

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 953 084**

51 Int. Cl.:

G10L 19/008 (2013.01)
G10L 19/06 (2013.01)
H04N 19/105 (2014.01)
H04N 19/146 (2014.01)
H04N 19/154 (2014.01)
H04N 19/156 (2014.01)
H04N 19/50 (2014.01)
H04N 19/70 (2014.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.02.2011** **E 20199069 (4)**

97 Fecha y número de publicación de la concesión europea: **21.06.2023** **EP 3779977**

54 Título: **Decodificador de audio para procesar audio estéreo usando una dirección de predicción variable**

30 Prioridad:

13.04.2010 US 32368310 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
08.11.2023

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastr. 27c
80686 München, DE**

72 Inventor/es:

**ROBILLIARD, JULIEN;
NEUSINGER, MATTHIAS;
HELMRICH, CHRISTIAN;
HILPERT, JOHANNES;
RETTTELACH, NIKOLAUS;
DISCH, SASCHA y
EDLER, BERND**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 953 084 T3

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

DESCRIPCIÓN

Decodificador de audio para procesar audio estéreo usando una dirección de predicción variable

- 5 La presente invención está relacionada con el procesamiento de audio y, en particular, con el procesamiento de una señal de audio estéreo que tiene dos señales de canal.

Se conoce en el campo de procesamiento multicanal o estéreo la aplicación de la llamada codificación estéreo media-lateral. En este concepto, se forma una combinación de la primera señal de canal de audio o señal izquierda y la segunda señal de canal de audio o señal derecha para obtener una señal media o mono M. Además, se forma una diferencia entre la primera señal de canal o señal izquierda y la segunda señal de canal o señal derecha para obtener la señal lateral S. Este método de codificación media-lateral tiene como resultado una ganancia de codificación significativa, cuando la señal izquierda y la señal derecha son bastante similares entre sí, dado que la señal lateral se volverá bastante pequeña. Por lo general, una ganancia de codificación de una etapa de un codificador entrópico/cuantificador se volverá más alta cuando el rango de valores a ser cuantificados/codificados entrópicamente se vuelva más pequeño. Por consiguiente, para un codificador entrópico de tipo PCM, basado en Huffman o de tipo aritmético, la ganancia de codificación aumenta cuando la señal lateral se vuelve más pequeña. Existen, no obstante, ciertas situaciones en las cuales la codificación media-lateral no tendrá como resultado una ganancia de codificación. La situación puede suceder cuando las señales en ambos canales tienen un desplazamiento de fase entre sí, por ejemplo, de 90°. Entonces, la señal media y la señal lateral pueden estar en un rango bastante similar y, por lo tanto, la codificación de la señal media y la señal lateral usando el codificador entrópico no producirá una ganancia de codificación y hasta puede resultar en una mayor velocidad de transferencia de bits. Por consiguiente, se puede aplicar una codificación media-lateral en modo selectivo de frecuencia a fin de desactivar la codificación media-lateral en bandas, donde la señal lateral no se vuelve más pequeña hasta cierto punto con respecto a la señal izquierda original, por ejemplo.

A pesar de que la señal lateral se volverá cero, cuando las señales izquierda y derecha son idénticas, dando como resultado una ganancia de codificación máxima debido a la eliminación de la señal lateral, la situación se vuelve otra vez diferente cuando la señal media y la señal lateral son idénticas con respecto a la forma de la forma de onda, pero la única diferencia entre ambas señales son sus amplitudes totales. En este caso, cuando además se supone que la señal lateral no tiene desplazamiento de fase respecto de la señal media, la señal lateral aumenta significativamente, aunque, por otro lado, la señal media no disminuye tanto con respecto a su rango de valor. Cuando ocurre una situación de ese tipo en una banda de frecuencia determinada, entonces se desactivaría nuevamente la codificación media-lateral debido a la falta de ganancia de codificación. La codificación media-lateral puede aplicarse de manera selectiva en cuanto a la frecuencia o puede aplicarse alternativamente en el dominio del tiempo.

Existen técnicas de codificación multicanal alternativas que no dependen de un tipo de aproximación de forma de onda como la codificación media-lateral, sino que dependen del procesamiento paramétrico basado en ciertos indicadores de referencia binaurales. Dichas técnicas se conocen bajo el término de "codificación indicador de referencia binaural", "codificación con estéreo paramétrico" o "codificación MPEG Envoltente". Aquí, se calculan ciertos indicadores de referencia para una pluralidad de bandas de frecuencia. Estos indicadores de referencia incluyen diferencias de nivel entre canales, medidas de coherencia entre canales, diferencias de tiempo entre canales y/o diferencias de fase entre canales. Estas aproximaciones comienzan a partir de la suposición de que una impresión multicanal percibida por el oyente no necesariamente depende de las formas de onda detalladas de los dos canales, sino que depende de los indicadores de referencia precisos proporcionados en modo selectivo de frecuencia o de información entre canales. Esto significa que, en una máquina de representación, se debe tener cuidado de representar señales multicanal que reflejen con precisión los indicadores de referencia, mientras que las formas de onda no tienen una importancia decisiva.

Esta aproximación puede ser compleja, especialmente en el caso en que el decodificador tiene que aplicar un proceso de decorrelación a fin de crear artificialmente señales estéreo que están decorrelacionadas entre sí, a pesar de que todos estos canales se derivan de uno y el mismo canal de mezcla descendente. Los decorrelacionadores con este fin son, según su implementación, complejos y pueden introducir artefactos especialmente en el caso de porciones de señales transitorias. Asimismo, en contraposición a la codificación de formas de onda, la aproximación de codificación paramétrica es una aproximación de codificación con pérdidas que inevitablemente tiene como resultado una pérdida de información que está introducida no solamente por la cuantificación típica sino también por concentrarse en los indicadores de referencia binaurales en vez de las formas de onda particulares. Esta aproximación tiene como resultado velocidades de transferencia de bits muy bajas pero puede incluir compromisos de calidad.

Existen desarrollos recientes para la codificación de voz y audio unificada (USAC) que se ilustran en la figura 7a. Un decodificador principal 700 realiza una operación de decodificación de la señal estéreo codificada en la entrada 701,

que puede tener codificación media-lateral. El decodificador principal produce una señal media en la línea 702 y una señal lateral o residual en la línea 703. Ambas señales son transformadas en un dominio QMF por bancos de filtros QMF 704 y 705. Luego, se aplica un decodificador MPEG Envolvente 706 para generar una señal de canal izquierdo 707 y una señal de canal derecho 708. Estas señales de banda baja son introducidas posteriormente en un decodificador de replicación de banda espectral (SBR) 709, que produce señales de banda ancha izquierda y derecha en las líneas 710 y 711, las cuales son transformadas luego en un dominio de tiempo por los bancos de filtros de síntesis QMF 712, 713 de manera que se obtengan las señales de banda ancha izquierda y derecha L, R.

La figura 7b ilustra la situación en la cual el decodificador MPEG Envolvente 706 realizaría una decodificación media-lateral. De manera alternativa, el bloque de decodificador MPEG Envolvente 706 podría realizar una decodificación paramétrica basada en indicadores de referencia binaurales para generar señales estéreo a partir de una señal de decodificador principal mono única. Naturalmente, el decodificador MPEG Envolvente 706 también podría generar una pluralidad de señales de salida de banda baja para ser ingresadas en el bloque de decodificador SBR 709 usando información paramétrica como diferencias de nivel entre canales, medidas de coherencia entre canales u otros parámetros de dicha información entre canales.

Cuando el bloque de decodificador MPEG Envolvente 706 realiza la decodificación media-lateral ilustrada en la figura 7b, se puede aplicar un factor de ganancia real g y DMX/RES y L/R son señales de mezcla descendente/residual e izquierda/derecha, respectivamente, representadas en el dominio QMF híbrido complejo.

El uso de una combinación de un bloque 706 y un bloque 709 provoca solo un pequeño aumento en complejidad computacional en comparación con un decodificador estéreo usado como base, porque la representación compleja QMF de la señal ya está disponible como parte del decodificador SBR. En una configuración que no es SBR, sin embargo, la codificación estéreo basada en QMF, según se propone en el contexto de USAC, tendría como resultado un aumento importante en complejidad computacional debido a los bancos QMF necesarios que requerirían en este ejemplo bancos de análisis de 64 bandas y bancos de síntesis de 64 bandas. Estos bancos de filtros tendrían que agregarse únicamente con el fin de la codificación estéreo.

En el sistema MPEG USAC en desarrollo, no obstante, también existen modos de codificación a altas velocidades de transferencia de bits donde por lo general no se usa SBR.

El documento WO 2008/014853 A1 da a conocer un dispositivo y método para procesar una señal de subbanda real para reducir los efectos de solapamiento. Se proporciona un ponderador para ponderar una señal de subbanda por un factor de ponderación determinado por una señal de subbanda. Además, se calcula un término de corrección mediante un determinador de término de corrección que se implementa para calcular el término de corrección utilizando por lo menos otra señal de subbanda y utilizando otro factor de ponderación que se proporciona para la otra señal de subbanda. El término de corrección se combina con la señal de subbanda ponderada para obtener la señal de subbanda corregida.

El documento WO 2009/141775 A1 da a conocer un aparato de mezcla ascendente estéreo paramétrico, un decodificador estéreo paramétrico, un aparato de mezcla descendente estéreo paramétrico y un codificador estéreo paramétrico. Se generan una señal izquierda y una señal derecha a partir de una señal de mezcla descendente mono basada en parámetros espaciales. La mezcla ascendente se caracteriza porque comprende un medio para predecir una señal de diferencia que comprende una diferencia entre la señal izquierda y la señal derecha basándose en la señal de mezcla descendente mono escalada con un coeficiente de predicción. El coeficiente de predicción se deriva de los parámetros espaciales. Adicionalmente, la señal izquierda y la señal derecha se derivan basándose en la suma y una diferencia de la señal de mezcla descendente mono y la señal de diferencia.

Es un objetivo de la presente invención proporcionar un concepto de procesamiento de audio o vídeo mejorado que, por un lado, produzca una alta ganancia de codificación y, por el otro, tenga como resultado una buena calidad de audio o vídeo y/o complejidad computacional reducida.

Este objetivo se logra por medio de un decodificador de audio o vídeo según la reivindicación 1.

La presente invención depende del hallazgo de que una ganancia de codificación de la aproximación de codificación de forma de onda de alta calidad puede mejorarse significativamente por una predicción de una segunda señal de combinación que use una primera señal de combinación, donde ambas señales de combinación se derivan de las señales originales usando una regla de combinación tal como la regla de combinación media-lateral. Se ha descubierto que esta información de predicción se calcula por un predictor en un codificador de audio o vídeo de modo que se cumpla con un objetivo de optimización, provoca solamente un pequeño margen de guarda, pero tiene como consecuencia una disminución significativa de velocidad de transferencia de bits requerida para la señal lateral sin perder calidad de audio o vídeo, ya que la predicción de la invención es, no obstante, una codificación basada en formas de onda y no una aproximación de codificación multicanal o estéreo basada en parámetros. Para reducir la

complejidad computacional, se prefiere realizar una codificación en el dominio de la frecuencia, donde la información de predicción se deriva de datos de entrada del dominio de la frecuencia en modo de selección de banda. El algoritmo de conversión para convertir la representación del dominio del tiempo en una representación espectral es preferentemente un proceso muestreado de forma crítica tal como una transformada discreta del coseno modificada (MDCT) o una transformada discreta del seno modificada (MDST), que difiere de una transformada compleja en que se calculan solo valores reales o solo valores imaginarios mientras que, en una transformada compleja, se calculan valores reales y complejos de un espectro cuyo resultado es un sobremuestreo doble.

Además, el concepto de conmutar la dirección de predicción logra un aumento en ganancia de predicción con un esfuerzo computacional mínimo. Con este fin, el codificador determina un indicador de dirección de predicción que indica una dirección de predicción asociada a la señal residual de predicción. En una realización, la primera señal de combinación tal como la señal media se usa para predecir la segunda señal de combinación tal como la señal lateral. Esta aproximación es útil cuando la energía de la señal media es mayor que la energía de la señal lateral. Sin embargo, cuando la energía de la segunda señal de combinación tal como la señal lateral es mayor que la energía de la primera señal de combinación, es decir, cuando la energía de la señal lateral es mayor que la energía de la señal media, entonces la dirección de predicción se invierte y la señal lateral se usa para predecir la señal media. En el primer caso, es decir, cuando se usa la señal media para predecir la señal lateral, la señal media, la señal residual, la información de predicción y el indicador de dirección de predicción que indica esta dirección de predicción se transmiten desde un codificador a un decodificador. En el segundo caso, cuando se usa la segunda señal de combinación para predecir la primera señal de combinación, es decir cuando la señal lateral es utilizada para predecir la señal media, la señal lateral (en vez de la señal media) se transmite junto con la señal residual, la información de predicción y el indicador de dirección de predicción que indica esta dirección inversa.

Este procedimiento permite un mejor enmascaramiento del ruido de cuantificación resultante. Para señales que tienen la mayor parte de su energía en la segunda señal de combinación que representa el canal lateral, la predicción de la primera señal de combinación que representa la señal media a partir de la señal lateral S permite un recorrido de la parte dominante del ruido de cuantificación de acuerdo con la fuente de sonido original. Esto, a su vez, tiene como resultado la distribución de error perceptivamente más adecuada en la última señal de salida.

Este procedimiento tiene la ventaja adicional de que proporciona una codificación estéreo eficaz. La elección adaptable a la señal de la dirección de predicción para codificación media-lateral (M/S) garantiza una mayor ganancia de predicción para señales con energía dominante en la señal lateral, con un aumento mínimo en complejidad computacional. Asimismo, se obtiene un enmascaramiento perceptivamente mejor adaptado del ruido de cuantificación resultante debido al recorrido de la dirección espacial percibida del ruido de cuantificación a la dirección de la señal principal. Adicionalmente, se reduce el rango de valores para los coeficientes de predicción que deben transmitirse en el flujo de bits, lo cual permite una codificación más eficaz de los coeficientes/información de predicción. Este procedimiento es útil para todo tipo de aplicaciones de procesamiento estéreo como la codificación de transformada conjunta de señales de audio y vídeo de doble canal o multicanal.

La MDCT usada permite una atenuación cruzada entre bloques subsiguientes sin ningún margen de guarda debido a la conocida propiedad de cancelación de solapamiento en el dominio del tiempo (TDAC) que se obtiene mediante un proceso de superposición-suma en el lado del decodificador.

La información de predicción calculada en el codificador, transmitida al decodificador y usada en el decodificador tiene un valor complejo y comprende una parte imaginaria que puede reflejar de manera ventajosa diferencias de fase entre dos canales de audio o vídeo en cantidades seleccionadas arbitrariamente entre 0° y 360° . La complejidad computacional se reduce significativamente debido al uso de la MDCT que proporciona un espectro con valor real.

A fin de hacer uso de esta información de predicción imaginaria que indica un desplazamiento de fase entre una banda determinada de la señal izquierda y una banda correspondiente de la señal derecha, se proporciona en el decodificador un convertidor real a imaginario para calcular una señal de predicción con rotación de fase a partir de la primera señal de combinación, que tiene rotación de fase con respecto a la señal de combinación original. Esta señal de predicción con rotación de fase se combina luego con la señal residual de predicción transmitida en el flujo de bits para regenerar una señal lateral que, finalmente, se combine con la señal media mediante una codificación media/lateral inversa y una MDCT inversa hasta obtener un resultado de la codificación media/lateral para obtener la señal de canal izquierdo decodificada y la señal de canal derecho decodificada.

Para aumentar la calidad de audio, el mismo convertidor real a imaginario que se aplica en el lado del decodificador se implementa en el lado del codificador también.

La presente invención es ventajosa en el sentido de que proporciona una calidad de audio mejorada y una velocidad de transferencia de bits reducida en comparación con sistemas que tienen la misma velocidad de transferencia de

bits o que tienen la misma calidad de audio o vídeo.

Además, se obtienen ventajas con respecto a la eficacia computacional de la codificación estéreo unificada que es útil en el sistema MPEG USAC a altas velocidades de transferencia de bits, donde por lo general no se usa SBR. En vez de procesar la señal en el dominio QMF híbrido complejo, estas aproximaciones implementan codificación estéreo predictiva con base residual en el dominio de MDCT nativo del codificador de transformada estéreo subyacente.

Posteriormente se tratan realizaciones preferidas, entre las cuales se encuentran las que son según las figuras 12b a 13b y la descripción asociada están cubiertas por la invención reivindicada, mientras que el resto no son según la invención reivindicada pero son útiles para resaltar aspectos específicos de la misma, con respecto a los dibujos adjuntos, en los que:

- 15 La figura 1 es un diagrama de bloques de una realización preferente de un decodificador de audio o vídeo;
- La figura 2 es un diagrama de bloques de una realización preferente de un codificador de audio o vídeo;
- La figura 3a ilustra una implementación de la calculadora del codificador de la figura 2;
- La figura 3b ilustra una implementación alternativa de la calculadora del codificador de la figura 2;
- La figura 3c ilustra una regla de combinación media-lateral que debe aplicarse del lado del codificador;
- La figura 4a ilustra una implementación de la calculadora del decodificador de la figura 1;
- 20 La figura 4b ilustra una implementación alternativa de la calculadora del decodificador en forma de una calculadora de matriz;
- La figura 4c ilustra una regla de combinación inversa media-lateral correspondiente a la regla de combinación ilustrada en la figura 3c;
- La figura 5a ilustra una realización de un codificador de audio o vídeo que funciona en el dominio de la frecuencia y que es preferentemente un dominio de la frecuencia con valor real;
- 25 La figura 5b ilustra una implementación de un decodificador de audio o vídeo que funciona en el dominio de la frecuencia;
- La figura 6a ilustra una implementación alternativa de un codificador de audio o vídeo que funciona en el dominio de MDCT y usa una transformada real a imaginaria.
- 30 La figura 6b ilustra un decodificador de audio o vídeo que funciona en el dominio de MDCT y usa una transformada real a imaginaria;
- La figura 7a ilustra un postprocesador de audio que usa un decodificador estéreo y un decodificador SBR conectado posteriormente;
- La figura 7b ilustra una matriz de mezcla ascendente media-lateral;
- 35 La figura 8a ilustra una vista detallada en el bloque de MDCT en la figura 6a;
- La figura 8b ilustra una vista detallada en el bloque de MDCT⁻¹ de la figura 6b;
- La figura 9a ilustra una implementación de un optimizador que funciona en resolución reducida con respecto a la salida de MDCT;
- La figura 9b ilustra una representación de un espectro de MDCT y las bandas de resolución más baja correspondientes en las cuales se calcula la información de predicción;
- 40 La figura 10a ilustra una implementación del transformador real a imaginario en la figura 6a o la figura 6b;
- La figura 10b ilustra una posible implementación de la calculadora de espectro imaginario de la figura 10a;
- La figura 11a ilustra una implementación preferida de un codificador de audio que tiene una dirección de predicción reversible;
- 45 La figura 11b ilustra una implementación preferida de un decodificador de audio relacionado que tiene la capacidad de procesar señales residuales generadas por direcciones de predicción reversibles;
- La figura 12a ilustra una realización preferida adicional de un codificador de audio o vídeo que tiene una dirección de predicción reversible;
- La figura 12b ilustra una realización adicional de un decodificador de audio o vídeo controlado por un indicador de dirección de predicción;
- 50 La figura 13a ilustra una bandera de dirección de predicción;
- La figura 13b ilustra una implementación de las distintas reglas de predicción del lado del codificador según el indicador de dirección de predicción;
- La figura 13c ilustra reglas de cálculo del lado del decodificador para un indicador de dirección de predicción que tiene un primer estado;
- 55 La figura 13d ilustra una regla de cálculo del lado del decodificador para un indicador de dirección de predicción distinto que tiene un segundo estado;
- La figura 13e ilustra multiplicaciones con valor complejo aplicadas en realizaciones del codificador de audio o vídeo o del decodificador de audio o vídeo; y
- 60 La figura 14 ilustra realizaciones para determinar el indicador de dirección de predicción.

La figura 1 ilustra un decodificador de audio o vídeo para decodificar una señal de audio multicanal codificada obtenida en una línea de entrada 100. La señal de audio multicanal codificada comprende una primera señal de

combinación codificada generada usando una regla de combinación para combinar una primera señal de canal y una segunda señal de canal que representan la señal de audio multicanal, una señal residual de predicción codificada e información de predicción. La señal multicanal codificada puede ser un flujo de datos como un flujo de bits que tiene los tres componentes en una forma multiplexada. Se puede incluir información lateral adicional en la señal multicanal codificada en la línea 100. La señal es ingresada en una interfaz de entrada 102. La interfaz de entrada 102 puede implementarse como un desmultiplexor de flujos de datos que produce la primera señal de combinación codificada en la línea 104, la señal residual codificada en la línea 106 y la información de predicción en la línea 108. Preferentemente, la información de predicción es un factor que tiene una parte real que no es igual a cero y/o una parte imaginaria distinta de cero. La señal de combinación codificada y la señal residual codificada se ingresan en un decodificador de señal 110 para decodificar la primera señal de combinación para obtener una primera señal de combinación decodificada en la línea 112. Asimismo, el decodificador de señal 110 se configura para decodificar la señal residual codificada para obtener una señal residual decodificada en la línea 114. Según el proceso de codificación en el lado de un codificador de audio, el decodificador de señal puede comprender un decodificador entrópico como un decodificador de Huffman, un decodificador aritmético o cualquier otro decodificador entrópico y una etapa de descuantificación conectada posteriormente para realizar una operación de descuantificación concordante con una operación de cuantificador en un codificador de audio asociado. Las señales en la línea 112 y 114 son ingresadas en una calculadora de decodificador 115, que produce la primera señal de canal en la línea 117 y una segunda señal de canal en la línea 118, donde estas dos señales son señales estéreo o dos canales de una señal de audio multicanal. Cuando, por ejemplo, la señal de audio multicanal comprende cinco canales, entonces las dos señales son dos canales de la señal multicanal. A fin de codificar totalmente una señal multicanal de ese tipo que tiene cinco canales, se pueden aplicar dos decodificadores de acuerdo con la figura 1, donde el primer decodificador procesa el canal izquierdo y el canal derecho, el segundo decodificador procesa el canal envolvente izquierdo y el canal envolvente derecho y un tercer decodificador mono se usaría para realizar una decodificación mono del canal central. Sin embargo, también se pueden aplicar otros modos de agrupar o combinaciones de codificadores de formas de onda y codificadores paramétricos. Una manera alternativa para generalizar el esquema de predicción a más de dos canales sería tratar tres (o más) señales al mismo tiempo, es decir, predecir una tercera señal de combinación a partir de una primera y una segunda señal usando dos coeficientes de predicción, de forma muy similar al módulo "dos a tres" en MPEG Envolvente.

Además, la señal de audio multicanal codificada obtenida en la línea de entrada 100 comprende un indicador de dirección de predicción. Este indicador de dirección de predicción, como una bandera de dirección de predicción, es extraído de la señal multicanal codificada por la interfaz de entrada 102 y es enviado a la calculadora del decodificador 116 de manera que la calculadora del decodificador calcule la señal multicanal decodificada según la información de predicción, la primera (o segunda) señal de combinación decodificada y el indicador de dirección de predicción proporcionados por la interfaz de entrada 102.

La calculadora del decodificador 116 está configurada para calcular una señal multicanal decodificada que tiene la primera señal de canal decodificada 117 y la segunda señal de canal decodificada 118 usando la señal residual decodificada 114, la información de predicción 108 y la primera señal de combinación decodificada 112. En particular, la calculadora del decodificador 116 está configurada para funcionar de modo tal que la primera señal de canal decodificada y la segunda señal de canal decodificada sean por lo menos una aproximación de una primera señal de canal y una segunda señal de canal de la señal multicanal ingresada en un codificador correspondiente, las cuales son combinadas por la regla de combinación al generar la primera señal de combinación y la señal residual de predicción. Específicamente, la información de predicción en la línea 108 comprende una parte con valor real distinta de cero y/o una parte imaginaria distinta de cero.

La calculadora del decodificador 116 puede implementarse de diferentes maneras. En la figura 4a se ilustra una primera implementación. Esta implementación comprende un predictor 1160, una calculadora de señal de combinación 1161 y un combinador 1162. El predictor recibe la primera señal de combinación decodificada 112 y la información de predicción 108 y produce una señal de predicción 1163. De manera más específica, el predictor 1160 está configurado para aplicar la información de predicción 108 a la primera señal de combinación decodificada 112 o una señal derivada de la primera señal de combinación decodificada. La regla de derivación para derivar la señal a la cual se aplica la información de predicción 108 puede ser una transformada real a imaginaria, o igualmente, una transformada imaginaria a real o una operación de ponderación o, según la implementación, una operación de desplazamiento de fase o una operación combinada de ponderación/desplazamiento de fase. La señal de predicción 1163 es ingresada junto con la señal residual decodificada en la calculadora de señal de combinación 1161 a fin de calcular la segunda señal de combinación decodificada 1165. Ambas señales 112 y 1165 son ingresadas en el combinador 1162, el cual combina la primera señal de combinación decodificada y la segunda señal de combinación para obtener la señal de audio multicanal decodificada que tiene la primera señal de canal decodificada y la segunda señal de canal decodificada en las líneas de salida 1166 y 1167, respectivamente. De manera alternativa, la calculadora del decodificador se implementa como una calculadora de matriz 1168 que recibe, como entrada, la primera señal de combinación decodificada o señal M, la señal residual decodificada o señal D y la información de predicción α 108. La calculadora de matriz 1168 aplica una matriz de transformada ilustrada como 1169 a las

señales M, D para obtener las señales de salida L, R, donde L es la primera señal de canal decodificada y R es la segunda señal de canal decodificada. La notación en la figura 4b se asemeja a una notación estéreo con un canal izquierdo L y un canal derecho R. Esta notación se ha aplicado para facilitar la comprensión, pero está claro para aquellos expertos en la técnica que las señales L, R pueden ser cualquier combinación de dos señales de canal en una señal multicanal que tiene más de dos señales de canal. La operación de matriz 1169 unifica las operaciones en los bloques 1160, 1161 y 1162 de la figura 4a en un tipo de cálculo de matriz “de una sola vez” y las entradas en el circuito de la figura 4a y las salidas desde el circuito de la figura 4a son idénticas a las entradas en la calculadora de matriz 1168 o las salidas desde la calculadora de matriz 1168.

La figura 4c ilustra un ejemplo para una regla de combinación inversa aplicada por el combinador 1162 en la figura 4a. En particular, la regla de combinación es similar a la regla de combinación del lado del decodificador en la codificación media-lateral conocida, donde $L = M + S$, y $R = M - S$. Cabe señalar que la señal S usada por la regla de combinación inversa en la figura 4c es la señal calculada por la calculadora de señal de combinación, es decir, la combinación de la señal de predicción en la línea 1163 y la señal residual decodificada en la línea 114. Cabe señalar que en esta especificación las señales en líneas se denominan a veces por los números de referencia para las líneas o a veces están indicadas por los números de referencia mismos, los cuales han sido atribuidos a las líneas. Por lo tanto, la notación es tal que una línea que tiene una señal determinada está indicando la señal misma. Una línea puede ser una línea física en una implementación cableada. En una implementación computarizada, sin embargo, no existe una línea física, sino que la señal representada por la línea es transmitida desde un módulo de cálculo al otro módulo de cálculo.

La figura 11b ilustra otra implementación preferida de la calculadora del decodificador que funciona dependiente del indicador de dirección de predicción proporcionado en una entrada de indicador de dirección de predicción 401. Según el estado del indicador de dirección de predicción, se aplica una primera regla de cálculo ilustrada en 402 o una segunda regla de cálculo ilustrada en 403. La otra regla de cálculo 402 proporciona, en una salida, la primera señal de canal y la segunda señal de canal y la primera regla de cálculo puede implementarse como se ilustra en la figura 13c que se describe más adelante. En una realización específica donde la primera señal de combinación es la señal media y la segunda señal de combinación es la señal lateral, el indicador de dirección de predicción tiene un valor de “0” y la predicción se lleva a cabo desde la primera señal de combinación a la segunda señal de combinación. En este caso, la entrada 404 tiene la señal media, es decir, la primera señal de combinación. No obstante, cuando el indicador de dirección de predicción es igual a “1”, entonces una conmutación 405 conecta la entrada 404 a la entrada del dispositivo de la segunda regla de cálculo 403. En este caso, se realiza una predicción desde la segunda señal de combinación como la señal lateral a la primera señal de combinación como la señal media y la entrada 404 tendrá la señal lateral en vez de la señal media. El dispositivo de la segunda regla de cálculo 403 producirá, nuevamente, la primera señal de canal y la segunda señal de canal, pero las reglas para calcular estas dos señales, es decir la señal izquierda y la señal derecha en una realización estéreo, serán diferentes. En la figura 13d, que se discute más adelante, se ilustra una realización específica para la segunda regla de cálculo.

La figura 2 ilustra un codificador de audio para codificar una señal de audio multicanal 200 que tiene dos o más señales de canal, donde una primera señal de canal se ilustra en 201 y en 202 se ilustra un segundo canal. Ambas señales se ingresan en una calculadora de codificador 203 para calcular una primera señal de combinación 204 y una señal residual de predicción 205 usando la primera señal de canal 201 y la segunda señal de canal 202 y la información de predicción 206, de modo que la señal residual de predicción 205, cuando se combina con una señal de predicción derivada de la primera señal de combinación 204 y la información de predicción 206 da como resultado una segunda señal de combinación, donde la primera señal de combinación y la segunda señal de combinación se pueden derivar de la primera señal de canal 201 y la segunda señal de canal 202 usando una regla de combinación.

La información de predicción es generada por un optimizador 207 para calcular la información de predicción 206 de manera que la señal residual de predicción cumpla con un objetivo de optimización 208. La primera señal de combinación 204 y la señal residual 205 se ingresan en un codificador de señal 209 para codificar la primera señal de combinación 204 para obtener una primera señal de combinación codificada 210 y para codificar la señal residual 205 para obtener una señal residual codificada 211. Ambas señales codificadas 210, 211 se ingresan en una interfaz de salida 212 para combinar la primera señal de combinación codificada 210 con la señal residual de predicción codificada 211 y la información de predicción 206 para obtener una señal multicanal codificada 213, que es similar a la señal multicanal codificada 100 ingresada en la interfaz de entrada 102 del decodificador de audio ilustrado en la figura 1.

Según la implementación, el optimizador 207 recibe, o bien la primera señal de canal 201 y la segunda señal de canal 202 o, como lo ilustran las líneas 214 y 215, la primera señal de combinación 214 y la segunda señal de combinación 215 derivadas de un combinador 2031 de la figura 3a, que se discutirá posteriormente.

En la figura 2 se ilustra un objetivo de optimización preferida, en la cual la ganancia de codificación está maximizada,

es decir, la velocidad de transferencia de bits se reduce todo lo posible. En este objetivo de optimización, la señal residual D está minimizada con respecto a α . Esto significa, en otras palabras, que la información de predicción α está elegida de manera que se minimice $\|S - \alpha M\|^2$. Esto da como resultado una solución para α que se ilustra en la figura 2. Las señales S, M están dadas a modo de bloques y son preferentemente señales del dominio espectral, donde la notación $\|\dots\|$ significa la norma 2 del argumento y donde $\langle \dots \rangle$ ilustra el producto escalar como siempre. Cuando la primera señal de canal 201 y la segunda señal de canal 202 son ingresadas en el optimizador 207, entonces el optimizador tendría que aplicar la regla de combinación, donde una regla de combinación ejemplar se ilustra en la figura 3c. Cuando, sin embargo, la primera señal de combinación 214 y la segunda señal de combinación 215 son ingresadas en el optimizador 207, el optimizador 207 no necesita implementar la regla de combinación por sí mismo.

Otros objetivos de optimización pueden relacionarse con la calidad perceptiva. Un objetivo de optimización puede ser obtener una máxima calidad perceptiva. Luego, el optimizador requeriría información adicional de un modelo perceptivo. Otras implementaciones del objetivo de optimización pueden relacionarse con obtener una velocidad de transferencia de bits mínima o fija. Luego, el optimizador 207 se implementaría para realizar una operación de cuantificación/codificación entrópica a fin de determinar la velocidad de transferencia de bits requerida para ciertos valores α de manera que α pueda fijarse para cumplir con los requisitos como una mínima velocidad de transferencia de bits o, de manera alternativa, una velocidad de transferencia de bits fija. Otras implementaciones del objetivo de optimización pueden relacionarse con un uso mínimo de recursos de codificador o decodificador. En caso de una implementación de un objetivo de optimización de ese tipo, la información sobre los recursos requeridos para una optimización determinada estaría disponible en el optimizador 207. Además, se puede aplicar una combinación de estos objetivos de optimización u otros objetivos de optimización para controlar el optimizador 207 que calcula la información de predicción 206.

El codificador de audio además comprende una calculadora de dirección de predicción 219 que proporciona, en su salida, el indicador de dirección de predicción que indica una dirección de predicción asociada a la señal residual de predicción 205 que produce la calculadora de codificador 203 en la figura 2. La calculadora de dirección de predicción 219 puede implementarse de distintas maneras, donde se discuten varios ejemplos en el contexto de la figura 14.

La calculadora del codificador 203 en la figura 2 puede implementarse de distintas maneras, donde una primera implementación ilustrativa se ilustra en la figura 3a, en la cual se realiza una regla de combinación explícita en el combinador 2031. En la figura 3b se ilustra una implementación ejemplar alternativa, donde se usa una calculadora de matriz 2039. El combinador 2031 en la figura 3a puede implementarse para realizar la regla de combinación ilustrada en la figura 3c, que es de manera ilustrativa la regla de codificación media-lateral conocida, donde se aplica un factor de ponderación de 0,5 a todas las ramificaciones. No obstante, según la implementación se pueden usar otros factores de ponderación o ningún factor de ponderación en absoluto (ponderación de unidad). Además, cabe señalar que se pueden aplicar otras reglas de combinación como otras reglas de combinación lineal o reglas de combinación no lineal, siempre que exista una regla de combinación inversa correspondiente que pueda aplicarse en el combinador de decodificador 1162 ilustrado en la figura 4a, que aplica una regla de combinación que es inversa a la regla de combinación aplicada por el codificador. Debido a la predicción de la invención, se puede usar cualquier regla de predicción invertible, dado que la influencia en la forma de onda está "equilibrada" por la predicción, es decir, en la señal residual transmitida se incluye cualquier error, ya que la operación de predicción realizada por el optimizador 207 en combinación con la calculadora del codificador 203 es un proceso de conservación de forma de onda.

El combinador 2031 produce la primera señal de combinación 204 y una segunda señal de combinación 2032. La primera señal de combinación es ingresada en un predictor 2033 y la segunda señal de combinación 2032 es ingresada en la calculadora residual 2034. El predictor 2033 calcula una señal de predicción 2035, que se combina con la segunda señal de combinación 2032 para obtener finalmente la señal residual 205. En particular, el combinador 2031 está configurado para combinar las dos señales de canal 201 y 202 de la señal de audio multicanal en dos formas distintas para obtener la primera señal de combinación 204 y la segunda señal de combinación 2032, donde las dos formas diferentes se ilustran en una realización ilustrativa en la figura 3c. El predictor 2033 está configurado para aplicar la información de predicción a la primera señal de combinación 204 o una señal derivada de la primera señal de combinación para obtener la señal de predicción 2035. La señal derivada de la señal de combinación puede derivarse por medio de cualquier operación lineal o no lineal, donde se prefiere una transformada real a imaginaria/transformada imaginaria a real, que se puede implementar usando un filtro lineal como un filtro FIR que realiza sumas ponderadas de ciertos valores.

La calculadora residual 2034 en la figura 3a puede realizar una operación de resta de manera que la señal de predicción se reste de la segunda señal de combinación. Sin embargo, otras operaciones son posibles en la calculadora residual. De la misma manera, la calculadora de señal de combinación 1161 en la figura 4a puede realizar una operación de suma donde la señal residual decodificada 114 y la señal de predicción 1163 se suman

juntas para obtener la segunda señal de combinación 1165.

- La figura 11a ilustra una implementación preferida de la calculadora de codificador. Según el indicador de dirección de predicción ingresado en la entrada de dirección de predicción 501, se selecciona o una primera regla de predicción 502 o una segunda regla de predicción 503 que está ilustrada por una conmutación de selección controlada 505. La primera regla de predicción puede ser similar a lo que se ilustra en la figura 13b, primera alternativa, y la segunda regla de predicción puede ser similar a lo que se ilustra en la figura 13b, segunda alternativa. La salida de los bloques 502, 503, es decir, una señal de combinación y la señal residual, pueden enviarse a la interfaz de salida, o en caso de una codificación de señal, al codificador de señal 209 en la figura 2.
- Asimismo, el indicador de dirección de predicción es ingresado en el flujo de bits de salida junto con la información de predicción, la señal residual codificada y la señal de combinación codificada que puede ser o bien la primera señal de combinación en caso de un indicador de dirección de predicción igual a "0" o una segunda señal de combinación en caso de un indicador de dirección de predicción igual a "1".
- La figura 5a ilustra una implementación preferida de un codificador de audio. En comparación con el codificador de audio ilustrado en la figura 3a, la primera señal de canal 201 es una representación espectral de una primera señal de canal en el dominio del tiempo 55a. De la misma manera, la segunda señal de canal 202 es una representación espectral de una señal de canal en el dominio del tiempo 55b. La conversión desde el dominio del tiempo en la representación espectral es realizada por un convertidor de tiempo/frecuencia 50 para la primera señal de canal y un convertidor de tiempo/frecuencia 51 para la segunda señal de canal. Preferentemente, pero no necesariamente, los convertidores espectrales 50, 51 se implementan como convertidores con valor real. El algoritmo de conversión puede ser una transformada discreta del coseno (DCT), una FFT donde solo se usa la parte real, una MDCT o cualquier otra transformada que proporcione valores espectrales con valor real.
- De manera alternativa, ambas transformadas pueden implementarse como una transformada imaginaria, como una DST, una MDST o una FFT donde solo se usa la parte imaginaria y la parte real se descarta. También se puede usar cualquier otra transformada que solo proporcione valores imaginarios. Un objetivo de usar una transformada con valor real pura o una transformada imaginaria pura es la complejidad computacional, ya que, para cada valor espectral, solamente se tiene que procesar un solo valor como magnitud o la parte real o, de manera alternativa, la fase o la parte imaginaria. En contraposición a esto, en una transformada totalmente compleja como una FFT, tendrían que procesarse dos valores, es decir, la parte real y la parte imaginaria para cada línea espectral, lo cual es un aumento de complejidad computacional por un factor de por lo menos 2. Otra razón para usar una transformada con valor real aquí es que dicha transformada por lo general está muestreada críticamente y, por lo tanto, proporciona un dominio adecuado (y normalmente utilizado) para cuantificación de señales y codificación entrópica (el paradigma estándar de "codificación de audio perceptiva" implementado en "MP3", AAC o sistemas de codificación de audio similares).
- La figura 5a además ilustra la calculadora residual 2034 como un sumador que recibe la señal lateral en su entrada "positiva" y que recibe la salida de la señal de predicción por el predictor 2033 en su entrada "negativa". Asimismo, la figura 5a ilustra la situación en que la información de control del predictor es enviada desde el optimizador al multiplexor 212 que produce un flujo de bits multiplexado que representa la señal de audio multicanal codificada. En particular, la operación de predicción se realiza de manera tal que la señal lateral sea predicha desde la señal media como se ilustra mediante las Ecuaciones a la derecha de la figura 5a.
- Mientras la figura 5a ilustra una predicción de M a S, es decir, la señal media predice la señal lateral, lo cual ocurre para un indicador de dirección de predicción igual a cero, se aplica una predicción inversa cuando el indicador de dirección de predicción es igual a 1. Entonces, se realiza una predicción de S a M. Esto se puede ilustrar por medio de intercambiar las salidas del bloque 2031 de modo que la salida superior tenga la señal lateral y la salida inferior tenga la señal media.
- Preferentemente, la información de control del predictor 206 es un factor como el que se ilustra a la derecha en la figura 3b. En una realización en la cual la información de control de predicción solamente comprende una porción real como la parte real de α con valor complejo o una magnitud de α con valor complejo, donde esta porción corresponde a un factor distinto de cero, se puede obtener una ganancia de codificación significativa cuando la señal media y la señal lateral son similares entre sí debido a su estructura de forma de onda, pero tienen amplitudes diferentes.
- Sin embargo, cuando la información de control de predicción solo comprende una segunda porción que puede ser la parte imaginaria de un factor con valor complejo o la información de fase del factor con valor complejo, donde la parte imaginaria o la información de fase es distinta de cero, la presente invención logra una ganancia de codificación significativa para señales que están desfasadas entre sí por un valor distinto de 0° o 180° y que tienen, además del desplazamiento de fase, características de forma de onda similares y relaciones de amplitud similares.

Preferentemente, la información de control de predicción tiene valor complejo. Luego, se puede obtener una ganancia de codificación significativa para señales que son distintas en amplitud y están desfasadas. En una situación en la cual las transformadas de tiempo/frecuencia proporcionan espectros complejos, la operación 2034 sería una operación compleja en la cual la parte real de la información de control del predictor se aplica a la parte real del espectro complejo M y la parte imaginaria de la información de predicción compleja se aplica a la parte imaginaria del espectro complejo. Luego, en el sumador 2034, el resultado de esta operación de predicción es un espectro real predicho y un espectro imaginario predicho, y el espectro real predicho se restaría del espectro real de la señal lateral S (respecto de la banda), y el espectro imaginario predicho se restaría de la parte imaginaria del espectro de S para obtener un espectro residual complejo D.

Las señales en el dominio del tiempo L y R son señales con valor real pero las señales en el dominio de la frecuencia pueden ser con valor real o complejo. Cuando las señales en el dominio de la frecuencia son con valor real entonces la transformada es una transformada con valor real. Cuando las señales en el dominio de la frecuencia son con valor complejo entonces la transformada es una transformada con valor complejo. Esto significa que la entrada a la transformada de tiempo a frecuencia y la salida de la transformada de frecuencia a tiempo son con valor real, mientras que las señales en el dominio de la frecuencia podrían ser por ejemplo señales en el dominio QMF con valor complejo.

La figura 5b ilustra un decodificador de audio que corresponde al codificador de audio ilustrado en la figura 5a. Los elementos similares con respecto al decodificador de audio de la figura 1 tienen números de referencia similares.

La salida del flujo de bits por el multiplexor de flujo de bits 212 en la figura 5a se ingresa en un desmultiplexor de flujo de bits 102 en la figura 5b. El desmultiplexor de flujo de bits 102 desmultiplexa el flujo de bits en la señal de mezcla descendente M y la señal residual D. La señal de mezcla descendente M se ingresa en un descuantificador 110a. La señal residual D se ingresa en un descuantificador 110b. Además, el desmultiplexor de flujo de bits 102 desmultiplexa una información de control del predictor 108 a partir del flujo de bits e ingresa la misma en el predictor 1160. El predictor 1160 produce una señal lateral predicha $\alpha \cdot M$ y el combinador 1161 combina la salida de señal residual mediante el descuantificador 110b con la señal lateral predicha para finalmente obtener la señal lateral reconstruida S. Luego se ingresa la señal en el combinador 1162 que realiza, por ejemplo, un procesamiento de suma/diferencia, como se ilustra en la figura 4c con respecto a la codificación media/lateral. En particular, el bloque 1162 realiza una decodificación media/lateral (inversa) para obtener una representación en el dominio de la frecuencia del canal izquierdo y una representación en el dominio de la frecuencia del canal derecho. Luego, se convierte la representación en el dominio de la frecuencia en una representación en el dominio del tiempo mediante los convertidores de frecuencia/tiempo correspondientes 52 y 53.

La figura 5b ilustra la situación donde se ha realizado la predicción, en el codificador, de la señal media M a la señal lateral S indicada por el indicador de dirección de predicción igual a cero. Sin embargo, cuando un indicador de dirección de predicción igual a 1 se transmite del codificador como en el codificador en la figura 5a al decodificador en la figura 5b, entonces debe realizarse una predicción inversa de S a M, es decir la regla de cálculo del decodificador es tal que M se calcula a partir de S en lugar del cálculo opuesto en el caso de un indicador de dirección de predicción igual a cero.

Según la implementación del sistema, los convertidores de frecuencia/tiempo 52, 53 son convertidores de frecuencia/tiempo con valor real cuando la representación en el dominio de la frecuencia es una representación con valor real o convertidores de frecuencia/tiempo con valor complejo cuando la representación en el dominio de la frecuencia es una representación con valor complejo.

Sin embargo, para mayor eficacia, se prefiere realizar una transformada con valor real como se ilustra en otra implementación en la figura 6a para el codificador y en la figura 6b para el decodificador. Las transformadas con valor real 50 y 51 se implementan mediante una MDCT. Además, la información de predicción se calcula como un valor complejo que tiene una parte real y una parte imaginaria. Dado que ambos espectros M, S son espectros con valor real y, dado que, por lo tanto, no existe ninguna parte imaginaria del espectro, se proporciona un convertidor real a imaginario 2070 que calcula un espectro imaginario estimado 600 a partir del espectro con valor real de la señal M. Este transformador real a imaginario 2070 es una parte del optimizador 207 y el espectro imaginario 600 estimado por el bloque 2070 se ingresa en la etapa del optimizador α 2071 junto con el espectro real M para calcular la información de predicción 206, que ahora tiene un factor con valor real indicado en 2073 y un factor imaginario indicado en 2074. Ahora, de conformidad con esta realización, el espectro con valor real de la primera señal de combinación M se multiplica por la parte real α_R 2073 para obtener la señal de predicción que luego se resta del espectro lateral con valor real. Además, el espectro imaginario 600 se multiplica por la parte imaginaria α_I ilustrada en 2074 para obtener la otra señal de predicción, donde esta señal de predicción luego se resta del espectro lateral con valor real como se indica en 2034b. Luego, la señal residual de predicción D se cuantifica en el cuantificador 209b mientras que el espectro con valor real de M se cuantifica/codifica en el bloque 209a. Además, se prefiere cuantificar y codificar la información de predicción α en el cuantificador/codificador entrópico 2072 para obtener el

valor α complejo codificado que se envía al multiplexor de flujo de bits 212 de la figura 5a, por ejemplo, y que finalmente se ingresa en un flujo de bits como información de predicción.

Con relación a la posición del módulo de cuantificación/codificación (Q/C) 2072 para α , se observa que los multiplicadores 2073 y 2074 preferentemente utilizan exactamente la misma α (cuantificada) que también se utilizará en el decodificador. Por lo tanto, uno podría mover 2072 directamente hacia la salida de 2071 o uno podría considerar que la cuantificación de α ya se tiene en cuenta en el proceso de optimización en 2071.

Aunque uno podría calcular un espectro complejo en el lado del codificador, dado que toda la información está disponible, se prefiere realizar la transformada real a compleja en el bloque 2070 en el codificador de manera que se producen condiciones similares con respecto a un decodificador ilustrado en la figura 6b. El decodificador recibe un espectro codificado con valor real de la primera señal de combinación y una representación espectral con valor real de la señal residual codificada. Además, se obtiene una información de predicción compleja codificada en 108 y se realiza una decodificación entrópica y una descuantificación en el bloque 65 para obtener la parte real α_R ilustrada en 1160b y la parte imaginaria α_I ilustrada en 1160c. La salida de señales medias por los elementos de ponderación 1160b y 1160c se agrega a la señal residual de predicción decodificada y descuantificada. Particularmente, los valores espectrales ingresados en el ponderador 1160c, donde la parte imaginaria del factor de predicción complejo se utiliza como el factor de ponderación, se derivan del espectro con valor real M mediante el convertidor real a imaginario 1160a, que se implementa preferentemente de la misma manera que el bloque 2070 de la figura 6a relacionado con el lado del codificador. En el lado del decodificador no hay una representación con valor complejo de la señal media o la señal lateral, a diferencia del lado del codificador. El motivo es que solo los espectros con valor real codificados se han transmitido del codificador al decodificador debido a la velocidad de transferencia de bits y motivos de complejidad.

La figura 6a y la figura 6b ilustran la situación, donde el indicador de dirección de predicción es igual a cero, es decir donde se realiza una predicción de M a S o un cálculo de S utilizando M y la información de predicción compleja α . Sin embargo, cuando el indicador de dirección de predicción es igual a 1 o, como se expresa generalmente, indica una dirección de predicción inversa, entonces puede aplicarse el mismo circuito pero las salidas del bloque 2031 se intercambian de manera que la línea superior tiene la señal lateral S y la línea inferior tiene la señal media M. En el lado del decodificador también se cambia la regla de cálculo del decodificador de manera que, en el caso de una dirección de predicción inversa, M se calcula a partir de S que también puede indicarse reemplazando la señal M en la figura 6b en la línea superior en la salida del bloque 110a mediante la señal lateral S. Esto tiene como resultado una señal media M en la salida del bloque 1161b y la señal lateral S en la entrada superior del bloque 1162. Por lo tanto, la regla aplicada por el bloque 1162 tiene que adaptarse a esta situación de entrada diferente o bien las señales M/S tienen que intercambiarse antes de ser ingresadas en el bloque 1162. En el último caso, es decir cuando se realiza un intercambio, el bloque 1162 es el mismo para ambos valores de indicador de dirección de predicción.

El transformador real a imaginario 1160a o el bloque correspondiente 2070 de la figura 6a puede implementarse como se publica en WO 2004/013839 A1 o WO 2008/014853 A1 o la Patente Estadounidense N.º 6.980.933. Según la señal o la implementación, la información de predicción α puede ser con valor real puramente o con valor imaginario puramente o puede ser un número complejo que tiene una parte real y una parte imaginaria. Sin embargo, si solo se implementa la predicción con valor real, la inversión de dirección de predicción ya proporcionará una realización mejorada con requisitos de cálculo adicionales muy limitados y tendrá como resultado una velocidad de transferencia de bits menor debido al hecho de que la señal residual tendrá menor energía, y lo mismo es cierto también para la información de predicción. Por lo tanto, la velocidad de transferencia de bits adicional requerida para transmitir el indicador de dirección de predicción al final tiene como resultado considerables ahorros de bits debido a la menor velocidad de transferencia de bits requerida para la señal residual y la información de predicción. Por lo tanto, la información de predicción puede comprender una porción con valor real distinto de cero y/o una porción imaginaria distinta de cero. De manera alternativa, puede aplicarse cualquier otra implementación conocida en la técnica y se describe una implementación preferida en los contextos de las figuras 10a, 10b.

Específicamente, como se ilustra en la figura 10a, el convertidor real a imaginario 1160a comprende un selector de trama espectral 1000 conectado a una calculadora de espectro imaginario 1001. El selector de trama espectral 1000 recibe una indicación de una trama actual i en la entrada 1002 y, según la implementación, información de control en una entrada de control 1003. Cuando, por ejemplo, la indicación en la línea 1002 indica que debe calcularse un espectro imaginario para una trama actual i, y cuando la información de control 1003 indica que debe utilizarse solo la trama actual para ese cálculo, entonces el selector de trama espectral 1000 solo selecciona la trama actual i y envía esta información a la calculadora de espectro imaginario. Entonces, la calculadora de espectro imaginario solo utiliza las líneas espectrales de la trama actual i para realizar una combinación ponderada de las líneas ubicadas en la trama actual (bloque 1008), con respecto a la frecuencia, cerca o alrededor de la línea espectral actual k, para la cual debe calcularse una línea imaginaria como se ilustra en 1004 en la figura 10b. Sin embargo, cuando el selector de trama espectral 1000 recibe una información de control 1003 que indica que también debe utilizarse la trama

precedente $i-1$ y la trama siguiente $i+1$ para el cálculo de espectro imaginario, entonces la calculadora de espectro imaginario además recibe los valores de las tramas $i-1$ y $i+1$ y realiza una combinación ponderada de las líneas en las tramas correspondientes como se ilustra en 1005 para la trama $i-1$ y en 1006 para la trama $i+1$. Los resultados de las operaciones de ponderación se combinan mediante una combinación ponderada en el bloque 1007 para finalmente obtener una línea imaginaria k para la trama f_i que luego se multiplica por la parte imaginaria de la información de predicción en el elemento 1160c para obtener la señal de predicción para esta línea que luego se suma a la línea correspondiente de la línea media en el sumador 1161b para el decodificador. En el codificador, se realiza la misma operación pero se hace una resta en el elemento 2034b.

- 10 Cabe señalar que la información de control 1003 puede indicar además que se utilicen más tramas que las dos tramas circundantes o que, por ejemplo, solo se utilice la trama actual y exactamente una o más tramas precedentes pero que no se utilicen tramas "futuras" para reducir el retardo sistemático.

Además, debe observarse que la combinación ponderada respecto de la etapa ilustrada en la figura 10b, en la cual, en una primera operación, las líneas de una trama se combinan y, posteriormente, los resultados de estas operaciones de combinación respecto de la trama se combinan entre sí, también puede realizarse en el otro orden. El otro orden significa que, en un primer paso, las líneas para la frecuencia actual k a partir de un número de tramas adyacentes indicadas por la información de control 103 se combinan mediante una combinación ponderada. Esta combinación ponderada se realiza para las líneas k , $k-1$, $k-2$, $k+1$, $k+2$ etc., según el número de líneas adyacentes a utilizar para estimar la línea imaginaria. Luego, los resultados de estas combinaciones "respecto del tiempo" se someten a una combinación ponderada en la "dirección de frecuencia" para finalmente obtener la línea imaginaria k para la trama f_i . Las ponderaciones se establecen para tener un valor entre -1 y 1 , preferentemente, y las ponderaciones pueden implementarse en una combinación de filtro IIR o FIR sencilla que realiza una combinación lineal de líneas espectrales o señales espectrales a partir de frecuencias diferentes y tramas diferentes.

25 Como se indica en las figuras 6a y 6b, el algoritmo de transformada preferido es el algoritmo de transformada MDCT que se aplica en la dirección hacia delante en los elementos 50 y 51 en la figura 6a y que se aplica en la dirección hacia atrás en los elementos 52, 53, posterior a la operación de combinación en el combinador 1162 que funciona en el dominio espectral.

30 La figura 8a ilustra una implementación más detallada del bloque 50 o 51. Particularmente, una secuencia de muestras de audio en el dominio del tiempo se ingresa en un modo de ventana de análisis 500 que realiza una operación de modo de ventana utilizando una ventana de análisis y, particularmente, realiza esta operación trama por trama pero utilizando un avance o superposición del 50 %. El resultado del modo de ventana de análisis, es decir, una secuencia de tramas de muestras en modo de ventana, se ingresa en un bloque de transformada MDCT 501, que produce la secuencia de tramas de MDCT con valor real, donde estas tramas están afectadas por solapamiento. De manera ilustrativa, el modo de ventana de análisis aplica ventanas de análisis que tienen una longitud de 2048 muestras. Luego, el bloque de transformada MDCT 501 produce espectros de MDCT que tienen 1024 líneas espectrales reales o valores de MDCT. Preferentemente, el modo de ventana de análisis 500 y/o el transformador de MDCT 501 se controlan mediante una longitud de ventana o control de longitud de transformada 502 de manera que, por ejemplo, para porciones transitorias en la señal, la longitud de ventana/longitud de transformada se reduce para obtener mejores resultados de codificación.

45 La figura 8b ilustra la operación de MDCT inversa realizada en los bloques 52 y 53. De manera ilustrativa, el bloque 52 comprende un bloque 520 para realizar una transformada MDCT inversa trama por trama. Cuando, por ejemplo, una trama de valores de MDCT tiene 1024 valores, entonces la salida de esta transformada inversa MDCT tiene 2048 muestras de tiempo afectadas por solapamiento. Dicha trama se proporciona a un modo de ventana de síntesis 521, que aplica una ventana de síntesis a esta trama de 2048 muestras. La trama en modo de ventana luego se envía a un procesador de superposición/suma 522 que, de manera ejemplar, aplica un 50% de superposición entre dos tramas subsiguientes y, luego, realiza una suma muestra por muestra de manera que un bloque de 2048 muestras finalmente tiene como resultado 1024 muestras nuevas de la señal de salida libre de solapamiento. Nuevamente, se prefiere aplicar un control de longitud de ventana/transformada utilizando información que es, por ejemplo, transmitida en la información lateral de la señal multicanal codificada como se indica en 523.

55 Los valores de predicción α pueden calcularse para cada línea espectral individual de un espectro de MDCT. Sin embargo, se ha descubierto que esto no es necesario y que puede ahorrarse una cantidad considerable de información lateral realizando un cálculo respecto de la banda de la información de predicción. Expresado de manera diferente, un convertidor espectral 50 ilustrado en la figura 9 que es, por ejemplo, un procesador de MDCT como se describe en el contexto de la figura 8a proporciona un espectro de resolución de alta frecuencia que tiene ciertas líneas espectrales ilustradas en la figura 9b. Este espectro de resolución de alta frecuencia es utilizado por un selector de línea espectral 90 que proporciona un espectro de resolución de baja frecuencia que comprende ciertas bandas B_1 , B_2 , B_3 , ..., B_N . Este espectro de resolución de baja frecuencia se envía al optimizador 207 para calcular la información de predicción de manera que no se calcula una información de predicción para cada línea

espectral sino solo para cada banda. Con este fin, el optimizador 207 recibe las líneas espectrales por banda y calcula la operación de optimización que comienza de la suposición de que el mismo valor α se utiliza para todas las líneas espectrales en la banda.

5 Preferentemente, las bandas se forman de manera psicoacústica de manera que el ancho de banda de las bandas aumenta desde frecuencias más bajas a frecuencias más altas como se ilustra en la figura 9b. De manera alternativa, aunque no preferida como la implementación de ancho de banda creciente, también podrían utilizarse bandas de frecuencia de igual tamaño, donde cada banda de frecuencia tiene por lo menos dos o por lo general muchas más, como por lo menos 30, líneas de frecuencia. Por lo general, para un espectro de 1024 líneas
10 espectrales, se calculan menos de 30 valores α complejos y, preferentemente, más de 5 valores α . Para espectros con menos de 1024 líneas espectrales (por ejemplo, 128 líneas), preferentemente, se utilizan menos bandas de frecuencia (por ejemplo, 6) para α .

15 Para calcular los valores α no se requiere necesariamente el espectro de MDCT de alta resolución. De manera alternativa, también puede utilizarse un banco de filtros que tiene una resolución de frecuencia similar a la resolución requerida para calcular los valores α . Cuando se deben implementar bandas con frecuencia creciente, entonces este banco de filtros debe tener ancho de banda variable. Sin embargo, cuando es suficiente un ancho de banda constante de frecuencias bajas a altas, entonces puede utilizarse un banco de filtros tradicional con subbandas de igual ancho.

20 Según la implementación, puede invertirse el signo del valor α indicado en la figura 3b o 4b. Sin embargo, para mantener la coherencia, es necesario que esta inversión del signo se utilice en el lado del codificador y también en el lado del decodificador. En comparación con la figura 6a, la figura 5a ilustra una vista generalizada del codificador, donde el ítem 2033 es un predictor que se controla mediante la información de control del predictor 206, que se
25 determina en el ítem 207 y que se inserta como información lateral en el flujo de bits. En lugar de la MDCT utilizada en la figura 6a en los bloques 50, 51, se utiliza una transformada de tiempo/frecuencia generalizada en la figura 5a según se describe. Como se describió anteriormente, la figura 6a es el proceso del codificador que corresponde al proceso del decodificador en la figura 6b, donde L significa señal de canal izquierdo, R significa señal de canal derecho, M significa señal media o señal de mezcla descendente, S significa señal lateral y D significa señal residual. De manera alternativa, también se denomina L a la primera señal de canal 201, también se denomina R a la segunda señal de canal 202, también se denomina M a la primera señal de combinación 204 y también se
30 denomina S a la segunda señal de combinación 2032.

35 Preferentemente, los módulos 2070 en el codificador y 1160a en el decodificador deben coincidir de manera exacta para asegurar la codificación de forma de onda correcta. Esto se aplica preferentemente al caso en el cual estos módulos utilizan alguna forma de aproximación como filtros truncados o cuando solo se utiliza una o dos en lugar de las tres tramas de MDCT, es decir, la trama de MDCT actual en la línea 60, la trama de MDCT precedente en la línea 61 y la trama de MDCT siguiente en la línea 62.

40 Además, se prefiere que el módulo 2070 en el codificador en la figura 6a utilice el espectro de MDCT no cuantificado M como entrada, aunque el módulo real a imaginario (R2I) 1160a en el decodificador tenga disponible solo el espectro de MDCT cuantificado como entrada. De manera alternativa, uno puede también utilizar una implementación en la cual el codificador utiliza los coeficientes de MDCT cuantificados como una entrada en el módulo 2070. Sin embargo, desde el punto de vista perceptivo el enfoque preferido es utilizar el espectro de MDCT
45 no cuantificado como entrada en el módulo 2070.

Posteriormente, se discuten en mayor detalle varios aspectos de realizaciones de la presente invención.

50 La codificación con estéreo paramétrico estándar, como la codificación estéreo basada en MPEG Envolvente (MPS) en el sistema USAC, depende de la capacidad del dominio QMF (híbrido) complejo sobremuestreado para permitir el procesamiento de señal motivado perceptivamente variable en tiempo y frecuencia sin introducir artefactos de solapamiento. Sin embargo, en caso de la codificación de mezcla descendente/residual (como la utilizada para las altas velocidades de transferencia de bits consideradas en el presente), el codificador estéreo unificado resultante actúa como un codificador de forma de onda.

55 Esto permite el funcionamiento en un dominio muestreado críticamente, como el dominio de MDCT, dado que el paradigma de codificación de forma de onda garantiza que la propiedad de cancelación de solapamiento de la cadena de procesamiento de MDCT-IMDCT se preserve suficientemente bien.

60 Sin embargo, para poder aprovechar la eficacia de codificación mejorada que puede lograrse en el caso de las señales estéreo con diferencias de tiempo o fase entre canales por medio de un coeficiente de predicción con valor complejo α , se requiere una representación en el dominio de la frecuencia con valor complejo de la señal de mezcla descendente DMX como entrada en la matriz de mezcla ascendente con valor complejo. Esto puede obtenerse

utilizando una transformada MDST además de la transformada MDCT para la señal DMX. El espectro de MDST puede calcularse (de manera exacta o como una aproximación) a partir del espectro de MDCT.

Asimismo, la parametrización de la matriz de mezcla ascendente puede simplificarse transmitiendo el coeficiente de predicción complejo α en lugar de los parámetros de MPS. Por lo tanto, se transmiten solo dos parámetros (parte real e imaginaria de α) en lugar de tres (ICC, CLD e IPD). Esto es posible debido a la redundancia en la parametrización de MPS en el caso de la codificación de mezcla descendente/residual. La parametrización de MPS incluye información sobre la cantidad relativa de decorrelación a agregar en el decodificador (es decir, la relación de energía entre las señales RES y DMX) y esta información es redundante cuando se transmiten las señales DMX y RES reales.

Por el mismo motivo, un factor de ganancia es obsoleto en el caso de la codificación de mezcla descendente/residual. Por lo tanto, la matriz de mezcla ascendente para la codificación de mezcla descendente/residual con predicción compleja ahora es:

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} 1-\alpha & 1 \\ 1+\alpha & -1 \end{bmatrix} \begin{bmatrix} DMX \\ RES \end{bmatrix}.$$

En comparación con la Ecuación 1169 en la figura 4b, el signo de α se invierte en esta ecuación y $DMX=M$ y $RES=D$. Esta es, por lo tanto, una implementación/anotación alternativa con respecto a la figura 4b.

Hay dos opciones para calcular la señal residual de predicción en el codificador. Una opción es utilizar los valores espectrales de MDCT cuantificados de la mezcla descendente. Esto tendría como resultado la misma distribución de error de cuantificación que en la codificación M/S dado que el codificador y el decodificador utilizan los mismos valores para generar la predicción. La otra opción es utilizar los valores espectrales de MDCT no cuantificados. Esto implica que el codificador y el decodificador no utilizarán los mismos datos para generar la predicción, lo cual permite la redistribución espacial del error de codificación de acuerdo con las propiedades de enmascaramiento instantáneas de la señal a costa de una ganancia de codificación algo reducida.

Es preferible calcular el espectro de MDST directamente en el dominio de la frecuencia por medio del filtrado FIR bidimensional de tres tramas de MDCT adyacentes como se describe. Esto último puede considerarse una transformada "real a imaginaria" (R2I). La complejidad del cálculo en el dominio de la frecuencia de la MDST puede reducirse de diferentes maneras, lo cual significa que se calcula solo una aproximación del espectro de MDST:

- Limitar el número de etapas de filtro FIR.
- Estimar la MDST a partir de la trama de MDCT actual solamente.
- Estimar la MDST a partir de la trama de MDCT actual y previa.

Las propiedades de codificación de forma de onda no se ven afectadas siempre que se utilice la misma aproximación en el codificador y el decodificador. Dichas aproximaciones del espectro de MDST, sin embargo, pueden causar una reducción en la ganancia de codificación lograda por la predicción compleja.

Si el codificador de MDCT subyacente admite el cambio de forma de ventana, los coeficientes del filtro FIR bidimensional utilizados para calcular el espectro de MDST deben adaptarse a las formas de ventana reales. Los coeficientes de filtro aplicados al espectro de MDCT de la trama actual dependen de la ventana completa, es decir, se requiere un conjunto de coeficientes para cada tipo de ventana para cada transición de ventana. Los coeficientes de filtro aplicados al espectro de MDCT de la trama previa/siguiente dependen solo de la ventana semisuperpuesta a la trama actual, es decir, para estos se requiere un conjunto de coeficientes solo para cada tipo de ventana (ningún coeficiente adicional para transiciones).

Si el codificador de MDCT subyacente utiliza el cambio de longitud de transformada, que incluye la trama de MDCT previa y/o siguiente en la aproximación se vuelve más complicada alrededor de las transiciones entre las longitudes de transformada diferentes. Debido al número diferente de coeficientes de MDCT en la trama actual y previa/siguiente, el filtro bidimensional es más complicado en este caso. Para evitar mayor complejidad computacional y estructural, puede excluirse la trama previa/siguiente del filtrado en transiciones de longitud de transformada, a costa de la precisión reducida de la aproximación para las respectivas tramas.

Asimismo, deben cuidarse especialmente las partes más inferiores y superiores del espectro de MDST (ceranas a CC y $fs/2$), donde hay menos coeficientes de MDCT circundantes para el filtrado FIR que los requeridos. Aquí el proceso de filtrado necesita adaptarse para calcular el espectro de MDST de manera correcta. Esto puede hacerse

utilizando una extensión simétrica del espectro de MDCT para los coeficientes faltantes (de acuerdo con la periodicidad de los espectros de señales discretas en el tiempo) o adaptando coeficientes de filtro de acuerdo con eso. El manejo de estos casos especiales puede simplificarse desde luego a costa de una precisión reducida en los alrededores de los límites del espectro de MDST.

5 Calcular el espectro de MDST exacto a partir de los espectros de MDCT transmitidos en el decodificador aumenta el retardo del codificador en una trama (que aquí se supone que es 1024 muestras). El retardo adicional puede evitarse utilizando una aproximación del espectro de MDST que no requiere el espectro de MDCT de la trama siguiente como entrada.

10 La siguiente lista con viñetas resume las ventajas de la codificación estéreo unificada basada en MDCT en comparación con la codificación estéreo unificada basada en QMF:

- 15 • Aumento solamente pequeño en complejidad computacional (cuando no se utiliza SBR).
- Escalas hasta la reconstrucción perfecta si los espectros de MDCT no se cuantifican. Obsérvese que esto no es así para la codificación estéreo unificada basada en QMF.
- 20 • Unificación y extensión naturales de codificación M/S y codificación estéreo de intensidad.
- Arquitectura más clara que simplifica la sintonización del codificador, dado que el procesamiento y la cuantificación/codificación de señales estéreo pueden estar fuertemente asociadas. Obsérvese que en la codificación estéreo unificada basada en QMF, las tramas de MPS y las tramas de MDCT no se alinean y las bandas de factor de escala no coinciden con las bandas de parámetros de MPS.
- 25 • La codificación eficiente de parámetros estéreo, dado que deben transmitirse solo dos parámetros (α complejo) en lugar de tres parámetros como en MPEG Envolvente (ICC, CLD, IPD).
- 30 • Ningún retardo adicional del decodificador si el espectro de MDST se calcula como una aproximación (sin utilizar la trama siguiente).

Las propiedades importantes de una implementación pueden resumirse de la siguiente manera:

- 35 a) Los espectros de MDST se calculan por medio de filtrado FIR bidimensional a partir de los espectros de MDCT actual, previo y siguiente. Diferentes equilibrios de complejidad/calidad para el cálculo de MDST (aproximación) son posibles reduciendo el número de etapas de filtro FIR y/o el número de tramas de MDCT utilizadas. En particular, si no hay una trama adyacente debido a la pérdida de trama durante la transmisión o cambio de longitud de transformada, esa trama en particular se excluye de la estimación de MDST. Para el caso de cambio de longitud de transformada la exclusión se señala en el flujo de bits.
- 40 b) Se transmiten solo dos parámetros, la parte real e imaginaria del coeficiente de predicción complejo α , en lugar de ICC, CLD e IPD. Las partes real e imaginaria de α se manejan independientemente, limitadas al rango $[-3,0; 3,0]$ y cuantificadas con un tamaño de paso de 0,1. Si un parámetro determinado (parte real o imaginaria de α) no está utilizándose en una trama dada, esto se señala en el flujo de bits y el parámetro irrelevante no se transmite. Los parámetros se codifican de manera diferencial respecto del tiempo o de manera diferencial respecto de la frecuencia y finalmente se aplica la codificación de Huffman utilizando un libro de códigos de factor de escala. Los coeficientes de predicción se actualizan cada segunda banda de factor de escala, lo cual tiene como resultado una resolución de frecuencia similar a la de MPEG Envolvente. Este esquema de cuantificación y codificación tiene como resultado una velocidad de transferencia de bits promedio de aproximadamente 2 kb/s para la información lateral estéreo dentro de una configuración típica que tiene una velocidad de transferencia de bits objetivo de 96 kb/s.
- 50

Los detalles de la implementación adicional o alternativa preferida comprenden:

- 55 c) Para cada uno de los dos parámetros de α , se puede elegir codificación no diferencial (PCM) o diferencial (DPCM) por trama o por flujo, señalada mediante un bit correspondiente en el flujo de bits. Para la codificación DPCM, es posible la codificación diferencial respecto de tiempo o frecuencia. Nuevamente, esto puede señalarse utilizando una bandera de un bit.
- 60 d) En lugar de reutilizar un libro de códigos predefinido como el libro de factor de escala de AAC, se puede también utilizar un libro de códigos no variante dedicado o adaptativo a señal para codificar los valores de parámetro α , o se puede recurrir a palabras de codificación complemento de dos o no señaladas de longitud fija (por ejemplo, 4 bits).

e) El rango de valores del parámetro α y el tamaño de paso de cuantificación de parámetro pueden elegirse arbitrariamente y optimizarse para las características disponibles de la señal.

f) El número y el ancho espectral y/o temporal de las bandas de parámetro α activas pueden elegirse arbitrariamente y optimizarse para las características de señal dadas. En particular, la configuración de banda puede señalarse por trama o por flujo.

g) Además o en lugar de los mecanismos descritos en a), anteriormente, puede señalarse explícitamente por medio de un bit por trama en el flujo de bits que solo se utiliza el espectro de MDCT de la trama actual para calcular la aproximación de espectro de MDST, es decir, que no se tienen en cuenta las tramas de MDCT adyacentes.

Las realizaciones se relacionan con un sistema inventivo para codificación estéreo unificada en el dominio de MDCT. Permite utilizar las ventajas de la codificación estéreo unificada en el sistema MPEG USAC incluso en velocidades de transferencia de bits más altas (cuando no se utiliza SBR) sin el aumento significativo de complejidad computacional que traería un enfoque basado en QMF.

Las dos listas siguientes resumen los aspectos de configuración preferidos descritos anteriormente, que pueden utilizarse alternativamente entre sí o además de otros aspectos:

1a) concepto general: predicción de valores complejos de MDCT lateral a partir de MDCT y MDST medias;

1b) calcular/aproximar MDST a partir de MDCT ("transformada R2I") en el dominio de la frecuencia utilizando una o más tramas (enfoque de 3 tramas introduce retardo);

1c) truncamiento de filtro (incluso hasta 1 trama 2 etapas, por ejemplo [-1 0 1]) para reducir la complejidad computacional;

1d) manejo adecuado de coeficientes de transformada alrededor de CC y $f_s/2$;

1e) manejo adecuado de conmutación de forma de ventana;

1f) no utilizar la trama previa/siguiente si tiene un tamaño de transformada diferente;

1g) predicción basada en coeficientes de MDCT no cuantificados o cuantificados en el codificador;

2a) cuantificar y codificar la parte real e imaginaria del coeficiente de predicción complejo directamente (es decir, parametrización de MPEG Envolverte);

2b) utilizar cuantificador uniforme para este (tamaño de paso por ejemplo 0,1);

2c) utilizar resolución de frecuencia apropiada para coeficientes de predicción (por ejemplo, 1 coeficiente cada 2 bandas de factor de escala);

2d) señalización económica en caso de que todos los coeficientes de predicción sean con valor real;

2e) bit explícito por trama para provocar la operación R2I de 1 trama, es decir, no utilizar la trama previa/siguiente.

En una realización, el codificador adicionalmente comprende: un convertidor espectral (50, 51) para convertir una representación en el dominio del tiempo de las dos señales de canal en una representación espectral de las dos señales de canal que tienen señales de subbanda para las dos señales de canal, en el cual el combinador (2031), el predictor (2033) y la calculadora de señal residual (2034) se configuran para procesar cada señal de subbanda de forma separada de manera que la primera señal combinada y la señal residual se obtienen para una pluralidad de subbandas, donde la interfaz de salida (212) se configura para combinar la primera señal combinada codificada y la señal residual codificada para la pluralidad de subbandas.

Aunque algunos aspectos se han descrito en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo se corresponde con un paso del método o una característica de un paso del método. De manera análoga, los aspectos descritos en el contexto de un paso del método también representan una descripción de un bloque o ítem o característica correspondiente de un aparato correspondiente.

En una realización de la presente invención, se aplica un manejo adecuado de la conmutación de forma de ventana.

Cuando se considera la figura 10a, se puede ingresar una información de forma de ventana 109 en la calculadora de espectro imaginario 1001. Específicamente, la calculadora de espectro imaginario que lleva a cabo la conversión de real a imaginario del espectro con valor real como el espectro MDCT (como el elemento 2070 en la figura 6a o el elemento 1160a en la figura 6b) puede implementarse como un filtro FIR o IIR. Los coeficientes FIR o IIR en este módulo real a imaginario 1001 dependen de la forma de ventana de la mitad izquierda y de la mitad derecha de la trama actual. Esta forma de ventana puede ser diferente para una ventana sinusoidal o una ventana KBD (Derivada de Kaiser Bessel) y, dependiendo de la configuración de secuencia de ventana dada, puede ser una ventana larga, una ventana de inicio, una ventana de parada, una ventana de inicio-parada o una ventana corta. El módulo real a imaginario puede comprender un filtro FIR bidimensional, en el cual una dimensión es la dimensión del tiempo donde dos tramas MDCT posteriores se ingresan en el filtro FIR, y la otra dimensión es la dimensión de la frecuencia, donde se ingresan los coeficientes de frecuencia de una trama.

La siguiente tabla ofrece distintos coeficientes de filtro MDST para una secuencia de ventana actual para distintas formas de ventana y distintas implementaciones de la mitad izquierda y la mitad derecha de la ventana.

Tabla A – Parámetros de Filtro MDST para Ventana Actual

Secuencia de Ventana Actual	Mitad Izq.: Forma Sinusoidal Mitad Der.: Forma Sinusoidal	Mitad Izq.: Forma KBD Mitad Der.: Forma KBD
ONLY_LONG_SEQUENCE, EIGHT_SHORT_SEQUENCE	[0,000000, 0,000000, 0,500000, 0,000000, -0,500000, 0,000000, 0,000000]	[0,091497, 0,000000, 0,581427, 0,000000, -0,581427, 0,000000, - 0,091497]
LONG_START_SEQUENCE	[0,102658, 0,103791, 0,567149, 0,000000, -0,567149, -0,103791, - 0,102658]	[0,150512, 0,047969, 0,608574, 0,000000, -0,608574, -0,047969, - 0,150512]
LONG_STOP_SEQUENCE	[0,102658, -0,103791, 0,567149, 0,000000, -0,567149, 0,103791, - 0,102658]	[0,150512, -0,047969, 0,608574, 0,000000, -0,608574, 0,047969, - 0,150512]
STOP_START_SEQUENCE	[0,205316, 0,000000, 0,634298, 0,000000, -0,634298, 0,000000, - 0,205316]	[0,209526, 0,000000, 0,635722, 0,000000, -0,635722, 0,000000, - 0,209526]
Secuencia de Ventana Actual	Mitad Izq.: Forma Sinusoidal Mitad Der.: Forma KBD	Mitad Izq.: Forma KBD Mitad Der.: Forma Sinusoidal
ONLY_LONG_SEQUENCE, EIGHT_SHORT_SEQUENCE	[0,045748, 0,057238, 0,540714, 0,000000, -0,540714, -0,057238, - 0,045748]	[0,045748, -0,057238, 0,540714, 0,000000, -0,540714, 0,057238, - 0,045748]
LONG_START_SEQUENCE	[0,104763, 0,105207, 0,567861, 0,000000, -0,567861, -0,105207, - 0,104763]	[0,148406, 0,046553, 0,607863, 0,000000, -0,607863, -0,046553, - 0,148406]
LONG_STOP_SEQUENCE	[0,148406, -0,046553, 0,607863, 0,000000, -0,607863, 0,046553, - 0,148406]	[0,104763, -0,105207, 0,567861, 0,000000, -0,567861, 0,105207, - 0,104763]
STOP_START_SEQUENCE	[0,207421, 0,001416, 0,635010, 0,000000, -0,635010, -0,001416, - 0,207421]	[0,207421, -0,001416, 0,635010, 0,000000, -0,635010, 0,001416, - 0,207421]

Asimismo, la información de forma de ventana 109 proporciona información de forma de ventana para la ventana anterior, cuando la ventana anterior se utiliza para calcular el espectro MDST a partir del espectro MDCT. Los coeficientes de filtro MDST correspondientes para la ventana anterior se dan en la siguiente tabla como una función de la forma y secuencia de ventana actual.

Tabla B – Parámetros de Filtro MDST para Ventana Anterior

Secuencia de Ventana Actual	Mitad Izquierda de Ventana Actual: Forma Sinusoidal	Mitad Izquierda de Ventana Actual: Forma KBD
ONLY_LONG_SEQUENCE, LONG_START_SEQUENCE, EIGHT_SHORT_SEQUENCE	[0,000000, 0,106103, 0,250000, 0,318310, 0,250000, 0,106103, 0,000000]	[0,059509, 0,123714, 0,186579, 0,213077, 0,186579, 0,123714, 0,059509]
LONG_STOP_SEQUENCE, STOP_START_SEQUENCE	[0,038498, 0,039212, 0,039645, 0,039790, 0,039645, 0,039212, 0,038498]	[0,026142, 0,026413, 0,026577, 0,026631, 0,026577, 0,026413, 0,026142]

Por lo tanto, dependiendo de la información de forma de ventana 109, la calculadora de espectro imaginario 1001 en la figura 10a se adapta mediante la aplicación de distintos conjuntos de coeficientes de filtro.

La información de forma de ventana que se usa en el lado del decodificador se calcula en el lado del codificador y se transmite como información lateral junto con la señal de salida del codificador. En el lado del decodificador, la información de forma de ventana 109 se extrae del flujo de bits mediante el desmultiplexor de flujo de bits (por ejemplo 102 en la figura 5b) y se proporciona a la calculadora de espectro imaginario 1001 como se ilustra en la figura 10a.

Cuando la información de forma de ventana 109 indica que la trama anterior tenía un tamaño de transformada distinto, es preferible que la trama anterior no se use para calcular el espectro imaginario a partir del espectro con valor real. Sucede lo mismo cuando se encuentra mediante la interpretación de la información de forma de ventana 109 que la trama siguiente tiene un tamaño de transformada distinto. Entonces, la trama siguiente no se usa para calcular el espectro imaginario a partir del espectro con valor real. En dicho caso, cuando, por ejemplo, la trama anterior tenía un tamaño de transformada distinto de la trama actual y cuando la trama siguiente tiene nuevamente un tamaño de transformada distinto en comparación con la trama actual, entonces solo la trama actual, es decir, los valores espectrales de la ventana actual, se utilizan para calcular el espectro imaginario.

La predicción en el codificador se basa en coeficientes de frecuencia no cuantificados o cuantificados como los coeficientes MDCT. Cuando la predicción ilustrada por el elemento 2033 en la figura 3a, por ejemplo, se basa en datos no cuantificados, entonces la calculadora residual 2034 preferentemente también opera sobre datos no cuantificados y la señal de salida de la calculadora residual, es decir, la señal residual 205, se cuantifica antes de ser codificada entrópicamente y transmitida a un decodificador. Sin embargo, en una realización alternativa, es preferible que la predicción esté basada en coeficientes MDCT cuantificados. Por lo tanto, la cuantificación puede darse antes del combinador 2031 en la figura 3a de manera que un primer canal cuantificado y un segundo canal cuantificado sean la base para calcular la señal residual. De manera alternativa, la cuantificación también puede realizarse posterior al combinador 2031 de manera que la primera señal de combinación y la segunda señal de combinación se calculen de forma no cuantificada y se cuantifiquen antes de calcular la señal residual. Nuevamente, de manera alternativa, el predictor 2033 puede funcionar en el dominio no cuantificado y la señal de predicción 2035 se cuantifica antes de ser ingresada en la calculadora residual. Entonces, es útil que la segunda señal de combinación 2032, que también se ingresa en la calculadora residual 2034, también se cuantifique antes de que la calculadora residual calcule la señal residual D en la figura 6a, que puede implementarse dentro del predictor 2033 en la figura 3a, opere sobre los mismos datos cuantificados que están disponibles en el lado del decodificador. Por consiguiente, puede garantizarse que el espectro MDST calculado en el codificador con el fin de realizar el cálculo de la señal residual sea exactamente el mismo que el espectro MDST en el lado del decodificador utilizado para realizar la predicción inversa, es decir, para calcular la señal lateral a partir de la señal residual. Con este fin, la primera señal de combinación como la señal M en la línea 204 en la figura 6a se cuantifica antes de ser ingresada en el bloque 2070. Entonces, el espectro MDST calculado mediante el uso del espectro MDCT cuantificado de la trama actual y, dependiendo de la información de control, el espectro MDCT cuantificado de la trama anterior o siguiente se ingresa en el multiplicador 2074, y el producto del multiplicador 2074 de la figura 6a será nuevamente un espectro no cuantificado. Este espectro no cuantificado será restado del ingreso de espectro en el sumador 2034b y el resultado será finalmente cuantificado en el cuantificador 209b.

En una realización, la parte real y la parte imaginaria del coeficiente de predicción complejo por banda de predicción se cuantifican y codifican directamente, es decir, sin por ejemplo la parametrización MPEG Envolvente. La cuantificación puede realizarse usando un cuantificador uniforme con un tamaño de paso, por ejemplo, de 0,1. Esto significa que cualquier tamaño de paso de cuantificación logarítmica o similar no se aplica aunque sí se aplica cualquier tamaño de paso lineal. En una implementación, el rango de valor para la parte real y la parte imaginaria del coeficiente de predicción complejo va desde -3 hasta 3, lo cual significa que se utilizan 60 o, dependiendo de los detalles de la implementación, 61 pasos de cuantificación para la parte real y la parte imaginaria del coeficiente de predicción complejo.

Preferentemente, la parte real aplicada en el multiplicador 2073 en la figura 6a y la parte imaginaria 2074 aplicada en la figura 6a se cuantifican antes de aplicarse de manera que, nuevamente, se utilice el mismo valor para la predicción en el lado del codificador que está disponible en el lado del decodificador. Esto garantiza que la señal residual de predicción cubra – además del error de cuantificación introducido – cualquier error que pueda producirse cuando un coeficiente de predicción no cuantificado se aplica en el lado del codificador mientras que un coeficiente de predicción cuantificado se aplica en el lado del decodificador. Preferentemente, la cuantificación se aplica de modo tal que – en la medida de lo posible – la misma situación y las mismas señales estén disponibles en el lado del codificador y en el lado del decodificador. Por lo tanto, es preferible cuantificar la entrada a la calculadora real a imaginario 2070 usando la misma cuantificación que se aplica en el cuantificador 209a. De manera adicional, es preferible cuantificar la parte real y la parte imaginaria del coeficiente de predicción α para realizar las multiplicaciones en el ítem 2073 y el ítem 2074. La cuantificación es la misma que la que se aplica en el

cuantificador 2072. De manera adicional, la señal lateral que sale del bloque 2031 en la figura 6a también puede cuantificarse antes de los sumadores 2034a y 2034b. Sin embargo, la realización de la cuantificación por parte del cuantificador 209b posterior a la suma donde la suma por parte de estos sumadores se aplica con una señal lateral no cuantificada no es problemática.

5 En una realización adicional de la presente invención, se aplica una señalización barata en caso de que todos los coeficientes de predicción sean reales. Se puede dar que todos los coeficientes de predicción para una determinada trama, es decir, para la misma porción de tiempo de la señal de audio, se calculen como reales. Dicha situación puede producirse cuando la señal media completa y la señal lateral completa no tengan desplazamiento de fase
10 entre sí o solo lo tengan levemente. Con el fin de ahorrar bits, un indicador real individual indica esta situación. Luego, la parte imaginaria del coeficiente de predicción no necesita señalizarse en el flujo de bits con una palabra de código que represente un valor cero. En el lado del decodificador, la interfaz de decodificador de flujo de bits, como un desmultiplexor de flujo de bits, interpretará este indicador real y, por consiguiente, no buscará palabras de código para una parte imaginaria sino que asumirá que todos los bits que se encuentran en la sección correspondiente del flujo de bits representan solamente los coeficientes de predicción con valor real. Asimismo, el predictor 2033, cuando reciba una indicación de que todas las partes imaginarias de los coeficientes de predicción en la trama son cero, no necesitará calcular un espectro MDST, o generalmente un espectro imaginario a partir del espectro MDCT con valor real. Por lo tanto, el elemento 1160a en el decodificador de la figura 6b se desactivará y la predicción inversa solo se dará usando el coeficiente de predicción con valor real aplicado en el multiplicador 1160b en la figura 6b. Lo mismo
20 ocurrirá en el lado del codificador donde el elemento 2070 se desactivará y la predicción solo se dará usando el multiplicador 2073. Esta información lateral se usa preferentemente como un bit por trama adicional y el decodificador leerá este bit trama por trama a fin de decidir si el convertidor real a imaginario 1160a estará activo para una trama o no. Por lo tanto, proporcionar esta información tiene como consecuencia un tamaño reducido del flujo de bits debido a la señalización más eficaz de todas las partes imaginarias de los coeficientes de predicción que son cero para una trama y, además, reduce en gran medida la complejidad para el decodificador para una trama así que inmediatamente da como resultado un menor consumo de la batería de dicho procesador implementado, por ejemplo, en un dispositivo móvil a batería.

30 La predicción estéreo compleja de acuerdo con realizaciones preferentes de la presente invención es una herramienta para la codificación eficaz de pares de canales con diferencias de nivel y/o fase entre los canales. Mediante el uso de un parámetro con valor complejo α , los canales izquierdo y derecho se reconstruyen mediante la siguiente matriz. dmx_{Im} denota la MDST correspondiente a la MDCT del canal de mezcla descendente dmx_{Re} .

$$\begin{bmatrix} r \\ l \end{bmatrix} = \begin{bmatrix} 1 - \alpha_{Re} & -\alpha_{Im} & 1 \\ 1 + \alpha_{Re} & \alpha_{Im} & -1 \end{bmatrix} \begin{bmatrix} dmx_{Re} \\ dmx_{Im} \\ res \end{bmatrix}$$

35 La ecuación anterior es otra representación, la cual está dividida respecto de la parte real y la parte imaginaria de α y representa la ecuación para una operación combinada de predicción/combinación, en la cual no necesariamente se calcula la señal predicha S.

40 Los siguientes elementos de datos se usan preferentemente para esta herramienta:

cplx_pred_all	0: Algunas bandas usan codificación L/R, según lo señala cplx_pred_used[] 1: Todas las bandas usan predicción estéreo compleja
cplx_pred_used[g][sfb]	Bandera de un bit por grupo de ventana g y banda de factor de escala sfb (posterior al mapeo a partir de bandas de predicción) indicando que 0: no se está usando predicción compleja, se está usando codificación L/R 1: se está usando predicción compleja
complex_coef	0: $\alpha_{Im} = 0$ para todas las bandas de predicción (predicción solo real) 1: α_{Im} se transmite para todas las bandas de predicción
use_prev_frame	0: Se usa solo la trama actual para la estimación de MDST 1: Se usa la trama actual y la anterior para la estimación de MDST
delta_code_time	0: Codificación de diferencial de frecuencia de los coeficientes de predicción 1: Codificación de diferencial de tiempo de los coeficientes de predicción
hcod_alpha_q_re	Código de Huffman de α_{Re}
hcod_alpha_q_im	Código de Huffman de α_{Im}

La figura 13a ilustra otro elemento de datos, sobre el cual se basa la presente invención, es decir, el indicador de dirección de predicción pred_dir. Este elemento de datos indica la dirección de la predicción de acuerdo con la tabla

en la figura 13a. Por lo tanto, un primer valor 0 significa una predicción de canal medio a lateral y un segundo valor como un valor "1" significa una predicción de canal lateral a medio.

Estos elementos de datos se calculan en un codificador y se ponen en la información lateral de una señal de audio estéreo o multicanal. Los elementos se extraen de la información lateral en el lado del decodificador mediante un extractor de información lateral y se utilizan para controlar la calculadora del decodificador a fin de realizar una acción correspondiente.

La predicción estéreo compleja requiere el espectro MDCT de mezcla descendente del par de canales actual y, en el caso de $\text{complex_coef} = 1$, una estimación del espectro MDST de mezcla descendente del par de canales actual, es decir, la contraparte imaginaria del espectro MDCT. La estimación de MDST de mezcla descendente se computa a partir de la mezcla descendente de MDCT de la trama actual y, en el caso de $\text{use_prev_frame} = 1$, la mezcla descendente de MDCT de la trama anterior. La mezcla descendente de MDCT de la trama anterior del grupo de ventana g y la ventana de grupo b se obtiene a partir de los espectros izquierdo y derecho reconstruidos de esa trama.

La computación de la estimación de MDST de mezcla descendente depende de la transformada MDCT, cuya longitud es pareja, en window_sequence , así como también en filter_coefs y filter_coefs_prev , que son conjuntos que contienen los núcleos de filtro y que se derivan de acuerdo con las tablas anteriores. Para todos los coeficientes de predicción la diferencia respecto de un valor precedente (en tiempo o frecuencia) se codifica usando un libro de códigos de Huffman. Los coeficientes de predicción no se transmiten para bandas de predicción para las cuales $\text{cplx_pred_used} = 0$.

Los coeficientes de predicción cuantificados inversos α_{re} y α_{im} son dados por

$$\alpha_{\text{re}} = \alpha_{\text{q_re}} * 0,1$$

$$\alpha_{\text{im}} = \alpha_{\text{q_im}} * 0,1$$

Sin la dirección de predicción se pueden producir problemas de inversión cuando la señal lateral S tiene una energía bastante alta en comparación con la señal de mezcla descendente M. En estos casos, puede ser difícil predecir la parte dominante de la señal presente en S, especialmente cuando M tiene un nivel muy bajo y, por consiguiente, consiste principalmente en componentes de ruido.

Además, el rango de valores para el coeficiente de predicción α puede ser muy amplio, lo cual lleva potencialmente a artefactos de codificación debido a la amplificación o paneo no deseado del ruido de cuantificación (por ejemplo, efectos de desenmascaramiento espacial).

A modo de ejemplo, se puede considerar una señal desfasada levemente paneada con $R = -0,9 \cdot L$

$$R = -0,9 \cdot L;$$

$$M = 0,5 \cdot (L + R) = 0,05 \cdot L;$$

$$S = 0,5 \cdot (L - R) = 0,95 \cdot L;$$

$$RES = S - (\alpha * M);$$

α óptimo :

$$\alpha = 19;$$

que lleva a un factor de predicción óptimo bastante grande de 19.

De acuerdo con la presente invención, la dirección de predicción se conmuta y esto da como resultado un incremento de la ganancia de predicción con mínimo esfuerzo computacional y un menor α .

En el caso de una señal lateral S con alta energía en comparación con la señal media M, se vuelve interesante revertir la dirección de la predicción de manera que M se prediga a partir de la representación de valor complejo de S como se ilustra, por ejemplo, en la figura 13b(2). Al conmutar la dirección de predicción, de manera que M se prediga a partir de S, se necesita preferentemente una MDST adicional para S pero no se requiere una MDST para M. Además, en este caso, en lugar de la señal media como en la primera alternativa en la figura 13b(1), la señal lateral (con valor real) tiene que transmitirse al decodificador junto con la señal residual y la información de predicción α .

La conmutación de la dirección de predicción se realiza sobre una base por trama, es decir, sobre el eje del tiempo.

Con este fin, se proporciona la calculadora de dirección de predicción 219, la cual se ilustra en la figura 12a. Como en otras figuras, la figura 12a ilustra una etapa de MDCT 50/51, una etapa de codificación media/lateral 2031, un convertidor real a complejo 2070, una calculadora de predicción de señal 2073/2074 y una calculadora de señal residual final 2034. Asimismo, se proporciona un intercambiador M/S de control de dirección de predicción 507 que está configurado y es útil para implementar las dos reglas de predicción diferentes 502, 503 que se ilustran en la figura 11a. La primera regla de predicción es que el intercambiador 507 esté en el primer estado, es decir, donde M y S no están intercambiadas. La segunda regla de predicción se implementa cuando el intercambiador 507 se encuentra en el estado de intercambio, es decir, cuando M y S se intercambian de la entrada a la salida. Esta implementación tiene la ventaja de que todo el circuito detrás del intercambiador 507 es el mismo para ambas direcciones de predicción.

De manera similar, las distintas reglas de decodificación 402, 403, es decir, las distintas reglas de cálculo del decodificador, también pueden implementarse mediante un intercambiador 407 en la entrada del combinador 1162 que, en la realización de la figura 12b, se implementa para realizar una codificación media/lateral inversa. El intercambiador 407 que también puede denominarse "conmutador de predicción" recibe, en su entrada, la señal de mezcla descendente DMX y una señal IPS, donde IPS significa señal predicha inversamente. Dependiendo del indicador de dirección de predicción, el intercambiador 407 conecta la DMX a M y la IPS a S o conecta la DMX a S y la IPS a M según se ilustra en la tabla que se encuentra arriba de la figura 12b.

La figura 13b ilustra una implementación de la primera regla de cálculo de la figura 11b, es decir, la regla ilustrada por el bloque 402. En la primera realización, la predicción inversa se realiza explícitamente de manera que la señal lateral se calcula explícitamente a partir de la señal residual y la señal media transmitida. Luego, en un paso posterior, L y R se calculan mediante las ecuaciones a la derecha de la ecuación de predicción inversa explícita en la figura 13. En una implementación alternativa, se realiza una predicción inversa implícita, donde la señal lateral S no se calcula explícitamente pero donde la señal izquierda L y la señal derecha R se calculan directamente a partir de la señal M transmitida y la señal residual transmitida usando la información de predicción α .

La figura 13d ilustra las ecuaciones para la otra dirección de predicción, es decir, cuando el indicador de dirección de predicción pred_dir es igual a 1. Nuevamente, una predicción inversa explícita para obtener M puede realizarse usando la señal residual transmitida y la señal lateral transmitida y se puede hacer un cálculo posterior de L y R usando la señal media y la señal lateral. De manera alternativa, se puede realizar una predicción inversa implícita de manera que se calculan L y R a partir de la señal transmitida S, la señal residual y la información de predicción α sin calcular explícitamente la señal media M.

Según se describe debajo en la figura 13b, el signo de α puede invertirse en todas las ecuaciones. Cuando se realiza esto, la figura 13b tiene, para el cálculo de la señal residual, una suma entre los dos términos. Entonces, la predicción inversa explícita se convierte en un cálculo de diferencia. Dependiendo de la implementación real, puede ser conveniente la notación descrita en las figura 13b a figura 13d o la notación inversa.

En las ecuaciones de la figura 13b a figura 13d, pueden producirse diversas multiplicaciones complejas. Estas multiplicaciones complejas pueden producirse para todos los casos, donde α es un número complejo. Por lo tanto, la aproximación compleja de M o S se requiere según se especifica en las ecuaciones. La multiplicación compleja provocará una diferencia entre la multiplicación real de la parte real de los dos factores y el producto de las partes imaginarias de los dos factores según se ilustra en la figura 13e para el caso de α solamente o para el caso de $(1 + \alpha)$.

La calculadora de dirección de predicción 219 puede implementarse de diversas maneras. La figura 14 ilustra dos formas básicas para calcular la dirección de predicción. Una forma es el cálculo de alimentación hacia delante, donde la señal M y la señal S, que son por lo general la primera señal de combinación y la segunda señal de combinación, se comparan mediante el cálculo de una diferencia de energía según lo indica el paso 550. Luego, en el paso 551 la diferencia se compara con un umbral, donde el umbral puede fijarse por medio de una línea de entrada de umbral o puede fijarse a un programa. Sin embargo, es preferible que haya algo de histéresis. Por consiguiente, como un criterio de decisión para la dirección de predicción real, se puede evaluar la diferencia de energía entre S y M. A fin de lograr la mejor calidad perceptiva, el criterio de decisión puede entonces estabilizarse mediante el uso de histéresis, es decir, distintos umbrales de decisión en base a la dirección de predicción de la última trama. Otro criterio concebible para la dirección de predicción sería la diferencia de fase entre canales de los canales de entrada. Respecto a la histéresis, el control del umbral puede realizarse de manera tal que se favorezcan escasos cambios de la dirección de predicción en un determinado intervalo de tiempo sobre numerosos cambios en este intervalo de tiempo. Por consiguiente, comenzando a partir de un determinado umbral, el umbral puede incrementarse en respuesta a un cambio de dirección de predicción. Luego, en base a este alto valor, el umbral puede reducirse cada vez más durante los períodos en los que no se calcula ningún cambio de dirección de

predicción. Luego, cuando el umbral se acerca a su valor antes del último cambio, el umbral permanece en el mismo nivel y el sistema está listo una vez más para cambiar la dirección de predicción. Este procedimiento permite cambios dentro de intervalos cortos solo cuando existe una gran diferencia entre S y M; sin embargo, permite cambios menos frecuentes cuando las diferencias de energía entre M y S no son tan altas.

5 De manera alternativa, o adicional, se puede realizar un cálculo de devolución, donde las señales residuales para ambas direcciones de predicción se calculan según se ilustra en el paso 552. Luego, en el paso 553, se calcula la dirección de predicción lo cual da como resultado una señal residual más pequeña o menos bits para la señal residual o la señal de mezcla descendente o una menor cantidad de bits generales o una mejor calidad de la señal
10 de audio o cualquier otra condición específica. Por consiguiente, la dirección de predicción que tiene como resultado un determinado objetivo de optimización se selecciona en este cálculo de devolución.

15 Debe tenerse en cuenta que la invención no solo se aplica a señales estéreo, es decir, señales multicanal que tienen solo dos canales, sino que también puede aplicarse a dos canales de una señal multicanal que tenga tres o más canales como una señal 5.1 o 7.1. Una realización para una implementación multicanal puede comprender la identificación de una pluralidad de pares de señales y el cálculo y transmisión paralela o almacenamiento de los datos para más de un par de señales.

20 En una realización del decodificador de audio, cada una de la primera señal de combinación codificada o decodificada 104 y la señal residual de predicción codificada o decodificada 106 comprende una primera pluralidad de señales subbanda, en la que la información de predicción comprende una segunda pluralidad de parámetros de información de predicción, siendo la segunda pluralidad menor que la primera pluralidad, en la que el predictor 1160 está configurado para aplicar el mismo parámetro de predicción a por lo menos dos señales subbanda distintas de la primera señal de combinación decodificada, en la que la calculadora del decodificador 116 o la calculadora de señal
25 de combinación 1161 o el combinador 1162 están configurados para realizar un procesamiento respecto de la subbanda; y en el que el decodificador de audio comprende además un banco de filtros de síntesis 52, 53 para combinar señales subbanda de la primera señal de combinación decodificada y la segunda señal de combinación decodificada para obtener una primera señal decodificada en el dominio del tiempo y una segunda señal decodificada en el dominio del tiempo.

30 En una realización del decodificador de audio, el predictor 1160 está configurado para recibir información de forma de ventana 109 y para usar distintos coeficientes de filtro para calcular un espectro imaginario, donde los distintos coeficientes de filtro dependen de distintas formas de ventana indicadas por la información de forma de ventana 109.

35 En una realización del decodificador de audio, la primera señal de combinación decodificada se asocia con distintas longitudes de transformada indicadas por un indicador de longitud de transformada incluido en la señal multicanal codificada 100, y donde el predictor 1160 está configurado para usar solamente una o más tramas de la primera señal de combinación que tiene la misma longitud de transformada asociada para estimar la parte imaginaria para una trama actual para una primera señal de combinación.

40 En una realización del decodificador de audio, el predictor 1160 está configurado para usar una pluralidad de subbandas de la primera señal de combinación decodificada adyacente en frecuencia, a fin de estimar la parte imaginaria de la primera señal de combinación, y donde, en caso de bajas o altas frecuencias, una extensión simétrica en frecuencia de la trama actual de la primera señal de combinación se utiliza para subbandas asociadas
45 con frecuencias menores o iguales a cero o mayores o iguales a una mitad de una frecuencia de muestreo sobre la cual se basa la trama actual, o donde los coeficientes de filtro de un filtro incluido en el predictor 1160a se fijan en distintos valores para las subbandas faltantes en comparación con las subbandas no faltantes.

50 En una realización del decodificador de audio, la información de predicción 108 se incluye en la señal multicanal codificada en una representación cuantificada y codificada entrópicamente, en el que el decodificador de audio comprende además un decodificador de información de predicción 65 para decodificar entrópicamente o descuantificar a fin de obtener información de predicción decodificada usada por el predictor 1160, o donde la señal de audio multicanal codificada comprende una unidad de datos que indica en el primer estado que el predictor 1160 va a utilizar por lo menos una trama que precede o sigue en el tiempo a una trama actual de la primera señal de combinación decodificada, e indica en el segundo estado que el predictor 1160 va a utilizar solamente una única
55 trama de la primera señal de combinación decodificada para una estimación de una parte imaginaria para la trama actual de la primera señal de combinación decodificada, y donde el predictor 1160 está configurado para detectar un estado de la unidad de datos y para funcionar en consecuencia.

60 En una realización del decodificador de audio, la información de predicción 108 comprende palabras de código de diferencias entre valores complejos adyacentes en la frecuencia o secuenciales en el tiempo, y en el que el decodificador de audio está configurado para llevar a cabo un paso de decodificación entrópica y un posterior paso de decodificación de diferencia para obtener valores de predicción complejos cuantificados secuenciales en el

tiempo o valores de predicción complejos para bandas de frecuencia adyacente.

En una realización del decodificador de audio, la señal multicanal codificada comprende, como información lateral, un indicador real que indica que todos los coeficientes de predicción para una trama de la señal multicanal codificada tienen valor real, donde el decodificador de audio está configurado para extraer el indicador real de la
5 señal de audio multicanal codificada 100, y en el que la calculadora del decodificador 116 está configurada para no calcular una señal imaginaria para una trama, para la cual el indicador real está indicando solo coeficientes de predicción con valor real.

En una realización del codificador de audio, el predictor 2033 comprende un cuantificador para cuantificar la primera señal de canal, la segunda señal de canal, la primera señal de combinación o la segunda señal de combinación a fin de obtener una o más señales cuantificadas, y en el que el predictor 2033 está configurado para calcular la señal residual mediante el uso de señales cuantificadas.

En una realización del codificador de audio, la primera señal de canal es una representación espectral de un bloque de muestras y la segunda señal de canal es una representación espectral de un bloque de muestras, en el que las representaciones espectrales son representaciones espectrales reales puras o representaciones espectrales imaginarias puras, donde el optimizador 207 está configurado para calcular la información de predicción 206 como un factor con valor real distinto de cero y/o como un factor imaginario distinto de cero, y donde la calculadora del
20 codificador 203 está configurada para calcular la primera señal de combinación y la señal residual de predicción de manera tal que la señal de predicción se derive de la representación espectral real pura o la representación espectral imaginaria pura usando el factor con valor real.

La señal de audio codificada inventiva puede almacenarse en un medio de almacenamiento digital o puede transmitirse en un medio de transmisión tal como un medio de transmisión inalámbrica o un medio de transmisión
25 alámbrica tal como Internet.

Aunque la presente invención se describe principalmente en el contexto de procesamiento de audio, debe enfatizarse que la invención también puede aplicarse a la codificación de señales de vídeo. La predicción compleja con dirección variable puede aplicarse a p. ej., la compresión de vídeo estéreo en 3D. En este ejemplo particular, se usa una MDCT en 2D. Un ejemplo de esta técnica es Google Web/VP8. Sin embargo, también pueden aplicarse otras implementaciones sin una MDCT en 2D.

Si bien se han descrito algunos aspectos en el contexto de un aparato, queda claro que estos aspectos representan además una descripción del método correspondiente, donde un bloque o dispositivo corresponde a un paso de método o una característica de un paso de método. De manera análoga, los aspectos descritos en el contexto de un paso de método también representan una descripción de un bloque o ítem o característica correspondiente de un aparato correspondiente.

Dependiendo de determinados requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede llevarse a cabo mediante el uso de un medio de almacenamiento digital, por ejemplo un disquete, un DVD, un CD, una memoria ROM, una memoria PROM, una memoria EPROM, una memoria EEPROM o una memoria FLASH, que tenga señales de control que puedan leerse electrónicamente almacenadas en el mismo, que cooperen (o que sean capaces de cooperar) con un sistema
45 informático programable de manera que se lleve a cabo el método respectivo.

Algunas realizaciones según la invención comprenden un portador de datos tangible o no transitorio que tiene señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de manera que se realiza uno de los métodos descritos en el presente documento.

En general, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo operativo el código de programa para realizar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse por ejemplo en un portador legible por máquina.

Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenado en un portador legible por máquina.

En otras palabras, una realización del método inventivo es, por tanto, un programa informático que tiene un código de programa para realizar uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador.

Una realización adicional de los métodos inventivos es, por tanto, un portador de datos (o un medio de

almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los métodos descritos en el presente documento.

5 Una realización adicional del método inventivo es, por tanto, un flujo de datos o un secuencia de señales que representa el programa informático para realizar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales puede configurarse por ejemplo para transferirse mediante una conexión de comunicación de datos, por ejemplo, mediante Internet.

10 Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los métodos que se describen en la presente.

Una realización adicional comprende un ordenador que tiene instalado el programa informático para llevar a cabo uno de los métodos que se describen en el presente documento.

15 En algunas realizaciones, un dispositivo lógico programable (por ejemplo una matriz de puertas de campo programable) puede utilizarse para llevar a cabo algunas o todas las funciones de los métodos que se describen en el presente documento. En algunas realizaciones, una matriz de puertas de campo programable puede cooperar con un microprocesador a fin de llevar a cabo uno de los métodos que se describen en el presente documento. Por lo general, los métodos se llevan a cabo preferentemente mediante cualquier aparato de hardware.

20 Las realizaciones descritas anteriormente son solamente ilustrativas de los principios de la presente invención. Cabe señalar que diversas modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para otros expertos en la técnica. Por consiguiente, solo estará limitada por el alcance de la reivindicación y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

25

REIVINDICACIONES

1. Decodificador de audio para decodificar una señal de audio estéreo codificada (100), comprendiendo la señal de audio estéreo codificada una primera señal de combinación codificada generada en base a una regla de combinación para combinar una primera señal de audio de canal y una segunda señal de audio de canal de una señal de audio estéreo, siendo la primera señal de combinación codificada una representación de la transformada discreta del coseno modificada (MDCT) con valor real de una señal de mezcla descendente generada mediante codificación media-lateral como la regla de combinación, siendo una señal residual de predicción codificada (106) una representación de la MDCT y comprendiendo los datos del flujo de bits un coeficiente de predicción complejo (108) y un indicador de dirección de predicción (501), comprendiendo el decodificador de audio:
 - un decodificador de señal (110) para decodificar la primera señal de combinación codificada (104) a fin de obtener una primera señal de combinación decodificada (112) que es una señal de mezcla descendente decodificada y para decodificar la señal residual de predicción codificada (106) a fin de obtener una señal residual decodificada (114); y
 - una calculadora de decodificador (116) para calcular una señal de audio estéreo decodificada que tiene una primera señal de canal decodificada (117) que es una señal de canal izquierdo de la señal de audio estéreo decodificada y una segunda señal de canal decodificada (118) que es una señal de canal derecho de la señal de audio estéreo decodificada usando la señal residual decodificada (114), el coeficiente de predicción complejo (108), la primera señal de combinación decodificada (112) y el indicador de dirección de predicción (501) que indica una dirección de predicción de una señal media a una señal lateral o de una señal lateral a una señal media, en donde la calculadora de decodificador (116) está configurada para:
 - generar (1160a) una señal de mezcla descendente compleja a partir de la primera señal de combinación decodificada (112),
 - multiplicar (1160b,c) la señal de mezcla descendente compleja y el coeficiente de predicción complejo para obtener un resultado, agregando (1161a,b) el resultado y la señal residual decodificada (114) para obtener una señal predicha inversamente, llevando a cabo (1162) una codificación media-lateral inversa, en la que, en un primer estado del indicador de dirección de predicción (501), la señal de mezcla descendente decodificada representa la señal media de la codificación media-lateral inversa y la señal predicha inversamente representa la señal lateral de la codificación media-lateral inversa y, en la que, en un segundo estado del indicador de dirección de predicción (501), la señal de mezcla descendente decodificada representa la señal lateral de la codificación media-lateral inversa y la señal predicha inversamente representa la señal media de la codificación media-lateral inversa, en donde un cambio de la dirección de predicción entre el primer estado y el segundo estado se realiza en un base por trama, y
 - llevar a cabo una MDCT inversa, IMDCT, hasta obtener un resultado de la codificación media-lateral inversa para obtener la señal de canal izquierdo de la señal de audio estéreo decodificada y la señal de canal derecho de la señal de audio estéreo decodificada.

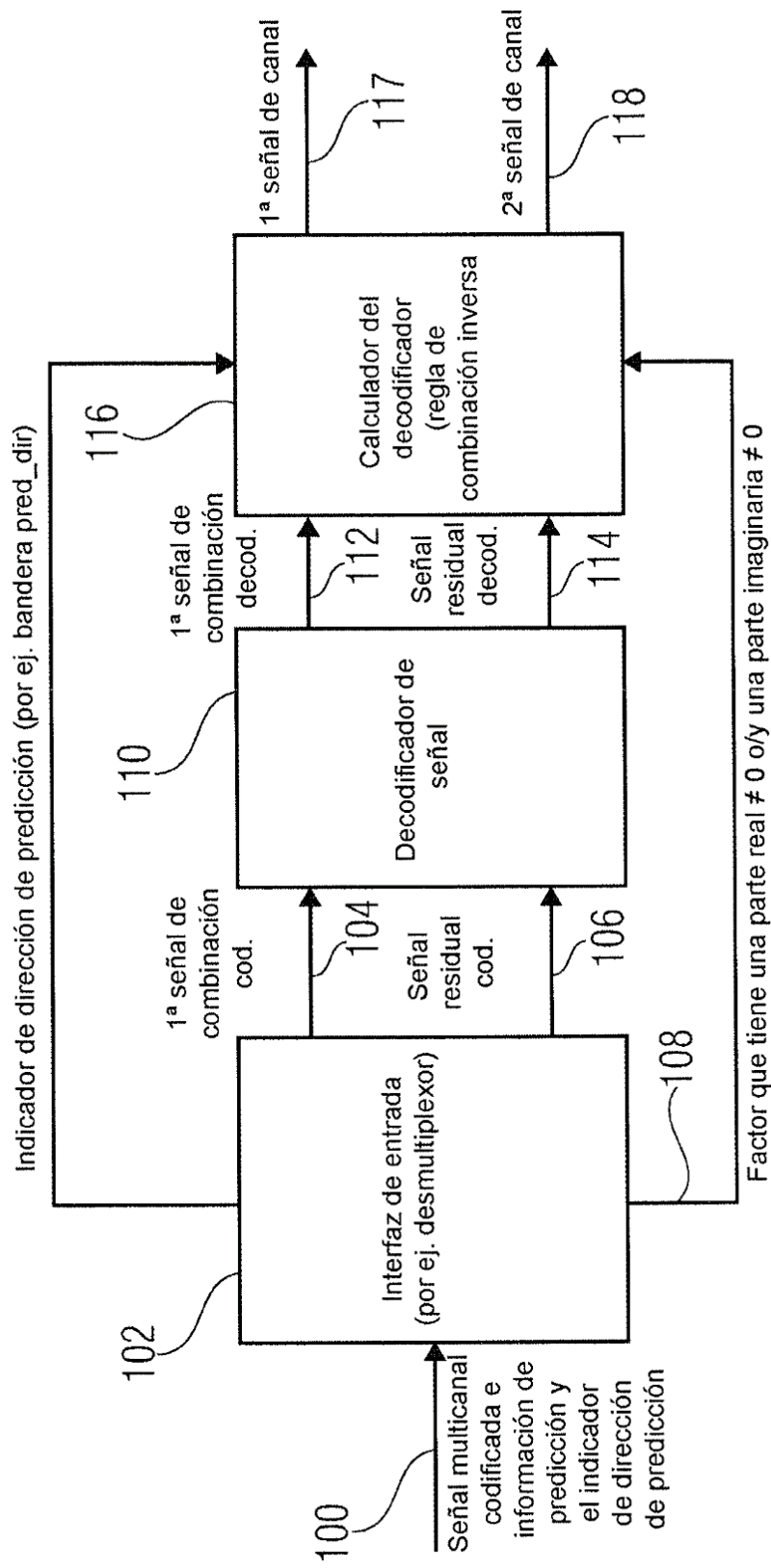


FIG 1

(Decodificador de Audio)

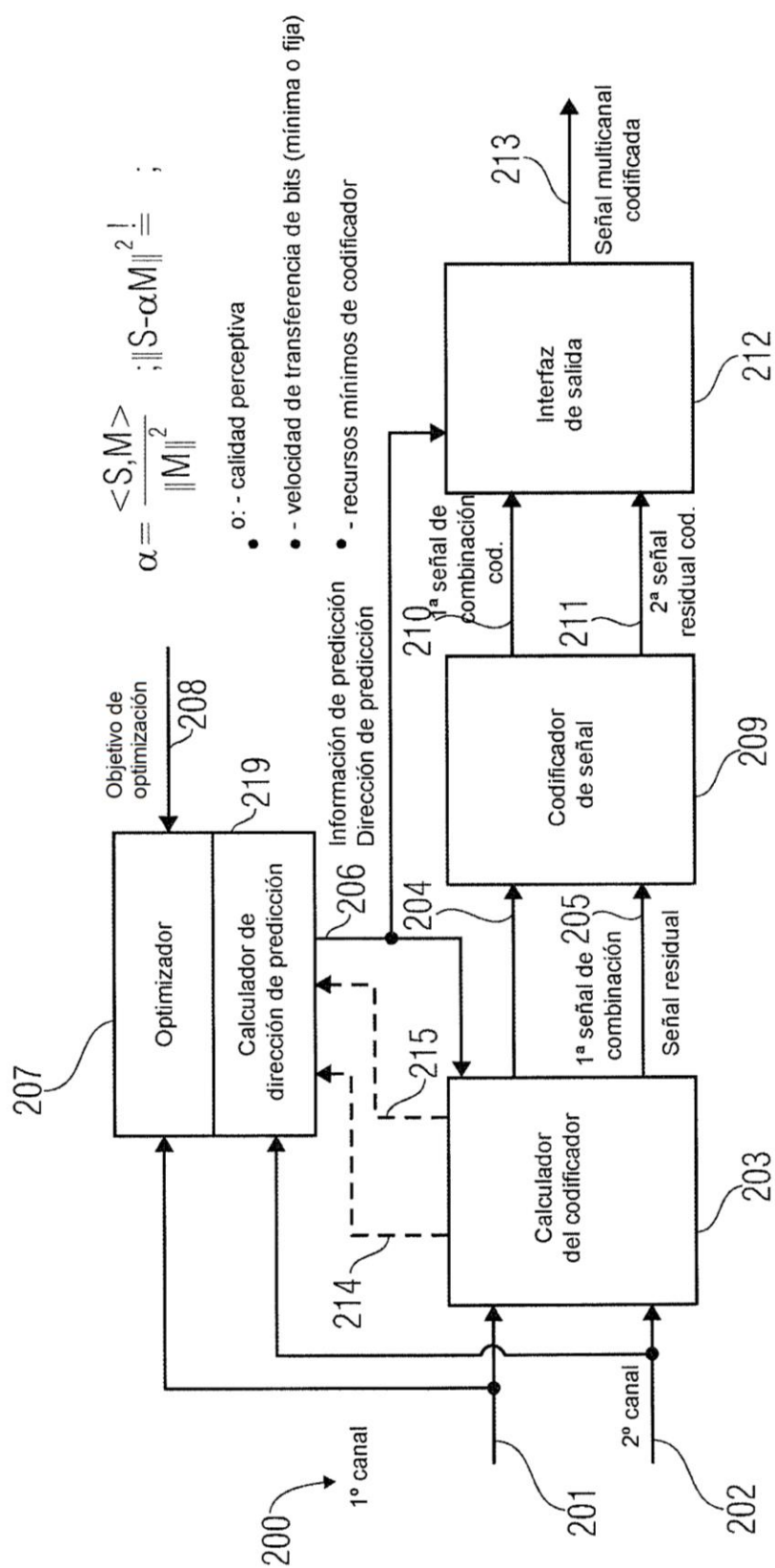


FIG 2

(Codificador de Audio)

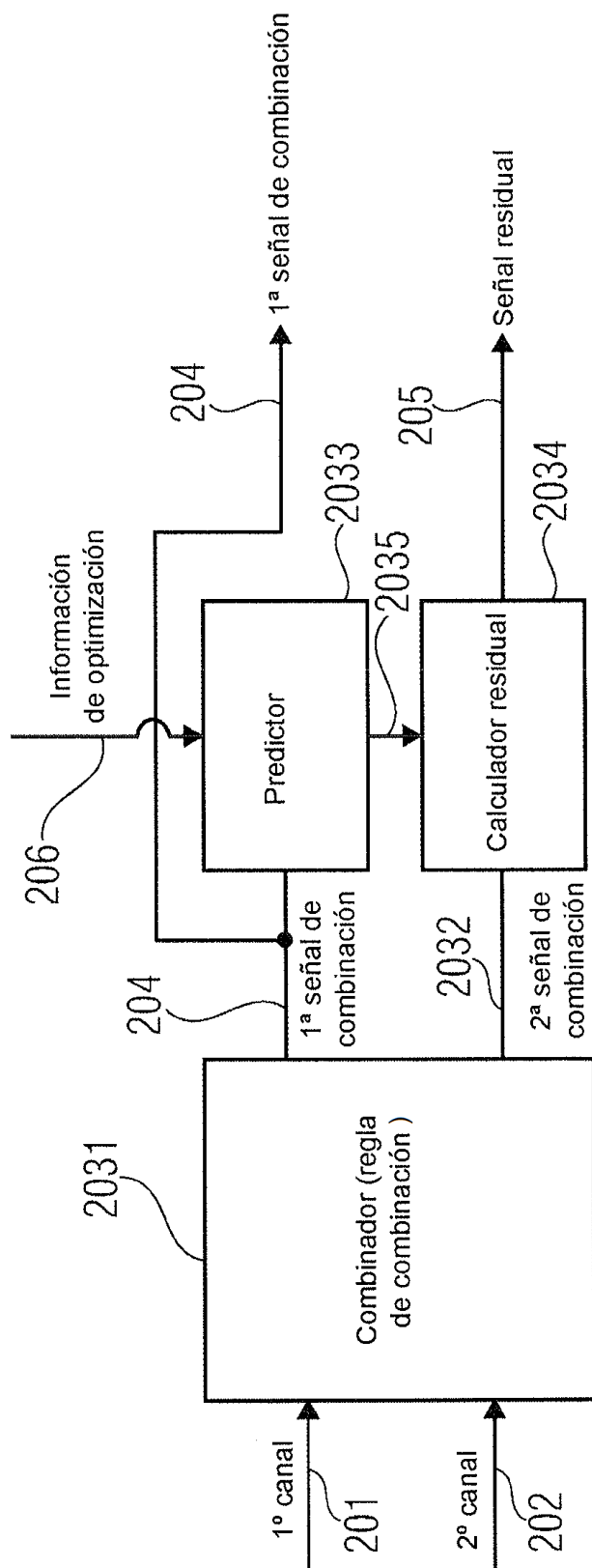
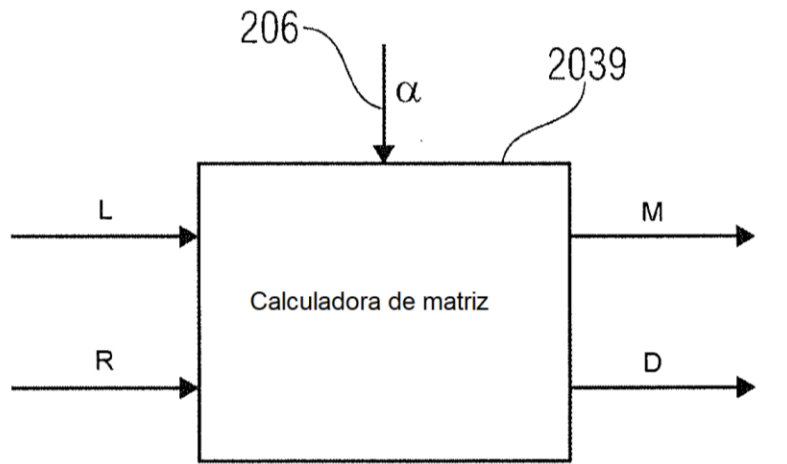


FIG 3A

(Lado del Codificador de Audio)



$$\alpha = RE_{\alpha} + jIm_{\alpha}$$

$$\alpha = |\alpha| \cdot e^{j\varphi_{\alpha}}$$

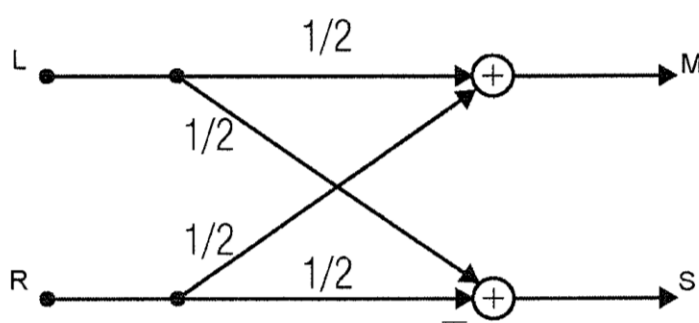
primera parte: RE_{α} $|\alpha|$

segunda parte: Im_{α} φ_{α}

$$\begin{bmatrix} M \\ D \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1-\alpha & -1-\alpha \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix}$$

FIG 3B

(Lado del Codificador de Audio)



$$M = \frac{1}{2} (L + R)$$

$$S = \frac{1}{2} (L - R)$$

FIG 3C

(Regla de Combinación)

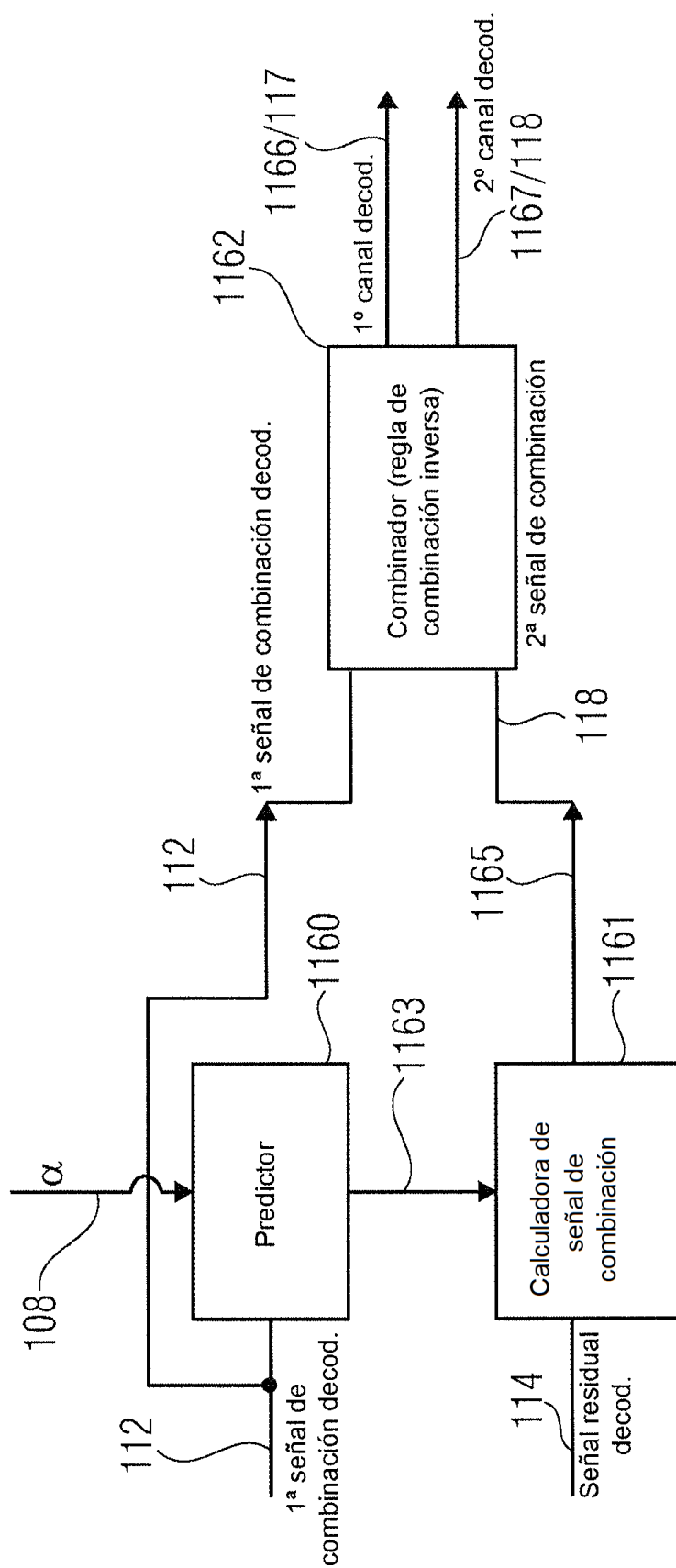


FIG 4A

(Lado del Decodificador de Audio)

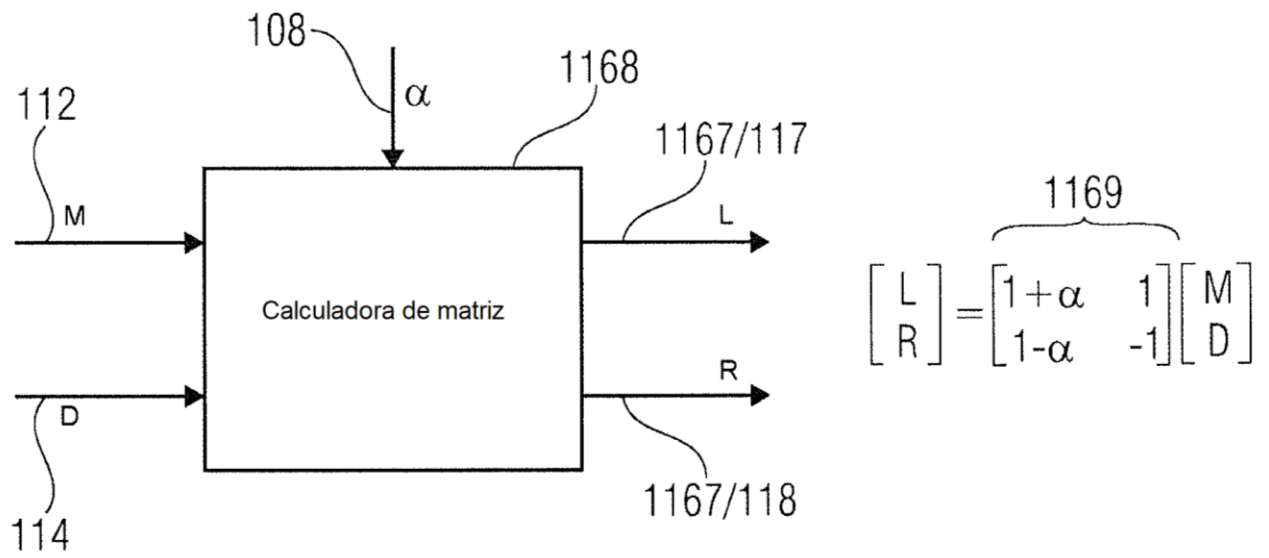


FIG 4B

(Lado del Decodificador de Audio)

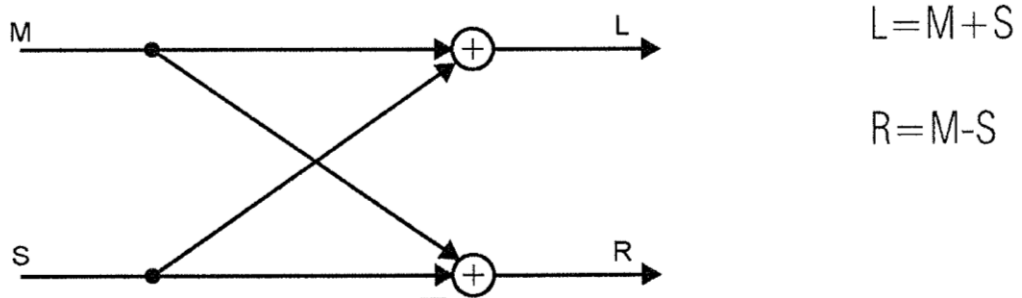


FIG 4C

(Regla de Combinación Inversa)

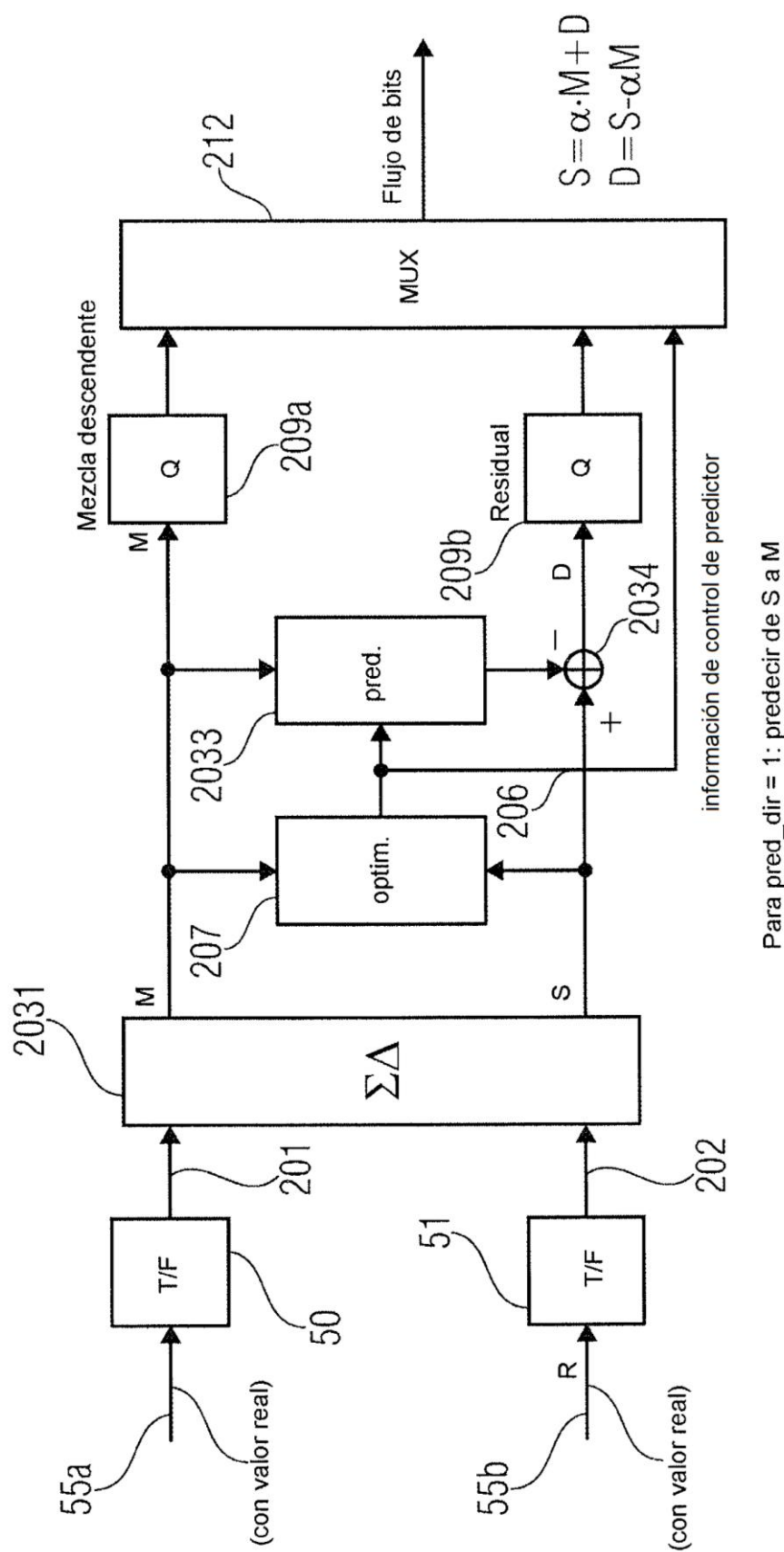


FIG 5A

(Lado del Codificador)

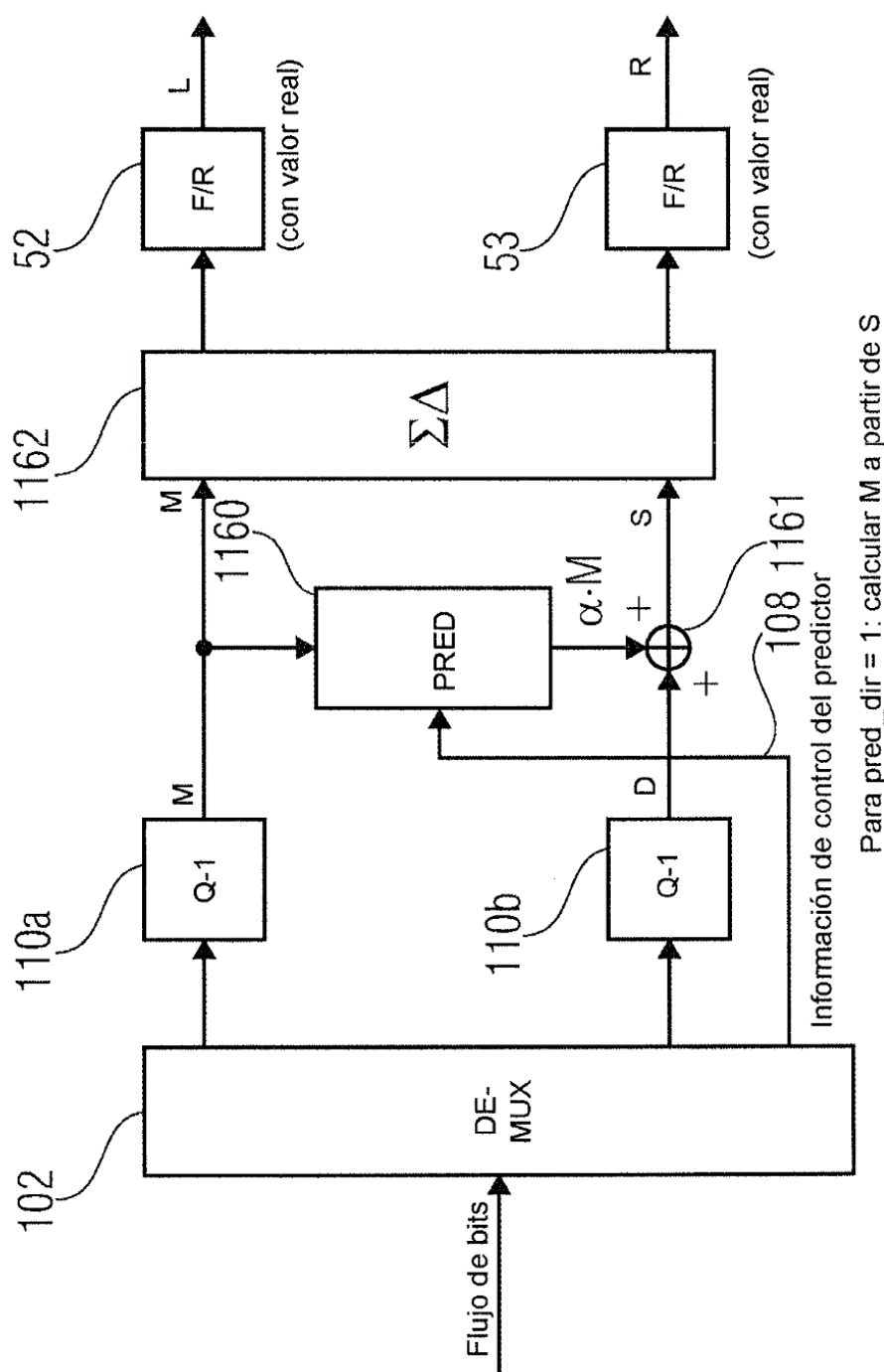


FIG 5B

(Lado del Decodificador)

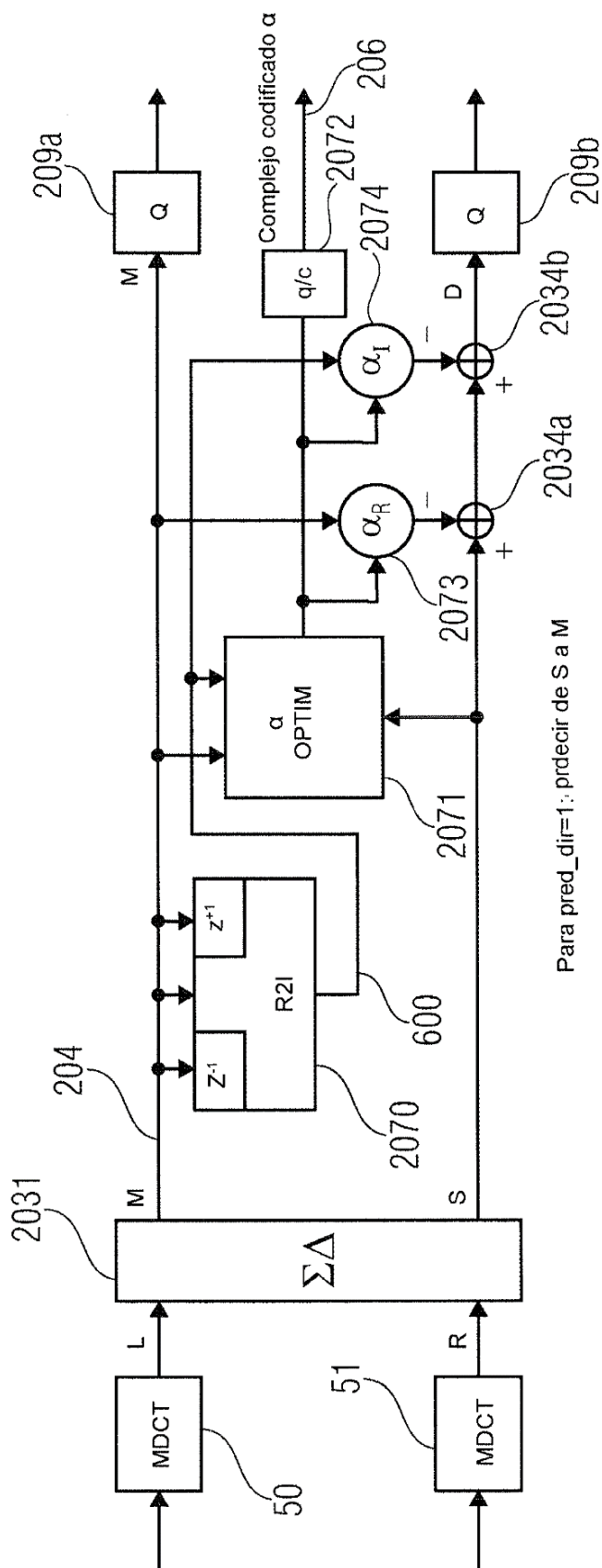


FIG 6A

(Lado del codificador)

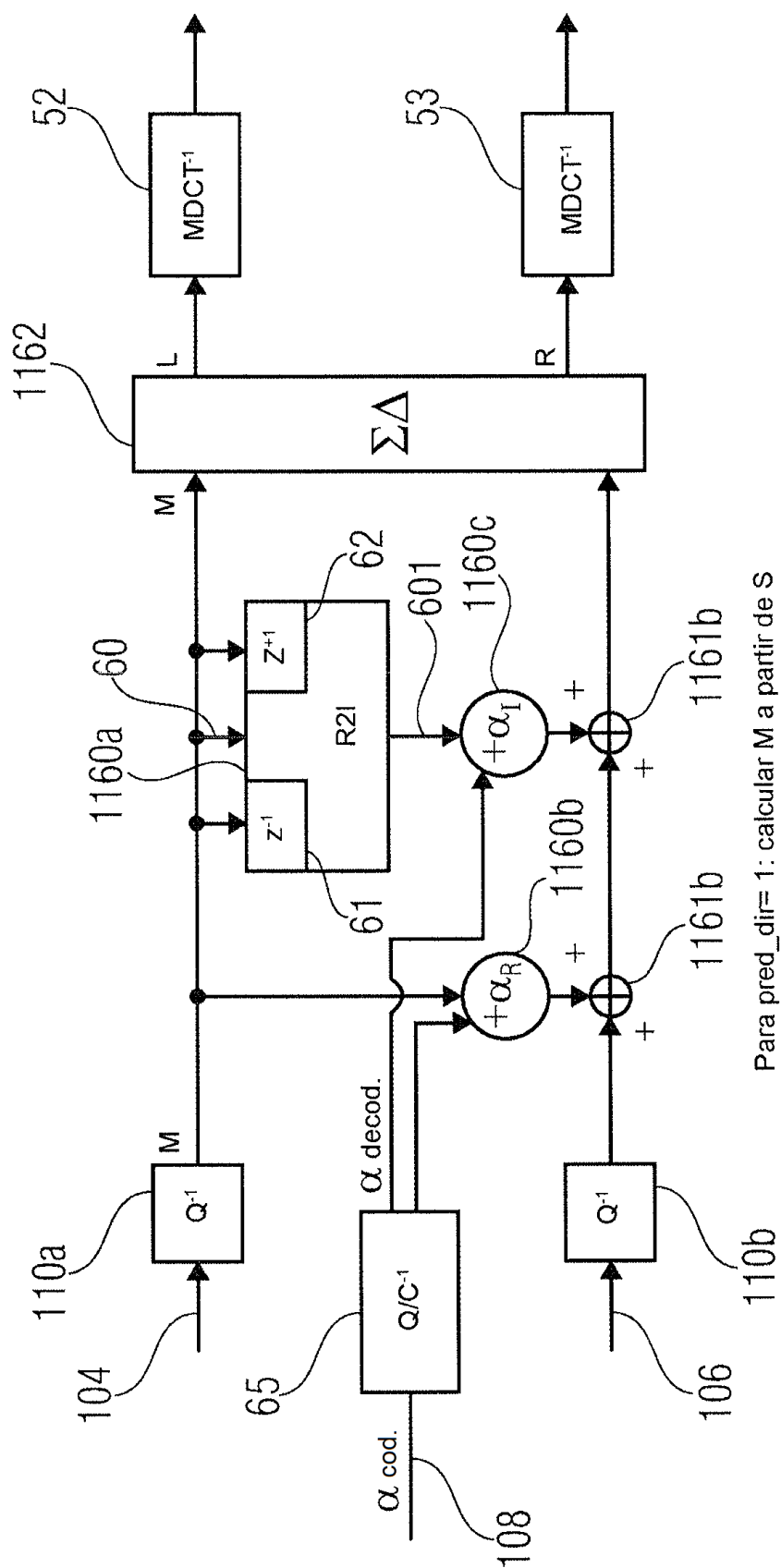


FIG 6B

(Lado del decodificador)

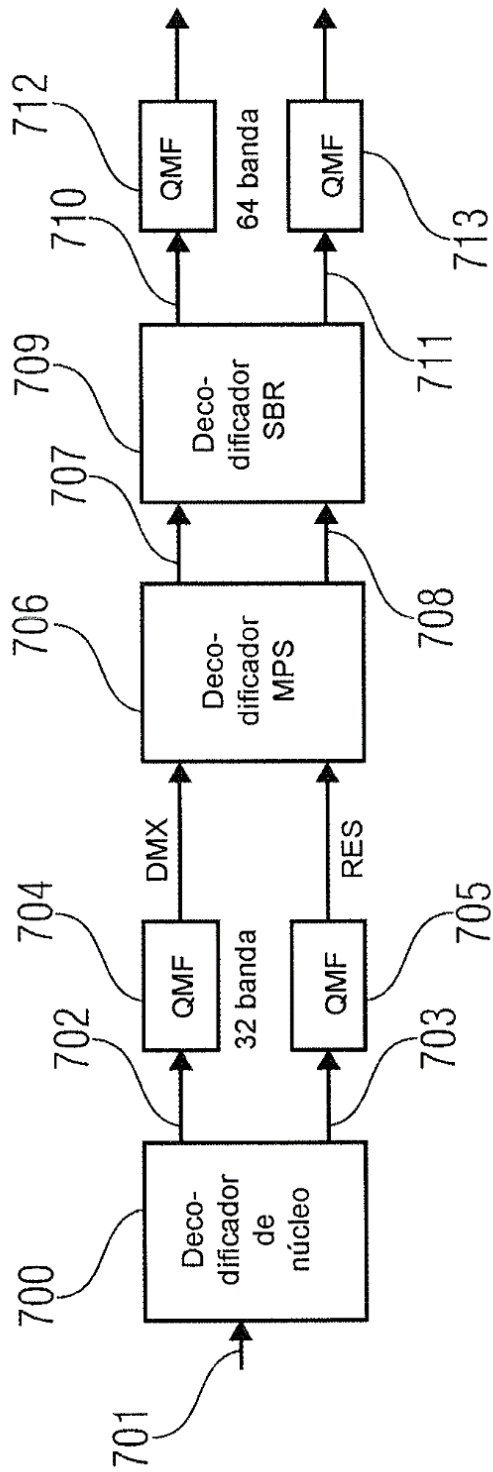


FIG 7A

$$\begin{bmatrix} L \\ R \end{bmatrix} = g \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} DMX \\ RES \end{bmatrix}$$

FIG 7B

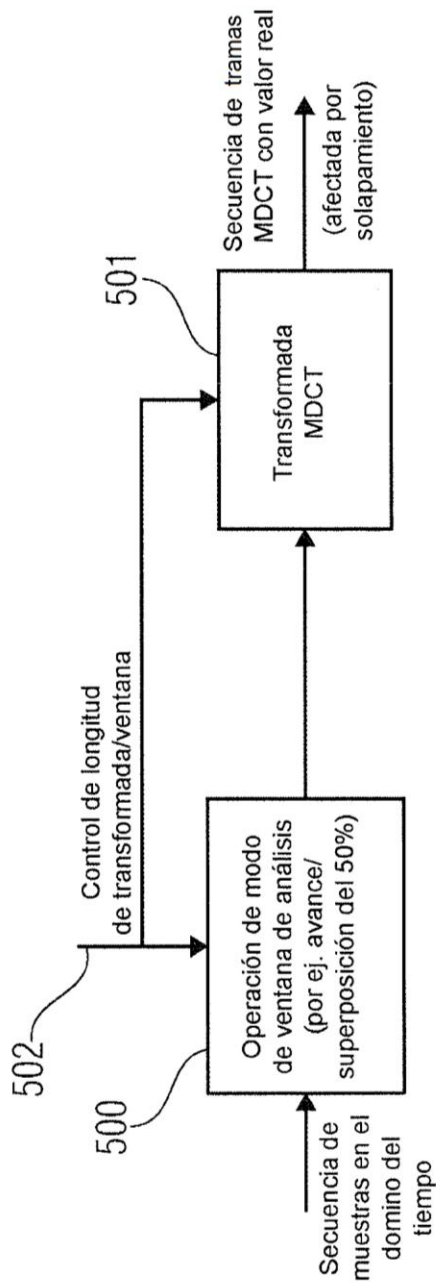


FIG 8A

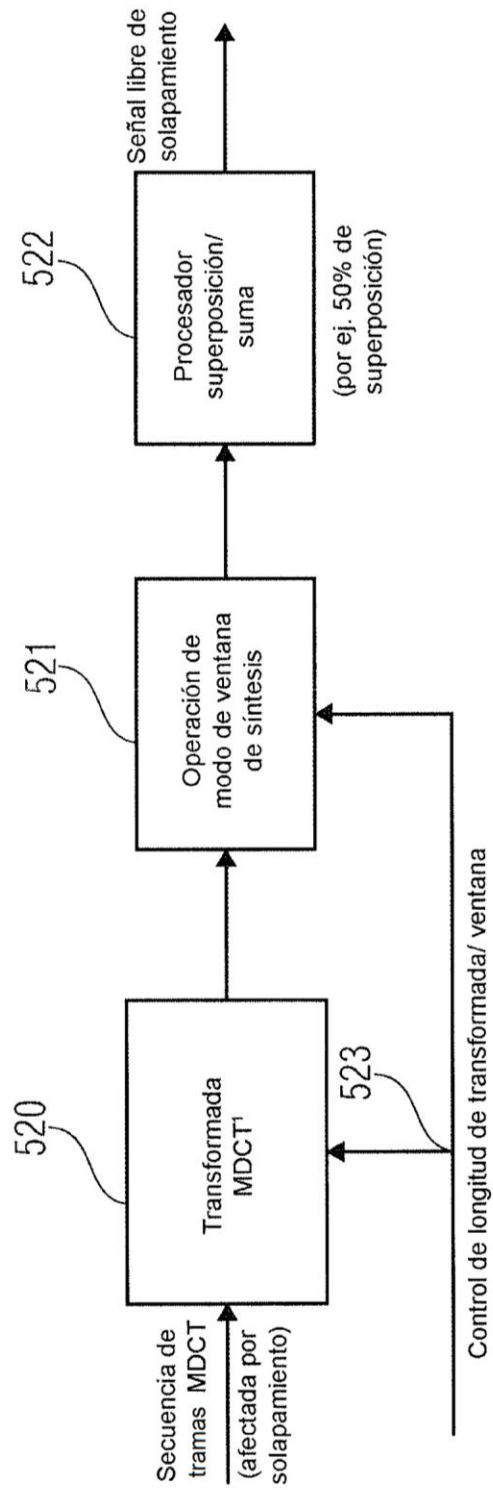


FIG 8B

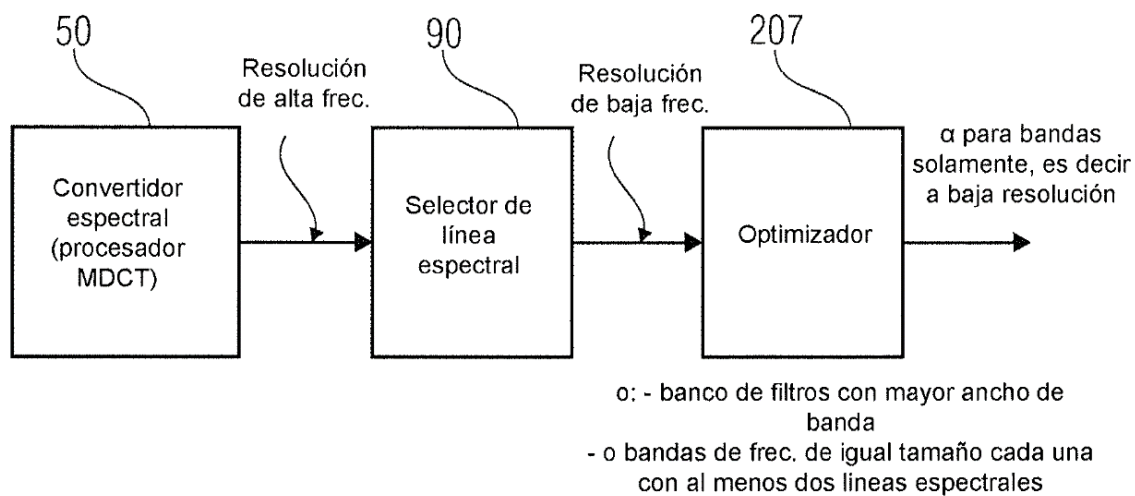


FIG 9A

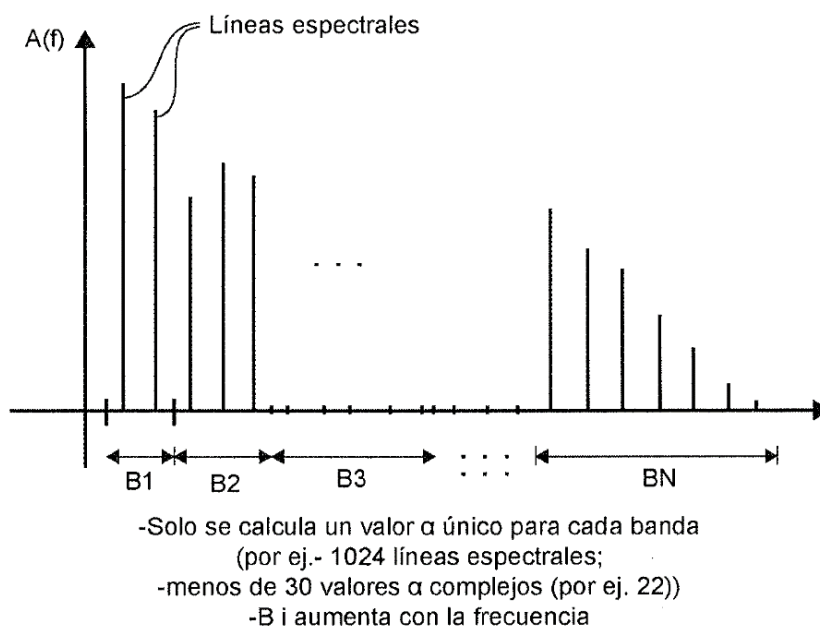
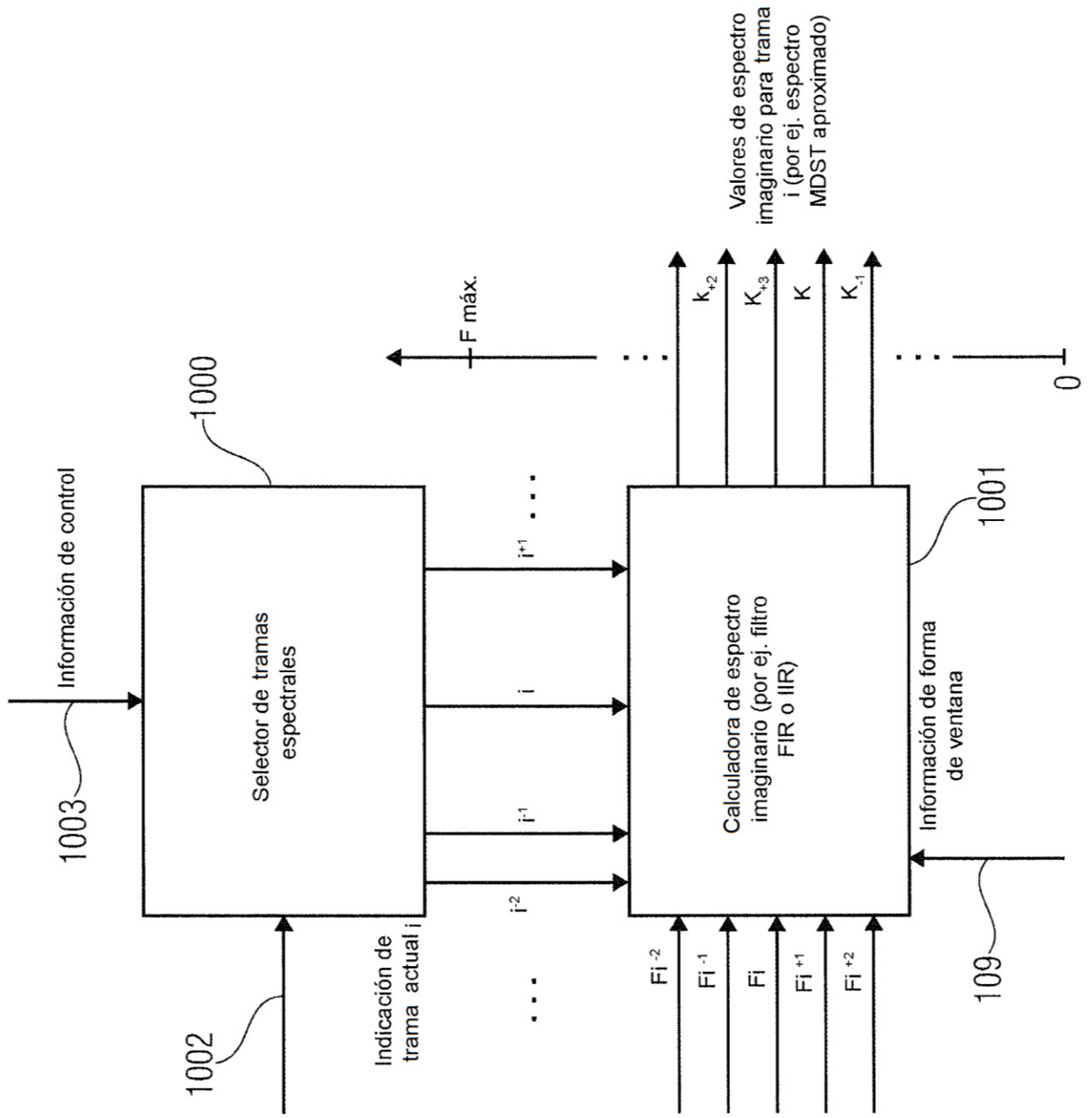


FIG 9B

FIG 10A



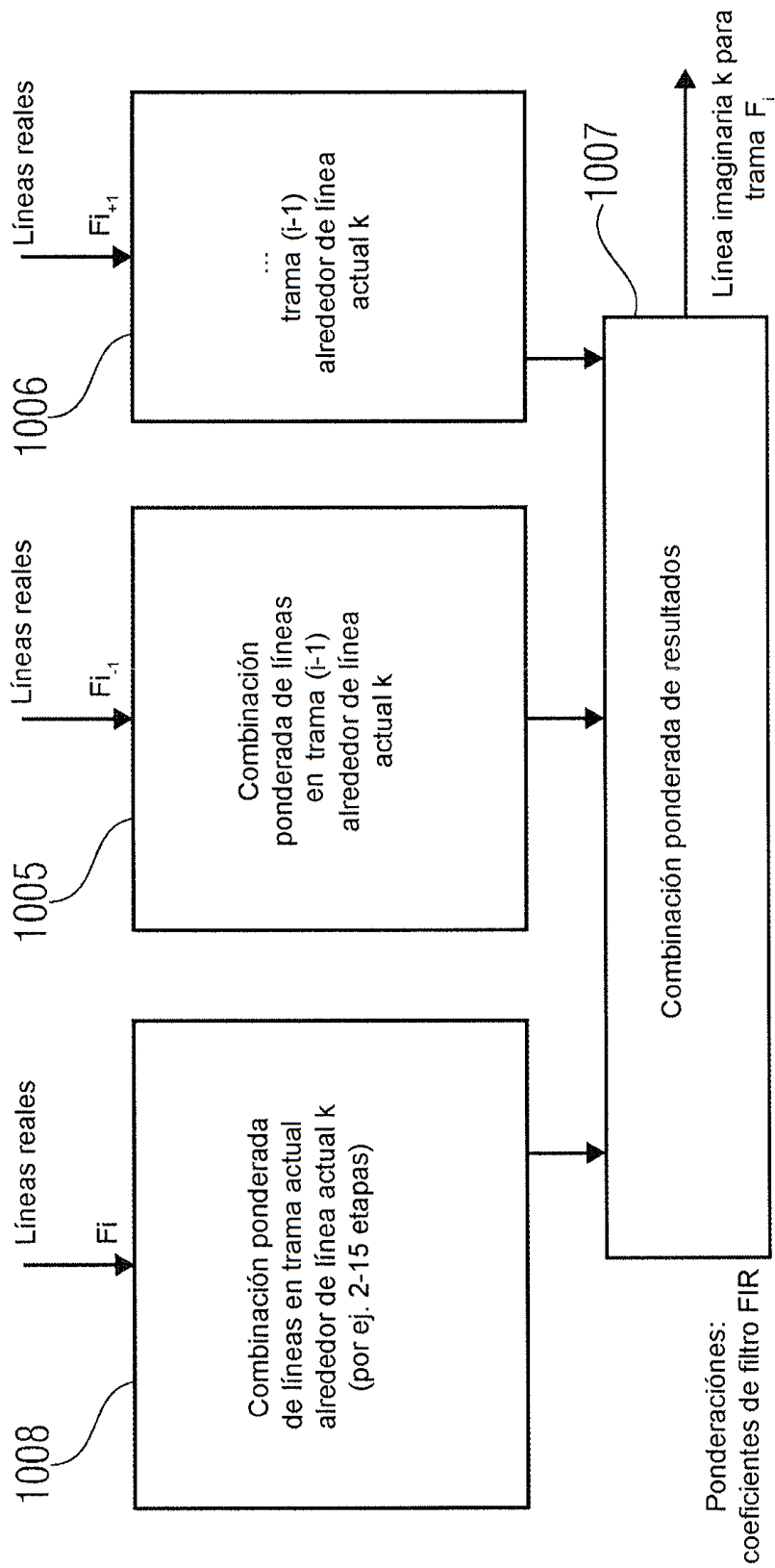


FIG 10B

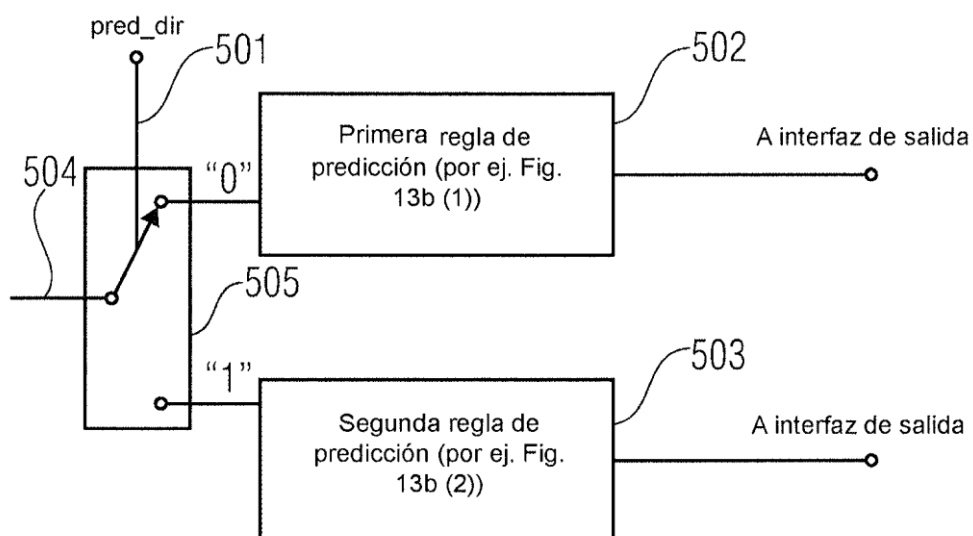


FIG 11A
(Lado del codificador)

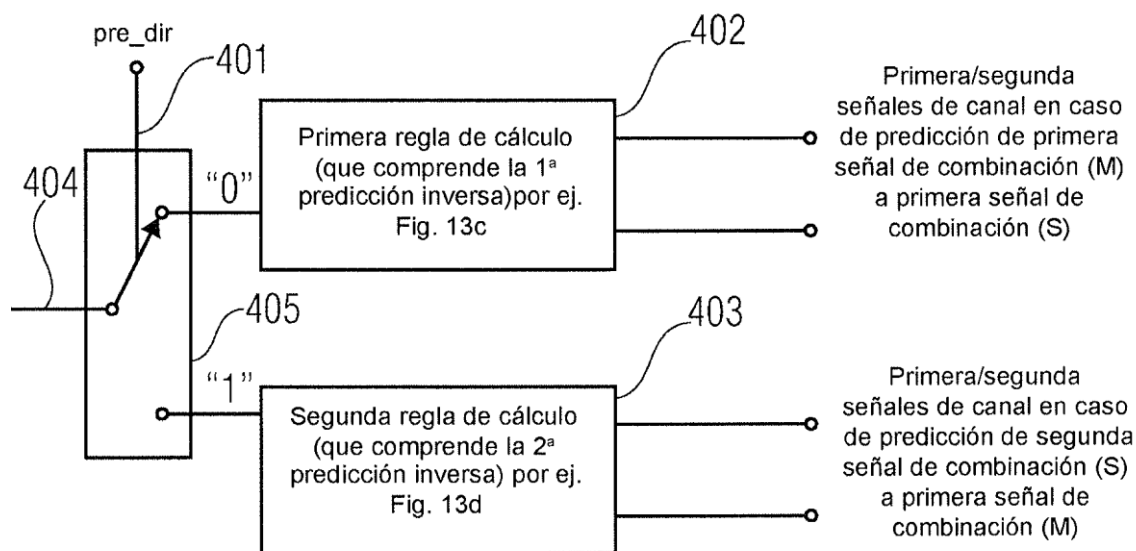


FIG 11B
(Lado del decodificador)

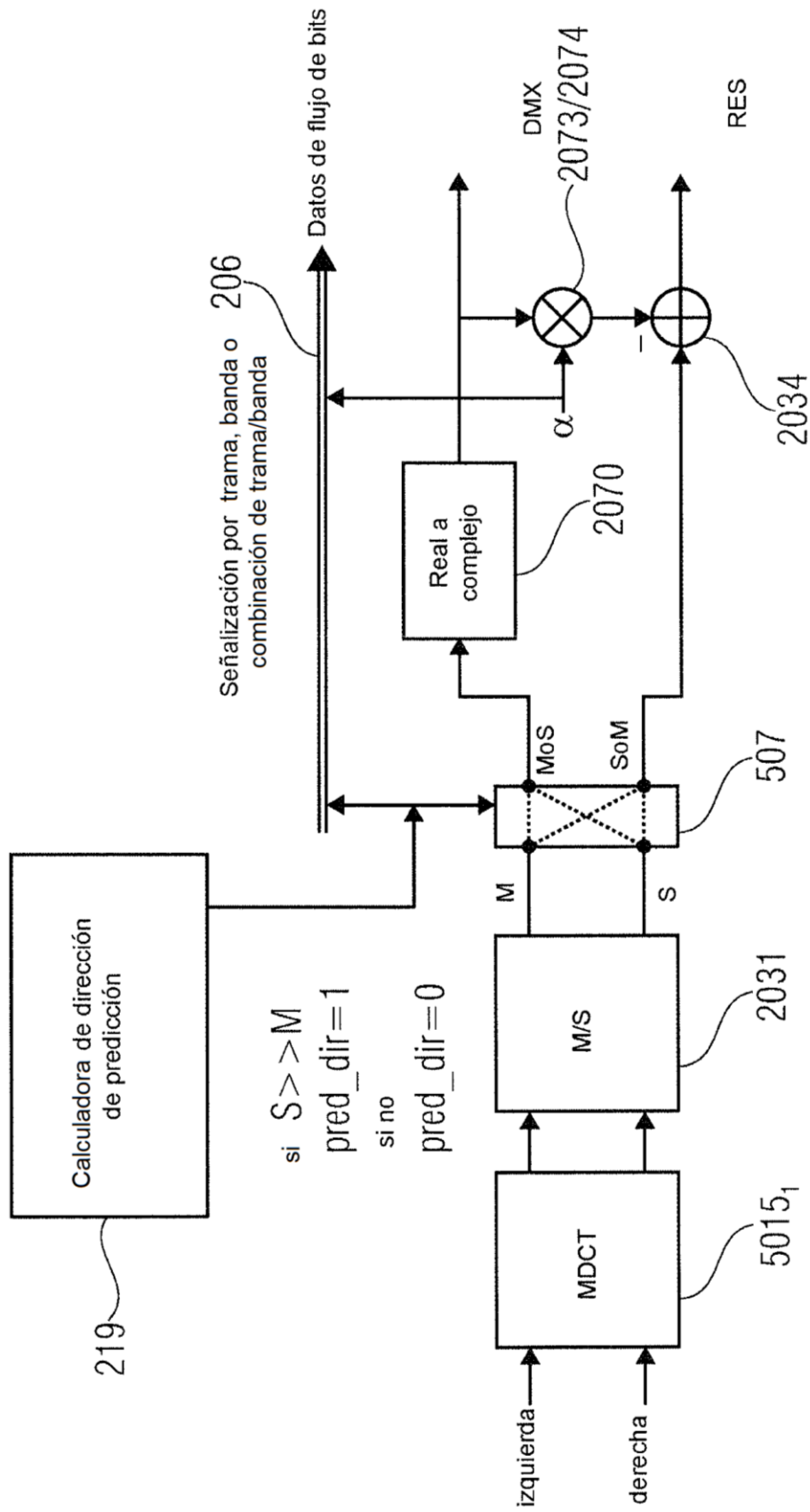
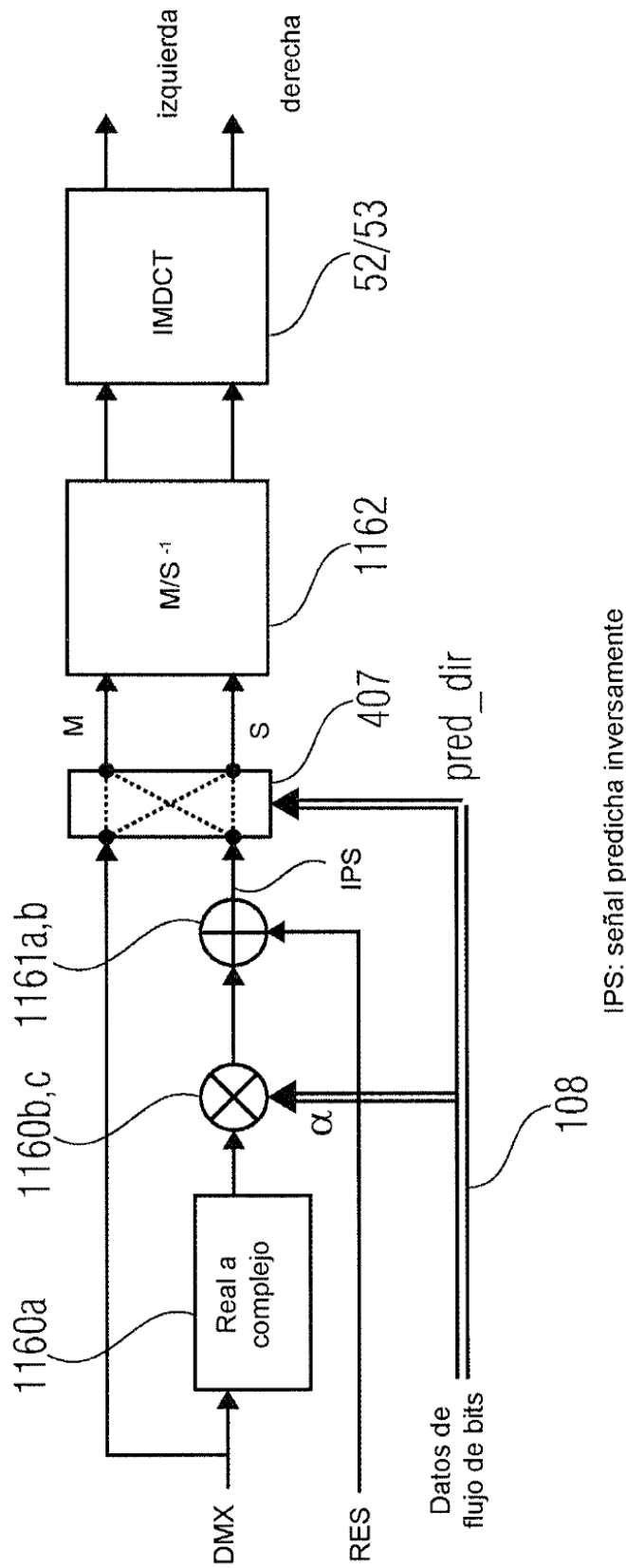


FIG 12A

(Lado del Codificador)

Estado de conmutación de predicción - Regla de cálculo
(1) DMX $\rightarrow$ M ; IPS $\rightarrow$ S
(2) DMX $\rightarrow$ S ; IPS $\rightarrow$ M



IPS: señal predicha inversamente

FIG 12B

(Lado del decodificador)

FIG 13A

pred_dir

Indica la dirección de predicción de acuerdo con:

pred_dir	Significado
0	predicción de canal medio a lateral
1	predicción de canal lateral a medio

FIG 13B

(1) pred_dir=0 : $RES = (S - \alpha \cdot M)$; transmitir : M, RES, α (2) pred_dir=1 : $RES = (M - \alpha \cdot S)$; transmitir : S, RES, α NOTA: el símbolo α se puede invertir en todas las ecuaciones.

FIG 13C

pred_dir=0:

predicción inversa explícita:

$$S = RES + \alpha \cdot M; \quad \begin{aligned} L &= \frac{1}{2}(M + S) \\ R &= \frac{1}{2}(M - S) \end{aligned}$$

predicción inversa implícita:

$$\begin{aligned} L &= \frac{1}{2}[M(1 - \alpha) + RES] \\ R &= \frac{1}{2}[M(1 + \alpha) - RES] \end{aligned}$$

FIG 13D

pred_dir=1:

predicción inversa explícita:

$$M = RES + \alpha \cdot S; \quad \begin{aligned} L &= \frac{1}{2}(M + S) \\ R &= \frac{1}{2}(M - S) \end{aligned}$$

predicción inversa implícita:

$$\begin{aligned} L &= \frac{1}{2}[S(\alpha + 1) + RES] \\ R &= \frac{1}{2}[S(\alpha - 1) + RES] \end{aligned}$$

$$\alpha \cdot M \hat{=} \alpha_{Re} \cdot M_{Re} - \alpha_{Im} \cdot M_{Im}$$

$$(1 + \alpha) \cdot M \hat{=} (1 + \alpha_{Re}) \cdot M_{Re} - \alpha_{Im} \cdot M_{Im}$$

productos
complejos,
si α es
compleja

FIG 13E

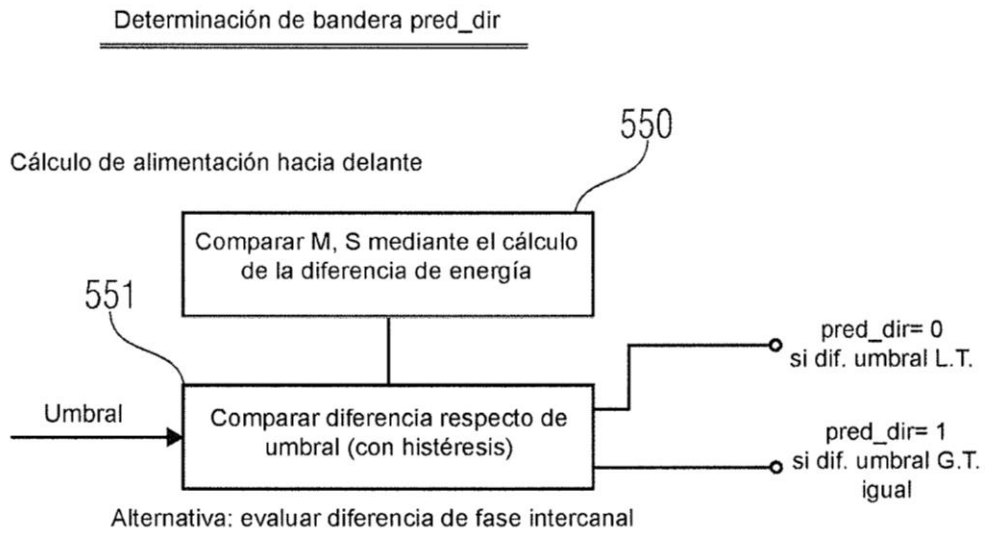


FIG 14A

Cálculo de devolución

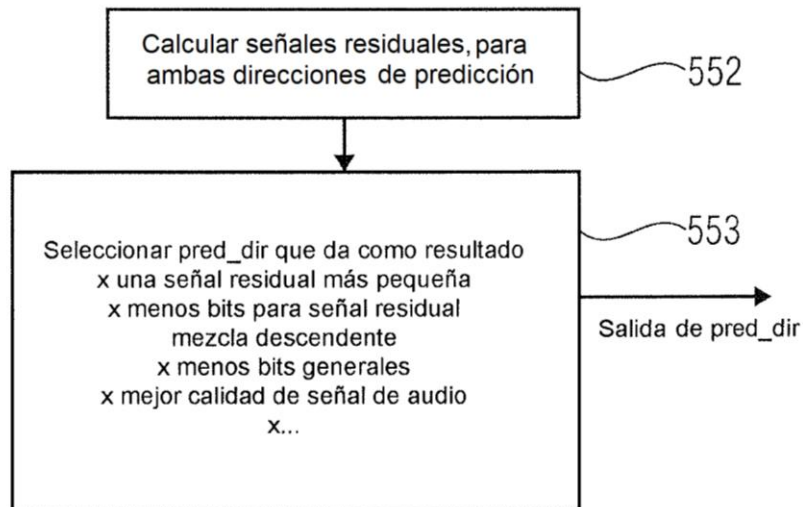


FIG 14B