

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 665 599**

51 Int. Cl.:

G10L 19/02 (2013.01)

G10L 19/06 (2013.01)

G10L 19/032 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.04.2014 PCT/EP2014/056851**

87 Fecha y número de publicación internacional: **09.10.2014 WO14161991**

96 Fecha de presentación y número de la solicitud europea: **04.04.2014 E 14715307 (6)**

97 Fecha y número de publicación de la concesión europea: **07.03.2018 EP 2981958**

54 Título: **Codificador y decodificador de audio**

30 Prioridad:

05.04.2013 US 201361808675 P
09.09.2013 US 201361875553 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.04.2018

73 Titular/es:

DOLBY INTERNATIONAL AB (100.0%)
Apollo Building, 3E, Herikerbergweg 1-35
1101 CN Amsterdam, NL

72 Inventor/es:

VILLEMoes, LARS;
KLEJSA, JANUSZ y
HEDELIN, PER

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 665 599 T3

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

DESCRIPCIÓN

Codificador y decodificador de audio

5 Campo técnico

El presente documento se refiere a un sistema de codificación y de decodificación de audio (denominado sistema de códecs de audio). En particular, el presente documento se refiere a un sistema de códecs de audio basado en transformada que, en particular, es muy adecuado para la codificación/decodificación de voz.

10

Antecedentes

Los codificadores de audio perceptuales de propósito general consiguen ganancias de codificación relativamente altas usando transformadas tales como la transformada discreta del coseno modificada (MDCT) con tamaños de bloque de muestras que abarcan varias decenas de milisegundos (por ejemplo, 20 ms). Un ejemplo de un sistema de códecs de audio basado en transformada de este tipo es la codificación de audio avanzada (AAC) o la AAC de alta eficacia (HE-ACC). Sin embargo, cuando se usan tales sistemas de códec de audio basados en transformada para señales de voz, la calidad de las señales de voz se degrada más rápidamente que la calidad de las señales musicales, llegando a velocidades binarias más bajas, especialmente en el caso de señales de voz seca (no reverberante). Por tanto, los sistemas de códecs de audio basados en transformada no son muy adecuados, en esencia, para la codificación de señales de voz o para la codificación de señales de audio que comprenden una componente de voz. Dicho de otro modo, los sistemas de códecs de audio basados en transformada presentan una asimetría en lo que respecta a la ganancia de codificación conseguida para las señales musicales en comparación con la ganancia de codificación conseguida para las señales de voz. Esta asimetría puede afrontarse añadiendo complementos a la codificación basada en transformada, donde los complementos tienen como objetivo una conformación espectral o una adaptación de señal mejoradas. Ejemplos de tales complementos son una conformación previa/posterior, una conformación de ruido temporal (TNS) y una MDCT deformada en el tiempo. Además, esta asimetría puede afrontarse mediante la incorporación de un codificador de voz clásico de dominio de tiempo basado en filtrado de predicción a corto plazo (LPC) o en predicción a largo plazo (LTP).

15

20

25

30

Puede observarse que las mejoras obtenidas proporcionando complementos a la codificación basada en transformada no son generalmente suficientes para igualar la diferencia de rendimiento entre la codificación de señales de música y de señales de voz. Por otro lado, la incorporación de un codificador de voz clásico de dominio de tiempo atenúa la diferencia de rendimiento hasta el punto en que la asimetría de rendimiento se invierte hacia el sentido contrario. Esto se debe a que los codificadores de voz clásicos de dominio de tiempo modelan el sistema de producción de voz humana y se han optimizado para la codificación de señales de voz.

35

En vista de lo anterior, puede usarse un códec de audio basado en transformada en combinación con un códec de voz clásico de dominio de tiempo, donde el códec de voz clásico de dominio de tiempo se usa para segmentos de voz de una señal de audio y donde el códec basado en transformada se usa para los segmentos restantes de la señal de audio. Sin embargo, la coexistencia de un códec de dominio de tiempo y de dominio de transformada en un único sistema de códecs de audio requiere herramientas fiables para la conmutación entre los diferentes códecs, en función de las propiedades de la señal de audio. Además, la conmutación real entre un códec de dominio de tiempo (para contenido de voz) y un códec de dominio de transformada (para el contenido restante) puede ser difícil de implementar. En particular, puede ser difícil garantizar una transición suave entre el códec de dominio de tiempo y el códec de dominio de transformada (y viceversa). Además, pueden requerirse modificaciones en el códec de dominio de tiempo para hacer que el códec de dominio de tiempo sea más robusto en lo que respecta a la inevitable codificación ocasional de señales que no son de voz, por ejemplo en la codificación de una voz que canta con un fondo instrumental. El presente documento afronta los problemas técnicos mencionados anteriormente de los sistemas de códecs de audio. En particular, el presente documento describe un sistema de códecs de audio que convierte solamente las características críticas de un códec de voz y, por lo tanto, consigue un rendimiento uniforme para la voz y la música manteniendo la arquitectura de un códec basado en transformada. Dicho de otro modo, el presente documento describe un códec de audio basado en transformada que es, en particular, muy adecuado para la codificación de señales de habla o de voz.

40

45

50

55

El documento de la técnica anterior titulado "*Digital Audio Coding: Dolby AC-3*" describe un codificador/decodificador de voz basado en transformada, que proporciona información de envolvente para una trama que consiste en varios bloques secuenciales de coeficientes de transformada.

60

Resumen

Según un aspecto, se describe un codificador de voz basado en transformada según la reivindicación 1 (codificador) y la reivindicación 7 (decodificador).

65

Breve descripción de las figuras

La invención se explica a continuación a modo de ejemplo con referencia a los dibujos adjuntos, en los que:

- la Fig. 1a muestra un diagrama de bloques de un codificador de audio de ejemplo que proporciona un flujo de bits a una velocidad binaria constante;
- 5 la Fig. 1b muestra un diagrama de bloques de un codificador de audio de ejemplo que proporciona un flujo de bits a una velocidad binaria variable;
- la Fig. 2 ilustra la generación de una envolvente de ejemplo según una pluralidad de bloques de coeficientes de transformada;
- la Fig. 3a ilustra envolventes de ejemplo de bloques de coeficientes de transformada;
- 10 la Fig. 3b ilustra la determinación de una envolvente interpolada de ejemplo;
- la Fig. 4 ilustra conjuntos de cuantificadores de ejemplo;
- la Fig. 5a muestra un diagrama de bloques de un descodificador de audio de ejemplo;
- la Fig. 5b muestra un diagrama de bloques de un descodificador de envolvente de ejemplo del descodificador de audio de la Fig. 5a;
- 15 la Fig. 5c muestra un diagrama de bloques de un predictor de subbanda de ejemplo del descodificador de audio de la Fig. 5a; y
- la Fig. 5d muestra un diagrama de bloques de un descodificador de espectro de ejemplo del descodificador de audio de la Fig. 5a.

Descripción detallada

20 Como se ha indicado en la sección de los antecedentes, es deseable proporcionar un códec de audio basado en transformada que tenga ganancias de codificación relativamente altas para señales de habla o de voz. Un códec de audio basado en transformada de este tipo puede denominarse códec de habla basado en transformada o códec de voz basado en transformada. Un códec de voz basado en transformada puede combinarse de manera conveniente con un códec de audio genérico basado en transformada, tal como un AAC o un HE-AAC, ya que también funciona en el dominio de transformada. Además, la clasificación de un segmento (por ejemplo, una trama) de una señal de audio de entrada como de voz o de no voz y la subsiguiente conmutación entre el códec de audio genérico y el códec de voz específico pueden simplificarse, debido a que ambos códecs funcionan en el dominio de transformada.

30 La Fig. 1a muestra un diagrama de bloques de un codificador de voz basado en transformada 100 de ejemplo. El codificador 100 recibe como entrada un bloque 131 de coeficientes de transformada (también denominado unidad de codificación). El bloque 131 del coeficiente de transformada puede haberse obtenido mediante una unidad de transformada configurada para transformar una secuencia de muestras de la señal de audio de entrada desde el dominio de tiempo al dominio de transformada. La unidad de transformada puede configurarse para realizar una MDCT. La unidad de transformada puede ser parte de un códec de audio genérico, tal como un AAC o un HE-AAC. 35 Tal códec de audio genérico puede usar diferentes tamaños de bloque, por ejemplo un bloque largo y un bloque corto. Tamaños de bloque de ejemplo son 1024 muestras para un bloque largo y 256 muestras para un bloque corto. Suponiendo una frecuencia de muestreo de 44,1 kHz y un solapamiento del 50%, un bloque largo abarca 20 ms aproximadamente de la señal de audio de entrada, y un bloque corto abarca 5 ms aproximadamente de la señal de audio de entrada. Los bloques largos se usan normalmente para segmentos estacionarios de la señal de audio de entrada, y los bloques cortos se usan normalmente para segmentos transitorios de la señal de audio de entrada.

45 Las señales de voz pueden considerarse estacionarias en segmentos temporales de 20 ms aproximadamente. En particular, la envolvente espectral de una señal de voz puede considerarse estacionaria en segmentos temporales de 20 ms aproximadamente. Para poder obtener estadísticas fiables en el dominio de transformada para tales segmentos de 20 ms, puede ser útil dotar al codificador de voz basado en transformada 100 de bloques cortos 131 de coeficientes de transformada (que tienen una longitud de, por ejemplo, 5 ms). Al hacer esto, una pluralidad de bloques cortos 131 pueden usarse para obtener estadísticas relacionadas con segmentos de tiempo de, por ejemplo, 20 ms (por ejemplo, el segmento de tiempo de un bloque largo o trama). Además, esto tiene la ventaja de 50 proporcionar una resolución de tiempo adecuada para las señales de voz.

Por tanto, la unidad de transformada puede configurarse para proporcionar bloques cortos 131 de coeficientes de transformada, si un segmento actual de la señal de audio de entrada se clasifica como de voz. El codificador 100 puede comprender una unidad de generación de tramas 101 configurada para extraer una pluralidad de bloques 131 de coeficientes de transformada, denominados conjunto 132 de bloques 131. El conjunto 132 de bloques también puede denominarse trama. A modo de ejemplo, el conjunto 132 de bloques 131 puede comprender cuatro bloques 55 cortos de 256 coeficientes de transformada, lo que abarca por tanto, aproximadamente, un segmento de 20 ms de la señal de audio de entrada.

El codificador de voz basado en transformada 100 puede estar configurado para funcionar en una pluralidad de modos diferentes, por ejemplo en un modo de corto espaciado y en modo de largo espaciado. Cuando funciona en el modo de corto espaciado, el codificador de voz basado en transformada 100 puede estar configurado para subdividir un segmento o una trama de la señal de audio (por ejemplo, la señal de voz) en un conjunto 132 de bloques cortos 131 (como se ha indicado anteriormente). Por otro lado, cuando funciona en el modo de largo espaciado, el codificador de voz basado en transformada 100 puede estar configurado para procesar directamente el segmento o la trama de la señal de audio.

A modo de ejemplo, cuando funciona en el modo de corto espaciado, el codificador 100 puede estar configurado para procesar cuatro bloques 131 por trama. Las tramas del codificador 100 pueden ser relativamente cortas en tiempo físico para ciertas configuraciones de una operación síncrona de trama de vídeo. Esto es particularmente el caso de una frecuencia de trama de vídeo aumentada (por ejemplo, 100 Hz frente a 50 Hz), lo que da lugar a una reducción de la longitud temporal del segmento o de la trama de la señal de voz. En tales casos, la subdivisión de la trama en una pluralidad de bloques (cortos) 131 puede ser desventajosa debido a la menor resolución en el dominio de transformada. Por tanto, puede usarse un modo de largo espaciado para utilizar solamente un bloque 131 por trama. El uso de un único bloque 131 por trama también puede ser beneficioso para codificar señales de audio que incluyen música (incluso en tramas relativamente largas). Los beneficios pueden deberse a la mayor resolución en el dominio de transformada cuando solo se usa un único bloque 131 por trama o cuando se usa un número reducido de bloques 131 por trama.

A continuación se describe en mayor detalle el funcionamiento del codificador 100 en el modo de corto espaciado. El conjunto 132 de bloques puede proporcionarse a una unidad de estimación de envolvente 102. La unidad de estimación de envolvente 102 puede configurarse para determinar una envolvente 133 en función del conjunto 132 de bloques. La envolvente 133 puede estar basada en valores de media cuadrática (RMS) de coeficientes de transformada correspondientes de la pluralidad de bloques 131 comprendidos en el conjunto 132 de bloques. Un bloque 131 proporciona normalmente una pluralidad de coeficientes de transformada (por ejemplo, 256 coeficientes de transformada) en una pluralidad correspondiente de celdas de frecuencia 301 (véase la Fig. 3a). La pluralidad de celdas de frecuencia 301 puede agruparse en una pluralidad de bandas de frecuencia 302. La pluralidad de bandas de frecuencia 302 puede seleccionarse según consideraciones psicoacústicas. A modo de ejemplo, las celdas de frecuencia 301 pueden agruparse en bandas de frecuencia 302 según una escala logarítmica o una escala de Bark. La envolvente 134 que se ha determinado en función de un conjunto actual 132 de bloques puede comprender una pluralidad de valores de energía para la pluralidad de bandas de frecuencia 302, respectivamente. Un valor de energía particular para un banda de frecuencia particular 302 puede determinarse en función de los coeficientes de transformada de los bloques 131 del conjunto 132, que corresponden a celdas de frecuencia 301 que están dentro de la banda de frecuencia particular 302. El valor de energía particular puede determinarse en función del valor RMS de estos coeficientes de transformada. De este modo, una envolvente 133 para un conjunto actual 132 de bloques (denominada envolvente actual 133) puede indicar una envolvente promedio de los bloques 131 de coeficientes de transformada comprendidos en el conjunto actual 132 de bloques, o puede indicar una envolvente promedio de bloques 132 de coeficientes de transformada usados para determinar la envolvente 133.

Debe observarse que la envolvente actual 133 puede determinarse en función de uno o más bloques adicionales 131 de coeficientes de transformada adyacentes al conjunto actual 132 de bloques. Esto se ilustra en la Fig. 2, donde la envolvente actual 133 (indicada por la envolvente actual cuantificada 134) se determina en función de los bloques 131 del conjunto actual 132 de bloques y en función del bloque 201 del conjunto de bloques anterior al conjunto actual 132 de bloques. En el ejemplo ilustrado, la envolvente actual 133 se determina en función de cinco bloques 131. Teniendo en cuenta bloques adyacentes cuando se determina la envolvente actual 133, puede garantizarse la continuidad de las envolventes de conjuntos adyacentes 132 de bloques.

Cuando se determina la envolvente actual 133 pueden ponderarse los coeficientes de transformada de los diferentes bloques 131. En particular, los bloques más externos 201, 202 que se tienen en cuenta para determinar la envolvente actual 133 pueden tener un menor peso que los bloques restantes 131. A modo de ejemplo, los coeficientes de transformada de los bloques más externos 201, 202 pueden ponderarse con 0,5, donde los coeficientes de transformada de los otros bloques 131 pueden ponderarse con 1.

Debe observarse que, de manera similar a la consideración de los bloques 201 de un conjunto previo 132 de bloques, uno o más bloques (denominados bloques de anticipación) de un conjunto directamente posterior 132 de bloques pueden considerarse para determinar la envolvente actual 133.

Los valores de energía de la envolvente actual 133 pueden representarse en una escala logarítmica (por ejemplo, en una escala de dB). La envolvente actual 133 puede proporcionarse a una unidad de cuantificación de envolvente 103 que está configurada para cuantificar los valores de energía de la envolvente actual 133. La unidad de cuantificación de envolvente 103 puede proporcionar una resolución de cuantificador predeterminada, por ejemplo una resolución de 3 dB. Los índices de cuantificación de la envolvente 133 pueden proporcionarse como datos de envolvente 161 en un flujo de bits generado por el codificador 100. Además, la envolvente cuantificada 134, es decir,

la envolvente que comprende los valores de energía cuantificados de la envolvente 133, puede proporcionarse a una unidad de interpolación 104.

5 La unidad de interpolación 104 está configurada para determinar una envolvente para cada bloque 131 del conjunto actual 132 de bloques en función de la envolvente actual cuantificada 134 y en función de la envolvente previa cuantificada 135 (que se ha determinado para el conjunto 132 de bloques directamente anterior al conjunto actual 132 de bloques). El funcionamiento de la unidad de interpolación 104 se ilustra en las Fig. 2, 3a y 3b. La Fig. 2 muestra una secuencia de bloques 131 de coeficientes de transformada. La secuencia de bloques 131 se agrupa en conjuntos subsiguientes 132 de bloques, donde cada conjunto 132 de bloques se usa para determinar una envolvente cuantificada, por ejemplo la envolvente actual cuantificada 134 y la envolvente previa cuantificada 135. La Fig. 3a muestra ejemplos de una envolvente previa cuantificada 135 y de una envolvente actual cuantificada 134. Como se ha indicado anteriormente, las envolventes pueden indicar la energía espectral 303 (por ejemplo, en una escala de dB). Valores de energía 303 correspondientes de la envolvente previa cuantificada 135 y de la envolvente actual cuantificada 134 para la misma banda de frecuencia 302 pueden interpolarse (por ejemplo, usando interpolación lineal) para determinar una envolvente interpolada 136. Dicho de otro modo, los valores de energía 303 de una banda de frecuencia particular 302 pueden interpolarse para proporcionar el valor de energía 303 de la envolvente interpolada 136 en la banda de frecuencia particular 302.

20 Debe observarse que el conjunto de bloques para el que las envolventes interpoladas 136 se determinan y aplican puede ser diferente del conjunto actual 132 de bloques, en función de lo cual se determina la envolvente actual cuantificada 134. Esto se ilustra en la Fig. 2, que muestra un conjunto desplazado 332 de bloques, que está desplazado en comparación con el conjunto actual 132 de bloques y que comprende los bloques 3 y 4 del conjunto previo 132 de bloques (indicados mediante los números de referencia 203 y 201, respectivamente) y los bloques 1 y 2 del conjunto actual 132 de bloques (indicados mediante los números de referencia 204 y 205, respectivamente). De hecho, las envolventes interpoladas 136 determinadas en función de la envolvente actual cuantificada 134 y en función de la envolvente previa cuantificada 135 pueden tener una mayor relevancia para los bloques del conjunto desplazado 332 de bloques, en comparación con la relevancia para los bloques del conjunto actual 132 de bloques.

30 Por tanto, las envolventes interpoladas 136 mostradas en la Fig. 3b pueden usarse para aplanar los bloques 131 del conjunto desplazado 332 de bloques. Esto se muestra en la Fig. 3b en combinación con la Fig. 2. Puede observarse que la envolvente interpolada 341 de la Fig. 3b puede aplicarse al bloque 203 de la Fig. 2, que la envolvente interpolada 342 de la Fig. 3b puede aplicarse al bloque 201 de la Fig. 2, que la envolvente interpolada 343 de la Fig. 3b puede aplicarse al bloque 204 de la Fig. 2, y que la envolvente interpolada 344 de la Fig. 3b (que en el ejemplo ilustrado corresponde a la envolvente actual cuantificada 136) puede aplicarse al bloque 205 de la Fig. 2. De este modo, el conjunto 132 de bloques para determinar la envolvente actual cuantificada 134 puede ser diferente del conjunto desplazado 332 de bloques para el que se determinan las envolventes interpoladas 136 y al que se aplican las envolventes interpoladas 136 (con fines de aplanamiento). En particular, la envolvente actual cuantificada 134 puede determinarse usando una cierta anticipación con respecto a los bloques 203, 201, 204, 205 del conjunto desplazado 332 de bloques, los cuales van a aplanarse usando la envolvente actual cuantificada 134. Esto es beneficioso desde el punto de vista de la continuidad.

45 La interpolación de los valores de energía 303 para determinar envolventes interpoladas 136 se ilustra en la Fig. 3b. Puede observarse que mediante la interpolación entre un valor de energía de la envolvente previa cuantificada 135 y el valor de energía correspondiente de la envolvente actual cuantificada 134, pueden determinarse valores de energía de las envolventes interpoladas 136 para los bloques 131 del conjunto desplazado 332 de bloques. En particular, para cada bloque 131 del conjunto desplazado 332 puede determinarse una envolvente interpolada 136, lo que proporciona una pluralidad de envolventes interpoladas 136 para la pluralidad de bloques 203, 201, 204, 205 del conjunto desplazado 332 de bloques. La envolvente interpolada 136 de un bloque 131 de coeficiente de transformada (por ejemplo, cualquiera de los bloques 203, 201, 204, 205 del conjunto desplazado 332 de bloques) puede usarse para codificar el bloque 131 de coeficientes de transformada. Debe observarse que los índices de cuantificación 161 de la envolvente actual 133 se proporcionan a un descodificador correspondiente en el flujo de bits. Por consiguiente, el descodificador correspondiente puede estar configurado para determinar la pluralidad de envolventes interpoladas 136 de una manera análoga con respecto a la unidad de interpolación 104 del codificador 100.

55 La unidad de generación de tramas 101, la unidad de estimación de envolvente 102, la unidad de cuantificación de envolvente 103 y la unidad de interpolación 104 actúan en un conjunto de bloques (es decir, el conjunto actual 132 de bloques y/o el conjunto desplazado 332 de bloques). Por otro lado, la codificación real del coeficiente de transformada puede realizarse bloque a bloque. A continuación se hace referencia a la codificación de un bloque actual 131 de coeficientes de transformada, que puede ser uno cualquiera de la pluralidad de bloques 131 del conjunto desplazado 332 de bloques (o posiblemente el conjunto actual 132 de bloques en otras implementaciones del codificador de voz basado en transformada 100).

65 Además, debe observarse que el codificador 100 puede hacerse funcionar en el denominado modo de largo espaciado. En este modo, una trama de segmento de la señal de audio no está subdividida y se procesa como un único bloque. Por tanto, solo se determina por trama un único bloque 131 de coeficientes de transformada.

Cuando funciona en el modo de largo espaciamiento, la unidad de generación de tramas 101 puede configurarse para extraer el único bloque actual 131 de coeficientes de transformada para el segmento o la trama de la señal de audio. La unidad de estimación de envolvente 102 puede configurarse para determinar la envolvente actual 133 para el bloque actual 131 y la unidad de cuantificación de envolvente 103 puede configurarse para cuantificar la única envolvente actual 133 para determinar la envolvente actual cuantificada 134 (y para determinar los datos de envolvente 161 para el bloque actual 131). Durante el modo de largo espaciamiento, la interpolación de envolvente está normalmente obsoleta. Por tanto, la envolvente interpolada 136 para el bloque actual 131 corresponde normalmente a la envolvente actual cuantificada 134 (cuando el codificador 100 está funcionando en el modo de largo espaciamiento).

La envolvente interpolada actual 136 para el bloque actual 131 puede proporcionar una aproximación de la envolvente espectral de los coeficientes de transformada del bloque actual 131. El codificador 100 puede comprender una unidad de aplanamiento previo 105 y una unidad de determinación de ganancia de envolvente 106 que están configuradas para determinar una envolvente ajustada 139 para el bloque actual 131 en función de la envolvente interpolada actual 136 y en función del bloque actual 131. En particular, una ganancia de envolvente para el bloque actual 131 puede determinarse de tal manera que se ajuste la varianza de los coeficientes de transformada aplanados del bloque actual 131. $X(k)$, $k = 1, \dots, K$ pueden ser los coeficientes de transformada del bloque actual 131 (donde, por ejemplo, $K = 256$), y $E(k)$, $k = 1, \dots, K$ pueden ser los valores de energía espectral promedio 303 de la envolvente interpolada actual 136 (donde los valores de energía $E(k)$ de una misma banda de frecuencia 302 son iguales). La ganancia de envolvente a puede determinarse de modo que se ajuste la varianza de los coeficientes de

transformada aplanados $\tilde{X}(k) = \frac{X(k)}{a \sqrt{E(k)}}$. En particular, la ganancia de envolvente a puede determinarse de manera que la varianza sea uno.

Debe observarse que la ganancia de envolvente a puede determinarse para un subintervalo del intervalo de frecuencias completo del bloque actual 131 de los coeficientes de transformada. Dicho de otro modo, la ganancia de envolvente a puede determinarse solamente en función de un subconjunto de las celdas de frecuencia 301 y/o solamente en función de un subconjunto de las bandas de frecuencia 302. A modo de ejemplo, la ganancia de envolvente a puede determinarse en función de las celdas de frecuencia 301 mayores que una celda de frecuencia de inicio 304 (siendo la celda de frecuencia de inicio mayor que 0 o 1). En consecuencia, la envolvente ajustada 139 para el bloque actual 131 puede determinarse aplicando la ganancia de envolvente a solamente a los valores de energía espectral promedio 303 de la envolvente interpolada actual 136 que están asociados a las celdas de frecuencia 301 que están por encima de la celda de frecuencia de inicio 304. Por tanto, la envolvente ajustada 139 para el bloque actual 131 puede corresponder a la envolvente interpolada actual 136, para celdas de frecuencia 301 en y por debajo de la celda de frecuencia de inicio, y puede corresponder a la envolvente interpolada actual 136 desfasada por la ganancia de envolvente a , para celdas de frecuencia 301 por encima de la celda de frecuencia de inicio. Esto se ilustra en la Fig. 3a mediante la envolvente ajustada 339 (mostrada en líneas discontinuas).

La aplicación de la ganancia de envolvente a 137 (que también se denomina ganancia de corrección de nivel) en la envolvente interpolada actual 136 corresponde a un ajuste o un desfase de la envolvente interpolada actual 136, lo que da como resultado una envolvente ajustada 139, como se ilustra en la Fig. 3a. La ganancia de envolvente a 137 puede codificarse como datos de ganancia 162 en el flujo de bits.

El codificador 100 puede comprender además una unidad de refinamiento de envolvente 107 que está configurada para determinar la envolvente ajustada 139 en función de la ganancia de envolvente a 137 y en función de la envolvente interpolada actual 136. La envolvente ajustada 139 puede usarse para el procesamiento de señales del bloque 131 de coeficientes de transformada. La ganancia de envolvente a 137 puede cuantificarse a una mayor resolución (por ejemplo, en etapas de 1 dB) en comparación con la envolvente interpolada actual 136 (que puede cuantificarse en etapas de 3dB). De este modo, la envolvente ajustada 139 puede cuantificarse a la mayor resolución de la ganancia de envolvente a 137 (por ejemplo, en etapas de 1 dB).

Además, la unidad de refinamiento de envolvente 107 puede configurarse para determinar una envolvente de asignación 138. La envolvente de asignación 138 puede corresponder a una versión cuantificada de la envolvente ajustada 139 (por ejemplo, cuantificada a niveles de cuantificación de 3 dB). La envolvente de asignación 138 puede usarse con fines de asignación de bits. En particular, la envolvente de asignación 138 puede usarse para determinar, para un coeficiente de transformada particular del bloque actual 131, un cuantificador particular a partir de un conjunto predeterminado de cuantificadores, donde el cuantificador particular va a usarse para cuantificar el coeficiente de transformada particular.

El codificador 100 comprende una unidad de aplanamiento 108 configurada para aplanar el bloque actual 131 usando la envolvente ajustada 139, proporcionando así el bloque 140 de coeficientes de transformada aplanados $\tilde{X}(k)$. El bloque 140 de coeficientes de transformada aplanados $\tilde{X}(k)$ pueden codificarse usando un bucle de predicción en el dominio de transformada. De este modo, el bloque 140 puede codificarse usando un predictor de subbanda 117. El bucle de predicción comprende una unidad de diferencia 115 configurada para determinar un bloque 141 de coeficientes de error de predicción $\Delta(k)$ en función del bloque 140 de coeficientes de transformada

aplanados $\tilde{X}(k)$ y en función de un bloque 150 de coeficientes de transformada estimados $\tilde{X}(k)$, por ejemplo $\Delta(k) = \tilde{X}(k) - X(k)$. Debe observarse que debido a que el bloque 140 comprende coeficientes de transformada aplanados, es decir, coeficientes de transformada que se han normalizado o aplanado usando los valores de energía 303 de la envolvente ajustada 139, el bloque 150 de coeficientes de transformada estimados también comprende estimaciones de coeficientes de transformada aplanados. Dicho de otro modo, la unidad de diferencia 115 funciona en el denominado dominio aplanado. En consecuencia, el bloque 141 de coeficientes de error de predicción $\Delta(k)$ se representa en el dominio aplanado. El bloque 141 de coeficientes de error de predicción $\Delta(k)$ puede presentar una varianza diferente al valor uno. El codificador 100 puede comprender una unidad de reescalado 111 configurada para reescalar los coeficientes de error de predicción $\Delta(k)$ para proporcionar un bloque 142 de coeficientes de error reescalados. La unidad de reescalado 111 puede usar una o más reglas heurísticas predeterminadas para realizar el reescalado. Como resultado, el bloque 142 de coeficientes de error reescalados presenta una varianza que (por término medio) está más próxima a uno (en comparación con el bloque 141 de coeficientes de error de predicción). Esto puede ser beneficioso para la cuantificación y codificación subsiguientes. El codificador 100 comprende una unidad de cuantificación de coeficientes 112 configurada para cuantificar el bloque 141 de coeficientes de error de predicción o el bloque 142 de coeficientes de error reescalados. La unidad de cuantificación de coeficientes 112 puede comprender o usar un conjunto de cuantificadores predeterminados. El conjunto de cuantificadores predeterminados puede proporcionar cuantificadores con diferentes grados de precisión o diferente resolución. Esto se ilustra en la Fig. 4, donde se ilustran diferentes cuantificadores 321, 322, 323. Los diferentes cuantificadores pueden proporcionar diferentes niveles de precisión (indicados mediante los diferentes valores en dB). Un cuantificador particular de la pluralidad de cuantificadores 321, 322, 323 puede corresponder a un valor particular de la envolvente de asignación 138. De este modo, un valor de energía de la envolvente de asignación 138 puede apuntar a un cuantificador correspondiente de la pluralidad de cuantificadores. De este modo, la determinación de una envolvente de asignación 138 puede simplificar el proceso de selección de un cuantificador que se usará para un coeficiente de error particular. Dicho de otro modo, la envolvente de asignación 138 puede simplificar el proceso de asignación de bits.

El conjunto de cuantificadores puede comprender uno o más cuantificadores 322 que añaden ruido aleatorio para aleatorizar el error de cuantificación. Esto se ilustra en la Fig. 4, que muestra un primer conjunto 326 de cuantificadores predeterminados que comprende un subconjunto 324 de cuantificadores con ruido aleatorio y un segundo conjunto 327 de cuantificadores predeterminados que comprende un subconjunto 325 de cuantificadores con ruido aleatorio. De este modo, la unidad de cuantificación de coeficientes 112 puede usar diferentes conjuntos 326, 327 de cuantificadores predeterminados, donde el conjunto de cuantificadores predeterminados que va a usarse por la unidad de cuantificación de coeficientes 112 puede depender de un parámetro de control 146 proporcionado por el predictor 117. En particular, la unidad de cuantificación de coeficientes 112 puede configurarse para seleccionar un conjunto 326, 327 de cuantificadores predeterminados para cuantificar el bloque 142 de coeficientes de error reescalados, en función del parámetro de control 146, donde el parámetro de control 146 puede depender de uno o más parámetros de predictor proporcionados por el predictor 117. El uno o más parámetros de predictor pueden indicar la calidad del bloque 150 de coeficientes de transformada estimados proporcionados por el predictor 117.

Los coeficientes de error cuantificados pueden codificarse por entropía usando, por ejemplo, un código de Huffman, proporcionando así datos de coeficientes 163 que se incluirán en el flujo de bits generado por el codificador 100.

El codificador 100 puede estar configurado para realizar un proceso de asignación de bits. Con este fin, el codificador 100 puede comprender unidades de asignación de bits 109, 110. La unidad de asignación de bits 109 puede estar configurada para determinar el número total de bits 143 que están disponibles para codificar el bloque actual 142 de coeficientes de error reescalados. El número total de bits 143 puede determinarse en función de la envolvente de asignación 138. La unidad de asignación de bits 110 puede estar configurada para proporcionar una asignación relativa de bits a los diferentes coeficientes de error reescalados, dependiendo del valor de energía correspondiente en la envolvente de asignación 138.

El proceso de asignación de bits puede usar un procedimiento de asignación iterativo. En el transcurso del procedimiento de asignación, la envolvente de asignación 138 puede desfasarse usando un parámetro de desfase, seleccionándose de este modo cuantificadores con una mayor/menor resolución. De este modo, el parámetro de desfase puede usarse para refinar o hacer más tosca la cuantificación global. El parámetro de desfase puede determinarse de manera que los datos de coeficientes 163, que se obtienen usando los cuantificadores dados por el parámetro de desfase y la envolvente de asignación 138, comprenden una pluralidad de bits que corresponden a (o no superan) el número total de bits 143 asignados al bloque actual 131. El parámetro de desfase usado por el codificador 100 para codificar el bloque actual 131 se incluye como datos de coeficientes 163 en el flujo de bits. En consecuencia, el descodificador correspondiente puede determinar los cuantificadores usados por la unidad de cuantificación de coeficientes 112 para cuantificar el bloque 142 de coeficientes de error reescalados.

Como resultado de la cuantificación de los coeficientes de error reescalados, se obtiene un bloque 145 de coeficientes de error cuantificados. El bloque 145 de coeficientes de error cuantificados corresponde al bloque de coeficientes de error que están disponibles en el descodificador correspondiente. En consecuencia, el bloque 145 de coeficientes de error cuantificados puede usarse para determinar un bloque 150 de coeficientes de transformada estimados. El codificador 100 puede comprender una unidad de reescalado inversa 113 configurada para realizar lo

inverso con respecto a las operaciones de reescalado realizadas por la unidad de reescalado 113, proporcionándose así un bloque 147 de coeficientes de error cuantificados escalados. Una unidad de adición 116 puede usarse para determinar un bloque 148 de coeficientes aplanados reconstruidos añadiendo el bloque 150 de coeficientes de transformada estimados al bloque 147 de coeficientes de error cuantificados escalados. Además, una unidad de aplanamiento inversa 114 puede usarse para aplicar la envolvente ajustada 139 al bloque 148 de coeficientes aplanados reconstruidos, lo que proporciona un bloque 149 de coeficientes reconstruidos. El bloque 149 de coeficientes reconstruidos corresponde a la versión del bloque 131 de coeficientes de transformada que está disponible en el descodificador correspondiente. En consecuencia, el bloque 149 de coeficientes reconstruidos puede usarse en el predictor 117 para determinar el bloque 150 de coeficientes estimados.

El bloque 149 de coeficientes reconstruidos se representa en el dominio no aplanado, es decir, el bloque 149 de coeficientes reconstruidos también representa la envolvente espectral del bloque actual 131. Como se indica posteriormente, esto puede ser beneficioso para el rendimiento del predictor 117.

El predictor 117 puede estar configurado para estimar el bloque 150 de coeficientes de transformada estimados en función de uno o más bloques previos 149 de coeficientes reconstruidos. En particular, el predictor 117 puede estar configurado para determinar uno o más parámetros de predictor de modo que se reduzca (por ejemplo, se minimice) un criterio de error de predicción predeterminado. A modo de ejemplo, el uno o más parámetros de predictor pueden determinarse de manera que se reduzca (por ejemplo, se minimice) la energía, o la energía ponderada de manera perceptual, del bloque 141 de coeficientes de error de predicción. El uno o más parámetros de predictor pueden incluirse como datos de predictor 164 en un flujo de bits generado por el codificador 100.

Los datos de predictor 164 pueden indicar el uno o más parámetros de predictor. Como se describirá en el presente documento, el predictor 117 solo puede usarse para un subconjunto de tramas o bloques 131 de una señal de audio. En particular, el predictor 117 puede no usarse para el primer bloque 131 de una trama I (trama independiente), que está codificada normalmente de manera independiente a un bloque anterior. Además, los datos de predictor 164 pueden comprender una o más banderas que indican la presencia de un predictor 117 para un bloque particular 131. En cuanto a los bloques, donde la contribución del predictor es apenas significativa (por ejemplo, cuando la ganancia del predictor se cuantifica a cero), puede ser beneficioso usar la bandera de presencia de predictor para indicar esta situación, que normalmente requiere un número considerablemente reducido de bits en comparación con la transmisión de la ganancia cero. Dicho de otro modo, los datos de predictor 164 para un bloque 131 pueden comprender una o más banderas de presencia de predictor que indican si se han determinado uno o más parámetros de predictor (y están comprendidos en los datos de predictor 164). El uso de una o más banderas de presencia de predictor puede usarse para ahorrar bits, si el predictor 117 no se usa para un bloque particular 131. Por tanto, dependiendo del número de bloques 131 que se codifican sin el uso de un predictor 117, el uso de una o más banderas de presencia de predictor puede ser más eficiente (por término medio) en lo que respecta a la velocidad binaria que la transmisión de parámetros de predictor por defecto (por ejemplo, con valor cero).

La presencia de un predictor 117 puede transmitirse explícitamente bloque a bloque. Esto permite ahorrar bits cuando no se usa la predicción. A modo de ejemplo, en lo que respecta a las tramas I, solo pueden usarse tres indicadores de presencia de predictor, ya que el primer bloque de la trama I no puede usar predicción. Dicho de otro modo, si se sabe que un bloque particular 131 es el primer bloque de una trama I, entonces puede no ser necesario transmitir ninguna bandera de presencia de predictor para este bloque particular 131 (ya que el descodificador correspondiente ya sabe que el bloque particular 131 no usa un predictor 117).

El predictor 117 puede usar un modelo de señal, como se describe en la solicitud de patente US61750052 y en las solicitudes de patente que reivindican la prioridad de la misma, cuyo contenido se incorpora como referencia. El uno o más parámetros de predictor pueden corresponder a uno o más parámetros de modelo del modelo de señal.

La Fig. 1b muestra un diagrama de bloques de un ejemplo adicional de un codificador de voz basado en transformada 170. El codificador de voz basado en transformada 170 de la Fig. 1b comprende muchos de los componentes del codificador 100 de la Fig. 1a. Sin embargo, el codificador de voz basado en transformada 170 de la Fig. 1b está configurado para generar un flujo de bits que tiene una velocidad binaria variable. Con este fin, el codificador 170 comprende una unidad de estados de velocidad binaria promedio (ABR) 172 configurada para realizar un seguimiento de la velocidad binaria usada por el flujo de bits en bloques anteriores 131. La unidad de asignación de bits 171 usa esta información para determinar el número total de bits 143 que están disponibles para codificar el bloque actual 131 de coeficientes de transformada. En términos generales, los codificadores de voz basados en transformada 100, 170 están configurados para generar un flujo de bits que indica o que comprende:

- datos de envolvente 161 que indican una envolvente actual cuantificada 134. La envolvente actual cuantificada 134 se usa para describir la envolvente de los bloques de un conjunto actual 132 o un conjunto desplazado 332 de bloques de coeficientes de transformada.
- Los datos de ganancia 162 indican una ganancia de corrección de nivel a para ajustar la envolvente interpolada 136 de un bloque actual 131 de coeficientes de transformada. Normalmente, una ganancia

diferente a se proporciona a cada bloque 131 del conjunto actual 132 o del conjunto desplazado 332 de bloques.

- Los datos de coeficientes 163 indican el bloque 141 de coeficientes de error de predicción para el bloque actual 131. En particular, los datos de coeficientes 163 indican el bloque 145 de coeficientes de error cuantificados. Además, los datos de coeficientes 163 pueden indicar un parámetro de desfase que puede usarse para determinar los cuantificadores para realizar una cuantificación inversa en el descodificador.
- Los datos de predictor 164 indican uno o más coeficientes de predictor que se usarán para determinar un bloque 150 de coeficientes estimados a partir de bloques anteriores 149 de coeficientes reconstruidos.

A continuación, un descodificador de voz basado en transformada 500 correspondiente se describe en el contexto de las Fig. 5a a 5d. La Fig. 5a muestra un diagrama de bloques de un descodificador de voz basado en transformada 500 de ejemplo. El diagrama de bloques muestra un banco de filtros de síntesis 504 (también denominado unidad de transformada inversa) que se usa para convertir un bloque 149 de coeficientes reconstruidos desde el dominio de transformada al dominio de tiempo, lo que proporciona muestras de la señal de audio descodificada. El banco de filtros de síntesis 504 puede usar una MDCT inversa con un espaciamiento predeterminado (por ejemplo, un espaciamiento de 5 ms o de 256 muestras aproximadamente). El bucle principal del descodificador 500 actúa en unidades de este espaciamiento. Cada etapa produce un vector de dominio de transformada (también denominado bloque) que presenta una longitud o dimensión que corresponde a un ajuste de ancho de banda predeterminado del sistema. Tras realizar un relleno con ceros hasta el tamaño de transformada del banco de filtros de síntesis 504, el vector de dominio de transformada se usará para sintetizar una actualización de señal de dominio de tiempo de una longitud predeterminada (por ejemplo, 5 ms) en el proceso de solapamiento/adición del banco de filtros de síntesis 504.

Como se ha indicado anteriormente, los códecs de audio genéricos basados en transformada utilizan normalmente tramas con secuencias de bloques cortos en el intervalo de 5 ms para la gestión de fenómenos transitorios. De este modo, los códecs de audio genéricos basados en transformada proporcionan las transformadas necesarias y herramientas de conmutación de ventana para una correcta coexistencia de bloques cortos y largos. Una sección de entrada espectral de voz definida omitiendo el banco de filtros de síntesis 504 de la Fig. 5a puede, por tanto, integrarse de manera adecuada en el códec de audio basado en transformada de propósito general sin necesidad de introducir herramientas de conmutación adicionales. Dicho de otro modo, el descodificador de voz basado en transformada 500 de la Fig. 5a puede combinarse de manera adecuada con un descodificador de audio genérico basado en transformada. En particular, el descodificador de voz basado en transformada 500 de la Fig. 5a puede usar el banco de filtros de síntesis 504 proporcionado por el descodificador de audio genérico basado en transformada (por ejemplo, el descodificador AAC o el descodificador HE-AAC).

A partir del flujo de bits entrante (en particular, a partir de los datos de envolvente 161 y de los datos de ganancia 162 comprendidos en el flujo de bits), un descodificador de envolvente 503 puede determinar una envolvente de señal. En particular, el descodificador de envolvente 503 puede configurarse para determinar la envolvente ajustada 139 en función de los datos de envolvente 161 y los datos de ganancia 162. De este modo, el descodificador de envolvente 503 puede realizar tareas similares a la de la unidad de interpolación 104 y la unidad de refinamiento de envolvente 107 del codificador 100, 170. Como se ha descrito anteriormente, la envolvente ajustada 109 representa un modelo de la varianza de señal en un conjunto de bandas de frecuencia predefinidas 302.

Además, el descodificador 500 comprende una unidad de aplanamiento inverso 114 que está configurada para aplicar la envolvente ajustada 139 a un vector de dominio aplanado, cuyas entradas pueden tener, de manera nominal, una varianza uno. El vector de dominio aplanado corresponde al bloque 148 de coeficientes aplanados reconstruidos descritos en el contexto del codificador 100, 170. En la salida de la unidad de aplanamiento inverso 114 se obtiene el bloque 149 de coeficientes reconstruidos. El bloque 149 de coeficientes reconstruidos se proporciona al banco de filtros de síntesis 504 (para generar la señal de audio descodificada) y al predictor de subbanda 517.

El predictor de subbanda 517 funciona de manera similar al predictor 117 del codificador 100, 170. En particular, el predictor de subbanda 517 está configurado para determinar un bloque 150 de coeficientes de transformada estimados (en el dominio aplanado) en función de uno o más bloques previos 149 de coeficientes reconstruidos (usando el uno o más parámetros de predictor señalizados en el flujo de bits). Dicho de otro modo, el predictor de subbanda 517 está configurado para proporcionar un vector de dominio aplanado predicho a partir de una memoria intermedia de vectores de salida descodificados previamente y envolventes de señal, en función de los parámetros de predictor tales como un retardo de predictor y una ganancia de predictor. El descodificador 500 comprende un descodificador de predictor 501 configurado para descodificar los datos de predictor 164 para determinar el uno o más parámetros de predictor.

El descodificador 500 comprende además un descodificador de espectro 502 que está configurado para suministrar una corrección aditiva al vector de dominio aplanado predicho, en función de, normalmente, la parte más grande del flujo de bits (es decir, en función de los datos de coeficientes 163). El proceso de descodificación de espectro se

controla principalmente mediante un vector de asignación que se obtiene a partir de la envolvente y de un parámetro de control de asignación transmitido (también denominado parámetro de desfase). Como se ilustra en la Fig. 5a, puede haber una dependencia directa del descodificador de espectro 502 con respecto a los parámetros de predictor 520. De este modo, el descodificador de espectro 502 puede estar configurado para determinar el bloque 147 de coeficientes de error cuantificados y escalados en función de los datos de coeficientes recibidos 163. Como se ha indicado en el contexto del codificador 100, 170, los cuantificadores 321, 322, 323 usados para cuantificar el bloque 142 de coeficientes de error reescalados dependen normalmente de la envolvente de asignación 138 (que puede obtenerse a partir de la envolvente ajustada 139) y del parámetro de desfase. Además, los cuantificadores 321, 322, 323 pueden depender de un parámetro de control 146 proporcionado por el predictor 117. El descodificador 500 puede obtener el parámetro de control 146 usando los parámetros de predictor 520 (de manera análoga al codificador 100, 170).

Como se ha indicado anteriormente, el flujo de bits recibido comprende datos de envolvente 161 y datos de ganancia 162 que pueden usarse para determinar la envolvente ajustada 139. En particular, la unidad 531 del descodificador de envolvente 503 puede estar configurada para determinar la envolvente actual cuantificada 134 a partir de los datos de envolvente 161. A modo de ejemplo, la envolvente actual cuantificada 134 puede tener una resolución de 3 dB en bandas de frecuencia predefinidas 302 (como se indica en la Fig. 3a). La envolvente actual cuantificada 134 puede actualizarse para cada conjunto 132, 332 de bloques (por ejemplo, cada cuatro unidades de codificación, es decir, bloques, o cada 20 ms), en particular para cada conjunto desplazado 332 de bloques. Las bandas de frecuencia 302 de la envolvente actual cuantificada 134 pueden comprender un mayor número de celdas de frecuencia 301 en función de la frecuencia, con el fin de adaptarse a las propiedades de la audición humana.

La envolvente actual cuantificada 134 puede interpolarse linealmente a partir de una envolvente previa cuantificada 135 en envolvente interpoladas 136 para cada bloque 131 del conjunto desplazado 332 de bloques (o, posiblemente, del conjunto actual 132 de bloques). Las envolventes interpoladas 136 pueden determinarse en el dominio cuantificado de 3 dB. Esto significa que los valores de energía interpolados 303 pueden redondearse al nivel de 3 dB más próximo. Una envolvente interpolada 136 de ejemplo se ilustra mediante el gráfico de línea de puntos de la Fig. 3a. Para cada envolvente actual cuantificada 134, cuatro ganancias de corrección de nivel a 137 (también denominadas ganancias de envolvente) se proporcionan como datos de ganancia 162. La unidad de descodificación de ganancia 532 puede estar configurada para determinar las ganancias de corrección de nivel a 137 a partir de los datos de ganancia 162. Las ganancias de corrección de nivel pueden cuantificarse en etapas de 1 dB. Cada ganancia de corrección de nivel se aplica a la envolvente interpolada correspondiente 136 para proporcionar las envolventes ajustadas 139 para los diferentes bloques 131. Debido a la mayor resolución de las ganancias de corrección de nivel 137, la envolvente ajustada 139 puede tener una mayor resolución (por ejemplo, una resolución de 1 dB). La Fig. 3b muestra una interpolación lineal o geométrica de ejemplo entre la envolvente previa cuantificada 135 y la envolvente actual cuantificada 134. Las envolventes 135, 134 pueden separarse en una parte de nivel promedio y una parte de forma del espectro logarítmico. Estas partes pueden interpolarse con estrategias independientes tales como una estrategia lineal, geométrica o armónica (resistencias paralelas). De este modo, pueden usarse diferentes esquemas de interpolación para determinar las envolventes interpoladas 136. El esquema de interpolación usado por el descodificador 500 corresponde normalmente al esquema de interpolación usado por el codificador 100, 170.

La unidad de refinamiento de envolvente 107 del descodificador de envolvente 503 puede estar configurada para determinar una envolvente de asignación 138 a partir de la envolvente ajustada 139 cuantificando la envolvente ajustada 139 (por ejemplo, en etapas de 3 dB). La envolvente de asignación 138 puede usarse junto con el parámetro de control de asignación o el parámetro de desfase (comprendidos en los datos de coeficientes 163) para crear un vector de asignación de enteros nominal usado para controlar la descodificación espectral, es decir, la descodificación de los datos de coeficientes 163. En particular, el vector de asignación de enteros nominal puede usarse para determinar un cuantificador para la cuantificación inversa de los índices de cuantificación comprendidos en los datos de coeficientes 163. La envolvente de asignación 138 y el vector de asignación de enteros nominal pueden determinarse de manera análoga en el codificador 100, 170 y en el descodificador 500.

Para permitir que un descodificador 500 se sincronice con un flujo de bits recibido, pueden transmitirse diferentes tipos de tramas. Una trama puede corresponder a un conjunto 132, 132 de bloques, en particular a un bloque desplazado 332 de bloques. En particular, pueden transmitirse las denominadas tramas P, que se codifican de manera relativa con respecto a una trama previa. En la descripción anterior se supuso que el descodificador 500 se ha percatado de la envolvente previa cuantificada 135. La envolvente previa cuantificada 135 puede proporcionarse en una trama previa, de manera que el conjunto actual 132 o el conjunto desplazado correspondiente 332 pueden corresponder a una trama P. Sin embargo, en un escenario de arranque, el descodificador 500 no se percata normalmente de la envolvente previa cuantificada 135. Con este fin, puede transmitirse una trama I (por ejemplo, tras el arranque o de manera habitual). La trama I puede comprender dos envolventes, una de las cuales se usa como la envolvente previa cuantificada 135 y la otra se usa como la envolvente actual cuantificada 134. Las tramas I pueden usarse para el caso de arranque de la sección de entrada espectral de voz (es decir, del descodificador de voz basado en transformada 500), por ejemplo tras una trama que utiliza un modo de codificación de audio diferente y/o como una herramienta para habilitar explícitamente un punto de unión del flujo de bits de audio.

El funcionamiento del predictor de subbanda 517 se ilustra en la Fig. 5d. En el ejemplo ilustrado, los parámetros de predictor 520 son un parámetro de retardo y un parámetro de ganancia de predictor g . Los parámetros de predictor 520 pueden determinarse a partir de los datos de predictor 164 usando una tabla predeterminada de posibles valores del parámetro de retardo y del parámetro de ganancia de predictor. Esto permite la transmisión eficaz, en cuanto a la velocidad binaria, de los parámetros de predictor 520.

El uno o más vectores de coeficientes de transformada descodificados previamente (es decir, el uno o más bloques previos 149 de coeficientes reconstruidos) pueden almacenarse en una memoria intermedia de señales de subbanda (o MDCT) 541. La memoria intermedia 541 puede actualizarse según el espaciamiento (por ejemplo, cada 5 ms). El extractor de predictor 543 puede configurarse para funcionar en la memoria intermedia 541 dependiendo de un parámetro de retardo normalizado T . El parámetro de retardo normalizado T puede determinarse normalizando el parámetro de retardo 520 con respecto a unidades de espaciamiento (por ejemplo, con respecto a unidades de espaciamiento de MDCT). Si el parámetro de retardo T es un entero, el extractor 543 puede almacenar una o más unidades de tiempo T de vectores de coeficientes de transformada previamente descodificados en la memoria intermedia 541. Dicho de otro modo, el parámetro de retardo T puede indicar cuáles de los uno o más bloques previos 149 de coeficientes reconstruidos van a usarse para determinar el bloque 150 de coeficientes de transformada estimados. Una descripción detallada relacionada con una posible implementación del extractor 543 se proporciona en la solicitud de patente US61750052 y en las solicitudes de patente que reivindican la prioridad de la misma, cuyo contenido se incorpora como referencia.

El extractor 543 puede actuar en vectores (o bloques) que transportan envolventes de señal completas. Por otro lado, el bloque 150 de coeficientes de transformada estimados (que serán proporcionados por el predictor de subbanda 517) se representa en el dominio aplanado. En consecuencia, la salida del extractor 543 puede conformarse en un vector de dominio aplanado. Esto puede conseguirse usando un conformador 544 que usa las envolventes ajustadas 139 del uno o más bloques previos 149 de coeficientes reconstruidos. Las envolventes ajustadas 139 del uno o más bloques previos 149 de coeficientes reconstruidos pueden envolverse en una memoria intermedia de envolventes 542. La unidad de conformación 544 puede estar configurada para introducir una envolvente de señal retardada que va a usarse en el aplanamiento de T_0 unidades de tiempo en la memoria intermedia de envolventes 542, donde T_0 es el entero más próximo a T . Después, el vector de dominio aplanado puede escalarse mediante el parámetro de ganancia g para proporcionar el bloque 150 de coeficientes de transformada estimados (en el dominio aplanado).

La unidad de conformación 544 puede configurarse para determinar un vector de dominio aplanado de manera que los vectores de dominio aplanado en la salida de la unidad de conformación 544 presentan una varianza unitaria en cada banda de frecuencia. La unidad de conformación 544 puede basarse totalmente en los datos de la memoria intermedia de envolventes 542 para conseguir este objetivo. A modo de ejemplo, la unidad de conformación 544 puede configurarse para seleccionar la envolvente de señal retardada de manera que los vectores de dominio aplanado en la salida de la unidad de conformación 544 presentan una varianza unitaria en cada banda de frecuencia. Además, o como alternativa, la unidad de conformación 544 puede estar configurada para medir la varianza de los vectores de dominio aplanado en la salida de la unidad de conformación 544 y para ajustar la varianza de los vectores en lo que respecta a la propiedad de varianza unitaria. Un posible tipo de normalización puede usar una única ganancia de banda ancha (por ranura) que normalice los vectores de dominio aplanado en un vector de varianza unitaria. Las ganancias pueden transmitirse desde un codificador 100 a un descodificador correspondiente 500 (por ejemplo, en una forma cuantificada y codificada) en el flujo de bits.

Como alternativa, el proceso de aplanamiento retardado llevado a cabo por el conformador 544 puede omitirse usando un predictor de subbanda 517 que funciona en el dominio aplanado, por ejemplo un predictor de subbanda 517 que actúa en los bloques 148 de coeficientes aplanados reconstruidos. Sin embargo, se ha observado que una secuencia de vectores (o bloques) de dominio aplanado no se correlacionan bien con señales de tiempo debido a los aspectos de solapamiento de tiempo de la transformada (por ejemplo, la transformada MDCT). En consecuencia se reduce el requisito de adecuación con el modelo de señales subyacente del extractor 543 y se produce un mayor nivel de ruido de codificación a partir de la estructura alternativa. Dicho de otro modo, se ha observado que los modelos de señales (por ejemplo, modelos sinusoidales o periódicos) usados por el predictor de subbanda 517 ofrecen un mayor rendimiento en el dominio no aplanado (en comparación con el dominio aplanado).

Debe observarse que en un ejemplo alternativo, la salida del predictor 517 (es decir, el bloque 150 de los coeficientes de transformada estimados) pueden añadirse en la salida de la unidad de aplanamiento inverso 114 (es decir, al bloque 149 de coeficientes reconstruidos) (véase la Fig. 5a). La unidad de conformación 544 de la Fig. 5c puede configurarse entonces para realizar la operación combinada de aplanamiento retardado y de aplanamiento inverso.

Elementos del flujo de bits recibido pueden controlar el vaciado ocasional de la memoria intermedia de subbandas 541 y de la memoria intermedia de envolventes 542, por ejemplo en caso de una primera unidad de codificación (es decir, un primer bloque) de una trama I . Esto permite la descodificación de una trama I sin conocer los datos anteriores. La primera unidad de codificación no podrá usar normalmente una contribución predictiva pero, sin embargo, puede usar un número relativamente menor de bits para transportar la información de predictor 520. La

pérdida de ganancia de predicción puede compensarse asignando más bits a la codificación de errores de predicción de esta primera unidad de codificación. Normalmente, la contribución de predictor es también sustancial para la segunda unidad de codificación (es decir, un segundo bloque) de una trama I. Debido a estos aspectos, la calidad puede mantenerse con un incremento relativamente pequeño en la velocidad binaria, incluso con un uso muy frecuente de tramas I.

Dicho de otro modo, los conjuntos 132, 332 de bloques (también denominados tramas) comprenden una pluralidad de bloques 131 que pueden codificarse usando codificación predictiva. Cuando se codifica una trama I, solamente el primer bloque 203 de un conjunto 332 de bloques no puede codificarse usando la ganancia de codificación conseguida por un codificador predictivo. El bloque inmediatamente posterior 201 puede beneficiarse ya de la codificación predictiva. Esto significa que las desventajas de una trama I en lo que respecta a la eficacia de codificación están limitadas a la codificación del primer bloque 203 de coeficientes de transformada de la trama 332 y no se aplican a los otros bloques 201, 204, 205 de la trama 332. Por tanto, el esquema de codificación de voz basado en transformada descrito en el presente documento permite un uso relativamente frecuente de tramas I sin un impacto significativo en la eficacia de codificación. De este modo, el esquema de codificación de voz basado en transformada ahora descrito es, en particular, adecuado en aplicaciones que requieren una sincronización relativamente rápida y/o relativamente frecuente entre el descodificador y el codificador. Como se ha indicado anteriormente, durante la inicialización de una trama I, la memoria intermedia de señal de predictor, es decir, la memoria intermedia de subbandas 541, puede vaciarse con ceros y la memoria intermedia de envolventes 542 puede llenarse solamente con una ranura de tiempo de valores, es decir, puede llenarse solamente con una única envolvente ajustada 139 (correspondiente al primer bloque 131 de la trama I). El primer bloque 131 de la trama I no usará normalmente predicción. El segundo bloque 131 tiene acceso solamente a dos ranuras de tiempo de la memoria intermedia de envolventes 542 (es decir, a las envolventes 139 del primer y del segundo bloque 131), el tercer bloque solamente a tres ranuras de tiempo (es decir, a las envolventes 139 de tres bloques 131), y el cuarto bloque 131 solamente a cuatro ranuras de tiempo (es decir, a envolventes 139 de cuatro bloques 131).

La regla de aplanamiento retardado del conformador espectral 544 (para identificar una envolvente para determinar el bloque 150 de coeficientes de transformada estimados (en el dominio aplanado)) está basada en un valor de retardo entero T_0 determinado mediante el redondeo del parámetro de retardo de predictor T , en unidades de tamaño de bloque K (donde la unidad de un tamaño de bloque puede denominarse ranura de tiempo o ranura), al entero más próximo. Sin embargo, en el caso de una trama I, este valor de retardo entero T_0 puede apuntar a entradas no disponibles de la memoria intermedia de envolventes 542. Por ello, el conformador espectral 544 puede estar configurado para determinar el valor de retardo entero T_0 de manera que el valor de retardo entero T_0 está limitado al número de envolventes 139 que están almacenadas en la memoria intermedia de envolventes 542, es decir, de manera que el valor de retardo entero T_0 no apunta a envolventes 139 que no están disponibles en la memoria intermedia de envolventes 542. Con este fin, el valor de retardo entero T_0 puede limitarse a un valor que depende del índice de bloque en la trama actual. A modo de ejemplo, el valor de retardo entero T_0 puede limitarse al valor de índice del bloque actual 131 (que va a codificarse) en la trama actual (por ejemplo, a 1 para el primer bloque 131, a 2 para el segundo bloque 131, a 3 para el tercer bloque 131 y a 4 para el cuarto bloque 131 de una trama). Al hacer esto pueden evitarse estados no deseables y/o distorsiones debidos al proceso de aplanamiento.

La Fig. 5d muestra un diagrama de bloques de un descodificador de espectro 502 de ejemplo. El descodificador de espectro 502 comprende un descodificador sin pérdidas 551 que está configurado para descodificar los datos de coeficientes codificados por entropía 163. Además, el descodificador de espectro 502 comprende un cuantificador inverso 552 que está configurado para asignar valores de coeficiente a los índices de cuantificación comprendidos en los datos de coeficientes 163. Como se ha descrito en el contexto del codificador 100, 170, diferentes coeficientes de transformada pueden cuantificarse usando diferentes cuantificadores seleccionados a partir de un conjunto de cuantificadores predeterminados, por ejemplo un conjunto finito de cuantificadores escalares basados en modelos. Como se muestra en la Fig. 4, un conjunto de cuantificadores 321, 322, 323 puede comprender diferentes tipos de cuantificadores. El conjunto de cuantificadores puede comprender un cuantificador 321 que proporciona síntesis de ruido (en caso de una velocidad binaria nula), uno o más cuantificadores con ruido aleatorio 322 (para relaciones de señal a ruido, SNR, relativamente bajas y para velocidades binarias intermedias) y/o uno o más cuantificadores planos 323 (para SNR relativamente altas y para velocidades binarias relativamente altas).

La unidad de refinamiento de envolvente 107 puede estar configurada para proporcionar la envolvente de asignación 138, que puede combinarse con el parámetro de desfase comprendido en los datos de coeficientes 163 para proporcionar un vector de asignación. El vector de asignación contiene un valor entero para cada banda de frecuencia 302. El valor entero para una banda de frecuencia particular 302 apunta al punto de distorsión de velocidad que se usará en la cuantificación inversa de los coeficientes de transformada de la banda particular 302. Dicho de otro modo, el valor entero para la banda de frecuencia particular 302 apunta al cuantificador que se usará para la cuantificación inversa de los coeficientes de transformada de la banda particular 302. Un incremento del valor entero en uno corresponde a un incremento de 1,5 dB en la SNR. En los cuantificadores con ruido aleatorio 322 y los cuantificadores planos 323, puede usarse un modelo de distribución de probabilidad de Laplace en la codificación sin pérdidas, que puede utilizar una codificación aritmética. Uno o más cuantificadores con ruido aleatorio 322 pueden usarse para conseguir una transición sin interrupciones entre casos de alta y baja velocidad

binaria. Cuantificadores con ruido aleatorio 322 pueden ser beneficiosos a la hora de crear una calidad de audio de salida suficientemente uniforme para señales estacionarias similares al ruido.

5 Dicho de otro modo, el cuantificador inverso 552 puede estar configurado para recibir los índices de cuantificación de coeficientes de un bloque actual 131 de coeficientes de transformada. El uno o más índices de cuantificación de coeficientes de una banda de frecuencia particular 302 se han determinado usando un cuantificador correspondiente a partir de un conjunto predeterminado de cuantificadores. El valor del vector de asignación (que puede determinarse desfasando la envolvente de asignación 138 con el parámetro de desfase) para la banda de frecuencia particular 302 indica el cuantificador que se ha usado para determinar el uno o más índices de cuantificación de coeficientes de la banda de frecuencia particular 302. Tras haberse identificado el cuantificador, el uno o más índices de cuantificación de coeficientes pueden cuantificarse de manera inversa para proporcionar el bloque 145 de coeficientes de error cuantificados.

15 Además, el descodificador espectral 502 puede comprender una unidad de reescalado inverso 113 para proporcionar el bloque 147 de coeficientes de error cuantificados. Las herramientas e interconexiones adicionales en torno al descodificador sin pérdidas 551 y al cuantificador inverso 552 de la Fig. 5d pueden usarse para adaptar la descodificación espectral a su uso en el descodificador global 500 mostrado en la Fig. 5a, donde la salida del descodificador espectral 502 (es decir, el bloque 145 de coeficientes de error cuantificados) se usa para proporcionar una corrección aditiva a un vector predicho de dominio aplanado (es decir, al bloque 150 de coeficientes de transformada estimados). En particular, las herramientas adicionales pueden garantizar que el procesamiento realizado por el descodificador 500 corresponde al procesamiento realizado por el codificador 100, 170.

25 En particular, el descodificador espectral 502 puede comprender una unidad de escalado heurístico 111. Como se muestra junto con el codificador 100, 170, la unidad de escalado heurístico 111 puede tener un impacto en la asignación de bits. En el codificador 100, 170, los bloques actuales 141 de coeficientes de error de predicción pueden escalarse hasta la varianza unitaria mediante una regla heurística. En consecuencia, la asignación por defecto puede dar lugar a una cuantificación muy precisa de la salida final escalada de manera descendente de la unidad de escalado heurístico 111. Por tanto, la asignación debería modificarse de manera similar a la modificación de los coeficientes de error de predicción. Sin embargo, como se describe posteriormente, puede ser beneficioso evitar la reducción de recursos de codificación para una o más de las celdas de baja frecuencia (o bandas de baja frecuencia). En particular, esto puede ser beneficioso para hacer frente a un artefacto de reverberación/ruido LF (baja frecuencia) que es más prominente en situaciones sonoras (es decir, para una señal que tiene un parámetro de control relativamente grande 146, rfu). De este modo, la selección de cuantificador / asignación de bits dependiente del parámetro de control 146, que se describe posteriormente, puede considerarse un "aumento adaptativo de la calidad de LF de una entonación".

40 El descodificador espectral puede depender de un parámetro de control 146 denominado rfu que puede ser una versión limitada de la ganancia de predictor g , por ejemplo $rfu = \min(1, \max(g, 0))$.

Pueden usarse procedimientos alternativos para determinar el parámetro de control 146, rfu. En particular, el parámetro de control 146 puede determinarse usando el seudocódigo dado en la Tabla 1.

Tabla 1

```
f_gain = f_pred_gain;
if (f_gain < -1,0)
    f_rfu=1,0;
else if (f_gain < 0,0)
    f_rfu = -f_gain;
else if (f_gain < 1,0)
    f_rfu = f_gain;
else if (f_gain < 2,0)
    f_rfu = 2,0 - f_gain;
else // f_gain >= 2,0
    f_rfu = 0,0.
```

45 Las variables f_gain y f_pred_gain pueden fijarse al mismo valor. En particular, la variable f_gain puede corresponder a la ganancia de predictor g . El parámetro de control 146, rfu, se denomina f_rfu en la Tabla 1. La ganancia f_gain puede ser un número real.

50 En comparación con la primera definición del parámetro de control 146, la segunda definición (según la Tabla 1) reduce el valor del parámetro de control 146, rfu, para ganancias de predictor superiores a 1 y aumenta el parámetro de control 146, rfu, para ganancias de predictor negativas.

Usando el parámetro de control 146, puede adaptarse el conjunto de cuantificadores usado en la unidad de cuantificación de coeficientes 112 del codificador 100, 170 y usados en el cuantificador inverso 552. En particular, el ruido del conjunto de cuantificadores puede adaptarse en función del parámetro de control 146. A modo de ejemplo, el valor del parámetro de control 146, rfu, próximo a 1, puede hacer que se limite el intervalo de niveles de asignación usando cuantificadores con ruido aleatorio y puede hacer que se reduzca la varianza del nivel de síntesis de ruido. En un ejemplo, puede fijarse un umbral de decisión de ruido aleatorio con rfu = 0,75 y una ganancia de ruido igual a 1-rfu. La adaptación de ruido aleatorio puede afectar tanto a la descodificación sin pérdidas como al cuantificador inverso, mientras que la adaptación de ganancia de ruido solo afecta normalmente al cuantificador inverso.

Puede suponerse que la contribución del predictor es sustancial en situaciones sonoras/tonales. De este modo, una ganancia de predictor relativamente alta *g* (es decir, un parámetro de control relativamente alto 146) puede indicar una señal de voz sonora o tonal. En tales situaciones, la adición de ruido aleatorio o explícito (caso de asignación cero) ha demostrado, empíricamente, ser contraproducente con respecto a la calidad percibida de la señal codificada. En consecuencia, el número de cuantificadores con ruido aleatorio 322 y/o el tipo de ruido usado en el cuantificador de síntesis de ruido 321 pueden adaptarse en función de la ganancia de predictor *g*, lo que mejora la calidad percibida de la señal de voz codificada.

De este modo, el parámetro de control 146 puede usarse para modificar el intervalo 324, 325 de SNR para el que se usan cuantificadores con ruido aleatorio 322. A modo de ejemplo, si el parámetro de control 146 rfu < 0,75, puede usarse el intervalo 324 de cuantificadores con ruido aleatorio. Dicho de otro modo, si el parámetro de control 146 es inferior a un umbral predeterminado, puede usarse el primer conjunto 326 de cuantificadores. Por otro lado, si el parámetro de control 146 rfu ≥ 0,75, puede usarse el intervalo 325 de cuantificadores con ruido aleatorio. Dicho de otro modo, si el parámetro de control 146 es superior o igual al umbral predeterminado, puede usarse el segundo conjunto 327 de cuantificadores.

Además, el parámetro de control 146 puede usarse para la modificación de la varianza y la asignación de bits. El motivo es que, normalmente, una predicción satisfactoria requerirá una menor corrección, especialmente en el intervalo de menor frecuencia de 0 a 1 kHz. Puede ser ventajoso hacer que el cuantificador se percate explícitamente de esta desviación con respecto al modelo de varianza unitaria con el fin de liberar recursos de codificación para bandas de frecuencia superiores 302. Esto se describe en el contexto del panel iii de la Figura 17c del documento WO2009/086918, cuyo contenido se incorpora a modo de referencia. En el descodificador 500, esta modificación puede implementarse modificando el vector de asignación nominal según una regla de escalado heurístico (aplicada usando la unidad de escalado 111) y, al mismo tiempo, escalando en el tiempo la salida del cuantificador inverso 552 según una regla heurística de escalado inverso usando la unidad de escalado inverso 113. Siguiendo la teoría del documento WO2009/086918, la regla de escalado heurístico y la regla heurística de escalado inverso deben estar estrechamente relacionadas. Sin embargo, se ha observado que es ventajoso, desde un punto de vista empírico, cancelar la modificación de asignación para la una o más bandas de frecuencia más bajas 302 con el fin de hacer frente a problemas ocasionales con ruido LF (de baja frecuencia) para componentes de señales sonoras. La cancelación de la modificación de asignación puede realizarse dependiendo del valor de la ganancia de predictor *g* y/o del parámetro de control 146. En particular, la cancelación de la modificación de asignación puede realizarse solamente si el parámetro de control 146 supera el umbral de decisión de ruido aleatorio.

Como se ha descrito anteriormente, un codificador 100, 170 y/o un descodificador 500 puede comprender una unidad de escalado 111 que está configurada para reescalar los coeficientes de error de predicción $\Delta(k)$ para proporcionar un bloque 142 de coeficientes de error reescalados. La unidad de reescalado 111 puede usar una o más reglas heurísticas predeterminadas para realizar el reescalado. En un ejemplo, la unidad de reescalado 111 puede usar una regla de escalado heurístico que comprende la ganancia $d(f)$, por ejemplo

$$d(f) = 1 + \frac{7 \cdot rfu^2}{1 + \left(\frac{f}{f_0}\right)^3}$$

donde una frecuencia de ruptura f_0 puede fijarse a, por ejemplo, 1000 Hz. Por tanto, la unidad de reescalado 111 puede configurarse para aplicar una ganancia de dependencia de frecuencia $d(f)$ a los coeficientes de error de predicción para proporcionar el bloque 142 de coeficientes de error reescalados. La unidad de reescalado inverso 113 puede configurarse para aplicar una inversa de la ganancia de dependencia de frecuencia $d(f)$. La ganancia de dependencia de frecuencia $d(f)$ puede depender del parámetro de control rfu 146. En el ejemplo anterior, la ganancia $d(f)$ presenta un carácter de paso bajo, de manera que los coeficientes de error de predicción se atenúan más a frecuencias superiores que a frecuencias inferiores, y/o de manera que los coeficientes de error de predicción se enfatizan más a frecuencias inferiores que a frecuencias superiores. La ganancia mencionada anteriormente $d(f)$ es siempre mayor o igual a uno. Por tanto, en una forma de realización preferida, la regla de escalado heurístico es tal que los coeficientes de error de predicción se enfatizan en un factor de uno o más (dependiendo de la frecuencia).

Debe observarse que la ganancia dependiente de la frecuencia puede indicar una potencia o una varianza. En tales casos, la regla de escalado y la regla de escalado inverso deben obtenerse en función de la raíz cuadrada de la

ganancia dependiente de la frecuencia, por ejemplo en función de $\sqrt{d(f)}$. El grado de énfasis y/o atenuación puede depender de la calidad de la predicción conseguida por el predictor 117. La ganancia de predictor g y/o el parámetro de control rfu 146 pueden indicar la calidad de la predicción. En particular, un valor relativamente bajo del parámetro de control rfu 146 (relativamente próximo a cero) puede indicar una calidad baja de predicción. En tales casos, cabe esperar que los coeficientes de error de predicción tengan valores (absolutos) relativamente altos en todas las frecuencias. Un valor relativamente alto del parámetro de control rfu 146 (relativamente próximo a uno) puede indicar una calidad alta de predicción. En tales casos, cabe esperar que los coeficientes de error de predicción tengan valores (absolutos) relativamente altos para altas frecuencias (que son más difíciles de predecir). Por tanto, para conseguir una varianza unitaria en la salida de la unidad de reescalado 111, la ganancia $d(f)$ puede ser tal que, en caso de una calidad de predicción relativamente baja, la ganancia $d(f)$ es sustancialmente plana para todas las frecuencias, mientras que en caso de una calidad de predicción relativamente alta, la ganancia $d(f)$ tiene un carácter de paso bajo con el fin de incrementar o maximizar la varianza a bajas frecuencias. Este es el caso para la ganancia dependiente de rfu , $d(f)$, mencionada anteriormente.

Como se ha indicado anteriormente, la unidad de asignación de bits 110 puede estar configurada para proporcionar una asignación relativa de bits a los diferentes coeficientes de error reescalados, dependiendo del valor de energía correspondiente en la envolvente de asignación 138. La unidad de asignación de bits 110 puede estar configurada para tener en cuenta la regla de reescalado heurística. La regla de reescalado heurística puede depender de la calidad de la predicción. En caso de una calidad de predicción relativamente alta, puede ser beneficioso asignar un número de bits relativamente mayor a la codificación de los coeficientes de error de predicción (o el bloque 142 de coeficientes de error reescalados) a altas frecuencias que con respecto a la codificación de los coeficientes a bajas frecuencias. Esto puede deberse a que en caso de una alta calidad de predicción, los coeficientes de baja frecuencia ya se han predicho correctamente, mientras que los coeficientes de alta frecuencia se predicen normalmente con peor precisión. Por otro lado, en caso de una calidad de predicción relativamente baja, la asignación de bits debe permanecer invariable.

El comportamiento anterior puede implementarse aplicando la inversa de las reglas heurísticas / ganancia $d(f)$ a la envolvente ajustada actual 139 con el fin de determinar una envolvente de asignación 138 que tenga en cuenta la calidad de predicción.

La envolvente ajustada 139, los coeficientes de error de predicción y la ganancia $d(f)$ pueden representarse en el dominio logarítmico o de dB. En tal caso, la aplicación de la ganancia $d(f)$ a los coeficientes de error de predicción puede corresponder a una operación "suma", y la aplicación de la inversa de la ganancia $d(f)$ con respecto a la envolvente ajustada 139 puede corresponder a una operación "resta".

Debe observarse que puede haber diversas variantes de las reglas heurísticas / ganancia $d(f)$. En particular, la curva dependiente de una frecuencia fija con un carácter de paso bajo $\left(1 + \left(\frac{f}{f_0}\right)^3\right)^{-1}$ puede sustituirse por una función que depende de los datos de envolvente (por ejemplo, de la envolvente ajustada 139 para el bloque actual 131). Las reglas heurísticas modificadas pueden depender tanto del parámetro de control rfu 146 como de los datos de envolvente.

A continuación se describen diferentes maneras de determinar una ganancia de predictor ρ , que puede corresponder a la ganancia de predictor g . La ganancia de predictor ρ puede usarse como una indicación de la calidad de la predicción. El vector residual de predicción (es decir, el bloque 141 de coeficientes de error de predicción z) puede venir dado por: $z = x - \rho y$, donde x es el vector objetivo (por ejemplo, el bloque actual 140 de coeficientes de transformada aplanados o el bloque actual 131 de coeficientes de transformada), y es un vector que representa el candidato elegido para la predicción (por ejemplo, un bloque previo 149 de coeficientes reconstruidos), y ρ es la ganancia de predicción (escalar).

$w \geq 0$ puede ser un vector de ponderación usado para la determinación de la ganancia de predictor ρ . En algunas formas de realización, el vector de ponderación depende de la envolvente de señal (por ejemplo, de la envolvente ajustada 139, que puede estimarse en el codificador 100, 170 y después transmitirse al descodificador 500). El vector de ponderación tiene normalmente la misma dimensión que el vector objetivo y el vector candidato. Una i -ésima entrada del vector x puede denotarse como x_i (por ejemplo, $i=1, \dots, K$).

Hay diferentes maneras de definir la ganancia de predictor ρ . En una forma de realización, la ganancia de predictor ρ es una ganancia de MMSE (error cuadrático medio mínimo) definida según el criterio de error cuadrático medio mínimo. En este caso, la ganancia de predictor ρ puede calcularse usando la siguiente fórmula:

$$\rho = \frac{\sum_i x_i y_i}{\sum_i y_i^2}.$$

$$D = \sum_i (x_i - \rho y_i)^2.$$

Tal ganancia de predictor ρ normalmente minimiza el error cuadrático medio definido como

A menudo es (perceptualmente) beneficioso introducir una ponderación en la definición del error cuadrático medio D . La ponderación puede usarse para enfatizar la importancia de una correspondencia entre x e y para porciones perceptualmente importantes del espectro de señal y dejar de enfatizar la importancia de una correspondencia entre x e y para porciones del espectro de señal que son relativamente menos importantes. Tal enfoque da como

resultado el siguiente criterio de error: $D = \sum_i (x_i - \rho y_i)^2 w_i$, lo que da lugar a la siguiente definición de la ganancia de predictor óptima (en el sentido del error cuadrático medio ponderado):

$$\rho = \frac{\sum_i w_i x_i y_i}{\sum_i w_i y_i^2}.$$

La anterior definición de la ganancia de predictor da normalmente como resultado una ganancia que no está acotada. Como se ha indicado anteriormente, los pesos w_i del vector de ponderación w pueden determinarse en función de la envolvente ajustada 139. Por ejemplo, el vector de ponderación w puede determinarse usando una función predefinida de la envolvente ajustada 139. La función predefinida puede conocerse en el codificador y en el descodificador (lo que también es el caso para la envolvente ajustada 139). Por tanto, el vector de ponderación puede determinarse de la misma manera en el codificador y en el descodificador.

Otra posible fórmula de ganancia de predictor viene dada por

$$\rho = \frac{2C}{E_x - E_y},$$

donde $C = \sum_i w_i x_i y_i$, $E_x = \sum_i w_i x_i^2$ y $E_y = \sum_i w_i y_i^2$. Esta definición de la ganancia de predictor da como resultado una ganancia que está siempre dentro del intervalo $[-1, 1]$. Una característica importante de la ganancia de predictor especificada por la segunda fórmula es que la ganancia de predictor ρ facilita una relación factible entre la energía de la señal objetivo x y la energía de la señal residual z . La energía residual LTP puede expresarse como:

$\sum_i w_i z_i^2 - E_x (1 - \rho^2)$. El parámetro de control rfu 146 puede determinarse en función de la ganancia de predictor g usando las fórmulas mencionadas anteriormente. La ganancia de predictor g puede ser igual a la ganancia de predictor ρ , determinada usando cualquiera de las fórmulas mencionadas anteriormente.

Como se ha mencionado anteriormente, el codificador 100, 170 está configurado para cuantificar y codificar el vector residual z (es decir, el bloque 141 de coeficientes de error de predicción). El proceso de cuantificación es guiado normalmente por la envolvente de señal (por ejemplo, por la envolvente de asignación 138) según un modelo perceptual subyacente con el fin de distribuir los bits disponibles entre las componentes espectrales de la señal de una manera perceptualmente significativa. El proceso de asignación de velocidad es guiado por la envolvente de señal (por ejemplo, por la envolvente de asignación 138), que se obtiene a partir de la señal de entrada (por ejemplo, a partir del bloque 131 de coeficientes de transformada). La operación del predictor 117 cambia normalmente la envolvente de señal. La unidad de cuantificación 112 usa normalmente cuantificadores que están diseñados suponiendo el funcionamiento en una fuente de varianza unitaria. Cabe señalar que en caso de una predicción de alta calidad (es decir, cuando el predictor 117 es satisfactorio), la propiedad de varianza unitaria puede no cumplirse, es decir, el bloque 141 de coeficientes de error de predicción puede no presentar una varianza unitaria.

Normalmente no es eficaz estimar la envolvente del bloque 141 de coeficientes de error de predicción (es decir, para la señal residual z) ni transmitir esta envolvente al descodificador (ni volver a aplanar el bloque 141 de coeficientes de error de predicción usando la envolvente estimada). En cambio, el codificador 100 y el descodificador 500 pueden usar una regla heurística para reescalar el bloque 141 de coeficientes de error de predicción (como se ha indicado anteriormente). La regla heurística puede usarse para reescalar el bloque 141 de coeficientes de error de predicción, de manera que el bloque 142 de coeficientes reescalados se aproxima a la varianza unitaria. En consecuencia, los resultados de la cuantificación pueden mejorarse (usando cuantificadores que suponen una varianza unitaria). Además, como ya se ha descrito anteriormente, la regla heurística puede usarse para modificar la envolvente de asignación 138, que se usa en el proceso de asignación de bits. La modificación de la envolvente de asignación 138 y el reescalado del bloque 141 de los coeficientes de error de predicción se realizan normalmente por el codificador 100 y por el descodificador 500 de la misma manera (usando la misma regla heurística).

Una posible regla heurística $d(f)$ se ha descrito anteriormente. A continuación se describe otro enfoque para determinar una regla heurística. Una inversa de la ganancia de predicción de energía de dominio ponderado puede obtenerse mediante $p \in [0,1]$, de manera que $\|z\|_w^2 = p\|x\|_w^2$, donde $\|z\|_w^2$ indica la energía elevada al cuadrado del vector residual (es decir, el bloque 141 de coeficientes de error de predicción) en el dominio ponderado y donde $\|x\|_w^2$ indica la energía elevada al cuadrado del vector objetivo (es decir, el bloque 140 de coeficientes de transformada aplanados) en el dominio ponderado.

Pueden adoptarse las siguientes suposiciones:

10 1. Las entradas del vector objetivo x tienen una varianza unitaria. Esto puede ser resultado del aplanamiento realizado por la unidad de aplanamiento 108. Esta suposición se satisface dependiendo de la calidad de la envolvente en función del aplanamiento realizado por la unidad de aplanamiento 108.

15 2. La varianza de las entradas del vector residual de predicción z son de la forma de $E\{z^2(i)} = \min\left\{\frac{t}{w(i)}, 1\right\}$ para $i = 1, \dots, K$ y para algunos valores de $t \geq 0$. Esta suposición está basada en la heurística de que una búsqueda de predictor orientada a mínimos cuadrados da lugar a una contribución de error distribuida de manera uniforme en el dominio ponderado, de manera que el vector residual $\sqrt{w}z$ es más o menos plano. Además, cabe esperar que el candidato a predictor sea casi plano, lo que da lugar a la relación razonable de $E\{z^2(i)} \leq 1$. Debe observarse que pueden usarse varias modificaciones de esta segunda suposición.

20 Para poder estimar el parámetro t , puede usarse las dos suposiciones antes mencionadas en la fórmula de error de predicción (por ejemplo, $D = \sum_i (x_i - \rho v_i)^2 w_i$) y proporcionar así la ecuación de "tipo de nivel de agua"

$$\sum_i \min\{t, w(i)\} = p \sum_i w(i)$$

25 Puede demostrarse que hay una solución a la anterior ecuación en el intervalo $t \in [0, \max(w(i))]$. La ecuación para encontrar el parámetro t puede resolverse usando rutinas de ordenación.

La regla heurística puede proporcionarse después mediante $d(i) = \max\left\{\frac{w(i)}{t}, 1\right\}$, donde $i = 1, \dots, K$ identifica la

30 celda de frecuencia. La inversa de la regla de escalado heurística viene dada por $\frac{1}{d(i)} = \min\left\{\frac{t}{w(i)}, 1\right\}$. La inversa de la regla de escalado heurística se aplica mediante la unidad de reescalado inverso 113. La regla de escalado dependiente de la frecuencia depende de los pesos $w(i) = w_i$. Como se ha indicado anteriormente, los pesos $w(i)$ pueden depender de o pueden corresponder al bloque actual 131 de coeficientes de transformada (por ejemplo, la envolvente ajustada 139, o alguna función predefinida de la envolvente ajustada 139).

35 Puede demostrarse que cuando se usa la fórmula $\rho = \frac{2C}{E_x + E_v}$ para determinar la ganancia de predictor, se aplica la siguiente relación: $p = 1 - \rho^2$.

40 Por tanto, una regla de escalado heurística puede determinarse de varias maneras diferentes. Se ha observado de manera experimental que la regla de escalado que se determina en función de las dos suposiciones antes mencionadas (denominadas como procedimiento de escalado B) es ventajosa en comparación con la regla de escalado fija $d(f)$. En particular, la regla de escalado que se determina en función de las dos suposiciones puede tener en cuenta el efecto de ponderación usado en el transcurso de una búsqueda de candidatos a predictor. El

45 procedimiento de escalado B se combina de manera conveniente con la definición de la ganancia $\rho = \frac{2C}{E_x + E_v}$, debido a la relación analíticamente factible entre la varianza de la señal residual y la varianza de la señal (lo que facilita la obtención de p como se ha descrito anteriormente).

A continuación se describe un aspecto adicional para mejorar el rendimiento del codificador de audio basado en transformada. En particular, se propone el uso de una denominada bandera de conservación de varianza. La bandera de conservación de varianza puede determinarse y transmitirse en cada bloque 131. La bandera de conservación de varianza puede indicar la calidad de la predicción. En una forma de realización, la bandera de conservación de varianza está inactiva en caso de una calidad de predicción relativamente alta, y la bandera de conservación de varianza está activa en caso de una calidad de predicción relativamente baja. La bandera de conservación de varianza puede determinarse por el codificador 100, 170, por ejemplo en función de la ganancia de

predictor p y/o en función de la ganancia de predictor g . A modo de ejemplo, la bandera de conservación de varianza puede fijarse a "activa" si la ganancia de predictor p o g (o un parámetro derivado de la misma) es inferior a un umbral predeterminado (por ejemplo, 2 dB) y viceversa. Como se ha observado anteriormente, la inversa de la ganancia de predicción de energía de dominio ponderado p depende normalmente de la ganancia de predictor, por ejemplo $p = 1 - \rho^2$. La inversa del parámetro p puede usarse para determinar un valor de la bandera de conservación de varianza. A modo de ejemplo, $1/p$ (por ejemplo, expresado en dB) puede compararse con un umbral predeterminado (por ejemplo, de 2dB), con el fin de determinar el valor de la bandera de conservación de varianza. Si $1/p$ es mayor que el umbral predeterminado, la bandera de conservación de varianza puede fijarse a "inactiva" (lo que indica una calidad de predicción relativamente alta), y viceversa.

La bandera de conservación de varianza puede usarse para controlar diferentes ajustes del codificador 100 y del descodificador 500. En particular, la bandera de conservación de varianza puede usarse para controlar el grado de ruido de la pluralidad de cuantificadores 321, 322, 323. En particular, la bandera de conservación de varianza puede afectar a uno o más de los siguientes ajustes:

- Ganancia de ruido adaptativa para la asignación de bits con valor cero. Dicho de otro modo, la ganancia de ruido del cuantificador de síntesis de ruido 321 puede verse afectada por la bandera de conservación de varianza.
- Intervalo de cuantificadores con ruido aleatorio. Dicho de otro modo, el intervalo 324, 325 de SNR para el que se usan cuantificadores con ruido aleatorio 322 puede verse afectada por la bandera de conservación de varianza.
- Posganancia de los cuantificadores con ruido aleatorio. Puede aplicarse una posganancia en la salida de los cuantificadores con ruido aleatorio, con el fin de afectar al rendimiento del error cuadrático medio de los cuantificadores con ruido aleatorio. La posganancia puede depender de la bandera de conservación de varianza.
- Aplicación de escalado heurístico. El uso de escalado heurístico (en la unidad de reescalado 111 y en la unidad de reescalado inverso 113) puede depender de la bandera de conservación de varianza.

Un ejemplo de cómo la bandera de conservación de varianza puede modificar uno o más ajustes del codificador 100 y/o del descodificador 500 se proporciona en la Tabla 2.

Tabla 2

Tipo de ajuste	Conservación de varianza inactiva	Conservación de varianza activa
Ganancia de ruido	$g_N = (1 - rfu)$	$g_N = \sqrt{(1 - rfu^2)}$
Intervalo de cuantificadores con ruido aleatorio	Depende del parámetro de control rfu	Está fijado a un intervalo relativamente grande (por ejemplo, al intervalo más grande posible)
Posganancia de los cuantificadores con ruido aleatorio $\gamma_o = \frac{\sigma_x^2}{\sigma_x^2 + 12} \Delta^2; \gamma_1 = \sqrt{\gamma_o}$	$\gamma = \gamma_o$	$\gamma = \max(\gamma_o, g_N \cdot \gamma_1)$
Regla de escalado heurística	activa	inactiva

En la fórmula para la posganancia, $\sigma_x^2 = E\{X^2\}$ es la varianza de uno o más de los coeficientes del bloque 141 de coeficientes de error de predicción (que van a cuantificarse), y Δ es un tamaño de etapa de cuantificador de un cuantificador escalar (612) del cuantificador con ruido aleatorio al que se aplica la posganancia.

Como puede observarse a partir del ejemplo de la Tabla 2, la ganancia de ruido g_N del cuantificador de síntesis de ruido 321 (es decir, la varianza del cuantificador de síntesis de ruido 321) puede depender de la bandera de conservación de varianza. Como se ha indicado anteriormente, el parámetro de control rfu 146 puede estar en el intervalo [0, 1], donde un valor relativamente bajo de rfu indica una calidad de predicción relativamente baja, y un valor relativamente alto de rfu indica una calidad de predicción relativamente alta. Para valores de rfu en el intervalo de [0, 1], la fórmula de la columna izquierda proporciona ganancias de ruido más bajas g_N que la fórmula de la columna derecha. Por tanto, cuando la bandera de conservación de varianza está activa (lo que indica una calidad

de predicción relativamente baja), se usa una ganancia de ruido más alta a cuando la bandera de conservación de varianza está inactiva (lo que indica una calidad de predicción relativamente alta). Se ha observado experimentalmente que esto mejora la calidad perceptiva global.

5 Como se ha descrito anteriormente, el intervalo de SNR de 324, 325 de los cuantificadores con ruido aleatorio 322 puede variar dependiendo del parámetro de control rfu. Según la tabla 2, cuando la bandera de conservación de
 10 varianza está activa (lo que indica una calidad de predicción relativamente baja), se usa un gran intervalo fijo de cuantificadores con ruido aleatorio 322 (por ejemplo, en el intervalo 324). Por otro lado, cuando la bandera de conservación de varianza está inactiva (lo que indica una calidad de predicción relativamente alta), se usan diferentes intervalos 324, 325, dependiendo del parámetro de control rfu.

La determinación del bloque 145 de coeficientes de error cuantificados puede implicar la aplicación de una posganancia γ en los coeficientes de error cuantificados, que se han cuantificado usando un cuantificador con ruido
 15 aleatorio 322. La posganancia γ puede obtenerse para mejorar el rendimiento MSE de un cuantificador con ruido aleatorio 322 (por ejemplo, un cuantificador que suprime el ruido aleatorio). La posganancia puede obtenerse mediante:

$$\gamma = \frac{\sigma_x^2}{\sigma_x^2 + \frac{\Delta^2}{12}}$$

20 Se ha observado experimentalmente que la calidad de codificación perceptual puede mejorarse cuando la posganancia depende de la bandera de conservación de varianza. Se usa la posganancia óptima MSE antes mencionada, cuando la bandera de conservación de varianza está inactiva (lo que indica una calidad de predicción
 25 relativamente alta). Por otro lado, cuando la bandera de conservación de varianza está activa (lo que indica una calidad de predicción relativamente baja), puede ser beneficioso usar una posganancia superior (determinada según la fórmula del lado derecho de la Tabla 2).

Como se ha indicado anteriormente, el escalado heurístico puede usarse para proporcionar bloques 142 de
 30 coeficientes de error reescalados que están más cerca de la propiedad de varianza unitaria que los bloques 141 de coeficientes de error de predicción. Las reglas de escalado heurísticas pueden depender del parámetro de control 146. Dicho de otro modo, las reglas de escalado heurísticas pueden depender de la calidad de predicción. El escalado heurístico puede ser particularmente beneficioso en caso de una calidad de predicción relativamente alta, mientras que los beneficios pueden limitarse en caso de una calidad de predicción relativamente baja. En vista de esto, puede ser beneficioso usar solamente el escalado heurístico cuando la bandera de conservación de varianza está inactiva (lo que indica una calidad de predicción relativamente alta).

35 En el presente documento se han descrito un codificador de voz basado en transformada 100, 170 y un descodificador de voz basado en transformada 500 correspondiente. El códec de voz basado en transformada puede usar varios aspectos que permiten mejorar la calidad de las señales de voz codificadas. El códec de voz puede usar bloques relativamente cortos (también denominados unidades de codificación), por ejemplo en el
 40 intervalo de 5 ms, lo que garantiza una resolución de tiempo apropiada y estadísticas útiles para señales de voz. Además, el códec de voz puede proporcionar una descripción adecuada de una envolvente espectral variable en el tiempo de las unidades de codificación. Además, el códec de voz puede usar predicción en el dominio de transformada, donde la predicción puede tener en cuenta las envolventes espectrales de las unidades de codificación. Por tanto, el códec de voz puede proporcionar actualizaciones predictivas de descubrimiento de
 45 envolvente a las unidades de codificación. Además, el códec de voz puede usar cuantificadores predeterminados que se adaptan a los resultados de la predicción. Dicho de otro modo, el códec de voz puede usar cuantificadores escalares adaptativos de predicción.

50 Los procedimientos y sistemas descritos en el presente documento pueden implementarse como software, firmware y/o hardware. Determinados componentes pueden implementarse, por ejemplo, como software que se ejecuta en un procesador o microprocesador de señales digitales. Otros componentes pueden implementarse, por ejemplo, como hardware y/o como circuitos integrados de aplicación específica. Las señales encontradas en los procedimientos y sistemas descritos pueden almacenarse en medios tales como una memoria de acceso aleatorio o medios de almacenamiento óptico. Pueden transferirse a través de redes, tales como redes radioeléctricas, redes por satélite,
 55 redes inalámbricas o redes cableadas, por ejemplo Internet. Dispositivos típicos que usan los procedimientos y sistemas descritos en el presente documento son dispositivos electrónicos portátiles u otros equipos de usuario que se usan para almacenar y/o reproducir señales de audio.

REIVINDICACIONES

1. Un codificador de voz basado en transformada (100, 170) configurado para codificar una señal de voz en un flujo de bits; comprendiendo el codificador (100, 170):

- 5
- una unidad de generación de tramas (101) configurada para recibir un conjunto (132, 332) de bloques; donde el conjunto (132, 332) de bloques comprende una pluralidad de bloques secuenciales (131) de coeficientes de transformada; donde la pluralidad de bloques (131) indica muestras de la señal de voz; donde un bloque (131) de coeficientes de transformada comprende una pluralidad de coeficientes de transformada para una pluralidad correspondiente de celdas de frecuencia (301);
 - 10 - una unidad de estimación de envolvente (102) configurada para determinar una envolvente actual (133) en función de la pluralidad de bloques secuenciales (131) de coeficientes de transformada; donde la envolvente actual (133) indica una pluralidad de valores de energía espectral (303) para la pluralidad correspondiente de celdas de frecuencia (301);
 - 15 - una unidad de cuantificación de envolvente (103) configurada para determinar una envolvente actual cuantificada (134) cuantificando la envolvente actual (133);
 - una unidad de interpolación de envolvente (104) configurada para determinar una pluralidad de envolventes interpoladas (136) para la pluralidad de bloques (131) de coeficientes de transformada, respectivamente, en función de la envolvente actual cuantificada (134) y en función de una envolvente previa cuantificada (134); y
 - 20 - una unidad de aplanamiento (108) configurada para determinar una pluralidad de bloques (140) de coeficientes de transformada aplanados mediante el aplanamiento de la pluralidad correspondiente de bloques (131) de coeficientes de transformada usando la pluralidad correspondiente de envolventes interpoladas (136), respectivamente; donde el flujo de bits se determina en función de la pluralidad de bloques (140) de coeficientes de transformada aplanados.

- 25
2. El codificador de voz basado en transformada (100, 170) según la reivindicación 1, en el que:
- 30 - el codificador de voz basado en transformada (100, 170) comprende además una unidad de determinación de ganancia de envolvente (105, 106) configurada para determinar una pluralidad de ganancias de envolvente (137) para la pluralidad de bloques (131) de coeficientes de transformada, respectivamente;
 - el codificador de voz basado en transformada (100, 170) comprende además una unidad de refinamiento de envolvente (107) configurada para determinar una pluralidad de envolventes ajustadas (139) desfasando valores de energía espectral (303) de la pluralidad de envolventes interpoladas (136) según la pluralidad de ganancias de envolvente (137), respectivamente;
 - 35 - la unidad de aplanamiento (108) está configurada para determinar la pluralidad de bloques (140) de coeficientes de transformada aplanados mediante el aplanamiento de la pluralidad correspondiente de bloques (131) de coeficientes de transformada usando la pluralidad correspondiente de envolventes ajustadas (139), respectivamente.

- 40
3. El codificador de voz basado en transformada (100, 170) según la reivindicación 2, en el que la unidad de determinación de ganancia de envolvente (105, 106) está configurada para determinar una primera ganancia de envolvente (137) de un primer bloque (131) de coeficientes de transformada, de manera que una varianza de los coeficientes de transformada aplanados de un primer bloque correspondiente (140) de coeficientes de transformada aplanados obtenidos usando una primera envolvente ajustada (139) se ajusta en comparación con una varianza de los coeficientes de transformada aplanados de un primer bloque correspondiente (140) de coeficientes de transformada aplanados obtenidos usando una primera envolvente interpolada (139), donde, opcionalmente, la unidad de determinación de ganancia de envolvente (105, 106) está configurada para determinar la primera ganancia de envolvente (137) para el primer bloque (131) de coeficientes de transformada, de manera que la varianza de los coeficientes de transformada aplanados del primer bloque correspondiente (140) de coeficientes de transformada aplanados obtenidos usando la primera envolvente ajustada (139) es uno.
- 50

4. El codificador de voz basado en transformada (100, 170) según cualquier reivindicación anterior, en el que:

- 55
- la envolvente actual (133) indica una pluralidad de valores de energía espectral (303) para una pluralidad correspondiente de bandas de frecuencia (302);
 - una banda de frecuencias (302) comprende una o más celdas de frecuencia (301);
 - la unidad de estimación de envolvente (102) está configurada para determinar el valor de energía espectral (303) para una banda de frecuencia particular (302) en función de los coeficientes de transformada de la pluralidad de bloques secuenciales (131) para la banda de frecuencia particular (302).

5. El codificador de voz basado en transformada (100, 170) según cualquier reivindicación anterior, en el que:

- un bloque (131) de coeficientes de transformada comprende coeficientes MDCT; y/o

5 - un bloque (131) de coeficientes de transformada comprende 256 coeficientes de transformada en 256 celdas de frecuencia (301); y/o

- un conjunto (132, 332) de bloques comprende cuatro o más bloques (131) de coeficientes de transformada.

6. El codificador de voz basado en transformada (100, 170) según cualquier reivindicación anterior, en el que:

10 - el codificador de voz basado en transformada (100, 170) está configurado para funcionar en una pluralidad de modos diferentes que comprenden un modo de corto espaciamiento y un modo de largo espaciamiento;

15 - la unidad de generación de tramas (101), la unidad de estimación de envolvente (102) y la unidad de interpolación de envolvente (104) están configuradas para procesar el conjunto (132, 332) de bloques que comprenden la pluralidad de bloques secuenciales (131) de coeficientes de transformada, cuando el codificador de voz basado en transformada (100, 170) se hace funcionar en el modo de corto espaciamiento; y

20 - la unidad de generación de tramas (101), la unidad de estimación de envolvente (102) y la unidad de interpolación de envolvente (104) están configuradas para procesar un conjunto (132, 332) de bloques que comprenden un único bloque (131) de coeficientes de transformada, cuando el codificador de voz basado en transformada (100, 170) se hace funcionar en el modo de largo espaciamiento, donde, opcionalmente, cuando en el modo de largo espaciamiento,

- la unidad de estimación de envolvente (102) está configurada para determinar una envolvente actual (133) del único bloque (131) de coeficientes de transformada comprendidos en el conjunto (132, 133) de bloques; y

25 - la unidad de interpolación de envolvente (104) está configurada para determinar una envolvente interpolada (136) del único bloque (131) de coeficientes de transformada como la envolvente actual (133) del único bloque (131) de coeficientes de transformada.

7. Un decodificador de voz basado en transformada (500) configurado para decodificar un flujo de bits para proporcionar una señal de voz reconstruida; comprendiendo el decodificador (500):

30 - una unidad de descodificación de envolvente (531), configurada para determinar una envolvente actual cuantificada (134) a partir de datos de envolvente (161) comprendidos en el flujo de bits; donde la envolvente actual cuantificada (134) indica una pluralidad de valores de energía espectral (303) para una pluralidad correspondiente de celdas de frecuencia (301); donde el flujo de bits comprende datos (163, 164) que indican una pluralidad de bloques secuenciales (148) de coeficientes de transformada aplanados reconstruidos; donde un bloque (148) de coeficientes de transformada aplanados reconstruidos comprende una pluralidad de coeficientes de transformada aplanados reconstruidos para la pluralidad correspondiente de celdas de frecuencia (301);

40 - una unidad de interpolación de envolvente (104) configurada para determinar una pluralidad de envolventes interpoladas (136) para la pluralidad de bloques (148) de coeficientes de transformada aplanados reconstruidos, respectivamente, en función de la envolvente actual cuantificada (134) y en función de una envolvente previa cuantificada (134); y

45 - una unidad de aplanamiento inverso (108) configurada para determinar una pluralidad de bloques (149) de coeficientes de transformada reconstruidos dotando a la pluralidad correspondiente de bloques (148) de coeficientes de transformada aplanados reconstruidos de una forma espectral, usando la pluralidad correspondiente de envolventes interpoladas (136), respectivamente; donde la señal de voz reconstruida se determina en función de la pluralidad de bloques (149) de coeficientes de transformada reconstruidos, donde, opcionalmente, la envolvente previa cuantificada (134) está asociada a una pluralidad de bloques previos (149) de coeficientes de transformada reconstruidos, directamente anteriores a la pluralidad de bloques (149) de coeficientes de transformada reconstruidos.

8. El decodificador de voz basado en transformada (500) según la reivindicación 7, en el que:

55 - la pluralidad de bloques secuenciales (148) de coeficientes de transformada aplanados reconstruidos comprende un primer bloque (148) de coeficientes de transformada aplanados reconstruidos en un primer instante de tiempo intermedio;

- la unidad de interpolación de envolvente (104) está configurada para determinar un valor de energía espectral (303) para una celda de frecuencia particular (301) de una primera envolvente interpolada (136)

interpolando los valores de energía espectral (303) para la celda de frecuencia particular (301) de la envolvente actual cuantificada (135) y de la envolvente previa cuantificada (134) en el primer instante de tiempo intermedio;

5 - la primera envolvente interpolada (136) está asociada al primer bloque (148) de coeficientes de transformada aplanados reconstruidos.

9. El descodificador de voz basado en transformada (500) según la reivindicación 8, en el que la unidad de interpolación de envolvente (104) está configurada para determinar el valor de energía espectral (303) para la celda de frecuencia particular (301) de la primera envolvente interpolada (136) cuantificando la interpolación entre los valores de energía espectral (303) para la celda de frecuencia particular (301) de la envolvente actual cuantificada (135) y de la envolvente previa cuantificada (134).

10. El descodificador de voz basado en transformada (500) según cualquiera de las reivindicaciones 8 o 9, en el que:

15 - la pluralidad de bloques secuenciales (148) de coeficientes de transformada aplanados reconstruidos comprende un segundo bloque (148) de coeficientes de transformada aplanados reconstruidos en un segundo instante de tiempo intermedio;

20 - la unidad de interpolación de envolvente (104) está configurada para determinar un valor de energía espectral (303) para la celda de frecuencia particular (301) de una segunda envolvente interpolada (136) interpolando los valores de energía espectral (303) para la celda de frecuencia particular (301) de la envolvente actual cuantificada (135) y de la envolvente previa cuantificada (134) en el segundo instante de tiempo intermedio;

- la segunda envolvente interpolada (136) está asociada al segundo bloque (148) de coeficientes de transformada aplanados reconstruidos;

25 - el segundo bloque (148) de coeficientes de transformada aplanados reconstruidos es subsiguiente al primer bloque (148) de coeficientes de transformada aplanados reconstruidos; y

30 - el segundo instante de tiempo intermedio es posterior al primer instante de tiempo intermedio, donde una diferencia entre el segundo instante de tiempo intermedio y el primer instante de tiempo intermedio corresponde a un intervalo de tiempo entre el segundo bloque de coeficientes de transformada aplanados reconstruidos y el primer bloque (148) de coeficientes de transformada aplanados reconstruidos.

11. El descodificador de voz basado en transformada (500) según cualquiera de las reivindicaciones 7 a 10, en el que:

35 - el flujo de bits indica una pluralidad de ganancias de envolvente (137) para la pluralidad de bloques (148) de coeficientes de transformada aplanados reconstruidos, respectivamente;

40 - el descodificador de voz basado en transformada (500) comprende además una unidad de refinamiento de envolvente (107) configurada para determinar una pluralidad de envolventes ajustadas (139) aplicando la pluralidad de ganancias de envolvente (137) a la pluralidad de envolventes interpoladas (136), respectivamente;

- la unidad de aplanamiento inverso (108) está configurada para determinar la pluralidad de bloques (149) de coeficientes de transformada reconstruidos dotando a la pluralidad correspondiente de bloques (148) de coeficientes de transformada aplanados reconstruidos de una forma espectral, usando la pluralidad correspondiente de envolventes interpoladas (139), respectivamente.

45 12. Un procedimiento para codificar una señal de voz en un flujo de bits; comprendiendo el procedimiento:

50 - recibir un conjunto (132, 332) de bloques; donde el conjunto (132, 332) de bloques comprende una pluralidad de bloques secuenciales (131) de coeficientes de transformada; donde la pluralidad de bloques secuenciales (131) indica muestras de la señal de voz; donde un bloque (131) de coeficientes de transformada comprende una pluralidad de coeficientes de transformada para una pluralidad correspondiente de celdas de frecuencia (301); determinar una envolvente actual (133) en función de la pluralidad de bloques secuenciales (131) de coeficientes de transformada; donde la envolvente actual (133) indica una pluralidad de valores de energía espectral (303) para la pluralidad correspondiente de celdas de frecuencia (301);

55 - determinar una envolvente actual cuantificada (134) cuantificando la envolvente actual (133);

- determinar una pluralidad de envolventes interpoladas (136) para la pluralidad de bloques (131) de coeficientes de transformada, respectivamente, en función de la envolvente actual cuantificada (133) y en función de una envolvente previa cuantificada (134);

- determinar una pluralidad de bloques (140) de coeficientes de transformada aplanados mediante el aplanamiento de la pluralidad correspondiente de bloques (131) de coeficientes de transformada usando la pluralidad correspondiente de envolventes interpoladas (136), respectivamente; y
 - determinar el flujo de bits en función de la pluralidad de bloques (140) de coeficientes de transformada aplanados.
- 5
13. Un procedimiento para descodificar un flujo de bits para proporcionar una señal de voz reconstruida, comprendiendo el procedimiento:
- determinar una envolvente actual cuantificada (134) a partir de datos de envolvente (161) comprendidos en el flujo de bits; donde la envolvente actual cuantificada (134) indica una pluralidad de valores de energía espectral (303) para una pluralidad correspondiente de celdas de frecuencia (301); donde el flujo de bits comprende datos (163, 164) que indican una pluralidad de bloques secuenciales (148) de coeficientes de transformada aplanados reconstruidos; donde un bloque (148) de coeficientes de transformada aplanados reconstruidos comprende una pluralidad de coeficientes de transformada aplanados reconstruidos para la pluralidad correspondiente de celdas de frecuencia (301);
 - determinar una pluralidad de envolventes interpoladas (136) para la pluralidad de bloques (148) de coeficientes de transformada aplanados reconstruidos, respectivamente, en función de la envolvente actual cuantificada (134) y en función de una envolvente previa cuantificada (134);
 - determinar una pluralidad de bloques (149) de coeficientes de transformada reconstruidos dotando a la pluralidad correspondiente de bloques (148) de coeficientes de transformada aplanados reconstruidos de una forma espectral, usando la pluralidad correspondiente de envolventes interpoladas (136), respectivamente; y
 - determinar la señal de voz reconstruida en función de la pluralidad de bloques (149) de coeficientes de transformada reconstruidos.
- 10
- 15
- 20
- 25
14. Un procedimiento para codificar una señal de audio que comprende un segmento de voz en un flujo de bits, donde el procedimiento comprende:
- identificar el segmento de voz a partir de la señal de audio;
 - determinar una pluralidad de bloques secuenciales (131) de coeficientes de transformada en función del segmento de voz, usando una unidad de transformada; donde un bloque (131) de coeficientes de transformada comprende una pluralidad de coeficientes de transformada para una pluralidad correspondiente de celdas de frecuencia (301); donde la unidad de transformada está configurada para determinar bloques largos que comprenden un primer número de coeficientes de transformada y bloques cortos que comprenden un segundo número de coeficientes de transformada; donde el primer número es mayor que el segundo número; donde los bloques (131) de la pluralidad de bloques secuenciales (131) son bloques cortos; y
 - codificar la pluralidad de bloques secuenciales (131) en el flujo de bits según la reivindicación 12.
- 30
- 35
- 40
15. Un procedimiento para descodificar un flujo de bits que indica una señal de audio que comprende un segmento de voz, comprendiendo el procedimiento:
- determinar una pluralidad de bloques secuenciales (149) de coeficientes de transformada reconstruidos en función de datos (161, 162, 163, 164) comprendidos en el flujo de bits según la reivindicación 13; y
 - determinar un segmento de voz reconstruido en función de la pluralidad de bloques secuenciales (149) de coeficientes de transformada reconstruidos, usando una unidad de transformada inversa; donde un bloque (149) de coeficientes de transformada reconstruidos comprende una pluralidad de coeficientes de transformada reconstruidos para una pluralidad correspondiente de celdas de frecuencia (301); donde la unidad de transformada inversa (504) está configurada para procesar bloques largos que comprenden un primer número de coeficientes de transformada reconstruidos y bloques cortos que comprenden un segundo número de coeficientes de transformada reconstruidos; donde el primer número es mayor que el segundo número; donde los bloques (149) de la pluralidad de bloques secuenciales (149) son bloques cortos.
- 45
- 50

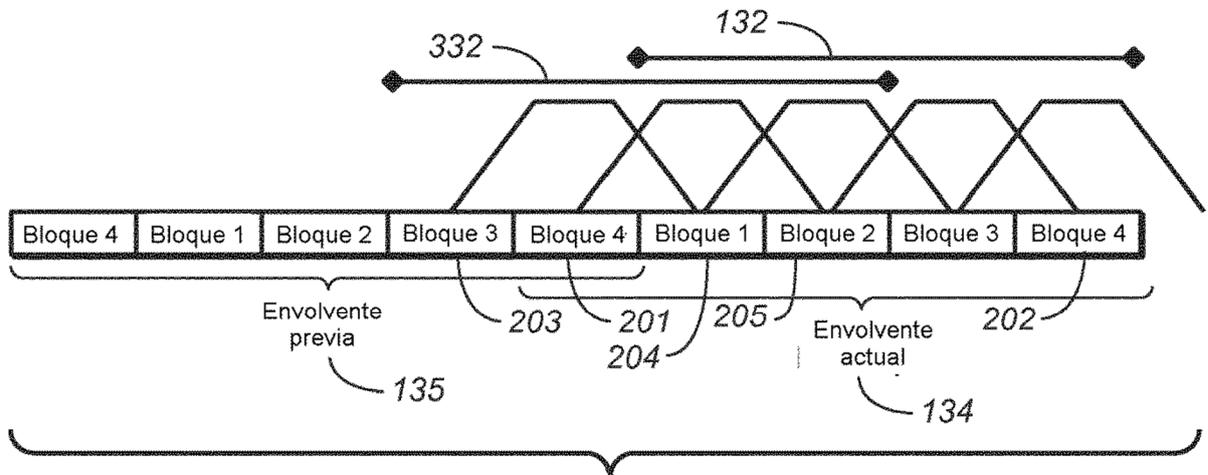


FIG. 2

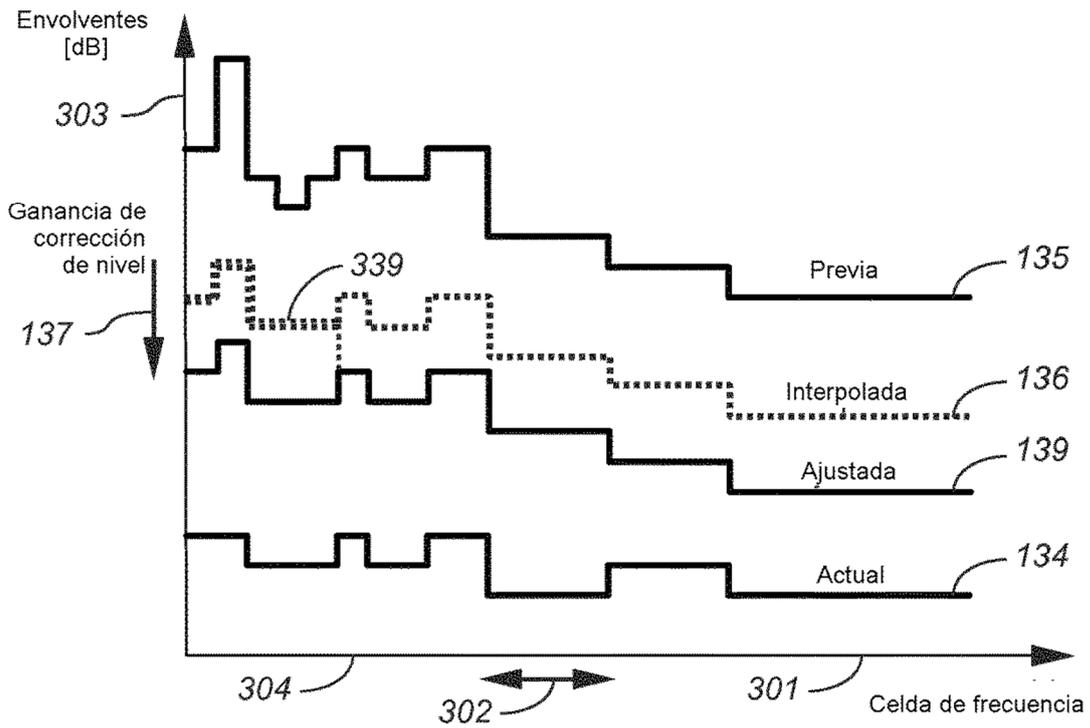


FIG. 3A

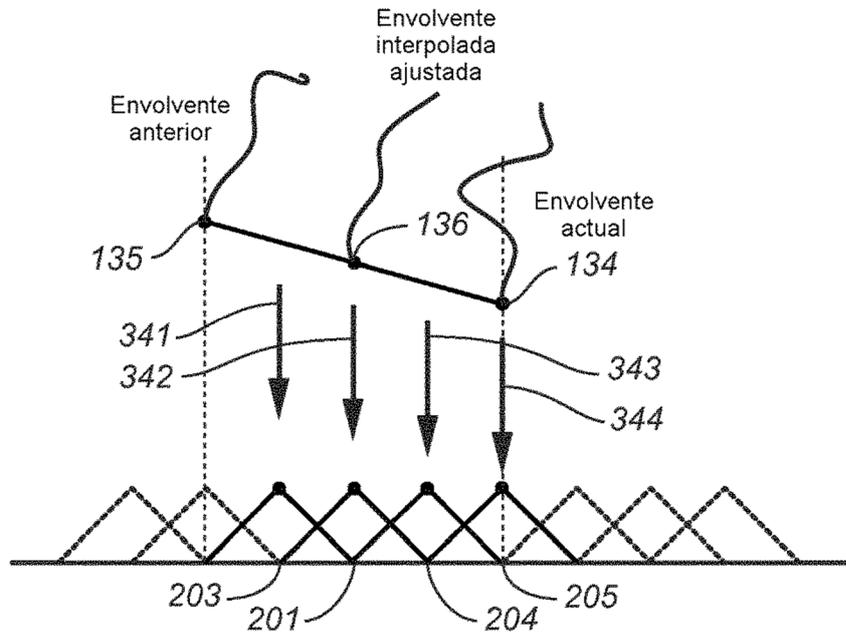


FIG. 3B

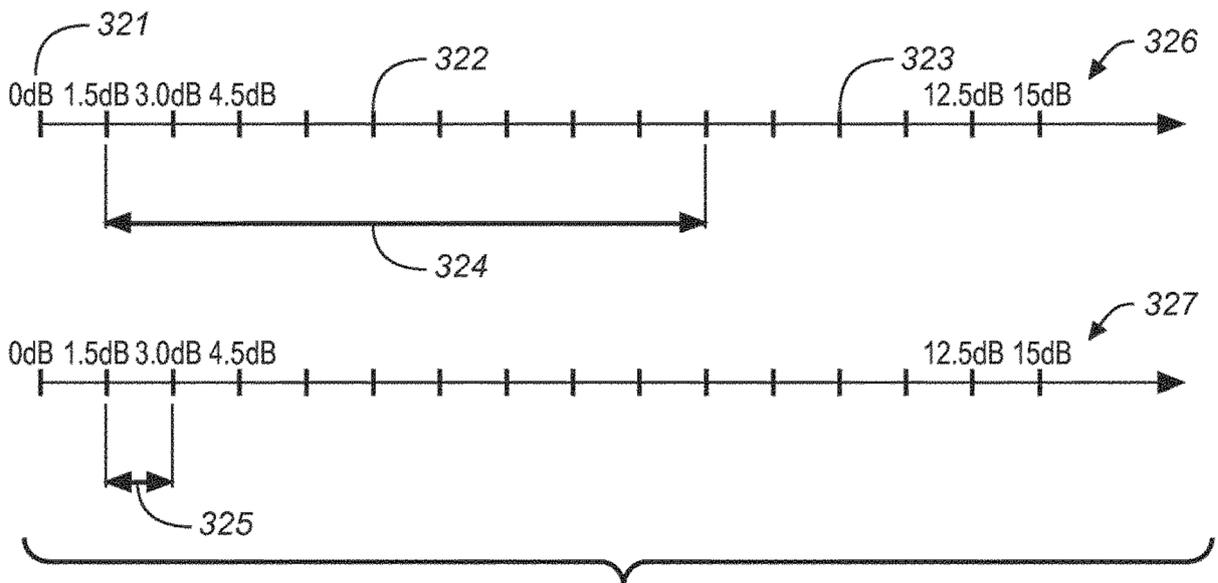


FIG. 4

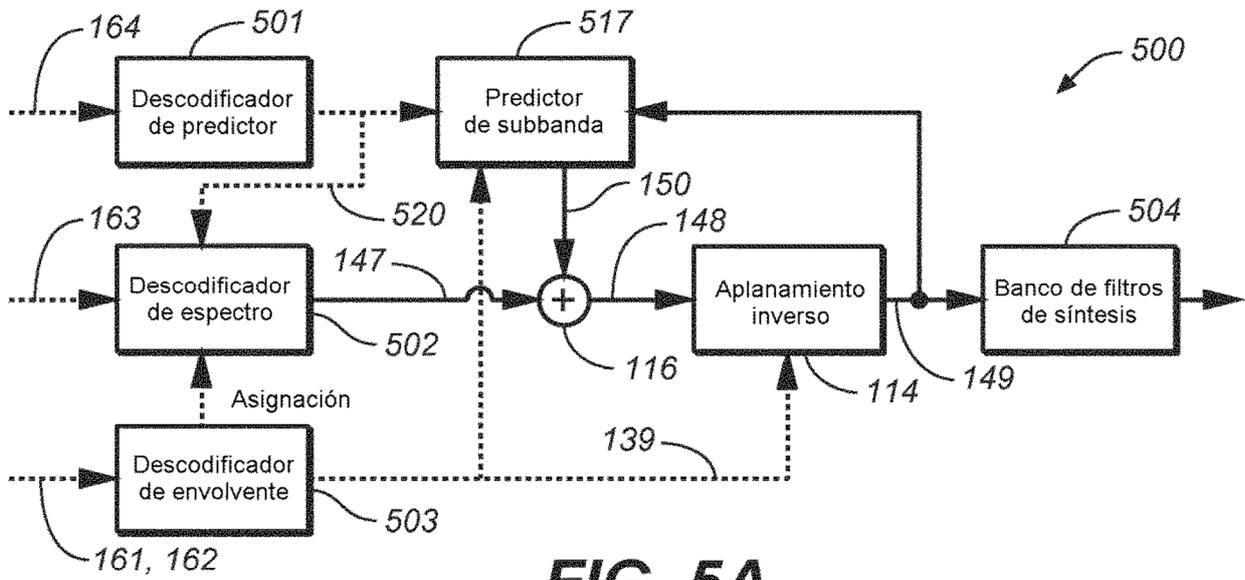


FIG. 5A

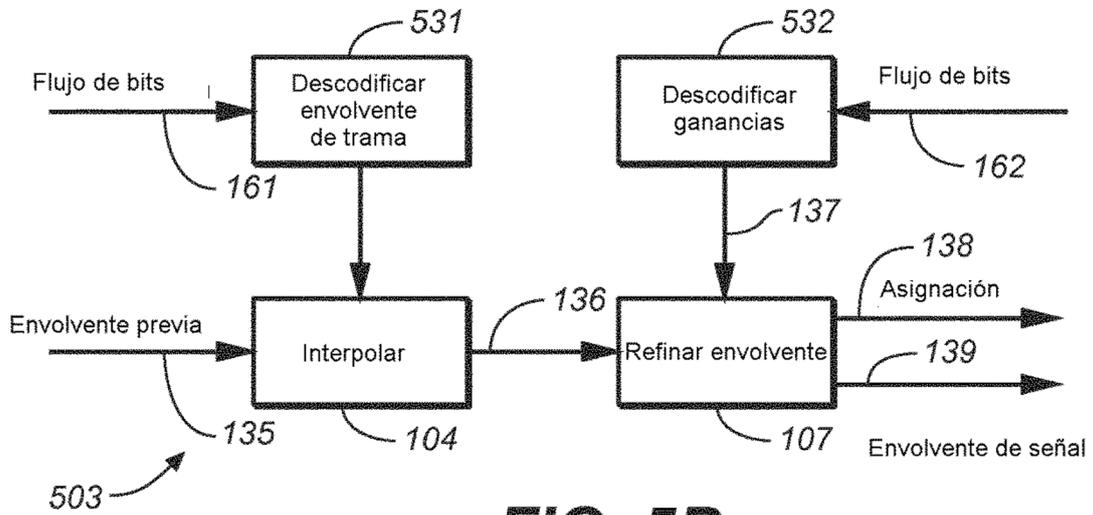


FIG. 5B

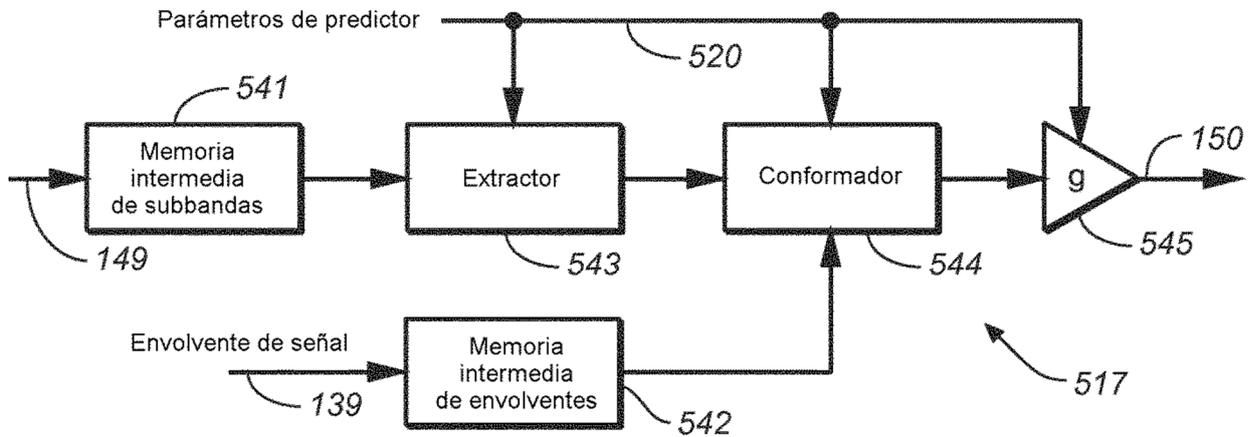


FIG. 5C

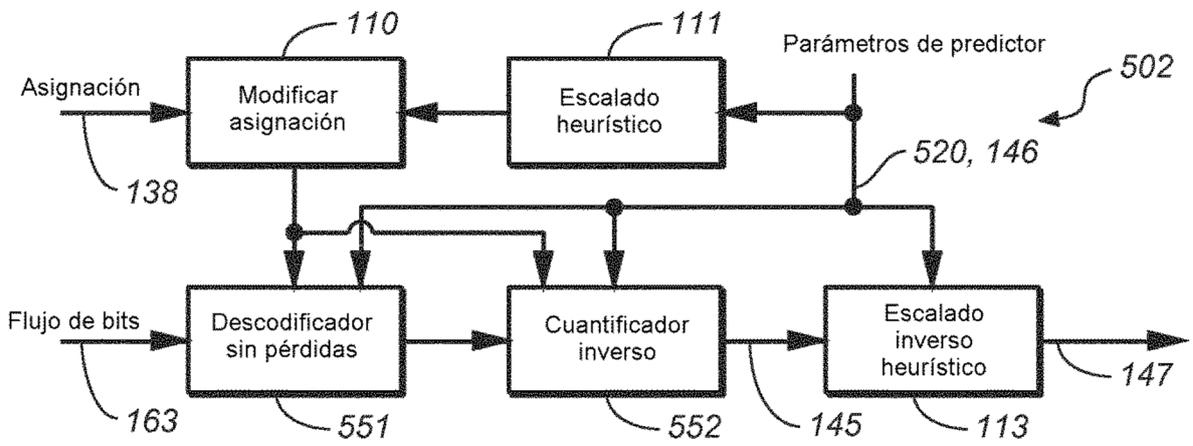


FIG. 5D