

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 932 053**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/22 (2013.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.01.2017 PCT/EP2017/051177**

87 Fecha y número de publicación internacional: **27.07.2017 WO17125544**

96 Fecha de presentación y número de la solicitud europea: **20.01.2017 E 17700980 (0)**

97 Fecha y número de publicación de la concesión europea: **28.09.2022 EP 3405950**

54 Título: **Codificación de audio estéreo con normalización basada en ild antes de la decisión media/lateral**

30 Prioridad:

22.01.2016 EP 16152457

22.01.2016 EP 16152454

21.11.2016 EP 16199895

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
09.01.2023

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)**

**Hansastr. 27c
80686 München, DE**

72 Inventor/es:

**RAVELLI, EMMANUEL;
SCHNELL, MARKUS;
DÖHLA, STEFAN;
JÄGERS, WOLFGANG;
DIETZ, MARTIN;
HELMRICH, CHRISTIAN;
MARKOVIC, GORAN;
FOTOPOULOU, ELENI;
MULTRUS, MARKUS;
BAYER, STEFAN;
FUCHS, GUILLAUME y
HERRE, JÜRGEN**

74 Agente/Representante:

PONTI & PARTNERS, S.L.P.

ES 2 932 053 T3

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

DESCRIPCIÓN

Codificación de audio estéreo con normalización basada en ild antes de la decisión media/lateral

- 5 **[0001]** La presente invención se refiere a la codificación de señales de audio y la decodificación de señales de audio y, en particular, con un aparato y procedimiento para Estéreo MDCT M/S (por sus siglas en inglés, Modified Discrete Cosine Transform MID/Side, Transformada de Coseno Discreta Modificada) con ILD (por sus siglas en inglés, (interaural level difference, diferencia de nivel interaural) Global con Detección Medio/lado mejorada.
- 10 **[0002]** El procesamiento M/S (M/S = Medio/lado) por bandas en los codificadores basados en MDCT (MDCT = Modified Discrete Cosine Transform, Transformada de Coseno Discreta Modificada) es un procedimiento conocido y eficaz para el procesamiento estéreo. Sin embargo, no es suficiente para las señales paneadas y es indispensable un procesamiento adicional, tal como la predicción compleja o una codificación de los ángulos entre un canal medio y uno lateral).
- 15 **[0003]** En [1], [2], [3] y [4], se describe el procesamiento M/S en señales formadas en ventanas y transformadas no normalizadas (no blanqueadas).
- [0004]** En [7], se describe la predicción entre los canales medio y laterales. En [7], se describe un codificador
20 que codifica una señal de audio sobre la base de una combinación de dos canales de audio. El codificador de audio obtiene una señal de combinación que es una señal media y, además obtiene una señal residual de predicción que es una señal lateral predicha derivada de la señal media. La primera señal de combinación y la señal residual de predicción se codifican y escriben en un flujo de datos junto con la información de predicción. Además, [7] describe un decodificador que genera primeros y segundos canales de audio decodificados utilizando la señal residual de
25 predicción, la primera señal de combinación y la información de predicción.
- [0005]** En [5], se describe la aplicación del acoplamiento estéreo M/S después de la normalización de manera independiente por cada banda. En particular, [5] se refiere al códec Opus. Opus codifica la señal media y la señal lateral como señales normalizadas $m = M/||M||$ y $s = S/||S||$. Para recuperar M y S de m y s, se codifica el ángulo $\theta_s =$
30 $\arctan(\arctan(s/||S||/||M||))$. Como N es el tamaño de la banda y a es el número total de bits disponibles para m y s, la distribución óptima para m es $a_{mid} = (a - (N - 1) \log_2 \tan \theta_s)/2$.
- [0006]** En las estrategias conocidas (por ejemplo en [2] y [4]), se combinan complicados bucles de tasa/distorsión con la decisión de qué canales de bandas se van a transformar (por ejemplo, usando M/S, a lo que
35 también puede seguir el cálculo residual de predicción de M a S de [7]), con el fin de reducir la correlación entre canales. Esta complicada estructura tiene un elevado coste informático.
- [0007]** La separación del modelo perceptual del bucle de tasa (como ocurre en [6a], [6b] y [13]) simplifica significativamente el sistema.
- 40 **[0008]** Además, la codificación de los coeficientes de predicción o los ángulos de cada banda requiere un considerable número de bits (como por ejemplo en [5] y [7]).
- [0009]** En [1], [3] y [5] solo se lleva a cabo una decisión todo a lo ancho del espectro para decidir si todo el
45 espectro debe ser codificado M/S (del inglés Mid/Side, Medio/Lateral) o L/R (del inglés, Left/Right, Izquierda/Derecha).
- [0010]** La codificación M/S no es eficiente si existe una ILD (diferencia de nivel interaural), es decir, si los canales están paneados.
- 50 **[0011]** Como se señalaba anteriormente, se sabe que el procesamiento M/S banda por banda en los codificadores basados en MDCT (por sus siglas en inglés, Modified Discrete Cosine Transform, Transformada de Coseno Discreta Modificada), es un procedimiento eficaz para el procesamiento estéreo. La ganancia de codificación por el procesamiento M/S varía del 0 % en el caso de los canales no correlacionados al 50 % en el caso de los monofónicos o de una diferencia de fase $\pi/2$ entre los canales. Debido al desenmascaramiento estéreo y al
55 desenmascaramiento inverso (véase [1]), es importante contar con una firme decisión M/S.
- [0012]** En [2], cada banda, en que los umbrales de enmascaramiento entre izquierdo y derecho varían en menos de 2dB, se opta por la codificación M/S como procedimiento de codificación.

- [0013]** En [1], la decisión M/S se basa en el consumo estimado de bits para la codificación M/S y para la codificación L/R (L/R = izquierda/derecha) de los canales. La demanda de tasa de bits para la codificación M/S y para la codificación L/R se estima según los espectros y los umbrales de enmascaramiento utilizando entropía perceptual (PE, por sus siglas en inglés). Se calculan los umbrales de enmascaramiento correspondientes al canal izquierdo y al derecho. Se presume que los umbrales de enmascaramiento correspondientes al canal medio y al canal lateral son el mínimo de los umbrales izquierdo y derecho.
- [0014]** Además, [1] describe cómo se derivan los umbrales de codificación de los canales individuales que se van a codificar. Específicamente, los umbrales de codificación correspondientes a los canales izquierdo y derecho se calculan mediante los respectivos modelos perceptuales para estos canales. En [1], los umbrales de codificación correspondientes al canal M y al canal S se eligen por igual y se derivan en términos de mínimo de los umbrales de codificación derecho e izquierdo.
- [0015]** Asimismo, [1] describe la decisión entre la codificación L/R y la codificación M/S de tal manera que se obtenga un buen rendimiento de codificación. Específicamente, se estima una entropía perceptual para la codificación L/R y la codificación M/S utilizando los umbrales.
- [0016]** En [1] y [2], así como en [3] y [4], el procesamiento M/S se lleva a cabo en una señal formada en ventana y transformada no normalizada (no blanqueada) y la decisión M/S se basa en el umbral de enmascaramiento y la estimación de entropía perceptual.
- [0017]** En [5], se codifica explícitamente la energía del canal izquierdo y del canal derecho y el ángulo codificado conserva la energía de la señal de diferencia. Se supone, en [5], que la codificación M/S es segura, incluso si la codificación L/R es más eficiente. Según [5], solo se elige la codificación L/R cuando la correlación entre los canales no es lo suficientemente fuerte.
- [0018]** Además, la codificación de los coeficientes de predicción o los ángulos en cada banda requiere un número significativo de bits (véase, por ejemplo, [5] y [7]).
- [0019]** El documento WO 2008/065487 A1 describe una técnica para realizar la codificación estéreo Media-Lateral (M/S), en la que se añade una etapa adicional al proceso de codificación, mediante el cual un parámetro que se utiliza para determinar cuándo se usarán las señales media y lateral en lugar de las señales de entrada izquierda y derecha se modifica antes de realizar la selección entre los pares de señales.
- [0020]** El documento US 2012/275604 A1 describe el procesamiento de una señal de audio estereofónica de entrada para generar de ese modo una señal de audio estereofónica convertida que representa la señal de audio estereofónica de entrada.
- [0021]** LINDBLOM J Y COL., "Codificación estéreo de suma-diferencia flexible basada en componentes de señal alineados en el tiempo", APLICACIONES DEL PROCESAMIENTO DE SEÑALES AL AUDIO Y LA ACÚSTICA, 2005. IEEE WORKSHOP ON NEW PALTZ, NY, EE. UU., 16 AL 19 DE OCTUBRE DE 2005, PISCATAWAY, NJ, EE. UU., IEEE, (20051016), doi:10.1109/ASPAA.2005.1540218, ISBN 978-0-7803-9154-3, páginas 255-258, describe una codificación flexible de señales de audio estéreo generales basada en la codificación de suma-diferencia.
- [0022]** El documento WO 2011/124608 A1 describe la codificación y decodificación estéreo utilizando predicción compleja en el dominio de la frecuencia.
- [0023]** HELMRICH CHRISTIAN R Y COL., "Codificación de transformación de audio estéreo conjunta semiparamétrica de baja complejidad", 2015 23ª CONFERENCIA EUROPEA DE PROCESAMIENTO DE SEÑALES (EUSIPCO), EURASIP, (20150831), doi:10.1109/EUSIP10 CO.2015.7362492, páginas 794 - 798 describe un concepto de codificación estéreo, en el que se deriva la información lateral paramétrica en el codificador.
- [0024]** El documento WO 2017/106041 A1 describe un combinador que está configurado para generar un marco en un codificador multicanal. El marco incluye un subconjunto de muestras de los datos de la primera parte de anticipación, una o más muestras de datos de muestra actualizados correspondientes al primer marco combinado, y un grupo de muestras de datos del segundo marco combinado correspondientes al segundo marco combinado.

[0025] El documento WO 2017/087073A1 describe la determinación de un valor de desajuste indicativo de una cantidad de desajuste temporal entre los dos canales de audio.

[0026] Por lo tanto, se apreciaría mucho que se dieran a conocer conceptos mejorados para la codificación de audio y la decodificación de audio.

[0027] El objeto de la presente invención consiste en proporcionar conceptos mejorados para la codificación de señales de audio, el procesamiento de señales de audio y la decodificación de señales de audio. El objeto de la presente invención se soluciona mediante las reivindicaciones independientes. Realizaciones adicionales se proporcionan en las reivindicaciones dependientes.

[0028] Según algunas realizaciones, se utiliza el FDNS (FDNS = Frequency Domain Noise Shaping, Modelado de Ruido en el Dominio de la Frecuencia) con el bucle de tasa de acuerdo con lo descrito en [6a] y [6b] combinado con la distorsión de la envolvente espectral de acuerdo con lo descrito en [8]. En algunas realizaciones, se utiliza un único parámetro de ILD en el espectro blanqueado con FDNS seguido por la decisión banda por banda, ya sea que se utilice la codificación M/S o la codificación L/R para la codificación. En algunas realizaciones, la decisión M/S se basa en el ahorro de bits estimado. En algunas realizaciones, la distribución de la tasa de bits entre los canales procesados M/S banda por banda puede depender, por ejemplo, de la energía.

[0029] Algunas realizaciones presentan una combinación de ILD (por sus siglas en inglés, interaural level difference, diferencia de nivel interaural) global única aplicada al espectro blanqueado, seguida por el procesamiento M/S banda por banda con un mecanismo eficiente de decisión M/S y con un bucle de tasa que controla la única ganancia global.

[0030] Algunas realizaciones emplean, entre otras, el FDNS con el bucle de tasa basándose, por ejemplo, en [6a] o [6b], combinados con la distorsión de la envolvente espectral basándose, por ejemplo en [8]. Estas realizaciones ofrecen una manera eficiente y muy eficaz de separar el modelado perceptual del ruido de cuantificación y el bucle de tasa. El uso del parámetro de ILD única en el espectro blanqueado por FDNS da lugar a una manera sencilla y eficaz de decidir si existe una ventaja en el procesamiento M/S de acuerdo con lo descrito anteriormente. El blanqueo del espectro y la eliminación de la ILD permite el procesamiento eficiente M/S. La codificación de la ILD global única en el caso del sistema descrito es suficiente y, por consiguiente, se obtiene un ahorro de bits a diferencia de las estrategias conocidas.

[0031] Según las realizaciones, el procesamiento M/S se realiza sobre la base de una señal perceptualmente blanqueada. Las realizaciones determinan umbrales de codificación y determinan, de manera óptima, una decisión de si se va a emplear la codificación L/R o una codificación M/S al procesar señales blanqueadas perceptualmente y con ILD compensada.

[0032] Además, según las realizaciones, se presenta una nueva estimación de la tasa de bits.

[0033] A diferencia de [1]-[5], en estas realizaciones, se separa el modelo perceptual del bucle de tasa como en [6a], [6b] y [13].

[0034] Pese a que la decisión M/S se basa en la tasa de bits estimada como se propone en [1], a diferencia de [1] la diferencia en la demanda de tasa de bits de la codificación M/S y la L/R no depende de los umbrales de enmascaramiento determinados por un modelo perceptual. Por el contrario, la demanda de tasa de bits se determina por medio de un codificador de entropía sin pérdidas que se utiliza. En otras palabras, en lugar de derivar la demanda de tasa de bits de la entropía perceptual de la señal original, se deriva la demanda de tasa de bits de la entropía de la señal perceptualmente blanqueada.

[0035] En oposición a [1]-[5], en las realizaciones, la decisión M/S se determina sobre la base de una señal perceptualmente blanqueada, y se obtiene una mejor estimación de la tasa de bits necesaria. Para este fin, se puede aplicar la estimación de consumo de bits del codificador aritmético descrita en [6a] o [6b]. No es necesario considerar explícitamente los umbrales de enmascaramiento.

[0036] En [1], se asume que los umbrales de enmascaramiento para los canales medio y laterales son el de los umbrales de enmascaramiento izquierdo y derecho. El modelado de ruido espectral se realiza en el canal medio y

el lateral y se puede basar, por ejemplo, en estos umbrales de enmascaramiento.

5 **[0037]** Según las realizaciones, el modelado de ruido espectral se puede llevar a cabo, por ejemplo, en el canal izquierdo y el derecho y, en esas realizaciones, la envolvente perceptual se puede aplicar exactamente cuando se la ha estimado.

[0038] Asimismo, las realizaciones se basan en el hallazgo de que la codificación M/S no es eficiente si existe ILD, es decir si los canales están paneados. Para evitar esto, las realizaciones utilizan un parámetro de ILD única en el espectro perceptualmente blanqueado.

10 **[0039]** Según algunas realizaciones, se dan a conocer nuevos conceptos para la decisión M/S que procesan una señal perceptualmente blanqueada.

[0040] Según algunas realizaciones, el códec utiliza nuevos conceptos que no son parte de los códecs de audio clásicos, por ejemplo, de acuerdo con lo descrito en [1].

[0041] Según algunas realizaciones, se utilizan señales perceptualmente blanqueadas para la codificación adicional, por ejemplo, de manera similar a su uso en un codificador de voz.

20 **[0042]** Esa estrategia tiene varias ventajas, por ejemplo, se simplifica la arquitectura del códec, se obtiene una representación compacta de las características de modelado de ruido y el umbral de enmascaramiento, por ejemplo, en forma de coeficientes de LPC (por sus siglas en inglés, *Linear Prediction Coding*, Codificación de Predicción Lineal). Además, se unifican las arquitecturas de códecs de transformación y voz y de esa manera se habilita una codificación combinada de audio y voz.

25 **[0043]** Algunas realizaciones emplean un parámetro de ILD global para codificar con eficiencia las fuentes paneadas.

[0044] En las realizaciones, el códec emplea el Modelado de Ruido en el Dominio de la Frecuencia (FDNS, por sus siglas en inglés *Frequency Domain Noise Shaping*) para blanquear perceptualmente la señal con el bucle de tasa, por ejemplo, como se describe en [6a] o [6b] combinado con la distorsión de la envolvente espectral de acuerdo con lo descrito en [8]. En esas realizaciones, el códec puede usar asimismo, por ejemplo, un parámetro de ILD único en el espectro blanqueado por FDNS, seguido por la decisión M/S o L/R banda por banda. La decisión M/S por banda se puede basar, por ejemplo, en la tasa de bits estimada en cada banda cuando se codifica en el modo L/R y en el M/S. Se elige el procedimiento que demande menos bits. La distribución de la tasa de bits entre los canales procesados en el modo M/S banda por banda se basa en la energía.

[0045] Algunas realizaciones aplican una decisión M/S banda por banda al espectro perceptualmente blanqueado y con ILD compensada utilizando el número estimado de bits por banda para un codificador entrópico.

40 **[0046]** En algunas realizaciones, se emplea el FDNS (por sus siglas en inglés, *Frequency Domain Noise Shaping*, Modelado de Ruido en el Dominio de la Frecuencia) con el bucle de tasa, por ejemplo, de acuerdo con lo descrito en [6a] o [6b] combinado con la distorsión de la envolvente espectral según lo descrito en [8]. Esto otorga una manera eficiente y muy eficaz de modelar el ruido de cuantificación y el bucle de tasa. El uso del parámetro de ILD única en el espectro blanqueado por FDNS ofrece una manera eficiente y eficaz de decidir si existe una ventaja en el procesamiento M/S descrito. El blanqueo del espectro y la eliminación de la ILD permite un procesamiento M/S eficiente.

[0047] La codificación de la ILD global única para el sistema descrito es suficiente y, de esa manera, se obtiene un ahorro de bits a diferencia de las estrategias conocidas.

[0048] Las realizaciones modifican los conceptos presentados en [1] en el procesamiento de señales perceptualmente blanqueadas y con ILD compensada. En particular, las realizaciones emplean una ganancia global igual para L, R, M y S (del inglés, *Left, Right, Mid y Side* (Izquierda, Derecha, Medio y Lateral) que, junto con el FDNS, forma los umbrales de codificación. La ganancia global se puede derivar de una estimación de SNR (por sus siglas en inglés *Signal–Noise Ratio*, Relación Señal a Ruido o de algún otro concepto).

[0049] La decisión M/S banda por banda propuesta estima con precisión el número de bits requeridos para

codificar cada banda con el codificador aritmético. Esto es posible porque la decisión M/S se realiza en el espectro blanqueado y seguido directamente por la cuantificación. No hay necesidad de búsqueda experimental de umbrales.

5 **[0050]** A continuación, se describen las realizaciones útiles para comprender la presente invención con más detalle con referencia a las figuras, en las cuales:

- Fig. 1a ilustra un aparato para codificación según una realización,
- Fig. 1b ilustra un aparato para codificación según otra realización, en la cual el aparato comprende además una unidad de transformación y una unidad de pre-procesamiento,
- 10 Fig. 1c ilustra un aparato para codificación según otra realización, en la cual el aparato comprende además una unidad de transformación,
- Fig. 1d ilustra un aparato para codificación según otra realización, en la cual el aparato comprende una unidad de pre-procesamiento y una unidad de transformación,
- Fig. 1e ilustra un aparato para codificación según otra realización, en la cual el aparato comprende asimismo un pre-
- 15 procesador en el dominio espectral,
- Fig. 1f ilustra un sistema para la codificación de cuatro canales de una señal de entrada de audio que comprende cuatro o más canales para obtener cuatro canales de una señal de audio codificada según una realización,
- Fig. 2a ilustra un aparato para decodificación según una realización,
- Fig. 2b ilustra un aparato para decodificación según una realización que comprende además una unidad de
- 20 transformación y una unidad de post-procesamiento,
- Fig. 2c ilustra un aparato para decodificación según con una realización, en la cual el aparato para decodificación comprende asimismo una unidad de transformación,
- Fig. 2d ilustra un aparato para decodificación según una realización, en la cual el aparato para decodificación comprende asimismo una unidad de post-procesamiento,
- 25 Fig. 2e ilustra un aparato para decodificación según una realización, en la cual el aparato comprende asimismo un post-procesador en el dominio espectral,
- Fig. 2f ilustra un sistema para la decodificación de una señal de audio codificada que comprende cuatro o más canales para obtener cuatro canales de una señal de audio decodificada que comprende cuatro o más canales según una realización,
- 30 Fig. 3 ilustra un sistema según una realización,
- Fig. 4 ilustra un aparato para codificación según otra realización,
- Fig. 5 ilustra módulos de procesamiento estéreo en un aparato para codificación según una realización,
- Fig. 6 ilustra un aparato para decodificación según otra realización,
- Fig. 7 ilustra el cálculo de una tasa de bits para la decisión M/S por bandas según una realización,
- 35 Fig. 8 ilustra una decisión del modo estéreo según una realización,
- Fig. 9 ilustra el procesamiento estéreo del lado de un codificador según las realizaciones, que emplean relleno estéreo,
- Fig. 10 ilustra el procesamiento estéreo del lado de un decodificador según las realizaciones, que emplean relleno estéreo,
- Fig. 11 ilustra relleno estéreo de una señal lateral del lado de un decodificador según algunas realizaciones específicas,
- 40 Fig. 12 ilustra el procesamiento estéreo del lado de un codificador según las realizaciones, que no emplean relleno estéreo, y
- Fig. 13 ilustra el procesamiento estéreo del lado de un decodificador según las realizaciones, que no emplean relleno estéreo.

45 **[0051]** La Fig. 1a ilustra un aparato para la codificación de un primer canal y un segundo canal de una señal de entrada de audio que comprende dos o más canales para obtener una señal de audio codificada según una realización.

[0052] El aparato comprende un normalizador 110 configurado para determinar un valor de normalización correspondiente a la señal de entrada de audio dependiendo del primer canal de la señal de entrada de audio y dependiendo del segundo canal de la señal de entrada de audio. El normalizador 110 está configurado para determinar un primer canal y un segundo canal de una señal de audio normalizada mediante la modificación, dependiendo del valor de normalización, de al menos uno del primer canal y el segundo canal de la señal de entrada de audio.

55 **[0053]** Por ejemplo, el normalizador 110 puede estar configurado, en una realización, por ejemplo, para determinar el valor de normalización correspondiente a la señal de entrada de audio; dependiendo de una pluralidad de bandas espectrales el primer canal y del segundo canal de la señal de entrada de audio, el normalizador 110 puede

estar configurado, por ejemplo, para determinar el primer canal y el segundo canal de la señal de audio normalizada mediante la modificación, dependiendo del valor de normalización, de la pluralidad de bandas espectrales de al menos uno del primer canal y el segundo canal de la señal de entrada de audio.

5 **[0054]** El aparato comprende una unidad de codificación 120 que está configurada para (cuando se elige un modo de codificación por banda) generar una señal de audio procesada que tiene un primer canal y un segundo canal, de tal manera que una o más bandas espectrales del primer canal de la señal de audio procesada sean una o más bandas espectrales del primer canal de la señal de audio normalizada, de tal manera que una o más bandas espectrales del segundo canal de la señal de audio procesada sean una o más bandas espectrales del segundo canal de la señal de audio normalizada, de tal manera que al menos una banda espectral del primer canal de la señal de audio procesada sea una banda espectral de una señal media dependiendo de una banda espectral del primer canal de la señal de audio normalizada y dependiendo de una banda espectral del segundo canal de la señal de audio normalizada, y de tal manera que al menos una banda espectral del segundo canal de la señal de audio procesada sea una banda espectral de una señal lateral dependiendo de una banda espectral del primer canal de la señal de audio normalizada y dependiendo de una banda espectral del segundo canal de la señal de audio normalizada. La unidad de codificación 120 está configurada para codificar la señal de audio procesada para obtener la señal de audio codificada.

20 **[0055]** La unidad de codificación 120 está configurada para elegir entre un modo de codificación medio-lateral total y un modo de codificación doble-mono total y un modo de codificación por banda dependiendo de una pluralidad de bandas espectrales de un primer canal de la señal de audio normalizada y dependiendo de una pluralidad de bandas espectrales de un segundo canal de la señal de audio normalizada.

25 **[0056]** La unidad de codificación 120 está configurada, si se elige el modo de codificación medio-lateral total, para generar una señal media desde el primer canal y desde el segundo canal de la señal de audio normalizada como un primer canal de una señal medio-lateral, para generar una señal lateral desde el primer canal y desde el segundo canal de la señal de audio normalizada como un segundo canal de la señal medio-lateral, y para codificar la señal medio-lateral con el fin de obtener la señal de audio codificada.

30 **[0057]** La unidad de codificación está configurada, en caso de elegirse el modo de codificación doble-mono total, para codificar la señal de audio normalizada con el fin de obtener la señal de audio codificada.

35 **[0058]** Además, la unidad de codificación 120 está configurada, si se elige el modo de codificación por banda, para generar la señal de audio procesada, de tal manera que una o más bandas espectrales del primer canal de la señal de audio procesada sean una o más bandas espectrales del primer canal de la señal de audio normalizada, de tal manera que una o más bandas espectrales del segundo canal de la señal de audio procesada sean una o más bandas espectrales del segundo canal de la señal de audio normalizada, de tal manera que al menos una banda espectral del primer canal de la señal de audio procesada sea una banda espectral de una señal media dependiendo de una banda espectral del primer canal de la señal de audio normalizada y dependiendo de una banda espectral del segundo canal de la señal de audio normalizada, y de tal manera que al menos una banda espectral del segundo canal de la señal de audio procesada sea una banda espectral de una señal lateral dependiendo de una banda espectral del primer canal de la señal de audio normalizada y dependiendo de una banda espectral del segundo canal de la señal de audio normalizada, donde la unidad de codificación 120 está configurada para codificar la señal de audio procesada para obtener la señal de audio codificada.

45 **[0059]** Según una realización, la señal de entrada de audio puede ser, por ejemplo, una señal estéreo de audio que comprende exactamente dos canales. Por ejemplo, el primer canal de la señal de entrada de audio puede ser, por ejemplo, un canal izquierdo de la señal estéreo de audio, y el segundo canal de la señal de entrada de audio puede ser, por ejemplo, un canal derecho de la señal estéreo de audio.

50 **[0060]** En una realización, la unidad de codificación 120 está configurada, por ejemplo, en caso de elegirse el modo de codificación por banda, para decidir, por cada banda espectral de una pluralidad de bandas espectrales de la señal de audio procesada, si se emplea la codificación medio-lateral o si se emplea la codificación doble-mono.

55 **[0061]** Si se emplea la codificación medio-lateral para dicha banda espectral, la unidad de codificación 120 puede estar configurada, por ejemplo, para generar dicha banda espectral del primer canal de la señal de audio procesada como banda espectral de una señal media sobre la base de dicha banda espectral del primer canal de la

señal de audio normalizada y sobre la base de dicha banda espectral del segundo canal de la señal de audio normalizada. La unidad de codificación 120 puede estar configurada, por ejemplo, para generar dicha banda espectral del segundo canal de la señal de audio procesada como banda espectral de una señal lateral sobre la base de dicha banda espectral del primer canal de la señal de audio normalizada y sobre la base de dicha banda espectral del
5 segundo canal de la señal de audio normalizada.

[0062] Si se emplea la codificación doble-mono para dicha banda espectral, la unidad de codificación 120 puede estar configurada, por ejemplo, para usar dicha banda espectral del primer canal de la señal de audio normalizada como banda espectral del primer canal de la señal de audio procesada, y puede estar configurada, por
10 ejemplo, para usar dicha banda espectral del segundo canal de la señal de audio normalizada como banda espectral del segundo canal de la señal de audio procesada.

[0063] bien la unidad de codificación 120 está configurada para usar dicha banda espectral del segundo canal de la señal de audio normalizada como banda espectral del primer canal de la señal de audio procesada, y puede
15 estar configurada, por ejemplo, para usar dicha banda espectral del primer canal de la señal de audio normalizada como banda espectral del segundo canal de la señal de audio procesada.

[0064] Según una realización, la unidad de codificación 120 puede estar configurada, por ejemplo, para elegir entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación
20 por banda mediante la determinación de una primera estimación que calcula un primer número de bits que se necesitan para la codificación cuando se emplea el modo de codificación medio-lateral total, mediante la determinación de una segunda estimación que calcula un segundo número de bits que se necesitan para la codificación cuando se emplea el modo de codificación doble-mono total, mediante la determinación de una tercera estimación que calcula un tercer
25 número de bits que se necesitan para la codificación cuando se emplea el modo de codificación por banda que se puede emplear, por ejemplo, y mediante la elección del modo de codificación entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda que tenga el menor número de bits entre la primera estimación y la segunda estimación y la tercera estimación.

[0065] En una realización, la unidad de codificación 120 puede estar configurada, por ejemplo, para estimar
30 la tercera estimación b_{BW} , que calcula el tercer número de bits que se necesitan para codificar cuando se emplea el modo de codificación por banda, según la fórmula:

$$b_{BW} = nBandas + \sum_{i=0}^{nBandas-1} \min(b_{bwLR}^i, b_{bwMS}^i) ,$$

35 donde $nBandas$ es un número de bandas espectrales de la señal de audio normalizada, donde b_{bwMS}^i es una estimación del número de bits que se necesitan para codificar una i^{th} banda espectral de la señal media y para codificar la i^{th} banda espectral de la señal lateral, y donde b_{bwLR}^i es una estimación del número de bits que se necesitan para codificar una i^{th} banda espectral de la primera señal y para codificar la i^{th} banda espectral de la segunda señal.

40 **[0066]** En las realizaciones, se puede emplear, por ejemplo, una medida objetiva de la calidad para elegir entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda.

[0067] Según una realización, la unidad de codificación 120 puede estar configurada, por ejemplo, para elegir
45 entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda mediante la determinación de una primera estimación que calcula un primer número de bits que se ahorran cuando se codifica en el modo de codificación medio-lateral total, mediante la determinación de una segunda estimación que calcula un segundo número de bits que se ahorran cuando se codifica en el modo de codificación doble-mono total, mediante la determinación de una tercera estimación que calcula un tercer número de bits que se
50 ahorran cuando se codifica en el modo de codificación por banda, y mediante la elección de ese modo de codificación entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda en el cual se ahorra el mayor número de bits entre la primera estimación y la segunda estimación y la tercera estimación.

[0068] En otra realización, la unidad de codificación 120 puede estar configurada, por ejemplo, para elegir entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda mediante la estimación de una primera relación señal a ruido que se produce cuando se emplea el modo de codificación medio-lateral total, mediante la estimación de una segunda relación señal a ruido que se produce cuando se emplea el modo de codificación doble-mono total, mediante la estimación de una tercera relación señal a ruido que se produce cuando se emplea el modo de codificación por banda, y mediante la elección de ese modo de codificación entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda en el cual se ahorra el mayor número de bits entre la primera relación señal a ruido y la segunda relación señal a ruido y la tercera relación señal a ruido.

[0069] La señal de entrada de audio está representada en un dominio espectral. El normalizador 110 está configurado para determinar el valor de normalización correspondiente a la señal de entrada de audio dependiendo de una pluralidad de bandas espectrales del primer canal de la señal de entrada de audio y dependiendo de una pluralidad de bandas espectrales del segundo canal de la señal de entrada de audio. Además, el normalizador 110 puede estar configurado para determinar la señal de audio normalizada mediante la modificación, dependiendo del valor de normalización, de la pluralidad de bandas espectrales de al menos uno del primer canal y el segundo canal de la señal de entrada de audio.

[0070] El normalizador 110 está configurado para determinar el valor de normalización basándose en las fórmulas:

$$NRG_L = \sqrt{\sum MDCT_{L,k}^2}$$

$$NRG_R = \sqrt{\sum MDCT_{R,k}^2}$$

$$ILD = \frac{NRG_L}{NRG_L + NRG_R}$$

en las cuales $MDCT_{L,k}$ es un k° coeficiente de un espectro de MDCT del primer canal de la señal de entrada de audio, y $MDCT_{R,k}$ es el k° coeficiente del espectro de MDCT del segundo canal de la señal de entrada de audio. El normalizador 110 puede estar configurado, por ejemplo, para determinar el valor de normalización mediante la cuantificación de la ILD .

[0071] Según una realización ilustrada en la Fig. 1b, el aparato para la codificación puede comprender asimismo, por ejemplo, una unidad de transformación 102 y una unidad de pre-procesamiento 105. La unidad de transformación 102 puede estar configurada, por ejemplo, para transformar una señal de audio en el dominio del tiempo de un dominio del tiempo a un dominio de la frecuencia para obtener una señal de audio transformada. La unidad de pre-procesamiento 105 puede estar configurada, por ejemplo, para generar el primer canal y el segundo canal de la señal de entrada de audio mediante la aplicación de una operación de modelado de ruido en el dominio de la frecuencia del lado del codificador a la señal de audio transformada.

[0072] En una realización específica, la unidad de pre-procesamiento 105 puede estar configurada, por ejemplo, para generar el primer canal y el segundo canal de la señal de entrada de audio mediante la aplicación de una operación de modelado de ruido temporal del lado del codificador en la señal de audio transformada antes de aplicar la operación de modelado de ruido en el dominio de la frecuencia del lado del codificador a la señal de audio transformada.

[0073] La Fig. 1c ilustra un aparato para codificación según otra realización que comprende además una unidad de transformación 115. El normalizador 110 puede estar configurado, por ejemplo, para determinar un valor de normalización correspondiente a la señal de entrada de audio que depende del primer canal de la señal de entrada de audio que está representado en el dominio del tiempo y que depende del segundo canal de la señal de entrada de audio que está representado en el dominio del tiempo. Además, el normalizador 110 puede estar configurado, por ejemplo, para determinar el primer canal y el segundo canal de la señal de audio normalizada mediante la modificación, dependiendo del valor de normalización, de al menos uno del primer canal y el segundo canal de la señal de entrada de audio que está representado en el dominio del tiempo. La unidad de transformación 115 puede estar configurada,

por ejemplo, para transformar la señal de audio normalizada del dominio del tiempo al dominio espectral para que la señal de audio normalizada esté representada en el dominio espectral. Además, la unidad de transformación 115 puede estar configurada, por ej., para alimentar la señal de audio normalizada que está representada en el dominio espectral a la unidad de codificación 120.

5

[0074] La Fig. 1d ilustra un aparato para codificación según otra realización, en la cual el aparato comprende además una unidad de pre-procesamiento 106 que está configurada para recibir una señal de audio en el dominio del tiempo que comprende un primer canal y un segundo canal. La unidad de pre-procesamiento 106 puede estar configurada, por ejemplo, para aplicar un filtro al primer canal de la señal de audio en el dominio del tiempo que produce un primer espectro perceptualmente blanqueado para obtener el primer canal de la señal de entrada de audio que está representado en el dominio del tiempo. Además, la unidad de pre-procesamiento 106 puede estar configurada, por ejemplo, para aplicar el filtro al segundo canal de la señal de audio en el dominio del tiempo que produce un segundo espectro perceptualmente blanqueado para obtener el segundo canal de la señal de entrada de audio que está representado en el dominio del tiempo.

15

[0075] En una realización, ilustrada en la Fig. 1e, la unidad de transformación 115 puede estar configurada, por ejemplo, para transformar la señal de audio normalizada del dominio del tiempo al dominio espectral para obtener una señal de audio transformada. En la realización de la Fig. 1e, el aparato comprende asimismo un pre-procesador en el dominio espectral 118 que está configurado para llevar a cabo el modelado de ruido temporal del lado del codificador en la señal de audio transformada para obtener la señal de audio normalizada que está representada en el dominio espectral.

20

[0076] Según una realización, la unidad de codificación 120 puede estar configurada, por ejemplo, para obtener la señal de audio codificada mediante la aplicación de relleno de huecos estéreo inteligente del lado del codificador en la señal de audio normalizada o en la señal de audio procesada.

25

[0077] En otra realización, ilustrada en la Fig. 1f, se da a conocer un sistema para codificar cuatro canales de una señal de entrada de audio que comprende cuatro o más canales para obtener una señal de audio codificada. El sistema comprende un primer aparato 170 según una de las realizaciones antes descritas para codificar un primer canal y un segundo canal de los cuatro o más canales de la señal de entrada de audio para obtener un primer canal y un segundo canal de la señal de audio codificada. Además, el sistema comprende un segundo aparato 180 según una de las realizaciones antes descritas para codificar un tercer canal y un cuarto canal de los cuatro o más canales de la señal de entrada de audio para obtener un tercer canal y un cuarto canal de la señal de audio codificada.

30

[0078] La Fig. 2a ilustra un aparato para decodificar una señal de audio codificada que comprende un primer canal y un segundo canal para obtener una señal de audio decodificada según una realización no reivindicada.

35

[0079] El aparato para decodificación comprende una unidad de decodificación 210 configurada para determinar, por cada banda espectral de una pluralidad de bandas espectrales, si dicha banda espectral del primer canal de la señal de audio codificada y dicha banda espectral del segundo canal de la señal de audio codificada se codificó empleando codificación doble-mono o utilizando la codificación medio-lateral.

40

[0080] La unidad de decodificación 210 está configurada para usar dicha banda espectral del primer canal de la señal de audio codificada como banda espectral de un primer canal de una señal de audio intermedia y está configurada para usar dicha banda espectral del segundo canal de la señal de audio codificada como banda espectral de un segundo canal de la señal de audio intermedia, en caso de haberse utilizado la codificación doble-mono.

45

[0081] Además, la unidad de decodificación 210 está configurada para generar una banda espectral del primer canal de la señal de audio intermedia sobre la base de dicha banda espectral del primer canal de la señal de audio codificada y sobre la base de dicha banda espectral del segundo canal de la señal de audio codificada, y para generar una banda espectral del segundo canal de la señal de audio intermedia sobre la base de dicha banda espectral del primer canal de la señal de audio codificada y sobre la base de dicha banda espectral del segundo canal de la señal de audio codificada, en caso de haberse utilizado la codificación medio-lateral.

50

[0082] Por otro lado, el aparato para decodificación comprende un desnormalizador 220 configurado para modificar, dependiendo de un valor de desnormalización, de al menos uno del primer canal y el segundo canal de la señal de audio intermedia para obtener el primer canal y el segundo canal de la señal de audio decodificada.

55

[0083] En una realización, la unidad de decodificación 210 puede estar configurada, por ejemplo, para determinar si la señal de audio codificada se va a codificar en un modo de codificación medio-lateral total o en un modo de codificación doble-mono total o en un modo de codificación por banda.

5

[0084] Además, en ese tipo de realización, la unidad de decodificación 210 puede estar configurada, por ejemplo, si se determina que la señal de audio codificada se codifica en el modo de codificación medio-lateral total, para generar el primer canal de la señal de audio intermedia a partir del primer canal y del segundo canal de la señal de audio codificada, y para generar el segundo canal de la señal de audio intermedia a partir del primer canal y del

10 segundo canal de la señal de audio codificada.

[0085] Según ese tipo de realización, la unidad de decodificación 210 puede estar configurada, por ejemplo, si se determina que la señal de audio codificada se va a codificar en el modo de codificación doble-mono total, para usar el primer canal de la señal de audio codificada como primer canal de la señal de audio intermedia, y para usar el

15 segundo canal de la señal de audio codificada como segundo canal de la señal de audio intermedia.

[0086] Por otro lado, en ese tipo de realización, la unidad de decodificación 210 puede estar configurada, por ejemplo, si se determina que se debe codificar la señal de audio codificada en el modo de codificación por banda,

20 - para determinar, por cada banda espectral de una pluralidad de bandas espectrales, si dicha banda espectral del primer canal de la señal de audio codificada y dicha banda espectral del segundo canal de la señal de audio codificada se codificó utilizando la codificación doble-mono o utilizando la codificación medio-lateral,

- para usar dicha banda espectral del primer canal de la señal de audio codificada como banda espectral del primer canal de la señal de audio intermedia y para usar dicha banda espectral del segundo canal de la señal de audio

25 codificada como banda espectral del segundo canal de la señal de audio intermedia, en caso de haberse utilizado la codificación doble-mono, y

- para generar una banda espectral del primer canal de la señal de audio intermedia sobre la base de dicha banda espectral del primer canal de la señal de audio codificada y sobre la base de dicha banda espectral del segundo canal de la señal de audio codificada, y para generar una banda espectral del segundo canal de la señal de audio intermedia

30 sobre la base de dicha banda espectral del primer canal de la señal de audio codificada y sobre la base de dicha banda espectral del segundo canal de la señal de audio codificada, en caso de haberse utilizado la codificación medio-lateral.

[0087] Por ejemplo, en el modo de codificación medio-lateral total, se pueden aplicar, las fórmulas:

35

$$L = (M + S) / \sqrt{2},$$

y

40

$$R = (M - S) / \sqrt{2}$$

para obtener, por ejemplo, el primer canal L de la señal de audio intermedia y para obtener el segundo canal R de la señal de audio intermedia, donde M es el primer canal de la señal de audio codificada y donde S es el segundo canal de la señal de audio codificada.

45

[0088] Según una realización, la señal de audio decodificada puede ser, por ejemplo, una señal estéreo de audio que comprende exactamente dos canales. Por ejemplo, el primer canal de la señal de audio decodificada puede ser, por ejemplo, un canal izquierdo de la señal estéreo de audio, y el segundo canal de la señal de audio decodificada puede ser, por ejemplo, un canal derecho de la señal estéreo de audio.

50

[0089] Según una realización, el des-normalizador 220 puede estar configurado, por ejemplo, para modificar, dependiendo del valor de des-normalización, la pluralidad de bandas espectrales de al menos uno del primer canal y el segundo canal de la señal de audio intermedia para obtener el primer canal y el segundo canal de la señal de audio decodificada.

55

[0090] En otra realización expuesta en la Fig. 2b, el des-normalizador 220 puede estar configurado, por ejemplo, para modificar, dependiendo del valor de des-normalización, la pluralidad de bandas espectrales de al menos uno del primer canal y el segundo canal de la señal de audio intermedia para obtener una señal de audio des-

normalizada. En ese tipo de realización, el aparato puede comprender además, por ejemplo, una unidad de post-procesamiento 230 y una unidad de transformación 235. La unidad de post-procesamiento 230 puede estar configurada, por ejemplo, para ejecutar al menos uno de modelado de ruido temporal del lado del decodificador y modelado de ruido en el dominio de la frecuencia del lado del decodificador en la señal de audio des-normalizada para
5 obtener una señal de audio post-procesada. La unidad de transformación (235) puede estar configurada, por ejemplo, para transformar la señal de audio post-procesada de un dominio espectral a un dominio del tiempo para obtener el primer canal y el segundo canal de la señal de audio decodificada.

[0091] Según una realización ilustrada en la Fig. 2c, el aparato comprende además una unidad de transformación 215 configurada para transformar la señal de audio intermedia de un dominio espectral a un dominio
10 del tiempo. El des-normalizador 220 puede estar configurado, por ejemplo, para modificar, dependiendo del valor de des-normalización, al menos uno del primer canal y el segundo canal de la señal de audio intermedia que está representado en el dominio del tiempo para obtener el primer canal y el segundo canal de la señal de audio decodificada.

[0092] En una realización similar que no se reivindica, ilustrada en la Fig. 2d, la unidad de transformación 215 puede estar configurada, por ejemplo, para transformar la señal de audio intermedia de un dominio espectral a un dominio del tiempo. El des-normalizador 220 puede estar configurado, por ejemplo, para modificar, dependiendo del
15 valor de des-normalización, al menos uno del primer canal y el segundo canal de la señal de audio intermedia que está representada en el dominio del tiempo para obtener una señal de audio des-normalizada. El aparato comprende además una unidad de post-procesamiento 235 que puede estar configurada, por ejemplo, para procesar la señal de audio des-normalizada, que es una señal de audio perceptualmente blanqueada, para obtener el primer canal y el segundo canal de la señal de audio decodificada.

[0093] Según otra realización, ilustrada en la Fig. 2e, el aparato comprende asimismo un post-procesador en el dominio espectral 212 que está configurado para realizar modelado de ruido temporal del lado del decodificador en la señal de audio intermedia. En ese tipo de realización, la unidad de transformación 215 está configurada para transformar la señal de audio intermedia del dominio espectral al dominio del tiempo, una vez realizado el modelado
25 de ruido temporal del lado del decodificador en la señal de audio intermedia.

[0094] En otra realización que no se reivindica, la unidad de decodificación 210 puede estar configurada, por ejemplo, para aplicar el relleno inteligente de huecos estéreo del lado del decodificador en la señal de audio codificada.

[0095] Además, como se ilustra en la Fig. 2f, se da a conocer un sistema para decodificar una señal de audio
35 codificada que comprende cuatro o más canales para obtener cuatro canales de una señal de audio decodificada que comprende cuatro o más canales. El sistema comprende un primer aparato 270 según una de las realizaciones antes descritas para decodificar un primer canal y un segundo canal de los cuatro o más canales de la señal de audio codificada para obtener un primer canal y un segundo canal de la señal de audio decodificada. Además, el sistema comprende un segundo aparato 280 según una de las realizaciones antes descritas para decodificar un tercer canal y
40 un cuarto canal de los cuatro o más canales de la señal de audio codificada para obtener un tercer canal y un cuarto canal de la señal de audio decodificada.

[0096] La Fig. 3 ilustra un sistema para generar una señal de audio codificada a partir de una señal de entrada de audio y para generar una señal de audio decodificada a partir de la señal de audio codificada según una realización.

[0097] El sistema comprende un aparato 310 para codificación según una de las realizaciones antes descritas, en la cual el aparato 310 para codificación está configurado para generar la señal de audio codificada a partir de la
45 señal de entrada de audio.

[0098] Además, el sistema comprende un aparato 320 para decodificación según lo descrito anteriormente. El aparato 320 para decodificación está configurado para generar la señal de audio decodificada a partir de la señal de audio codificada.

[0099] De manera similar, se da a conocer un sistema para generar una señal de audio codificada a partir de
55 una señal de entrada de audio y para generar una señal de audio decodificada a partir de la señal de audio codificada. El sistema comprende un sistema según la realización de la Fig. 1f, en la cual el sistema según la realización de la Fig. 1f está configurado para generar la señal de audio codificada a partir de la señal de entrada de audio, y un sistema

según la realización de la Fig. 2f, en la cual el sistema de la realización de la Fig. 2f está configurado para generar la señal de audio decodificada a partir de la señal de audio codificada.

[0100] En lo sucesivo se describen las realizaciones preferidas.

5

[0101] La Fig. 4 ilustra un aparato para codificación según otra realización. Entre otros elementos, se ilustra una unidad de pre-procesamiento 105 y una unidad de transformación 102 de acuerdo con una realización específica. La unidad de transformación 102 está configurada, entre otras cosas, para ejecutar una transformación de la señal de entrada de audio de un dominio del tiempo al dominio espectral, y la unidad de transformación está configurada para llevar a cabo el modelado de ruido temporal del lado del codificador y el modelado de ruido en el dominio de la frecuencia del lado del codificador en la señal de entrada de audio.

10

[0102] Además, la Fig. 5 ilustra módulos de procesamiento estéreo en un aparato para codificación según una realización. La Fig. 5 ilustra un normalizador 110 y una unidad de codificación 120.

15

[0103] Por otro lado, la Fig. 6 ilustra un aparato para decodificación según otra realización. Entre otras cosas, la Fig. 6 ilustra una unidad de post-procesamiento 230 según una realización específica. La unidad de post-procesamiento 230 está configurada, entre otras cosas, para obtener una señal de audio procesada del des-normalizador 220, y la unidad de post-procesamiento 230 está configurada para realizar al menos uno de modelado de ruido temporal del lado del decodificador y modelado de ruido en el dominio de la frecuencia del lado del decodificador en la señal de audio procesada.

20

[0104] Se puede efectuar la Detección de Transitorios en el Dominio del Tiempo (del inglés, Time Domain Transient Detector (TD TD)), la formación en ventanas, MDCT, MDST (por sus siglas en inglés, Modified Discrete Sine Transform, Transformada de Seno Discreta Modificada) y OLA (por sus siglas en inglés, Optical Linea Amplifier, Amplificador de Línea Óptica), por ejemplo, según lo descrito en [6a] o [6b]. MDCT y MDST forman la Transformada Solapada Modulada Compleja (del inglés, Modulated Complex Lapped Transform (MCLT)); la realización por separado de la MDCT y la MDST es equivalente a la realización de la MCLT; "MCLT a MDCT" representa tomar sólo la parte de MDCT de la MCLT y descartar la MDST (véase [12]).

30

[0105] La elección de diferentes longitudes de ventana en los canales izquierdo y derecho puede forzar, por ejemplo, la codificación doble mono en esa trama.

[0106] El Modelado de Ruido Temporal (TNS, por sus siglas en inglés) se puede efectuar, por ejemplo, de manera similar a la descrita en [6a] o [6b].

35

[0107] El modelado de ruido en el dominio de la frecuencia (FDNS) y el cálculo de los parámetros de FDNS pueden ser, por ejemplo, similares al procedimiento descrito en [8]. Una diferencia puede ser, por ejemplo, que los parámetros de FDNS correspondientes a las tramas donde el TNS está inactivo se calculan a partir del espectro de MCLT (por sus siglas en inglés, Modulated Complex Lapped Transform, Transformada Traslapada Modulada Compleja). En las tramas donde el TNS está activo, se puede estimar la MDST (por sus siglas en inglés Modified Discrete Sine Transform, Transformada de Seno Discreta Modificada), por ejemplo, a partir de la MDCT (por sus siglas en inglés Modified Discrete Cosine Transform, Transformada Discreta de Coseno Modificada).

40

[0108] También se puede reemplazar el FDNS con el blanqueo del espectro perceptual en el dominio del tiempo (como se describe, por ejemplo, en [13]).

45

[0109] El procesamiento estéreo consiste en un procesamiento de ILD (por sus siglas en inglés Interaural Level Difference, Diferencia de Nivel Interaural) global, el procesamiento M/S por banda, la distribución de la tasa de bits entre los canales.

50

[0110] La *ILD* global única se calcula de la siguiente manera

$$NRG_L = \sqrt{\sum MDCT_{L,k}^2}$$

55

$$NRG_R = \sqrt{\sum MDCT_{R,k}^2}$$

$$ILD = \frac{NRG_L}{NRG_L + NRG_R}$$

donde $MDCT_{L,k}$ es el k° coeficiente del espectro de MDCT en el canal izquierdo y $MDCT_{R,k}$ es el k° coeficiente del espectro de MDCT en el canal derecho. La ILD global se cuantifica de manera uniforme:

$$\hat{ILD} = \max\left(1, \min\left(ILD_{\text{intervalo}} - 1, \lfloor ILD_{\text{intervalo}} \cdot ILD + 0,5 \rfloor\right)\right)$$

$$ILD_{\text{intervalo}} = 1 \ll ILD_{\text{bits}}$$

donde ILD_{bits} es el número de bits utilizado para codificar la ILD global. $\hat{ILD}$ se almacena en el flujo de bits.

$\ll$ es una operación de desplazamiento de bits y desplaza los bits ILD_{bits} a la izquierda mediante la inserción de 0 bits.

En otras palabras: $ILD_{\text{intervalo}} = 2^{ILD_{\text{bits}}}$.

A continuación, la relación de energía de los canales es:

$$relación_{ILD} = \frac{ILD_{\text{intervalo}}}{\hat{ILD}} - 1 \approx \frac{NRG_R}{NRG_L}$$

Si $relación_{ILD} > 1$ entonces se escala el canal derecho con $\frac{1}{relación_{ILD}}$, de lo contrario se escala el canal izquierdo con $relación_{ILD}$. Esto significa, efectivamente, que se escala el canal más sonoro.

Si se utiliza el blanqueo del espectro perceptual en el dominio del tiempo (como se describe, por ejemplo, en [13]), también se puede calcular y aplicar la ILD global única en el dominio del tiempo, antes de la transformación al dominio de la frecuencia (es decir, antes de la MDCT). O bien, por otro lado, tras el blanqueo del espectro perceptual se puede realizar la transformación del dominio del tiempo a la frecuencia seguida por la ILD global única en el dominio de la frecuencia. Por otro lado, se puede calcular y aplicar la ILD global única en el dominio del tiempo antes de la transformación del dominio del tiempo a la frecuencia en el dominio de la frecuencia después de la transformación del dominio del tiempo a la frecuencia.

Los canales de $MDCT_{M,k}$ medio y $MDCT_{S,k}$ lateral se forman utilizando la $MDCT_{L,k}$ del canal izquierdo y la $MDCT_{R,k}$ del canal derecho en términos de $MDCT_{M,k} = \frac{1}{\sqrt{2}}(MDCT_{L,k} + MDCT_{R,k})$ y $MDCT_{S,k} = \frac{1}{\sqrt{2}}(MDCT_{L,k} - MDCT_{R,k})$. El espectro se divide en bandas y por cada banda se decide si se utiliza el canal izquierdo, derecho, medio o lateral.

Se estima una ganancia global G_{est} de la señal que comprende los canales Izquierdo y Derecho concatenados. Esto se diferencia de [6b] y [6a]. Se puede usar, por ejemplo, la primera estimación de la ganancia descrita en el capítulo 5.3.3.2.8.1.1 "Global gain estimator" (Estimador de ganancia global) de [6b] o de [6a], suponiendo una ganancia de SNR de 6 dB por muestra por bit a partir de la cuantificación escalar.

Se puede multiplicar la ganancia estimada por una constante para obtener una subestimación o una sobreestimación en la G_{est} final. A continuación se cuantifican las señales de los canales izquierdo, derecho, medio y lateral utilizando G_{est} , es decir que el tamaño de la etapa de cuantificación es $1/G_{est}$.

Seguidamente se codifican las señales cuantificadas utilizando un codificador aritmético, un codificador de Huffman u otro codificador entrópico, con el fin de obtener el número de bits necesario. Por ejemplo, se puede utilizar el codificador aritmético basado en el contexto descrito en el capítulo 5.3.3.2.8.1.3 – capítulo 5.3.3.2.8.1.7 de [6b] o de [6a]. Dado que se debe ejecutar el bucle de tasa (por ej. 5.3.3.2.8.1.2 de [6b] o de [6a]) después de la codificación estéreo, basta con una estimación de los bits necesarios.

Por ejemplo, se estima el número de bits necesarios para la codificación aritmética basada en contexto por cada canal cuantificado de acuerdo con lo descrito en el capítulo 5.3.3.2.8.1.3 – capítulo 5.3.3.2.8.1.7 de [6b] o

de [6a].

[0120] Según una realización, se determina la estimación de bits por cada canal cuantificado (izquierdo, derecho, medio o lateral) basándose en el siguiente código ejemplo:

```

5
  int context_based_arithmetic_coder_estimate (
    int spectrum[],
    int start_line,
    int end_line,
10  int lastnz, // lastnz = last non-zero spectrum line
    int & ctx, // ctx = context
    int & probability, // 14 bit fixed point probability
    const unsigned int cum_freq[N_CONTEXTS] []
    // cum_freq = cumulative frequency tables, 14 bit fixed point
15  )
  { int nBits = 0;
    for (int k = start_line; k < min(lastnz, end_line); k+=2)
    { int a1 = abs(spectrum[k]);
      int b1 = abs(spectrum[k+1]);
20    /* Signs Bits */
      nBits += min(a1, 1);
      nBits += min(b1, 1);
      while (max(a1, b1) >= 4)
      { probability *= cum_freq[ctx][VAL_ESC];
25      int nlz = Number_of_leading_zeros(probability);
        nBits += 2 + nlz;
        probability >>= 14 - nlz;
        a1 >>= 1;
        b1 >>= 1;
30      ctx = update_context(ctx, VAL_ESC);
    }
    int symbol = a1 + 4*b1;
    probability *= (cum_freq [ctx] [symbol] -
                    cum_freq[ctx] [symbol+1]) ;
35  int nlz = Number_of_leading_zeros(probability);
    nBits += nlz;
    hContextMem->proba >>= 14 - nlz;
    ctx = update_context(ctx, a1+b1);
  }
40  return nBits;
  }

```

donde se establece spectrum para apuntar al espectro cuantificado que se va a codificar, se establece start_line en 0, se establece end_line según la longitud del espectro, se establece lastnz al índice del último elemento no cero del
45 espectro, se establece ctx en 0 y se establece la probabilidad en una notación puntual fija de 1 a 14 bits (16384=1<<14).

[0121] Según se ha señalado, se puede emplear el código ejemplo anterior, por ejemplo, para obtener una estimación de bits correspondiente al menos a uno del canal izquierdo, el canal derecho, el canal medio y el canal
50 lateral.

[0122] Algunas realizaciones emplean un codificador aritmético de acuerdo con lo descrito en [6b] y [6a]. Se pueden encontrar más detalles, por ejemplo, en el capítulo 5.3.3.2.8 "Arithmetic encoder" (Codificador aritmético) de [6b].

55 **[0123]** A continuación, el número estimado de bits para "doble mono total" (b_{LR}) es igual a la suma de los bits necesarios para los canales derecho e izquierdo.

[0124] A continuación, el número estimado de bits para "M/S total" (b_{MS}) es igual a la suma de los bits
60 necesarios para el canal medio y el lateral.

[0125] En una realización alternativa, que es alternativa al código de ejemplo anterior, se puede utilizar, por

ejemplo, la fórmula:

$$b_{LR} = \sum_{i=0}^{nBandas-1} b_{bwLR}^i$$

5 para calcular un número de bits estimado para “doble mono total” (b_{LR}).

[0126] Más aun, en una realización alternativa, que es una alternativa del código ejemplo anterior, se puede emplear, por ejemplo, la fórmula:

$$b_{MS} = \sum_{i=0}^{nBandas-1} b_{bwMS}^i$$

10

para calcular un número de bits estimado para el “M/S total” (b_{MS}).

[0127] Por cada banda i con los bordes $[lb_i, ub_i]$, se verifica cuántos bits se utilizarían para codificar la señal cuantificada en la banda en el modo L/R (b_{bwLR}^i) y en el M/S (b_{bwMS}^i). En otras palabras, se lleva a cabo una estimación de bits banda por banda para el modo L/R por cada banda i : b_{bwLR}^i , lo que da lugar a la estimación de bits por banda en el modo L/R en el caso de la banda i , y una estimación de bits por banda en el modo M/S por cada banda i , lo que da lugar a la estimación de bits por banda en el modo M/S respecto de la banda i : b_{bwMS}^i .

20 **[0128]** Se elige el modo con menor número de bits para la banda. Se estima el número de bits necesarios para la codificación aritmética de acuerdo con lo descrito en el capítulo 5.3.3.2.8.1.3 – capítulo 5.3.3.2.8.1.7 de [6b] o de [6a]. El número total de bits necesarios para codificar el espectro en el modo “M/S por banda” (b_{BW}) es igual a la suma de $\min(b_{bwLR}^i, b_{bwMS}^i)$:

$$b_{BW} = nBandas + \sum_{i=0}^{nBandas-1} \min(b_{bwLR}^i, b_{bwMS}^i)$$

[0129] El modo “M/S por banda” requiere bits adicionales de $nBandas$ para señalar, en cada banda, si se utiliza la codificación L/R o la M/S. La elección entre el “M/S por banda”, el “doble mono total” y el “M/S total” se puede codificar, por ejemplo, como modo estéreo en el flujo de bits y, a continuación, el “doble mono total” y el “M/S total” no necesitan bits adicionales, en comparación con el “M/S por banda”, para la señalización.

[0130] En el caso del codificador aritmético basado en el contexto, b_{bwLR}^i usado en el cálculo de bLR no es igual a b_{bwLR}^i utilizado en el cálculo de bBW, ni b_{bwMS}^i usado en el cálculo de bMS es igual a b_{bwMS}^i usado en el cálculo de bBW, ya que el b_{bwLR}^i y el b_{bwMS}^i dependen de la elección del contexto correspondiente a los b_{bwLR}^j y b_{bwMS}^j anteriores, donde $j < i$. bLR se puede calcular en términos de suma de los bits para el canal izquierdo y para el derecho y bMS se puede calcular en términos de suma de los bits para el canal medio y el lateral, donde se pueden calcular los bits correspondientes a cada canal empleando el código ejemplo context_based_arithmetic_coder_estimate_bandwise donde start_line se fija en 0 y end_line se fija en lastnz.

40 **[0131]** En una realización alternativa, que es una alternativa del código ejemplo antes expuesto, se puede emplear, por ejemplo, la fórmula:

$$b_{LR} = nBandas + \sum_{i=0}^{nBandas-1} b_{bwLR}^i$$

para calcular el número de bits estimado para el “doble mono total” (b_{LR}) y señalar, en cada banda, la codificación L/R.

[0132] Además, en una realización alternativa, que es una alternativa del código ejemplo antes expuesto, se puede emplear, por ejemplo, la fórmula:

$$b_{MS} = nBandas + \sum_{i=0}^{nBandas-1} b_{bWMS}^i$$

para calcular un número estimado de bits para el “M/S total” (b_{MS}) y señalar, en cada banda, la codificación M/S.

[0133] En algunas realizaciones, en primer lugar, se puede estimar, por ejemplo, una ganancia G y se puede estimar, por ejemplo, un tamaño de la etapa de cuantificación, para lo cual se espera que haya suficientes bits para codificar los canales en L/R.

[0134] A continuación, se exponen realizaciones que describen diferentes maneras de determinar una estimación de bits por banda; por ejemplo, se describe cómo determinar b_{bWLR}^i y b_{bWMS}^i de acuerdo con realizaciones específicas.

[0135] Como ya se mencionara, según una realización específica, por cada canal cuantificado, se estima el número de bits necesario para la codificación aritmética, por ejemplo, de acuerdo con lo descrito en el capítulo 5.3.3.2.8.1.7 “Bit consumption estimation” (Estimación del consumo de bits) de [6b] o del capítulo similar de [6a].

[0136] Según una realización, la estimación de bits por banda se determina empleando context_based_arithmetic_coder_estimate para calcular cada uno de b_{bWLR}^i y b_{bWMS}^i por cada i , estableciendo start_line en lb_i , end_line en ub_i , lastnz en el índice del último elemento no cero de spectrum.

[0137] Se inicializan cuatro contextos (ctx_L , ctx_R , ctx_M , ctx_S) y cuatro probabilidades (p_L , p_R , p_M , p_S) y, a continuación, se actualizan repetidamente.

[0138] Al comienzo de la estimación (en el caso de $i = 0$) se establece cada contexto (ctx_L , ctx_R , ctx_M , ctx_S) en 0 y se establece cada probabilidad (p_L , p_R , p_M , p_S) en 1 en la notación puntual fija de 14bits ($16384=1<<14$).

b_{bWLR}^i se calcula en términos de la suma de b_{bWL}^i y b_{bWR}^i , donde b_{bWL}^i se determina usando context_based_arithmetic_coder_estimate estableciendo que spectrum apunte al espectro izquierdo cuantificado que se va a codificar, ctx se fija en ctx_L y la probabilidad se fija en p_L ; además b_{bWR}^i se determina usando context_based_arithmetic_coder_estimate estableciendo que spectrum apunte al espectro derecho cuantificado que se va a codificar, ctx se fija en ctx_R y la probabilidad se fija en p_R .

b_{bWMS}^i se calcula en términos de suma de b_{bWM}^i y b_{bWS}^i , donde b_{bWM}^i se determina empleando context_based_arithmetic_coder_estimate estableciendo que spectrum apunte al espectro medio que se va a codificar, se fija ctx en ctx_M y se establece que la probabilidad es p_M y se determina b_{bWS}^i utilizando context_based_arithmetic_coder_estimate estableciendo que spectrum apunte al espectro lateral cuantificado por codificar, se fija ctx en ctx_S y la probabilidad se fija en p_S .

[0139] Si $b_{bWLR}^i < b_{bWMS}^i$, entonces se establece que ctx_L es ctx_M , se establece que ctx_R es ctx_S , se establece que p_L es p_M , se establece que p_R es p_S .

[0140] Si $b_{bWLR}^i \geq b_{bWMS}^i$ entonces se establece que ctx_M es ctx_L , se establece que ctx_S es ctx_R , se establece que p_M es p_L , se establece que p_S es p_R .

[0141] En una realización alternativa, se obtiene la estimación de bits por banda de la siguiente manera: Se divide el espectro en bandas y, por cada banda, se decide si se debe efectuar el procesamiento M/S. En el caso de todas las bandas en que se utiliza M/S, se reemplazan $MDCT_{L,k}$ y $MDCT_{R,k}$ por $MDCT_{M,k} = 0.5(MDCT_{L,k} + MDCT_{R,k})$ y $MDCT_{S,k} = 0.5(MDCT_{L,k} - MDCT_{R,k})$.

[0142] La decisión M/S por banda vs L/R se puede basar, por ejemplo, en la estimación de ahorro de bits con el procesamiento M/S:

$$bitsSaved_i = nlines_i \cdot \log_2 \sqrt{\frac{NRG_{R,i} NRG_{L,i}}{NRG_{M,i} NRG_{S,i}}}$$

donde $NRG_{R,i}$ es la energía en la i^a banda del canal derecho, $NRG_{L,i}$ es la energía en la i^a banda del canal izquierdo, $NRG_{M,i}$ es la energía en la i^a banda del canal medio, $NRG_{S,i}$ es la energía en la i^a banda del canal lateral y $nlines_i$ es el número de coeficientes espectrales en la i^a banda. El canal medio es la suma del canal izquierdo y el derecho, el canal lateral es la diferencia de los canales izquierdo y derecho.

$bitsSaved_i$ está limitado por el número de bits estimado que se debe utilizar para la i^a banda:

$$maxBits_{LR} = \left(\frac{NRG_{R,i}}{NRG_R} + \frac{NRG_{L,i}}{NRG_L} \right) \cdot bitsAvailable$$

$$maxBits_{MS} = \left(\frac{NRG_{M,i}}{NRG_M} + \frac{NRG_{S,i}}{NRG_S} \right) \cdot bitsAvailable$$

$$\overline{bitsSaved_i} = \max \left(maxBits_{LR}, \min(-maxBits_{MS}, bitsSaved_i) \right)$$

[0143] La Fig. 7 ilustra el cálculo de una tasa de bits para la decisión M/S por bandas según una realización.

[0144] En particular, en la Fig. 7, se ilustra el proceso para el cálculo de b_{BW} . Para reducir la complejidad, se guarda el contexto del codificador aritmético para codificar el espectro hasta la banda $i - 1$ y se reutiliza en la banda i .

[0145] Cabe señalar que, en el caso del codificador aritmético basado en el contexto, b_{BW}^{LR} y b_{BW}^{MS} dependen del contexto del codificador aritmético, que depende de la elección de M/S o L/R en todas las bandas $j < i$, de acuerdo con lo descrito, por ejemplo, anteriormente.

[0146] La Fig. 8 ilustra una decisión del modo estéreo según una realización.

[0147] Si se elige "doble mono total", entonces el espectro completo consiste en $MDCT_{L,k}$ y $MDCT_{R,k}$. Si se elige "M/S total", entonces el espectro completo consiste en $MDCT_{M,k}$ y $MDCT_{S,k}$. Si se elige "M/S por banda", entonces algunas bandas del espectro consisten en $MDCT_{L,k}$ y $MDCT_{R,k}$ y otras bandas consisten en $MDCT_{M,k}$ y $MDCT_{S,k}$.

[0148] El modo estéreo se codifica en el flujo de bits. En el modo "M/S por banda" también la decisión M/S por banda se codifica en el flujo de bits.

[0149] Los coeficientes del espectro en los dos canales después del procesamiento estéreo se indican como $MDCT_{LM,k}$ y $MDCT_{RS,k}$. $MDCT_{LM,k}$ es igual a $MDCT_{M,k}$ en las bandas de M/S o a $MDCT_{L,k}$ en las bandas de L/R y $MDCT_{RS,k}$ es igual a $MDCT_{S,k}$ en las bandas de M/S o a $MDCT_{R,k}$ en las bandas de L/R, dependiendo del modo estéreo y de la decisión M/S por banda. Se puede hacer referencia al espectro que consiste en $MDCT_{LM,k}$, por ejemplo, como canal 0 codificado en forma conjunta (Cn 0 Conjunto) o se le puede denominar, por ejemplo, primer canal, y se puede hacer referencia al espectro que consiste en $MDCT_{RS,k}$, por ejemplo, como canal 1 codificado en forma conjunta (Cn 1 Conjunto) o se le puede denominar, por ejemplo, segundo canal.

[0150] La relación de división de la tasa de bits se calcula empleando las energías de los canales procesados estéreo:

$$NRG_{LM} = \sqrt{\sum MDCT_{LM,k}^2}$$

$$NRG_{RS} = \sqrt{\sum MDCT_{RS,k}^2}$$

$$r_{split} = \frac{NRG_{LM}}{NRG_{LM} + NRG_{RS}}$$

5 **[0151]** Se cuantifica la relación de división de la tasa de bits de manera uniforme:

$$r_{split}^{\wedge} = \max\left(1, \min(rsplit_{range} - 1, \lfloor rsplit_{range} \cdot r_{split} + 0.5 \rfloor)\right)$$

$$rsplit_{range} = 1 \ll rsplit_{bits}$$

10

donde $rsplit_{bits}$ es el número de bits utilizado para codificar la relación de división de la tasa de bits. Si

$$r_{split} < \frac{8}{9}$$

y

$$r_{split}^{\wedge} > \frac{9rsplit_{range}}{16}$$

15 entonces

$$r_{split}^{\wedge}$$

se reduce para

$$\frac{rsplit_{range}}{8}.$$

si

20

$$r_{split} > \frac{1}{9}$$

y

$$r_{split}^{\wedge} < \frac{7rsplit_{range}}{16}$$

entonces se incrementa

$$r_{split}^{\wedge}$$

25 para

$$\frac{rsplit_{range}}{8}.$$

se almacena $r_{split}^{\wedge}$ en la tasa de bits.

30 **[0152]** La distribución de la tasa de bits entre los canales es:

$$bits_{LM} = \lfloor \frac{r_{split}^{\wedge}}{rsplit_{range}} (totalBitsAvailable - stereoBits) \rfloor$$

$$bits_{RS} = (totalBitsAvailable - stereoBits) - bits_{LM}$$

35 **[0153]** Además, se verifica que haya suficientes bits para el codificador entrópico en cada canal controlando que $bits_{LM} - sideBits_{LM} > minBits$ y $bits_{RS} - sideBits_{RS} > minBits$, donde $minBits$ es el número mínimo de bits que requiere el codificador entrópico. Si no hay suficientes bits para el codificador entrópico, entonces se aumenta/reduce $r_{split}^{\wedge}$ en 1 hasta que se cumpla $bits_{LM} - sideBits_{LM} > minBits$ y $bits_{RS} - sideBits_{RS} > minBits$.

40 **[0154]** La cuantificación, el relleno de ruido y la codificación entrópica, incluyendo el bucle de tasa, son como se describe en 5.3.3.2 "General encoding procedure" (Procedimiento de Codificación General) de 5.3.3 "MDCT based TCX" (TCX (por sus siglas en inglés, Transformed Coded Excitation, Excitación Codificada Transformada) basada en MDCT) de [6b] o de [6a]. El bucle de tasa se puede optimizar empleando la G_{est} estimada. Se utiliza el espectro de energía P (magnitud de la MCLT) para las medidas de tonalidad/ruido en la cuantificación y el Relleno Inteligente de

Huecos (IGF, por sus siglas en inglés), de acuerdo con lo descrito en [6a] o [6b]. Dado que se utiliza el espectro de MDCT blanqueado y procesado M/S por banda para el espectro de energía, se debe realizar el mismo FDNS y procesamiento M/S en el espectro de MDST. Se debe efectuar el mismo escalado basado en la ILD global del canal más sonoro para la MDST que para la MDCT. En el caso de las tramas en que el TNS está activo, se utiliza el espectro de MDST para el cálculo del espectro de energía a partir del espectro de MDCT blanqueado y procesado M/S: $P_k = MDCT_k^2 + (MDCT_{k+1} - MDCT_{k-1})^2$.

[0155] El proceso de decodificación se inicia con la decodificación y cuantificación inversa del espectro de los canales codificados de forma conjunta, seguidas por el relleno de ruido descrito en 6.2.2 "MDCT based TCX" (TCX basada en MDCT) de [6b] o [6a]. Se determina el número de bits asignado a cada canal sobre la base de la longitud de ventana, el modo estéreo y la relación de división de la tasa de bits que se codifican en el flujo de bits. Se debe saber el número de bits asignado a cada canal antes de decodificar por completo el flujo de bits.

[0156] En el bloque de relleno inteligente de huecos (IGF, por sus siglas en inglés), se rellenan las líneas cuantificadas en cero en un cierto intervalo del espectro denominado mosaico objetivo con un contenido procesado de un intervalo diferente del espectro, denominado mosaico de origen. Debido al procesamiento estéreo por banda, la representación estéreo (es decir ya sea L/R o M/S) podría diferir entre el mosaico de origen y la objetivo. Para garantizar una buena calidad, si la representación del mosaico de origen es diferente de la representación del mosaico objetivo, se procesa el mosaico de origen para transformarlo a la representación del mosaico objetivo antes del relleno de huecos en el decodificador. Este procedimiento ya se ha descrito en [9]. El IGF en sí se aplica, a diferencia de [6a] y [6b], en el dominio espectral blanqueado en lugar del dominio espectral original. A diferencia de los códecs estéreo conocidos (por ejemplo, [9]), el IGF se aplica en el dominio espectral blanqueado con ILD compensada.

[0157] Basándose en el modo estéreo y en la decisión M/S por banda, se construye el canal izquierdo y el derecho a partir de los canales codificados en forma conjunta: $MDCT_{L,k} = \frac{1}{\sqrt{2}}(MDCT_{LM,k} + MDCT_{RS,k})$ y $MDCT_{R,k} = \frac{1}{\sqrt{2}}(MDCT_{LM,k} - MDCT_{RS,k})$.

[0158] Si $ratio_{ILD} > 1$, entonces se escala el canal derecho con $ratio_{ILD}$, de lo contrario se escala el canal izquierdo con $\frac{1}{ratio_{ILD}}$.

[0159] Por cada caso en que se pueda dar la división por 0, se añade una pequeña épsilon al denominador.

[0160] En el caso de las tasas de bits intermedias, por ejemplo 48 kbps, la codificación basada en MDCT puede llevar, por ejemplo, a una cuantificación demasiado basta del espectro para llegar al objetivo de consumo de bits. Eso suscita la necesidad de codificación paramétrica que, combinada con la codificación discreta en la misma región espectral, adaptada a la realización trama por trama, aumenta la fidelidad.

[0161] En lo sucesivo se describen aspectos de algunas de las realizaciones que emplean relleno estéreo. Se debe tener en cuenta que en las realizaciones antes expuestas no es necesario emplear ese relleno estéreo. Por ello, sólo algunas de las realizaciones anteriormente descritas emplean relleno estéreo. Otras realizaciones de las realizaciones antes descritas no emplean relleno estéreo en absoluto.

[0162] El relleno de frecuencia estéreo en el estéreo en el dominio de la frecuencia de MPEG-H (por sus siglas en inglés, Moving Picture Expert Group-High efficiency, Grupo de Expertos en Imágenes en Movimiento de Alta Eficiencia, ha sido descrito, por ejemplo, en [11]. En [11] se alcanza la energía objetivo por cada banda aprovechando la energía de banda enviada desde el codificador en forma de factores de escala (por ejemplo en AAC (por sus siglas en inglés, Advanced Audio Coding, Codificación Avanzada de Audio). Si se aplica el modelado de ruido en el dominio de la frecuencia (FDNS) y se codifica la envolvente espectral utilizando las LSFs (siglas en inglés de line spectral frequencies, frecuencias espectrales de líneas) (véase [6a], [6b], [8]) no es posible cambiar el escalado sólo para algunas bandas de frecuencia (bandas espectrales) necesario para el algoritmo de relleno estéreo descrito en [11].

[0163] En primer lugar se presenta cierta información de antecedentes.

[0164] Cuando se emplea la codificación media/lateral, es posible codificar las señales laterales de diferentes maneras.

[0165] Según un primer grupo de realizaciones, se codifica una señal lateral S de igual modo que una señal media M . Se lleva a cabo la cuantificación, pero no se realiza ninguna otra etapa para reducir la tasa de bits necesaria. En general, esa estrategia apunta a permitir una reconstrucción precisa de la señal lateral S del lado del decodificador aunque, por otro lado, requiere una gran cantidad de bits para codificar.

5

[0166] Según un segundo grupo de realizaciones, se genera una señal lateral residual S_{res} a partir de la señal lateral original S basada en la señal M . En una realización, se puede calcular la señal lateral residual, por ejemplo, según la fórmula:

10

$$S_{\text{res}} = S - g \cdot M.$$

[0167] Otras realizaciones pueden emplear, por ejemplo, otras definiciones de la señal lateral residual.

[0168] La señal residual S_{res} se cuantifica y transmite al decodificador, junto con el parámetro g . Al cuantificar la señal residual S_{res} en lugar de la señal lateral original S , en general, se cuantifican más valores espectrales en cero. En general, esto ahorra la cantidad de bits necesaria para codificar y transmitir en comparación con la señal lateral original cuantificada S .

[0169] En algunas de estas realizaciones del segundo grupo de realizaciones, se determina un parámetro único g para el espectro completo y se transmite al decodificador. En otras realizaciones del segundo grupo de realizaciones, cada una de una pluralidad de bandas de frecuencia/ bandas espectrales del espectro de frecuencia puede comprender, por ejemplo, dos o más valores espectrales, y se determina un parámetro g por cada una de las bandas de frecuencia/ bandas espectrales y se transmite al decodificador.

[0170] La Fig. 12 ilustra el procesamiento estéreo del lado de un codificador según el primero o el segundo grupo de realizaciones, que no emplean relleno estéreo.

[0171] La Fig. 13 ilustra el procesamiento estéreo del lado de un decodificador según el primero o el segundo grupo de realizaciones, que no emplean relleno estéreo.

30

[0172] Según un tercer grupo de realizaciones, se emplea el relleno estéreo. En algunas de estas realizaciones, del lado del decodificador, se genera la señal lateral S correspondiente a un cierto punto de tiempo t a partir de una señal media del punto de tiempo inmediatamente anterior $t-1$.

[0173] La generación de la señal lateral S para un cierto punto de tiempo t a partir de una señal media del punto de tiempo inmediatamente anterior $t-1$ del lado del decodificador se puede realizar, por ejemplo, según la fórmula:

$$S(t) = h_b \cdot M(t-1).$$

[0174] Del lado del codificador, se determina el parámetro h_b por cada banda de frecuencia de una pluralidad de bandas de frecuencia del espectro. Después de determinar los parámetros h_b , el codificador transmite los parámetros h_b al decodificador. En algunas realizaciones, no se transmiten los valores espectrales de la señal lateral S en sí o de una residual de ésta al decodificador. Ese tipo de estrategia apunta a ahorrar el número de bits necesario.

[0175] En algunas otras realizaciones del tercer grupo de realizaciones, al menos en el caso de las bandas de frecuencia en que la señal lateral es más intensa que la señal media, se codifican explícitamente y se envían al decodificador los valores espectrales de la señal lateral de esas bandas de frecuencia.

[0176] Según un cuarto grupo de realizaciones, algunas de las bandas de frecuencia de la señal lateral S se codifican codificando explícitamente la señal lateral original S (véase el primer grupo de realizaciones) o una señal lateral residual S_{res} , mientras que, para las demás bandas de frecuencia, se emplea el relleno estéreo. Esa estrategia combina el primero o el segundo grupo de realizaciones, con el tercer grupo de realizaciones, que utiliza relleno estéreo. Por ejemplo, las bandas de frecuencia más baja se pueden codificar, por ejemplo, cuantificando la señal lateral original S o la señal lateral residual S_{res} , mientras que para las otras bandas, de frecuencia más alta, se puede emplear, por ejemplo, el relleno estéreo.

[0177] La Fig. 9 ilustra el procesamiento estéreo del lado de un codificador según el tercero o el cuarto grupo de realizaciones, que emplean relleno estéreo.

[0178] La Fig. 10 ilustra el procesamiento estéreo del lado de un decodificador según el tercero o el cuarto grupo de realizaciones, que emplean relleno estéreo.

5 **[0179]** Las realizaciones antes descritas que sí emplean relleno estéreo, pueden emplear, por ejemplo, el relleno estéreo según lo descrito en MPEG-H; véase estéreo MPEG-H en el dominio de la frecuencia (véase, por ejemplo, [11]).

[0180] Algunas de las realizaciones que emplean relleno estéreo pueden aplicar, por ejemplo, el algoritmo de
10 relleno estéreo descrito en [11] a sistemas en los cuales se codifica la envolvente espectral en forma de LSF (por sus siglas en inglés, Low Sampling Frequency, Baja Frecuencia de Muestreo) combinada con relleno de ruido. La codificación de la envolvente espectral se puede implementar, por ejemplo, de la manera descrita por ejemplo, en [6a], [6b], [8]. El relleno de ruido se puede implementar, por ejemplo, de acuerdo con lo descrito en [6a] y [6b].

15 **[0181]** En algunas realizaciones específicas, el procesamiento con relleno estéreo que incluye el cálculo de los parámetros de relleno estéreo se puede llevar a cabo, por ejemplo, en las bandas M/S dentro de la región de frecuencia, por ejemplo, desde una frecuencia más baja, como de $0,08 F_s$ (F_s = frecuencia de muestreo) a, por ejemplo, una frecuencia más elevada, por ejemplo, la frecuencia de cruce de IGF (por sus siglas en inglés, Intelligent Gap Filling, Relleno Inteligente de Huecos).

20 **[0182]** Por ejemplo, en el caso de las porciones de frecuencia más baja que la frecuencia inferior (por ejemplo, $0,08 F_s$), se puede cuantificar, por ejemplo, la señal lateral original S o una señal lateral residual derivada de la señal lateral original S, se puede transmitir, por ejemplo, al decodificador. En el caso de las porciones de frecuencia mayor que la frecuencia superior (por ejemplo, la frecuencia de cruce de IGF), se puede realizar, por ejemplo, el Relleno
25 Inteligente de Huecos (IGF).

[0183] Más específicamente, en algunas realizaciones, el canal lateral (el segundo canal), en el caso de las bandas de frecuencia dentro del intervalo del relleno estéreo (por ejemplo, $0,08$ veces la frecuencia de muestreo hasta la frecuencia de cruce de IGF) que se cuantifican por completo en cero se pueden rellenar, por ejemplo, usando un
30 “copiado y pegado” de la mezcla descendente del espectro de MDCT blanqueado de la trama anterior (IGF = Intelligent Gap Filling, Relleno Inteligente de Huecos). El “copiado y pegado” se puede aplicar, por ejemplo, como complemento del relleno de ruido y escalar de manera acorde dependiendo de los factores de corrección que se envían desde el codificador. En otras realizaciones, la frecuencia más baja puede exhibir valores diferentes de $0,08 F_s$.

35 **[0184]** En lugar de ser de $0,08 F_s$, en algunas realizaciones, la frecuencia más baja puede ser, por ejemplo, un valor en el intervalo de 0 a $0,50 F_s$. En realizaciones específicas, la frecuencia más baja puede ser un valor en el intervalo de $0,01 F_s$ a $0,50 F_s$. Por ejemplo, la frecuencia más baja puede ser, por ejemplo, de $0,12 F_s$ o $0,20 F_s$ o $0,25 F_s$.

40 **[0185]** En otras realizaciones, además o en lugar de emplear el Relleno Inteligente de Huecos, en el caso de las frecuencias superiores a la frecuencia superior, se puede llevar a cabo el Relleno de Ruido, por ejemplo.

[0186] En otras realizaciones, no hay una frecuencia superior y se lleva a cabo el relleno estéreo por cada porción de frecuencia superior a la frecuencia más baja.

45 **[0187]** En otras realizaciones adicionales, no hay frecuencia inferior y se lleva a cabo el relleno estéreo por las porciones de frecuencia que abarcan desde la banda de frecuencia más baja hasta la frecuencia superior.

[0188] En otras realizaciones más, no hay una frecuencia más baja y una frecuencia superior y se lleva a cabo
50 el relleno estéreo en todo el espectro de frecuencia.

[0189] A continuación se describen realizaciones específicas, que emplean relleno estéreo.

[0190] En particular, se describe el relleno estéreo con factores de corrección según realizaciones específicas.
55 Se puede emplear el relleno estéreo con factores de corrección, por ejemplo, en las realizaciones de los bloques de procesamiento de relleno estéreo de la Fig. 9 (lado del codificador) y de la Fig. 10 (lado del decodificador).

[0191] En lo sucesivo,

- Dmx_R puede indicar, por ejemplo, la señal media del espectro de MDCT blanqueado,
- S_R puede indicar, por ejemplo, la señal lateral del espectro de MDCT blanqueado,
- Dmx_I puede indicar, por ejemplo, la señal media del espectro de MDST blanqueado,
- 5 - S_I puede indicar, por ejemplo, la señal lateral del espectro de MDST blanqueado,
- $prevDmx_R$ puede indicar, por ejemplo, la señal media del espectro de MDCT blanqueado
- $prevDmx_I$ puede indicar, por ejemplo, la señal media del espectro de MDST blanqueado

[0192] La codificación con relleno estéreo se puede aplicar cuando la decisión estéreo es M/S para todas las
10 bandas (M/S total) o M/S para todas las bandas con relleno estéreo (M/S por banda).

[0193] Cuando se determina la aplicación del procesamiento doble mono total se pasa por alto el relleno
estéreo. Además, cuando se opta por la codificación L/R para algunas de las bandas espectrales (bandas de
frecuencia), también se omite el relleno estéreo para estas bandas espectrales.

15 **[0194]** Se consideran ahora realizaciones específicas que emplean relleno estéreo. En ellas, el procesamiento
dentro del bloque se puede llevar a cabo, por ejemplo, de la siguiente manera:

[0195] En el caso de las bandas de frecuencia (fb) que están dentro de la región de frecuencia a partir de la
20 frecuencia más baja (por ejemplo, $0,08 F_s$ (F_s = frecuencia de muestreo)), hasta la frecuencia superior, (por ejemplo,
la frecuencia de cruce de IGF):

- Se calcula una Res_R residual de la señal lateral S_R , por ejemplo, según:

25
$$Res_R = S_R - a_R Dmx_R - a_I Dmx_I .$$

donde a_R es la parte real y a_I es la parte imaginaria del coeficiente de predicción compleja (véase [10]).

Se calcula una residual Res_I de la señal lateral S_I , por ejemplo, según:

30
$$Res_I = S_I - a_R Dmx_R - a_I Dmx_I .$$

- Se calculan las energías, por ejemplo, las energías de valor complejo, de la residual Res y de la mezcla descendente
de la trama anterior (señal media) $prevDmx$:

35
$$ERes_{fb} = \sum_{fb} Res_R^2 + \sum_{fb} Res_I^2 ,$$

$$EprevDmx_{fb} = \sum_{fb} prevDmx_R^2 + \sum_{fb} prevDmx_I^2$$

[0196] En las fórmulas anteriores:

40
$$\sum_{fb} Res_R^2$$

suma los cuadrados de todos los valores espectrales dentro de la banda de frecuencia fb de Res_R .

$$\sum_{fb} Res_I^2$$

45 suma los cuadrados de todos los valores espectrales dentro de la banda de frecuencia fb de Res_I .

$$\sum_{fb} prevDmx_R^2$$

suma los cuadrados de todos los valores espectrales dentro de la banda de frecuencia fb de $prevDmx_R$.

50

$$\sum_{fb} prevDmx_i^2$$

suma los cuadrados de todos los valores espectrales dentro de la banda de frecuencia fb de $prevDmx_i$.

- 5 De estas energías calculadas, ($ERes_{fb}$, $EprevDmx_{fb}$), se calculan los factores de corrección de relleno estéreo y se transmiten como información lateral o secundaria al decodificador:

$$correction_factor_{fb} = ERes_{fb} / (EprevDmx_{fb} + \varepsilon)$$

- 10 **[0197]** En una realización, $\varepsilon = 0$. En otras realizaciones, por ejemplo, $0,1 > \varepsilon > 0$, por ejemplo, para evitar la división por 0.

- Se puede calcular, por ejemplo, un factor de escala por banda, dependiendo de los factores de corrección de relleno estéreo calculados, por ejemplo, por cada banda espectral para la cual se emplea el relleno estéreo. Se introduce el escalado por banda de las señales Media y Lateral (residual) de salida por un factor de escala con el fin de compensar la pérdida de energía, ya que no hay ninguna operación de predicción compleja inversa para reconstruir la señal lateral a partir de la residual del lado del decodificador ($a_R = a_I = 0$).

- 20 **[0198]** En una realización específica, el factor de escala por banda se puede calcular, por ejemplo, según:

$$factor_escala_{fb} = \sqrt{\frac{\sum_{fb} (S_R - a_R Dmx_R)^2 + \sum_{fb} (S_I - a_I Dmx_I)^2 + EDmx_{fb}}{ERes_{fb} + EDmx_{fb} + \varepsilon}}$$

donde $EDmx_{fb}$ es la energía (por ejemplo, compleja) de la mezcla descendente de la trama actual (que se puede calcular, por ejemplo, de la manera antes descrita).

- 25 - En algunas realizaciones, después del procesamiento de relleno estéreo en el bloque de procesamiento estéreo y antes de la cuantificación, los bins (cajas) de la residual que estén dentro del intervalo de frecuencia de relleno estéreo se pueden ajustar, por ejemplo, a cero, si la mezcla descendente correspondiente a la banda equivalente (Media) es más intensa que la residual (Lateral):

$$30 \quad \frac{E_{fb}^M}{E_{fb}^S} > umbral$$

$$E_{fb}^M = \sum_{fb} Dmx_R^2$$

$$E_{fb}^S = \sum_{fb} Res_R^2$$

- 35 **[0199]** Por lo tanto, se consumen más bits en la codificación de la mezcla descendente y los bins de frecuencia más baja de la residual, mejorando la calidad general.

- [0200]** En realizaciones alternativas, se pueden ajustar todos los bits de la residual (Lateral), por ejemplo, a cero. Tales realizaciones alternativas se pueden basar, por ejemplo, en la presunción de que la mezcla descendente es, en la mayoría de los casos, más intensa que la residual.

- [0201]** La Fig. 11 ilustra el relleno estéreo de una señal lateral según algunas realizaciones específicas del lado del decodificador.

- 45 **[0202]** El relleno estéreo se aplica al canal lateral después de la decodificación, la cuantificación inversa y el relleno de ruido. En el caso de las bandas de frecuencia, dentro del intervalo de relleno estéreo, que se cuantifican a cero, se puede aplicar un "copiado y pegado" de la mezcla descendente del espectro de MDCT blanqueado de la última trama, por ejemplo, (como se aprecia en la Fig. 11), si la energía de la banda después del relleno de ruido no

alcanza la energía pretendida. La energía objetivo por banda de frecuencia se calcula a partir de los factores de corrección estéreo que se envían en forma de parámetros desde el codificador, por ejemplo, según la fórmula.

$$ET_{fb} = correction_factor_{fb} \cdot EprevDmx_{fb} .$$

5

[0203] La generación de la señal lateral del lado del decodificador (a la que se puede hacer referencia, por ejemplo, como “copiado y pegado” de la mezcla descendente anterior) se lleva a cabo, por ejemplo, según la fórmula:

$$S_i = N_i + facDmx_{fb} \cdot prevDmx_i, i \in [fb, fb + 1],$$

10

donde i denota los bins de frecuencia (valores espectrales) dentro de la banda de frecuencia fb , N es el espectro con relleno de ruido y $facDmx_{fb}$ es un factor que se aplica a la mezcla descendente anterior, que depende de los factores de corrección de relleno estéreo enviados desde el codificador.

15 **[0204]**

Se puede calcular $facDmx_{fb}$, en una realización específica, por ejemplo, por cada banda de frecuencia fb como:

$$facDmx_{fb} = \sqrt{correction_factor_{fb} - \frac{EN_{fb}}{(EprevDmx_{fb} + \varepsilon)}}$$

20 donde EN_{fb} , es la energía del espectro con relleno de ruido en la banda fb y $EprevDmx_{fb}$, es la respectiva energía de mezcla descendente de la trama anterior.

[0205] Del lado del codificador, otras realizaciones no tienen en cuenta el espectro de MDST (o el espectro de MDCT). En esas realizaciones, se adapta el procedimiento del lado del codificador, por ejemplo, de la siguiente manera:

25

En el caso de las bandas de frecuencia (fb) que están dentro de la región de frecuencia que comienza con la frecuencia más baja (por ejemplo, $0,08 F_s$ (F_s = frecuencia de muestreo)), hasta la frecuencia superior, (por ejemplo, la frecuencia de cruce de IGF):

30 - Se calcula una residual Res de la señal lateral S_R , por ejemplo, según:

$$Res = S_R - a_R Dmx_R - a_I Dmx_I,$$

donde a_R es un coeficiente de predicción (por ejemplo, real).

35 - Se calculan las energías de la residual Res y de la mezcla descendente de la trama anterior (señal media) $prevDmx$:

$$ERes_{fb} = \sum_{fb} Res_R^2 + \sum_{fb} Res_I^2,$$

$$EprevDmx_{fb} = \sum_{fb} prevDmx_R^2 + \sum_{fb} prevDmx_I^2$$

40

- De estas energías calculadas, ($ERes_{fb}$, $EprevDmx_{fb}$), se calculan los factores de corrección de relleno estéreo y se transmiten como información lateral al decodificador:

$$correction_factor_{fb} = ERes_{fb} / (EprevDmx_{fb} + \varepsilon)$$

45

[0206] En una realización, $\varepsilon = 0$. En otras realizaciones, por ejemplo, $0,1 > \varepsilon > 0$, por ejemplo para evitar la división por 0.

- Se puede calcular el factor de escala por banda, por ej., dependiendo de los factores de corrección de relleno estéreo calculados, por ejemplo, por cada banda espectral para la cual se emplea el relleno estéreo.

50

[0207] En una realización específica, se puede calcular el factor de escala por cada banda, por ejemplo, según:

$$factor_escala_{fb} = \sqrt{\frac{\sum_{fb} (S_R - a_R Dmx_R)^2 + EDmx_{fb}}{ERes_{fb} + EDmx_{fb} + \varepsilon}}$$

donde $EDmx_{fb}$ es la energía de la mezcla descendente de la trama actual (que se puede calcular, por ejemplo, de la manera descrita anteriormente).

- 5 - En algunas realizaciones, después del procesamiento de relleno estéreo en el bloque de procesamiento estéreo y con anterioridad a la cuantificación, se pueden ajustar a cero los bins de la residual que están dentro del intervalo de frecuencia de relleno estéreo, por ejemplo, a cero, si en la banda equivalente la mezcla descendente (Medio) es más intensa que la residual (Lateral):

$$10 \quad \frac{E_{fb}^M}{E_{fb}^S} > umbral$$

$$E_{fb}^M = \sum_{fb} Dmx_R^2$$

$$E_{fb}^S = \sum_{fb} Res_R^2$$

15

[0208] Por lo tanto, se consumen más bits en la codificación de la mezcla descendente y los bins de frecuencia más baja de la residual, mejorando así la calidad general.

- [0209]** En realizaciones alternativas, se pueden ajustar todos los bits de la residual (Lateral), por ejemplo, a
20 cero. Esas realizaciones alternativas se pueden basar, por ejemplo, en la presunción de que, en la mayoría de los casos, la mezcla descendente es más intensa que la residual.

- [0210]** Según algunas realizaciones, se pueden incluir medios, por ejemplo, para aplicar relleno estéreo en sistemas con FDNS, donde la envolvente espectral se codifica utilizando LSF (o una codificación similar en que no es
25 posible cambiar el escalado en las bandas individuales de forma independiente). Según algunas de las realizaciones, se pueden incluir medios, por ejemplo, para aplicar relleno estéreo en sistemas sin la predicción compleja/real.

- [0211]** Algunas de las realizaciones pueden emplear, por ejemplo, relleno estéreo paramétrico, en el sentido de que se envían parámetros explícitos (factores de corrección de relleno estéreo) del codificador al decodificador,
30 para controlar el relleno estéreo (por ejemplo, con la mezcla descendente de la trama anterior) del espectro de MDCT blanqueado izquierdo y derecho.

[0212] Más en general:

- En algunas de las realizaciones, la unidad de codificación 120 de la Fig. 1a - Fig. 1e se puede configurar, por ejemplo,
35 para generar la señal de audio procesada, de tal manera que al menos dicha banda espectral del primer canal de la señal de audio procesada sea dicha banda espectral de dicha señal media, y de tal manera que al menos dicha banda espectral del segundo canal de la señal de audio procesada sea dicha banda espectral de dicha señal lateral. Para obtener la señal de audio codificada, la unidad de codificación 120 se puede configurar, por ejemplo, para codificar dicha banda espectral de dicha señal lateral mediante la determinación de un factor de corrección para dicha banda
40 espectral de dicha señal lateral. La unidad de codificación 120 se puede configurar, por ejemplo, para determinar dicho factor de corrección para dicha banda espectral de dicha señal lateral dependiendo de una residual y dependiendo de una banda espectral de una señal media anterior, que corresponde a dicha banda espectral de dicha señal media, donde la señal media anterior precede a dicha señal media en el tiempo. Además, la unidad de codificación 120 se puede configurar, por ej., para determinar la residual dependiendo de dicha banda espectral de dicha señal lateral, y
45 dependiendo de dicha banda espectral de dicha señal media.

[0213] Según algunas de las realizaciones, la unidad de codificación 120 se puede configurar, por ejemplo, para determinar dicho factor de corrección para dicha banda espectral de dicha señal lateral según la fórmula

$$50 \quad correction_factor_{fb} = ERes_{fb} / (EprevDmx_{fb} + \varepsilon)$$

donde $correction_factor_{fb} = ERes_{fb} / (EprevDmx_{fb} + \varepsilon)$ indica dicho factor de corrección para dicha banda espectral de dicha señal lateral, donde $correction_factor_{fb} = ERes_{fb} / (EprevDmx_{fb} + \varepsilon)$ indica una energía residual dependiendo de una energía de una banda espectral de dicha residual, que corresponde a dicha banda espectral de
5 dicha señal media, donde $correction_factor_{fb} = ERes_{fb} / (EprevDmx_{fb} + \varepsilon)$ indica una energía anterior dependiendo de una energía de la banda espectral de la señal media anterior, y donde $\varepsilon = 0$, o donde $0,1 > \varepsilon > 0$.

[0214] En algunas de las realizaciones, dicha residual se puede definir, por ejemplo, según

$$10 \quad Res_R = S_R - a_R Dmx_R,$$

donde Res_R es dicha residual, donde S_R es dicha señal lateral, donde a_R es un coeficiente (por ejemplo, real) (por ejemplo, un coeficiente de predicción), donde Dmx_R es dicha señal media, donde la unidad de codificación (120) se configura para determinar dicha energía residual según

$$15 \quad ERes_{fb} = \sum_{fb} Res_R^2.$$

[0215] Según algunas de las realizaciones, dicha residual se define según

$$20 \quad Res_R = S_R - a_R Dmx_R - a_I Dmx_I,$$

donde Res_R es dicha residual, donde S_R es dicha señal lateral, donde a_R es una parte real de un coeficiente (de predicción) complejo, y donde a_I es una parte imaginaria de dicho coeficiente (de predicción) complejo, donde Dmx_R es dicha señal media, donde Dmx_I es otra señal media que depende del primer canal de la señal de audio normalizada
25 y que depende del segundo canal de la señal de audio normalizada, donde otra residual de otra señal lateral S_I que depende del primer canal de la señal de audio normalizada y que depende del segundo canal de la señal de audio normalizada se define según

$$30 \quad Res_I = S_I - a_R Dmx_R - a_I Dmx_I,$$

donde la unidad de codificación 120 se puede configurar, por ejemplo, para determinar dicha energía residual según

$$ERes_{fb} = \sum_{fb} Res_R^2 + \sum_{fb} Res_I^2$$

donde la unidad de codificación 120 se puede configurar, por ejemplo, para determinar la energía anterior dependiendo
35 de la energía de la banda espectral de dicha residual, que corresponde a dicha banda espectral de dicha señal media, y dependiendo de una energía de una banda espectral de dicha residual adicional, que corresponde a dicha banda espectral de dicha señal media.

[0216] En algunas de las realizaciones, la unidad de decodificación 210 de la Fig. 2a - Fig. 2e se puede
40 configurar, por ejemplo, para determinar, por cada banda espectral de dicha pluralidad de bandas espectrales, si dicha banda espectral del primer canal de la señal de audio codificada y dicha banda espectral del segundo canal de la señal de audio codificada se codificó empleando codificación doble-mono o utilizando la codificación medio-lateral. Además, la unidad de decodificación 210 se puede configurar, por ejemplo, para obtener dicha banda espectral del segundo canal de la señal de audio codificada mediante la reconstrucción de dicha banda espectral del segundo canal. Si se
45 ha utilizado la codificación medio-lateral, dicha banda espectral del primer canal de la señal de audio codificada es una banda espectral de una señal media, y dicha banda espectral del segundo canal de la señal de audio codificada es una banda espectral de una señal lateral. Además, si se ha utilizado la codificación medio-lateral, la unidad de decodificación 210 se puede configurar, por ejemplo, para reconstruir dicha banda espectral de la señal lateral dependiendo de un factor de corrección para dicha banda espectral de la señal lateral y dependiendo de una banda
50 espectral de una señal media anterior, que corresponde a dicha banda espectral de dicha señal media, donde la señal media anterior precede a dicha señal media en el tiempo.

[0217] Según algunas de las realizaciones, si se ha utilizado la codificación medio-lateral, la unidad de

decodificación 210 se puede configurar, por ejemplo, para reconstruir dicha banda espectral de la señal lateral, mediante la reconstrucción de valores espectrales de dicha banda espectral de la señal lateral según

$$S_i = N_i + facDmx_{fb} \cdot prevDmx_i$$

5

donde S_i indica los valores espectrales de dicha banda espectral de la señal lateral, donde $prevDmx_i$ indica los valores espectrales de la banda espectral de dicha señal media anterior, donde N_i indica los valores espectrales de un espectro con relleno de ruido, donde $facDmx_{fb}$ se define según

10

$$facDmx_{fb} = \sqrt{correction_factor_{fb} - \frac{EN_{fb}}{(E_{prevDmx_{fb}} + \varepsilon)}}$$

donde $correction_factor_{fb}$ es dicho factor de corrección para dicha banda espectral de la señal lateral, donde EN_{fb} , es una energía del espectro con relleno de ruido, donde $E_{prevDmx_{fb}}$ es una energía de dicha banda espectral de dicha señal media anterior, y donde $\varepsilon = 0$, o donde $0,1 > \varepsilon > 0$.

15

[0218] En algunas de las realizaciones, se puede derivar una residual, por ejemplo, de un algoritmo de predicción compleja estéreo en el codificador, mientras que no haya una predicción estéreo (real o compleja) del lado del decodificador.

20

[0219] Según algunas de las realizaciones, se puede recurrir al escalado corrector de la energía del espectro del lado del codificador, por ejemplo, para compensar la falta de procesamiento de predicción inversa del lado del decodificador.

25

[0220] Aunque se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del procedimiento correspondiente, en el cual un bloque o dispositivo corresponde a una etapa del procedimiento o a una característica de una etapa del procedimiento. De manera análoga, los aspectos descritos en el contexto de una etapa del procedimiento también representan una descripción de un bloque o elemento correspondiente o de una característica de un aparato correspondiente. Algunas o todas las etapas del procedimiento se pueden ejecutar por (o utilizando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, una cualquiera o más de las etapas más importantes del procedimiento se pueden ejecutar por ese tipo de aparato.

30

[0221] Dependiendo de ciertos requisitos de implementación, las formas se pueden implementar en hardware o en software o al menos parcialmente en software, o al menos parcialmente en hardware o al menos parcialmente en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo, un disco blando, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales de control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema de computación programable de tal manera que se ejecute el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por un ordenador.

40

[0222] Algunas realizaciones según la invención comprenden un soporte no transitorio de datos que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se ejecute uno de los procedimientos descritos en esta invención.

45

[0223] En general, las realizaciones de la presente invención se pueden implementar en forma de producto programa informático con un código de programa, donde el código de programa cumple la función de ejecutar uno de los procedimientos al ejecutarse el programa informático en un ordenador. El código de programa se puede almacenar, por ejemplo, en un soporte legible por una máquina.

50

[0224] Otras realizaciones comprenden el programa informático para ejecutar uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por una máquina.

55

[0225] En otras palabras, una realización del procedimiento de la invención consiste, por lo tanto, en un programa informático que consta de un código de programa para ejecutar uno de los procedimientos descritos en esta invención al ejecutar el programa informático en un ordenador.

[0226] Otra realización de los procedimientos de la invención consiste, por lo tanto, en un soporte de datos (o medio de almacenamiento digital, o medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para ejecutar uno de los procedimientos descritos en esta invención. El soporte de datos, el medio de almacenamiento digital o el medio grabado son por lo general tangibles y/o no transitorios.

[0227] Otra realización del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para ejecutar uno de los procedimientos descritos en la invención. El flujo de datos o la secuencia de señales se puede configurar, por ejemplo, para ser transferida a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

[0228] Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los procedimientos descritos en esta invención.

[0229] Otra realización comprende un ordenador en la que se ha instalado el programa informático para ejecutar uno de los procedimientos descritos en esta invención.

[0230] Otra de las realizaciones según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, por vía electrónica u óptica) un programa informático para ejecutar uno de los procedimientos descritos en esta invención a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir un programa informático al receptor.

[0231] En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programables en el campo) para ejecutar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas realizaciones, una matriz de puertas programables en el campo puede cooperar con un microprocesador para ejecutar uno de los procedimientos descritos en esta invención. Por lo general, los procedimientos son ejecutados preferentemente por cualquier aparato de hardware.

[0232] El aparato descrito en esta invención se puede implementar empleando un aparato de hardware o utilizando un ordenador, o utilizando una combinación de aparato de hardware y un ordenador.

[0233] Los procedimientos descritos en esta invención se pueden poner en práctica empleando un aparato de hardware o utilizando un ordenador, o utilizando una combinación de aparato de hardware y un ordenador.

[0234] Las realizaciones antes descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en esta invención deben resultar obvios para otras personas expertas en la materia. Por lo tanto, sólo se pretende quedar limitados por el alcance de las siguientes reivindicaciones de patente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de esta invención.

Bibliografía

[0235]

[1] J. Herre, E. Eberlein y K. Brandenburg, "Combined Stereo Coding," in 93rd AES Convention, San Francisco, 1992.

[2] J. D. Johnston y A. J. Ferreira, "Sum-difference stereo transform coding," en Proc. ICASSP, 1992.

[3] ISO/IEC 11172-3, Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Parte 3: Audio, 1993.

[4] ISO/IEC 13818-7, Information technology - Generic coding of moving pictures and associated audio information - Parte 7: Advanced Audio Coding (AAC), 2003.

[5] J.-M. Valin, G. Maxwell, T. B. Terriberry y K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec," en Proc. AES 135th Convention, New York, 2013.

[6a] 3GPP TS 26.445, Codec for Enhanced Voice Services (EVS); Detailed algorithmic description, V 12.5.0, diciembre de 2015.

[6b] 3GPP TS 26.445, Codec for Enhanced Voice Services (EVS); Detailed algorithmic description, V 13.3.0, septiembre de 2016.

- [7] H. Purnhagen, P. Carlsson, L. Villemoes, J. Robilliard, M. Neusinger, C. Helmrich, J. Hilpert, N. Rettelbach, S. Disch y B. Edler, "Audio encoder, audio decoder and related methods for processing multi-channel audio signals using complex prediction". Patente de EE. UU. 8.655.670 B2, 18 de febrero de 2014.
- 5 [8] G. Markovic, F. Guillaume, N. Rettelbach, C. Helmrich y B. Schubert, "Linear prediction based coding scheme using spectral domain noise shaping". Patente europea 2676266 B1, 14 de febrero de 2011.
- [9] S. Disch, F. Nagel, R. Geiger, B. N. Thoshkahna, K. Schmidt, S. Bayer, C. Neukam, B. Edler y C. Helmrich, "Audio Encoder, Audio Decoder and Related Methods Using Two-Channel Processing Within an Intelligent Gap Filling Framework". Patente internacional PCT/EP2014/065106, 15 07 2014.
- 10 [10] C. Helmrich, P. Carlsson, S. Disch, B. Edler, J. Hilpert, M. Neusinger, H. Purnhagen, N. Rettelbach, J. Robilliard y L. Villemoes, "Efficient Transform Coding Of Two-channel Audio Signals By Means Of Complex-valued Stereo Prediction," in Acoustics, Speech y Signal Processing (ICASSP), 2011 IEEE International Conference on, Praga, 2011.
- [11] C. R. Helmrich, A. Niedermeier, S. Bayer y B. Edler, "Low-complexity semi-parametric joint-stereo audio transform coding," en Signal Processing Conference (EUSIPCO), 2015 23rd European, 2015.
- 15 [12] H. Malvar, "A Modulated Complex Lapped Transform and its Applications to Audio Processing" in Acoustics, Speech, and Signal Processing (ICASSP), 1999. Proceedings., 1999 IEEE International Conference on, Fenix, AZ, 1999.
- [13] B. Edler y G. Schuller, "Audio coding using a psychoacoustic pre- y post-filter," Acoustics, Speech, y Signal Processing, 2000. ICASSP '00.
- 20

REIVINDICACIONES

1. Un aparato para codificar un primer canal y un segundo canal de una señal de entrada de audio que comprende dos o más canales para obtener una señal de audio codificada, donde el aparato comprende:

- 5 un normalizador (110) configurado para determinar un valor de normalización para la señal de entrada de audio dependiendo del primer canal de la señal de entrada de audio y dependiendo del segundo canal de la señal de entrada de audio, donde el normalizador (110) está configurado para determinar un primer canal y un segundo canal de una señal de audio normalizada mediante la modificación, dependiendo del valor de normalización, de al menos uno del primer canal y el segundo canal de la señal de entrada de audio,
- 10 una unidad de codificación (120) configurada para elegir un modo entre un modo de codificación medio-lateral total y un modo de codificación doble-mono total y un modo de codificación por banda dependiendo de una pluralidad de bandas espectrales de un primer canal de la señal de audio normalizada y dependiendo de una pluralidad de bandas espectrales de un segundo canal de la señal de audio normalizada,
- 15 donde la unidad de codificación (120) está configurada, si se elige el modo de codificación por banda, para generar una señal de audio procesada que tiene un primer canal y un segundo canal, de tal forma que una o más bandas espectrales del primer canal de la señal de audio procesada son una o más bandas espectrales del primer canal de la señal de audio normalizada, de tal forma que una o más bandas espectrales del segundo canal de la señal de audio procesada son una o más bandas espectrales del segundo canal de la señal de audio normalizada, de tal forma que al menos una banda espectral del primer canal de la señal de audio procesada es una banda espectral de una señal media dependiendo de una banda espectral del primer canal de la señal de audio normalizada y dependiendo de una banda espectral del segundo canal de la señal de audio normalizada, y de tal forma que al menos una banda espectral del segundo canal de la señal de audio procesada es una banda espectral de una señal lateral dependiendo de una banda espectral del primer canal de la señal de audio normalizada y dependiendo de una banda espectral del segundo canal de la señal de audio normalizada, donde la unidad de codificación (120) está configurada para codificar la señal de audio procesada para obtener la señal de audio codificada,
- 20 donde la unidad de codificación (120) está configurada para generar la señal de audio procesada dependiendo de una decisión media-lateral por banda para cada banda espectral de una pluralidad de bandas espectrales de la señal de audio normalizada, donde la decisión media-lateral por banda depende de una estimación de bits por banda para un modo izquierdo/derecho para cada banda espectral de la pluralidad de bandas espectrales de la señal de audio normalizada y depende de una estimación de bits por banda para un modo medio/lateral para cada banda espectral de la pluralidad de bandas espectrales, donde se elige el modo con menos bits para la banda espectral,
- 30 donde la unidad de codificación (120) está configurada, si se elige el modo de codificación medio-lateral total, para generar una señal media desde el primer canal y desde el segundo canal de la señal de audio normalizada como un primer canal de señal media-lateral, para generar una señal lateral desde el primer canal y desde el segundo canal de la señal de audio normalizada como un segundo canal de la señal media-lateral, y para codificar la señal media-lateral para obtener la señal de audio codificada,
- 35 donde la unidad de codificación (120) está configurada, si se elige el modo de codificación doble-mono total, para codificar la señal de audio normalizada para obtener la señal de audio codificada,
- 40 donde la señal de entrada de audio se representa en un dominio espectral, donde el normalizador (110) está configurado para determinar el valor de normalización para la señal de entrada de audio dependiendo de una pluralidad de bandas espectrales del primer canal de la señal de entrada de audio y dependiendo de una pluralidad de bandas espectrales del segundo canal de la señal de entrada de audio, y
- 45 donde el normalizador (110) está configurado para determinar la señal de audio normalizada modificando, dependiendo del valor de normalización, la pluralidad de bandas espectrales de al menos uno del primer canal y el segundo canal de la señal de entrada de audio, donde el normalizador (110) está configurado para determinar el valor de normalización basado en las fórmulas:

50

$$NRG_L = \sqrt{\sum MDCT_{L,k}^2}$$

$$NRG_R = \sqrt{\sum MDCT_{R,k}^2}$$

$$ILD = \frac{NRG_L}{NRG_L + NRG_R}$$

donde $MDCT_{L,k}$ es un k-ésimo coeficiente de un espectro MDCT del primer canal de la señal de entrada de audio, y

$MDCT_{R,k}$ es el k-ésimo coeficiente del espectro MDCT del segundo canal de la señal de entrada de audio, y donde el normalizador (110) está configurado para determinar el valor de normalización mediante la cuantificación de ILD .

2. Un aparato según la reivindicación 1,

donde la unidad de codificación (120) está configurada, si se opta por el modo de codificación por banda, para decidir, por cada banda espectral de una pluralidad de bandas espectrales de la señal de audio procesada, si se emplea la codificación medio-lateral o si se emplea la codificación doble-mono, donde, si se emplea la codificación medio-lateral para dicha banda espectral, la unidad de codificación (120) está configurada para generar dicha banda espectral del primer canal de la señal de audio procesada como banda espectral de una señal media sobre la base de dicha banda espectral del primer canal de la señal de audio normalizada y sobre la base de dicha banda espectral del segundo canal de la señal de audio normalizada, y la unidad de codificación (120) está configurada para generar dicha banda espectral del segundo canal de la señal de audio procesada como banda espectral de una señal lateral sobre la base de dicha banda espectral del primer canal de la señal de audio normalizada y sobre la base de dicha banda espectral del segundo canal de la señal de audio normalizada, y

donde, si se emplea la codificación doble-mono para dicha banda espectral,

la unidad de codificación (120) está configurada para usar dicha banda espectral del primer canal de la señal de audio normalizada como dicha banda espectral del primer canal de la señal de audio procesada, y está configurada para usar dicha banda espectral del segundo canal de la señal de audio normalizada como dicha banda espectral del segundo canal de la señal de audio procesada, o

la unidad de codificación (120) está configurada para usar dicha banda espectral del segundo canal de la señal de audio normalizada como dicha banda espectral del primer canal de la señal de audio procesada, y está configurada para usar dicha banda espectral del primer canal de la señal de audio normalizada como dicha banda espectral del segundo canal de la señal de audio procesada.

3. Un aparato según la reivindicación 1 o 2, donde la unidad de codificación (120) está configurada para elegir entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda mediante la determinación de una primera estimación que calcula un primer número de bits que se necesitan para la codificación cuando se emplea el modo de codificación medio-lateral total, mediante la determinación de una segunda estimación que calcula un segundo número de bits que se necesitan para la codificación cuando se emplea el modo de codificación doble-mono total, mediante la determinación de una tercera estimación que calcula un tercer número de bits que se necesitan para la codificación cuando se emplea el modo de codificación por banda, y mediante la elección de ese modo de codificación entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda que tenga un menor número de bits entre la primera estimación y la segunda estimación y la tercera estimación.

4. Un aparato según la reivindicación 3,

donde la unidad de codificación (120) está configurada para estimar la tercera estimación b_{BW} , que estima el tercer número de bits que se necesitan para la codificación cuando se emplea el modo de codificación por banda, según la fórmula:

$$b_{BW} = nBandas + \sum_{i=0}^{nBandas-1} \min(b_{bwLR}^i, b_{bwMS}^i) ,$$

donde $nBandas$ es un número de bandas espectrales de la señal de audio normalizada,

5 donde b_{bwMS}^i es una estimación del número de bits que se necesitan para codificar una *iésima* banda espectral de la señal media y para codificar la *iésima* banda espectral de la señal lateral, y

donde b_{bwLR}^i es una estimación del número de bits que se necesitan para codificar una *iésima* banda espectral de la primera señal y para codificar la *iésima* banda espectral de la segunda señal.

10 5. Un aparato según la reivindicación 1 o 2, donde la unidad de codificación (120) está configurada para elegir entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda mediante la determinación de una primera estimación que calcula un primer número de bits que se ahorran cuando se codifica en el modo de codificación medio-lateral total, mediante la determinación de una segunda estimación que calcula un segundo número de bits que se ahorran cuando se codifica en el modo de codificación doble-mono total, mediante la determinación de una tercera estimación que calcula un tercer número de bits que se ahorran cuando se codifica en el modo de codificación por banda, y mediante la elección de ese modo de codificación entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda que ofrece el ahorro del mayor número de bits entre la primera estimación y la segunda estimación y la tercera estimación.

20 6. Un aparato según la reivindicación 1 o 2, donde la unidad de codificación (120) está configurada para elegir entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda mediante la estimación de una primera relación señal a ruido que se produce cuando se emplea el modo de codificación medio-lateral total, mediante la estimación de una segunda relación señal a ruido que se produce cuando se emplea el modo de codificación doble-mono total, mediante la estimación de una tercera relación señal a ruido que se produce cuando se emplea el modo de codificación por banda, y mediante la elección de ese modo de codificación entre el modo de codificación medio-lateral total y el modo de codificación doble-mono total y el modo de codificación por banda con la mayor relación señal a ruido entre la primera relación señal a ruido y la segunda relación señal a ruido y la tercera relación señal a ruido.

30 7. Un aparato según una de las reivindicaciones anteriores,

donde el aparato para codificación comprende además una unidad de transformación (102) y una unidad de preprocesamiento (105),

35 donde la unidad de transformación (102) está configurada para transformar una señal de audio en el dominio del tiempo de un dominio de tiempo a un dominio de frecuencia para obtener una señal de audio transformada, donde la unidad de preprocesamiento (105) está configurada para generar el primer canal y el segundo canal de la señal de entrada de audio aplicando una operación de formación de ruido en el dominio de la frecuencia del lado del codificador sobre la señal de audio transformada.

40 8. Un aparato según la reivindicación 7, donde la unidad de preprocesamiento (105) está configurada para generar el primer canal y el segundo canal de la señal de entrada de audio aplicando una operación de conformación de ruido temporal del lado del codificador en la señal de audio transformada antes de aplicar la operación de conformación de ruido del dominio de frecuencia del lado del codificador en la señal de audio transformada.

9. Un aparato según una de las reivindicaciones anteriores, donde la unidad de codificación (120) está configurada para obtener la señal de audio codificada aplicando Relleno Inteligente de Huecos Estéreo del lado del codificador en la señal de audio normalizada o en la señal de audio procesada.

10. Un aparato según una de las reivindicaciones anteriores, donde la señal de entrada de audio es una señal de audio estéreo que comprende exactamente dos canales.

55 11. Un sistema para codificar cuatro canales de una señal de entrada de audio que comprende cuatro o más canales para obtener una señal de audio codificada, donde el sistema comprende:

un primer aparato (170) según una de las reivindicaciones 1 a 9, para codificar un primer canal y un segundo canal

de los cuatro o más canales de la señal de entrada de audio para obtener un primer canal y un segundo canal de la señal de audio codificada, y

un segundo aparato (180) según una de las reivindicaciones 1 a 9, para codificar un tercer canal y un cuarto canal de los cuatro o más canales de la señal de entrada de audio para obtener un tercer canal y un cuarto canal de la señal de audio codificada.

12. Un sistema para generar una señal de audio codificada a partir de una señal de entrada de audio y para generar una señal de audio decodificada a partir de la señal de audio codificada, que comprende:

un aparato (310) según una de las reivindicaciones 1 a 10, donde el aparato (310) según una de las reivindicaciones 1 a 10 está configurado para generar la señal de audio codificada a partir de la señal de entrada de audio, y un aparato (320) para decodificar la señal de audio codificada que comprende un primer canal y un segundo canal para obtener un primer canal y un segundo canal de una señal de audio decodificada que comprende dos o más canales, donde el aparato comprende una unidad de decodificación (210) configurada para determinar para cada banda espectral de una pluralidad de bandas espectrales, si dicha banda espectral del primer canal de la señal de audio codificada y dicha banda espectral del segundo canal de la señal de audio codificada se codificaron mediante el uso de codificación dual-mono o mediante el uso de codificación media-lateral, donde la unidad de decodificación (210) está configurada para utilizar dicha banda espectral del primer canal de la señal de audio codificada como banda espectral de un primer canal de una señal de audio intermedia y está configurada para utilizar dicha banda espectral del segundo canal de la señal de audio codificada como una banda espectral de un segundo canal de la señal de audio intermedia, si se utilizó la codificación dual-mono, donde la unidad de decodificación (210) está configurada para generar una banda espectral del primer canal de la señal de audio intermedia en base a dicha banda espectral del primer canal de la señal de audio codificada y en base a dicha banda espectral del segundo canal de la señal de audio codificada, y generar una banda espectral del segundo canal de la señal de audio intermedia en base a dicha banda espectral del primer canal de la señal de audio codificada y en base a dicha banda espectral del segundo canal de la señal de audio codificada, si se utilizó la codificación media-lateral, y donde el aparato comprende un desnormalizador (220) configurado para modificar, dependiendo de un valor de desnormalización, al menos uno del primer canal y el segundo canal de la señal de audio intermedia para obtener el primer canal y el segundo canal de la señal de audio decodificada.

13. Un procedimiento para codificar un primer canal y un segundo canal de una señal de entrada de audio que comprende dos o más canales para obtener una señal de audio codificada, donde el procedimiento comprende:

determinar un valor de normalización para la señal de entrada de audio dependiendo del primer canal de la señal de entrada de audio y dependiendo del segundo canal de la señal de entrada de audio, determinar un primer canal y un segundo canal de una señal de audio normalizada modificando, dependiendo del valor de normalización, al menos uno del primer canal y el segundo canal de la señal de entrada de audio, elegir un modo entre un modo de codificación medio-lateral total y un modo de codificación doble-mono total y un modo de codificación por banda dependiendo de una pluralidad de bandas espectrales de un primer canal de la señal de audio normalizada y dependiendo de una pluralidad de bandas espectrales de un segundo canal de la señal de audio normalizada, donde, si se elige el modo de codificación por bandas, generar una señal de audio procesada que tenga un primer canal y un segundo canal, de modo que una o más bandas espectrales del primer canal de la señal de audio procesada sean una o más bandas espectrales del primer canal de la señal de audio normalizada, de manera que una o más bandas espectrales del segundo canal de la señal de audio procesada sean una o más bandas espectrales del segundo canal de la señal de audio normalizada, de tal forma que al menos una banda espectral del primer canal de la señal de audio procesada es una banda espectral de una señal media que depende de una banda espectral del primer canal de la señal de audio normalizada y que depende de una banda espectral del segundo canal de la señal de audio normalizada, y de tal forma que al menos una banda espectral del segundo canal de la señal de audio procesada es una banda espectral de una señal lateral que depende de una banda espectral del primer canal de la señal de audio normalizada y que depende de una banda espectral del segundo canal de la señal de audio normalizada, y codificar la señal de audio procesada para obtener la señal de audio codificada, donde la generación de la señal de audio procesada se lleva a cabo dependiendo de una decisión media-lateral por banda para cada banda espectral de una pluralidad de bandas espectrales de la señal de audio normalizada, donde la decisión media-lateral por banda depende de una estimación de bits por banda para un modo izquierdo/derecho para cada banda espectral de la pluralidad de bandas espectrales de la señal de audio normalizada y depende de una estimación de bits por banda para un modo medio-lateral para cada banda espectral de la pluralidad de bandas espectrales, donde se elige el modo con menos bits para la banda espectral, donde, si se elige el modo de codificación medio-lateral total, generar una señal media desde el primer canal y desde el segundo canal de la señal de audio normalizada como un primer canal de una señal media-lateral, para generar una señal lateral a partir del primer canal y a partir del segundo canal de la señal de audio normalizada como segundo canal de la señal media-lateral, y codificar la señal media-lateral para obtener la señal de audio codificada,

donde, si se elige el modo de codificación doble-mono total, codificar la señal de audio normalizada para obtener la señal de audio codificada,

donde la señal de entrada de audio se representa en un dominio espectral,

- 5 donde la determinación del valor de normalización para la señal de entrada de audio se realiza dependiendo de una pluralidad de bandas espectrales del primer canal de la señal de entrada de audio y dependiendo de una pluralidad de bandas espectrales del segundo canal de la señal de entrada de audio, y
donde la determinación de la señal de audio normalizada se lleva a cabo modificando, dependiendo del valor de normalización, la pluralidad de bandas espectrales de al menos uno del primer canal y el segundo canal de la señal de entrada de audio,

- 10 donde la determinación del valor de normalización se lleva a cabo en base a las fórmulas:

$$NRG_L = \sqrt{\sum MDCT_{L,k}^2}$$

$$NRG_R = \sqrt{\sum MDCT_{R,k}^2}$$

$$ILD = \frac{NRG_L}{NRG_L + NRG_R}$$

15

donde $MDCT_{L,k}$ es un k-ésimo coeficiente de un espectro MDCT del primer canal de la señal de entrada de audio, y $MDCT_{R,k}$ es el k-ésimo coeficiente del espectro MDCT del segundo canal de la señal de entrada de audio, y donde la determinación del valor de normalización se lleva a cabo mediante la cuantificación de ILD .

- 20 14. Un programa informático para implementar el procedimiento de la reivindicación 13 cuando se ejecuta en un ordenador o procesador de señales.

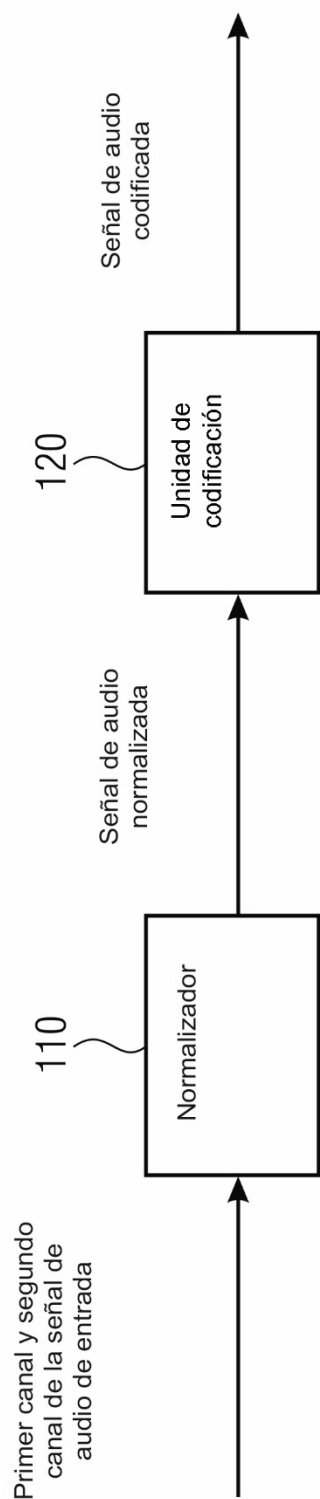


Fig. 1a

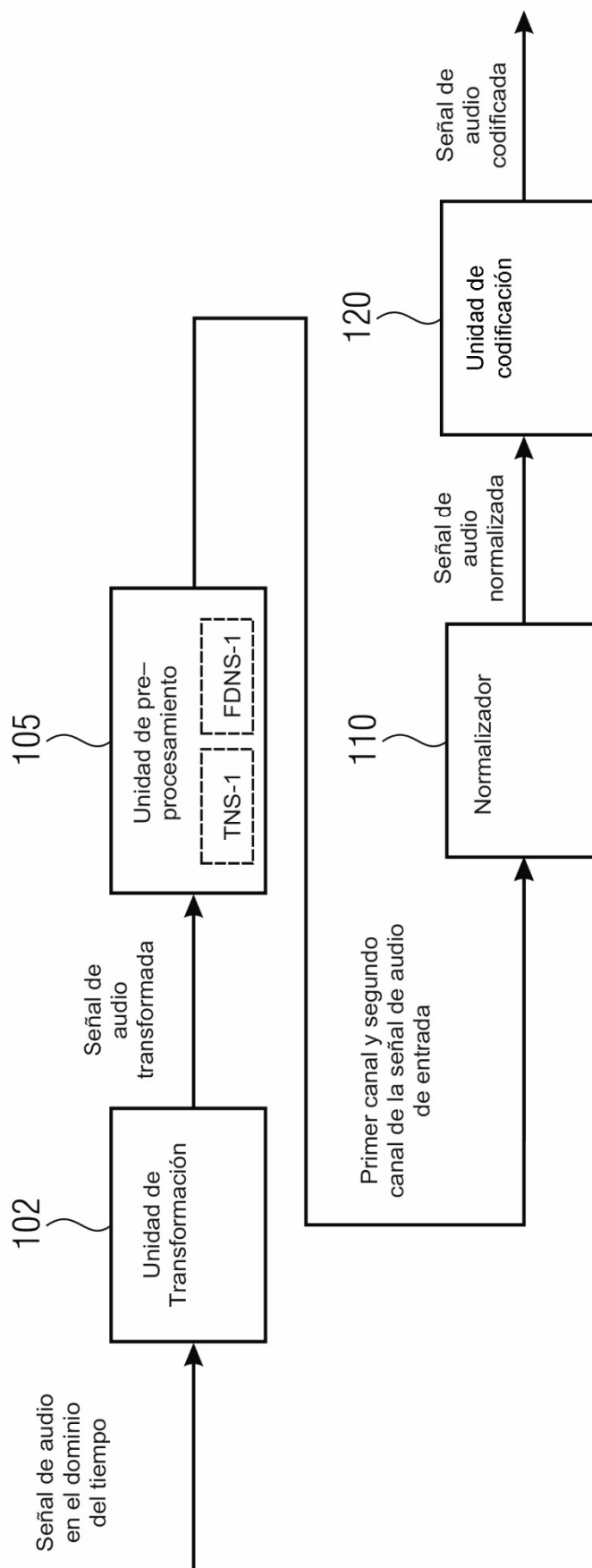
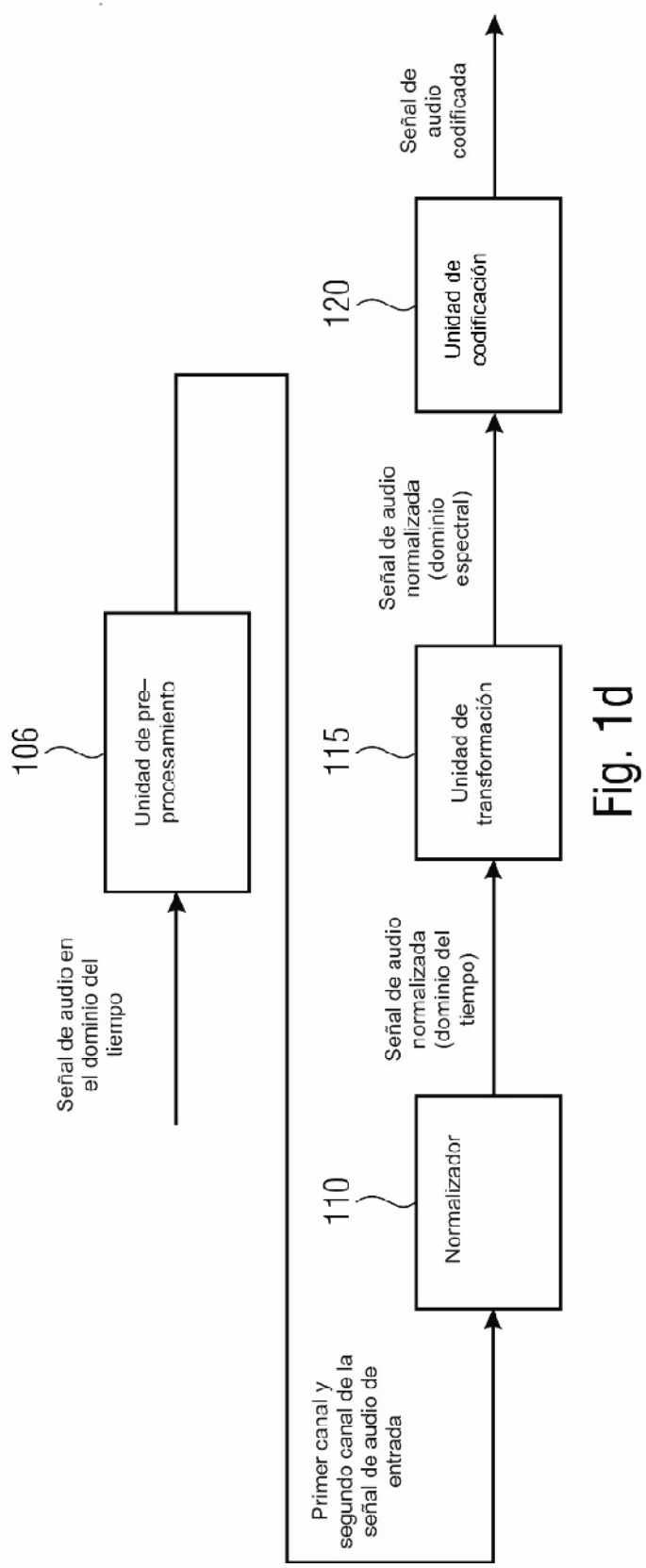
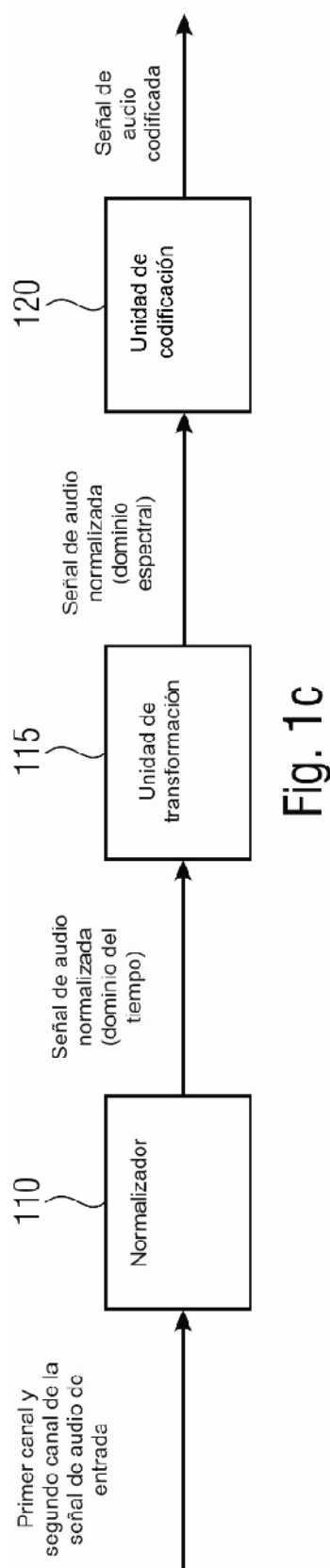


Fig. 1b



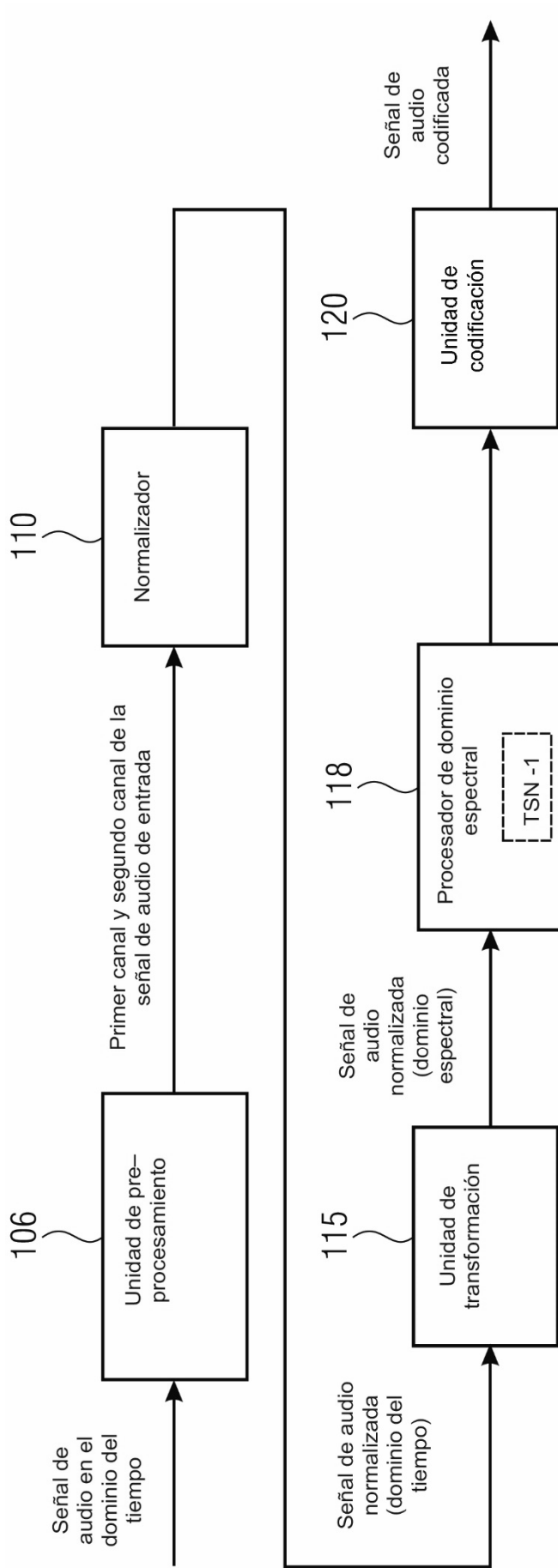


Fig. 1e

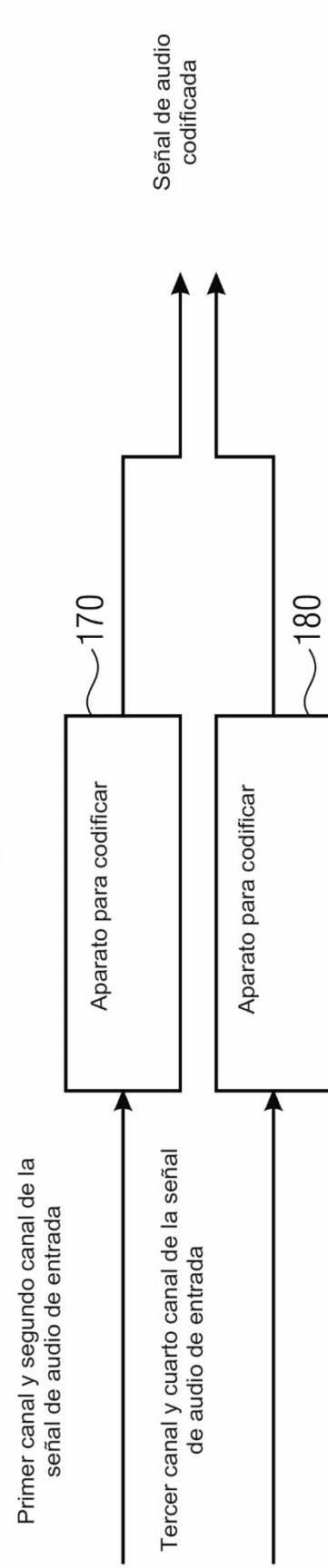


Fig. 1f

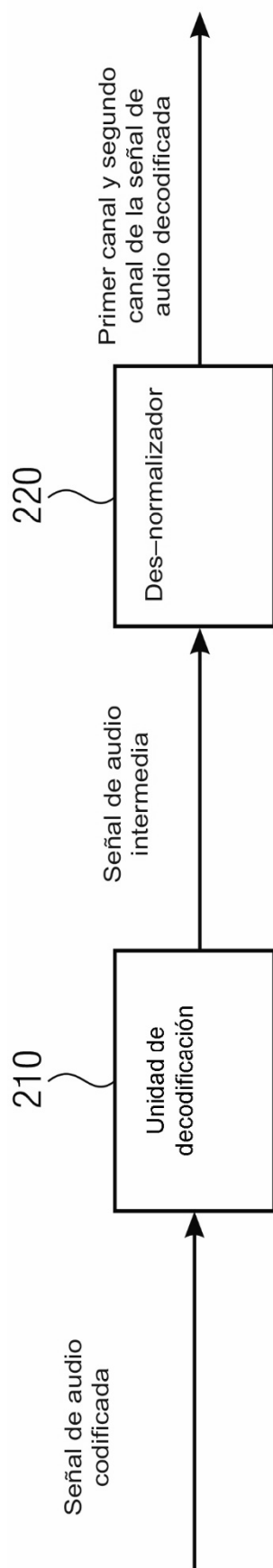


Fig. 2a

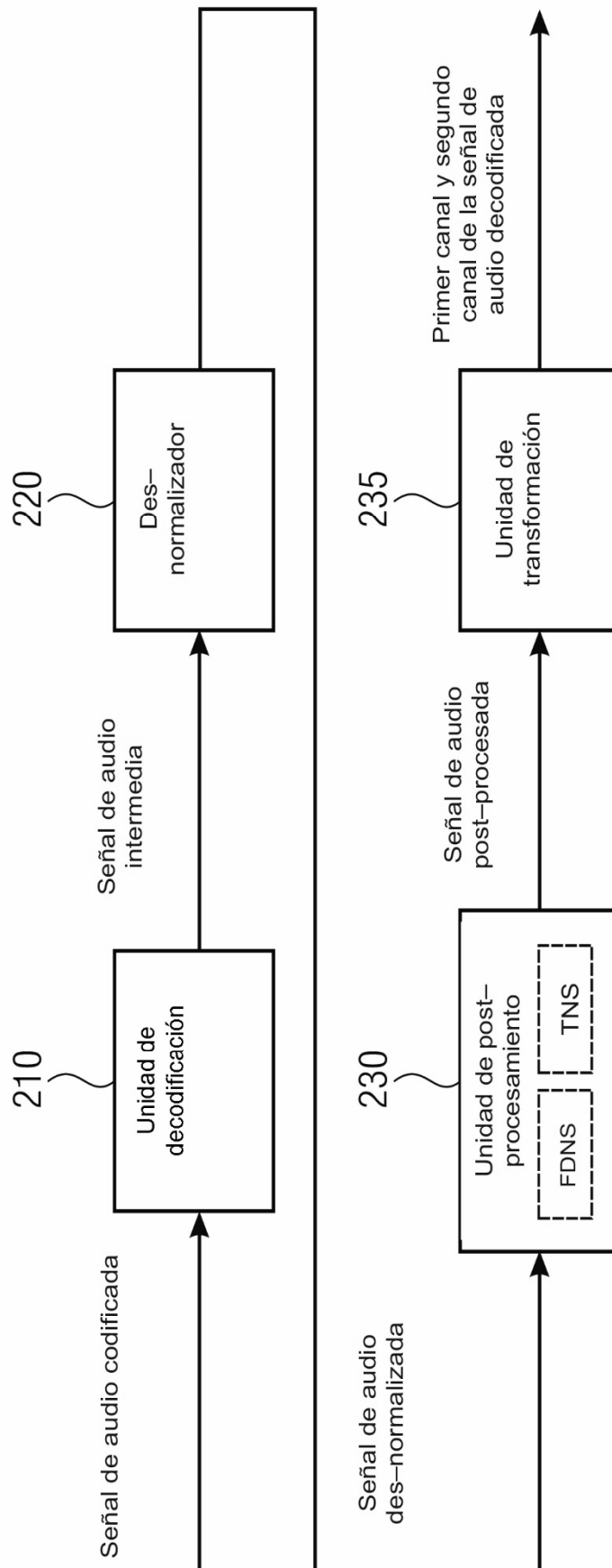
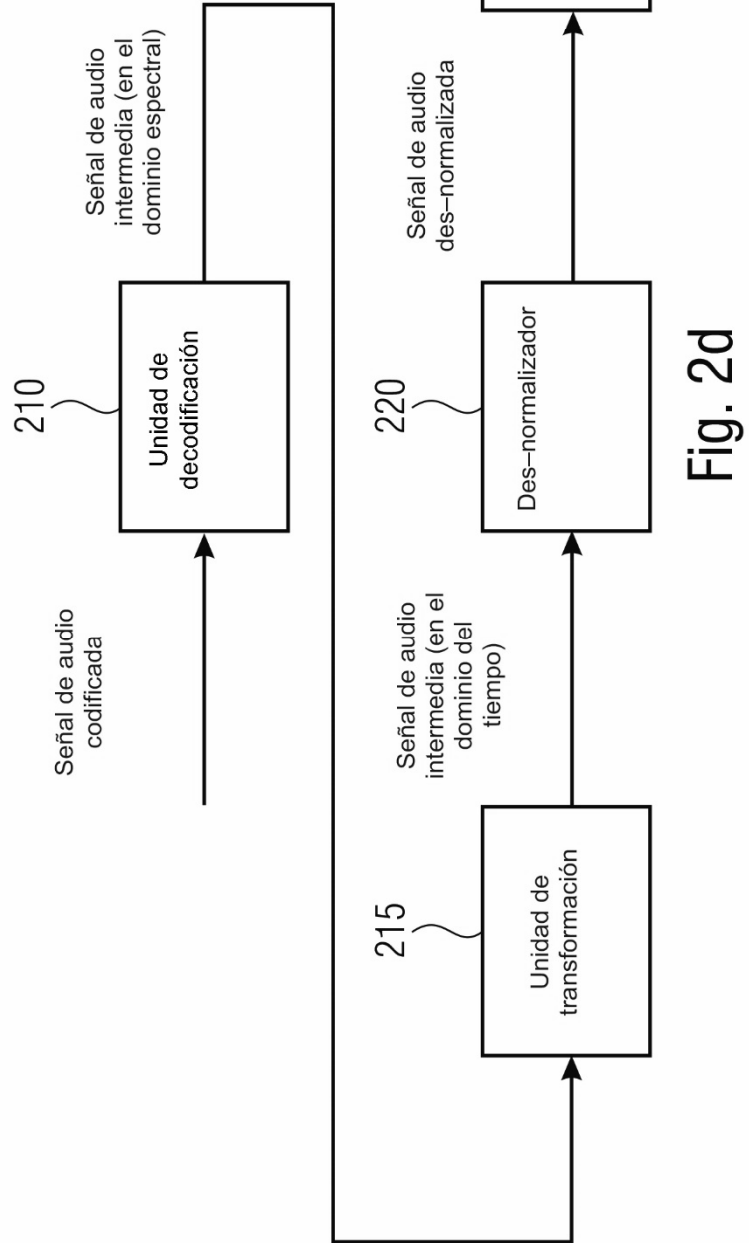
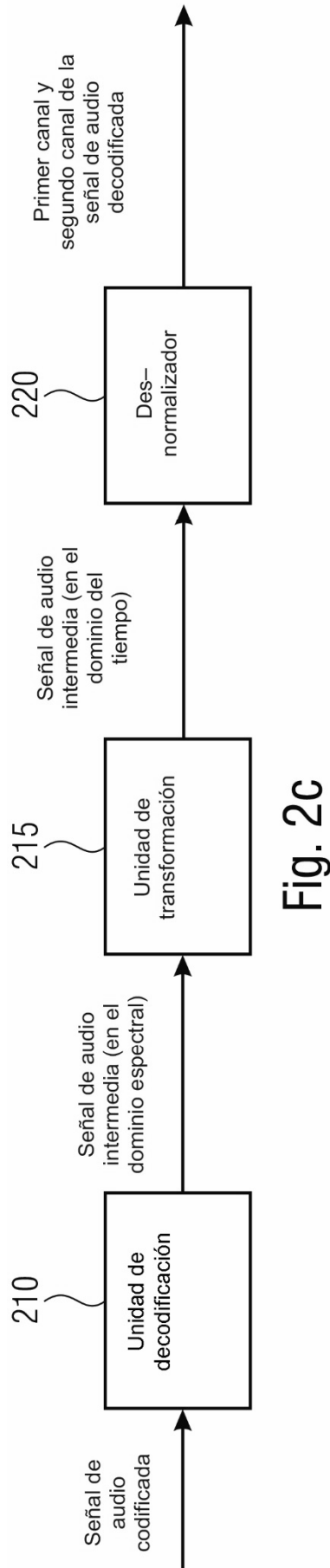


Fig. 2b



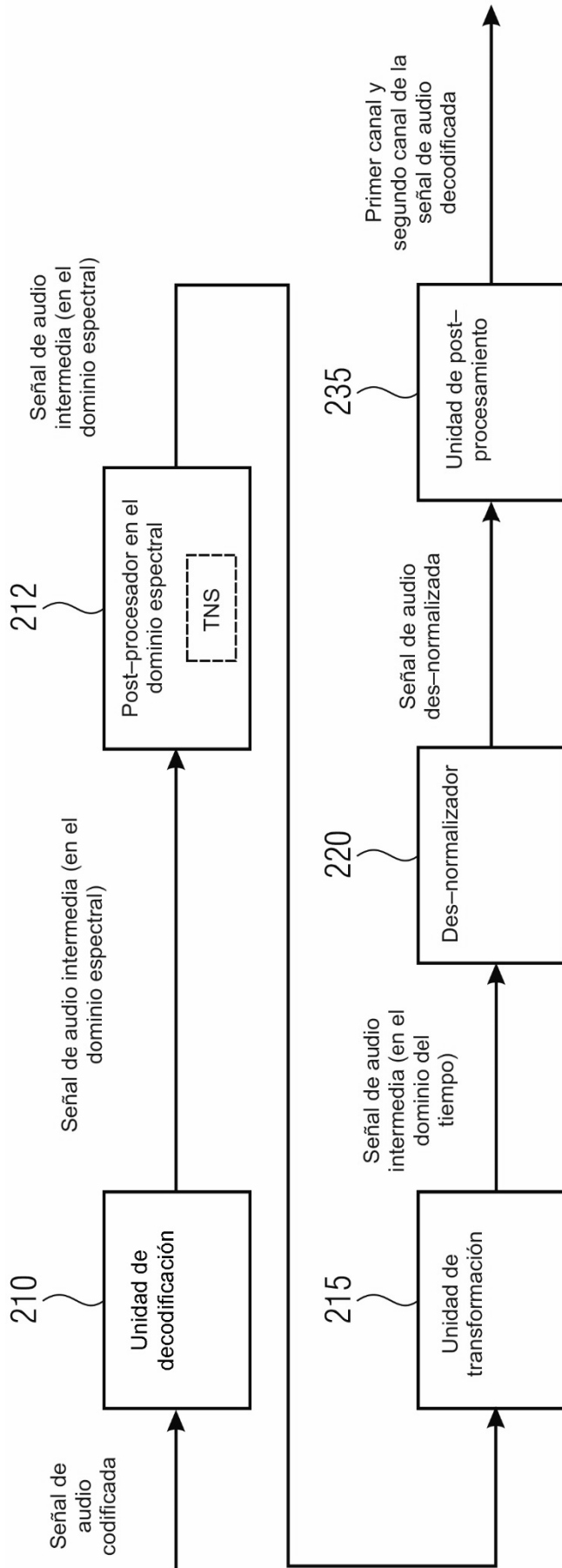


Fig. 2e

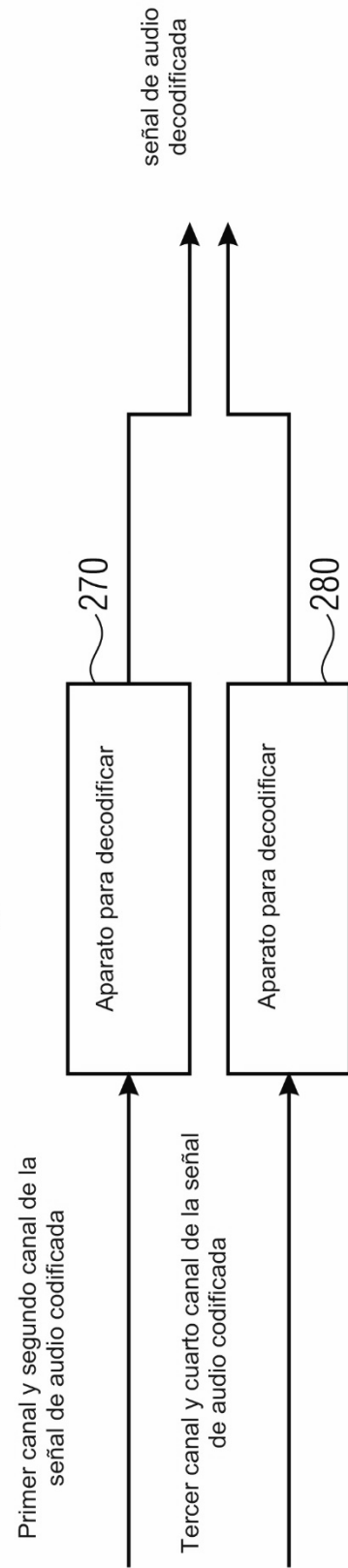


Fig. 2f

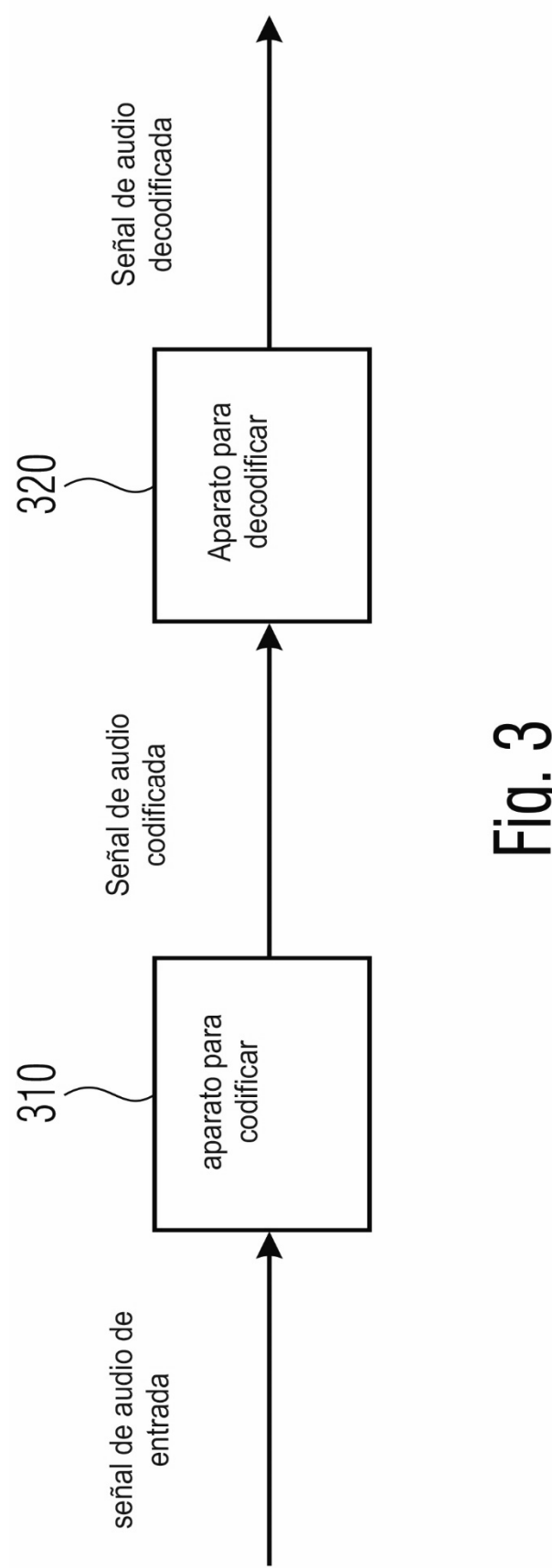


Fig. 3

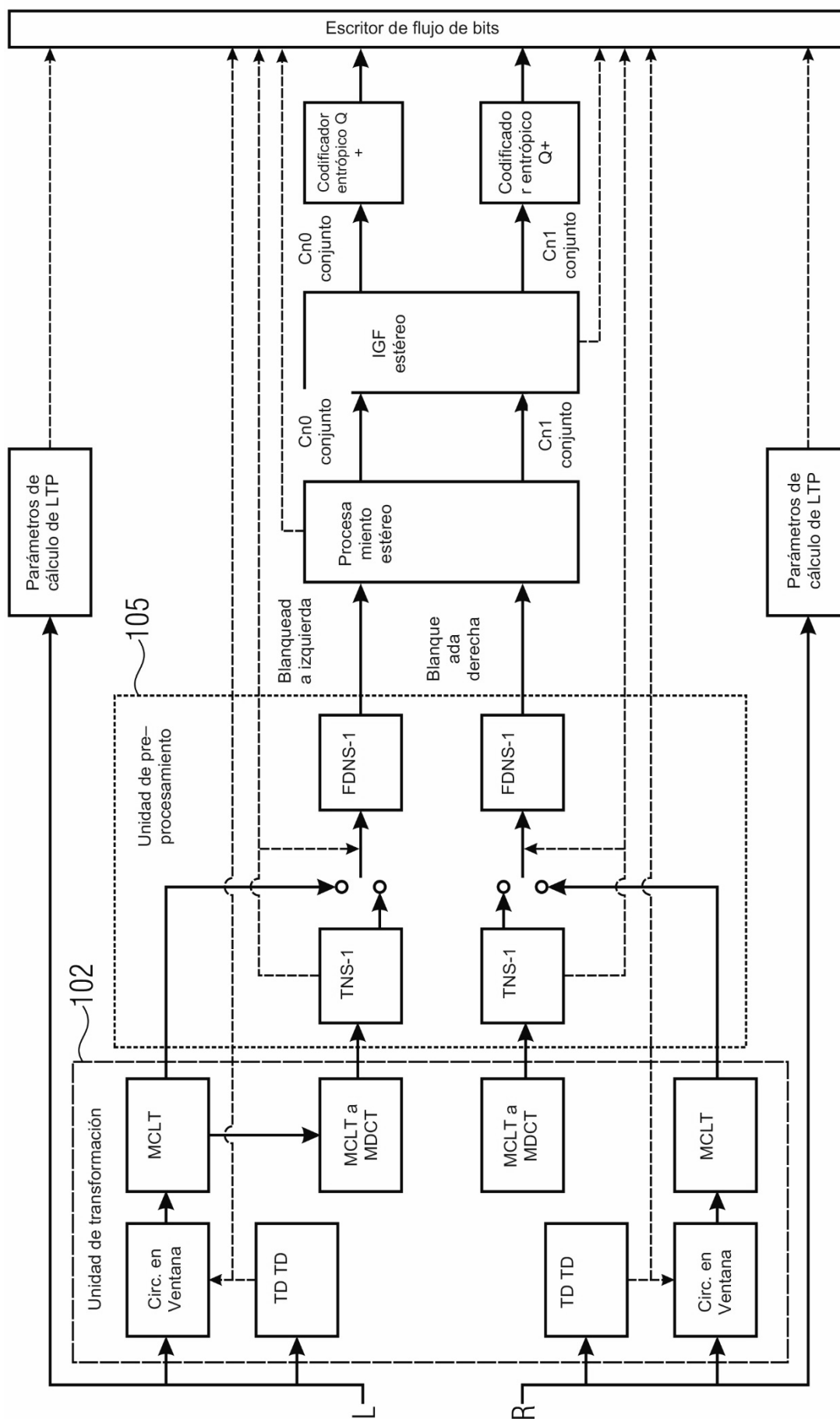


Fig. 4

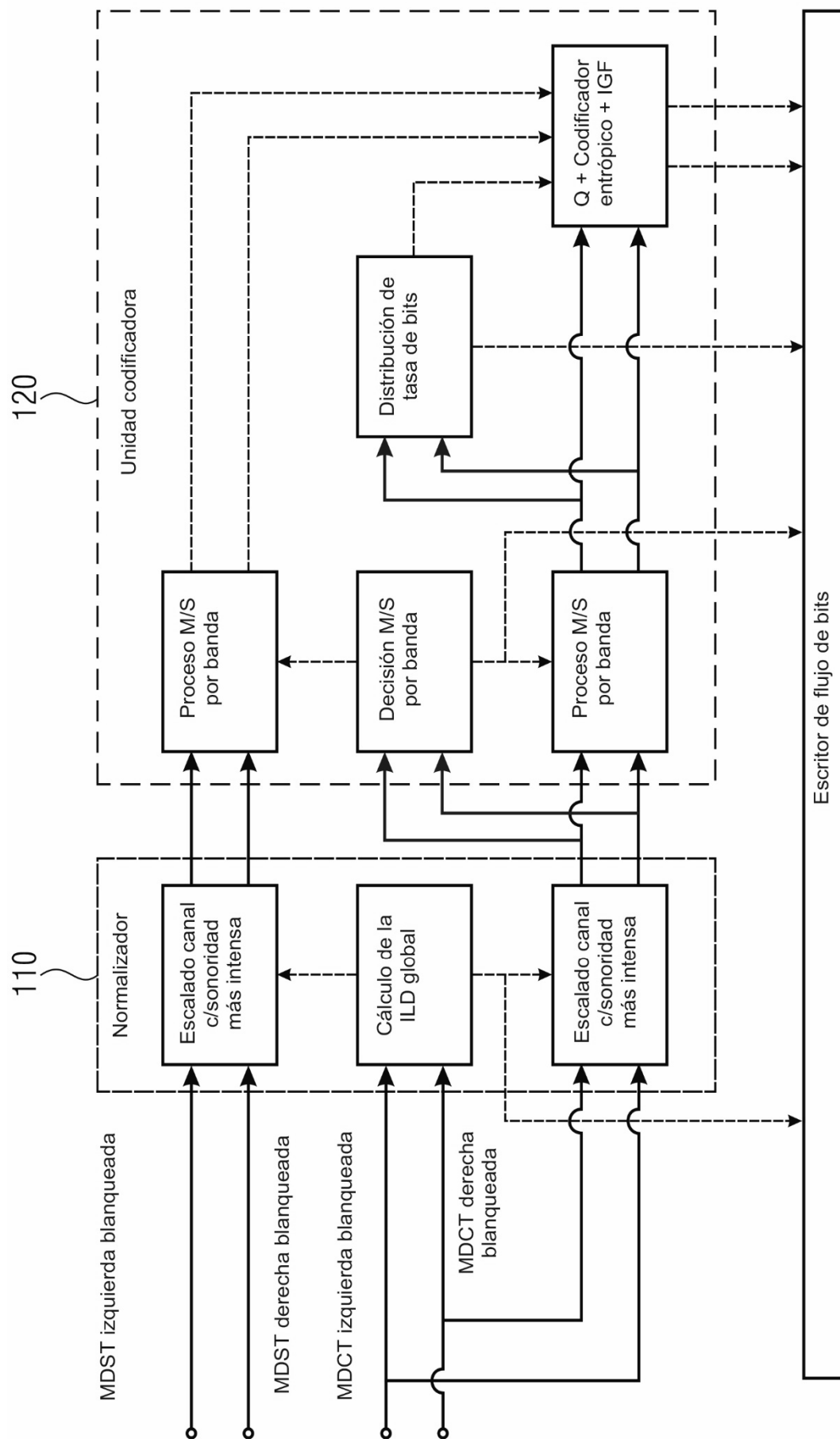


Fig. 5

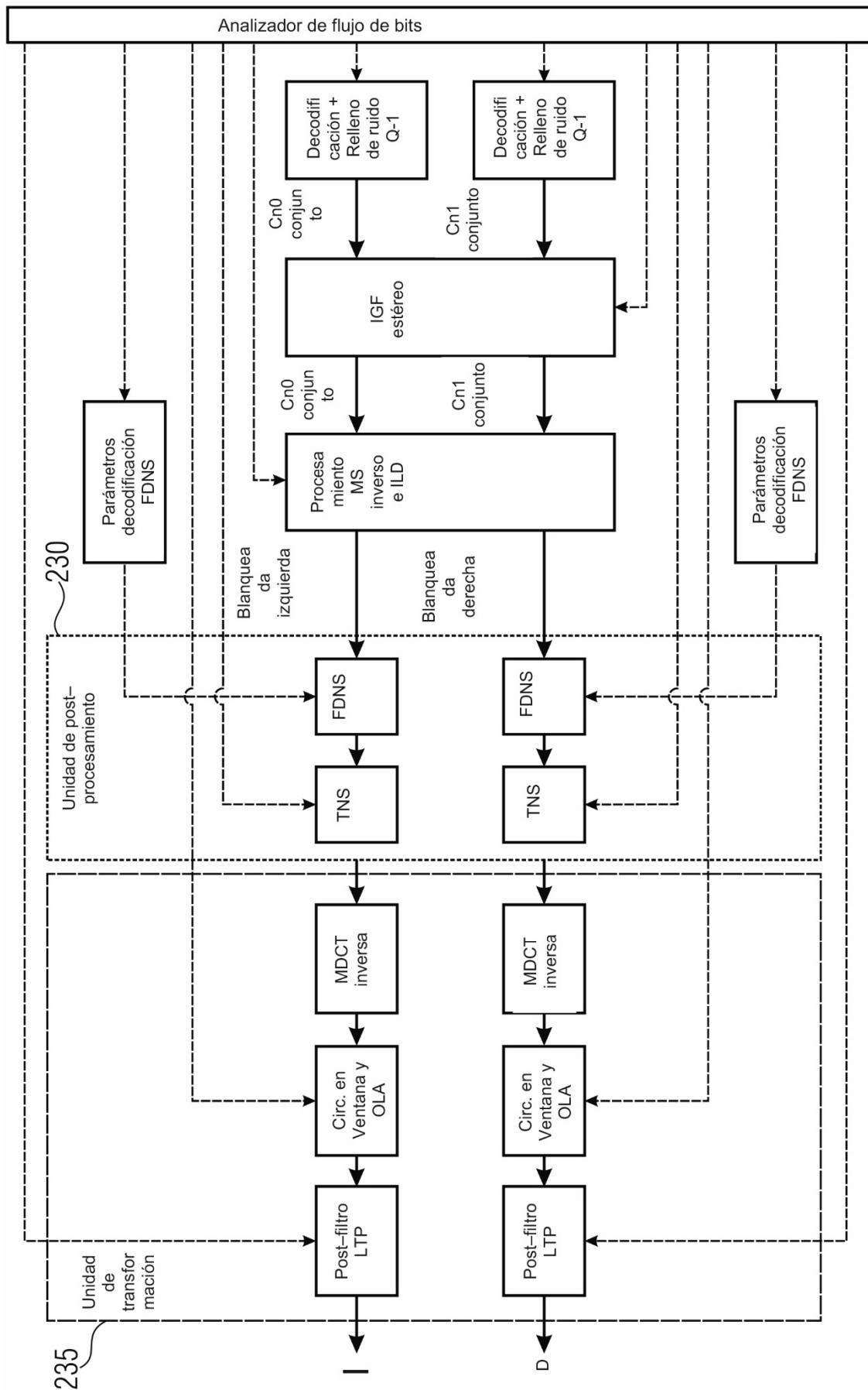


Fig. 6

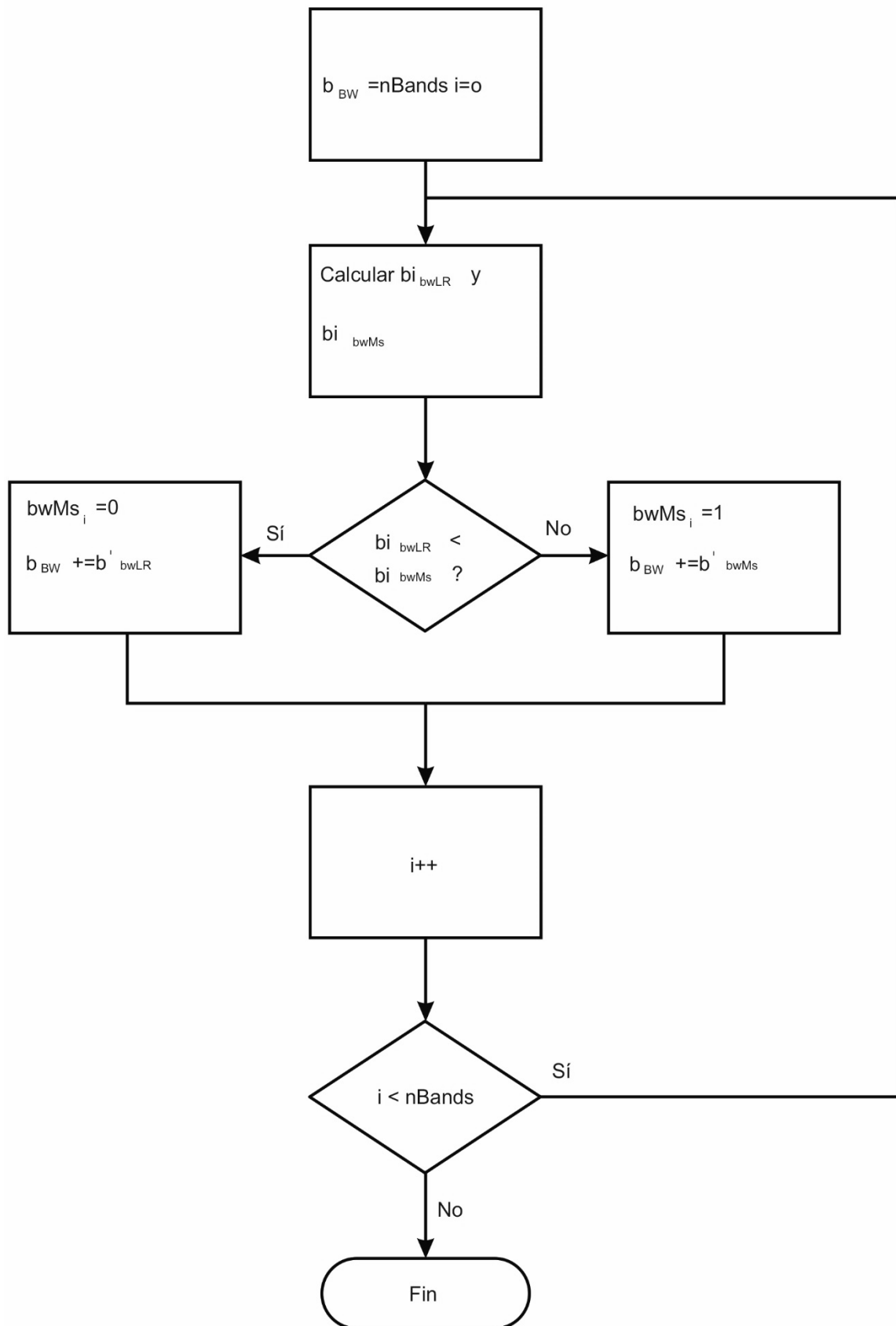


Fig. 7

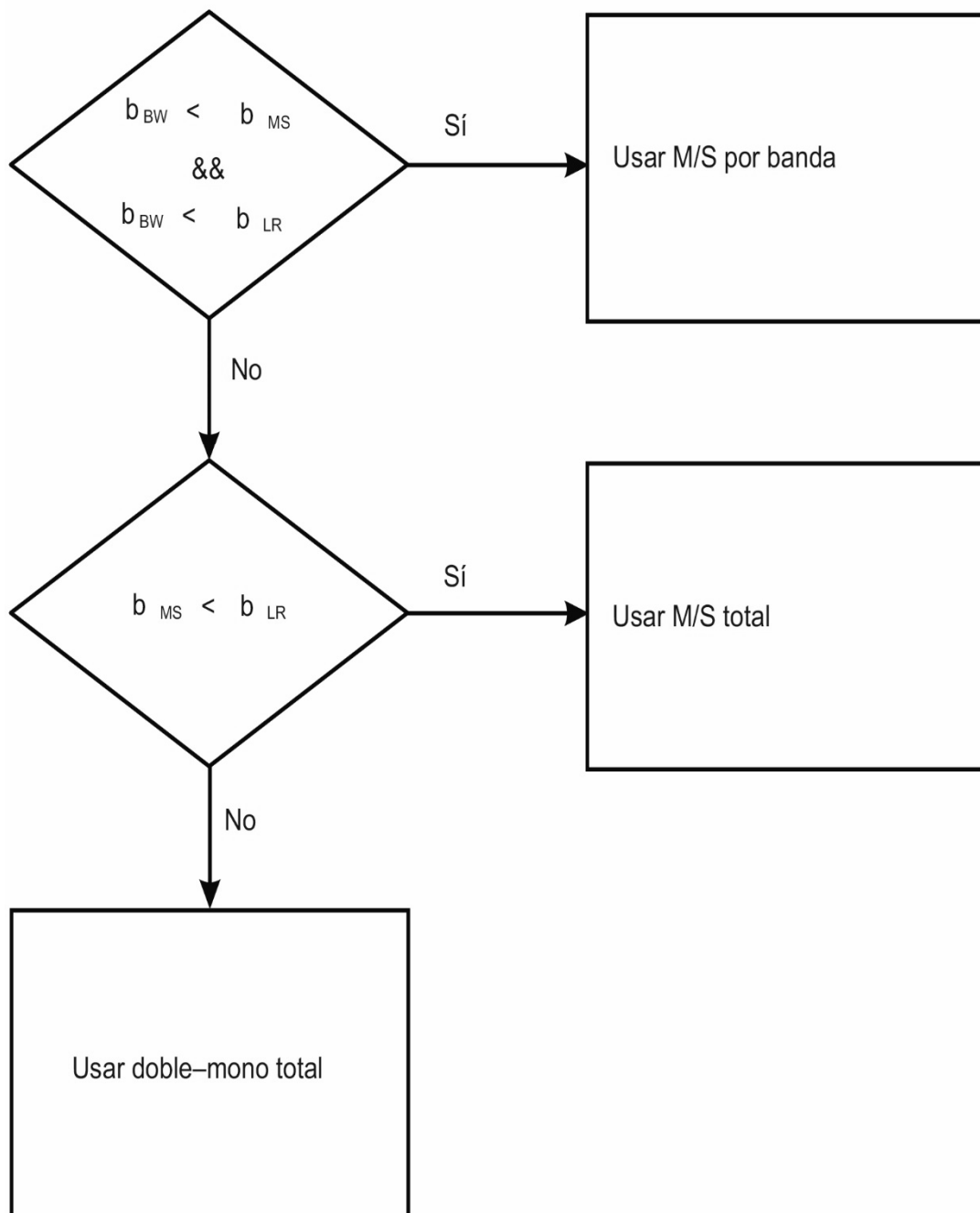


Fig. 8

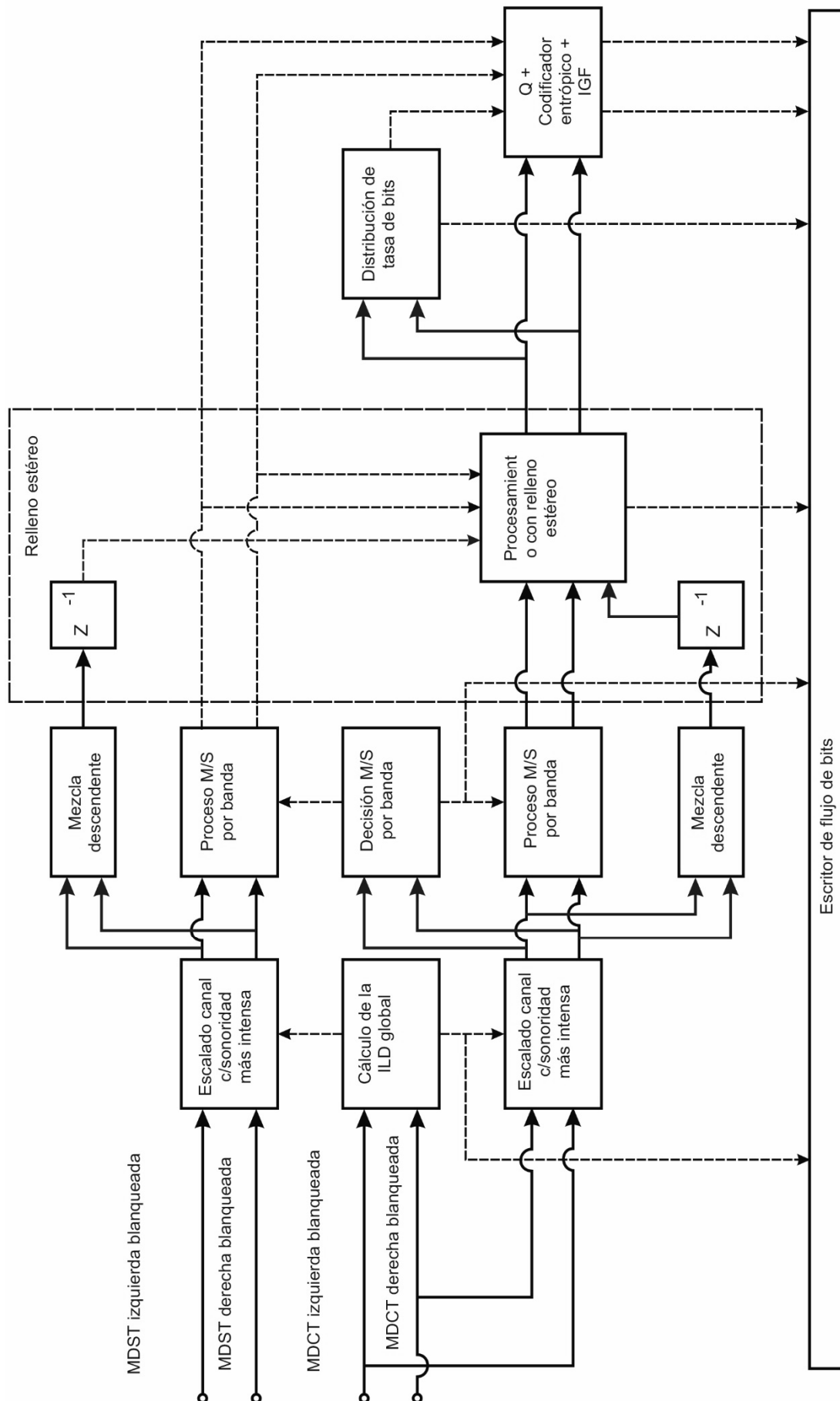


Fig. 9

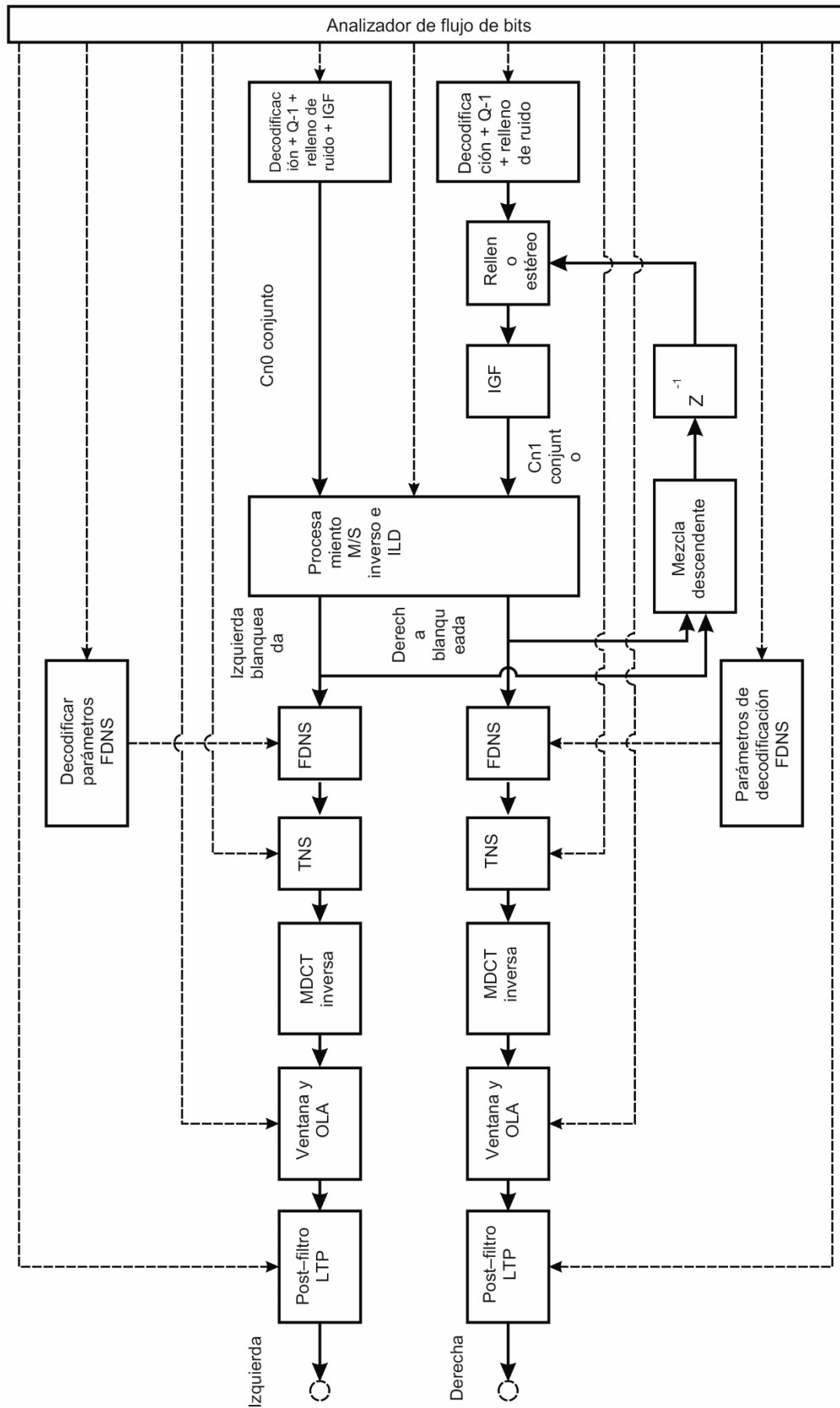


Fig. 10

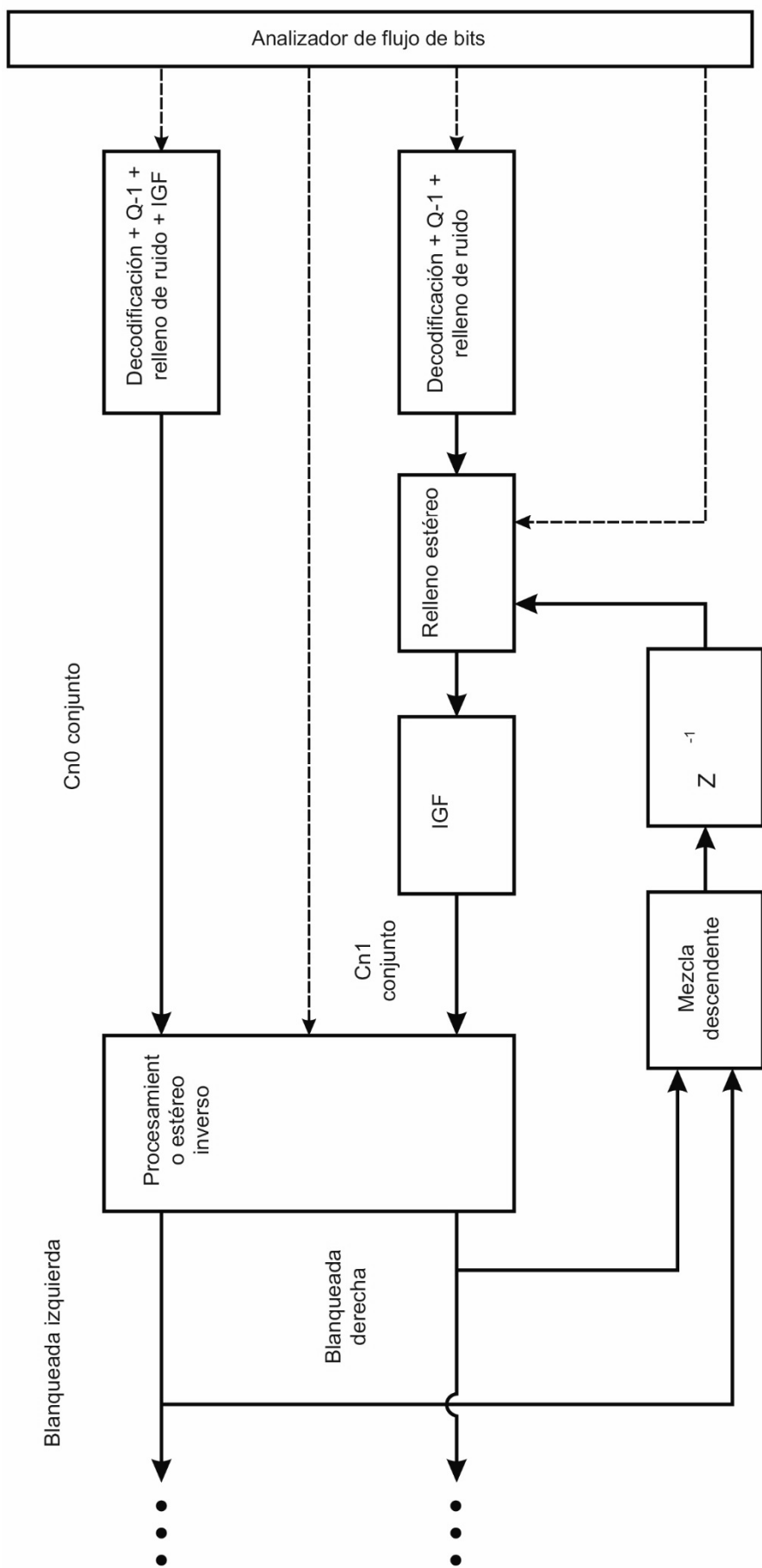


Fig. 11

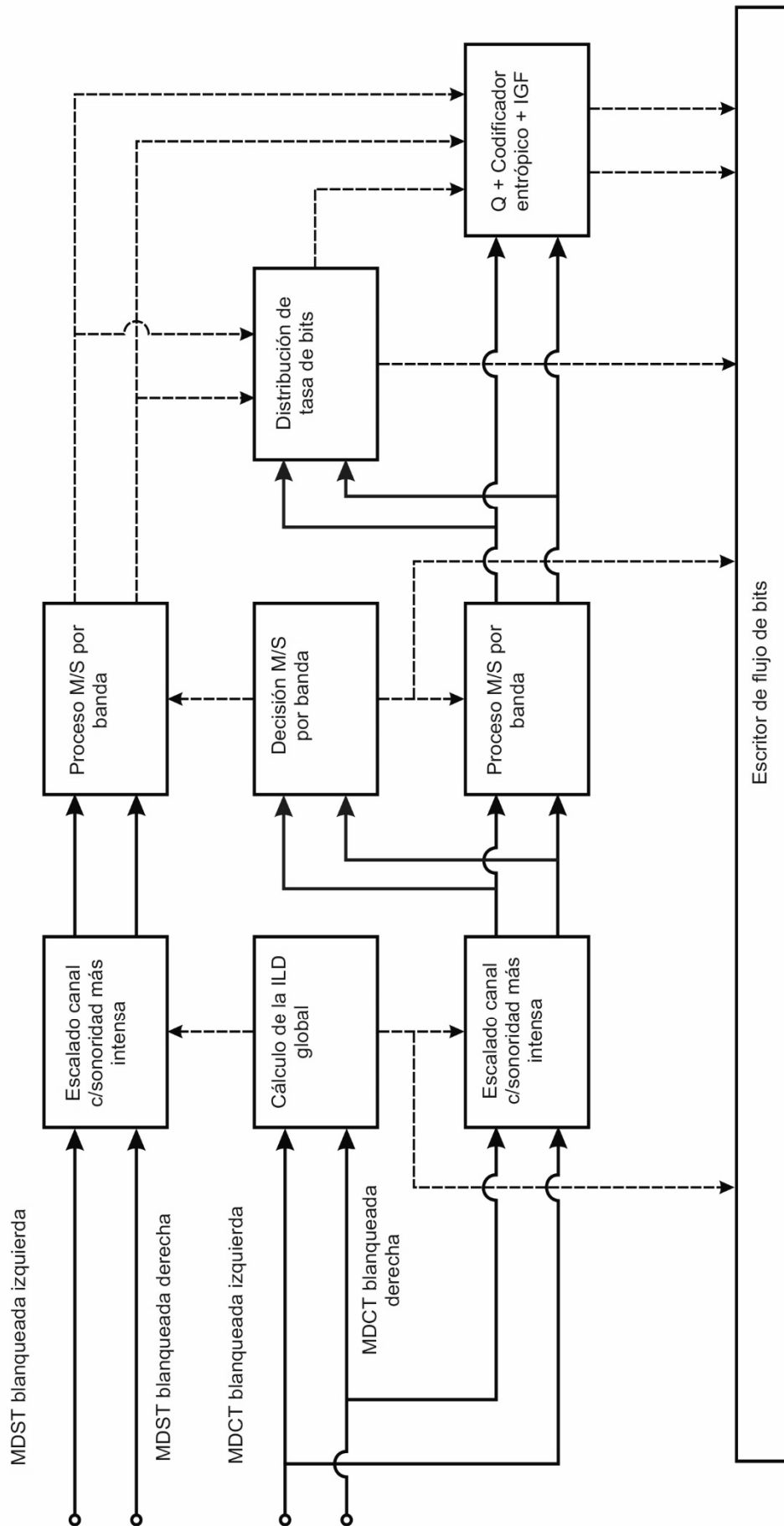


Fig. 12

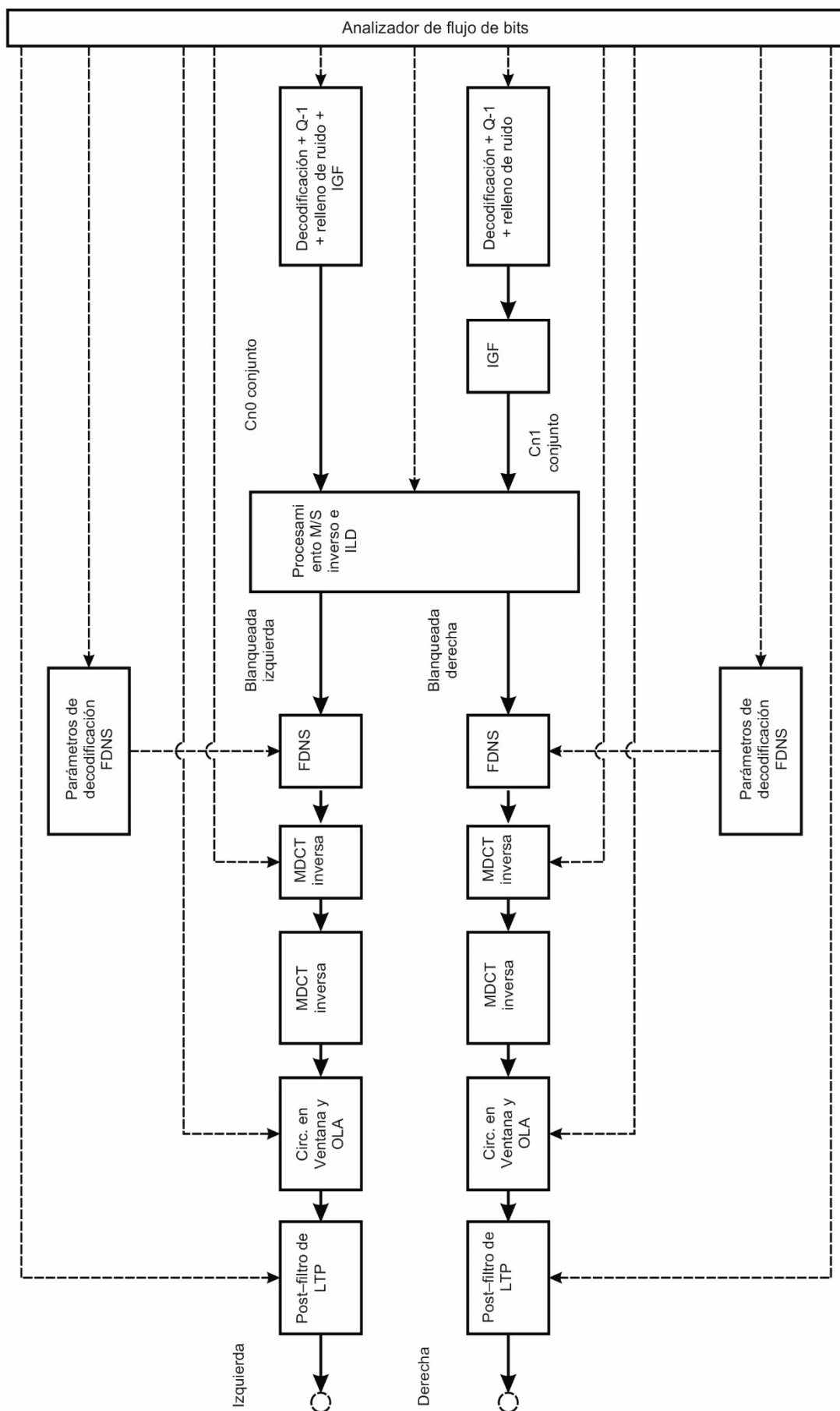


Fig. 13