CELP central. En este caso particular, se utiliza siempre el

libro de códigos fijo para actualizar el contenido del libro de códigos adaptativo. Sin embargo, el modo de codificación en el dominio de frecuencia puede aplicar a toda la trama. Este enfoque integrado funciona para tasas de bits en torno 12 kbps y superiores.

1) Clasificación del tipo de sonido

5 La figura 1 es un diagrama esquemático de bloques que muestra una visión general de un codificador CELP mejorado 100, por ejemplo un codificador ACELP. Por supuesto, se pueden implementar otros tipos de codificadores CELP utilizando el mismo concepto. La figura 2 es un diagrama esquemático de bloques de una estructura más detallada del codificador CELP mejorado de la figura 100.

10 El codificador CELP 100 comprende un preprocesador 102 (figura 1) para analizar parámetros de la señal de sonido de entrada 101 (figuras 1 y 2). Haciendo referencia a la figura 2, el preprocesador 102 comprende un analizador LP 201 de la señal de sonido de entrada 101, un analizador espectral 202, un analizador de tono en bucle abierto 203 y un clasificador de señal 204. Los analizadores 201 y 202 llevan a cabo el análisis LP y espectral normalmente realizado en codificación CELP, tal como se describe por ejemplo en la recomendación ITU-T G.718, secciones 6.4 y 6.1.4 y, por lo tanto, no se describirán en mayor detalle en la presente invención.

15 El preprocesador 102 realiza un primer nivel de análisis para clasificar la señal de sonido de entrada 101 entre de voz y no de voz (audio genérico (música o voz reverberante)), por ejemplo de una manera similar a la descrita en la referencia [T. Vaillancourt et al., "Inter-tone noise reduction in a low bit rate CELP decoder", Proc. IEEE ICASSP, Taipei, Taiwan, abril de 2009, páginas 4113-16], cuyo contenido íntegro se incorpora como referencia a la presente memoria, o con cualesquiera otros procedimientos fiables de discriminación de voz/no voz.

20 Después de este primer nivel de análisis, el preprocesador 102 lleva a cabo un segundo nivel de análisis de parámetros de la señal de entrada para permitir la utilización de codificación CELP en el dominio de tiempo (sin codificación en el dominio de frecuencia) sobre algunas señales de sonido con fuertes características no de voz, pero que se siguen codificando mejor con un enfoque en el dominio de tiempo. Cuando se produce una variación importante de energía, este segundo nivel de análisis permite que el codificador CELP 100 conmute a un modo de codificación en el dominio de tiempo sin memoria, denominado generalmente modo de transición en la referencia [Eksler, V., y Jelínek, M (2008), "Transition mode coding for source controlled CELP codecs", Actas del IEEE de la Conferencia internacional sobre acústica, voz y procesamiento de señal, marzo-abril, páginas 4001-40043], cuyo contenido íntegro se incorpora como referencia a la presente memoria.

30 Durante este segundo nivel de análisis, el clasificador de señal 204 calcula y utiliza una variación σ_C de una versión suavizada C_{st} de la correlación de tono en bucle abierto procedente del analizador de tonos bucle abierto 203, una energía total de la trama actual E_{tot} y una diferencia entre la energía total de la trama actual y la energía de trama total anterior E_{diff} . En primer lugar, la variación de la correlación de tono en bucle abierto suavizada se calcula como:

$$\sigma_C = \sqrt{\sum_{i=0}^{i=10} \left(\frac{C_{st}(i) - \bar{C}_{st}}{10} \right)^2}$$

donde:

35 C_{st} es la correlación de tono en bucle abierto suavizada, definida como:

$$\underline{C_{st} = 0.9 \cdot C_{ol} + 0.1 \cdot C_{st};}$$

C_{ol} es la correlación de tono en bucle abierto calculada por el analizador 203 utilizando un procedimiento conocido por los expertos en la materia de codificación CELP, por ejemplo, tal como se describe en la recomendación ITU-T G.718, Sección 6.6;

40 C_{st} es el promedio sobre las últimas 10 tramas de la correlación de tono en bucle abierto suavizada C_{st} ;

σ_C es la variación de la correlación de tono en bucle abierto suavizada.

45 Cuando, durante el primer nivel de análisis, el clasificador de señal 204 clasifica una trama como no de voz, se llevan a cabo las siguientes verificaciones mediante el clasificador de señal 204 para determinar, en el segundo nivel de análisis, si es realmente seguro utilizar un modo de codificación mixta en el dominio de tiempo/dominio de frecuencia. Sin embargo, en ocasiones es mejor codificar toda la trama solamente con el modo de codificación en el dominio de tiempo, utilizando uno de los enfoques del dominio de tiempo estimados por la función de preprocesamiento del modo de codificación en el dominio de tiempo. En particular, puede ser mejor utilizar el modo de codificación en el dominio de tiempo sin memoria para reducir al mínimo cualquier posible pre-eco que se pueda introducir con el modo de codificación mixta en el dominio de tiempo/dominio de frecuencia.

50 Como una primera verificación de si se debería utilizar la codificación mixta en el dominio de tiempo/dominio de frecuencia, el clasificador de señal 204 calcula la diferencia entre la energía total de la trama actual y la energía total de la trama anterior. Cuando la diferencia E_{tot} entre la energía total de la trama actual E_{tot} y la energía total de la

trama anterior es mayor de 6 dB, esto corresponde a un denominado "ataque temporal" en la señal de sonido de entrada. En una situación de este tipo, se sobrescribe la decisión de voz/no de voz y el modo de codificación seleccionado y se fuerza un modo de codificación en el dominio de tiempo sin memoria. Más específicamente, el codificador CELP mejorado 100 comprende un selector de codificación de solamente tiempo/tiempo-frecuencia 103 (figura 1) que comprende por sí mismo un selector de audio de voz/genérico 205 (figura 2), un detector de ataque temporal 208 (figura 2) y un selector 206 del modo de codificación en el dominio de tiempo sin memoria. En otras palabras, en respuesta a una determinación de señal no de voz (audio genérico) por el detector 205 y una detección de un ataque temporal en la señal de sonido de entrada por el detector 208, el selector 206 fuerza a un codificador CELP de bucle cerrado 207 (figura 2) a utilizar el modo de codificación en el dominio de tiempo sin memoria. El codificador CELP en bucle cerrado 207 forma parte del codificador solamente del dominio de tiempo 104 de la figura 1.

Como una segunda verificación, cuando la diferencia E_{diff} entre la energía total de la trama actual E_{tot} y la energía total de la trama anterior es menor o igual que 6 dB, pero:

- la correlación de tono en bucle abierto suavizada C_{st} es mayor que 0,96; o
 - 15 - la correlación de tono en bucle abierto suavizada C_{st} es mayor que 0,85 y la diferencia E_{diff} entre la energía total de la trama actual E_{tot} y la energía total de la trama anterior está por debajo de 0,3 dB; o
 - la variación de la correlación de tono en bucle abierto suavizada σ_c está por debajo de 0,1 y la diferencia E_{diff} entre la energía total de la trama actual E_{tot} y la última energía total de la trama anterior está por debajo de 0,6 dB; o
 - la energía total de la trama actual E_{tot} está por debajo de 20 dB;
- 20 y esta es por lo menos la segunda trama consecutiva ($cnt \geq 2$) donde va a ser modificada la decisión del primer nivel del análisis, entonces el selector de audio de voz/genérico 205 determina que la trama actual se codificará utilizando un modo solamente del dominio de tiempo utilizando el codificador CELP genérico de bucle cerrado 207 (figura 2).

De lo contrario, el selector de codificación de tiempo/tiempo-frecuencia 103 selecciona un modo de codificación mixta en el dominio de tiempo/dominio de frecuencia que es llevado a cabo por un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia dado a conocer en la siguiente descripción.

Esto se puede resumir, por ejemplo cuando la señal de sonido no de voz es música, con el siguiente pseudocódigo:

si (audio genérico)

si ($E_{diff} > 6dB$)

modo de codificación = dominio de tiempo sin memoria

$cnt = 1$

si no, *si* ($C_{st} > 0.96 \mid (C_{st} > 0.85 \ \& \ E_{diff} < 0.3dB)$) ($\sigma_c < 0.1 \ \& \ E_{diff} < 0.6dB$) ($E_{tot} < 20dB$)

$cnt ++$

si ($cnt \geq 2$)

modo de codificación = dominio de tiempo

si no

modo de codificación = mezcla dominio de tiempo/frecuencia

$cnt = 0$

Donde E_{tot} es una energía de trama actual, expresada como:

$$E_{tot} = 10 \log \left(\frac{\sum_{i=0}^{t=N} x(i)^2}{N} \right)$$

30 (donde $x(i)$ representa las muestras de la señal de sonido de entrada en la trama) y E_{diff} es la diferencia entre la energía total de la trama actual E_{tot} y la última energía total de la trama anterior.

2) Decisión sobre la longitud de la subtrama

En un CELP típico, las muestras de señal de sonido de entrada se procesan en tramas de 10 a 30 ms, y estas tramas se dividen en varias subtramas para análisis de libro de códigos adaptativo y de libro de códigos fijo. Por ejemplo, una trama de 20 ms (256 muestras cuando la frecuencia de muestreo interna es 12,8 kHz) se puede utilizar

y dividir en 4 subtramas de 5 ms. Una longitud de subtrama variable es una característica utilizada para obtener una integración completa del dominio de tiempo y el dominio de frecuencia en un modo de codificación. La longitud de la subtrama puede variar de un típico 1/4 de la longitud de trama a la mitad de una trama, o a una longitud de trama completa. Por supuesto, se puede implementar la utilización de otro número de subtramas (longitud de la subtrama).

- 5 La decisión acerca de la longitud de las subtramas (el número de subtramas), o el soporte temporal, es determinada por un calculador del número de subtramas 210 en base a la tasa de bits disponible y al análisis de la señal de entrada en el preprocesador 102, en particular, la dinámica espectral de alta frecuencia de la señal de sonido de entrada 101 procedente de un analizador 209 y el análisis del tono en bucle abierto que incluye la correlación de tono en bucle abierto suavizada procedente del analizador 203. En analizador 209 es sensible a la información
10 procedente del analizador espectral 202 para determinar la dinámica espectral de alta frecuencia de la señal de entrada 101. La dinámica espectral se calcula a partir de una característica descrita en la recomendación ITU-T G.718, sección 6.7.2.2, como el espectro de entrada sin su ruido de fondo proporcionando una representación de la dinámica del espectro de entrada. Cuando la dinámica espectral promedio de la señal de sonido de entrada 101 en la banda de frecuencia entre 4,4 kHz y 6,4 kHz determinada por el analizador 209 está por debajo de 9,6 dB y se ha
15 considerado que la última trama tenía una alta dinámica espectral, se deja de considerar que la señal de entrada 101 tiene contenido de alta dinámica espectral en frecuencias superiores. En este caso, se pueden asignar más bits a las frecuencias por debajo, por ejemplo, de 4 kHz, añadiendo más subtramas al modo de codificación en el dominio de tiempo o forzando más pulsos en la parte de frecuencia inferior de la contribución del dominio de frecuencia.
- 20 Por otra parte, si el aumento de la dinámica promedio del contenido de frecuencia superior de la señal de entrada 101 frente a la dinámica espectral promedio de la última trama que no se ha considerado que tenga una alta dinámica espectral, determinado por el analizador 209, es mayor, por ejemplo, que 4,5 dB, se considera que la señal de entrada de sonido 101 tiene un contenido de alta dinámica espectral por encima, por ejemplo, de 4 kHz. En ese caso, dependiendo de la tasa de bits disponible, se utilizan algunos bits adicionales para codificar las altas
25 frecuencias de la señal de sonido de entrada 101 con el fin de permitir codificar uno o varios pulsos de frecuencia.

La longitud de la subtrama determinada por el calculador 210 (figura 2) depende asimismo del presupuesto de bits disponible. A una tasa de bits muy baja, por ejemplo a tasas de bits por debajo de 9 kbps, solamente está disponible una subtrama para la codificación en el dominio de tiempo, o de lo contrario el número de bits disponibles será
30 insuficiente para la codificación en el dominio de frecuencia. Para tasas de bits medias, por ejemplo tasas de bits entre 9 kbps y 16 kbps, se utiliza una subtrama para el caso en que las altas frecuencias contienen un alto contenido espectral dinámico, y dos subtramas en caso contrario. Para tasas de bits medias-altas, por ejemplo tasas de bits en torno a 16 kbps y superiores, se hace disponible asimismo el caso de cuatro (4) subtramas si la correlación de tono en bucle abierto suavizada C_{st} , tal como se define en el párrafo [0037] de la sección de clasificación del tipo de sonido, es mayor que 0,8.

35 Aunque el caso con una o dos subtramas limita la codificación en el dominio de tiempo a solamente una contribución de libro de códigos adaptativo (con retardo de tono y ganancia de tono codificados), es decir no se utiliza ningún libro de códigos fijo en este caso, las cuatro (4) subtramas permiten contribuciones de libros de códigos adaptativos y fijos si el presupuesto de bits es suficiente. Se permite que comience el caso de cuatro (4) subtramas aproximadamente de 16 kbps en adelante. Debido a las limitaciones del presupuesto de bits, la excitación del dominio de tiempo consiste solamente en la contribución del libro de códigos adaptativo a tasas de bits menores. Se
40 puede añadir una simple contribución del libro de códigos fijo para tasas de bits superiores, por ejemplo a partir de 24 kbps. Para todos los casos, la eficiencia de codificación en el dominio de tiempo será evaluada después para decidir hasta qué frecuencia es valiosa dicha codificación en el dominio de tiempo.

3) Análisis de tono en bucle cerrado

45 Cuando se utiliza un modo de codificación mixta en el dominio de tiempo/dominio de frecuencia, se lleva a cabo un análisis de tono de bucle cerrado seguido, si es necesario, por una búsqueda de libro de códigos algebraico fijo. Para este propósito, un codificador CELP 100 (figura 1) comprende un calculador de la contribución de excitación del dominio de tiempo 105 (figuras 1 y 2). Este calculador comprende además un analizador 211 (figura 2) sensible al análisis de tono en bucle abierto realizado en el analizador de tono en bucle abierto 203 y a la determinación de la
50 longitud de la subtrama (o del número de subtramas en una trama) en el calculador 210, para llevar a cabo un análisis de tono en bucle cerrado. El análisis de tono en bucle cerrado es bien conocido por los expertos en la materia, y se describe un ejemplo de implementación, por ejemplo, en la referencia [recomendación ITU-T G.718; sección 6.8.4.1.4.1], cuyo contenido íntegro se incorpora como referencia a la presente memoria. El análisis de tono en bucle cerrado tiene como resultado el cálculo de los parámetros de tono, conocidos asimismo como parámetros de libro de códigos adaptativo, que consisten principalmente en un retardo de tono (índice de libro de códigos adaptativo T) y una ganancia de tono (o ganancia de libro de códigos adaptativo b). La contribución del libro de códigos adaptativo es normalmente la excitación anterior en el retardo T o una versión interpolada de la misma. El índice de libro de códigos adaptativo T es codificado y transmitido a un descodificador distante. La ganancia de tono
55 b es asimismo cuantificada y transmitida al descodificador distante.

5 Cuando el análisis de tono en bucle cerrado se ha completado, el codificador CELP 100 comprende un libro de códigos fijo 212 buscado para encontrar los mejores parámetros del libro de códigos fijo, que comprenden normalmente un índice del libro de códigos fijo y una ganancia del libro de códigos fijo. El índice y la ganancia del libro de códigos fijo forman la contribución del libro de códigos fijo. El índice del libro de códigos fijo es codificado y transmitido al descodificador distante. La ganancia del libro de códigos fijo es asimismo cuantificada y transmitida al descodificador distante. Se considera que el libro de códigos algebraico fijo y la búsqueda del mismo son conocidos por los expertos en la materia de la codificación CELP, y por lo tanto no se describirán en mayor detalle en la presente invención.

10 El índice y la ganancia del libro de códigos adaptativo y el índice y la ganancia del libro de códigos fijo forman una contribución de excitación CELP del dominio de tiempo.

4) Transformada de frecuencia de la señal de interés

15 Durante la codificación en el dominio de frecuencia del modo de codificación mixta en el dominio de tiempo/dominio de frecuencia, es necesario representar dos señales en un dominio de transformada, por ejemplo en el dominio de frecuencia. En una realización, la transformada de tiempo a frecuencia se puede conseguir utilizando una DCT (Discrete Cosine Transform, transformada de coseno discreta) de tipo II (o de tipo IV) de 256 puntos, que proporciona una resolución de 25 Hz con una frecuencia de muestreo interna de 12,8 kHz, pero se podría utilizar cualquier otra transformada. En caso de que se utilice otra transformada, puede ser necesario revisar en correspondencia la resolución de frecuencia (definida anteriormente), el número de bandas de frecuencia y el número de cajas de frecuencia por banda (definido en mayor detalle a continuación). A este respecto, el codificador CELP 100 comprende un calculador 107 (figura 1) de una contribución de excitación del dominio de frecuencia en respuesta al residuo LP de entrada $r_{es}(n)$ resultante del análisis LP de la señal de sonido de entrada mediante el analizador 201. Tal como se muestra en la figura 2, el calculador 107 puede calcular una DCT 213, por ejemplo una DCT de tipo II del residuo LP de entrada $r_{es}(n)$. El codificador CELP 100 comprende asimismo un calculador 106 (figura 1) de una transformada de frecuencia de la contribución de excitación del dominio de tiempo. Tal como se muestra en la figura 2, el calculador 106 puede calcular una DCT 214, por ejemplo una DCT de tipo II de la contribución de excitación del dominio de tiempo. La transformada de frecuencia del residuo LP de entrada r_{es} y la contribución de excitación CELP del dominio de tiempo f_{exc} se pueden calcular utilizando las expresiones siguientes:

$$f_{res}(k) = \begin{cases} \sqrt{\frac{1}{N}} \cdot \sum_{n=0}^{N-1} r_{es}(n) \cdot \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right), & k = 0 \\ \sqrt{\frac{2}{N}} \cdot \sum_{n=0}^{N-1} r_{es}(n) \cdot \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right), & 1 \leq k < N-1 \end{cases}$$

y:

$$f_{exc}(k) = \begin{cases} \sqrt{\frac{1}{N}} \cdot \sum_{n=0}^{N-1} e_{td}(n) \cdot \cos\left(\frac{\pi}{N}\left(n - \frac{1}{2}\right)k\right), & k = 0 \\ \sqrt{\frac{2}{N}} \cdot \sum_{n=0}^{N-1} e_{td}(n) \cdot \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right), & 1 \leq k < N-1 \end{cases}$$

30 donde $r_{es}(n)$ es el residuo LP de entrada, $e_{td}(n)$ es la contribución de excitación del dominio de tiempo y N es la longitud de trama. En una posible implementación, la longitud de trama es de 256 muestras para una correspondiente frecuencia de muestreo interna de 12,8 kHz. La contribución de excitación del dominio de tiempo está dada por la siguiente relación:

$$35 \quad e_{td}(n) = bv(n) + gc(n)$$

donde $v(n)$ es la contribución del libro de códigos adaptativo, b es la ganancia del libro de códigos adaptativo, $c(n)$ es la contribución del libro de códigos fijo y g es la ganancia del libro de códigos fijo. Se debe observar que la contribución de excitación del dominio de tiempo puede consistir solamente en la contribución del libro de códigos adaptativo, tal como se ha descrito en la descripción anterior.

40 5) Frecuencia de corte de la contribución del dominio de tiempo

45 Con muestras de audio genérico, la contribución de excitación del dominio de tiempo (la combinación de libros de códigos algebraicos adaptativo y/o fijo) no siempre contribuye mucho a la mejora de la codificación comparada con la codificación en el dominio de frecuencia. A menudo, esto mejora la codificación de la parte inferior del espectro, mientras que la mejora de la codificación de la parte superior del espectro es mínima. El codificador CELP 100 comprende un determinador de una frecuencia de corte y un filtro 108 (figura 1), que es la frecuencia en la que la

mejora de codificación proporcionada por la contribución de excitación del dominio de tiempo se hace demasiado baja para ser valiosa. El determinador y filtro 108 comprenden un calculador de la frecuencia de corte 215 y el filtro 216 de la figura 2. La frecuencia de corte de la contribución de excitación del dominio de tiempo es primero estimada por el calculador 215 (figura 2) utilizando un calculador 303 (figuras 3 y 4) de correlación cruzada normalizada, para cada banda de frecuencia entre el residuo LP de entrada transformado en frecuencia procedente del calculador 107 y la contribución de excitación del dominio de tiempo transformada en frecuencia procedente del calculador 106, denominadas respectivamente f_{res} y f_{exc} , que se definen en la sección anterior 4. La última frecuencia L_f incluida en cada una de, por ejemplo, las dieciséis (16) bandas de frecuencia se define en Hz como:

$$L_f = \left\{ \begin{array}{l} 175, 375, 775, 1175, 1575, 1975, 2375, 2775, \\ 3175, 3575, 3975, 4375, 4775, 5175, 5575, 6375 \end{array} \right\}$$

10 Para este ejemplo ilustrativo, el número de cajas de frecuencia por banda B_b , las cajas de frecuencia acumulativas por banda C_{Bb} y la correlación cruzada normalizada por banda de frecuencia $C_c(t)$ se definen como sigue, para una trama de 20 ms a una frecuencia de muestreo de 12,8 kHz:

$$B_b = \left\{ \begin{array}{l} 8, 8, 16, 16, 16, 16, 16, 16, \\ 16, 16, 16, 16, 16, 16, 16, 32 \end{array} \right\}$$

$$C_{Bb} = \left\{ \begin{array}{l} 0, 8, 16, 32, 48, 64, 80, 96, \\ 112, 128, 144, 160, 176, 192, 208, 224 \end{array} \right\}$$

$$15 \quad C_c(i) = \frac{\sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{exc}(j) \cdot f_{res}(j)}{\sqrt{(S'_{f_{exc}}(i) \cdot S'_{f_{res}}(i))}}$$

Donde

$$S'_{f_{exc}}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{exc}(j)^2$$

y

$$S'_{f_{res}}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{res}(j)^2$$

20 donde B_b es el número de cajas de frecuencia por banda B_b , C_{Bb} son las cajas de frecuencia acumulativas por banda, $C_{Bb}C_c(i)C_c(i)$ es la correlación cruzada normalizada por banda de frecuencia, $S'_{f_{exc}}$ es la energía de excitación para una banda y análogamente $S'_{f_{res}}$ es la energía de residuo por banda.

25 El calculador de la frecuencia de corte 215 comprende un elemento de suavizado 304 (figuras 3 y 4) de correlación cruzada a través de las bandas de frecuencia, que lleva a cabo algunas operaciones para suavizar el vector de correlación cruzada entre las diferentes bandas de frecuencia. Más específicamente, el elemento de suavizado 304 de correlación cruzada a través de las bandas calcula un nuevo vector de correlación cruzada C_{c2} utilizando la siguiente relación:

$$C_{c2}(i) = \left\{ \begin{array}{l} 2 \cdot (\min(0.5, \alpha \cdot C_c(0) + \delta C_c(1)) - 0.5) \quad \text{para } i = 0 \\ 2 \cdot (\min(0.5, \alpha \cdot C_c(i) + \beta C_c(i+1) + \beta C_c(i-1)) - 0.5) \quad \text{para } 1 \leq i < N_b \end{array} \right\}$$

donde

$$30 \quad \alpha = 0.95; \quad \delta = (1 - \alpha); \quad N_b = 13; \quad \beta = \frac{\delta}{2}$$

El calculador de frecuencia de corte 215 comprende además un calculador 305 (figuras 3 y 4) de un promedio del nuevo vector de correlación cruzada C_{c2} sobre las primeras N_b bandas ($N_b = 13$ representa 5575 Hz).

El calculador 215 de la frecuencia de corte comprende asimismo un módulo de frecuencia de corte 306 (figura 3) que incluye un limitador 406 (figura 4) de la correlación cruzada, un normalizador 407 de la correlación cruzada y un determinador 408 de la banda de frecuencia donde la correlación cruzada es la mínima. Más específicamente, el limitador 406 limita el promedio del vector de correlación cruzada a un valor mínimo de 0,5 y el normalizador 408 normaliza el promedio limitado del vector de correlación cruzada entre 0 y 1. El determinador 408 obtiene una primera estimación de la frecuencia de corte encontrando la última frecuencia de una banda de frecuencia L_f que minimiza la diferencia entre dicha última frecuencia de una banda de frecuencia L_f y el promedio normalizado C_{c2} del vector de correlación cruzada C_{c2} multiplicado por la anchura $F/2$ del espectro de la señal de entrada.

$$i_{min} = \min_{0 \leq i < N_b} \left(L_f(i) - \overline{C_{c2}} \cdot \left(\frac{F_s}{2} \right) \right) \quad y \quad f_{tc1} = L_f(i_{min})$$

10 donde

$$F_s = 12800 \text{ Hz} \quad y \quad \overline{C_{c2}} = \frac{\sum_{i=0}^{i=N_b-1} (C_{c2}(i))}{N_b}$$

f_{tc1} es la primera estimación de la frecuencia de corte.

A una tasa de bits baja, donde el promedio normalizado C_{c2} nunca es realmente alto, o para aumentar artificialmente el valor de f_{tc1} para proporcionar algo más de peso a la contribución del dominio de tiempo, es posible sobredimensionar el valor de C_{c2} con un factor de escala fijo, por ejemplo, a una tasa de bits por debajo de 8 kbps, f_{tc1} se multiplica por 2 todo el tiempo en la implementación de ejemplo.

La precisión de la frecuencia de corte se puede aumentar añadiendo un siguiente componente al cálculo. A tal efecto, el calculador 215 de la frecuencia de corte comprende un extrapolador 410 (figura 4) del octavo armónico calculado a partir del valor de retardo de tono mínimo o más bajo de la contribución de excitación del dominio de tiempo de todas las subtramas, utilizando la siguiente relación:

$$h_{8th} = \frac{8 \cdot F_s}{\min_{0 \leq i < N_{sub}} (T(i))}$$

donde $F_s = 12800 \text{ Hz}$, N_{sub} es el número de subtramas y $T(i)$ es el índice de libro de códigos adaptativo o el retardo de tono para la subtrama i .

El calculador 215 de la frecuencia de corte comprende asimismo un determinador 409 (figura 4) de la banda de frecuencia en la que está situado el octavo armónico h_{8th} . Más específicamente, para todo $i < N_b$, el determinador 409 busca la banda de frecuencia más alta para la que se sigue verificando la siguiente desigualdad:

$$(h_{8th} \geq L_f(i)) \quad h_{8th} \geq L_f(i)$$

El índice de dicha banda se puede denominar i_{8th} e indica la banda en la que probablemente está situado el octavo armónico.

El calculador 215 de la frecuencia de corte comprende finalmente un selector 411 (figura 4) de la frecuencia de corte final f_{tc} . Más específicamente, el selector 411 retiene la frecuencia superior entre la primera estimación f_{tc1} de la frecuencia de corte procedente del determinador 408 y la última frecuencia de la banda de frecuencia en la que está situado el octavo armónico ($L_f(i_{8th})$), utilizando la siguiente relación:

$$f_{tc} = \max(L_f(i_{8th}), f_{tc1})$$

35 Tal como se muestra en las figuras 3 y 4,

- el calculador 215 de la frecuencia de corte comprende además un decisor 307 (figura 3) sobre el número de cajas de frecuencia a reducir a cero, que incluye por sí mismo un analizador 415 (figura 4) de parámetros y un selector 416 (figura 4) de cajas de frecuencia a reducir a cero; y

40 - el filtro 216 (figura 2), que funciona en el dominio de frecuencia, comprende un reductor a cero 308 (figura 3) de las cajas de frecuencia que se ha decidido reducir a cero. El reductor a cero puede reducir a cero todas las cajas de frecuencia (reductor a cero 417 en la figura 4), o (filtro 418 en la figura 4) tan sólo algunas de las cajas de frecuencia superiores situadas por encima de la frecuencia de corte f_{tc} complementada con una zona de transición suave. La zona de transición está situada por encima de la frecuencia de corte f_{tc} y por debajo de las cajas reducidas a cero, y permite una transición espectral suave entre el espectro no modificado por debajo de f_{tc} y las cajas reducidas a cero en frecuencias superiores.

Para el ejemplo ilustrativo, cuando la frecuencia de corte f_{tc} del selector 411 es menor o igual que 775 Hz, el analizador 415 considera que el coste de la contribución de excitación del dominio de tiempo es demasiado elevado. El selector 416 selecciona todas las cajas de frecuencia de la representación en frecuencia de la contribución de excitación del dominio de tiempo a reducir a cero, y el reductor a cero 417 fuerza la reducción a cero de todas las cajas de frecuencia y fuerza asimismo la reducción a cero de la frecuencia de corte f_{tc} . Todos los bits asignados a la contribución de excitación del dominio de tiempo se asignan a continuación al modo de codificación en el dominio de frecuencia. De lo contrario, el analizador 415 fuerza el selector 416 a elegir las cajas de frecuencia altas por encima de la frecuencia de corte f_{tc} para reducir a cero por el reductor a cero 418.

Finalmente, el calculador 215 de frecuencia de corte comprende un cuantificador 309 (figuras 3 y 4) de la frecuencia de corte f_{tc} en una versión cuantificada f_{tcQ} de esta frecuencia de corte. Si se asocian tres (3) bits al parámetro de la frecuencia de corte, un posible conjunto de valores de salida se puede definir (en Hz) como sigue:

$$f_{tcQ} = \{0, 1175, 1575, 1975, 2375, 2775, 3175, 3575\}$$

Se podrían utilizar muchos mecanismos para estabilizar la elección de la frecuencia de corte final f_{tc} con el fin de impedir que la versión cuantificada f_{tcQ} conmute entre 0 y 1175 en un segmento de señal inapropiado. Para conseguir esto, el analizador 415 de esta implementación de ejemplo es sensible a la ganancia de tono promedio a largo plazo G_{lt} 412 procedente del analizador de tono de bucle cerrado 211 (figura 2), a la correlación de bucle abierto C_{ol} 413 procedente del analizador de tono de bucle abierto 203 y a la correlación de bucle abierto suavizada C_{st} . Para impedir la conmutación a una codificación de frecuencia completa, cuando se cumplen las siguientes condiciones el analizador 415 no permite la codificación de sólo frecuencia, es decir, f_{tcQ} no se puede ajustar a 0:

$$f_{tc} > 2375 \text{ Hz}$$

o

$$f_{tc} > 1175 \text{ Hz} \text{ y } C_{ol} > 0.7 \text{ y } G_{lt} \geq 0.6$$

o

$$f_{tc} \geq 1175 \text{ Hz} \text{ y } C_{st} > 0.8 \text{ y } G_{lt} \geq 0.4$$

o

$$f_{tcQ}(t-1) \neq 0 \text{ y } C_{ol} > 0.5 \text{ y } C_{st} > 0.5 \text{ y } G_{lt} \geq 0.6$$

donde C_{ol} es la correlación de tono en bucle abierto 413 y C_{st} corresponde la versión suavizada de la correlación de tono en bucle abierto 414 definida como $C_{st} = 0.9 \cdot C_{ol} + 0.1 \cdot C_{st}$. Además, G_{lt} (elemento 412 de la figura 4) corresponde al promedio a largo plazo de la ganancia de tono obtenida mediante el analizador de tono de bucle cerrado 211 dentro de la contribución de excitación del dominio de tiempo. El promedio a largo plazo de la ganancia de tono 412 se define como $G_{lt} = 0.9 \cdot G_{lt} + 0.1 \cdot G_{lt}$ y G_{lt} es la ganancia de tono promedio sobre la trama actual. Para reducir más la velocidad de conmutación entre codificación de sólo frecuencia y codificación mixta en el dominio de tiempo/dominio de frecuencia, se puede añadir un residuo.

6) Codificación en el dominio de frecuencia

Creación de un vector de diferencia

Una vez se ha definido la frecuencia de corte de la contribución de excitación del dominio de tiempo, se lleva a cabo la codificación en el dominio de frecuencia. El codificador CELP 100 comprende un substractor o calculador 109 (figuras 1, 2, 5 y 6) para formar una primera parte de un vector de diferencia f_d con la diferencia entre la transformada de frecuencia 502 (figuras 5 y 6) (u otra representación en frecuencia) del residuo LP de entrada procedente de la DCT 213 (figura 2) y la transformada de frecuencia f_{exc} 501 (figuras 5 y 6) (u otra representación en frecuencia) de la contribución de excitación del dominio de tiempo procedente de la DCT 214 (figura 2), desde cero hasta la frecuencia de corte f_{tc} de la contribución de excitación del dominio de tiempo. Un factor de reducción de escala 603 (figura 6) se aplica a la transformada de frecuencia f_{exc} 501 para la siguiente zona de transición de $f_{trans}=2$ kHz (80 cajas de frecuencia en esta implementación de ejemplo) antes de su resta de la respectiva parte espectral de la transformada de frecuencia f_{res} . El resultado de esta resta constituye la segunda parte del vector de diferencia f_d que representa el intervalo de frecuencias desde la frecuencia de corte f_{tc} hasta $f_{tc}+f_{trans}$. La transformada de frecuencia f_{res} 502 del residuo LP de entrada se utiliza para la restante tercera parte del vector f_d . La parte en escala reducida del vector f_d resultante de la aplicación del factor de reducción de escala 603 se puede llevar a cabo con cualquier tipo de función de desvanecimiento, se puede acortar a solamente unas pocas cajas de frecuencia, pero podría asimismo omitirse cuando se estima que el presupuesto de bits es suficiente para impedir artefactos de oscilación de energía cuando la frecuencia de corte f_{tc} está cambiando. Por ejemplo, con una resolución de 25 Hz, correspondiente a 1 caja de frecuencia $f_{bin} = 25$ Hz en DCT de 256 puntos a 12,8 kHz, el vector de diferencia se puede construir como:

$$f_d(k) = f_{res}(k) - f_{exc}(k)$$

donde $0 \leq k \leq f_{tc} / f_{bin}$

$$f_d(k) = f_{res}(k) - f_{exc}(k) \cdot \left(1 - \text{sen} \left(\frac{\pi}{2} \cdot \frac{f_{bin}}{f_{trans}} \cdot \left(k - \frac{f_{tc}}{f_{bin}} \right) \right) \right)$$

donde $f_{tc} / f_{bin} < k \leq (f_{tc} + f_{trans}) / f_{bin}$

5 $f_d(k) = f_{res}(k)$, si no

donde f_{res} , f_{exc} y f_{ts} se han definido en las anteriores acciones 4 y 5.

Búsqueda de pulsos de frecuencia

El codificador CELP 100 comprende un cuantificador de frecuencia 110 (figuras 1 y 2) del vector de diferencia f_d . El vector de diferencia f_d se puede cuantificar utilizando varios procedimientos. En todos los casos, es necesario buscar y cuantificar pulsos de frecuencia. En un posible procedimiento simple, la codificación en el dominio de frecuencia comprende una búsqueda de la mayor parte de los pulsos energéticos del vector de diferencia f_d a través del espectro. El procedimiento para buscar los pulsos puede ser tan simple como dividir el espectro en bandas de frecuencia y permitir un cierto número de pulsos por bandas de frecuencia. El número de pulsos por bandas de frecuencia depende del presupuesto de bits disponible y de la posición de la banda de frecuencia en el interior del espectro. Habitualmente, se asignan más pulsos a las frecuencias bajas.

Vector de diferencia cuantificado

Dependiendo de la tasa de bits disponible, la cuantificación de los pulsos de frecuencia se puede llevar a cabo utilizando diferentes técnicas. En una realización, a una tasa de bits por debajo de 12 kbps, se puede utilizar un esquema de búsqueda y cuantificación simple para codificar la posición y el signo de los pulsos. Este esquema se describe a continuación.

Por ejemplo, para frecuencias por debajo de 3175 Hz, este esquema de búsqueda y cuantificación simple utiliza un enfoque basado en codificación factorial de pulsos (FPC, factorial pulse coding) que se describe en la bibliografía, por ejemplo en la referencia [Mittal, U., Ashley, J.P., y Cruz-Zeno, E.M (2007), "Low Complexity Factorial Pulse Coding of MDCT Coefficients using Approximation of Combinatorial Functions", Actas del IEEE sobre acústica, voz y procesamiento de señal, volumen 1, abril, páginas 289-292], la totalidad de cuyo contenido se incorpora como referencia a la presente memoria.

Más específicamente, un selector 504 (figuras 5 y 6) determina que todo el espectro no es cuantificado utilizando FPC. Tal como se muestra en la figura 5, la codificación FPC y la codificación de la posición y del signo de los pulsos se llevan a cabo en un codificador 506. Tal como se muestra en la figura 6, el codificador 506 comprende un buscador 609 de pulsos de frecuencia. La búsqueda se realiza a través de todas las bandas de frecuencia para las frecuencias menores que 3175 Hz. Un codificador FPC 610 procesa a continuación los pulsos de frecuencia. El codificador 506 comprende asimismo un determinador 611 de la mayor parte de los pulsos energéticos para frecuencias iguales y mayores que 3175 Hz, y un cuantificador 612 de la posición y del signo de los pulsos más energéticos encontrados. Si se permite más de un (1) pulso dentro de una banda de frecuencia, entonces la amplitud del pulso encontrado anteriormente se divide por 2 y la búsqueda se realiza de nuevo sobre toda la banda de frecuencia. Cada vez que se encuentra un pulso, su posición y su signo se almacenan para cuantificación y para la etapa de empaquetamiento de bits. El siguiente pseudocódigo muestra este esquema de búsqueda y cuantificación simple:

```

para k = 0 : NBD
    para i = 0 : Np
        pmax = 0
        para j = CBb(k) : CBb(k) + Bb(k)
            si fd(j)2 > pmax
                pmax = fd(j)2
                fd(j) =  $\frac{f_d(j)}{2}$ 
                pp(i) = j
                ps(i) = sign(fd(j))
            fin
        fin
    fin
fin

```

Donde N_{BD} es el número de bandas de frecuencia ($N_{BD} = 16$ en el ejemplo ilustrativo), N_p es el número de pulsos a codificar en una banda de frecuencia k , B_b es el número de cajas de frecuencia por banda de frecuencia B_b , C_{Bb} son las cajas de frecuencia acumulativas por banda, tal como se ha definido anteriormente en la sección 5, p_p representa el vector que contiene la posición de pulso encontrada, $p_s p_s$ representa el vector que contiene el signo del pulso encontrado y $p_{max} \square p_{max}$ representa la energía del pulso encontrado.

A una tasa de bits por encima de 12 kbps, el selector 504 determina que todo el espectro se tiene que cuantificar utilizando FPC. Tal como se muestra en la figura 6, la codificación FPC se lleva a cabo en un codificador 505. Tal como se muestra en la figura 6, el codificador 505 comprende un buscador 607 de pulsos de frecuencia. La búsqueda se realiza a través de todas las bandas de frecuencia. Un procesador FPC 610 codifica con FPC a continuación los pulsos de frecuencia encontrados.

A continuación, el vector de diferencia cuantificado f_{dQ} se obtiene sumando el número de pulsos nb_pulses con el signo del pulso p_s a cada una de las posiciones p_p encontradas. Para cada banda, el vector de diferencia cuantificado f_{dQ} se puede escribir con el siguiente pseudocódigo:

```

para j = 0, ..., j < nb_pulses
    fdQ(pp(j)) += ps(j)

```

Relleno con ruido

Todas las bandas de frecuencia son cuantificadas con más o menos precisión; el procedimiento de cuantificación descrito en la sección anterior no garantiza que se cuantifiquen todas las cajas de frecuencia dentro de las bandas de frecuencia. Este es el caso sobretodo a tasas de bits bajas donde el número de pulsos cuantificados por banda de frecuencia es relativamente bajo. Para impedir la aparición de artefactos audibles debido a estas cajas no cuantificadas, un elemento de relleno con ruido 507 (figura 5) añade algún ruido para rellenar estos huecos. Esta adición de ruido se lleva a cabo sobre todo el espectro a una tasa de bits por debajo de 12 kbps, por ejemplo, pero se puede aplicar solamente por encima de la frecuencia de corte f_{fc} de la contribución de excitación del dominio de tiempo para tasas de bits superiores. Por simplicidad, la intensidad del ruido varía solamente con la tasa de bits disponible. A tasas de bits altas el nivel de ruido es bajo, pero el nivel de ruido es superior a tasas de bits bajas.

El dispositivo de relleno de ruido 504 comprende un sumador 613 (figura 6) que añade ruido al vector de diferencia cuantificado f_{dQ} después de que se ha determinado la intensidad o el nivel de energía de dicho ruido añadido, en un estimador 614, y antes de que se haya determinado la ganancia por banda en un calculador 615. En la realización ilustrativa, el nivel de ruido está relacionado directamente con la tasa de bits codificada. Por ejemplo, a 6,60 kbps el nivel de ruido N'_L es 0,4 veces la amplitud de los pulsos espectrales codificados en una banda específica, y desciende progresivamente a un valor de 0,2 veces la amplitud de los pulsos espectrales codificados en una banda a 24 kbps. El ruido se suma solamente a la sección o secciones del espectro donde un determinado número de cajas de frecuencia consecutivas tienen una energía muy baja, por ejemplo, cuando el número de cajas consecutivas de muy baja energía N_z es la mitad del número de cajas incluidas en la banda de frecuencia. Para una banda específica i , el ruido se inyecta como:

```

para j = CBb(i), ..., j < CBb(i) + Bb(i)

```

$$\text{si } \sum_{k=j}^{j+N_z} f_{dQ}(k)^2 < 0,5$$

$$\text{para } k = j, \dots, k < j + N_z$$

$$f_{dQ}(k) = f_d(k) + N'_L(i) \cdot r_{and}(\)$$

$$j+ = N_z$$

$$N_z = \frac{B_b(i)}{2}$$

5 Donde

donde, para una banda i , C_{Bb} es el número acumulativo de cajas por bandas, B_b es el número de cajas en una banda específica i , N'_L es el nivel de ruido y r_{and} es un generador numérico aleatorio que está limitado entre -1 y 1

7) Cuantificación de ganancia por banda

10 El cuantificador de frecuencia 110 comprende un calculador/cuantificador de ganancia por banda 508 (figura 5) que incluye un calculador 615 (figura 6) de ganancia por banda y cuantificador 616 (figura 6) de la ganancia por banda calculada. Una vez encontrado el vector de diferencia cuantificado f_{dQ} , incluyendo el relleno de ruido si es necesario, el calculador 615 calcula la ganancia por banda para cada banda de frecuencia. La ganancia por banda para una banda específica $G_b(i)$ se define como la relación entre la energía de la señal del vector de diferencia no cuantificado f_d y la energía del vector de diferencia cuantificado f_{dQ} en el dominio logarítmico, como:

$$15 \quad G_b(i) = \log_{10} \left(\frac{S_{f'_d}(i)}{S_{f_{dQ}}(i)} \right)$$

$$\text{Donde } S_{f'_d}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_d(j)^2 \quad \text{y} \quad S_{f_{dQ}}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{dQ}(j)^2$$

donde C_{Bb} y B_b se han definido anteriormente en la sección 5.

20 En la realización de las figuras 5 y 6, el cuantificador de ganancia por banda 616 cuantifica vectorialmente las ganancias de frecuencia por banda. Antes de la cuantificación vectorial, a baja tasa de bits, la última ganancia (correspondiente a la última banda de frecuencia) se cuantifica por separado, y la totalidad de las quince (15) ganancias restantes se dividen por la última ganancia cuantificada. A continuación, las quince (15) restantes ganancias normalizadas son cuantificadas vectorialmente. A una tasa superior, la media de las ganancias por banda se cuantifica en primer lugar y a continuación se elimina de todas las ganancias por banda de, por ejemplo, dieciséis (16) bandas de frecuencia antes de la cuantificación vectorial de dichas ganancias por banda. La cuantificación vectorial que se está utilizando puede ser una minimización estándar en el dominio logarítmico de la distancia entre el vector que contiene las ganancias por banda y las entradas de un libro de códigos específico.

25 En el modo de codificación en el dominio de frecuencia, se calculan ganancias en el calculador 615 para cada banda de frecuencia con el fin de ajustar la energía del vector no cuantificado f_d al vector cuantificado f_{dQ} . Las ganancias son cuantificadas vectorialmente en el cuantificador 616 y aplicadas por banda al vector cuantificado f_{dQ} por medio de un multiplicador 509 (figuras 5 y 6).

30 Alternativamente, es posible asimismo utilizar el esquema de codificación FPC a una tasa por debajo de 12 kbps para todo el espectro, seleccionando solamente algunas de las bandas de frecuencia a cuantificar. Antes de llevar a cabo la selección de las bandas de frecuencia, se cuantifica la energía E_d de las bandas de frecuencia del vector de diferencia no cuantificado f_d . La energía se calcula como:

$$35 \quad E_d(i) = \log_{10} (S_d(i))$$

$$S_d(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_d(j)^2$$

donde

donde C_{Bb} y B_b se han definido anteriormente en la sección 5.

5 Para llevar a cabo la cuantificación de la energía de banda de frecuencia E_d , en primer lugar la energía promedio sobre las primeras 12 bandas entre las dieciséis bandas es cuantificada y restada de la totalidad de las dieciséis (16) energías de banda. A continuación todas las bandas de frecuencia son cuantificadas vectorialmente por cada grupo de 3 o 4 bandas. La cuantificación vectorial que se está utilizando puede ser una minimización estándar en el dominio logarítmico de la distancia entre el vector que contiene las ganancias por banda y las entradas de un libro de códigos específico. Si no hay suficientes bits disponibles, es posible cuantificar solamente las primeras 12 bandas y extrapolar las últimas 4 bandas utilizando el promedio de las anteriores 3 bandas, o mediante cualesquiera otros procedimientos.

10 Una vez que las energías de las bandas de frecuencia del vector de diferencia no cuantificado se cuantifican, es posible ordenar la energía en orden decreciente de tal modo que sería aplicable en el lado del descodificador. Durante la clasificación, se mantienen siempre todas las bandas de energía por debajo de 2 kHz, y entonces solamente las bandas más energéticas se transferirán al FPC para codificar amplitudes y signos de pulsos. Con este enfoque, el esquema FPC codifica un vector más pequeño pero abarca un intervalo de frecuencias más amplio. En otras palabras, cuesta menos bits cubrir eventos de energía importantes sobre todo el espectro.

15 Después del proceso de cuantificación de pulsos, es necesario un relleno con ruido similar al que se ha descrito anteriormente. A continuación, un factor de ajuste de ganancia G_a se calcula por cada banda de frecuencia, para ajustar la energía E_{dQ} del vector de diferencia cuantificado f_{dQ} a la energía cuantificada E_d del vector de diferencia no cuantificado f_d . Entonces, este factor de ajuste de la ganancia por banda se aplica al vector de diferencia cuantificado f_{dQ} .

20
$$G_a(i) = 10^{E_d(i) - E_{dQ}(i)}$$

donde

$$E_{dQ}(i) = \log_{10} \left(\sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{dQ}(j)^2 \right)$$

y E_d es la energía cuantificada por banda del vector de diferencia no cuantificado f_d que se ha definido anteriormente.

25 Después de la finalización de la etapa de codificación en el dominio de frecuencia, se obtiene la excitación total del dominio de tiempo/dominio de frecuencia sumando por medio de un sumador 111 (figuras 1, 2, 5 y 6) el vector de diferencia cuantificado en frecuencia f_{dQ} a la contribución de excitación del dominio de tiempo transformada en frecuencia filtrada f_{excF} . Cuando el codificador CELP mejorado 100 cambia su asignación de bits desde un modo de codificación solamente en el dominio de tiempo a un modo de codificación mixta en el dominio de tiempo/dominio de frecuencia, la energía del espectro de excitación por banda de frecuencia del modo de codificación solamente en el dominio de tiempo no se ajusta a la energía del espectro de excitación por banda de frecuencia del modo de codificación mixta en el dominio de tiempo/dominio de frecuencia. Este desajuste de energía puede crear artefactos de conmutación que son más audibles a una tasa de bits baja. Para reducir cualquier degradación audible creada por esta reasignación de bits, se puede calcular una ganancia a largo plazo para cada banda y se puede aplicar a la excitación sumada, con el fin de corregir la energía de cada banda de frecuencia para unas pocas tramas después de la reasignación. A continuación, la suma del vector de diferencia cuantificado en frecuencia f_{dQ} y la contribución de excitación del dominio de tiempo transformada en frecuencia y filtrada f_{excF} se transforma de nuevo al dominio de tiempo en un convertidor 112 (figuras 1, 5 y 6) que comprende por ejemplo una IDCT (DCT inversa) 220.

40 Finalmente, la señal sintetizada se calcula filtrando la señal de excitación total procedente de la IDCT 220 por medio de un filtro de síntesis LP 113 (figuras 1 y 2).

45 La suma del vector de diferencia cuantificado en frecuencia f_{dQ} y la contribución de excitación del dominio de tiempo transformada en frecuencia y filtrada f_{excF} forma la excitación mixta en el dominio de tiempo/dominio de frecuencia transmitida a un descodificador distante (no mostrado). El descodificador distante comprenderá asimismo el convertidor 112 para transformar la excitación mixta en el dominio de tiempo/dominio de frecuencia de nuevo al dominio de tiempo utilizando, por ejemplo, la IDCT (DCT inversa) 220. Finalmente, la señal sintetizada es calculada en el descodificador filtrando la señal de excitación total de la IDCT 220, es decir, la excitación mixta en el dominio de tiempo/dominio de frecuencia, por medio del filtro de síntesis LP 113 (figuras 1 y 2).

50 En una realización, si bien las memorias de codificación CELP se actualizan en un esquema por subtramas utilizando solamente la contribución de excitación del dominio de tiempo, se utiliza la excitación total para actualizar estas memorias en los límites de trama. En otra posible implementación, las memorias de codificación CELP se actualizan en un esquema por subtrama y asimismo en los límites de trama utilizando solamente la contribución de excitación del dominio de tiempo. Esto tiene como resultado una estructura integrada donde la señal cuantificada en el dominio de frecuencia constituye una capa de cuantificación superior independiente de la capa CELP central. Esto

presenta ventajas en determinadas aplicaciones. En este caso particular, el libro de códigos fijo se utiliza siempre para mantener una buena calidad perceptual, y el número de subtramas es siempre de cuatro (4) por la misma razón. No obstante, el análisis en el dominio de frecuencia puede aplicar a toda la trama. Este enfoque integrado funciona para tasas de bits en torno 12 kbps y superiores.

- 5 La discusión anterior se refiere a realizaciones ilustrativas, no restrictivas, y estas realizaciones se pueden modificar a voluntad, dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia para codificar una señal de sonido de entrada (101), **caracterizado por que** comprende:
- 5 un calculador (105) de una contribución de excitación del dominio de tiempo en respuesta a la señal de sonido de entrada (101);
- un calculador (215) de una frecuencia de corte para la contribución de excitación del dominio de tiempo en respuesta a la señal de sonido de entrada (101);
- un filtro (216) sensible la frecuencia de corte, para ajustar una extensión de frecuencia de la contribución de excitación del dominio de tiempo;
- 10 un calculador (107) de una contribución de excitación del dominio de frecuencia en respuesta a la señal de sonido de entrada (101); y
- un sumador (111) de la contribución de excitación del dominio de tiempo filtrada y la contribución de excitación del dominio de frecuencia en el dominio de frecuencia, para formar una excitación mixta en el dominio de tiempo/dominio de frecuencia que constituye una versión codificada de la señal de sonido de entrada (101).
- 15 2. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 1, **caracterizado por que** la contribución de excitación del dominio de tiempo incluye (a) solamente una contribución de libro de códigos adaptativo, o (b) una contribución de libro de códigos adaptativo y una contribución de libro de códigos fijo.
3. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 1 o 2, **caracterizado por que** comprende un calculador (210) de un número de subtramas a utilizar en una trama actual, el calculador (210) del número de subtramas en la trama actual es sensible a por lo menos uno de un presupuesto de bits disponible y una dinámica espectral de alta frecuencia de la señal de sonido de entrada (101), y el calculador (105) de la contribución de excitación del dominio de tiempo utiliza en la trama actual el número de subtramas determinado por el calculador del número de subtramas (210) para dicha trama actual.
- 20 4. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 1 a 3, **caracterizado por que** el calculador (107) de la contribución de excitación del dominio de frecuencia realiza una transformada de frecuencia (213) de un residuo LP obtenido a partir de un análisis LP (201) de la señal de sonido de entrada (101) para producir una representación en frecuencia del residuo LP.
- 25 5. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 4, **caracterizado por que** el calculador (215) de la frecuencia de corte comprende un calculador (303) de correlación cruzada, para cada una de una serie de bandas de frecuencia, entre la representación en frecuencia del residuo LP y una representación en frecuencia de la contribución de excitación del dominio de tiempo, y el dispositivo de codificación comprende un determinador (408) de una estimación de la frecuencia de corte, en respuesta a la correlación cruzada.
- 30 6. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 4 o 5, **caracterizado por que** comprende un elemento de suavizado (304) de la correlación cruzada a través de las bandas de frecuencia para producir un vector de correlación cruzada, un calculador (305) de un promedio del vector de correlación cruzada sobre las bandas de frecuencia y un normalizador (407) del promedio del vector de correlación cruzada, y el determinador (408) de la estimación de la frecuencia de corte determina una primera estimación de la frecuencia de corte encontrando una última frecuencia de una de las bandas de frecuencia que minimiza una diferencia entre dicha última frecuencia y el promedio normalizado del vector de correlación cruzada multiplicado por un valor de anchura del espectro.
- 35 7. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 6, **caracterizado por que** el calculador (215) de la frecuencia de corte comprende un determinador (409) de una de las bandas de frecuencia en la que está situado un armónico calculado a partir de la contribución de excitación del dominio de tiempo, y un selector (411) de la frecuencia de corte como la frecuencia superior entre dicha primera estimación de la frecuencia de corte y una última frecuencia de la banda de frecuencia en la que está situado dicho armónico.
- 45 8. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 1 a 7, **caracterizado por que** el filtro (216) comprende un reductor a cero (418) de cajas de frecuencia, que fuerza la reducción a cero de las cajas de frecuencia de una serie de bandas de frecuencia por encima de la frecuencia de corte.
- 50

9. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 1 a 8, **caracterizado por que** el filtro (216) comprende un reductor a cero (417) de cajas de frecuencia, que fuerza la reducción a cero de todas las cajas de frecuencia de una serie de bandas de frecuencia cuando la frecuencia de corte está por debajo de un determinado valor.
- 5 10. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 1 a 9, **caracterizado por que** el calculador (107) de la contribución de excitación del dominio de frecuencia comprende un calculador (109) de una diferencia entre una representación en frecuencia de un residuo LP de la señal de sonido de entrada (101) y una representación en frecuencia filtrada de la contribución de excitación del dominio de tiempo.
- 10 11. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 4, **caracterizado por que** el calculador (107) de la contribución de excitación del dominio de frecuencia comprende un calculador (109) de la diferencia entre la representación en frecuencia del residuo LP y la representación en frecuencia de la contribución de excitación del dominio de tiempo hasta la frecuencia de corte para formar una primera parte de un vector de diferencia, se aplica un factor de reducción de escala (603) a la representación en frecuencia de la contribución de excitación del dominio de tiempo en un intervalo de frecuencia determinado a continuación de la frecuencia de corte para formar una segunda parte del vector de diferencia, y el vector de diferencia se forma mediante la representación en frecuencia (604) del residuo LP para una tercera parte restante por encima del intervalo de frecuencia determinado.
- 15 12. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 11, **caracterizado por que** comprende un cuantificador (110) del vector de diferencia, y el sumador (111) suma, en el dominio de frecuencia, el vector de diferencia cuantificado y una versión transformada en frecuencia de la contribución de excitación del dominio de tiempo, filtrada, para formar la excitación mixta en el dominio de tiempo/dominio de frecuencia.
- 20 13. Un dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 1 a 12, **caracterizado por que** comprende medios para asignar dinámicamente un presupuesto de bits entre una contribución de excitación del dominio de tiempo y la contribución de excitación del dominio de frecuencia.
- 25 14. Un codificador (100) que utiliza el modelo de dominio de tiempo y dominio de frecuencia, **caracterizado por que** comprende:
- 30 un clasificador (204) de una señal de sonido de entrada (101) como de voz o no de voz;
- un codificador solamente del dominio de tiempo (104);
- el dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 1 a 13; y
- 35 un selector (103) de uno del codificador solamente del dominio de tiempo y el dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia para codificar la señal de sonido de entrada (101) en función de la clasificación de la señal de sonido de entrada.
- 40 15. Un codificador según la reivindicación 14, **caracterizado por que** comprende un selector (206) de un modo de codificación en el dominio de tiempo sin memoria que, cuando el clasificador (204) clasifica la señal de sonido de entrada (101) como no de voz y detecta un ataque temporal en la señal de sonido de entrada (101), fuerza el modo de codificación en el dominio de tiempo sin memoria para codificar la señal de sonido de entrada (101) en el codificador solamente del dominio de tiempo (207).
- 45 16. Un descodificador para descodificar una señal de sonido codificada utilizando el dispositivo de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 1 a 13, **caracterizado por que** comprende:
- un convertidor de la excitación mixta en el dominio de tiempo/dominio de frecuencia, según cualquiera de las reivindicaciones 1 a 13, al dominio de tiempo; y
- un filtro de síntesis para sintetizar la señal de sonido en respuesta a la excitación mixta en el dominio de tiempo/dominio de frecuencia convertida al dominio de tiempo.
- 50 17. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia para codificar una señal de sonido de entrada (101), **caracterizado por que** comprende:
- calcular (105) una contribución de excitación del dominio de tiempo en respuesta a la señal de sonido de entrada (101);

calcular (215) una frecuencia de corte para la contribución de excitación del dominio de tiempo en respuesta a la señal de sonido de entrada (101);

en respuesta a la frecuencia de corte, ajustar (216) una extensión de frecuencia de la contribución de excitación del dominio de tiempo;

5 calcular (107) una contribución de excitación del dominio de frecuencia en respuesta a la señal de sonido de entrada (101); y

sumar (111) la contribución de excitación del dominio de tiempo ajustada y la contribución de excitación del dominio de frecuencia en el dominio de frecuencia para formar una excitación mixta en el dominio de tiempo/dominio de frecuencia que constituye una versión codificada de la señal de sonido de entrada (101).

10 18. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 17, **caracterizado por que** la contribución de excitación del dominio de tiempo incluye (a) solamente una contribución de libro de códigos adaptativo, o (b) una contribución de libro de códigos adaptativo y una contribución de libro de códigos fijo.

15 19. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 17 o 18, **caracterizado por que** comprende calcular (210) un número de subtramas para utilizar en una trama actual en respuesta a por lo menos uno de un presupuesto de bits disponible y una dinámica espectral de alta frecuencia de la señal de sonido de entrada (101), y calcular (105) la contribución de excitación del dominio de tiempo comprende utilizar en la trama actual el número de subtramas determinado para dicha trama actual.

20 20. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 19, **caracterizado por que** calcular (107) la contribución de excitación del dominio de frecuencia comprende llevar a cabo una transformada de frecuencia (213) de un residuo LP obtenido a partir de un análisis LP de la señal de sonido de entrada (101) para producir una representación en frecuencia del residuo LP.

25 21. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 20, **caracterizado por que** calcular (215) la frecuencia de corte comprende calcular (303) una correlación cruzada, para cada una de una serie de bandas de frecuencia, entre la representación en frecuencia del residuo LP y una representación en frecuencia de la contribución de excitación del dominio de tiempo, y el procedimiento de codificación comprende encontrar (408) una estimación de la frecuencia de corte en respuesta a la correlación cruzada.

30 22. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 21, **caracterizado por que** comprende suavizar (304) la correlación cruzada a través de las bandas de frecuencia para producir un vector de correlación cruzada, calcular (305) un promedio del vector de correlación cruzada sobre las bandas de frecuencia y normalizar (407) el promedio de el vector de correlación cruzada, y encontrar (408) la estimación de la frecuencia de corte comprende determinar una primera estimación de la frecuencia de corte encontrando una última frecuencia de una de las bandas de frecuencia que minimiza la diferencia entre dicha última frecuencia y el promedio normalizado del vector de correlación cruzada multiplicado por un valor de anchura del espectro.

35 23. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 22, **caracterizado por que** calcular (215) la frecuencia de corte comprende encontrar (409) una de las bandas de frecuencia en las que está situado un armónico calculado a partir de la contribución de excitación del dominio de tiempo, y seleccionar (411) la frecuencia de corte como la frecuencia superior entre dicha primera estimación de la frecuencia de corte y una última frecuencia de la banda de frecuencia en la que está situado dicho armónico.

40 24. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 23, **caracterizado por que** ajustar (216) la extensión de frecuencia de la contribución de excitación del dominio de tiempo comprende reducir a cero (418) cajas de frecuencia para forzar a cero las cajas de frecuencia de una serie de bandas de frecuencia por encima de la frecuencia de corte.

45 25. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 24, **caracterizado por que** ajustar (216) la extensión de frecuencia de la contribución de excitación del dominio de tiempo comprende reducir a cero (417) cajas de frecuencia para forzar a cero todas las cajas de frecuencia de una serie de bandas de frecuencia cuando la frecuencia de corte está por debajo de un determinado valor.

50 26. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 25, **caracterizado por que** calcular (107) la contribución de excitación del dominio de frecuencia comprende calcular (109) una diferencia entre una representación en frecuencia de un residuo LP de la señal de sonido de entrada (101) y una representación en frecuencia filtrada de la contribución de excitación del dominio de tiempo.

55

27. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 25, **caracterizado por que** calcular (107) la contribución de excitación del dominio de frecuencia comprende calcular (109) la diferencia entre la representación en frecuencia del residuo LP y la representación en frecuencia de la contribución de excitación del dominio de tiempo hasta la frecuencia de corte para formar una primera parte de un vector de diferencia, se aplica un factor de reducción de escala (603) a la representación en frecuencia de la contribución de excitación del dominio de tiempo en un intervalo de frecuencia determinado a continuación de la frecuencia de corte para formar una segunda parte del vector de diferencia, y el vector de diferencia se forma con la representación en frecuencia (604) del residuo LP para una tercera parte restante por encima del intervalo de frecuencia determinado.
28. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según la reivindicación 27, **caracterizado por que** comprende cuantificar (110) el vector de diferencia, y sumar (111) la contribución de excitación del dominio de tiempo ajustada y la contribución de excitación del dominio de frecuencia para formar la excitación mixta en el dominio de tiempo/dominio de frecuencia comprende sumar, en el dominio de frecuencia, el vector de diferencia cuantificado y una versión transformada en frecuencia de la contribución de excitación del dominio de tiempo ajustada.
29. Un procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 28, **caracterizado por que** comprende medios para asignar dinámicamente un presupuesto de bits entre una contribución de excitación del dominio de tiempo y la contribución de excitación del dominio de frecuencia.
30. Un procedimiento (100) de codificación que utiliza un modelo de dominio de tiempo y dominio de frecuencia, **caracterizado por que** comprende:
 clasificar (204) una señal de sonido de entrada como de voz o no de voz;
 proporcionar un procedimiento de codificación solamente en el dominio de tiempo (104);
 proporcionar el procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 29; y
 seleccionar (103) uno del procedimiento de codificación solamente en el dominio de tiempo y el procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia, para codificar la señal de sonido de entrada (101), en función de la clasificación de la señal de sonido de entrada (101).
31. Un procedimiento de codificación según se define en la reivindicación 30, **caracterizado por que** comprende seleccionar (206) un modo de codificación en el dominio de tiempo sin memoria que, cuando la señal de sonido de entrada (101) se clasifica (204) como no de voz y se detecta (208) un ataque temporal en la señal de sonido de entrada (101), fuerza el modo de codificación en el dominio de tiempo sin memoria para codificar la señal de sonido de entrada (101) utilizando el procedimiento de codificación solamente en el dominio de tiempo (207).
32. Un procedimiento de descodificación de una señal de sonido codificada utilizando el procedimiento de codificación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 31, **caracterizado por que** comprende:
 convertir la excitación mixta en el dominio de tiempo/dominio de frecuencia según cualquiera de las reivindicaciones 17 a 31 al dominio de tiempo; y
 sintetizar la señal de sonido a través de un filtro de síntesis en respuesta a la excitación mixta en el dominio de tiempo/dominio de frecuencia convertida al dominio de tiempo.

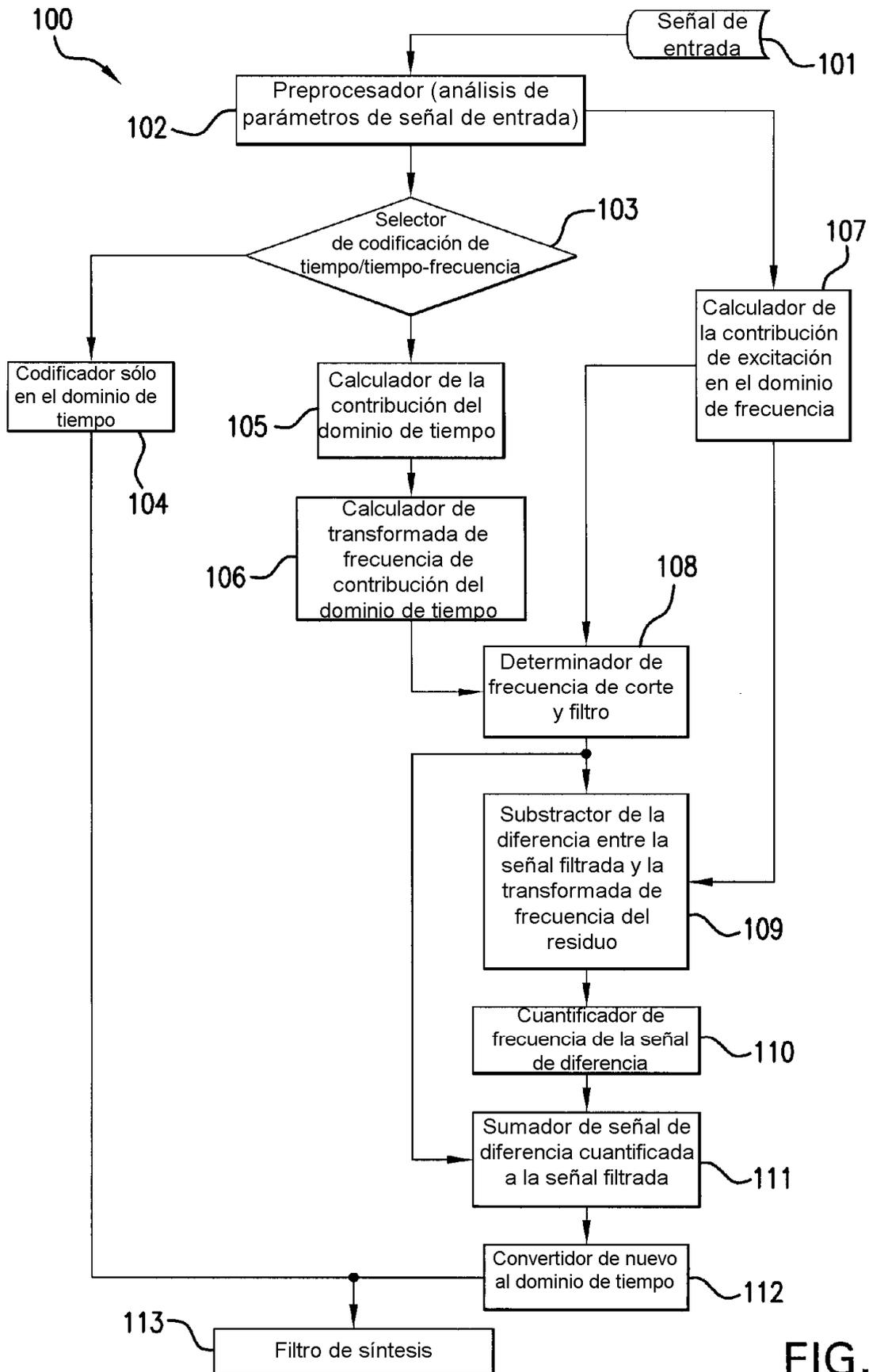


FIG. 1

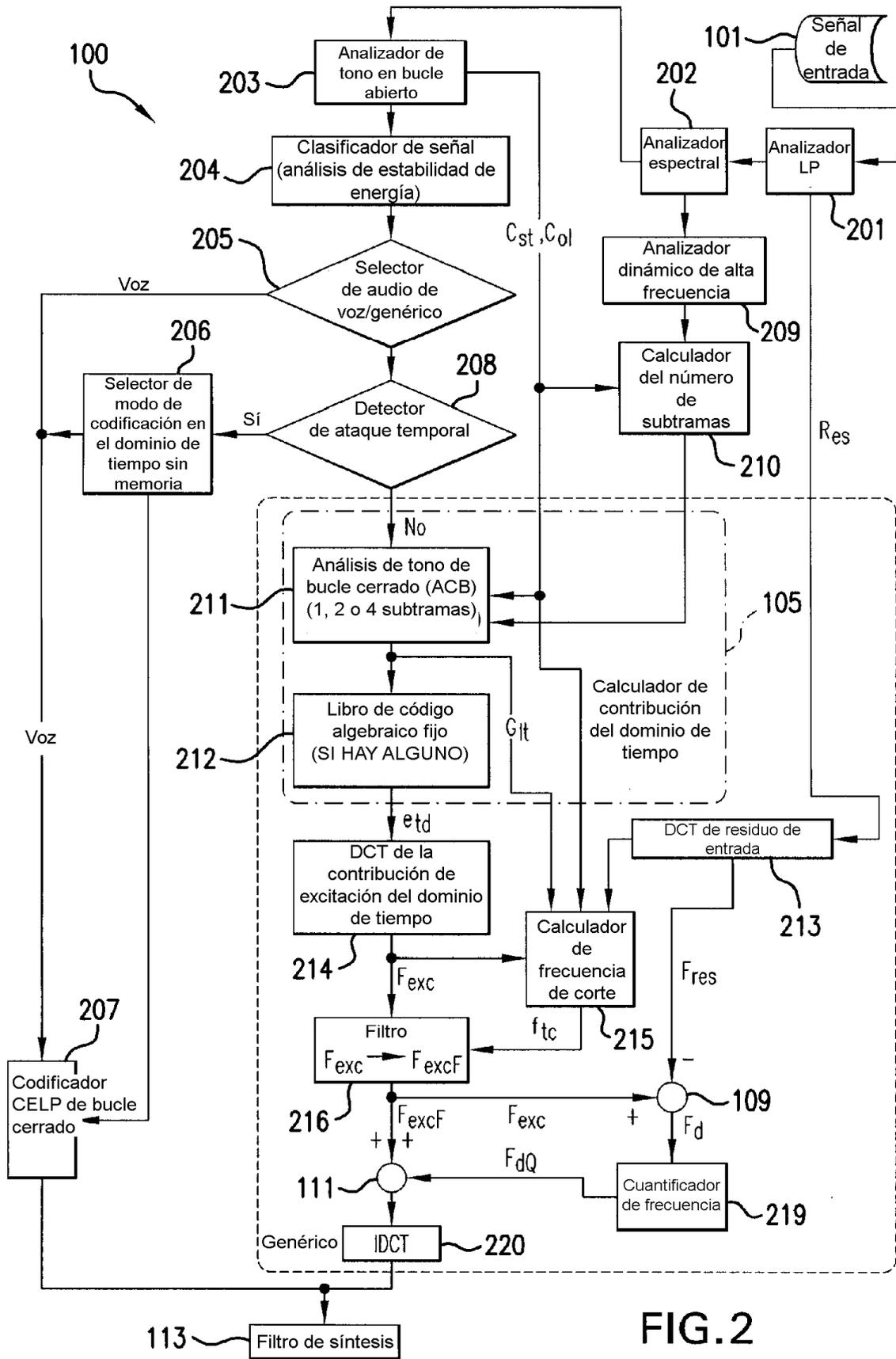


FIG. 2

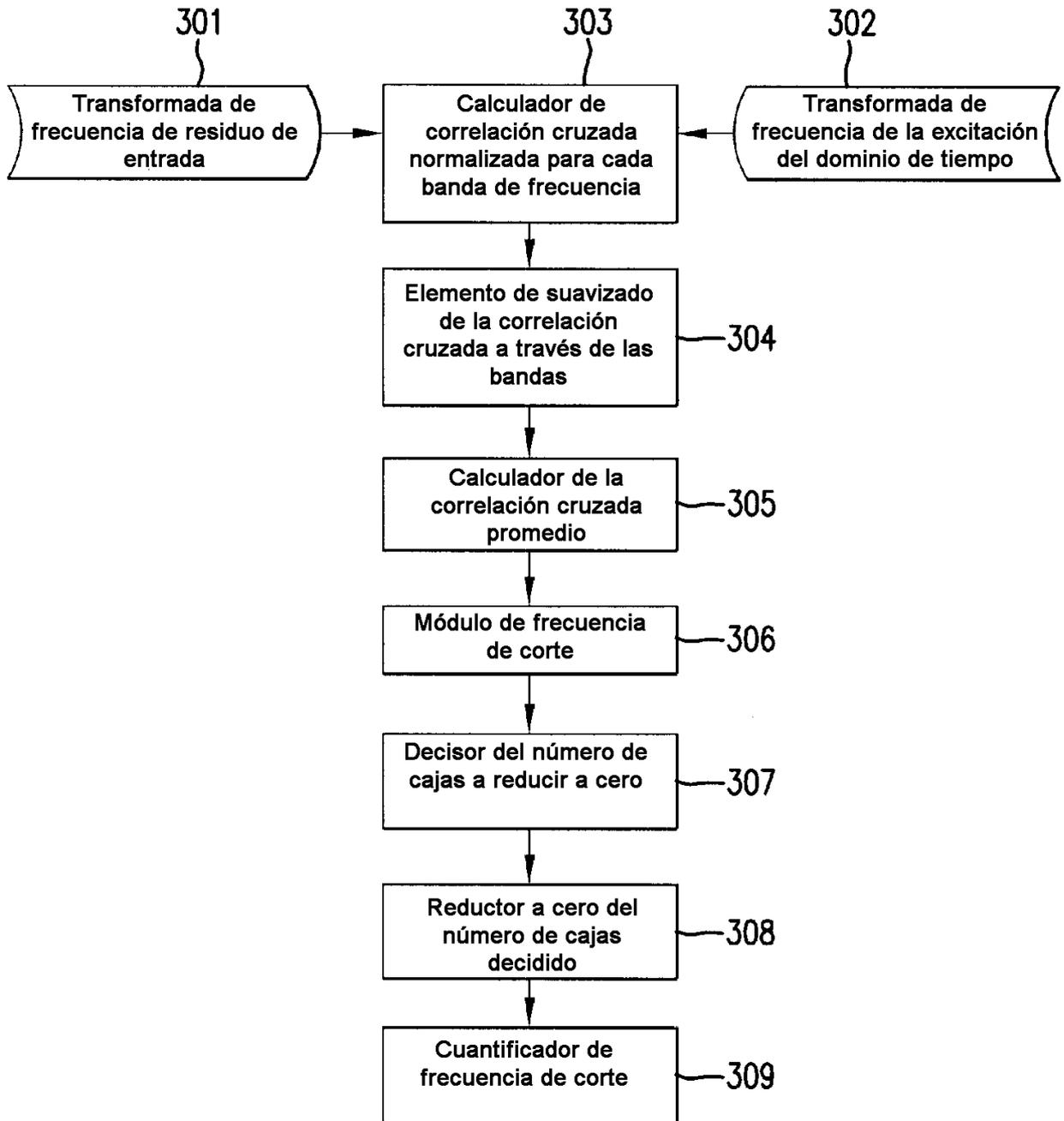


FIG. 3

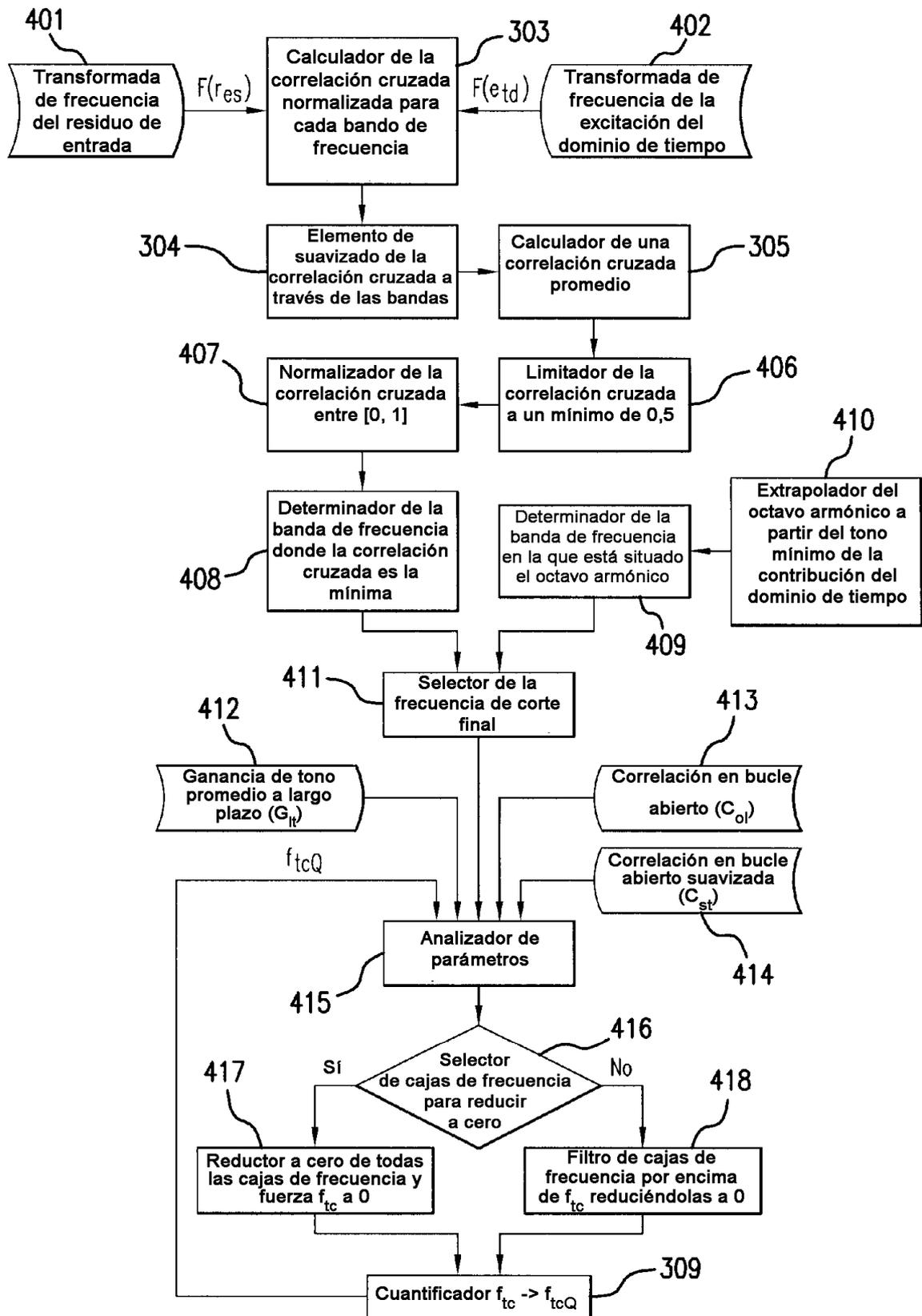


FIG. 4

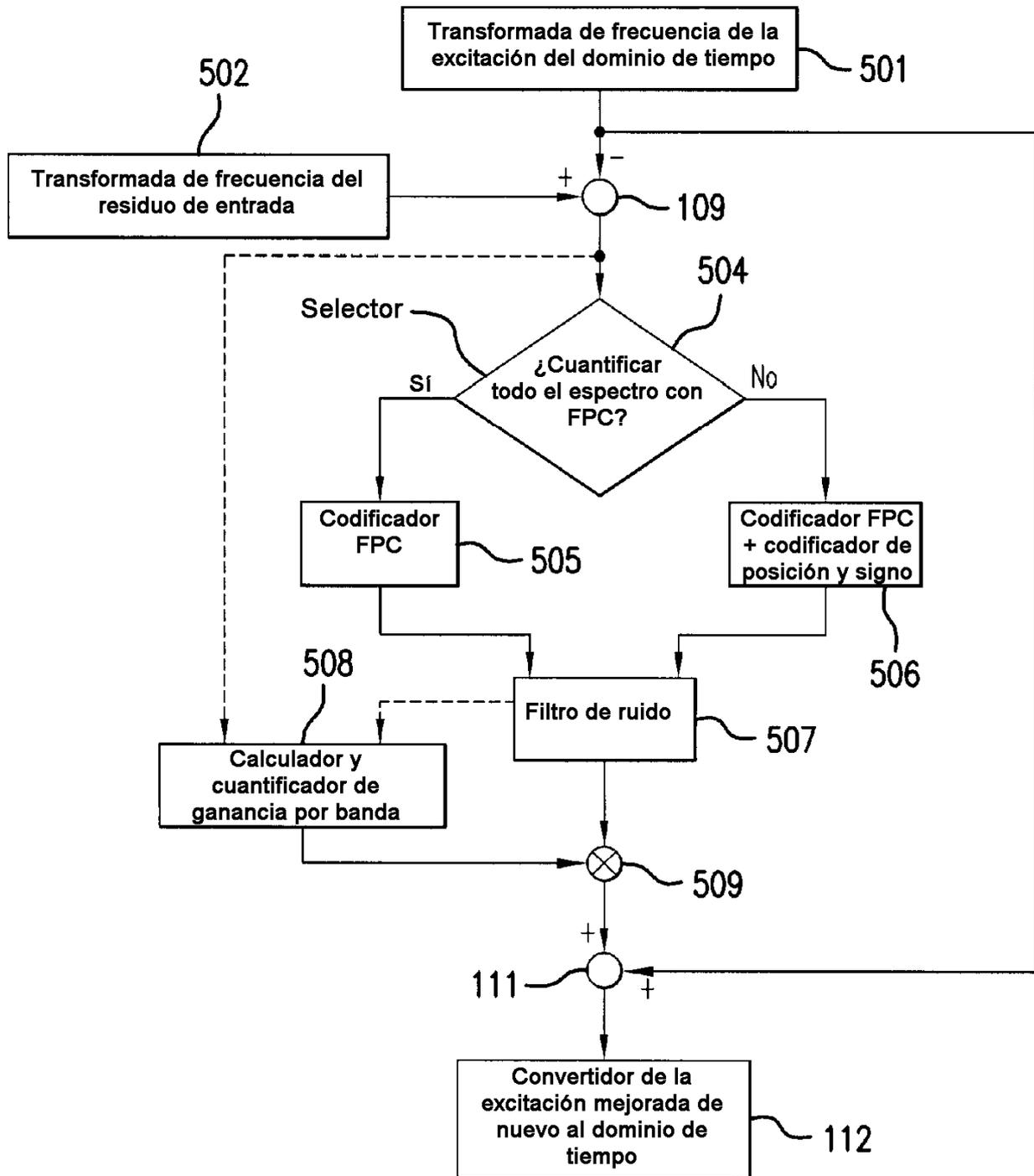


FIG.5

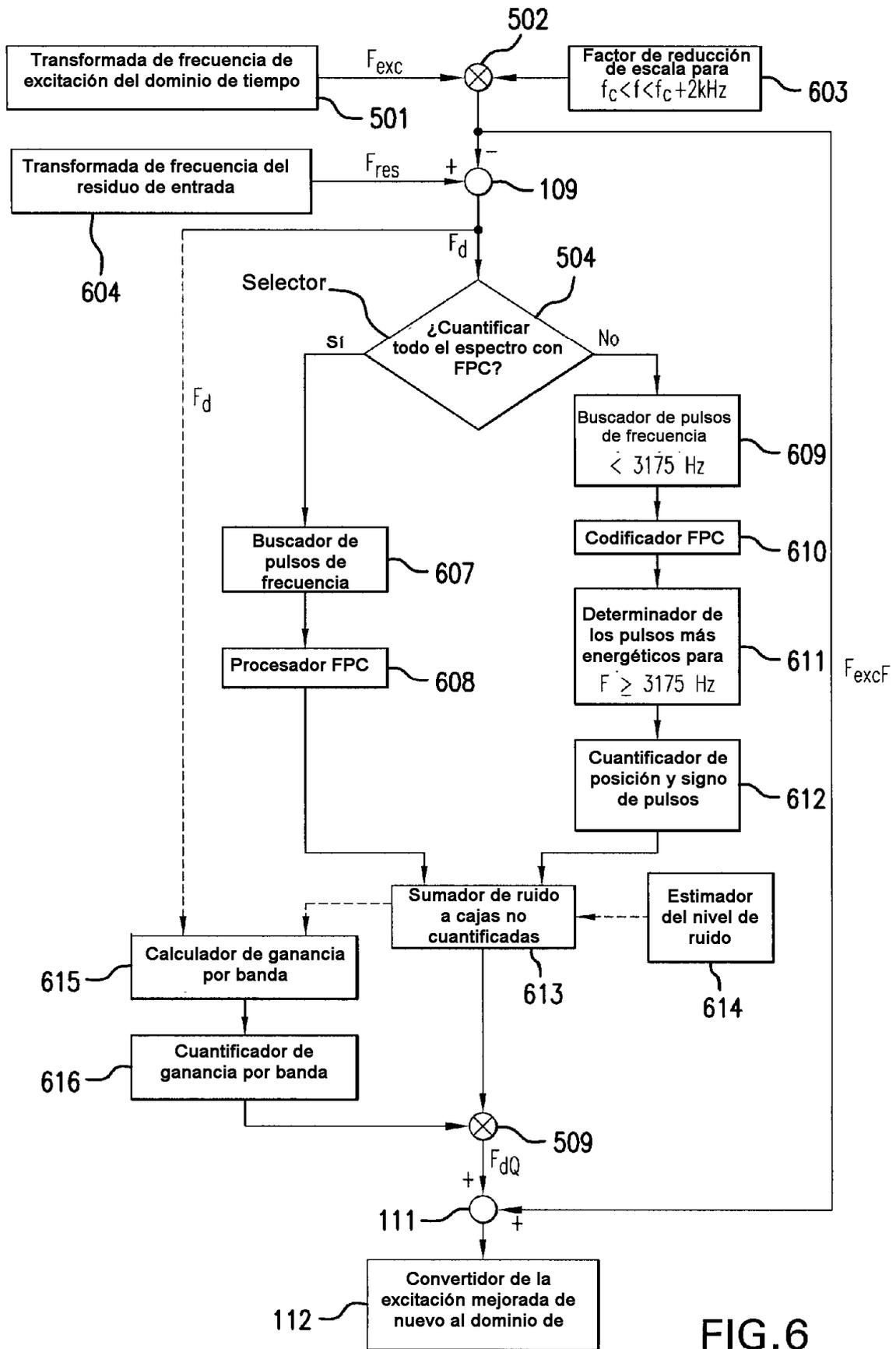


FIG. 6