

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 002 807**

51 Int. Cl.:

**G10L 19/00** (2013.01)

**H04B 1/66** (2006.01)

**G10L 19/18** (2013.01)

**G10L 19/008** (2013.01)

**G10L 25/12** (2013.01)

**G10L 19/02** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.04.2011** **E 22188294 (7)**

97 Fecha y número de publicación de la concesión europea: **18.12.2024** **EP 4120246**

54 Título: **Decodificación estéreo de predicción compleja basada en MDCT**

30 Prioridad:

**09.04.2010 US 32245810 P**

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

**07.03.2025**

73 Titular/es:

**DOLBY INTERNATIONAL AB (100.00%)**  
**77 Sir John Rogerson's Quay, Block C, Grand**  
**Canal Docklands**  
**Dublin, D02 VK60, IE**

72 Inventor/es:

**PURNHAGEN, HEIKO;**  
**CARLSSON, PONTUS y**  
**VILLEMOES, LARS**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 3 002 807 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

Decodificación estéreo de predicción compleja basada en MDCT

### Referencia cruzada a solicitud relacionada

Esta solicitud es una solicitud divisional europea de la solicitud de patente europea EP19184726.8 (referencia D10010AEP02) cuyo formulario de la EPO 1001 se presentó el 5 de julio de 2019.

### Campo técnico

La invención descrita en la presente memoria se relaciona de manera general con la codificación de audio estéreo.

### Antecedentes de la invención

La codificación conjunta de los canales izquierdo (L) y derecho (R) de una señal estéreo permite una codificación más eficiente comparada con la codificación independiente de L y R. Un enfoque común para la codificación estéreo conjunta es la codificación central/lateral (M/S). Aquí, una señal central (M) es formada mediante la suma de las señales L y R, por ejemplo, la señal M puede tener la forma

$$M = (L + R)/2$$

También, una señal lateral (S) está formada mediante la resta de los dos canales L y R, por ejemplo, la señal S puede tener la forma

$$S = (L - R)/2$$

En el caso de una codificación M/S, se codifican las señales M y S en lugar de las señales L y R.

En el estándar (véase el documento estándar ISO/IEC 13818-7) AAC (Codificación de Audio Avanzado) de MPEG (Grupo de Expertos de Imágenes en Movimiento), la codificación estéreo L/R y la codificación estéreo M/S se pueden elegir de una manera variante en el tiempo y variante en la frecuencia. Por tanto, el codificador estéreo puede aplicar la codificación L/R para algunas bandas de frecuencias de la señal estéreo, mientras que la codificación M/S se usa para codificar otras bandas de frecuencias de la señal estéreo (variante en frecuencia). Además, el codificador puede conmutar en el tiempo entre la codificación (variante en el tiempo) L/R y M/S. En la AAC de MPEG, la codificación estéreo se lleva a cabo en el dominio de la frecuencia, más concretamente en el dominio MDCT (transformada de coseno discreta modificada). Esto permite elegir de manera adaptativa ya sea la codificación L/R o la M/S de una manera variable en la frecuencia y también en el tiempo.

La codificación estéreo paramétrica es una técnica para codificar de manera eficiente una señal de audio estéreo como una señal monoaural más una pequeña cantidad de información lateral para los parámetros estéreo. Es parte del estándar de Audio MPEG-4 (véase el documento estándar ISO/IEC 14496-3). La señal monoaural se puede codificar usando cualquier codificador de audio. Los parámetros estéreo se pueden incrustar en la parte auxiliar del flujo de bits mono, alcanzando así una compatibilidad hacia delante y hacia atrás total. En el decodificador, es la señal monoaural la que se decodifica primero, después de lo cual se reconstruye la señal estéreo con la ayuda de los parámetros estéreo. Una versión decorrelacionada de la señal monodecodificada, que tiene una correlación de cruces por cero con la señal mono, es generada por medio de un decorrelacionador, por ejemplo, un filtro paso todo apropiado que puede incluir una o más líneas de retardo. Esencialmente la señal decorrelacionada tiene la misma distribución de energía espectral y temporal que la señal mono. La señal monoaural junto con la señal decorrelacionada se introducen al proceso de mezcla ascendente que está controlado por los parámetros estéreo y que reconstruye la señal estéreo. Para más información, vea el documento "Codificación Estéreo Paramétrica de Baja Complejidad en MPEG-4", H. Purnhagen, Proc. de la 7ª Conferencia Int. sobre Efectos de Audio Digital (DAFx'04), Naples, Italia, 5-8 de Octubre, 2004, páginas 163-168.

Envolvente MPEG (MPS; véase la ISO/IEC 23003-1 y el documento "Envolvente MPEG – el Estándar ISO/MPEG para la Codificación de Audio de Multi Canal Eficiente y Compatible", J. Herre et al., Documento 7084 de la Convención de Ingeniería de Audio, 122ª Convención, 5-8 Mayo, 2007) permite combinar los principios de la codificación estéreo paramétrica con la codificación residual, sustituyendo la señal decorrelacionada con una residual transmitida y mejorando por lo tanto la calidad perceptual. La codificación residual se puede conseguir mediante una mezcla descendente de una señal multi canal y, de manera opcional, extrayendo las señales espaciales. Durante el proceso de mezcla descendente, se calculan las señales residuales que representan la señal de error y después de codifican y transmiten. Estas toman el lugar de las señales decorrelacionadas en el decodificador. En un enfoque híbrido, pueden reemplazar las señales decorrelacionadas en ciertas bandas de frecuencias, preferiblemente en las bandas relativamente bajas.

Según el sistema actual de Codificación de Audio y Voz Unificado MPEG (USAC), del cual se muestran dos ejemplos en la figura 1, el decodificador comprende un banco de filtros espejo en cuadratura de valor complejo (QMF) ubicado hacia abajo del decodificador de núcleo. La representación QMF obtenida como la salida del banco

de filtros es compleja – por tanto sobre muestreada por un factor de 2 – y se puede disponer como una señal  $M$  de mezcla descendente (o, de manera equivalente, una señal central) y una señal  $D$  residual, a la cual se aplica una matriz de mezcla ascendente con entradas complejas. Las señales  $L$  y  $R$  (en el dominio QMF) se obtienen como:

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

donde  $g$  es un factor de ganancia de valor real y  $\alpha$  es un coeficiente de predicción de valor complejo. Preferiblemente,  $\alpha$  se escoge de manera tal que se minimice la energía de la señal  $D$  residual. El factor de ganancia puede ser determinado mediante normalización, esto es, para asegurar que la energía de la señal de suma es igual que la suma de las energías de las señales izquierda y derecha. Las partes real e imaginaria de cada una de las señales izquierda y derecha son mutuamente redundantes – en principio, cada una de ellas se puede calcular en base a la otra – pero son beneficiosas para permitir la aplicación posterior de un decodificador de replicación de la banda espectral (SBR) sin que se produzcan artefactos de solapamiento audibles. El uso de una representación de señal sobre muestreada puede también, por razones similares, ser escogido con el objetivo de evitar artefactos conectados con otro procesamiento de señal adaptativo en el tiempo o la frecuencia (no mostrado), tal como la mezcla ascendente de mono a estéreo. El filtrado QMF inverso es el último paso de procesamiento en el decodificador. Se observa que la representación QMF limitada en banda de la señal permite técnicas residuales limitadas en banda y técnicas de “relleno residual”, que se pueden integrar dentro de los decodificadores de este tipo.

La estructura de codificación anterior es bien adecuada para tasas de bits bajas, normalmente por debajo de 80 kb/s, pero no es óptima con respecto a la complejidad computacional para tasas de bits más altas. Más precisamente, a tasas de bits mayores, la herramienta SBR normalmente no se utiliza (ya que no mejoraría la eficiencia de codificación). Entonces, en un decodificador sin una etapa SBR, sólo la matriz de mezcla ascendente de valor complejo motiva la presencia del banco de filtros QMF, que es computacionalmente demandante e introduce un retardo (con una longitud de trama de 1024 muestras, el banco de filtros de análisis/síntesis QMF introduce un retardo de 961 muestras). Esto indica claramente una necesidad de una estructura de codificación más eficiente. Una descripción técnica de una propuesta para la codificación estéreo unificada en USAC se proporciona por H. Purnhagen et al: "Technical description of proposed Unified Stereo Coding in USAC", 90. Reunión MPEG; 26.10.2009 - 30.10.2009; Xian, China; Grupo de expertos en imágenes en movimiento ISO/IEC JTC1/SC29/WG11, MPEG2009/M16921.

### Sumario de la invención

Es un objeto de la presente invención proporcionar métodos y aparatos para codificación estéreo que son computacionalmente eficientes también en el intervalo de alta tasa de bits.

La invención cumple este objeto proporcionando un sistema decodificador según la reivindicación 1 y un método decodificador según la reivindicación 7. Se definen aspectos preferibles adicionales por las reivindicaciones dependientes.

En un primer aspecto, la invención proporciona un sistema decodificador para proporcionar una señal estéreo mediante codificación estéreo de predicción compleja, comprendiendo el sistema decodificador:

una mezcla ascendente adaptada para generar la señal estéreo basándose en primeras representaciones en el dominio de la frecuencia de una señal de mezcla descendente ( $M$ ) y una señal residual ( $D$ ), comprendiendo cada una de las primeras representaciones en el dominio de la frecuencia primeras componentes espectrales que representan contenido espectral de la señal correspondiente expresada en un primer subespacio de un espacio multidimensional, comprendiendo la etapa de mezcla ascendente:

un módulo para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente basándose en la primera representación en el dominio de la frecuencia de la misma, comprendiendo la segunda representación en el dominio de la frecuencia segundos componentes espectrales que representan contenido espectral de la señal expresada en un segundo subespacio del espacio multidimensional que incluye una parte del espacio multidimensional no incluida en el primer subespacio;

un sumador ponderado para calcular una señal lateral combinando la primera representación en el dominio de la frecuencia de la señal residual, la primera representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor real de un coeficiente de predicción complejo codificado en una señal de flujo de bits, y la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor imaginario del coeficiente de predicción complejo; y

una etapa de suma y diferencia para calcular la señal estéreo basándose en la primera representación en el dominio de la frecuencia de la señal de mezcla descendente y la señal lateral,

en la que la etapa de mezcla ascendente es operable además en un modo de paso, en el que dichas señales de mezcla descendente y residual se suministran directamente a la suma y diferencia.

En un segundo aspecto relacionado (pero no abarcado por las reivindicaciones), se proporciona un sistema codificador para codificar una señal estéreo mediante una señal de flujo de bits mediante codificación estéreo de predicción compleja, que incluye:

un estimador para estimar un coeficiente de predicción complejo;

5 una etapa de codificación que funciona para:

(a) transformar la señal estéreo en una representación en el dominio de la frecuencia de una mezcla descendente y una señal residual, en una relación determinada por el valor del coeficiente de predicción complejo;

y

10 un multiplexor para recibir la salida de la etapa de codificación y del estimador y codificar esta por dicha señal de flujo bits.

En un tercer aspecto relacionado (pero no abarcado por las reivindicaciones), se proporciona un método para codificar una señal estéreo en un flujo de bits y en un cuarto aspecto, se proporciona un método para decodificar un flujo de bits en al menos una señal estéreo. Las características técnicas de cada método son análogas a las del sistema codificador y el sistema decodificador, respectivamente. En un quinto y sexto aspecto (no abarcados en las reivindicaciones), se proporciona un producto de programa informático que contiene instrucciones para ejecutar cada uno de los métodos en un ordenador.

La invención se beneficia de las ventajas de la codificación estéreo unificada en el sistema USAC MPEG. Estas ventajas se mantienen también con tasas de bits mayores, en las que normalmente no se utiliza la SBR, sin el aumento significativo en la complejidad computacional que acompañaría a un enfoque basado en QMF. Esto es posible ya que la transformada MDCT muestreada de manera crítica, que es la base del sistema de codificación de transformada USAC MPEG, se puede usar para la codificación estéreo por predicción compleja que se proporciona por la invención, al menos en los casos en los que los anchos de banda de audio codificado de los canales de mezcla descendente son los mismos y el proceso de mezcla ascendente no incluye la decorrelación. Esto significa que no se requiere más una transformada QMF adicional. Una implementación representativa de la codificación estéreo por predicción compleja en el dominio QMF aumentaría realmente el número de operaciones por unidad de tiempo de manera significativa comparado con la tradicional estéreo L/R o M/S. Por tanto, el aparato de codificación según la invención parece ser competitivo a dichas tasas de bits, proporcionando una alta calidad de audio con un moderado gasto computacional.

Como se darán cuenta las personas expertas en la materia, el hecho de que la etapa de mezcla ascendente sea más operable en un modo de paso permite al decodificador decodificar de manera adaptativa según la codificación directa o conjunta convencional, y la codificación de predicción compleja, como se determina en el lado del decodificador. Por tanto, en estos casos en los que el decodificador no puede aumentar de manera positiva el nivel de calidad más allá del convencional de la codificación estéreo L/R directa o la codificación M/S conjunta, puede garantizar al menos que se mantiene el mismo nivel. Por tanto, un decodificador según este aspecto de la invención puede, a partir de un punto de vista funcional, considerarse como un súper conjunto en relación con los antecedentes de la técnica.

Como una ventaja sobre la señal estéreo codificada por predicción basada en QMF, es posible una perfecta reconstrucción de la señal (aparte de los errores de cuantificación, que se pueden hacer arbitrariamente pequeños).

Por tanto, la invención proporciona un aparato de codificación para una codificación estéreo basada en transformada mediante predicción compleja. Preferiblemente, el aparato según la invención no se limita a la codificación estéreo de predicción compleja, sino que es operable también en un régimen de codificación estéreo L/R directo o de codificación estéreo M/S conjunta según los antecedentes de la técnica, de modo que es posible seleccionar el método de codificación más adecuado para una aplicación concreta o durante un intervalo de tiempo concreto.

Una representación sobre muestreada (por ejemplo, compleja) de la señal, que incluye tanto dicho primer como dicho segundo componentes espectrales, se usa como base para la predicción compleja según la invención, y por lo tanto, los módulos para calcular dicha representación sobre muestreada se disponen en el sistema codificador relacionado (no abarcado por las reivindicaciones) y en el sistema decodificador reivindicado. Los componentes espectrales se refieren al primer y segundo sub espacios de un espacio multidimensional, que puede ser el conjunto de funciones dependientes del tiempo en un intervalo de una longitud dada (por ejemplo, una longitud de trama de tiempo predefinida) muestreado a una frecuencia de muestreo finita. Es bien conocido que las funciones en este espacio multi dimensional concreto se pueden aproximar mediante una suma ponderada finita de funciones base.

Como apreciará una persona experta en la técnica, se equipa un codificador (no abarcado por las reivindicaciones) adaptado para cooperar con un decodificador con módulos equivalentes para proporcionar la representación sobre muestreada sobre la que se basa la codificación de predicción, para de este modo permitir una fiel reproducción de la señal codificada. Dichos módulos equivalentes pueden ser módulos idénticos o similares o módulos que tengan

características de transferencia idénticas o similares. En concreto, los módulos en el codificador y en el decodificador, respectivamente, pueden ser unidades de procesamiento similares o distintas que ejecutan respectivos programas informáticos que realizan conjuntos equivalentes de operaciones matemáticas.

En algunas realizaciones del sistema decodificador reivindicado o del sistema codificador relacionado (no abarcado por las reivindicaciones), los primeros componentes espectrales tienen valores reales expresados en el primer sub espacio, y los segundos componentes espectrales tienen valores imaginarios expresados en el segundo sub espacio. Los primeros y los segundos componentes forman juntos una representación espectral compleja de la señal. El primer sub espacio puede ser el espacio lineal de un primer conjunto de funciones base, mientras que el segundo sub espacio puede ser el espacio lineal de un conjunto de segundas funciones base, algunas de las cuales son linealmente independientes del primer conjunto de funciones base.

En una realización, el módulo para calcular la representación compleja es una transformada de real a imaginario, esto es, un módulo para calcular las partes imaginarias del espectro de una señal de tiempo discreto en base a una representación espectral real de la señal. La transformada puede estar basada en relaciones matemáticas exactas o aproximadas, tales como fórmulas de análisis armónico o relaciones heurísticas.

En algunas realizaciones del sistema decodificador reivindicado o del sistema codificador relacionado (no abarcado por las reivindicaciones), las primeras componentes espectrales se obtienen mediante una transformada desde el dominio del tiempo al de la frecuencia, preferiblemente una transformada de Fourier, de una señal discreta en el dominio del tiempo, tal como mediante una transformada de coseno discreta (DCT), una transformada de coseno discreta modificada (MDCT), una transformada de seno discreta (DST), una transformada de seno discreta modificada (MDST), una transformada rápida de Fourier (FFT), un algoritmo de Fourier basado en un factor primo o similar. En los primeros cuatro casos, los segundos componentes espectrales se pueden obtener después mediante la DST, la MDST, la DCT y la MDCT, respectivamente. Como es bien sabido, el espacio lineal de cosenos que son periódicos en el intervalo de la unidad forma un sub espacio que no está enteramente contenido en el espacio lineal de senos periódicos en el mismo intervalo. Preferiblemente, los primeros componentes espectrales son obtenibles mediante la MDCT y los segundos componentes espectrales son obtenibles mediante la MDCT.

En una realización, el sistema decodificador incluye al menos un módulo de conformado de ruido temporal (módulo TNS, o filtro TNS), que se adapta por encima de la etapa de mezcla ascendente. Hablando de manera general, el uso del TNS aumenta la calidad de audio percibida para las señales con componentes transitorios, y esto también aplica a las realizaciones del sistema decodificador inventivo que presenta el TNS. En la codificación estéreo L/R y M/S convencional, el filtro TNS se puede aplicar como un último paso de procesamiento en el dominio de la frecuencia, directamente antes de la transformada inversa. En el caso de la codificación estéreo de predicción compleja, sin embargo, es ventajoso a menudo aplicar el filtro TNS en la mezcla descendente y las señales residuales, esto es, antes de la matriz de mezcla ascendente. Dicho de otra manera, se aplica el TNS a las combinaciones lineales de los canales izquierdo y derecho, lo que tiene varias ventajas. En primer lugar, puede resultar en una situación dada de que el TNS es beneficioso sólo para la señal de la mezcla descendente. Entonces, el filtrado TNS se puede suprimir u omitir para la señal residual y, lo que puede implicar un uso más económico del ancho de banda disponible, los coeficientes del filtro TNS sólo necesitan ser transmitidos para la señal de la mezcla descendente. En segundo lugar, el cálculo de la representación sobre muestreada de la señal del enlace descendente (por ejemplo, los datos de la MDST que se derivan de los datos de la MDCT para formar una representación en el dominio de la frecuencia compleja), que es necesaria en la codificación de predicción compleja, puede requerir que la representación en el dominio del tiempo de la señal de mezcla descendente sea computable. Esto implica a su vez que la señal de mezcla descendente esté disponible preferiblemente como una secuencia de tiempo del espectro MDCT obtenida de una manera uniforme. Si el filtro TNS fuera aplicado en el decodificador después de la matriz de mezcla ascendente, que convierte una representación de la mezcla descendente/residual en una representación izquierda/derecha, sólo una secuencia del espectro MDCT residual de TNS de la señal de la mezcla descendente estaría disponible. Esto haría el cálculo eficiente del espectro MDST correspondiente muy desafiante, especialmente si los canales izquierdo y derecho están usando filtros TNS con diferentes características.

Se enfatiza que la disponibilidad de una secuencia de tiempo del espectro MDCT no es un criterio absoluto para obtener una representación MDST ajustada para servir como base para la codificación de predicción compleja. Además de la evidencia experimental, este hecho se puede explicar mediante el TNS que se aplica de manera general sólo a las frecuencias mayores, tales como por encima de unos pocos kilohercios, por lo que la señal residual filtrada por el TNS corresponde aproximadamente a la señal no filtrada residual para las frecuencias menores. Por tanto, la invención se puede realizar como un decodificador para la codificación estéreo de predicción compleja, en la que los filtros TNS tienen una colocación diferente que por encima de la etapa de mezcla ascendente, como se indica más adelante.

En una realización, el sistema decodificador incluye al menos un módulo TNS adicional ubicado por debajo de la etapa de mezcla ascendente. Por medio de una disposición selectora, ya sea el módulo o módulos TNS por encima de la etapa de mezcla ascendente o el módulo o módulos TNS por debajo de la etapa de mezcla ascendente. Bajo ciertas circunstancias, el cálculo de la representación compleja en el dominio de la frecuencia no requiere que sea computable una representación en el dominio del tiempo de la señal de mezcla descendente. Además, como se expuso anteriormente, el decodificador puede ser operable de manera selectiva en un modo de codificación directo o

conjunto, no aplicando codificación de predicción compleja, y entonces puede ser más adecuado aplicar la ubicación convencional de los módulos TNS, esto es, como uno de los últimos pasos de procesamiento en el dominio de la frecuencia.

En una realización, el sistema decodificador se adapta para economizar los recursos de procesamiento, y posiblemente la energía, desactivando el módulo para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente cuando la última no es necesaria. Se supone que la señal de mezcla descendente se parte en bloques de tiempo sucesivos, cada uno de los cuales se asocia con un valor de coeficiente de predicción complejo. Este valor puede ser determinado mediante una decisión tomada para cada bloque de tiempo mediante un codificador que coopera con el decodificador reivindicado. Además, en esta realización, el módulo para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente se adapta para desactivarse a sí mismo si, durante un bloque de tiempo dado, el valor absoluto de la parte imaginaria del coeficiente de predicción complejo es cero o es menor que una tolerancia predeterminada. La desactivación del módulo puede implicar que no se calcule una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente para este bloque de tiempo. Si la desactivación no tiene lugar, la segunda representación en el dominio de la frecuencia (por ejemplo, un conjunto de coeficientes MDST) sería multiplicada por cero o por un número sustancialmente del mismo orden de magnitud que la  $\epsilon$  máquina (unidad de redondeo) del decodificador o algún otro valor de umbral adecuado.

En un desarrollo adicional de la realización anterior, se consigue la economización de recursos de procesamiento en un sub nivel del bloque de tiempo dentro del cual se parte la señal de mezcla descendente. Por ejemplo, dicho sub nivel dentro de un bloque de tiempo puede ser una banda de frecuencias, en donde el codificador (no abarcado por las reivindicaciones) determina un valor del coeficiente de predicción complejo para cada banda de frecuencias dentro de un bloque de tiempo. De manera similar, el módulo para producir una segunda representación en el dominio de la frecuencia se adapta para eliminar su funcionamiento para una banda de frecuencias en un bloque de tiempo para el que el coeficiente de predicción complejo es cero o tiene una magnitud menor que una tolerancia.

En una realización, los primeros componentes espectrales son coeficientes de transformadas dispuestos en uno o más bloques de tiempo de los coeficientes de transformada, cada bloque generado mediante la aplicación de una transformada a un segmento de tiempo de una señal en el dominio del tiempo. Además, el módulo para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente se adapta para:

- derivar uno o más primeros componentes intermedios a partir de al menos alguno de los componentes espectrales;
- formar una combinación de dichos uno o más primeros componentes espectrales según al menos una parte de una o más respuestas al impulso para obtener uno o más segundos componentes intermedios; y
- derivar dichos uno o más segundos componentes espectrales a partir de dichos uno o más segundos componentes intermedios.

Este procedimiento logra el cálculo de la segunda representación en el dominio de la frecuencia de manera directa a partir de la primera representación en el dominio de la frecuencia, como se describe con mayor detalle en el documento de los EE.UU. 6.980.933 B2, de manera notable en las columnas 8-28 y en concreto en la ecuación 41 del mismo. Como se darán cuenta las personas expertas, no se realiza el cálculo a través del dominio del tiempo, al contrario de, por ejemplo, la transformada inversa seguida por una transformada diferente.

Para una transformación ejemplar de la codificación estéreo de predicción compleja según la invención, se ha estimado que la complejidad computacional aumenta sólo ligeramente (de manera significativamente menor que el aumento provocado por la codificación estéreo de predicción compleja en el dominio QMF) comparada con el estéreo L/R o M/S tradicional. Una realización de este tipo que incluye el cálculo exacto de los segundos componentes espectrales introduce un retardo que es normalmente unos pocos por cientos más largo que el introducido por una implementación basada en QMF (suponiendo que la longitud del bloque de tiempo es de 1024 muestras y comparando con el retardo del banco de filtros de análisis/síntesis QMF híbrido, que es de 961 muestras).

De manera adecuada, en al menos alguna de las realizaciones anteriores, las respuestas al impulso se adaptan a la transformada mediante la cual es obtenible la primera representación en el dominio de la frecuencia, y más concretamente de acuerdo con la respuesta en frecuencia característica de la misma.

En algunas realizaciones, la primera representación en el dominio de la frecuencia de la señal de mezcla descendente es obtenida mediante una transformada que se aplica en conexión con una o más funciones de ventana de análisis (o funciones de corte, por ejemplo, la ventana rectangular, la ventana sinusoidal, la ventana Kaiser-Bessel-derivada, etc), un objetivo de la cual es lograr una segmentación temporal sin introducir una cantidad perjudicial de ruido o cambiar el espectro de una manera indeseable. Posiblemente, dichas funciones de ventana se solapan de manera parcial. Entonces, preferiblemente, las características de respuesta en frecuencia de la transformada son dependientes de las características de dichas una o más funciones de ventana de análisis.

Aún referente a las realizaciones que realizan el cálculo de la segunda representación en el dominio de la frecuencia

dentro del dominio de la frecuencia, es posible disminuir la carga computacional involucrada mediante el uso de una segunda representación aproximada en el dominio de la frecuencia. Dicha aproximación se puede lograr no requiriendo la información completa sobre la cual basar el cálculo. Mediante las enseñanzas del documento de los EE.UU. 6.980.933 B2, por ejemplo, se requieren los primeros datos en el dominio de la frecuencia de tres bloques de tiempo para el cálculo exacto de la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente en un bloque, es decir un bloque contemporáneo con el bloque de salida, un bloque anterior y un bloque posterior. Con el propósito de la codificación de predicción compleja según la presente invención, se pueden obtener aproximaciones adecuadas omitiendo – o reemplazando por cero – los datos que se obtienen a partir del bloque posterior (por lo cual la operación de módulo puede resultar causal, esto es, no contribuye al retardo) y/o a partir del bloque anterior, para el cálculo de la segunda representación en el dominio de la frecuencia se basa en los datos de uno o más bloques de tiempo sólo. Se observa que incluso aunque la omisión de los datos de entrada implique un reescalado de la segunda representación en el dominio de la frecuencia – en el sentido de que, por ejemplo, no representa más la misma energía – se puede usar ya como una base para la codificación de predicción compleja siempre que se calcule de una manera equivalente en tanto los extremos del codificador (no abarcado por las reivindicaciones) como del decodificador reivindicado, como se observó anteriormente. En efecto, un posible reescalado de este tipo será compensado mediante el cambio correspondiente del valor del coeficiente de predicción.

Todavía otro método aproximado para calcular el componente espectral que forma parte de la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente puede incluir la combinación de al menos dos componentes de la primera representación en el dominio de la frecuencia. Los últimos componentes pueden ser adyacentes con respecto al tiempo y/o la frecuencia. Como alternativa, pueden ser combinados mediante un filtrado de respuesta al impulso finita (FIR), con relativamente pocas pulsaciones. Por ejemplo, en un sistema que aplica un tamaño de bloque de tiempo de 1024, dichos filtros FIR pueden incluir 2, 3, 4 etc pulsaciones. Las descripciones de los métodos de cálculo aproximados de esta naturaleza se pueden encontrar, por ejemplo, en el documento de los EE.UU. 2005/0197831 A1. Si se usa una función de ventana que da relativamente menores pesos a los alrededores de cada límite de bloque de tiempo, por ejemplo, una función no rectangular, puede ser conveniente basar los segundos componentes espectrales en un bloque de tiempo sólo en las combinaciones de los primeros componentes espectrales en el mismo bloque de tiempo, lo que implica que no está disponible la misma cantidad de información para los componentes más exteriores. El error de aproximación posiblemente introducido por dicha práctica es hasta cierto punto suprimido o cancelado por la forma de la función ventana.

En una realización de un decodificador, que se diseña para emitir una señal estéreo en el dominio del tiempo, se incluye una posibilidad de conmutación entre la codificación directa o conjunta y la codificación de predicción compleja. Esto se logra mediante la provisión de:

- un conmutador que se puede operar de manera selectiva ya sea como una etapa de paso (que no modifica las señales) o como una transformada de suma y diferencia;
- una etapa de transformada inversa para realizar una transformada de frecuencia a tiempo; y
- una disposición selectora para alimentar la etapa de transformada inversa con ya sea una señal codificada de manera directa (o conjunta) o con una señal codificada mediante predicción compleja.

Como se darán cuenta las personas expertas, dicha flexibilidad en la parte del decodificador da al codificador libertad para elegir entre la codificación convencional directa o conjunta y la codificación por predicción compleja. Por lo tanto, en los casos en los que el nivel de calidad de la codificación estéreo L/R directa o la codificación estéreo M/S conjunta convencionales no se pueda superar, esta realización puede garantizar al menos que se mantiene el mismo nivel. Por tanto, el decodificador según esta realización se puede considerar como un súper conjunto con respecto a la técnica relacionada.

Otro grupo de realizaciones del sistema decodificador realizan el cálculo de los segundos componentes espectrales en la segunda representación en el dominio de la frecuencia a través del dominio del tiempo. Más precisamente, se aplica la inversa de la transformada mediante la cual se obtuvieron (o son obtenibles) los primeros componentes espectrales y es seguida por una transformada diferente que tiene como salida los segundos componentes espectrales. En concreto, la MDCT inversa puede ser seguida por una MDST. Para reducir el número de transformadas y de transformadas inversas, la salida de la MDCT inversa puede, en dicha realización, ser proporcionada a tanto la MDST como los terminales de salida (posiblemente precedidos por pasos de procesamiento adicionales) del sistema de decodificación.

Para una implementación ejemplar de la codificación estéreo de predicción compleja según la invención, se ha estimado que la complejidad computacional aumenta sólo ligeramente (aun significativamente menos que el aumento provocado por la codificación estéreo de predicción compleja en el dominio QMF) comparado con la estéreo L/R o M/S convencional.

Como un desarrollo adicional de la realización referida en el párrafo anterior, la etapa de mezcla ascendente puede comprender una etapa de transformada inversa adicional para procesar la señal lateral. Entonces, la etapa de suma y diferencia se suministra con una representación en el dominio del tiempo de la señal lateral, generada por dicha

etapa de transformada inversa adicional, y una representación en el dominio del tiempo de la señal de mezcla descendente, generada por la etapa de transformada inversa ya referida. Se recuerda que, de manera ventajosa a partir del punto de vista de la complejidad computacional, la última señal se suministra a tanto la etapa de suma y diferencia como a dicha etapa de transformada diferente referida a la anterior.

En una realización, un decodificador diseñado para emitir una señal estéreo en el dominio del tiempo incluye la posibilidad de conmutar entre la codificación estéreo L/R directa la codificación estéreo M/S conjunta y la codificación estéreo de predicción compleja. Esto se logra mediante la provisión de:

- un conmutador operable ya sea como una etapa de paso o como una etapa de suma y diferencia;
- una etapa de transformada inversa adicional para calcular una representación en el dominio del tiempo de la señal lateral;
- una disposición selectora para conectar las etapas de transformada inversa a ya sea una etapa de suma y diferencia adicional conectada a un punto por encima de la etapa de mezcla ascendente y por debajo del conmutador (preferiblemente cuando el conmutador ha sido accionado para funcionar como un filtro de paso, como puede ser el caso en la decodificación de una señal estéreo generada mediante codificación de predicción compleja) o una combinación de una señal de mezcla descendente desde el conmutador y una señal lateral desde el sumador pesado (preferiblemente cuando el conmutador se ha accionado para funcionar como una etapa de suma y diferencia, como puede ser el caso en la decodificación como una señal estéreo codificada directamente).

Como se darán cuenta las personas expertas, esto da al codificador la libertad para elegir entre la codificación directa o conjunta convencional y la codificación de predicción compleja lo que significa que se puede garantizar un nivel de calidad al menos equivalente a la de la codificación estéreo directa o conjunta,

En una realización, del sistema codificador según el segundo aspecto (no abarcado por las reivindicaciones) se puede comprender un estimador para estimar el coeficiente de predicción complejo con el objetivo de reducir o minimizar la energía de señal o la energía media de señal de la señal residual. La minimización puede tener lugar sobre un intervalo de tiempo, preferiblemente un segmento de tiempo o un bloque de tiempo o una trama de tiempo de la señal a ser codificada. El cuadrado de la amplitud se puede tomar como una medida de la energía momentánea de la señal, y una integral a lo largo del intervalo de tiempo de la amplitud al cuadrado (forma de onda) puede ser tomada como una medida de la energía de la señal media en ese intervalo. De manera adecuada, el coeficiente de predicción complejo se determina sobre una base de bloque de tiempo y banda de frecuencias, esto es, su valor se establece de manera tal que reduce la energía media (esto es, la energía total) de la señal residual en ese bloque de tiempo y banda de frecuencias. En concreto, los módulos para estimar los parámetros de la codificación estéreo paramétrica tales como la IID, la ICC y la IPD o similares, pueden proporcionar la salida sobre la que el coeficiente de predicción complejo puede ser calculado según las relaciones matemáticas conocidas por personas expertas.

En una realización, la etapa de codificación del sistema codificador (no abarcado por las reivindicaciones) es operable, además, para funcionar como una etapa de paso para permitir una codificación estéreo directa. Seleccionando la codificación estéreo directa en las situaciones en las que se espera que proporcione una mayor calidad, el sistema codificador puede garantizar que la señal estéreo codificada tiene al menos la misma calidad que en la codificación directa. De manera similar, en las situaciones en las que el mayor esfuerzo computacional incurrido por la codificación de predicción compleja no está motivado por un aumento significativo de la calidad, hay por tanto una opción disponible para economizar los recursos computacionales fácilmente para el sistema codificador. La decisión entre la codificación conjunta, directa, de predicción real y de predicción compleja en el codificador se basa generalmente en una optimización de la tasa/distorsión.

En una realización, el sistema codificador (no abarcado por las reivindicaciones) puede comprender un módulo para calcular directamente una segunda representación en el dominio de la frecuencia (esto es, sin aplicar una transformada inversa dentro del dominio del tiempo y sin usar los datos del dominio del tiempo de la señal) en base a los primeros componentes espectrales. En relación con las realizaciones correspondientes del sistema decodificador descrito anteriormente, este módulo puede tener una estructura análoga, es decir comprender las operaciones de procesamiento análogas pero en un orden diferente, por lo que el codificador se adapta para emitir los datos adecuados como entrada en el lado del decodificador. Con el propósito de ilustrar esta realización, se supone que la señal estéreo a ser codificada comprende los canales central y lateral, o se ha transformado dentro de esta estructura, y la etapa de codificación se adapta para recibir una primera representación en el dominio de la frecuencia. La etapa de codificación comprende un módulo para calcular una segunda representación en el dominio de la frecuencia del canal central. (La primera y segunda representaciones en el dominio de la frecuencia referidas aquí son como se define anteriormente; en concreto las primeras representaciones en el dominio de la frecuencia pueden ser representaciones MDCT y la segunda representación en el dominio de la frecuencia puede ser una representación MDST). La etapa de codificación comprende además un sumador ponderado para calcular la señal residual como una combinación lineal formada a partir de la señal lateral y las dos representaciones en el dominio de la frecuencia de la señal central ponderadas mediante las partes real e imaginaria, respectivamente, del coeficiente

de predicción complejo. La señal central, o de manera adecuada la primera representación en el dominio de la frecuencia de la misma, se puede usar de manera directa como una señal de mezcla descendente. En esta realización, además, el estimador determina el valor del coeficiente de predicción complejo con el objetivo de minimizar la energía o la energía media de la señal residual. La operación final (optimización) se puede efectuar ya sea mediante control de retroalimentación, en donde el estimador puede recibir la señal residual obtenida mediante los valores del coeficiente de predicción actuales a ser ajustador de manera adicional si fuera necesario, o, de una manera de alimentación hacia adelante, mediante los cálculos efectuados directamente o en los canales izquierdo/derecho de una señal estéreo original o los canales central/lateral. Se prefiere el método de alimentación hacia adelante, mediante el cual el coeficiente de predicción complejo se determina de manera directa (concretamente, de una manera no iterativa o no retroalimentada) en base a la primera y a la segunda representaciones en el dominio de la frecuencia de la señal central y la primera representación en el dominio de la frecuencia de la señal lateral. Se observa que la determinación del coeficiente de predicción complejo puede ser seguido por una decisión de ya sea aplicar la codificación directa, conjunta, de predicción real o de predicción compleja, en donde se considera la calidad resultante (preferiblemente la calidad perceptual, teniendo en cuenta, por ejemplo, los efectos señal a máscara) de cada opción disponible; por tanto los enunciados anteriores no se interpretan con el efecto de que exista ningún mecanismo de retroalimentación en el codificador.

En una realización, el sistema codificador (no abarcado por las reivindicaciones) comprende los módulos para calcular una segunda representación en el dominio de la frecuencia de la señal central (o de mezcla descendente) a través del dominio del tiempo. Se entiende que los detalles de la implementación en relación a esta realización, al menos hasta lo que al cálculo de la segunda representación en el dominio de la frecuencia se refiere, son similares o se pueden resolver de manera análoga a las correspondientes realizaciones del decodificador. En esta realización, la etapa de codificación comprende:

- una etapa de suma y diferencia para convertir la señal estéreo en una forma que comprende los canales central y lateral;
- una etapa de transformada para proporcionar una representación en el dominio de la frecuencia del canal lateral y una representación en el dominio de la frecuencia del valor complejo (y por lo tanto sobre muestreado) del canal central; y
- un sumador pesado para calcular una señal residual, en donde el coeficiente de predicción complejo se usa como un peso.

Aquí, el estimador puede recibir la señal residual y determinar, posiblemente en un modo de control de la retroalimentación, el coeficiente de predicción complejo para reducir o minimizar la energía o la media de la señal residual. Preferiblemente, sin embargo, el estimador recibe la señal estéreo a codificar y determina el coeficiente de predicción en base a esto. Es ventajoso a partir del punto de vista de la economía computacional usar una representación en el dominio de la frecuencia muestreada de manera crítica del canal lateral, ya que ésta última no estará sometida a la multiplicación por un número complejo en esta realización. De manera adecuada, la etapa de transformada puede comprender una etapa MDCT y una etapa MDST en paralelo, teniendo ambas la representación en el dominio del tiempo del canal central como entrada. Por tanto, se producen una representación en el dominio de la frecuencia sobre muestreada del canal central y una representación en el dominio de la frecuencia muestreada de manera crítica del canal lateral.

Se observa que los métodos y aparatos descritos en esta sección se pueden aplicar, después de las modificaciones apropiadas dentro de las capacidades de las personas expertas incluyendo la experimentación de rutina, para la codificación de señales que tengan más de dos canales. Las modificaciones en tal operatividad multicanal puede proceder, por ejemplo, a lo largo de las líneas de las secciones 4 y 5 en el documento de J. Herre et al. citado anteriormente.

Las características de las dos o más realizaciones indicadas anteriormente se pueden combinar, a menos que sean claramente complementarias, en realizaciones adicionales. El hecho de que dos características se reciten en reivindicaciones diferentes no impide que se puedan combinar como una ventaja. Así mismo, las realizaciones adicionales pueden proporcionar la omisión de ciertas características que no son necesarias o no son esenciales para el propósito deseado. Como un ejemplo, el sistema de decodificación según la invención se puede realizar sin una etapa de descuantificación en los casos en los que la señal codificada a ser procesada no se cuantifique o esté ya disponible en una forma adecuada para el procesamiento por la etapa de mezcla ascendente.

### Breve descripción de los dibujos

La invención será además ilustrada ahora por las realizaciones descritas en la siguiente sección, haciéndose referencia a los dibujos adjuntos, en los que:

la figura 1 consiste en dos diagramas de bloques generalizados que muestran los decodificadores basados en el QMF según los antecedentes de la técnica;

la figura 2 es un diagrama de bloques generalizado de un sistema decodificador estéreo basado en la MDCT con

predicción compleja, según una realización de la presente invención, en la que la representación compleja de una canal de la señal a ser decodificada se calcula en el dominio de la frecuencia;

la figura 3 es un diagrama de bloques generalizado de un sistema decodificador estéreo basado en la MDCT con predicción compleja, según una realización de la presente invención, en la que la representación compleja de una canal de la señal a ser decodificada se calcula en el dominio del tiempo;

la figura 4 muestra una realización alternativa del sistema decodificador de la figura 2, en el que la ubicación de la etapa TNS activa es seleccionable;

la figura 5 comprende diagramas de bloques generalizados que muestran los sistemas codificadores estéreo basados en la MDCT con predicción compleja, según las realizaciones de otro aspecto (no abarcado por las reivindicaciones);

la figura 6 es un diagrama de bloques generalizado de un codificador estéreo basado en la MDCT con predicción compleja, según una realización (no abarcada por las reivindicaciones), en la que se calcula una representación de un canal de la señal a ser codificada en base a la representación en el dominio del tiempo del mismo;

la figura 7 muestra una realización alternativa del sistema codificador (no abarcado por las reivindicaciones) de la figura 6, que es operable también en un modo de codificación L/R directo;

la figura 8 es un diagrama de bloques generalizado de un sistema codificador estéreo basado en la MDCT con predicción compleja (no abarcado por las reivindicaciones) en la que se calcula una representación de un canal de la señal a ser codificada en base a una primera representación en el dominio de la frecuencia del mismo, cuyo sistema decodificador es operable también en un modo de codificación L/R directo;

la figura 9 muestra una realización alternativa del sistema codificador (no abarcado por las reivindicaciones) de la figura 7, que incluye además una etapa TNS dispuesta debajo de la etapa de codificación;

la figura 10 muestra realizaciones alternativas de la parte etiquetada A en las figuras 2 y 8;

la figura 11 muestra una realización alternativa del sistema codificador (no abarcado por las reivindicaciones) de la figura 8, que además incluye dos dispositivos de modificación en el dominio de la frecuencia dispuestos respectivamente por debajo y por encima de la etapa de codificación;

la figura 12 es una representación gráfica de los resultados de la prueba de escucha a 96 kb/s a partir de seis sujetos que muestran diferentes opciones de compensación de complejidad-calidad para el cálculo o la aproximación del espectro de la MDST, en donde los puntos de datos etiquetados como "+" se refieren a la referencia oculta, "x" se refiere a una sujeción limitada en banda de 3.5 kHz, "\*" se refiere al estéreo tradicional USAC (M/S o L/R), "□" se refiere a la codificación estéreo unificada en el dominio MDCT mediante la predicción compleja con la parte imaginaria del coeficiente de predicción deshabilitada (esto es, predicción de valor real, que no requiere la MDST), "■" se refiere a la codificación estéreo unificada en el dominio MDCT mediante predicción compleja usando una trama MDCT actual para calcular una aproximación de la MDST, "○" se refiere a la codificación estéreo unificada en el dominio MDCT mediante predicción compleja que usa las tramas MDCT actual y anterior para calcular una aproximación de la MDST y "●" se refiere a la codificación estéreo unificada en el dominio MDCT mediante la predicción compleja que usa las tramas MDCT actual, anterior y siguiente para calcular la MDST;

la figura 13 presenta los datos de la figura 12, sin embargo como puntuaciones diferenciales en relación a la codificación estéreo unificada en el dominio MDCT mediante la predicción compleja que usa la trama MDCT actual para calcular una aproximación de la MDST;

la figura 14 comprende los diagramas de bloques generalizados que muestran tres realizaciones de un sistema decodificador según las realizaciones de la invención;

la figura 15 es un diagrama de flujo que muestra un método de decodificación según una realización de la invención; y

la figura 16 es un diagrama de flujo que muestra un método de codificación según una realización (no abarcada por las reivindicaciones).

## Descripción detallada de las realizaciones

### I. Sistemas decodificadores

La Figura 2 muestra, en forma de un diagrama de bloques generalizado, un sistema de decodificación para decodificar un flujo de bits que comprende al menos un valor de un coeficiente de predicción complejo  $\alpha = \alpha_R + i\alpha_I$  y una representación de la MDCT de una señal estéreo que tiene unos canales M de mezcla descendente y D residual. Las partes real e imaginaria  $\alpha_R$ ,  $\alpha_I$  del coeficiente de predicción se pueden cuantificar y/o codificar de

manera conjunta. Sin embargo preferiblemente, las partes real e imaginaria se cuantifican de manera independiente y uniforme, normalmente con un tamaño de paso de 0,1 (número sin dimensión). La resolución de la banda de frecuencias usada para el coeficiente de predicción complejo no es necesariamente la misma que la resolución para las bandas de los factores de escala (sfb; esto es, un grupo de las líneas de la MDCT que usan el mismo tamaño de paso de cuantificación y rango de cuantificación de la MDCT) según el estándar MPEG. En concreto, la resolución de la banda de frecuencias para el coeficiente de predicción puede ser una que esté justificada de manera psicoacústica, tal como la escala Bark. El demultiplexor 201 se adapta para extraer estas representaciones de la MDCT y el coeficiente de predicción (parte de la Información de Control como se indica en la figura) a partir del flujo de bits que se suministra a éste. En efecto, se puede codificar más información de control que simplemente el coeficiente de predicción complejo en el flujo de bits, por ejemplo, las instrucciones de si el flujo de bits se ha de decodificar en el modo de predicción o de no predicción, la información TNS, etc. La información TNS puede incluir valores de los parámetros TNS a ser aplicados por los filtros TNS (de síntesis) del sistema decodificador. Si se han de usar conjuntos idénticos de parámetros TNS para varios filtros TNS, tal como para ambos canales, es económico recibir esta información en la forma de un bit que indica dicha identidad de los conjuntos de parámetros en lugar de recibir los dos conjuntos de parámetros de manera independiente. Se puede incluir también información de si aplicar el TNS antes o después de la etapa de mezcla ascendente, según sea apropiado en base a, por ejemplo, una evaluación psicoacústica de las dos opciones disponibles. Además, entonces la información de control puede indicar anchos de banda limitados de manera individual para la mezcla descendente y para las señales residuales. Para cada canal, las bandas de frecuencias por encima de un límite de ancho de banda no serán decodificadas pero se establecerán a cero. En ciertos casos, las bandas de frecuencias más altas tienen un contenido de energía tan bajo que se cuantifican como cero. La práctica normal (cf, el parámetro max\_sfb en el estándar MPEG) ha sido usar la misma limitación de ancho de banda para tanto la mezcla descendente como las señales residuales. Sin embargo, la señal residual, en una mayor extensión que la señal de mezcla descendente, tiene su contenido de energía ubicado en las bandas de frecuencias inferiores. Por lo tanto, colocando un límite de ancho de banda superior dedicado en la señal residual, es posible una reducción de la tasa de bits con una pérdida de calidad no significativa. Por ejemplo, esto puede ser determinado por dos parámetros max\_sfb independientes codificados en el flujo de bits, uno para la señal de mezcla descendente y otro para la señal residual.

En esta realización, la representación de la MDCT de la señal estéreo se segmenta en sucesivas tramas de tiempo (o bloques de tiempo) que comprenden un número fijo de puntos de datos (por ejemplo, 1024 puntos), uno de entre varios números fijos de puntos de datos (por ejemplo, 128 o 1024 puntos) o un número variable de puntos. Como es sabido por aquellos expertos en la técnica, la MDCT se muestrea de manera crítica. La salida del sistema de decodificación, indicada en la parte derecha del dibujo, es una señal estéreo en el dominio del tiempo que tiene canales izquierdo L y derecho R. Los módulos 202 de descuantificación se adaptan para manejar la entrada de flujo de bits al sistema de decodificación o, donde sea apropiado, dos flujos de bits obtenidos después de la demultiplexación de un flujo de bits original y correspondiente a cada uno de entre la mezcla descendente y los canales residuales. Las señales del canal descuantificado se proporcionan a un ensamblaje 203 de conmutación operable ya sea en un modo de paso o un modo de suma y diferencia correspondiente a las matrices de transformación respectivas

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

y

$$\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Como se discutirá de manera adicional en el siguiente párrafo, el sistema decodificador incluye un segundo ensamblaje 205 de conmutación. Tanto el ensamblaje 203 como 205 de conmutación, como muchos otros conmutadores y ensamblajes de conmutación, son operables de una forma selectiva en frecuencia. Esto permite la decodificación de una gran variedad de modos de decodificación, por ejemplo, la decodificación L/R o M/S dependiente de la frecuencia, como es sabido en la técnica relacionada. Por tanto, el decodificador según la invención se puede considerar un súper conjunto en relación con la técnica relacionada.

Suponiendo por ahora que el ensamblaje 203 de conmutación está en el modo de paso, las señales del canal descuantificado son pasadas, en esta realización, a través de los respectivos filtros 204 TNS. Los filtros 204 TNS no son esenciales en el funcionamiento del sistema de decodificación y pueden ser reemplazados por elementos de paso. Después de esto, se suministra la señal al segundo ensamblaje 205 de conmutación que tiene la misma funcionalidad que el ensamblaje 203 de conmutación ubicado encima. Con las señales de entrada como se describe anteriormente y con el segundo ensamblaje 205 de conmutación fijado en su modo de paso, la salida del anterior es el canal de mezcla descendente y la señal del canal residual. La señal de mezcla descendente, aún representada mediante sus espectros MDCT sucesivos en el tiempo, se suministra a una transformada 206 de real a imaginario adaptada para calcular, basada en esto, los espectros MDST de la señal de mezcla descendente. En esta realización, una trama MDST se basa en tres tramas MDCT, una trama anterior, una trama actual (o contemporánea) y una trama posterior. Se indica de manera simbólica ( $Z^{-1}$ ,  $Z$ ) que el lado de la entrada de la transformada 206 de real a imaginario comprende los componentes de retardo.

La representación MDST de la señal de mezcla descendente obtenida a partir de la transformada 206 de real a imaginario es ponderada por la parte imaginaria  $\alpha_I$  del coeficiente de predicción y se añade a la representación MDCT de la señal de mezcla descendente ponderada mediante la parte real  $\alpha_R$  del coeficiente de predicción y la representación MDCT de la señal residual. Las dos sumas y multiplicaciones se realizan mediante los multiplicadores y sumadores 210, 211, que forman juntos (de manera funcional) un sumador ponderado, que se suministra con el valor del coeficiente  $\alpha$  de predicción complejo codificado en el flujo de bits recibido inicialmente por el sistema decodificador. El coeficiente de predicción complejo se puede determinar una vez por cada trama de tiempo. Se puede determinar también más a menudo, tal como una vez por cada trama de tiempo. Se puede determinar también más a menudo, tal como una vez por cada banda de frecuencias dentro de una trama, siendo las bandas de frecuencias una partición motivada de manera psico acústica. Se puede determinar también de manera menos frecuente, como se discutió anteriormente en conexión con los sistemas de codificación según la invención. La transformada 206 de real a imaginario se sincroniza con el sumador ponderado de tal manera que una trama MDST actual de la señal del canal de mezcla descendente se combina con unas tramas contemporáneas de la MDCT de cada señal del canal de mezcla descendente y de la señal del canal residual. La suma de estas tres señales es una señal lateral  $S = \text{Re}\{\alpha M\} + D$ . En esta expresión,  $M$  incluye tanto las representaciones MDCT como MDST de la señal de mezcla descendente, es decir  $M = M_{\text{MDCT}} - iM_{\text{MDST}}$ , mientras que  $D = D_{\text{MDCT}}$  es de valor real. Por tanto, se obtiene una señal estéreo que tiene un canal de mezcla descendente y un canal lateral, a partir del cual una transformada 207 de suma y diferencia restaura los canales izquierdo y derecho como sigue:

$$\begin{cases} L = M + S \\ R = M - S \end{cases}$$

Estas señales se representan en el dominio MDCT. El último paso del sistema de decodificación es aplicar una MDCT 209 inversa a cada uno de los canales, mediante lo cual se obtiene una representación en el dominio del tiempo de la señal estéreo izquierda/derecha.

Una posible implementación de la transformada 206 de real a imaginario se describe además en la patente de los EE.UU. 6.980.933 B2 del solicitante, como se observó anteriormente. Mediante la fórmula 41, la transformada se puede expresar como un filtro de respuesta al impulso finita, por ejemplo, para los puntos pares,

$$S(2v) = \frac{2}{N} \sum_{p=0}^{N-1} [(-1)^{p+1} X_I(p) + X_{III}(p)] h_{I,III}(2v - p) + \frac{4}{N} \sum_{l=1}^{N-1} X_{II}(2l+1) h_{II}(2v - 2l - 1),$$

donde  $S(2v)$  es el  $2v$ -ésimo punto de datos MDST,  $X_I$ ,  $X_{II}$ ,  $X_{III}$  son los datos MDCT de cada una de las tramas y  $N$  es la longitud de la trama. Además,  $h_I$ ,  $h_{II}$ ,  $h_{III}$  son respuestas al impulso que dependen de la función ventana aplicada, y son por lo tanto determinadas por cada elección de la función ventana, tal como rectangular, sinusoidal y Kaiser-Bessel-derivada, y para cada longitud de trama. La complejidad de este cálculo se puede reducir omitiendo aquellas respuestas al impulso que tienen un relativamente menor contenido de energía y contribuyen relativamente menos a los datos MDST. Como una alternativa o extensión a esta simplificación, las respuestas al impulso en sí se pueden acortar, por ejemplo, desde la longitud de trama total  $N$  a un número menor de puntos. Como ejemplo, la longitud de la respuesta al impulso se puede disminuir desde 1024 puntos (pulsaciones) a 10 puntos. El truncamiento más extremo que se puede considerar significativo es

$$S(v) = X_{II}(v+1) - X_{II}(v-1).$$

Otros enfoques directos se pueden encontrar en el documento de los EE.UU. 2005/0197831 A1.

Es posible además reducir la cantidad de datos de entrada sobre los cuales se basa el cálculo. Para ilustrar, la transformada 206 de real a imaginario y sus conexiones por encima, que se indican como la parte denotada por "A" en el dibujo, pueden ser reemplazadas por una variable simplificada, dos de las cuales  $A'$  y  $A''$  se muestran en la figura 10. La variable  $A'$  proporciona una representación imaginaria aproximada de la señal. Aquí, el cálculo de la MDST sólo toma la trama actual y la siguiente en cuenta. Con referencia a la fórmula anterior a este párrafo, esto puede ser efectuado mediante el establecimiento de  $X_{III}(p) = 0$  para  $p = 0, \dots, N-1$  (el índice  $III$  denota la trama de tiempo siguiente). Ya que la variable  $A'$  no requiere el espectro MDCT de la trama siguiente como entrada, el cálculo de la MDST no incurre en ningún retardo de tiempo. Claramente, esta aproximación reduce en cierto modo la precisión de la señal MDST obtenida, pero puede implicar también que la energía de esta señal resulta reducida; el último hecho puede ser compensado mediante un aumento de  $\alpha_I$  como resultado de la naturaleza de la codificación de predicción.

También se muestra en la figura 10 la variable  $A''$ , que usa como entrada sólo los datos MDCT para la trama de tiempo actual. La variable  $A''$  produce probablemente una representación de la MDST menos precisa que la variable  $A'$ . Por otro lado, funciona con retardo cero, como la variable  $A'$ , y tiene una menor complejidad computacional. Como ya se mencionó, las propiedades de la codificación de forma de onda no se ven afectadas siempre que se use

la misma aproximación en el sistema codificador y en el sistema decodificador.

Se ha de observar, independientemente de si se usa la variable  $A$ ,  $A'$  o  $A''$  o cualquier desarrollo adicional de la misma, que sólo aquellas partes del espectro MDST para las cuales la parte imaginaria del coeficiente de predicción complejo no es cero  $\alpha_i \neq 0$  necesitan ser calculadas. En circunstancias prácticas esto significará que el valor absoluto  $|\alpha_i|$  de la parte imaginaria del coeficiente es mayor que un valor de umbral predeterminado, lo que puede estar relacionado con el redondeo unitario del hardware usado. En caso de que la parte imaginaria del coeficiente sea cero para todas las bandas de frecuencias dentro de una trama de tiempo, no hay necesidad de calcular ningún dato MDST para esa trama. Por tanto, de manera adecuada, la transformada 206 de real a imaginario se adapta para responder a las ocurrencias de valores muy pequeños de  $|\alpha_i|$  no generando la salida de la MDST, mediante lo cual se pueden economizar los recursos de cálculo. En las realizaciones donde se usan más tramas que la actual para producir una trama de datos MDST, sin embargo, cualesquiera unidades por encima de la transformada 206 deberían continuar funcionando de manera adecuada incluso aunque no sea necesario ningún espectro MDST – en concreto, el segundo ensamblaje 205 de conmutación debería seguir enviando espectros MDCT – por lo que hay disponibles suficientes datos de entrada para la transformada 206 de real a imaginario ya cuando ocurra la siguiente trama de tiempo asociada con un coeficiente de predicción distinto de cero; esto puede por supuesto ser el siguiente bloque de tiempo.

Volviendo a la figura 2, la función del sistema de decodificación se ha descrito bajo la suposición de que ambos ensamblajes 203, 205 de conmutación se fijan en sus respectivos modos de paso. Como se discutirá ahora, el sistema decodificador puede decodificar también las señales que no están codificadas por predicción. Para este uso, el segundo ensamblaje 205 de conmutación se fijará en su modo suma y diferencia y de manera adecuada, como se indica en el dibujo, una disposición 208 selectora se fijará en su posición inferior, asegurando de este modo que las señales se proporcionen de manera directa a la transformada 209 inversa desde un punto de origen entre los filtros 204 de TNS y el segundo ensamblaje 205 de conmutación. Para asegurar la correcta decodificación, la señal adecuada tiene una forma L/R en el punto de origen. Por lo tanto, para asegurar que la transformada de real a imaginario se suministra con la señal central (esto es, la mezcla descendente) correcta en todo momento (en lugar de, de manera intermitente mediante una señal izquierda), el segundo ensamblaje 205 de conmutación se fija preferiblemente en su modo de suma y diferencia durante la decodificación de una señal estéreo no codificada por predicción. Como se observa anteriormente, la codificación de predicción puede ser reemplazada mediante codificación directa o conjunta convencional para ciertas tramas en base a, por ejemplo, una decisión entre la tasa de datos o la calidad del audio. El resultado de dicha decisión se puede comunicar desde el codificador al decodificador de varias maneras, por ejemplo, mediante el valor de un bit indicador dedicado en cada trama, o mediante la ausencia o la presencia de un valor del coeficiente de predicción. Habiendo establecido estos hechos, el rol del primer ensamblaje 203 de conmutación se puede conseguir fácilmente. De hecho, en el modo de codificación de no predicción, el sistema decodificador puede procesar ambas señales según la codificación estéreo (L/R) directa o la codificación (M/S) conjunta, y haciendo funcionar el primer ensamblaje 203 de conmutación ya sea en el modo de paso o de suma y diferencia, es posible asegurar que el punto de origen esté siempre provisto con una señal codificada de manera directa. Claramente, el ensamblaje 203 de conmutación cuando funciona como una etapa de suma y diferencia convierte una señal de entrada en forma M/S en una señal de salida (suministrada a los filtros 204 TNS opcionales) en forma L/R.

El sistema decodificador recibe una señal si una trama de tiempo concreta ha de ser decodificada por el sistema decodificador en el modo de codificación de predicción o el modo de codificación de no predicción. El modo de no predicción puede ser señalizado por el valor de un bit indicador dedicado en cada trama o mediante la ausencia (o el valor cero) del coeficiente de predicción. El modo de predicción se puede comunicar de manera análoga. Una implementación particularmente ventajosa, que permite retroceder sin ninguna sobrecarga, hace uso de un cuarto valor reservado del campo de dos bits `ms_mascara_presente` (véase la AAC MPEG-2, documento ISO/IEC 13818-7), que se transmite por trama de tiempo y se define como sigue:

Tabla 1: Definición de `ms_mascara_presente` en USAC

| Valor | Significado                                                    |
|-------|----------------------------------------------------------------|
| 00    | Codificación L/R para todas las bandas de frecuencias          |
| 01    | Se usa un bit de señalización por banda para indicar L/R o M/S |
| 10    | Codificación M/S para todas las bandas de frecuencias          |
| 11    | Reservado                                                      |

Redefiniendo el valor 11 para significar “codificación de predicción compleja”, el decodificador puede funcionar en todos los modos tradicionales, concretamente con codificación M/S y L/R, sin ninguna penalización y es capaz aún de recibir una señal que indique el modo de codificación de predicción compleja para las tramas relevantes.

La figura 4 muestra un sistema decodificador de la misma estructura general como el mostrado en la figura 2 pero

incluyendo, sin embargo, al menos dos estructuras diferentes. En primer lugar, el sistema de la figura 4 incluye los conmutadores 404, 411 que permiten la aplicación de algún paso de procesamiento que involucre la modificación en el dominio de la frecuencia por encima y por debajo de la etapa de mezcla ascendente. Esto se logra, por un lado, mediante un primer conjunto de modificadores 403 en el dominio de la frecuencia (dibujados con filtros de síntesis TNS en esta figura) proporcionado junto con el primer conmutador 404 por debajo de los módulos 401 de descuantificación y un primer ensamblaje 402 de conmutación pero por encima de un segundo ensamblaje 405 de conmutación dispuesto inmediatamente por encima de la etapa 406, 407, 408, 409 de mezcla ascendente. Por otro lado, el sistema decodificador incluye un segundo conjunto de modificadores 410 en el dominio de la frecuencia proporcionados junto con un segundo conmutador 411 por debajo de la etapa 406, 407, 408, 409 de mezcla ascendente pero por encima de una etapa 412 de transformada inversa. De manera ventajosa, como se indica en el dibujo, cada modificador en el dominio de la frecuencia se dispone en paralelo con una línea de paso que se conecta por encima del lado de entrada del modificador en el dominio de la frecuencia y se conecta por debajo del conmutador asociado. En virtud de esta estructura, el modificador en el dominio de la frecuencia se suministra con los datos de señal en todo momento, permitiendo el procesamiento en el dominio de la frecuencia en base a más tramas de tiempo que sólo la actual. La decisión de si aplicar el primer 403 o el segundo conjunto de modificadores 410 en el dominio del tiempo puede ser tomada por el codificador (y transportada en el flujo de bits), o puede estar basada en si se aplica la codificación de predicción, o puede estar basada en algún otro criterio encontrado adecuado en las circunstancias prácticas. Como ejemplo, si el modificador en el dominio de la frecuencia son filtros TNS, entonces puede ser ventajoso usar el primer conjunto 403 para algunos tipos de señales, mientras que el segundo conjunto 410 puede ser ventajoso para otros tipos de señales. Si el resultado de esta selección se codifica en el flujo de bits, entonces por consiguiente el sistema decodificador activará el conjunto respectivo de filtros TNS.

Para facilitar el entendimiento del sistema decodificador mostrado en la figura 4, se observa de manera explícita que la decodificación de una señal codificada de manera directa (L/R) tiene lugar cuando  $\alpha = 0$  (lo que implica que la pseudo-L/R y la L/R son idénticas y que los canales lateral y residual no difieren), el primer ensamblaje 402 de conmutación está en el modo de paso, el segundo ensamblaje de conmutación está en el modo de suma y diferencia, provocando de este modo que la señal tenga la forma M/S entre el segundo ensamblaje 405 de conmutación y la etapa 409 de suma y diferencia de la etapa de mezcla ascendente. Ya que la etapa de mezcla ascendente efectivamente será una etapa de paso, es inmaterial si (usando los respectivos conmutadores 404, 411) se activa el primer o el segundo conjunto de modificadores en el dominio de la frecuencia.

La figura 3 ilustra un sistema decodificador según una realización de la invención que, en relación con las figuras 2 y 4, representa un enfoque diferente a la provisión de los datos MDST requeridos para la mezcla ascendente. Como los sistemas decodificadores ya descritos, el sistema de la figura 3 comprende los módulos 301 de descuantificación, un primer ensamblaje 302 de conmutación operable en ya sea un modo de paso o de suma y diferencia y los filtros 303 TNS (de síntesis), que se disponen en serie desde el extremo de la entrada del sistema decodificador. Los módulos por debajo de este punto se utilizan de manera selectiva por medio de dos segundos conmutadores 305, 310, que se hacen funcionar preferiblemente de manera conjunta por lo que ambos están ya sea en sus posiciones superiores o en sus posiciones inferiores, como se indica en la figura. En el extremo de salida del sistema decodificador, hay una etapa 312 de suma y diferencia e, inmediatamente por encima de esta, dos módulos 306, 311 de MDCT inversa para transformar una representación en el dominio MDCT de cada canal en una representación en el dominio del tiempo.

En la decodificación de predicción compleja, donde el sistema decodificador se suministra con una codificación de flujo de bits una señal de mezcla descendente/residual estéreo y valores de un coeficiente de predicción complejo, el primer ensamblaje 302 de conmutación se fija en su modo de paso y los segundos conmutadores 305, 310 se fijan en su posición superior. Por debajo de los filtros TNS, los dos canales de la señal estéreo (descuantificada, filtrada TNS, MDCT) se procesan de diferentes maneras. El canal de mezcla descendente se proporciona, por otro lado, a un multiplicador y un sumador 308, que suma la representación MDCT del canal de mezcla descendente ponderada por la parte real  $\alpha_R$  del coeficiente de predicción a la representación MDCT del canal residual, y, por otro lado, a uno 306 de los módulos de transformada MDCT. La representación en el dominio del tiempo del canal M de mezcla descendente, que es la salida del módulo 306 de transformada MDCT inversa, se suministra tanto a la etapa 312 de suma y diferencia final como al módulo 307 de transformada MDST. Este doble uso de la representación en el dominio del tiempo del canal de mezcla descendente es ventajoso desde el punto de vista de la complejidad computacional. La representación MDST del canal de mezcla descendente así obtenida se suministra a un multiplicador y sumador 309 adicional, que después de ponderar mediante la parte imaginaria  $\alpha_I$  del coeficiente de predicción añade esta señal a la combinación lineal emitida desde el sumador 308; por tanto, la salida del sumador 309 es una señal de canal lateral,  $S = \text{Re}\{\alpha M\} + D$ . De manera similar al sistema decodificador mostrado en la figura 2, los multiplicadores y sumadores 308, 309 pueden ser combinados fácilmente para formar un sumador multi señal ponderado con entradas las representaciones MDCT y MDST de la señal de mezcla descendente, la representación MDCT de la señal residual y el valor del coeficiente de predicción complejo. Por debajo de este punto en la presente realización, sólo se mantiene un paso a través del módulo 311 de transformada MDCT antes de que se suministre la señal del canal lateral a la etapa 312 final de suma y diferencia.

La sincronización necesaria en el sistema decodificador se puede lograr aplicando las mismas longitudes de transformadas y formas de ventana a ambos módulos 306, 311 de transformada MDCT inversa, como es ya la práctica en la codificación M/S y L/R selectiva en frecuencia. Un retardo de una trama es introducido mediante la

combinación de ciertas realizaciones del módulo 306 MDCT inversa y las realizaciones del módulo 307 MDST. Por lo tanto, se proporcionan cinco bloques 313 de retardo opcionales (o instrucciones de software a este efecto en una implementación informática), por lo que la parte del sistema ubicada a la derecha de la línea discontinua se puede retrasar en una trama en relación a la parte izquierda cuando sea necesario. Aparentemente todas las intersecciones entre la línea discontinua y las líneas de conexión se proporcionan con bloques de retardo, con la excepción de la línea de conexión entre el módulo 306 de MDCT inversa y el módulo 307 de transformada MDST, que es donde surge el retardo que requiere compensación.

El cálculo de los datos MDST para una trama de tiempo requiere los datos de una trama de la representación en el dominio del tiempo. Sin embargo, la transformada MDCT inversa se basa en una (la actual), dos (preferiblemente; la anterior y la actual) o tres (preferiblemente: la anterior, la actual y la posterior) tramas consecutivas. En virtud de la bien conocida cancelación del efecto de solape en el dominio del tiempo (TDAC) asociado con la MDCT, la opción de tres tramas consigue un solapamiento completo de las tramas de entrada y por tanto proporciona la mejor (y posiblemente perfecta) precisión, al menos en las tramas que contienen el efecto de solape en el dominio del tiempo. Claramente, la MDCT inversa de tres tramas funciona con una trama de retardo. Aceptando usar una aproximación en el dominio del tiempo aproximada como entrada a la transformada MDST, se puede evitar este retardo y de este modo la necesidad de compensar los retardos entre las diferentes partes del sistema decodificador. En la opción de dos tramas, se produce el solapamiento/adición que permite la TDAC en la primera mitad de la trama, y el efecto de solape puede estar presente sólo en la última mitad. En la opción de una trama, la ausencia de la TDAC implica que el efecto de solape puede ocurrir en toda la trama; sin embargo, una representación MDST conseguida de esta manera, y usada como una señal intermedia en la codificación de predicción compleja, puede aún proporcionar una calidad satisfactoria.

El sistema de decodificación ilustrado en la figura 3 se puede hacer funcionar también en dos modos de decodificación de no predicción. Para decodificar una señal estéreo codificada L/R de manera directa, los segundos conmutadores 305, 310 se fijan en la posición inferior y el primer ensamblaje 302 de conmutación se fija en el modo de paso. Por tanto, la señal tiene la forma L/R por encima de la etapa 304 de suma y diferencia, que la convierte en la forma M/S, tras lo cual tiene lugar una transformada MDCT inversa y una operación de suma y diferencia final. Para decodificar una señal estéreo proporcionada en la forma codificada M/S de manera conjunta, el primer ensamblaje 302 de conmutación se fija en cambio en su modo de suma y diferencia, por lo que la señal tiene la forma L/R entre el primer ensamblaje 302 de conmutación y la etapa 304 de suma y diferencia, que es a menudo más adecuado desde el punto de vista del filtrado TNS de lo que sería una forma M/S. El procesamiento por debajo de la etapa 304 de suma y diferencia es idéntico al del caso de la decodificación L/R directa.

La figura 14 se compone de tres diagramas de bloques generalizados de decodificadores según las realizaciones de la invención. Al contrario de los otros diversos diagramas de bloques que acompañan esta solicitud, una línea de conexión en la figura 14 puede simbolizar una señal multi canal. En concreto, dicha línea de conexión se puede disponer para transmitir una señal estéreo que comprende los canales izquierdo/derecho, central/lateral, de mezcla descendente/residual, pseudo izquierdo/pseudo derecho y otras combinaciones.

La figura 14A muestra un sistema decodificador para decodificar una representación en el dominio de la frecuencia (indicada, para el propósito de esta figura, como una representación MDCT) de una señal de entrada. El sistema decodificador se adapta para suministrar como su salida una representación en el dominio del tiempo de una señal estéreo, que se genera en base a la señal de entrada. Para ser capaz de decodificar una señal de entrada codificada mediante codificación estéreo de predicción compleja, se proporciona el sistema decodificador con una etapa 1410 de mezcla ascendente. Sin embargo, es capaz también de manejar una señal de entrada codificada en otros formatos y posiblemente, que alterne entre varios formatos de codificación en el tiempo, por ejemplo, una secuencia de tramas de tiempo codificadas mediante codificación de predicción compleja puede ser seguida por una parte de tiempo codificada mediante codificación izquierda/derecha directa. La capacidad del sistema decodificador para manejar los diferentes formatos de codificación es lograda mediante la provisión de una línea de conexión (paso) dispuesta en paralelo con dicha etapa 1410 de mezcla ascendente. Por medio de un conmutador 1411 es posible seleccionar si la salida de la etapa 1410 de mezcla ascendente (posición de conmutador inferior en la figura) o la señal no procesada disponible en la línea de conexión (posición de conmutador superior en la figura) se ha de suministrar a los módulos decodificadores dispuestos por debajo. En esta realización, se dispone un módulo 1412 de MDCT inversa por debajo del conmutador, que transforma la representación MDCT de una señal en una representación en el dominio del tiempo. Como ejemplo, la señal suministrada en la etapa 1410 de mezcla ascendente puede ser una señal estéreo en forma de mezcla descendente/residual. La etapa 1410 de mezcla ascendente se adapta para derivar una señal lateral y para realizar una operación de suma y diferencia de modo que se emita una señal estéreo izquierda/derecha (en el dominio MDCT).

La figura 14B muestra un sistema decodificador similar al de la figura 14A. El presente sistema se adapta para recibir un flujo de bits como su señal de entrada. El flujo de bits es inicialmente procesado mediante un módulo 1420 demultiplexor y de descuantificación combinado, que proporciona, como una primera señal de salida, una representación MDCT de una señal estéreo multi canal para un tratamiento adicional, como es determinado por la posición de un conmutador 1422 que tiene una funcionalidad análoga que el conmutador 1411 de la figura 14A. Más precisamente, el conmutador 1422 determina si la primera salida desde el demultiplexor y descuantificador ha de ser procesada por una etapa 1421 de mezcla ascendente y un módulo 1423 de MDCT inversa (posición inferior) o por el

módulo 1423 de MDCT inversa sólo (posición superior). El módulo 1420 demultiplexor y de descuantificación combinado emite la información de control también. En el caso actual, la información de control asociada con la señal estéreo puede incluir los datos que indican si es adecuada la posición superior o inferior del conmutador 1422 para decodificar la señal o, de manera más abstracta, según a qué formato de codificación se ha de decodificar la señal estéreo. La información de control puede incluir parámetros para ajustar las propiedades de la etapa 1421 de mezcla ascendente, por ejemplo, un valor del coeficiente de predicción complejo  $\alpha$  usado en la codificación de predicción compleja como ya se describió anteriormente.

La figura 14C muestra un sistema decodificador que, además de las entidades análogas a aquellas de la figura 14B, comprende un primer y un segundo dispositivos 1431, 1435 de modificación en el dominio de la frecuencia dispuestos respectivamente por encima y por debajo de la etapa 1433 de mezcla ascendente. Para el propósito de esta figura, cada dispositivo de modificación en el dominio de la frecuencia es ilustrado por un filtro TNS. Sin embargo, por el término dispositivo de modificación en el dominio de la frecuencia se podría entender también otros procesos distintos del filtrado TNS que son susceptibles de ser aplicados bien antes o después de la etapa de mezcla ascendente. Ejemplos de modificaciones en el dominio de la frecuencia incluyen la predicción, la adición de ruido, la extensión del ancho de banda, y el procesamiento no lineal. Las consideraciones psico acústicas y razones similares, que posiblemente incluyen las propiedades de la señal a ser procesada y/o la configuración o los ajustes de dicho dispositivo de modificación en el dominio de la frecuencia, a veces indican que es ventajoso aplicar dicha modificación en el dominio de la frecuencia por encima de la etapa 1433 de mezcla ascendente en lugar de por debajo. En otros casos, se puede establecer mediante consideraciones similares que es preferible la posición por debajo de la modificación en el dominio de la frecuencia a la de por encima. Por medio de los conmutadores 1432, 1436, los dispositivos 1431, 1435 de modificación en el dominio de la frecuencia se pueden activar de manera selectiva de modo que, en respuesta a la información de control, el sistema decodificador puede seleccionar la configuración deseada. Como ejemplo, la figura 14C muestra una configuración en la que la señal estéreo del módulo 1430 demultiplexador y de descuantificación combinado es procesado inicialmente mediante un primer dispositivo 1431 de modificación en el dominio de la frecuencia, es después suministrado a la etapa 1433 de mezcla ascendente y es finalmente enviado de manera directa a un módulo 1437 de MDCT inversa, sin pasar a través del segundo dispositivo 1435 de modificación en el dominio de la frecuencia. Como se explica en la sección Compendio, esta configuración se prefiere sobre la opción de realizar el TNS después de la mezcla superior en la codificación de predicción compleja.

## II. Sistemas codificadores

Esta sección describe diversos sistemas codificadores. La descripción de los diversos sistemas codificadores es solamente con fines ilustrativos, y los diversos sistemas codificadores no están abarcados por las reivindicaciones. Un sistema codificador se describirá ahora con referencia a la figura 5, que es un diagrama de bloques generalizado de un sistema codificador para codificar una señal estéreo izquierda/derecha (L/R) como un flujo de bits de salida mediante la codificación de predicción compleja. El sistema codificador recibe una representación en el dominio del tiempo o en el dominio de la frecuencia de la señal y la suministra a tanto una etapa de mezcla descendente como a un estimador de coeficientes de predicción. Las partes real e imaginaria de los coeficientes de predicción se proporcionan a la etapa de mezcla descendente para controlar la conversión de los canales izquierdo y derecho en los canales de mezcla descendente y residual, que se suministran después a un multiplexador MUX final. Si la señal no se suministró como una representación en el dominio de la frecuencia al codificador, se transforma en dicha representación en la etapa de mezcla descendente o en el multiplexador.

Uno de los principios de la codificación de predicción es convertir la señal izquierda/derecha en la forma central/lateral, esto es,

$$\begin{cases} M = (L + R)/2 \\ S = (L - R)/2 \end{cases}$$

y después hacer uso de la correlación restante entre estos canales, es decir fijando

$$S = \text{Re}\{\alpha M\} + D,$$

donde  $\alpha$  es el coeficiente de predicción complejo a determinar y D es la señal residual. Es posible elegir  $\alpha$  para que la energía de la señal residual  $D = S - \text{Re}\{\alpha M\}$  se minimice. La minimización de la energía se puede efectuar con respecto a la energía momentánea, una energía a corto o largo plazo (energía media), que en el caso de una señal discreta equivale a la optimización en el sentido cuadrático medio.

Las partes real e imaginaria  $\alpha_R$ ,  $\alpha_I$  del coeficiente de predicción se pueden cuantificar y/o codificar de manera conjunta. Preferiblemente sin embargo, las partes real e imaginaria se cuantifican de manera independiente y uniforme, normalmente con un tamaño de paso de 0,1 (número sin dimensión). La resolución de la banda de frecuencias usada para el coeficiente de predicción complejo no es necesariamente la misma que la resolución para los factores de escala de las bandas (sfb; esto es, un grupo de líneas MDCT que usan el mismo tamaño de paso de cuantificación y rango de cuantificación) según el estándar MPEG. En concreto, la resolución de la banda de

frecuencias para el coeficiente de predicción puede ser una que esté justificada psico acústicamente, tal como la escala Bark. Se observa que la resolución de la banda de frecuencias puede variar en los casos en los que varíe la longitud de la transformada.

Como se observa ya, el sistema codificador puede tener la libertad de si aplicar codificación estéreo de predicción o no, implicando el último caso retroceder a la codificación L/R o M/S. Dicha decisión se puede tomar en base a la trama de tiempo o con más precisión, en base a la banda de frecuencias dentro de una trama de tiempo. Como se observa anteriormente, un resultado negativo de la decisión se puede comunicar a la entidad de decodificación de varias maneras, por ejemplo, mediante el valor de un bit indicador dedicado en cada trama, o mediante la ausencia (o con el valor cero) de un valor del coeficiente de predicción. La decisión positiva se puede comunicar de manera análoga. Una implementación particularmente ventajosa, que permite el retroceso sin ninguna sobrecarga, hace uso de un cuarto valor reservado del campo de dos bits `ms_mascara_presente` (véase la AAC MPEG-2, documento ISO/IEC 131818-7) que se transmite por trama de tiempo y se define como sigue:

Tabla 1: Definición de `ms_mascara_presente` en USAC

| Valor | Significado                                                    |
|-------|----------------------------------------------------------------|
| 00    | Codificación L/R para todas las bandas de frecuencias          |
| 01    | Se usa un bit de señalización por banda para indicar L/R o M/S |
| 10    | Codificación M/S para todas las bandas de frecuencias          |
| 11    | Reservado                                                      |

Redefiniendo el valor 11 para significar “codificación de predicción compleja”, el codificador se puede hacer funcionar en todos los modos tradicionales, concretamente con codificación M/S y L/R, sin ninguna penalización y es capaz aún de la codificación de predicción compleja para aquellas tramas en las que sea ventajoso.

La decisión sustancial se puede basar en la relación entre la tasa de datos a calidad de audio. Como una medición de la calidad, se pueden usar los datos obtenidos usando un modelo psico acústico incluido en el codificador (como a menudo es el caso de los codificadores de audio basados en las MDCT disponibles). En concreto, algunas realizaciones del codificador proporcionan una selección optimizada de la tasa de distorsión del coeficiente de predicción. Por consiguiente, en dichas realizaciones, la parte imaginaria – y posiblemente la parte real también – del coeficiente de predicción se fija a cero si el aumento en la ganancia de predicción no economiza suficientes bits para la codificación de la señal residual como para justificar el gasto de los bits requeridos para codificar el coeficiente de predicción.

Las realizaciones del codificador pueden codificar la información relacionada con el TNS en el flujo de bits. Dicha información puede incluir los valores de los parámetros TNS a aplicar por los filtros TNS (de síntesis) en el lado del decodificador. Si se han de usar idénticos conjuntos de parámetros TNS para ambos canales, es económico incluir un bit de señalización que indique esta identidad de los conjuntos de parámetros en lugar de transmitir los dos conjuntos de parámetros de manera independiente. Se puede incluir información también de si aplicar el TNS antes o después de la etapa de mezcla ascendente, según se basa de manera apropiada en, por ejemplo, la evaluación psico acústica de las dos opciones disponibles.

Como aún otra característica opcional, que es potencialmente beneficiosa desde el punto de vista de la complejidad y la tasa de bits, el codificador se puede adaptar para usar un ancho de banda limitado de manera individual para la codificación de la señal residual. Las bandas de frecuencias por encima de este límite no serán transmitidas al decodificador pero se fijarán a cero. En ciertos casos, las bandas de mayor frecuencia tienen un contenido de energía tan bajo que ya se cuantifican a cero. La práctica normal (cf. el parámetro `max_sfb` en el estándar MPEG) ha supuesto usar la misma limitación de ancho de banda para tanto la mezcla descendente como las señales residuales. Por lo tanto, colocando una banda dedicada superior con límite en la señal residual, es posible una reducción de la tasa de bits con una pérdida de calidad no significativa.

Por ejemplo, esto se puede lograr transmitiendo dos parámetros `max_sfb` independientes, uno para la señal de mezcla descendente y otro para la señal residual.

Se señala que si bien los problemas de la determinación óptima del coeficiente de predicción, la cuantificación y la codificación del mismo, el retroceso al modo M/S o L/R, el filtrado TNS y la limitación del ancho de banda superior etc. se discutieron con referencia al sistema decodificador mostrado en la figura 5, los mismos hechos son igualmente aplicables a las realizaciones que serán descritas a continuación con referencia a las figuras posteriores.

La figura 6 muestra otro sistema codificador adaptado para realizar la codificación estéreo de predicción compleja. El sistema recibe como entrada una representación en el dominio del tiempo de una señal estéreo segmentada en tramas de tiempo sucesivas, posiblemente solapadas, y que comprende los canales izquierdo y derecho. Una etapa 601 de suma y diferencia convierte la señal en los canales central y lateral. El canal central se suministra a tanto el

módulo 602 MDCT como al módulo 603 MDST, mientras que el canal lateral se suministra a un módulo 604 MDCT sólo. El estimador 605 del coeficiente de predicción estima para cada trama de tiempo – y posiblemente para las bandas de frecuencias individuales dentro de una trama – un valor del coeficiente  $\alpha$  de predicción complejo como se explica anteriormente. El valor del coeficiente  $\alpha$  se suministra como un peso a los sumadores 606, 607 ponderados, que forman una señal D residual como una combinación lineal de las representaciones MDCT y MDST de la señal central y la representación MDCT de la señal lateral. Preferiblemente, el coeficiente de predicción complejo se suministra como un peso a los sumadores 606, 607 ponderados representados mediante el mismo esquema de cuantificación que se usará cuando se codifique en el flujo de bits; esto de manera obvia proporciona una reconstrucción más fiel, ya que tanto el codificador como el decodificador aplican el mismo valor del coeficiente de predicción. La señal residual, la señal central (que se puede llamar de manera más apropiada señal de mezcla descendente cuando aparezca en combinación con una señal residual) y el coeficiente de predicción se suministran a una etapa 608 de cuantificación y multiplexor, que codifica esta y posible información adicional como un flujo de bits de salida.

La figura 7 muestra una variación al sistema codificador de la figura 6. Como está claro a partir de la similitud de los símbolos en la figura, tiene una estructura similar pero también la funcionalidad añadida de funcionar en un modo de retroceso de codificación L/R directa. El sistema codificador se acciona entre el modo de codificación de predicción complejo y el modo de retroceso por medio de un conmutador 710 proporcionado inmediatamente por encima de la etapa 709 de cuantificación y multiplexora combinada. En su posición superior, como se muestra en la figura, el conmutador 710 hará que el codificador funcione en el modo de retroceso. A partir de los puntos inmediatamente por debajo de los módulos 702, 704 MDCT, la señal central/lateral se suministra a una etapa 705 de suma y diferencia, que después de convertirla en la forma izquierda/derecha la pasa al conmutador 710, que la conecta a la etapa 709 de cuantificación y multiplexora combinada.

La figura 8 muestra un sistema codificador. Al contrario de los sistemas codificadores de las figuras 6 y 7, esta realización deriva los datos MDST requeridos para la codificación de predicción compleja directamente desde los datos MDCT, esto es, mediante una transformada de real a imaginario en el dominio de la frecuencia. La transformada de real a imaginario aplica cualquiera de los enfoques discutidos en conexión con los sistemas decodificadores de las figuras 2 y 4. Es importante hacer coincidir el método de cálculo del decodificador con el del codificador, para que se pueda lograr una decodificación fiel; preferiblemente, se usan métodos de transformada de real a imaginario idénticos en el lado del codificador y el en lado del decodificador. Como para las realizaciones del decodificador, la parte A encerrada por la línea discontinua y que comprende la transformada 804 de real a imaginario puede ser reemplazada por variantes aproximadas o por el uso de menos tramas de tiempo de entrada como entrada. Del mismo modo, la codificación se puede simplificar usando cualquiera de los otros enfoques de aproximación descritos anteriormente.

En un nivel superior, el sistema codificador de la figura 8 tiene una estructura que difiere de la que probablemente seguiría por una acción directa de reemplazo del módulo MDST en la figura 7 por un módulo de real a imaginario (conectado de manera adecuada). La presente arquitectura es limpia y logra la funcionalidad de conmutación entre la codificación de predicción y la codificación L/R directa de una manera robusta y computacionalmente económica. La señal estéreo de entrada es suministrada a los módulos 801 de transformada MDCT, que emiten una representación en el dominio de la frecuencia de cada canal. Esta se proporciona tanto a un conmutador 808 final para accionar el sistema codificador entre sus modos de codificación de predicción y directa como a una etapa 802 de suma y diferencia. En la codificación L/R directa o la codificación M/S conjunta – que se lleva a cabo en la trama de tiempo para la cual se fija el coeficiente de predicción  $\alpha$  a cero - esta realización somete la señal de entrada a sólo la transformación, cuantificación y multiplexación, MDCT estando los dos últimos pasos efectuados por una etapa 807 de cuantificación y multiplexora combinada dispuesta en el extremo de salida del sistema, donde se suministra el flujo de bits. En la codificación de predicción, cada uno de los canales sufren un procesamiento adicional entre la etapa 802 de suma y diferencia y el conmutador 808. A partir de la representación MDCT de la señal central, la transformada 804 de real a imaginario deriva los datos MDST y los envía a tanto el estimador 803 de coeficientes de predicción como el sumador 806 pesado. Como en los sistemas codificadores mostrados en las figuras 6 y 7, se usa un sumador 805 pesado adicional para combinar la señal lateral con las representaciones MDCT y MDST ponderadas de la señal central para formar una señal del canal residual, que se codifica junto con el canal central (esto es, de mezcla descendente) y el coeficiente de predicción mediante el módulo 807 de cuantificación y multiplexor combinado.

Volviendo ahora a la figura 9, se ilustrará que cada una de las realizaciones del sistema codificador se puede combinar con uno o más filtros TNS (de análisis). De acuerdo con las discusiones anteriores, a menudo es ventajoso aplicar un filtrado TNS a la señal en su forma de mezcla descendente. Por tanto, como se muestra en la figura 9, la adaptación del sistema codificador de la figura 7 para incluir el TNS es efectuado mediante la adición de los filtros 911 TNS inmediatamente por encima del módulo 909 de cuantificación y multiplexor combinado.

En lugar del filtro 911 b TNS derecho/residual, se pueden proporcionar dos filtros TNS separados (no mostrados) de manera inmediatamente por encima de la parte del conmutador 910 adaptados para manejar el canal derecho o residual. Por tanto, cada uno de los dos filtros TNS se suministrará con los respectivos datos de la señal de canal en todos los tiempos, permitiendo un filtrado TNS basado en más tramas de tiempo que sólo la actual. Como ya se ha observado, los filtros TNS son sólo un ejemplo de dispositivos de modificación en el dominio de la frecuencia,

dispositivos que de manera notable basan su procesamiento en más tramas que la actual, lo cual puede beneficiar de dicha ubicación tanto o más que hace el filtro TNS.

Como otra posible alternativa a la realización mostrada en la figura 9, los filtros TNS para la activación selectiva se pueden disponer en más de un punto para cada canal. Esto es similar a la estructura del sistema decodificador mostrado en la figura 4, donde los diferentes conjuntos de filtros TNS pueden ser conectados por medio de conmutadores. Esto permite la selección de la etapa disponible más adecuada para el filtrado TNS para cada trama de tiempo. En concreto, puede ser ventajoso para conmutar entre diferentes ubicaciones en conexión con la conmutación entre la codificación estéreo de predicción compleja y otros modos de codificación.

La figura 11 muestra una variación basada en el sistema codificador de la figura 8, en el que la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente es derivada por medio de una transformada 1105 de real a imaginario. De manera similar al sistema decodificador mostrado en la figura 4, este sistema decodificador incluye también módulos modificadores en el dominio de la frecuencia activables de manera selectiva, uno 1102 proporcionado por encima de la etapa de mezcla descendente y otro 1109 proporcionado por debajo de la misma. Los módulos 1102, 1109 en el dominio de la frecuencia, que se han ejemplificado en esta figura mediante filtros TNS, se pueden conectar en cada una de las rutas de señal usando los cuatro conmutadores 1103a, 1103b, 1109a y 1109b.

### III. Realizaciones de no aparatos

Esta sección describe métodos para decodificar y codificar. La descripción del método para codificar en esta sección es solamente con fines ilustrativos, y el método de codificación en esta sección no está abarcado por las reivindicaciones. Las realizaciones del tercer y el cuarto aspecto referidos previamente se muestran en las figuras 15 y 16. La figura 15 muestra un método para decodificar un flujo de bits en una señal estéreo, que comprende los siguientes pasos:

1. Un flujo de bits como entrada.
2. El flujo de bits se des cuantifica, mediante lo cual se obtienen una primera representación en el dominio de la frecuencia de los canales de mezcla descendente y residual de una señal estéreo.
3. Se calcula una segunda representación en el dominio de la frecuencia de la mezcla descendente.
4. Se calcula una señal de canal lateral en base a las tres representaciones en el dominio de la frecuencia de los canales.
5. Se calcula una señal estéreo, preferiblemente en la forma izquierda/derecha, en base a los canales lateral y de mezcla descendente.
6. Se emite la señal estéreo así obtenida.

Los pasos 3 a 5 se pueden considerar como un proceso de mezclado ascendente. Cada uno de los pasos 1 a 6 es análogo a la funcionalidad correspondiente en cualquiera de los sistemas decodificadores descritos en las partes anteriores de este texto, y los detalles adicionales en relación a su implementación se pueden recuperar en las mismas partes.

La figura 16 muestra un método (no abarcado por las reivindicaciones) para codificar una señal estéreo como una señal de flujo de bits, que comprende los siguientes pasos:

1. Se introduce una señal estéreo.
2. La señal estéreo se transforma en una primera representación en el dominio de la frecuencia.
3. Se determina un coeficiente de predicción complejo.
4. Se mezcla de manera descendente la representación en el dominio de la frecuencia.
5. Los canales de mezcla descendente y residual se codifican como un flujo de bits junto con el coeficiente de predicción complejo.
6. Se emite el flujo de bits.

Cada uno de los pasos 1 a 5 es análogo a la funcionalidad correspondiente en cualquiera de los sistemas codificadores en las partes anteriores de este texto, y detalles adicionales en relación a su implementación se pueden recuperar en las mismas partes.

Ambos métodos se pueden expresar como instrucciones legibles por un ordenador en la forma de programas de software y pueden ser ejecutadas por un ordenador. El alcance de protección de esta invención se extiende a dicho software y productos de programas informáticos para distribuir dicho software.

#### IV. Evaluación empírica

Varias de las realizaciones descritas en la presente memoria se han evaluado de manera empírica. Las partes más importantes del material experimental obtenido en este proceso se resumirán en esta subsección.

La realización usada para los experimentos tuvo las siguientes características.

- 5 (i) Cada espectro MDST (para una trama de tiempo) se calculó mediante un filtrado de respuesta al impulso finita de dos dimensiones de los espectros actual, anterior y siguiente .
- (ii) Se usó un modelo psico acústico del codificador estéreo USAC.
- 10 (iii) Las partes real e imaginaria del coeficiente  $\alpha$  de predicción complejo se transmitieron en lugar de los parámetros PS ICC, CLD e IPD. Las partes real e imaginaria se manejaron de manera independiente, fueron limitadas al rango  $[-3,0, 3,0]$  y se cuantificaron usando un tamaño de paso de 0,1. Después se codificaron de manera diferencial en el tiempo y finalmente se les realizó una codificación Huffman usando el libro de códigos de factores de escala de la USAC. Los coeficientes de predicción se actualizaron cada segundo en la banda del factor de escala, lo que resultó en una resolución de frecuencia similar a la de la Envolvente MPEG (véase, por ejemplo, ISO/IEC 23003-1). Este esquema de cuantificación y codificación resultó en una tasa media de bits de aproximadamente 2 kb/s para esta información lateral estéreo en una configuración típica con una tasa de bits objetivo de 96 kb/s.
- 15 (iv) El formato de flujo de bits se modificó sin romper los flujos de bits USAC actuales, ya que el elemento del flujo de bits de 2 bits `ms_mascara_presente` actualmente tiene sólo tres posibles valores. Usando un cuarto valor para indicar la predicción compleja permite un modo de retroceso de la codificación central/lateral básica sin ningún desperdicio de bits (para más detalles acerca de este tema, véase la subsección anterior de esta descripción).
- 20

Las pruebas de escucha se cumplieron según la metodología MUSHRA, implicando en concreto la reproducción sobre auriculares y el uso de 8 elementos de prueba con una tasa de muestreo de 48 kHz. Tres, cinco o seis sujetos de prueba participaron en cada prueba.

- 25 El impacto de las diferentes aproximaciones MDST se evaluó para ilustrar la compensación práctica entre complejidad y calidad que existe entre estas opciones. Los resultados se encontraron en las figuras 12 y 13, la primera mostrando las puntuaciones absolutas obtenidas y la última mostrando las puntuaciones diferenciales en relación a 96s USAC cp1f, esto es, la codificación estéreo unificada en el dominio MDCT mediante predicción compleja usando una trama MDCT actual para calcular una aproximación de la MDST. Se puede ver que la ganancia de calidad lograda por la codificación estéreo unificada basada en MDCT aumenta cuando se aplican enfoques computacionalmente más complejos para calcular el espectro MDST. Considerando la media sobre todas las pruebas, el sistema 96s USAC cp1f basado en trama única proporciona un aumento significativo en la eficiencia de codificación sobre la codificación estéreo convencional. En cambio, se obtiene incluso significativamente mejores resultados para la 96s USAC cp3f, es decir la codificación estéreo unificada en el dominio MDCT mediante predicción compleja usando las tramas MDCT actual, anterior y siguiente para calcular la MDST.
- 30
- 35

#### V. Realizaciones

Además, la invención se puede realizar como un sistema decodificador para decodificar una señal de flujo de bits en una señal estéreo mediante codificación estéreo de predicción compleja, comprendiendo el sistema decodificador:

- 40 una etapa (202; 401) de descuantificación para proporciona unas primeras representaciones en el dominio de la frecuencia de una señal (M) de mezcla descendente y una señal (D) residual en base a la señal de flujo de bits, comprendiendo cada una de las primeras representaciones en el dominio de la frecuencia primeros componentes espectrales que representan el contenido espectral de la correspondiente señal expresada en un primer subespacio de un espacio multidimensional, en donde las primeras componentes espectrales son coeficientes de transformada dispuestos en una o más tramas de tiempo de coeficientes de transformada, cada bloque generado mediante la aplicación de una transformada a un segmento de tiempo de una señal en el dominio del tiempo; y
- 45

una etapa (206, 207, 210, 211; 406, 407, 408, 409) de mezcla ascendente dispuesta debajo de la etapa de descuantificación, adaptada para generar la señal estéreo en base a la señal de mezcla descendente y la señal residual y que comprende:

- 50 un módulo (206; 408) para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente en base a la primera representación en el dominio de la frecuencia de la misma, comprendiendo la segunda representación en el dominio de la frecuencia segundos componentes espectrales que representan el contenido espectral de la señal expresada en un segundo subespacio del espacio multidimensional que incluye una parte del espacio multidimensional no incluida en el primer subespacio, estando dicho módulo adaptado para:

derivar uno o más primeros componentes intermedios desde al menos algunos de los primeros componentes espectrales;

formar una combinación de dichos uno o más primeros componentes espectrales según al menos una parte de una o más respuestas al impulso para obtener uno o más segundos componentes intermedios; y

5 derivar dicho uno o más segundos componentes espectrales de dichos uno o más segundos componentes intermedios; y

10 un sumador (210, 211; 406, 407) ponderado para calcular una señal lateral (S) en base a la primera y segunda representaciones en el dominio de la frecuencia de la señal de mezcla descendente, la primera representación en el dominio de la frecuencia de la señal residual y un coeficiente ( $\alpha$ ) de predicción complejo codificado en la señal de flujo de bits; y

una etapa de suma y diferencia (207; 409) para calcular la señal estéreo en base a la primera representación en el dominio de la frecuencia de la señal de mezcla descendente y la señal lateral.

15 Aún además, la invención se puede realizar como un sistema decodificador para decodificar una señal de flujo de bits en una señal estéreo mediante codificación estéreo de predicción compleja, comprendiendo el sistema decodificador:

20 una etapa (301) de descuantificación para proporcionar las primeras representaciones en el dominio de la frecuencia de una señal (M) de mezcla descendente y una señal (D) residual en base a la señal de flujo de bits, comprendiendo cada una de las primeras representaciones en el dominio de la frecuencia primeros componentes espectrales que representan el contenido espectral de la señal correspondiente expresada en un primer subespacio de un espacio multidimensional; y

una etapa (306, 307, 308, 309, 312) de mezcla ascendente dispuesta debajo de la etapa de descuantificación, adaptada para generar la señal estéreo en base a la señal de mezcla descendente y la señal residual y que comprende:

25 un módulo (306, 307) para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente en base a la primera representación en el dominio de la frecuencia de la misma, comprendiendo la segunda representación en el dominio de la frecuencia segundos componentes espectrales que representan el contenido espectral de la señal expresada en un segundo subespacio del espacio multidimensional que incluye una parte del espacio multidimensional no incluida en el primer subespacio, comprendiendo el módulo:

30 una etapa (306) de transformada inversa para calcular una representación en el dominio del tiempo de la señal de mezcla descendente en base a la primera representación en el dominio de la frecuencia de la señal de mezcla descendente en el primer subespacio del espacio multidimensional; y

una etapa (307) de transformada para calcular la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente en base a la representación en el dominio del tiempo de la señal;

35 un sumador ponderado para calcular una señal lateral combinando la primera representación en el dominio de la frecuencia de la señal residual, la primera representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor real de un coeficiente de predicción complejo codificado en una señal del flujo de bits; y la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor imaginario del coeficiente de predicción complejo; y

40 una etapa (312) de suma y diferencia para calcular la señal estéreo en base a la primera representación de la señal de mezcla descendente y la señal lateral.

De manera adicional, la invención se puede realizar como un sistema decodificador con las características expuestas en la reivindicación independiente del sistema decodificador, en donde el módulo para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente comprende:

45 una etapa (306) de transformada inversa para calcular una representación en el dominio del tiempo de la señal de mezcla descendente y/o la señal lateral en base a la primera representación en el dominio de la frecuencia de la respectiva señal en el primer subespacio del espacio multidimensional; y

una etapa (307) de transformada para calcular la segunda representación en el dominio de la frecuencia de la señal respectiva en base a la representación en el dominio del tiempo de la señal,

50 en donde, preferiblemente, la etapa (306) de transformada inversa realiza una transformada de coseno discreta modificada y la etapa de transformada realiza una transformada de seno discreta modificada.

En el sistema decodificador anterior, la señal estéreo se puede representar en el dominio del tiempo y el sistema

decodificador puede comprender, además:

un ensamblaje (302) de conmutación dispuesto entre dicha etapa de descuantificación y dicha etapa de mezcla ascendente, operable para funcionar como bien:

- (a) una etapa de paso, para usar en codificación estéreo conjunta; o
- 5 (b) una etapa de suma y diferencia, para usar en codificación estéreo directa;

una etapa (311) de transformada inversa adicional dispuesta en la etapa de mezcla ascendente, para calcular una representación en el dominio del tiempo de la señal lateral;

una disposición (305, 310) de selector dispuesta por encima de las etapas (306, 301) de transformada inversa, adaptada para conectar de manera selectiva éstas a bien:

- 10 (a) una etapa (304) de suma y diferencia adicional que se conecta a su vez a un punto por debajo del ensamblaje (302) de conmutación y por encima de la etapa de mezcla ascendente; o
- (b) una señal de mezcla descendente obtenida desde el ensamblaje (302) de conmutación y una señal lateral obtenida desde el sumador (308, 309) ponderado.

## VI. Observaciones finales

15 Las realizaciones adicionales de la presente invención resultarán evidentes para una persona experta en la técnica después de la lectura de la descripción anterior. Incluso aunque la presente descripción y los dibujos describen las realizaciones y ejemplos, la invención no se restringe a estos ejemplos específicos. Se pueden hacer numerosas modificaciones y variaciones sin salir del alcance de la presente invención, que se define por las reivindicaciones adjuntas.

20 Se observa que los métodos y aparatos descritos en esta solicitud se pueden aplicar, después de las modificaciones apropiadas dentro de las capacidades de una persona experta que incluyen la experimentación de rutina, para la codificación de señales que tienen más de dos canales. Se enfatiza de manera particular que cualesquiera señales, parámetros y matrices mencionadas en las conexiones con las realizaciones descritas pueden ser variantes en frecuencia o invariantes en frecuencia y/o variantes en el tiempo o invariantes en el tiempo. Los pasos de  
25 computación descritos se pueden llevar a cabo para una frecuencia o para todas las bandas de frecuencias en cada momento, y todas las entidades se pueden realizar para tener una acción selectiva en frecuencia. Para los propósitos de la solicitud, cualquier esquema de cuantificación se puede adaptar según los modelos psico acústicos. Se observa además que las diversas conversiones de suma y diferencia, esto es, la conversión de la forma de mezcla descendente/residual a la forma pseudo L/R así como la conversión de L/R a M/S y la conversión de M/S a  
30 L/R, son todas de la forma

$$g \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

donde, principalmente, el factor g de ganancia puede variar. Por tanto, ajustando los factores de ganancia de manera individual, es posible compensar una cierta ganancia de codificación mediante la elección apropiada de la ganancia de decodificación. Además, como se darán cuenta las personas expertas en la técnica, un número par de  
35 transformadas de suma y diferencia dispuestas en serie tienen el efecto de una etapa de paso, posiblemente con una ganancia no unitaria.

Los sistemas y métodos descritos aquí anteriormente se pueden implementar como software, firmware, hardware o una combinación de los mismos. Ciertos componentes o todos los componentes se pueden implementar como un software ejecutado mediante un procesador o microprocesador de señal digital, o ser implementados como  
40 hardware o como un circuito integrado de aplicación específica. Dicho software se puede distribuir en medios legibles por ordenador, que pueden comprender medios de almacenamiento y medios de comunicación informáticos. Como es sabido por una persona experta en la técnica, los medios de almacenamiento informáticos incluyen tanto los volátiles como los no volátiles, los medios extraíbles como los no extraíbles implementados en cualquier método o tecnología para el almacenamiento de la información tal como las instrucciones, las estructuras de datos, los  
45 módulos de programa u otros datos legibles por un ordenador. Los medios de almacenamiento informáticos incluyen, pero no se limitan a, una RAM, una ROM, una EEPROM, una memoria flash u otra tecnología de memoria, un CD-ROM, discos versátiles digitales (DVD) u otro almacenamiento en discos ópticos, casetes magnéticos, una cinta magnética, un almacenamiento de discos magnéticos u otros dispositivos de almacenamiento magnético, o cualquier otro medio que se pueda usar para almacenar la información deseada y que pueda ser accedido por un  
50 ordenador. Además, es sabido por una persona experta que los medios de comunicación normalmente realizan instrucciones, estructuras de datos, módulos de programa u otros datos legibles por un ordenador en una señal de datos modulada tal como una onda portadora u otro mecanismo de transporte e incluye cualquier medio de entrega de información.

## REIVINDICACIONES

1. Un sistema decodificador para proporcionar una señal estéreo mediante codificación estéreo de predicción compleja, comprendiendo el sistema decodificador:

- 5 una etapa (206, 207, 210, 211; 306, 307, 308, 309, 312; 406, 407, 408, 409; 1410; 1421; 1433) de mezcla ascendente adaptada para generar la señal estéreo en base a las primeras representaciones en el dominio de la frecuencia de una señal (M) de mezcla descendente y una señal (D) residual, comprendiendo cada una de las representaciones en el dominio de la frecuencia primeros componentes espectrales que representan el contenido espectral de la señal correspondiente expresada en un primer subespacio de un espacio multidimensional, comprendiendo la etapa de mezcla ascendente:
- 10 un módulo (206; 306, 307; 408) para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente en base a la primera representación en el dominio de la frecuencia de la misma, comprendiendo la segunda representación en el dominio de la frecuencia los segundos componentes espectrales que representan el contenido espectral de la señal expresada en un segundo subespacio del espacio mutidimensional que incluye una parte del espacio multidimensional no incluida en el primer subespacio;
- 15 un sumador (210, 211; 308, 309; 406, 407) ponderado para calcular una señal lateral (S) combinando la primera representación en el dominio de la frecuencia de la señal residual, la primera representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor real de un coeficiente de predicción complejo codificado en una señal del flujo de bits; y la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor imaginario del coeficiente de predicción complejo; y
- 20 una etapa (207; 312; 409) de suma y diferencia para calcular la señal estéreo en base a la primera representación en el dominio de la frecuencia de la señal de mezcla descendente y la señal lateral,
- 25 en donde la etapa de mezcla ascendente es operable además en un modo de paso, en el cual las señales de mezcla descendente y residual se suministran a la etapa de suma y diferencia de manera directa.

2. El sistema decodificador de la reivindicación 1, en donde dichas señales de mezcla descendente y residual se segmentan en tramas de tiempo,

- 30 estando dicha etapa de mezcla ascendente adaptada además para recibir, para cada trama de tiempo, un campo de datos de dos bits asociado con esa trama y para operar, en respuesta al valor del campo de datos, en su modo activo o su modo de paso.

3. El sistema decodificador de la reivindicación 1, en donde dichas señales de mezcla descendente y residual son segmentadas en tramas de tiempo,

- 35 estando dicha etapa de mezcla ascendente adaptada además para recibir en un flujo de bits MPEG, para cada trama de tiempo, un campo ms\_mascara\_presente asociado con esa trama y para operar, en respuesta al valor del campo ms\_mascara\_presente, en su modo activo o su modo de paso.

4. El sistema decodificador de las reivindicaciones anteriores, en donde:

- los primeros componentes espectrales tienen valores reales expresados en el primer subespacio; y
- los segundos componentes espectrales tienen valores imaginarios expresados en el segundo subespacio;
- opcionalmente, los primeros componentes espectrales se pueden obtener mediante uno de lo siguiente:
- 40 una transformada de coseno discreta, DCT, o
- una transformada de coseno discreta modificada, MDCT,
- y, de manera opcional, los segundos componentes espectrales se pueden obtener mediante uno de lo siguiente:
- una transformada de seno discreta, DST, o
- una transformada de seno discreta modificada, MDST.

45 5. El sistema decodificador de cualquiera de las reivindicaciones anteriores, en donde:

- los primeros componentes espectrales son coeficientes de transformada dispuestos en una o más tramas de tiempo de coeficientes de transformada, cada bloque generado mediante la aplicación de una transformada a un segmento de tiempo de una señal en el dominio del tiempo; y

el módulo para calcular una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente está adaptado para:

derivar uno o más primeros componentes intermedios desde al menos algunos primeros componentes espectrales;

5 formar una combinación de dichos uno o más primeros componentes espectrales según al menos una parte de una o más respuestas al impulso para obtener uno o más segundos componentes intermedios; y

derivar dicho uno o más segundos componentes espectrales de dicho uno o más segundos componentes intermedios.

10 6. El sistema decodificador de cualquiera de las reivindicaciones anteriores, estando dicha señal estéreo representada en el dominio del tiempo y comprendiendo además el sistema decodificador:

un ensamblaje (203) de conmutación dispuesto entre dicha etapa de descuantificación y dicha etapa de mezcla ascendente, operable para funcionar como bien:

(a) una etapa de paso, o

(b) una etapa de suma y diferencia,

15 permitiendo de este modo la conmutación entre las señales de entrada estéreo codificadas de manera directa y conjunta;

una etapa (209) de transformada inversa adaptada para calcular una representación en el dominio del tiempo de la señal estéreo; y

20 una disposición (208) de selector dispuesta por encima de la etapa de transformada inversa, adaptada para conectar de manera selectiva éste a bien:

(a) un punto debajo de la etapa de mezcla ascendente, por el cual se suministra la señal estéreo obtenida mediante predicción compleja a la etapa de transformada inversa; o

25 (b) un punto debajo del ensamblaje (203) de conmutación y por encima de la etapa de mezcla ascendente, por el cual se suministra una señal estéreo obtenida mediante codificación estéreo directa a la etapa de transformada inversa.

7. Un método de decodificación para proporcionar una señal estéreo mediante codificación estéreo de predicción compleja, comprendiendo el método los pasos de:

30 recepción de las primeras representaciones en el dominio de la frecuencia de una señal (M) de mezcla descendente y una señal (D) residual, comprendiendo cada una de las primeras representaciones en el dominio de la frecuencia las primeras componentes espectrales que representan el contenido espectral de la señal correspondiente expresada en un primer subespacio de un espacio multidimensional;

recepción de una señal de control; y

mezclado ascendente de dichas señales de mezcla descendente y residual, usando una etapa de mezcla ascendente, para obtener dicha señal estéreo mediante, en respuesta al valor de dicha señal de control:

35 (a) la realización de los sub-pasos de:

cálculo de una segunda representación en el dominio de la frecuencia de la señal de mezcla descendente en base a la primera representación en el dominio de la frecuencia de la misma, comprendiendo la segunda representación en el dominio de la frecuencia los segundos componentes espectrales que representan el contenido espectral de la señal expresada en un segundo subespacio del espacio multidimensional que incluye una parte del espacio multidimensional no incluido en el primer subespacio;

40 cálculo, mediante un sumador ponderado, de una señal lateral (S) combinando la primera representación en el dominio de la frecuencia de la señal residual, la primera representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor real de un coeficiente de predicción complejo y la segunda representación en el dominio de la frecuencia de la señal de mezcla descendente ponderada por una parte de valor imaginario del coeficiente de predicción complejo; y

45 cálculo de dicha señal estéreo aplicando una transformada de suma y diferencia a la primera representación en el dominio de la frecuencia de la señal de mezcla descendente y la señal lateral, o

(b) el cálculo de dicha señal estéreo aplicando una transformada de suma y diferencia de manera directa a las primeras representaciones en el dominio de la frecuencia de la señal de mezcla descendente y la señal residual.

50

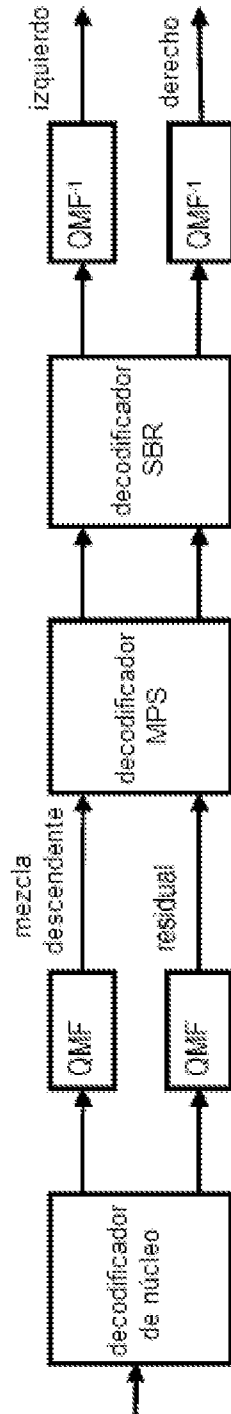


Fig. 1A

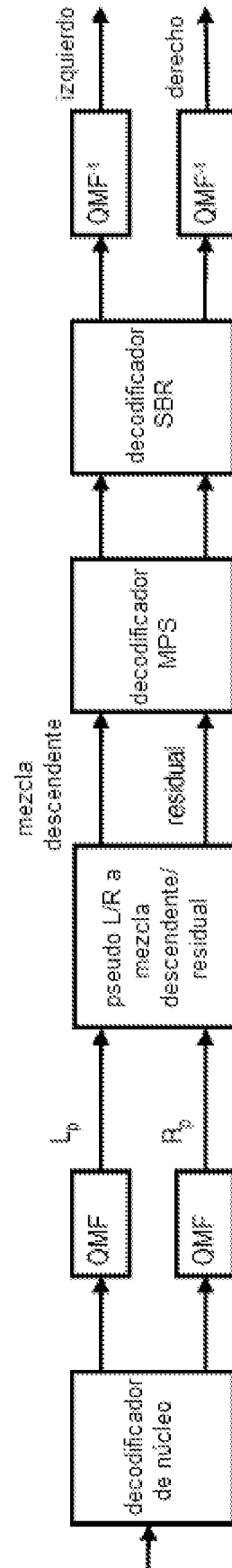


Fig. 1B

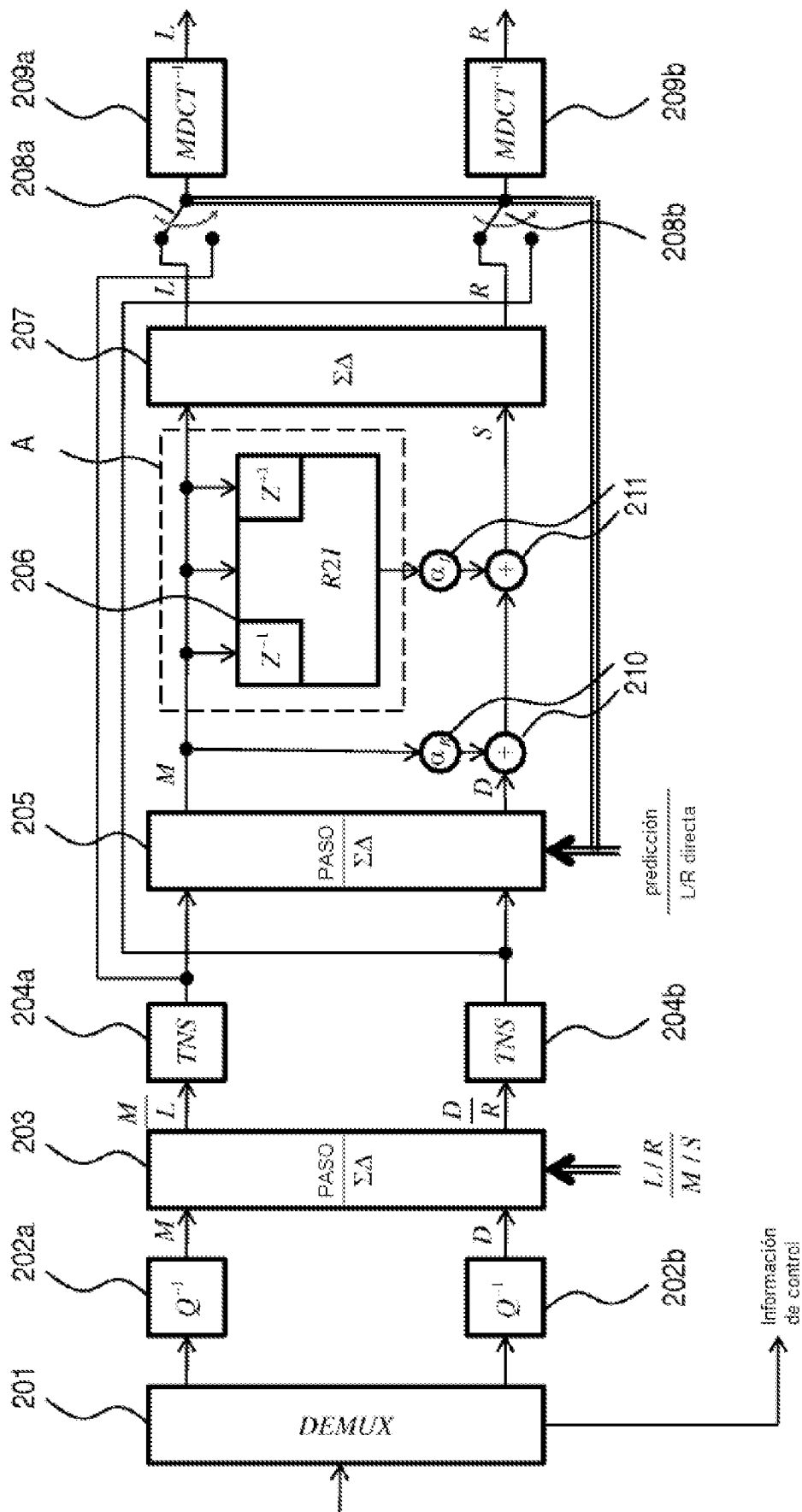


Fig. 2

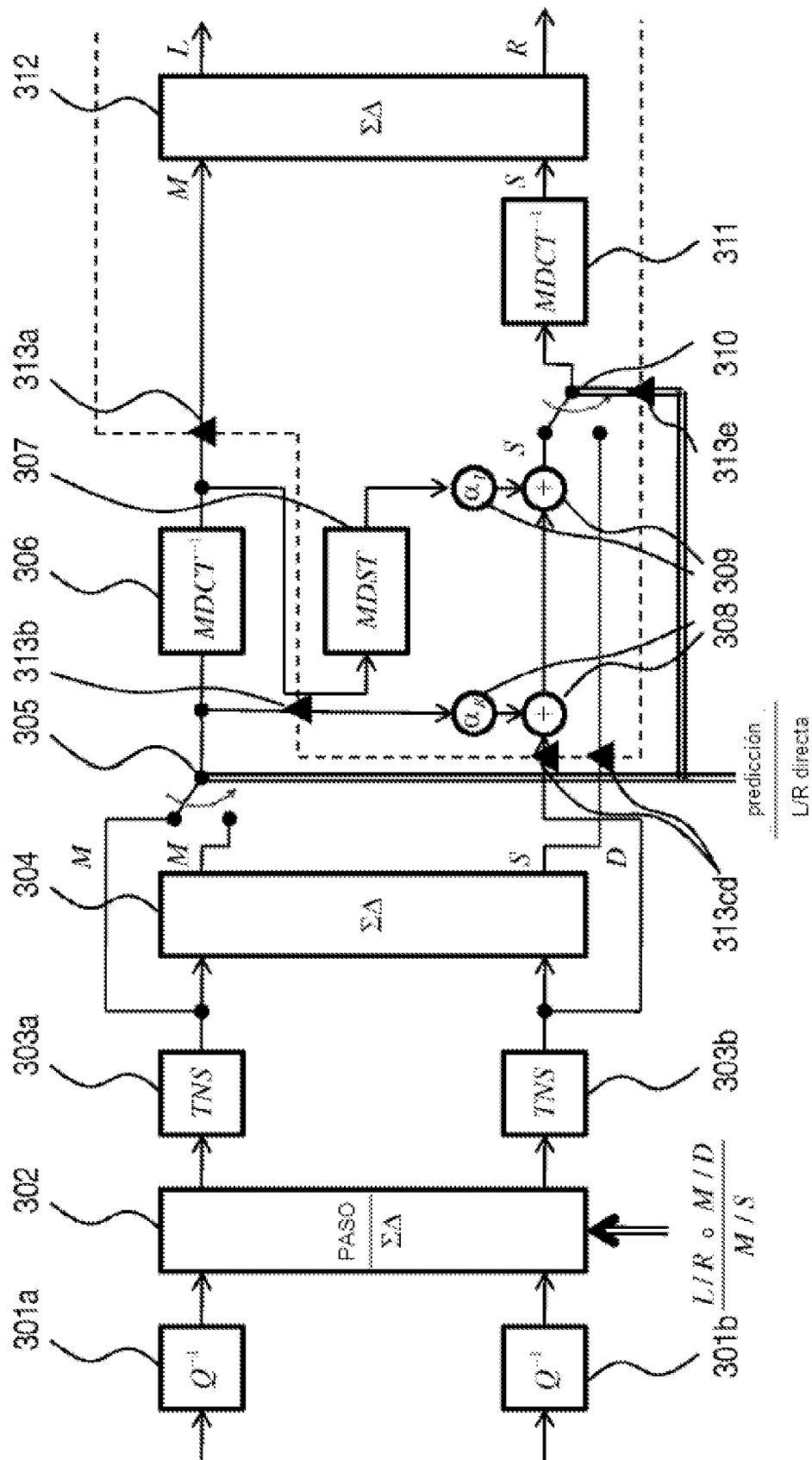


Fig. 3

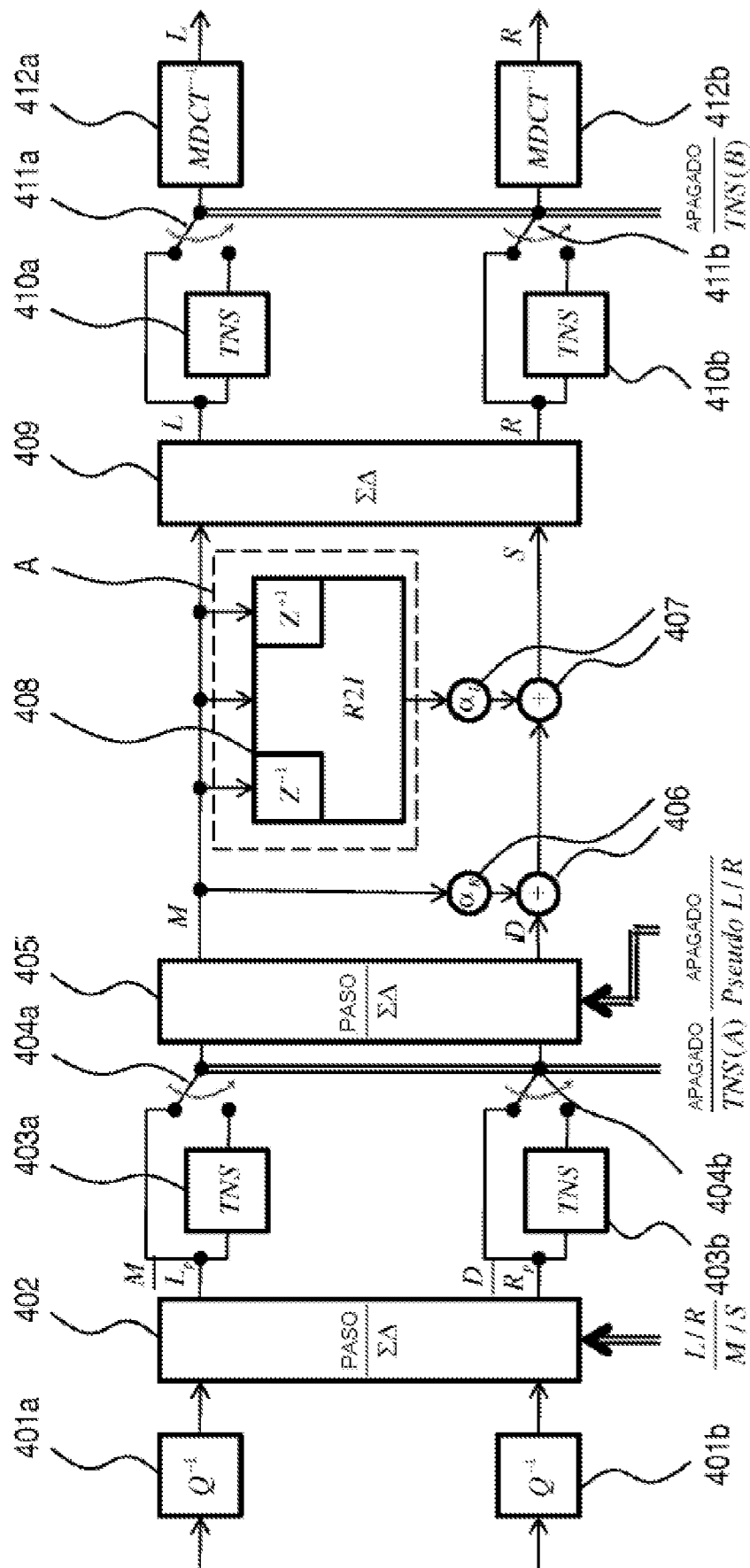


**Fig. 4**

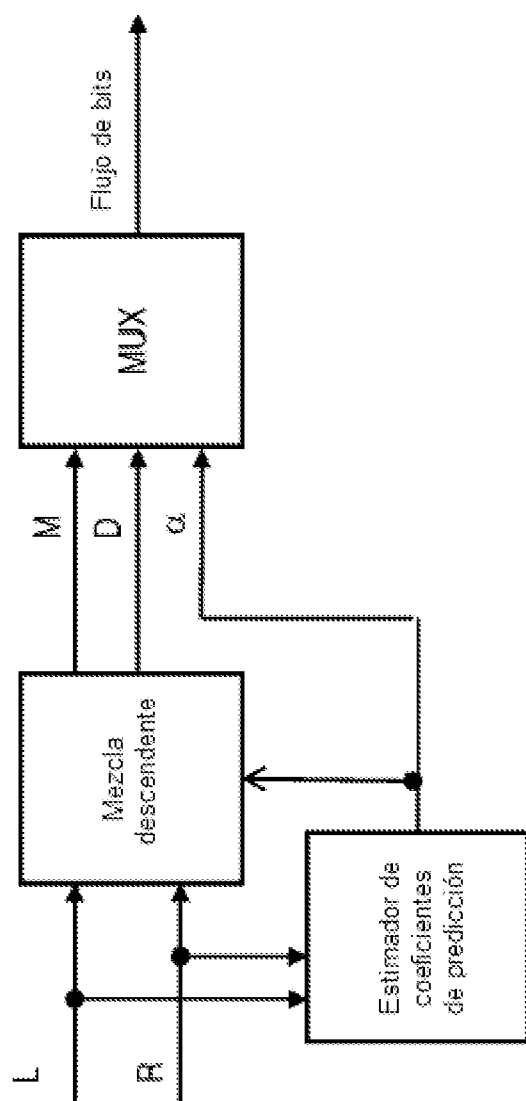


Fig. 5

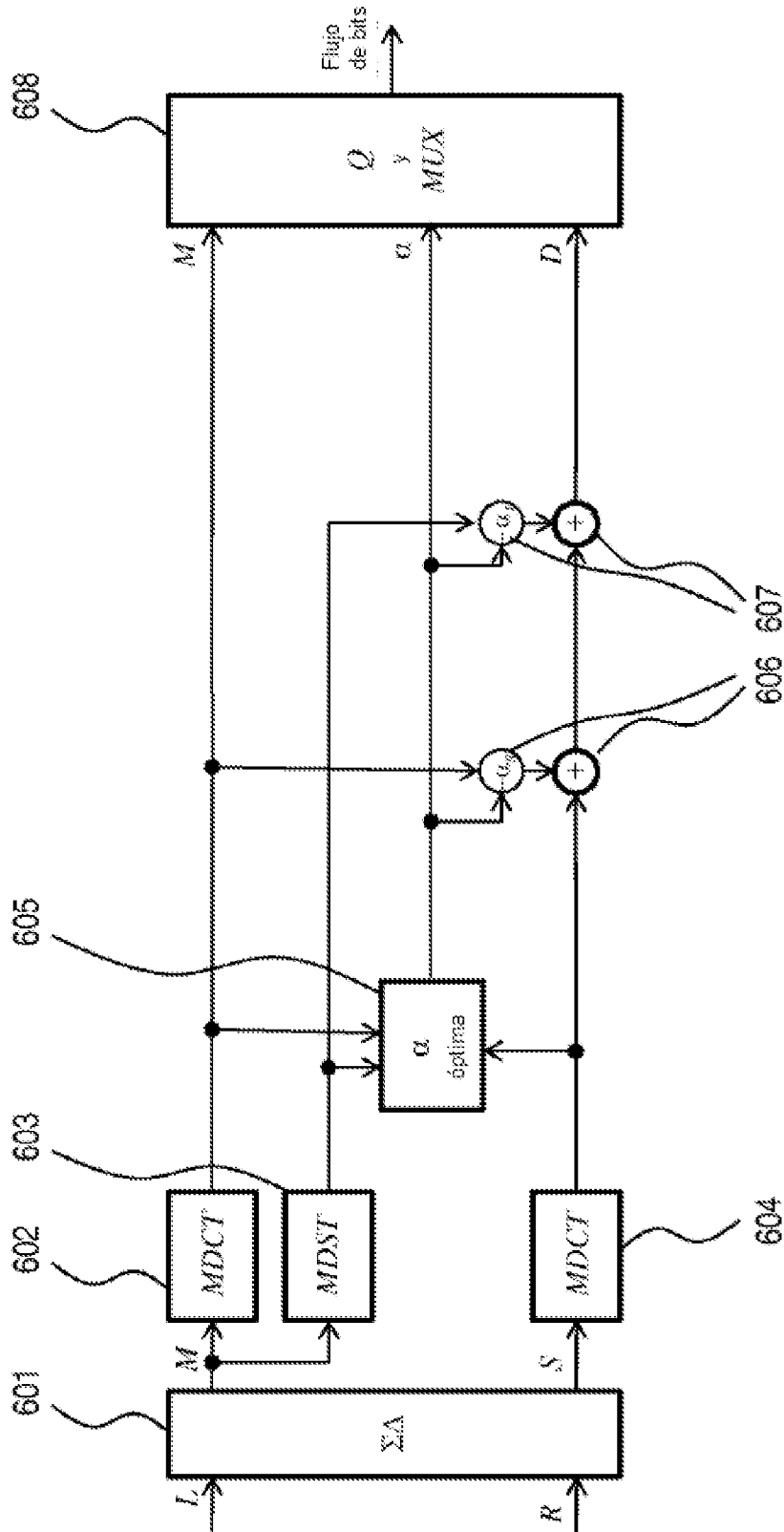


Fig. 6

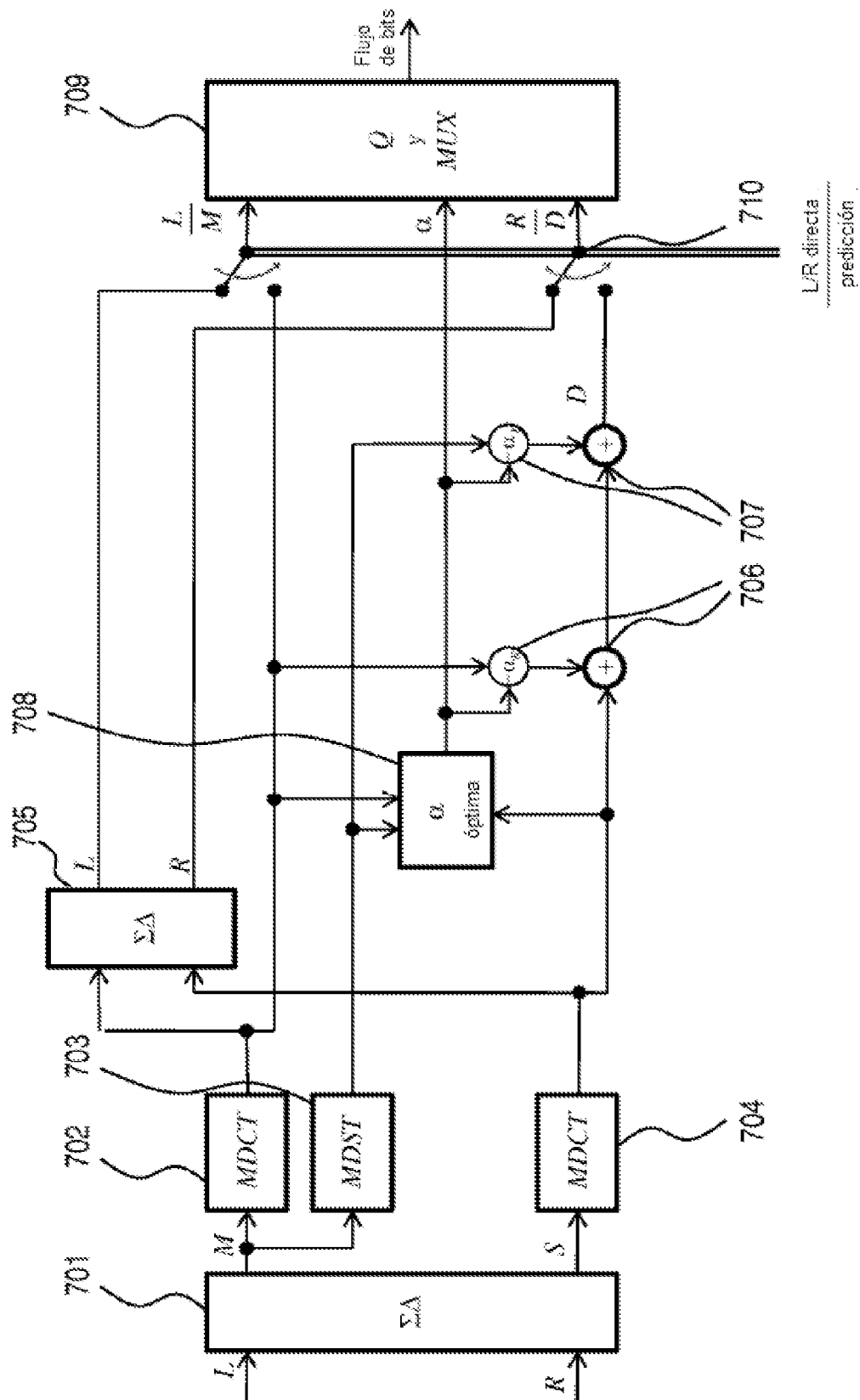


Fig. 7

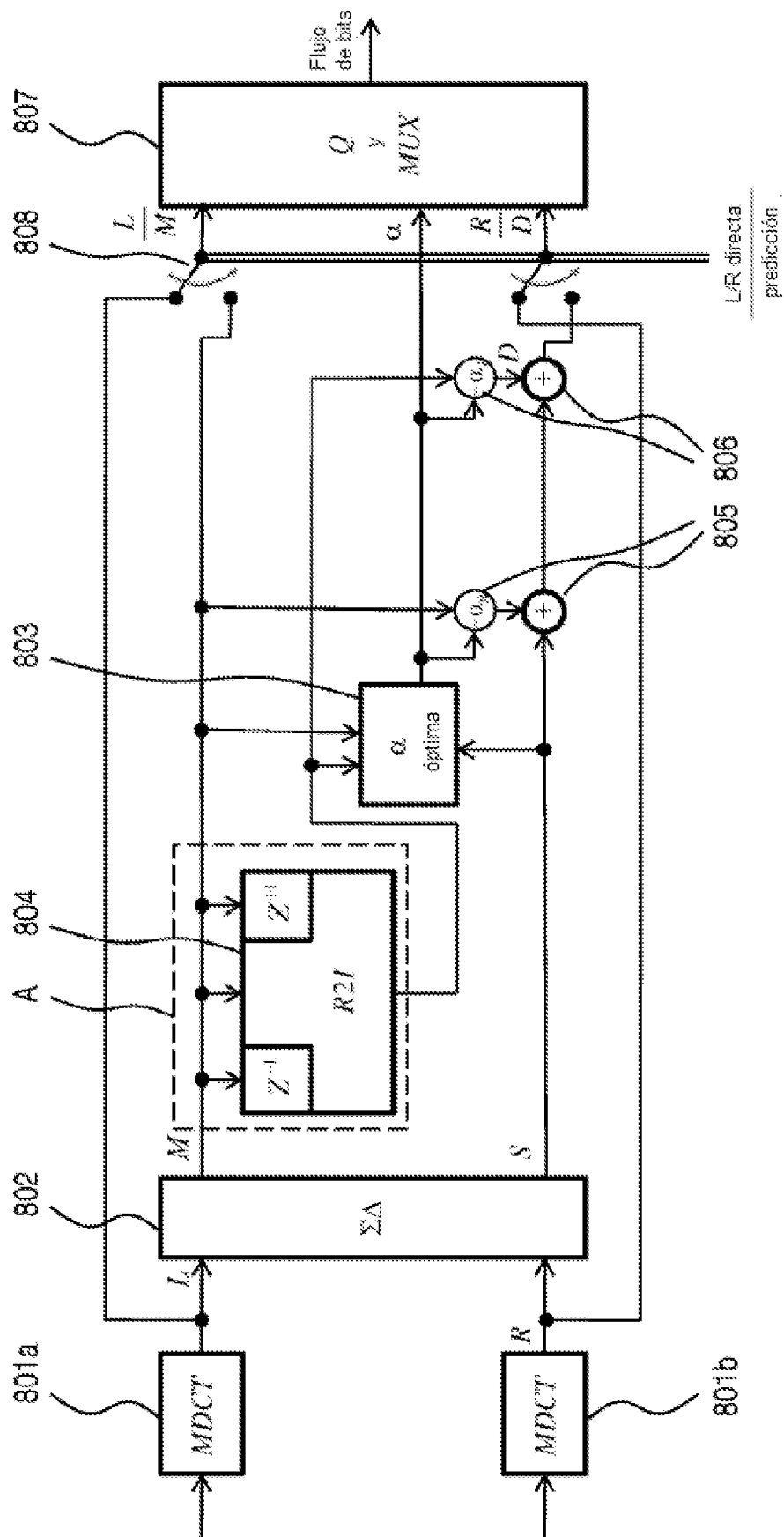


Fig. 8

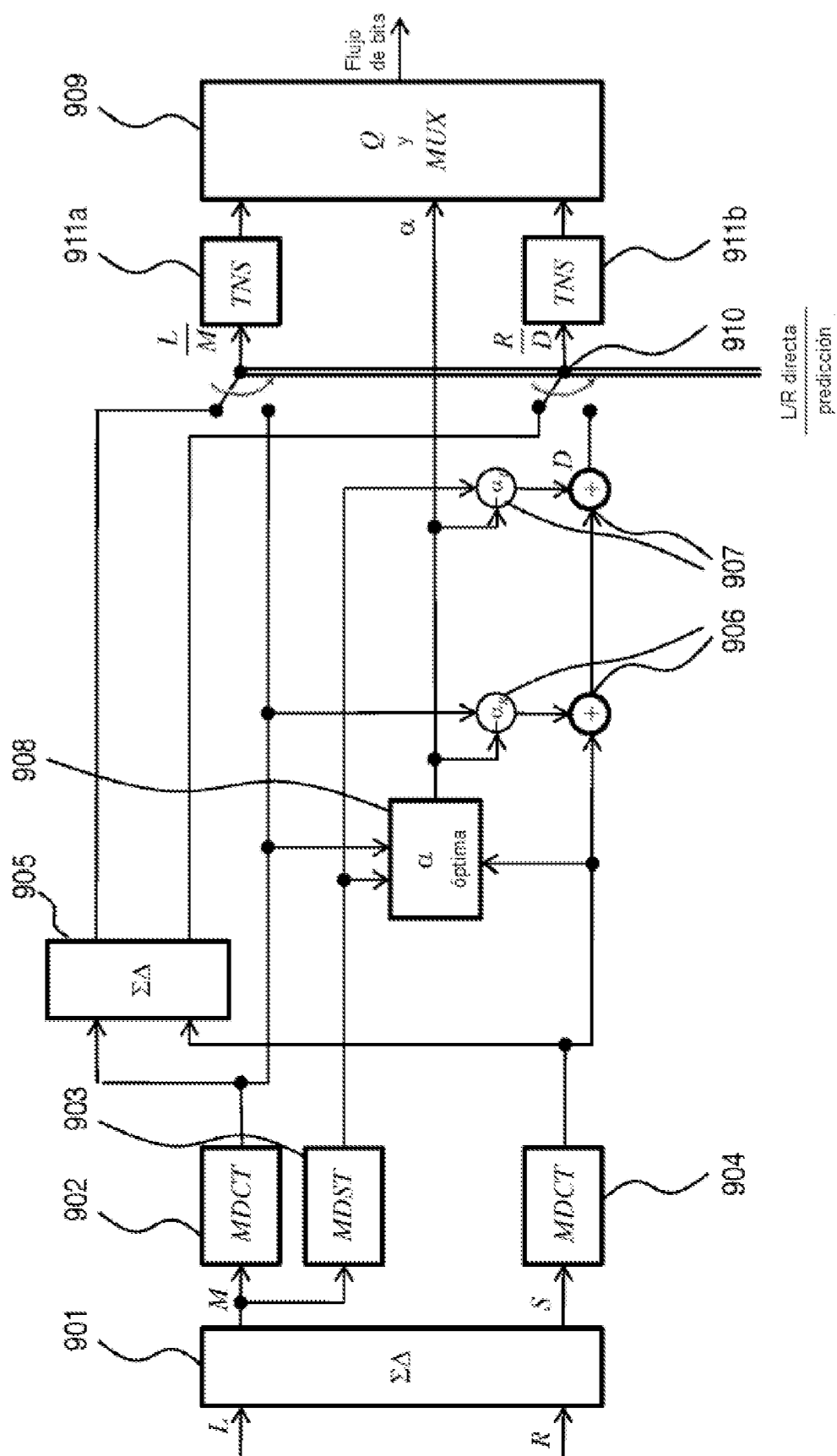


Fig. 9

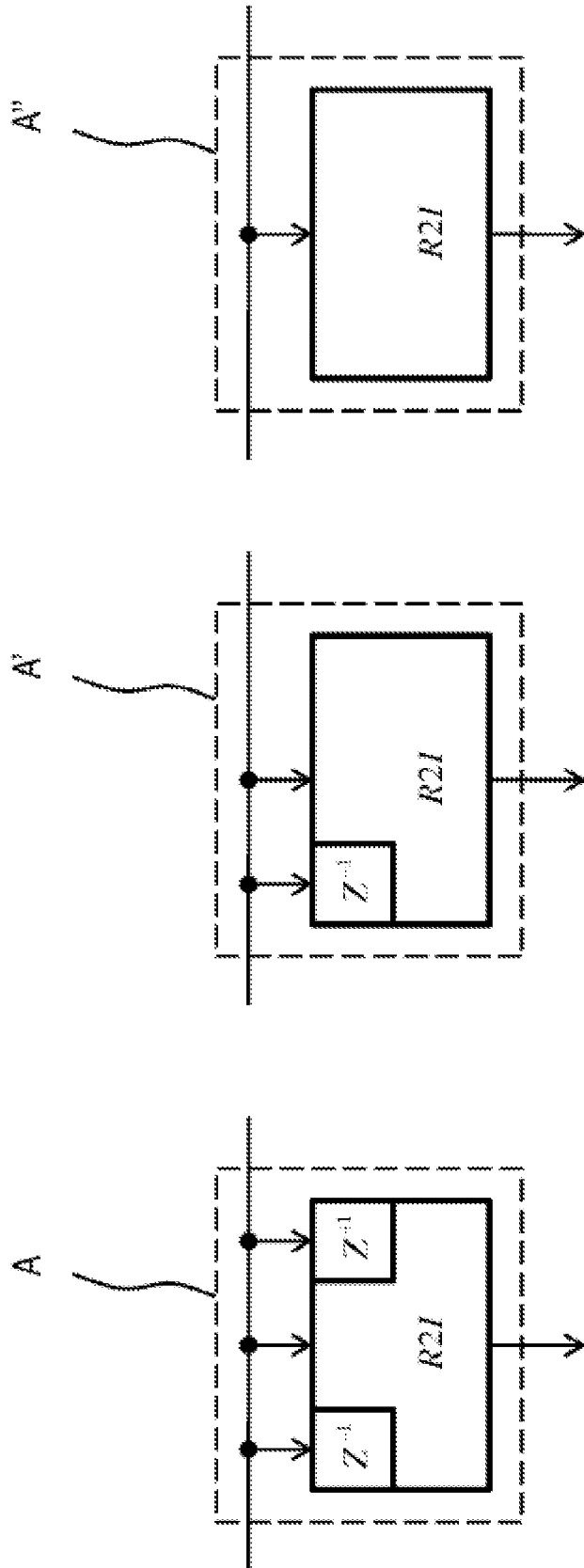


Fig. 10

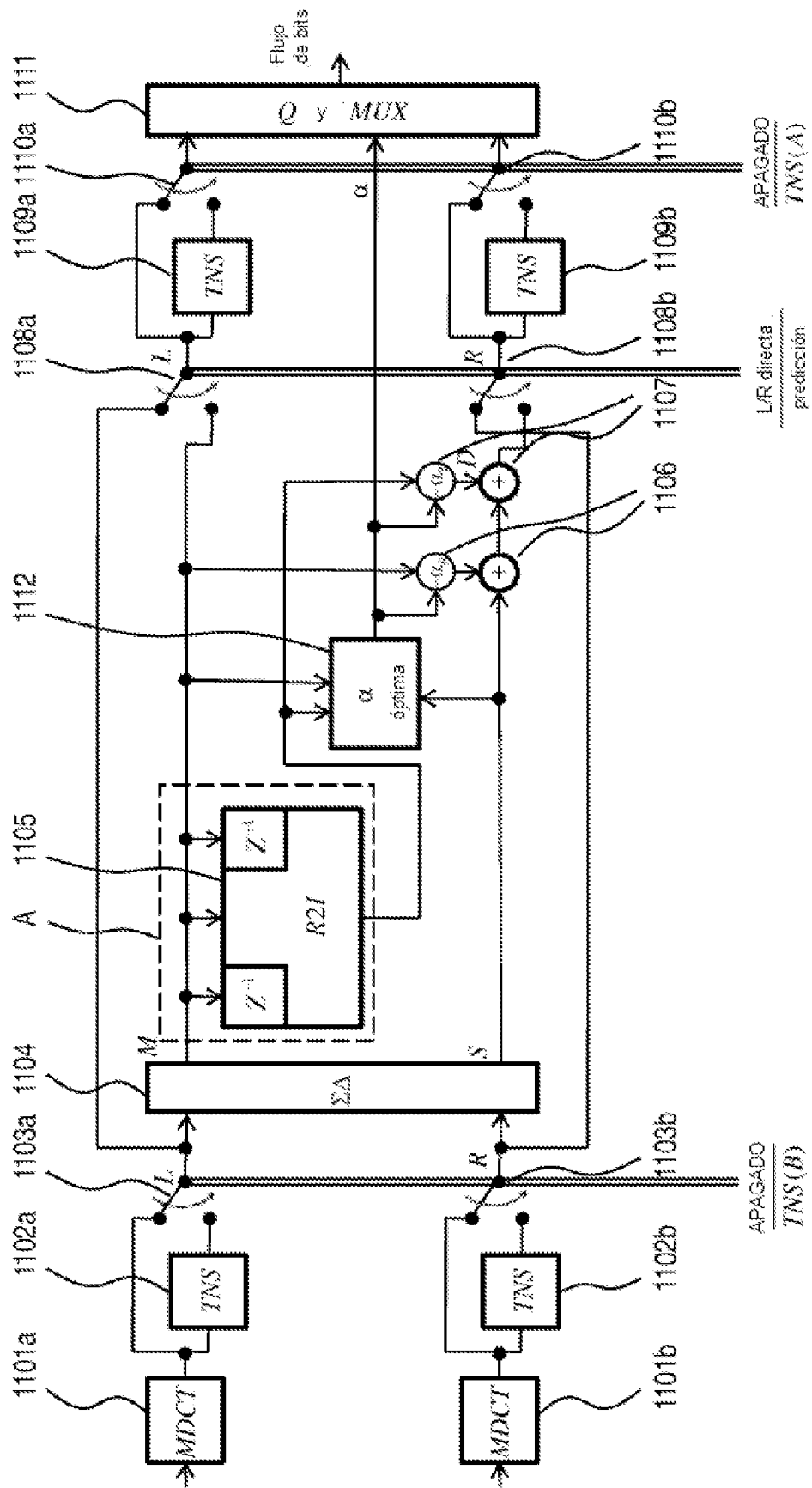


Fig. 11

MUSHRA, puntuaciones absolutas, 6 sujetos

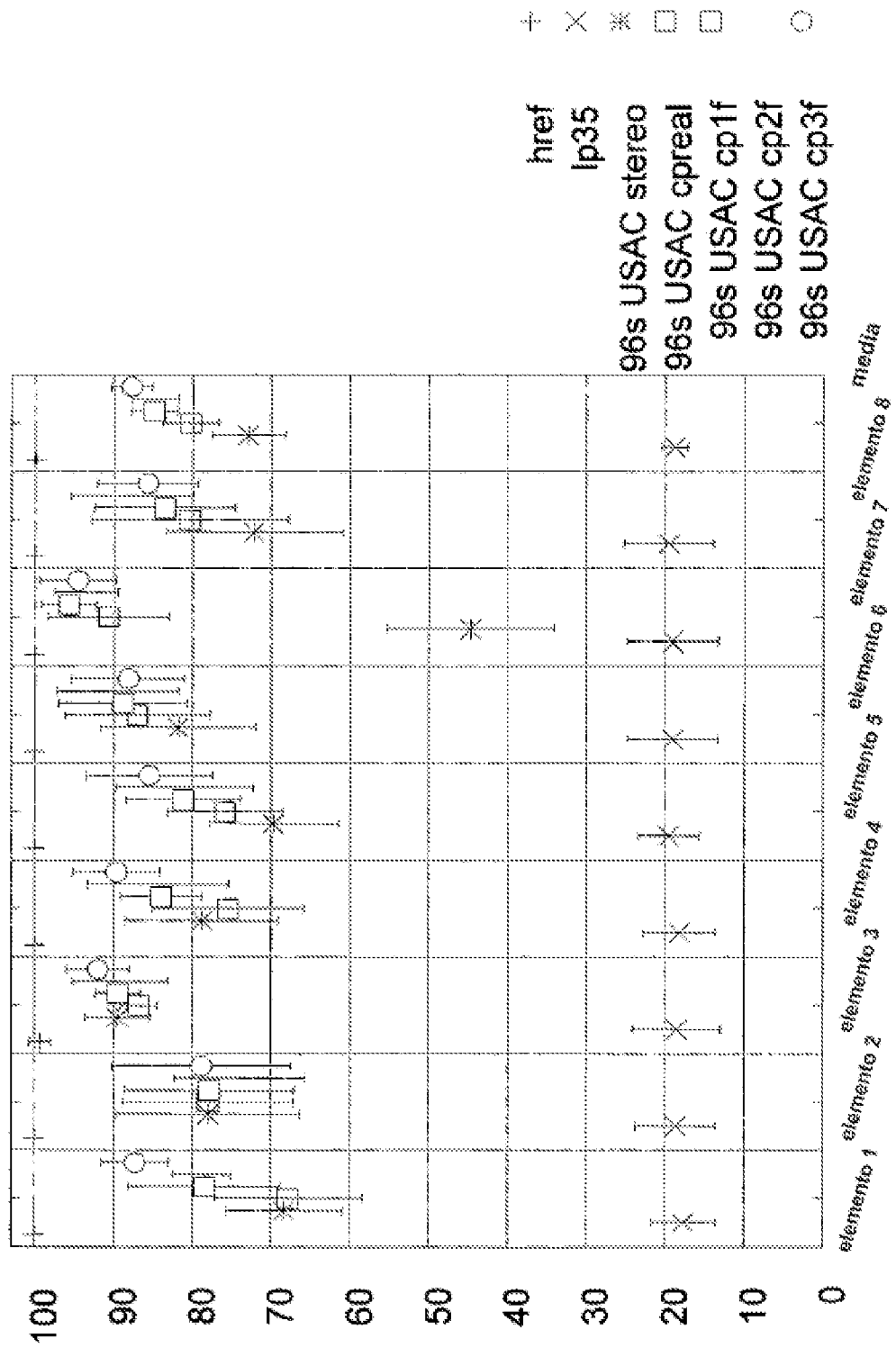


Fig. 12

MUSHRA, puntuaciones diferenciales, 6 sujetos

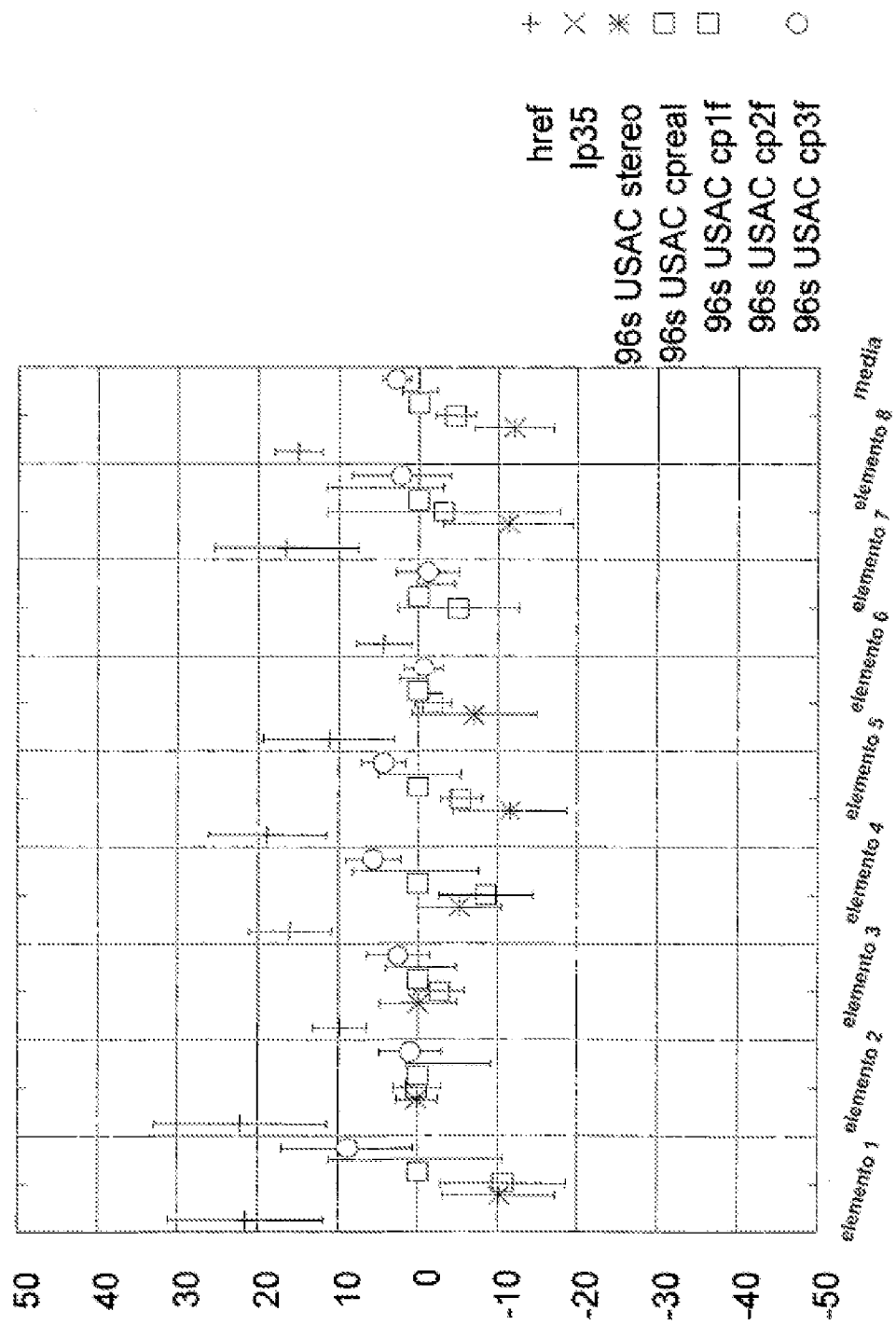


Fig. 13

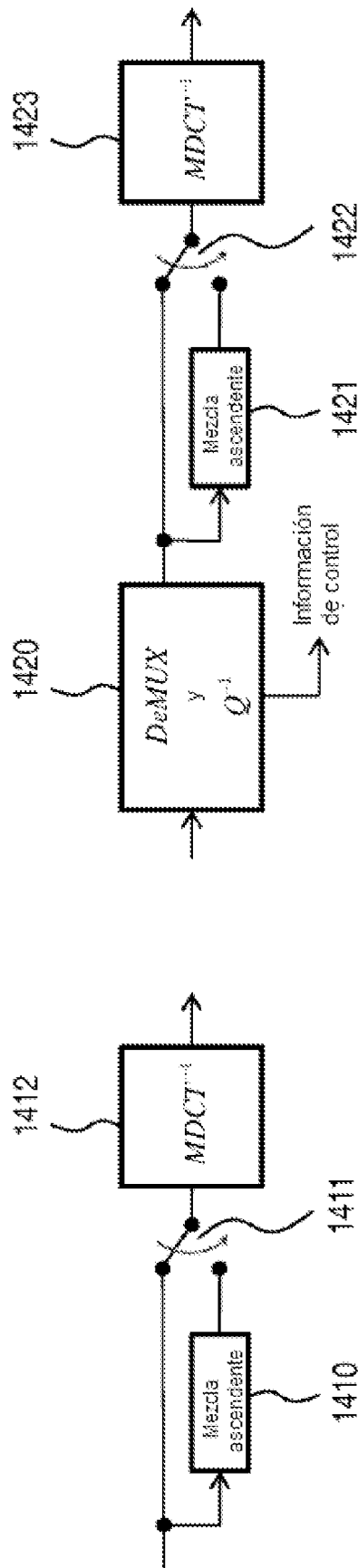


Fig. 14B

Fig. 14A

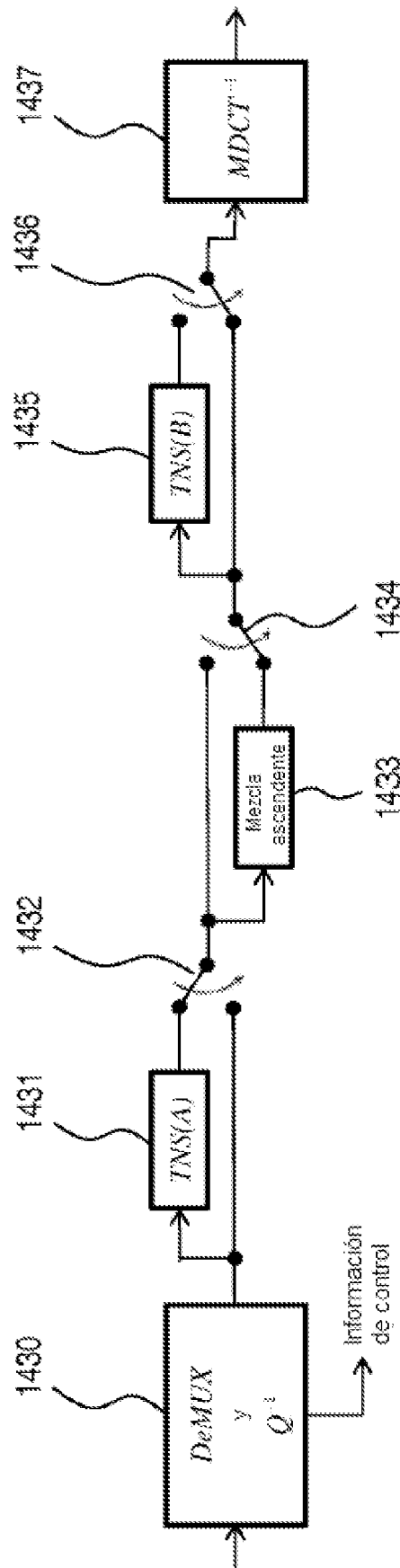


Fig. 14C

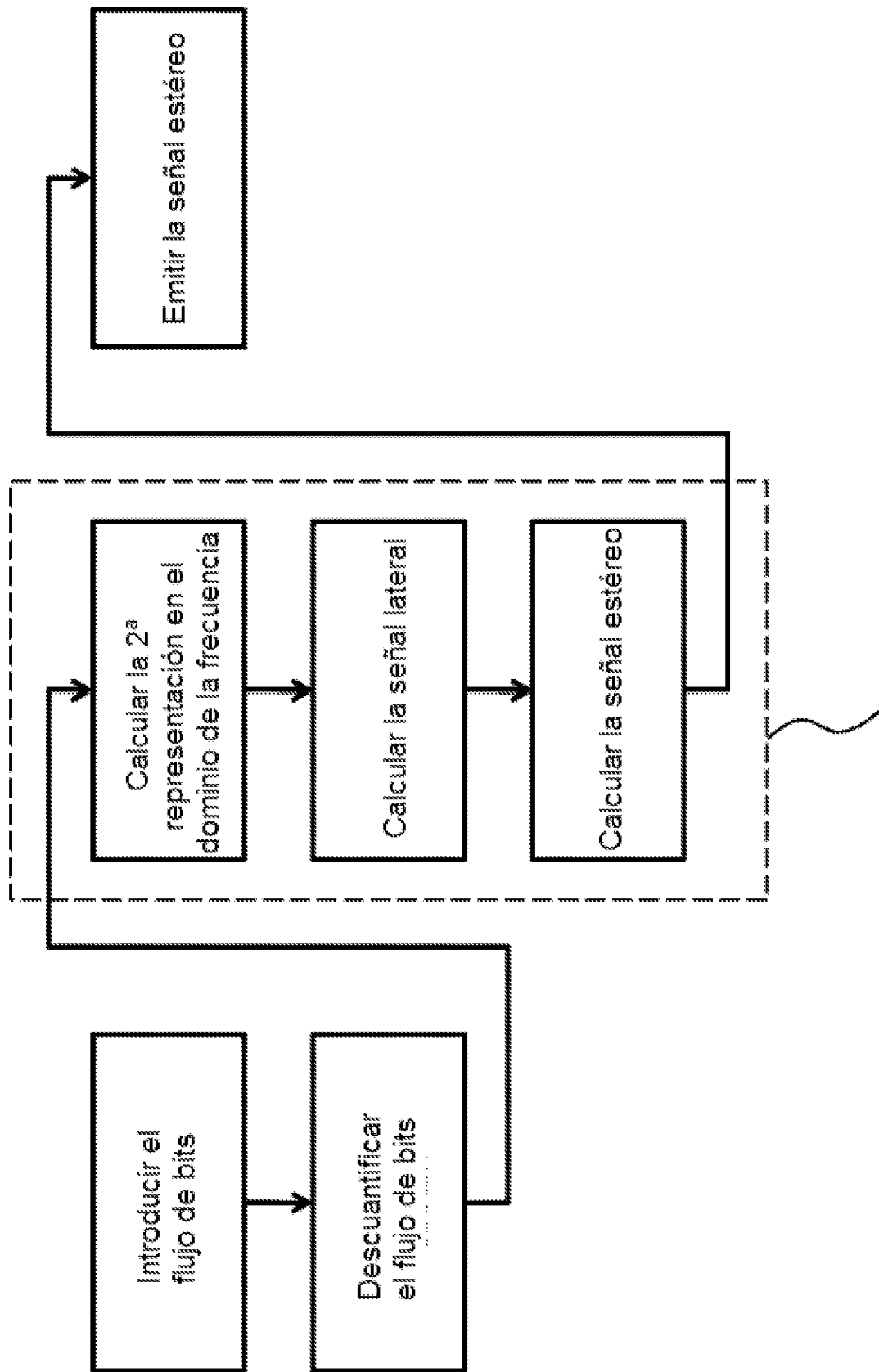


Fig. 15

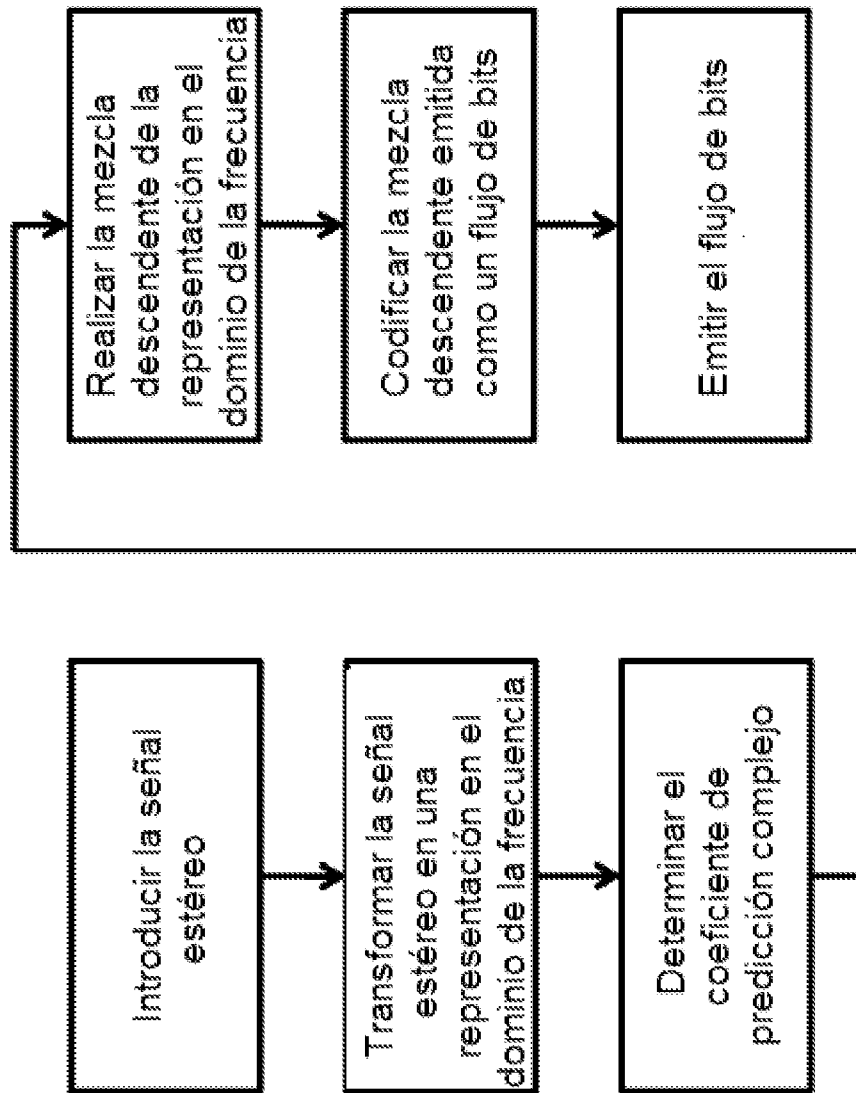


Fig. 16