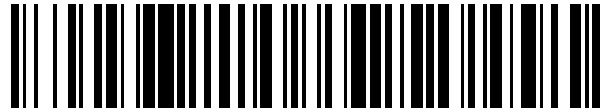


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 956 344**

51 Int. Cl.:

H04S 7/00 (2006.01)

G10L 19/02 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.08.2016 E 20187841 (0)**

97 Fecha y número de publicación de la concesión europea: **16.08.2023 EP 3748994**

54 Título: **Descodificador de audio y procedimiento de descodificación**

30 Prioridad:

25.08.2015 US 201562209742 P
08.10.2015 EP 15189008

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
19.12.2023

73 Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (50.0%)
1275 Market Street
San Francisco, CA 94103, US y
DOLBY INTERNATIONAL AB (50.0%)

72 Inventor/es:

BREEBAART, DIRK JEROEN;
COOPER, DAVID MATTHEW y
LEIF JONAS, SAMUELSSON

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 956 344 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

Descodificador de audio y procedimiento de descodificación

Sector técnico de la invención

5 La presente invención se refiere al sector técnico del procesamiento de señales y, en particular, a descodificar una señal de audio codificada.

Antecedentes de la invención

Ninguna discusión de la técnica de los antecedentes en toda la memoria descriptiva se deberá considerar como una admisión de que dicha técnica es ampliamente conocida o forma parte de un conocimiento general común en el sector técnico.

10 La creación, codificación, distribución y reproducción de contenidos de audio se llevan a cabo habitualmente en un formato basado en canal, es decir, se contempla un sistema de reproducción objetivo específico para contenido en todo el ecosistema de contenidos. Ejemplos de dichos formatos de audio de sistemas de reproducción objetivo son mono, estéreo, 5.1, 7.1 y similares.

15 Si se va reproducir contenido en un sistema de reproducción diferente al previsto, se puede aplicar un proceso de mezcla descendente o de mezcla ascendente. Por ejemplo, el contenido 5.1 se puede reproducir en un sistema de reproducción estéreo utilizando ecuaciones específicas de mezcla descendente. Otro ejemplo es la reproducción de contenido codificado en estéreo sobre el sistema de altavoces 7.1, que puede comprender un denominado proceso de mezcla ascendente, que podría o no estar guiado por información presente en la señal estéreo. Un sistema apto para mezcla ascendente es Dolby Pro Logic from Dolby Laboratories Inc (Roger Dressler, "Dolby Pro Logic Surround Decoder, Principles of Operation", www.Dolby.com).

20 Cuando se va reproducir contenido estéreo o multicanal sobre auriculares, a menudo es deseable simular una configuración de altavoz multicanal por medio de respuestas de impulsos relacionadas con la cabeza (HRIR, head-related impulse responses) o respuestas de impulsos de habitación binaural (BRIR, binaural room impulse responses), que simulan el trayecto acústico desde cada altavoz hasta los tímpanos, en un entorno anecoico o ecoico (simulado), respectivamente. En particular, las señales de audio se pueden convolucionar con HRIR o BRIR para restablecer diferencias de nivel interaural (ILD, inter-aural level differences), diferencias de tiempo interaural (ITD, inter-aural time differences) y entradas espectrales que permiten al oyente determinar la posición de cada canal individual. La simulación de un entorno acústico (reverberación) ayuda asimismo a conseguir una determinada distancia percibida.

30 Localización de la fuente de sonido y simulación de altavoz virtual

35 Cuando se va a reproducir en auriculares contenido estéreo, multicanal o basado en objetos, a menudo es deseable simular una configuración de altavoz multicanal o un conjunto de objetos acústicos virtuales discretos, por medio de una convolución con respuestas de impulsos relacionadas con la cabeza (HRIR), o respuestas de impulsos de habitación binaural (BRIR), que simulan el trayecto acústico desde cada altavoz hasta los tímpanos, en un entorno anecoico o ecoico (simulado), respectivamente.

40 En particular, las señales de audio se pueden convolucionar con HRIR o BRIR para restablecer diferencias de nivel interaural (ILD, inter-aural level differences), diferencias de tiempo interaural (ITD, inter-aural time differences) y entradas espectrales que permiten al oyente determinar la posición de cada objeto o canal individual. La simulación de un entorno acústico (reflexiones tempranas y reverberación posterior) ayuda a conseguir una determinada distancia percibida.

45 Pasando a la figura 1, se muestra 10 una visión general esquemática del flujo de procesamiento para entregar dos señales de canal o de objeto x_i 13, 11, que se leen desde un almacenamiento de contenidos 12 para su procesamiento por 4 HRIR, por ejemplo, 14 La salida de las HRIR se suman a continuación 15, 16, para cada señal de canal, con el fin de producir salidas de altavoces de auriculares para reproducir para un oyente por medio de los auriculares 18. El principio básico de las HRIR, se explica, por ejemplo, en el documento de Wightman et al (1989).

50 El enfoque de convolución HRIR/BRIR conlleva varios inconvenientes, siendo uno de estos la cantidad sustancial de procesamiento que se necesita para reproducción en auriculares. La convolución HRIR o BRIR tiene que ser aplicada para cada canal u objeto de entrada por separado, y por lo tanto habitualmente la complejidad crece linealmente con el número de canales u objetos. Dado que los auriculares se utilizan habitualmente junto con dispositivos portátiles alimentados por batería, no es deseable una gran complejidad computacional dado que reducirá sustancialmente la vida de la batería. Además, con la introducción de contenido de audio basado en objetos, que puede constar de más de 100 objetos activos simultáneamente, la complejidad de la convolución HRIR puede ser sustancialmente mayor que para un contenido tradicional basado en canales.

Técnicas de codificación paramétrica

La complejidad computacional no es el único problema para el suministro de contenido basado en canales u objetos, dentro de un ecosistema que involucra creación, distribución y reproducción de contenidos. En muchas situaciones prácticas, y especialmente para aplicaciones móviles, la velocidad de datos disponible para suministro de contenidos está seriamente limitada. Los consumidores, difusores y proveedores de contenidos han estado suministrando contenido de audio estéreo (dos canales), utilizando códecs de audio perceptual con pérdidas, con tasas de bits típicas entre 48 y 192 kbits/s. Estos códecs de audio convencionales basados en canales, tales como MPEG-1 capa 3 (Brandenberg et al., 1994), MPEG AAC (Bosi et al., 1997) y Dolby Digital (Andersen et al., 2004) tienen una tasa de bits que escala de manera aproximadamente lineal con el número de canales. Como resultado, el suministro de decenas o incluso de cientos de objetos tiene como resultado tasas de bits que son poco factibles o incluso inviables para propósitos de distribución a consumidores.

Para permitir el suministro de contenido complejo, basado en objetos, a tasas de bits que sean compatibles con las tasas de bits requeridas para suministro de contenido estéreo utilizando códecs convencionales de audio perceptual, durante la última de cada se han sometido a estudio y desarrollo los denominados procedimientos paramétricos. Estos procedimientos paramétricos permiten la reconstrucción de un gran número de canales u objetos a partir de un número pequeño de señales de base. Estas señales de base pueden ser transportadas del emisor al receptor utilizando códecs de audio convencionales, aumentados con información (paramétrica) adicional para permitir la reconstrucción de los canales u objetos originales. Ejemplos de dichas técnicas son Estéreo paramétrico (Schuijers et al., 2004), MPEG Surround (Herre et al., 2008), y codificación de objetos de audio espaciales MPEG (Herre et al., 2012).

Un aspecto importante de las técnicas, tales como estéreo paramétrico y MPEG Surround es que estos procedimientos están dirigidos a una reconstrucción paramétrica de una única presentación predeterminada (por ejemplo, altavoces estéreo en estéreo paramétrico, y altavoces 5.1 en MPEG Surround). En el caso de MPEG Surround, puede estar integrado en el descodificador un elemento de virtualización de auriculares que genera una configuración de altavoces 5.1 virtuales para auriculares, en la que los altavoces 5.1 virtuales corresponden a la configuración de altavoces 5.1 para reproducción en altavoces. Por consiguiente, estas presentaciones no son independientes porque la presentación de auriculares representa la misma distribución de altavoces (virtuales) que la presentación de altavoces. Por otra parte, la codificación de objetos de audio espaciales MPEG está dirigida a la reconstrucción de objetos que requieren una entrega posterior.

Pasando a continuación a la figura 2, se describirá, en una visión general, un sistema paramétrico 20 que soporta canales y objetos. El sistema está dividido en partes de codificador 21 y descodificador 22. El codificador 21 recibe canales y objetos 23 como entradas, y genera una mezcla descendente 24 con un número limitado de señales de base. Adicionalmente, se calcula una serie de parámetros de reconstrucción de objeto/canal 25. Un codificador de señales 26 codifica las señales de base procedentes del mezclador descendente 24, e incluye los parámetros calculados 25, así como metadatos de objetos 27 que indican cómo se deberían entregar los objetos en el flujo de bits resultante.

El descodificador 22 descodifica primero 29 las señales de base, seguido por la reconstrucción de canales y/o de objetos 30 con la ayuda de los parámetros de reconstrucción transmitidos 31. Las señales resultantes se pueden reproducir directamente (si son canales) o se pueden entregar 32 (si son objetos). Para el segundo caso, cada señal de objeto reconstruida es entregada en función de sus metadatos de objeto asociados 33. Un ejemplo de dichos metadatos es un vector de posición (por ejemplo, una coordenada x, y y z del objeto en un sistema de coordenadas tridimensional).

Matrización del descodificador

La reconstrucción de objetos y/o canales 30 se puede conseguir mediante operaciones matriciales de variación de tiempo y frecuencia. Si las señales de base descodificadas 35 se denominan $z_s[n]$, siendo s el índice de señal de base, y n el índice de la muestra, la primera etapa comprende habitualmente la transformación de las señales de base por medio de una transformada o banco de filtros.

Se puede utilizar una amplia gama de transformadas y bancos de filtros, tales como una transformada de Fourier discreta (DFT, Discrete Fourier Transform), una transformada de coseno discreta modificada (MDCT, Modified Discrete Cosine Transform) o un banco de filtros espejo en cuadratura (QMF, Quadrature Mirror Filter). La salida de semejante transformada o del banco de filtros se denomina $Z_s[k, b]$, siendo p el índice de sub-banda o espectral, y k el tiempo de la trama, intervalo o sub-banda o el índice de la muestra.

En la mayor parte de los casos, las sub-bandas o índices espectrales se mapean a un conjunto menor de bandas paramétricas p que comparten parámetros comunes de reconstrucción de objetos/canales. Esto se puede indicar como que $b \in B(p)$. En otras palabras, $B(p)$ representa un conjunto de sub-bandas consecutivas b que pertenecen a un índice de banda paramétrica p. A la inversa, $p(b)$ se refiere al índice de banda paramétrica p a la que se mapea la sub-banda b. Los objetos o canales reconstruidos en el dominio de transformada o sub-banda \hat{Y}_j se obtienen a continuación mediante señales de matrización Z_i con matrices $M[p(b)]$:

$$\begin{bmatrix} \hat{Y}_1[k, b] \\ \vdots \\ \hat{Y}_J[k, b] \end{bmatrix} = M[p(b)] \begin{bmatrix} Z_1[k, b] \\ \vdots \\ Z_S[k, b] \end{bmatrix}$$

Las señales de objetos y/o canales reconstruidos en el dominio de tiempo $y_j[n]$ se obtienen a continuación mediante una transformada inversa o banco de filtros de síntesis.

5 El proceso anterior se aplica habitualmente a un determinado rango limitado de muestras de sub-banda, intervalos o tramas k . En otras palabras, habitualmente las matrices $M[p(b)]$ son actualizadas/modificadas a lo largo del tiempo. Por simplicidad de notación, estas actualizaciones no se denotan aquí. Sin embargo, se considera que el procesamiento de un conjunto de muestras k asociadas con una matriz $M[p(b)]$ puede ser un proceso variable en el tiempo.

10 En algunos casos, en los que el número de señales reconstruidas J es significativamente mayor que el número de señales de base S , a menudo es útil utilizar salidas de decorrelador opcionales $D_m[k, b]$ que funcionan sobre una o varias señales de base que se pueden incluir en las señales de salida reconstruidas.

$$\begin{bmatrix} \hat{Y}_1[k, b] \\ \vdots \\ \hat{Y}_J[k, b] \end{bmatrix} = M[p(b)] \begin{bmatrix} Z_1[k, b] \\ \vdots \\ Z_S[k, b] \\ D_1[k, b] \\ \vdots \\ D_M[k, b] \end{bmatrix}$$

15 La figura 3 muestra esquemáticamente en mayor detalle una clase de unidad de reconstrucción de canales u objetos 30 de la figura 2. Las señales de entrada 35 se procesan primero mediante bancos de filtros de análisis 41, seguido por decorrelación opcional (D_1 , D_2) 44 y matrización 42, y un banco de filtros de síntesis 43. La manipulación de la matriz $M[p(b)]$ se controla mediante reconstrucción de parámetros 31.

Predicción de mínimo error cuadrático medio (MMSE, minimum mean square error) para reconstrucción de objetos/canales

20 Aunque existen diferentes estrategias y procedimientos para reconstruir objetos o canales a partir de un conjunto de señales de base $Z_s[k, b]$, un procedimiento particular se denomina a menudo un predictor de mínimo error cuadrático medio (MMSE), que utiliza correlaciones y matrices de covarianza para obtener coeficientes de matriz M que minimizan la norma L_2 entre una señal deseada y reconstruida. Para este método, se supone que las señales de base $z_s[n]$ se generan en el mezclador descendente 24 del codificador como una combinación lineal de señales de canal u objeto de entrada $x_i[n]$:

25
$$z_s[n] = \sum_i g_{i,s} x_i[n]$$

30 Para contenido de entrada basado en canal, las ganancias de panoramización de amplitud $g_{i,s}$ son habitualmente constantes, mientras que para contenido basado en objetos, en los que la posición prevista de un objeto es proporcionada por metadatos de objeto variables a lo largo del tiempo, las ganancias $g_{i,s}$ pueden, por consiguiente, ser variables a lo largo del tiempo. Esta ecuación se puede formular asimismo en el dominio de sub-banda o transformada, en cuyo caso se utiliza un conjunto de ganancias $g_{i,s}[k]$ para cada banda/tramo de frecuencia k y, de este modo, las ganancias $g_{i,s}[k]$ se pueden hacer variables en frecuencia:

$$Z_s[k, b] = \sum_i g_{i,s}[k] X_i[k, b]$$

La matriz del descodificador 42, ignorando los decorreladores por ahora, produce:

$$\begin{bmatrix} \hat{Y}_1[k, b] \\ \vdots \\ \hat{Y}_J[k, b] \end{bmatrix}^T = \begin{bmatrix} Z_1[k, b] \\ \vdots \\ Z_S[k, b] \end{bmatrix}^T M[p(b)]$$

35 o, en formulación matricial, omitiendo el índice de sub-banda b y el índice de banda paramétrica p por claridad:

$$Y = ZM$$

$$Z = XG$$

El criterio para calcular los coeficientes de matriz M mediante el codificador es minimizar el error cuadrático medio E que representa el error cuadrático entre las salidas del descodificador \hat{Y}_j y los objetos/canales de entrada originales X_j :

$$E = \sum_{j,k,b} (\hat{Y}_j[k, b] - X_j[k, b])^2$$

Los coeficientes de matriz que minimizan E se proporcionan a continuación en notación matricial mediante:

$$M = (Z^*Z + \epsilon I)^{-1}Z^*X$$

siendo épsilon una constante de regularización, y (*) el operador de traspuesta conjugada compleja. Esta operación se puede realizar para cada banda paramétrica p independientemente, produciendo una matriz M[p(b)].

Predicción de error mínimo cuadrático medio (MMSE) para transformación de representación

Además de la reconstrucción de objetos y/o canales, se pueden utilizar técnicas paramétricas para transformar una representación en otra representación. Un ejemplo de dicha transformación de representación es convertir una mezcla estéreo prevista para reproducción en altavoces en una representación binaural para auriculares, o viceversa.

La figura 4 muestra el flujo de control para un procedimiento 50 para semejante transformación de representación. Primero se procesa audio de objeto o de canal en un codificador 52 mediante un banco de análisis de filtros espejo en cuadratura híbridos 54. Se calcula una matriz de entrega de los altavoces C y se aplica 55 a las señales de objeto X_i almacenadas en el medio de almacenamiento 51, en base a los metadatos de objeto utilizando técnicas de panoramización de amplitud, con el resultado de una presentación de altavoces estéreo Z_s. Esta presentación de altavoces se puede codificar con un codificador de audio 57.

Adicionalmente, se genera una matriz de entrega binaural H y se aplica 58 utilizando una base de datos HRTF 59. Esta matriz H se utiliza para calcular señales binaurales Y_j que permiten la reconstrucción de una mezcla binaural utilizando la mezcla de altavoces estéreo como entrada. Los coeficientes de matriz M son codificados por el codificador de audio 57.

La información transmitida se transmite del codificador 52 al descodificador 53, donde se desempaqueta 61 para incluir componentes M y Z_s. Si se utilizan altavoces como sistema de reproducción, la presentación de altavoces se reproduce utilizando información de canal Z_s y, de ese modo, se desechan los coeficientes de matriz M. Para reproducción en auriculares, por otra parte, la presentación de altavoces se transforma 62 primero en una presentación binaural aplicando la matriz variable en tiempo y frecuencia M antes de síntesis de QMF híbridos y reproducción 60.

Si la salida binaural deseada procedente del elemento de matrización 62 se describe en notación matricial como:

$$Y = XH$$

entonces se pueden obtener los coeficientes de matriz M en el codificador 52 como:

$$M = (G^*X^*XG + \epsilon I)^{-1}G^*X^*XH$$

En esta aplicación, los coeficientes de la matriz del codificador H aplicados en 58 tienen habitualmente valores complejos, por ejemplo, teniendo un elemento de retardo o modificación de fase, para permitir el restablecimiento de diferencias temporales interaurales, que son perceptualmente muy relevantes para la ubicación de la fuente del sonido con auriculares. En otras palabras, la matriz de entrega binaural H tiene valores complejos, y por lo tanto la matriz de transformación M tiene valores complejos. Para un restablecimiento transparente perceptualmente de entradas de localización de la fuente de sonido, se ha demostrado que es deseable una resolución en frecuencia que reproduzca la resolución en frecuencia del sistema auditivo humano (Breebaart 2010).

En las secciones anteriores, se utiliza un criterio de error mínimo cuadrático medio para determinar los coeficientes de matriz M. Sin pérdida de generalidad, se pueden utilizar análogamente otros criterios o procedimientos bien conocidos, para calcular los coeficientes de matriz, con el fin de sustituir o complementar el principio del error mínimo cuadrático medio. Por ejemplo, los coeficientes de matriz M se pueden calcular utilizando términos de error de orden superior, o mediante minimización de una norma L1 (por ejemplo, criterio de desviación mínima absoluta). Además, se pueden utilizar diversos procedimientos incluyendo técnicas de optimización o factorización no negativa, estimadores no paramétricos, estimadores de máxima probabilidad y similares. Adicionalmente, los coeficientes de matriz se pueden calcular utilizando procesos iterativos o de gradiente descendente, procedimientos de interpolación, procedimientos heurísticos, programación dinámica, aprendizaje automático, optimización borrosa, recocido simulado o soluciones de forma cerrada, y técnicas de análisis por síntesis. Por último, pero no menos importante, la estimación de coeficientes de matriz se puede restringir de varios modos, por ejemplo limitando el intervalo de valores, los términos de regularización, la superposición de requisitos de conservación de energía y similares.

Requisitos de transformada y de banco de filtros

Dependiendo de la aplicación, y de si se tienen que reconstruir objetos o canales, determinados requisitos se pueden superponer sobre la resolución en frecuencia de banco de filtros o de transformada para la unidad de banco de filtros 41 de la figura 3. En la mayor parte de las aplicaciones prácticas, la resolución en frecuencia se hace corresponder con la resolución asumida del sistema auditivo humano, para proporcionar la mejor calidad de audio percibida para una determinada tasa de bits (determinada por el número de parámetros) y una determinada complejidad. Se sabe que el sistema auditivo humano se puede considerar como un banco de filtros con una resolución en frecuencia no lineal. Estos filtros se denominan bandas críticas (Zwicker, 1961) y son de naturaleza aproximadamente logarítmica. A frecuencias bajas, las bandas críticas tienen menos de 100 Hz de anchura, mientras que, a frecuencias altas, se puede encontrar que las bandas tienen una anchura mayor de 1 kHz.

Este comportamiento no lineal puede plantear desafíos en materia de diseño de bancos de filtros. Las transformadas y los bancos de filtros se pueden implementar de manera muy eficiente utilizando simetrías en su estructura de procesamiento, siempre que la resolución en frecuencia sea constante a través de la frecuencia.

Esto implica que la longitud de la transformada, o el número de sub-bandas, se determinará mediante el ancho de banda crítico a bajas frecuencias, y se puede utilizar mapeo de tramos DFT sobre denominadas bandas paramétricas, para reproducir una resolución en frecuencia no lineal. Dicho proceso de mapeo se explica, por ejemplo, en Breebaart et al., (2005) y Breebaart et al., (2010). Un inconveniente de este enfoque es que se requiere una transformada muy larga para satisfacer la limitación de ancho de banda crítico de baja frecuencia, mientras que la transformada es relativamente larga (o ineficiente) a frecuencias altas. Una solución alternativa para mejorar la resolución en frecuencia a frecuencias bajas es utilizar una estructura de banco de filtros híbridos. En dicha estructura, se utiliza una cascada de dos bancos de filtros, en la que el segundo banco de filtros mejora la resolución del primero, pero solamente en unas pocas de las sub-bandas más bajas (Schuijers et al., 2004).

La figura 5 muestra una clase de estructura de banco de filtros híbridos 41, similar a la establecida en los documentos de Schuijers et al. La señal de entrada $z[n]$ es procesada primero por un banco de análisis de filtros espejo en cuadratura (CQMF, complex-valued Quadrature Mirror Filter) de valores complejos 71. A continuación, las señales se someten a muestreo descendente mediante un factor Q , por ejemplo, 72, con el resultado de señales de sub-banda $Z[k, b]$ siendo k el índice muestra de sub-banda y b el índice de frecuencia de sub-banda. Además, por lo menos una de las señales de sub-banda resultantes es procesada por un segundo banco de filtros (Nyquist) 74, mientras que las señales de sub-banda restantes son retardadas 75 para compensar el retardo introducido por el banco de filtros de Nyquist. En este ejemplo particular, la cascada de bancos de filtros tiene como resultado 8 sub-bandas ($b = 1, \dots, 8$) que son mapeadas sobre 6 bandas paramétricas $p = (1, 6)$ con una resolución en frecuencia no lineal. Fusionándose conjuntamente las bandas 76 para formar una única banda paramétrica ($p=6$).

El beneficio de este enfoque es una complejidad menor comparada con la utilización de un único banco de filtros con muchas más sub-bandas (más estrechas). Sin embargo, el inconveniente es que el retardo del sistema global aumenta significativamente y, por consiguiente, la utilización de memoria es asimismo significativamente superior, lo que provoca un aumento en el consumo de energía.

Limitaciones de la técnica anterior

Volviendo a la figura 4, se sugiere que la técnica anterior utiliza el concepto de matrización 62, posiblemente aumentado con la utilización de decorreladores, para reconstruir los canales, objetos o señales de presentación \hat{Y}_j partir de un conjunto de señales de base Z_s . Esto conduce a la siguiente formulación matricial para describir la técnica anterior de manera genérica:

$$\begin{bmatrix} \hat{Y}_1[k, b] \\ \vdots \\ \hat{Y}_j[k, b] \end{bmatrix}^T = \begin{bmatrix} Z_1[k, b] \\ \vdots \\ Z_s[k, b] \\ D_1[k, b] \\ \vdots \\ D_M[k, b] \end{bmatrix}^T M[p(b)]$$

Los coeficientes de matriz M se transmiten directamente del codificador al decodificador, o bien se obtienen a partir de parámetros de localización de las fuentes de sonido, por ejemplo tal como se describe en el documento de Breebaart et al 2005 para codificación estéreo paramétrica o de Herre et al., (2008) para decodificación multicanal. Además, este enfoque se puede utilizar asimismo para restablecer diferencias de fase entre canales utilizando coeficientes de matriz de valores complejos (ver los documentos de Breebaart et al., 2010, y de Breebaart., 2005, por ejemplo).

Tal como se muestra en la figura 6, en la práctica, utilizar coeficientes de matriz de valores complejos implica que un retardo deseado 80 se representa por una aproximación de fase constante por piezas 81. Asumiendo que la respuesta de fase deseada es un retardo puro 80 con una fase decreciente linealmente con la frecuencia (línea de trazos), la operación de matrización de valores complejos de la técnica anterior tiene como resultado una aproximación constante por piezas 81 (línea continua). La aproximación se puede mejorar aumentando la resolución de la matriz M . Sin embargo, esto tiene dos desventajas importantes. Requiere un aumento de la resolución del

banco de filtros, que provoca una mayor utilización de la memoria, una mayor complejidad computacional, una latencia más larga y, por lo tanto, un mayor consumo de energía. Requiere asimismo que se envíen más parámetros, provocando una mayor tasa de bits.

5 Todas estas desventajas son especialmente problemáticas para dispositivos móviles y alimentados por batería. Sería ventajoso que estuviera disponible una solución más óptima.

Compendio de la invención

Un objetivo de la invención, en su forma preferida, es dar a conocer una forma mejorada de descodificación de señales de audio para su reproducción en diferentes presentaciones.

10 De acuerdo con un primer aspecto de la presente invención, se da a conocer un procedimiento para descodificar una señal de audio codificada de acuerdo con la reivindicación 1.

15 Los coeficientes de matriz pueden representar un filtro de respuesta finita al impulso (FIR). El conjunto de señales de base son divididas preferiblemente en una serie de segmentos temporales, y se puede proporcionar un conjunto de parámetros de transformación para cada segmento temporal. Los coeficientes de filtro pueden incluir al menos un coeficiente que puede ser de valor complejo. La primera o segunda representación puede estar destinada a reproducción en auriculares.

20 En algunas realizaciones, los parámetros de transformación asociados con frecuencias mayores no modifican la fase de la señal, mientras que, para frecuencias menores, los parámetros de transformación modifican la fase de la señal. El conjunto de coeficientes de filtro puede funcionar preferentemente para procesar una matriz de convolución de múltiples tomas. El conjunto de coeficientes de filtro puede ser utilizado preferentemente para procesar una banda de baja frecuencia.

25 El conjunto de señales de base y el conjunto de parámetros de transformación se combinan para formar el flujo de datos. Los parámetros de transformación pueden incluir coeficientes de matriz de audio de alta frecuencia para manipulación de matrices de una parte de alta frecuencia del conjunto de señales de base. En algunas realizaciones, para una parte de frecuencia intermedia de la parte de alta frecuencia del conjunto de señales de base, la manipulación de matrices puede incluir preferentemente parámetros de transformación de valores complejos.

De acuerdo con otro aspecto de la presente invención, se da a conocer un descodificador para descodificar un flujo de bits de entrada de acuerdo con la reivindicación 5.

30 Las unidades de multiplicación de matrices pueden modificar la fase de los componentes de baja frecuencia de las señales de base de audio. En algunas realizaciones, los parámetros de transformación de matriz de convolución de múltiples tomas son preferentemente de valores complejos. Los parámetros de transformación de audio de alta frecuencia tienen asimismo, preferentemente, valores complejos. El conjunto de parámetros de transformación puede comprender además parámetros de transformación de audio de frecuencia superior de valores reales. En algunas realizaciones, el descodificador puede incluir además filtros para separar las señales de base de audio en los componentes de baja frecuencia y los componentes de alta frecuencia.

35 En algunas realizaciones, la señal codificada puede comprender múltiples segmentos temporales, pudiendo incluir además preferentemente el procedimiento las etapas de: interpolar parámetros de transformación de múltiples segmentos temporales de la señal codificada, para producir parámetros de transformación interpolados, incluyendo parámetros de transformación de audio de baja frecuencia interpolados; y convolucionar múltiples segmentos temporales de los componentes de baja frecuencia de las señales de base de audio con los parámetros de transformación de audio de baja frecuencia interpolados, para producir múltiples segmentos temporales de los componentes de baja frecuencia convolucionados.

40 El conjunto de parámetros de transformación de la señal de audio codificada puede ser preferentemente variable a lo largo del tiempo, y el procedimiento puede incluir además preferentemente las etapas de: convolucionar los componentes de baja frecuencia con los parámetros de transformación de baja frecuencia para múltiples segmentos temporales para producir múltiples conjuntos de componentes de baja frecuencia convolucionados intermedios; interpolar los múltiples conjuntos de componentes de baja frecuencia convolucionados intermedios para producir los componentes de baja frecuencia convolucionados.

45 La interpolación puede utilizar un procedimiento de solapar y añadir de los múltiples conjuntos de componentes de baja frecuencia convolucionados intermedios.

Breve descripción de los dibujos

A continuación se describirán realizaciones de la invención, solamente a modo de ejemplo, haciendo referencia a los dibujos adjuntos, en los cuales:

la figura 1 muestra una visión general esquemática del proceso de convolución HRIR para dos objetos fuente,

siendo cada canal u objeto procesado por un par de HRIR/BRIR;

la figura 2 muestra esquemáticamente un sistema genérico de codificación paramétrica que soporta canales y objetos;

5 la figura 3 muestra esquemáticamente una clase de unidad de reconstrucción de canales u objetos 30 de la figura 2, en mayor detalle;

la figura 4 muestra el flujo de datos de un procedimiento para transformar una presentación de altavoces estéreo en una presentación de auriculares binaural;

la figura 5 muestra esquemáticamente la estructura de banco de filtros de análisis híbridos, según la técnica anterior;

10 la figura 6 muestra una comparación de la respuesta de fase deseada (línea de trazos) y real (línea continua) obtenida con la técnica anterior;

la figura 7 muestra esquemáticamente un sistema a modo de ejemplo de banco de filtros y mapeo de parámetros del codificador, de acuerdo con una realización de la invención;

la figura 8 muestra esquemáticamente el banco de filtros y mapeo de parámetros del descodificador, de acuerdo con una realización; y

15 la figura 9 muestra un codificador para transformación de presentaciones estéreo a binaurales.

La figura 10 muestra esquemáticamente un descodificador para transformación de presentaciones estéreo a binaurales.

Referencias

20 Wightman, F. L., y Kistler, D. J. (1989). "Headphone simulation of free-field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. 85, 858-867.

Schuijers, Erik, et al. (2004). "Low complexity parametric stereo coding." Audio Engineering Society Convention 116. Audio Engineering Society.

25 Herre, J., Kjörling, K., Breebaart, J., Faller, C., Disch, S., Purnhagen, H., ... & Chong, K. S. (2008). "MPEG surround-the ISO/MPEG standard for efficient and compatible multichannel audio coding". Journal of the Audio Engineering Society, 56(11), 932-955.

Herre, J., Purnhagen, H., Koppens, J., Hellmuth, O., Engdegård, J., Hilpert, J., & Oh, H. O. (2012). "MPEG Spatial Audio Object Coding-the ISO/MPEG standard for efficient coding of interactive audio scenes". Journal of the Audio Engineering Society, 60(9), 655-673.

30 Brandenburg, K., & Stoll, G. (1994). "ISO/MPEG-1 audio: A generic standard for coding of high-quality digital audio". Journal of the Audio Engineering Society, 42(10), 780-792.

Bosi, M., Brandenburg, K., Quackenbush, S., Fielder, L., Akagiri, K., Fuchs, H., & Dietz, M. (1997). "ISO/IEC MPEG-2 advanced audio coding". Journal of the Audio engineering society, 45(10), 789-814.

35 Andersen, R. L., Crockett, B. G., Davidson, G. A., Davis, M. F., Fielder, L. D., Turner, S. C., ... & Williams, P. A. (2004, octubre). "Introduction to Dolby digital plus, an enhancement to the Dolby digital coding system". En Audio Engineering Society Convention 117. Audio Engineering Society.

Zwicker, E. (1961). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)". The Journal of the Acoustical Society of America, (33 (2)), 248.

Breebaart, J., van de Par, S., Kohlrausch, A., & Schuijers, E. (2005). "Parametric coding of stereo audio". EURASIP Journal on Applied Signal Processing, 2005, 1305-1322.

40 Breebaart, J., Nater, F., & Kohlrausch, A. (2010). "Spectral and spatial parameter resolution requirements for parametric, filter-bank-based HRTF processing". Journal of the Audio Engineering Society, 58(3), 126-140.

Breebaart, J., van de Par, S., Kohlrausch, A., & Schuijers, E. (2005). "Parametric coding of stereo audio". EURASIP Journal on Applied Signal Processing, 2005, 1305-1322.

Descripción detallada

45 Esta realización da a conocer un procedimiento para reconstruir objetos, canales o 'presentaciones' a partir de un conjunto de señales de base que se pueden aplicar en bancos de filtros con una baja resolución en frecuencia. Un ejemplo es la transformación de una presentación estéreo en una presentación binaural destinada a reproducción en

auriculares, que se puede aplicar sin un banco de filtros (híbridos) de Nyquist. La resolución en frecuencia del descodificador reducida se compensa mediante una matriz de convolución, multi-toma. Esta matriz de convolución requiere solamente unas pocas tomas (por ejemplo, dos) y, en casos prácticos, se requiere solamente a bajas frecuencias. Este procedimiento (1) reduce la complejidad computacional de un descodificador, (2) reduce la
 5 utilización de memoria de un descodificador y (3) reduce la tasa de bits paramétrica.

En la realización preferida se da a conocer un sistema y un procedimiento para superar los requisitos no deseables de complejidad computacional y memoria del lado del descodificador. Esto se implementa disponiendo una alta resolución de frecuencia en un codificador, utilizando una resolución en frecuencia limitada (menor) en el descodificador (por ejemplo, utilizar una resolución en frecuencia que es significativamente peor que la utilizada en el codificador correspondiente) y utilizando una matriz (de convolución) multi-toma para compensar la resolución en
 10 frecuencia, reducida, del descodificador.

Típicamente, dado que se requiere una resolución de matriz de alta frecuencia solamente a bajas frecuencias, la matriz (de convolución) multi-toma se puede usar a bajas frecuencias, mientras que puede utilizarse una matriz (sin estado) convencional para las frecuencias (mayores) restantes. En otras palabras, a bajas frecuencias, la matriz
 15 representa un conjunto de filtros FIR que funcionan en cada combinación de entrada y salida, mientras que a frecuencias altas, se utiliza una matriz sin estado.

Banco de filtros y mapeo de parámetros del codificador (no es parte de la invención reivindicada)

La figura 7 muestra 90 un sistema de banco de filtros y mapeo de parámetros del codificador a modo de ejemplo, de acuerdo con una realización. En esta realización de ejemplo 90, 8 sub-bandas ($b = 1, 8$), por ejemplo 91, son generadas inicialmente por medio de un banco de filtros híbridos (en cascada) 92 y un banco de filtros de Nyquist 93. A continuación, las primeras cuatro sub-bandas son mapeadas 94 sobre una misma banda paramétrica ($p = 1$) para calcular una matriz de convolución $M[k, p = 1]$, por ejemplo, la matriz tiene ahora un índice adicional k . Las sub-bandas restantes ($b = 5, \dots, 8$) se mapean sobre bandas paramétricas ($p = 2, 3$) utilizando matrices sin estado $M[p(b)]$ 95, 96.
 20

Banco de filtros y mapeo de parámetros del descodificador

La figura 8 muestra el correspondiente sistema de banco de filtros y mapeo de parámetros 100 del descodificador, a modo de ejemplo. Por contraste con el codificador, no está presente ningún banco de filtros de Nyquist, ni hay ningún retardo para compensar el retardo del banco de filtros de Nyquist. El banco de filtros de análisis del descodificador 101 genera solamente 5 sub-bandas ($b = 1, \dots, 5$), por ejemplo 102, que son sometidas a muestreo descendente por un factor Q . La primera sub-banda es procesada por una matriz de convolución $M[k, p = 1]$ 103, mientras que las bandas restantes son procesadas por matrices sin estado 104, 105, de acuerdo con la técnica anterior.
 25
 30

Aunque el ejemplo anterior aplica un banco de filtros de Nyquist en el codificador 90 y una correspondiente matriz de convolución solamente para la primera sub-banda CQMF en el descodificador 100, el mismo proceso se puede aplicar a multitud de sub-bandas, sin limitarse necesariamente a solamente la sub-banda o sub-bandas más bajas.

Realización del codificador (es consistente con, pero no necesariamente muestra explícitamente todas las características de, las reivindicaciones independientes)
 35

Una realización que es especialmente útil está en la transformación de la presentación de altavoz a una presentación binaural. La figura 9 muestra un codificador 110 que utiliza el procedimiento propuesto para la transformación de presentación. Un conjunto de canales u objetos de entrada $x_i[n]$ se transforma primero utilizando un banco de filtros 111. El banco de filtros 111 es un banco de filtros espejo en cuadratura complejos híbridos (HCQMF, hybrid complex quadrature Mirror Filter), pero se pueden utilizar igualmente otras estructuras de banco de filtros. Las representaciones de sub-banda resultantes $X_i[k, b]$ se procesan dos veces 112, 113.
 40

En primer lugar 113, para generar un conjunto de señales de base $Z_s[k, b]$ 113 destinado a la salida del codificador. Esta salida se puede generar, por ejemplo, utilizando técnicas de panoramización de amplitud, de tal forma que las señales resultantes están destinadas a reproducción en altavoces.
 45

En segundo lugar 112, para generar un conjunto de señales transformadas deseadas $Y_j[k, b]$ 112. Esta salida se puede generar, por ejemplo, utilizando procesamiento HRIR, de tal modo que las señales resultantes están destinadas a reproducción en auriculares. Dicho procesamiento HRIR se puede utilizar en el dominio de bancos de filtros, pero puede funcionar igualmente en el dominio de tiempo por medio de convolución HRIR. Las HRIR se obtienen de una base de datos 114.
 50

La matriz de convolución $Z_s[k, b]$ se obtiene a continuación alimentando las señales de base $Z_s[k, b]$ a través de una línea de retardo derivada 116. Cada una de las tomas de estas líneas de retardo sirve como entrada adicional a una etapa de predictor MMSE 115. La etapa de predictor MMSE calcula la matriz de convolución $M[k, p]$ que minimiza el error entre las señales transformadas deseadas $Y_j[k, b]$ y la salida del descodificador 100 de la figura 8, aplicando matrices de convolución. Se sigue entonces que los coeficientes de matriz $M[k, p]$ están dados por:
 55

$$M = (Z^*Z + \epsilon I)^{-1}Z^*Y$$

En esta formulación, la matriz Z contiene todas las entradas de las líneas de retardo derivadas.

Tomando inicialmente el caso para la reconstrucción de una señal $\hat{Y}[k]$ para una sub-banda determinada b, donde existen A entradas desde las líneas de retardo derivadas, se tiene:

$$Z = \begin{bmatrix} Z_1[0, b] & \dots & Z_1[-(A-1), b] & Z_s[0, b] & \dots & Z_s[-(A-1), b] \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Z_1[K-1, b] & \dots & Z_1[K-1-(A-1), b] & Z_s[K-1, b] & \dots & Z_s[K-1-(A-1), b] \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_1[0, b] \\ \vdots \\ Y_1[K-1, b] \end{bmatrix}$$

$$M = \begin{bmatrix} m_1[0, b] & \dots & m_s[0, b] \\ \vdots & \ddots & \vdots \\ m_1[A-1, b] & \dots & m_s[A-1, b] \end{bmatrix} = (Z^*Z + \epsilon I)^{-1}Z^*Y$$

Los coeficientes de matriz de convolución resultantes M [k, p] son cuantificados, codificados y transmitidos junto con las señales de base $Z_s[n]$. El descodificador puede utilizar a continuación un proceso de convolución para reconstruir $\hat{Y}[k, b]$ a partir de señales de entrada $Z_s [k, b]$.

$$\hat{Y}[k, b] = \sum_s Z_s [k, b] * m_s[\cdot, b]$$

o escrito de otra manera, utilizando una expresión de convolución:

$$\hat{Y}[k, b] = \sum_s \sum_{a=0}^{A-1} Z_s [k-a, b] m_s [a, b]$$

El enfoque de convolución se puede mezclar con un proceso de matriz lineal (sin estado).

Se puede realizar otra distinción entre matrización sin estado de valores complejos y de valores reales. A frecuencias bajas (habitualmente por debajo de 1 kHz), es preferible el proceso de convolución ($A>1$) para permitir la reconstrucción de propiedades intercanal en línea con una escala de frecuencia perceptual. A frecuencias medias, hasta 2 o 3 kHz, el sistema auditivo humano es sensible a diferencias de fase intercanal, pero no requiere una alta resolución de frecuencia para la reconstrucción de dicha fase. Esto implica que basta una única matriz de valor complejo, de una sola toma (sin estado). Para frecuencias superiores, el sistema auditivo humano es virtualmente insensible a la fase de estructura fina de la forma de onda, y basta con una matrización sin estado, de valores reales. Con frecuencias crecientes, el número de salidas de banco de filtros mapeadas sobre una banda paramétrica se incrementa habitualmente para reflejar la resolución en frecuencia no lineal del sistema auditivo humano.

En otra realización, la primera y la segunda presentaciones en el codificador se intercambian, por ejemplo la primera presentación se destina a reproducción en auriculares, y la segunda presentación se destina a reproducción en altavoces. En esta realización, la presentación de altavoces (segunda presentación) se genera aplicando parámetros de transformación dependientes del tiempo en, por lo menos, dos bandas de frecuencia a la primera presentación, donde se especifica además que los parámetros de transformación incluyen un conjunto de coeficientes de filtro para por lo menos una de las bandas de frecuencia.

En algunas realizaciones, la primera presentación se puede dividir temporalmente en una serie de segmentos, con un conjunto independiente de parámetros de transformación para cada segmento. En un refinamiento adicional, cuando no hay parámetros de transformación disponibles, los parámetros se pueden interpolar desde coeficientes anteriores.

Realización del descodificador

La figura 10 muestra una realización del descodificador 120. El flujo de bits de entrada 121 se divide en un flujo de bits de señal de base 131 y datos de parámetros de transformación 124. A continuación, un descodificador de señal de base 123 descodifica las señales de base $Z[n]$, que se procesan a continuación mediante un banco de filtros de análisis 125. Las señales resultantes del dominio de frecuencia $Z[k, b]$ con sub-banda $b = 1, 5$ son procesadas por unidades de multiplicación de matrices 126, 129 y 130. En particular, la unidad de multiplicación de matrices 126 aplica una matriz de convolución de valores complejos $M[k, p=1]$ a la señal del dominio de frecuencia $Z[k, b=1]$. Además, la unidad de multiplicador de matrices 129 aplica coeficientes de matriz de valores complejos, de una sola toma, $M[p=2]$, a la señal $Z[k, b=2]$. Por último, la unidad de multiplicación de matrices 130 aplica coeficientes de

matriz de valores reales $M[p=3]$ a señales del dominio de frecuencia $Z[k, b=3\dots 5]$. Las señales de salida de la unidad de multiplicación de matrices se convierten en una salida en el dominio de tiempo 128 por medio de un banco de filtros de síntesis 127. Las referencias a $z[n]$, $Z[k]$, etc., se refieren al conjunto de señales de base, no a una señal de base específica. Por lo tanto, $z[n]$, $Z[k]$, etc., se pueden interpretar como $z_s[n]$, $Z_s[k]$, etc., donde $0 \leq s < N$ y N es el número de señales de base.

En otras palabras, la unidad de multiplicación de matrices 126 determina muestras de salida de la sub-banda $b=1$ de una señal de salida $\hat{Y}_j[k]$ a partir de combinaciones ponderadas de muestras actuales de la sub-banda $b=1$ de señales de base $Z[k]$ y muestras anteriores de la sub-banda $b=1$ de señales de base $Z[k]$, (por ejemplo $Z[k-a]$, donde $0 < a < A$, y A es mayor que 1). Las ponderaciones utilizadas para determinar las muestras de salida de la sub-banda $b=1$ de la señal de salida $\hat{Y}_j[k]$ corresponden a la matriz de convolución de valores complejos $M[k, p=1]$ para la señal.

Además, la unidad de multiplicador de matrices 129 determina muestras de salida de la sub-banda $b=2$ de la señal de salida $\hat{Y}_j[k]$ a partir de combinaciones ponderadas de muestras actuales de sub-banda $b=2$ de las señales de base $Z[k]$. Las ponderaciones utilizadas para determinar las muestras de salida de la sub-banda $b=2$ de la señal de salida $\hat{Y}_j[k]$ corresponden a los coeficientes de matriz de una sola toma, de valores complejos, $M[p=2]$.

Finalmente, la unidad de multiplicador de matrices 130 determina muestras de salida de las sub-bandas $b=3\dots 5$ de la señal de salida $\hat{Y}_j[k]$ a partir de combinaciones ponderadas de muestras actuales de sub-bandas $b=3\dots 5$ de las señales de base $Z[k]$. Las ponderaciones utilizadas para determinar muestras de salida de las sub-bandas $b=3\dots 5$ de la señal de salida $\hat{Y}_j[k]$ corresponden a los coeficientes de matriz de valores reales $M[p=3]$.

En algunos casos, el descodificador de señal de base 123 puede funcionar sobre señales a la misma resolución en frecuencia que la proporcionada por el banco de filtros de análisis 125. En dichos casos, el descodificador de señal de base 125 puede estar configurado para entregar señales en el dominio de frecuencia $Z[k]$ en lugar de señales en el dominio de tiempo $z[n]$, en cuyo caso se puede omitir el banco de filtros de análisis 125. Además, en algunos casos, puede ser preferible aplicar coeficientes de matriz de una sola toma, de valores complejos, en lugar de coeficientes de matriz de valores reales, a señales en el dominio de frecuencia $Z[k, b = 3\dots 5]$.

En la práctica, los coeficientes de matriz M se pueden actualizar a lo largo del tiempo; por ejemplo, asociando tramas individuales de las señales de base con coeficientes de matriz M . Alternativa o adicionalmente, los coeficientes de matriz M se complementan con marcas de tiempo, que indican en qué tiempo o intervalo de señales de base $z[n]$ se deberían aplicar las matrices. Para reducir la tasa de bits de transmisión asociada con actualizaciones de matrices, el número de actualizaciones se limita de forma ideal, con el resultado de una distribución de actualizaciones de matriz dispersa en el tiempo. Dichas actualizaciones infrecuentes de matrices requieren procesamiento dedicado para garantizar transiciones suaves de una instancia de la matriz a la siguiente. Las matrices M se pueden proporcionar asociadas con segmentos de tiempo (tramas) específicos y/o zonas de frecuencia de las señales de base Z . El descodificador puede utilizar diversos procedimientos de interpolación para garantizar una transición suave desde instancias subsiguientes de la matriz M a lo largo del tiempo. Un ejemplo de dicho procedimiento de interpolación es calcular tramas en ventana, solapadas, de las señales Z y calcular un correspondiente conjunto de señales de salida Y para cada una de dichas tramas utilizando los coeficientes de matriz M asociados con dicha trama particular. Las tramas subsiguientes pueden a continuación agregarse utilizando una técnica de solapar-añadir que proporciona una transición suave de desvanecimiento cruzado. Alternativamente, el descodificador puede recibir marcas de tiempo asociadas con matrices M , que describen los coeficientes de matriz deseados en instancias de tiempo específicas. Para muestras de audio entre marcas de tiempo, los coeficientes de matriz de la matriz M se pueden interpolar utilizando medios lineales, cúbicos, limitados en banda u otros, para que la interpolación garantice transiciones suaves. Junto a la interpolación a través del tiempo, se pueden utilizar técnicas similares para interpolar coeficientes de matriz a través de la frecuencia.

Por lo tanto, el presente documento describe un procedimiento (y un correspondiente codificador 90) para representar una segunda presentación de canales u objetos de audio X_i como un flujo de datos que se tiene que transmitir o proporcionar a un correspondiente descodificador 100. El procedimiento comprende la etapa de proporcionar señales de base Z_s , representando dichas señales de base una primera presentación de los canales u objetos de audio X_i . Tal como se ha esbozado anteriormente, las señales de base Z_s se pueden determinar a partir de canales u objetos de audio X_i utilizando primeros parámetros de entrega G (es decir, utilizando notablemente una primera matriz de ganancia, por ejemplo, para panoramización de amplitud). La primera presentación puede estar destinada a reproducción en altavoces o a reproducción en auriculares. Por otra parte, la segunda presentación puede estar destinada a reproducción en auriculares o a reproducción en altavoces. Por lo tanto, se puede llevar a cabo una transformación de reproducción en altavoces a reproducción en auriculares (o viceversa).

El procedimiento comprende además proporcionar parámetros de transformación M (notablemente, una o varias matrices de transformación), estando previstos dichos parámetros de transformación M para transformar las señales de base Z_s de dicha primera presentación en señales de salida \hat{Y}_j de dicha segunda presentación. Los parámetros de transformación se pueden determinar tal como se esboza en el presente documento. En particular, las señales de salida deseadas Y_j para la segunda presentación se pueden determinar a partir de los canales u objetos de audio X_i utilizando segundos parámetros de entrega C (tal como se esboza en el presente documento). Los parámetros de

transformada M se pueden determinar minimizando la desviación de las señales de salida \hat{Y}_j respecto de las señales de salida deseadas Y_j (por ejemplo, utilizando un criterio de mínimo error cuadrático medio).

Aún más particularmente, los parámetros de transformada M se pueden determinar en el dominio de sub-banda (es decir, para diferentes bandas de frecuencia). Para este propósito, las señales de base en el dominio de sub-banda $Z[k,b]$ se pueden determinar para P bandas de frecuencia utilizando un banco de filtros del codificador 92, 93. El número B de bandas de frecuencia es mayor que uno, por ejemplo B es igual o mayor que 4, 6, 8, 10. En los ejemplos descritos en el presente documento, $B=8$ o $B=5$. Tal como se ha esbozado anteriormente, el banco de filtros del codificador 92, 93 puede comprender un banco de filtros híbridos que proporciona bandas de baja frecuencia B , teniendo las B bandas de frecuencia una resolución en frecuencia mayor que las bandas de alta frecuencia de las B bandas de frecuencia. Además, se pueden determinar señales de salida deseadas en el dominio de sub-banda $Y[k,b]$ para las B bandas de frecuencia. Los parámetros de transformada M para una o varias bandas de frecuencia se pueden determinar minimizando la desviación de las señales de salida \hat{Y}_j respecto de las señales de salida deseadas Y_j dentro de una o varias bandas de frecuencia (por ejemplo, utilizando un criterio de error cuadrático medio mínimo).

Por lo tanto, se puede especificar cada uno de los parámetros de transformación M para, por lo menos, dos bandas de frecuencia (notablemente, para P bandas de frecuencia). Además, los parámetros de transformación pueden incluir un conjunto de parámetros de matriz de convolución de múltiples tomas para por lo menos una de las bandas de frecuencia.

Por lo tanto, se describe un procedimiento (y un correspondiente descodificador) para determinar señales de salida de una segunda presentación de canales/objetos de audio a partir de señales de base de una primera presentación de los objetos/canales de audio. La primera presentación se puede utilizar para reproducción en altavoces y la segunda presentación se puede utilizar para reproducción en auriculares (o viceversa). Las señales de salida se determinan utilizando parámetros de transformación para diferentes bandas de frecuencia, donde los parámetros de transformación para por lo menos una de las bandas de frecuencia comprenden parámetros de matriz de convolución de múltiples tomas. Como resultado de utilizar parámetros de matriz de convolución de múltiples tomas para por lo menos una de las bandas de frecuencia, se puede reducir la complejidad computacional de un descodificador 100, notablemente reduciendo la resolución en frecuencia de un banco de filtros utilizado por el descodificador.

Por ejemplo, determinar una señal de salida para una primera banda de frecuencia utilizando parámetros de matriz de convolución de múltiples tomas puede comprender determinar una muestra actual de la primera banda de frecuencia de la señal de salida como una combinación ponderada de muestras actual, y una o varias anteriores, de la primera banda de frecuencia de las señales de base, donde las ponderaciones utilizadas para determinar la combinación ponderada corresponden a los parámetros de matriz de convolución de múltiples tomas para la primera banda de frecuencia. Habitualmente, uno o varios de los parámetros de matriz de convolución de múltiples tomas para la primera banda de frecuencia tienen valores complejos.

Además, determinar una señal de salida para una segunda banda de frecuencia puede comprender determinar una muestra actual de la segunda banda de frecuencia de la señal de salida como una combinación ponderada de muestras actuales de la segunda banda de frecuencia de las señales de base (y no basada en muestras anteriores de la segunda banda de frecuencia de la señal de base), donde las ponderaciones utilizadas para determinar la combinación ponderada corresponden a parámetros de transformación para la segunda banda de frecuencia. Los parámetros de transformación para la segunda banda de frecuencia pueden tener valores complejos, o alternativamente pueden tener valores reales.

En particular, se puede determinar el mismo conjunto de parámetros de matriz de convolución de múltiples tomas para, por lo menos, dos bandas de frecuencia adyacentes de las B bandas de frecuencia. Tal como se muestra en la figura 7, se puede determinar un único conjunto de parámetros de matriz de convolución de múltiples tomas para las bandas de frecuencia proporcionadas por el banco de filtros de Nyquist (es decir, para las bandas de frecuencia que tienen una relativamente alta resolución de frecuencia). Haciendo esto, se puede omitir la utilización de un banco de filtros de Nyquist dentro del descodificador 100, reduciendo de ese modo la complejidad computacional del descodificador 100 (mientras se mantiene la calidad de las señales de salida para la segunda presentación).

Además, se puede determinar el mismo parámetro de transformada de valores reales para, por lo menos, dos bandas de alta frecuencia adyacentes (tal como se muestra en el contexto de la figura 7). Con ello, se puede reducir más la complejidad computacional del descodificador 100 (manteniendo al mismo tiempo la calidad de las señales de salida para la segunda presentación).

Interpretación

La referencia en toda esta memoria descriptiva a "una realización" o "algunas realizaciones" significa que un aspecto particular, estructura o característica descritas en relación con la realización se incluye, por lo menos, en una realización de la presente invención. Por lo tanto, las apariciones de las expresiones "en una realización" o "en algunas realizaciones" en varios lugares en toda esta memoria descriptiva no se refieren necesariamente todas a la

misma realización.

Tal como se utiliza en la presente memoria, salvo que se especifique lo contrario, la utilización de los adjetivos ordinales "primero", "segundo", "tercero", etc., para describir un objeto común, indican tan sólo que se está aludiendo a diferentes instancias de objetos similares, y no pretenden implicar que los objetos así descritos tengan que encontrarse en una secuencia determinada, ya sea temporal, espacialmente, en orden o de cualquier otro modo.

5 En las siguientes reivindicaciones y la descripción de la presente memoria, cualesquiera de las expresiones comprende, comprendido en, o que comprende, es una expresión abierta que significa que incluye por lo menos los elementos/características que siguen, pero sin excluir otros. Por lo tanto, la expresión que comprende, cuando se utiliza en las reivindicaciones, no se deberá interpretar como que se limita a los medios o elementos o etapas enumerados a continuación. Por ejemplo, el alcance de la expresión un dispositivo que comprende A y B no se deberá limitar a dispositivos que consisten solamente en los elementos A y B. Cualesquiera de las expresiones incluye, o que incluye, tal como se utilizan en la presente memoria, es asimismo una expresión abierta que significa asimismo que incluye por lo menos los elementos/características que siguen la expresión, pero sin excluir otros. Por lo tanto, que incluye es sinónimo de un medio que comprende.

15 Tal como se utiliza en la presente memoria, la expresión "a modo de ejemplo" se utiliza en el sentido de proporcionar ejemplos, no indicando calidad. Es decir, una "realización a modo de ejemplo" es una realización proporcionada como ejemplo, sin que sea necesariamente una realización de calidad ejemplar.

20 Se apreciará que en la descripción anterior de realizaciones a modo de ejemplo de la invención, en ocasiones se agrupan diversas características de la invención conjuntamente en una única realización, figura o descripción de la misma con el objeto de simplificar la descripción y de ayudar a la comprensión de uno o varios de los diversos aspectos inventivos. Sin embargo, no se debe interpretar que este método de la invención refleja la intención de que la invención reivindicada requiere más características de las expresamente enumeradas en cada reivindicación. Por el contrario, tal como reflejan las siguientes reivindicaciones, los aspectos inventivos residen en menos de todas las características de una única realización dada a conocer en lo anterior.

25 Además, algunas de las realizaciones se describen en la presente memoria como un procedimiento o combinación de elementos de un procedimiento que se puede implementar mediante un procesador de un sistema informático, o por otros medios para llevar a cabo la función. Por lo tanto, un procesador con las instrucciones necesarias para llevar a cabo dicho procedimiento o elemento de un procedimiento, forma un medio para llevar a cabo el procedimiento o elemento de un procedimiento. Además, un elemento descrito en la presente memoria de una
30 realización de aparato es un ejemplo de un medio para llevar a cabo la función realizada por el elemento con el propósito de llevar a cabo la invención.

En la descripción dada a conocer en la presente memoria, se exponen numerosos detalles específicos. Sin embargo, se comprende que las realizaciones de la invención se pueden practicar sin estos detalles específicos. En otros casos, procedimientos, estructuras y técnicas bien conocidas no se han mostrado en detalle para no
35 obscurecer la comprensión de esta descripción.

De manera similar, cabe destacar que no se debe interpretar que la expresión acoplado, cuando se utiliza en las reivindicaciones, se limita solamente a conexiones directas. Se pueden utilizar las expresiones "acoplado" y "conectado", junto con sus derivados. Se debe entender que estas expresiones no se consideran sinónimos entre sí. Por lo tanto, el alcance de la expresión un dispositivo A acoplado a un dispositivo B no se debe limitar a dispositivos o sistemas en los que una salida de un dispositivo A esté conectada directamente una entrada del dispositivo B. Esto significa que existe un trayecto entre una salida de A y una entrada de B que puede ser un trayecto que incluya otros dispositivos o medios. "Acoplado" puede significar que dos o más elementos están en contacto directo físico o eléctrico, o bien que dos o más elementos no están en contacto directo entre sí pero siguen no obstante cooperando o interactuando entre sí.

45 Por lo tanto, aunque se han descrito las que se consideran realizaciones preferidas de la invención, los expertos en la materia reconocerán que se pueden realizar a aquellas otras modificaciones sin salir del alcance la invención, tal como se define en las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un procedimiento para descodificar una señal de audio codificada, que comprende:
 - recibir, mediante un descodificador (120), un flujo de bits de entrada (121);
 - 5 dividir el flujo de bits de entrada (121) en un flujo de bits de señal de base (131) y datos de parámetros de transformación (124);
 - descodificar, mediante un descodificador de señal de base (123), el flujo de bits de señal de base (131) para generar señales de base que representan una primera presentación de la señal de audio;
 - en el que dichos datos de parámetros de transformación están para transformar dichas señales de base en señales de salida que representan una segunda presentación de la señal de audio;
 - 10 procesar las señales de base mediante un banco de filtros de análisis (125) para generar señales de dominio de frecuencia que tienen una pluralidad de sub-bandas;
 - aplicar, mediante una primera unidad de multiplicación de matrices (126), una matriz de convolución de valores complejos a una primera sub-banda de las señales de dominio de frecuencia;
 - 15 aplicar, mediante una segunda unidad de multiplicación de matrices (129), coeficientes de matriz de valores complejos a una segunda sub-banda de las señales de dominio de frecuencia;
 - aplicar, mediante una tercera unidad de multiplicación de matrices (130), coeficientes de matriz de valores reales a una o más de las restantes sub-bandas de las señales de dominio de frecuencia; y
 - convertir, mediante un banco de filtros de síntesis (127), señales de salida desde las unidades de multiplicación de matrices en una salida de dominio de tiempo (128),
 - 20 en el que dichos datos de parámetros de transformación (124) incluyen coeficientes de dicha matriz de convolución de valores complejos, dichos coeficientes de matriz de valores complejos y dichos coeficientes de matriz de valores reales.
2. El procedimiento de la reivindicación 1, en el que el descodificador de señal de base funciona en señales en la misma resolución de frecuencia que la proporcionada mediante banco de filtros de análisis.
- 25 3. El procedimiento de la reivindicación 1, que comprende actualizar los coeficientes de matriz a lo largo del tiempo asociando tramas individuales de las señales de base con coeficientes de matriz.
4. El procedimiento de la reivindicación 1, que comprende aumentar los coeficientes de matriz con marcas de tiempo que indican en qué momento o intervalo de las señales de base se deben aplicar las matrices.
5. Un descodificador (120) para descodificar un flujo de bits de entrada (121), que comprende:
 - 30 un demultiplexador para dividir el flujo de bits de entrada (121) en un flujo de bits de señal de base (131) y datos de parámetros de transformación (124);
 - un descodificador de señal de base (123) para descodificar el flujo de bits de señal de base (131) para generar señales de base que representan una primera presentación de la señal de audio;
 - 35 en el que dichos datos de parámetros de transformación son para transformar dichas señales de base en señales de salida que representan una segunda presentación de la señal de audio;
 - un banco de filtros de síntesis (125) para procesar las señales de base para generar señales de dominio de frecuencia que tienen una pluralidad de sub-bandas;
 - una primera unidad de multiplicación de matrices (126) para aplicar unos coeficientes de matriz de convolución de valores complejos a una primera sub-banda de las señales de dominio de frecuencia;
 - 40 una segunda unidad de multiplicación de matrices (129) para aplicar coeficientes de matriz de valores complejos a una segunda sub-banda de señales de dominio de frecuencia;
 - una tercera unidad de multiplicación de matrices (130) para aplicar coeficientes de matriz de valores reales a una o más sub-bandas restantes de las señales de dominio de frecuencia; y
 - un banco de filtros de síntesis (127) para convertir señales de salida de las unidades de multiplicación de matrices en una salida de dominio de tiempo (128),
 - 45 donde dichos datos de parámetros de transformación (124) incluyen coeficientes de dicha matriz de convolución de

valores complejos, dichos coeficientes de matriz de valores complejos, y dichos coeficientes de matriz de valores reales.

6. El descodificador de la reivindicación 5, donde el descodificador de señal de base está configurado para funcionar en señales en la misma resolución de frecuencia como la proporcionada por el banco de filtros de análisis.
- 5 7. El descodificador de la reivindicación 5, donde los coeficientes de matriz están actualizados a lo largo del tiempo al asociar tramas individuales de las señales de base con coeficientes de matriz.
8. El descodificador de la reivindicación 5, donde los coeficientes de matriz están aumentados con marcas de tiempo que indican en qué momento o intervalo de las señales de base deben aplicarse las matrices.

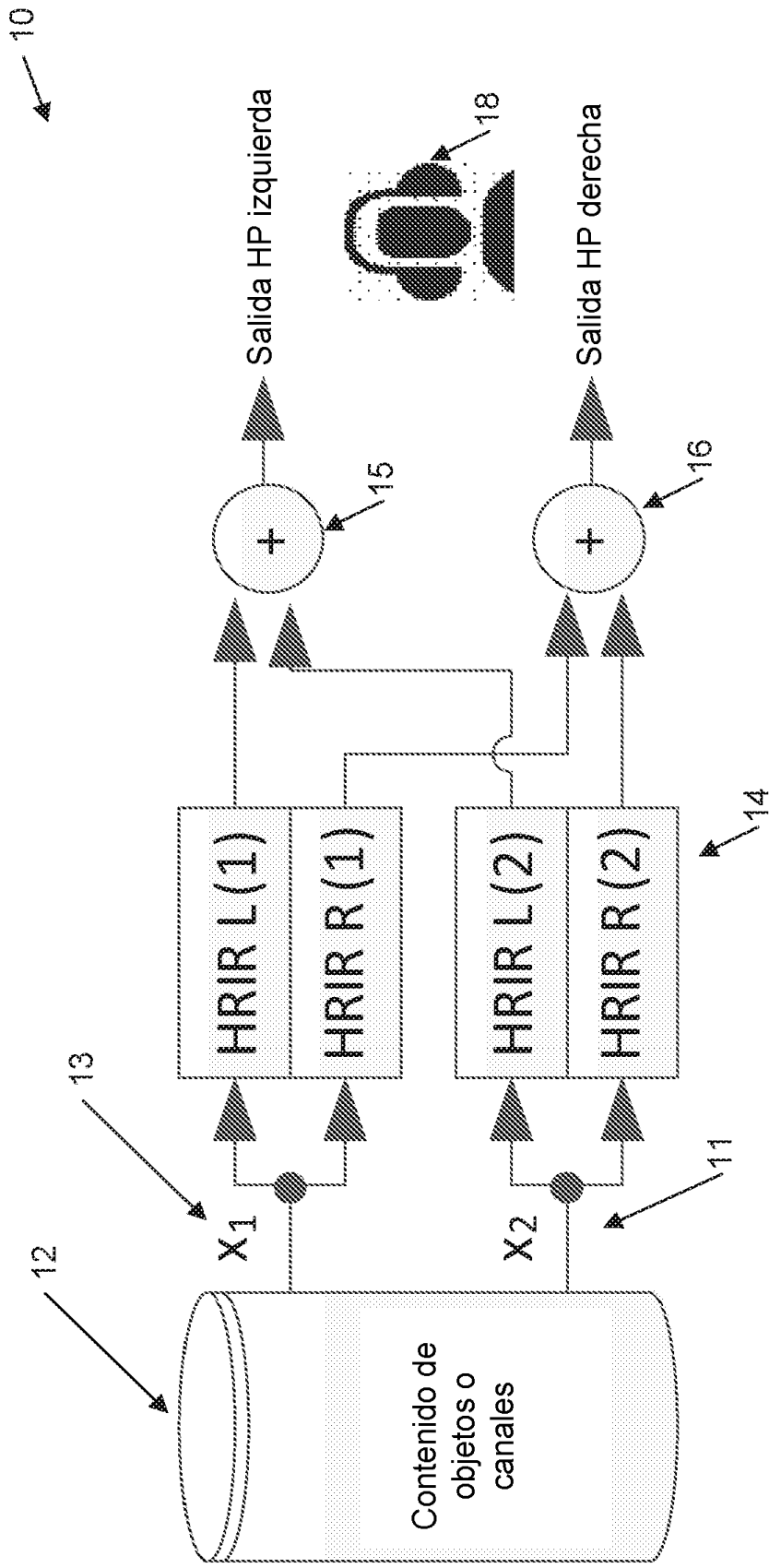


Fig. 1 (Técnica anterior)

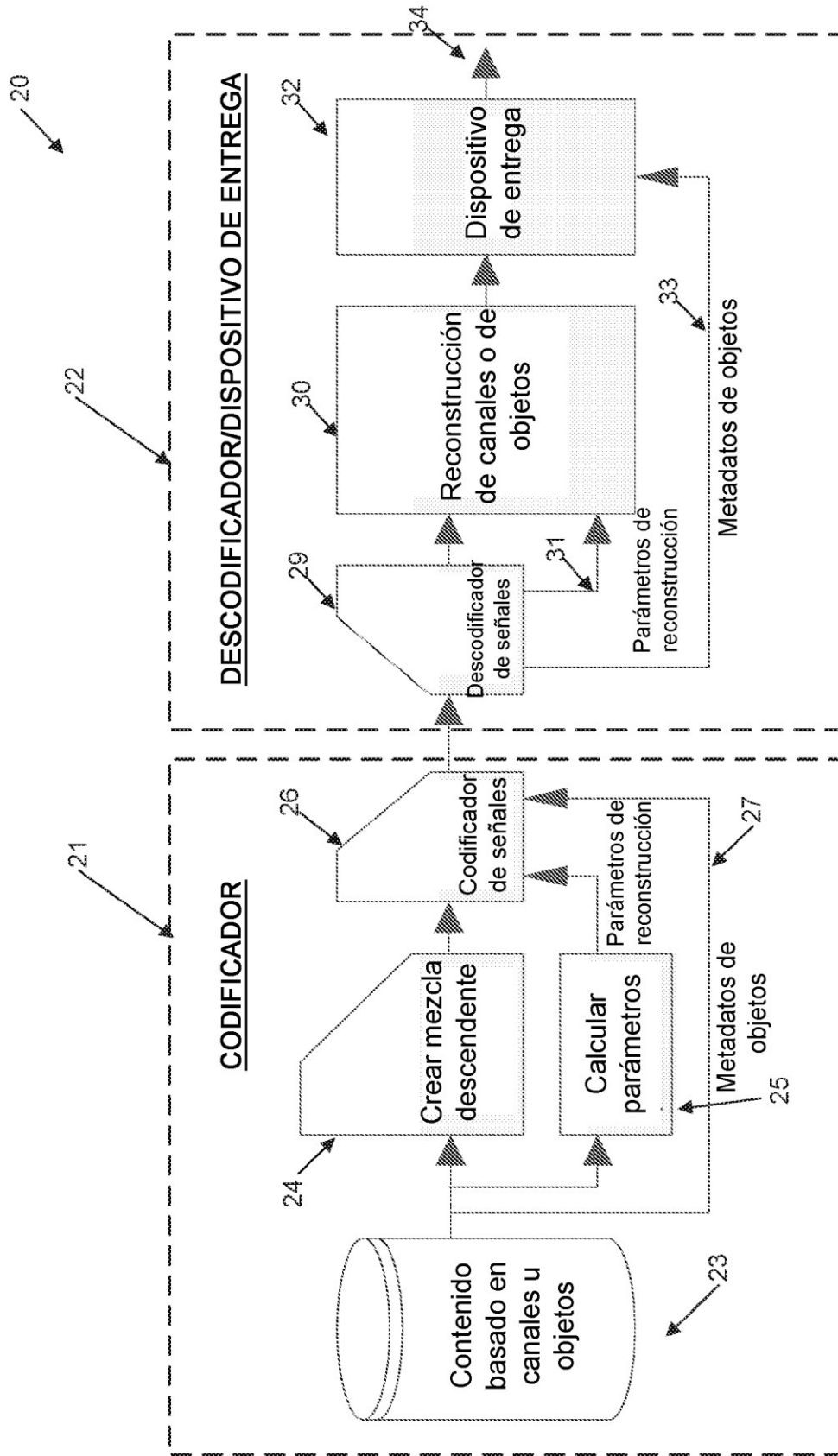


Fig. 2 (Técnica anterior)

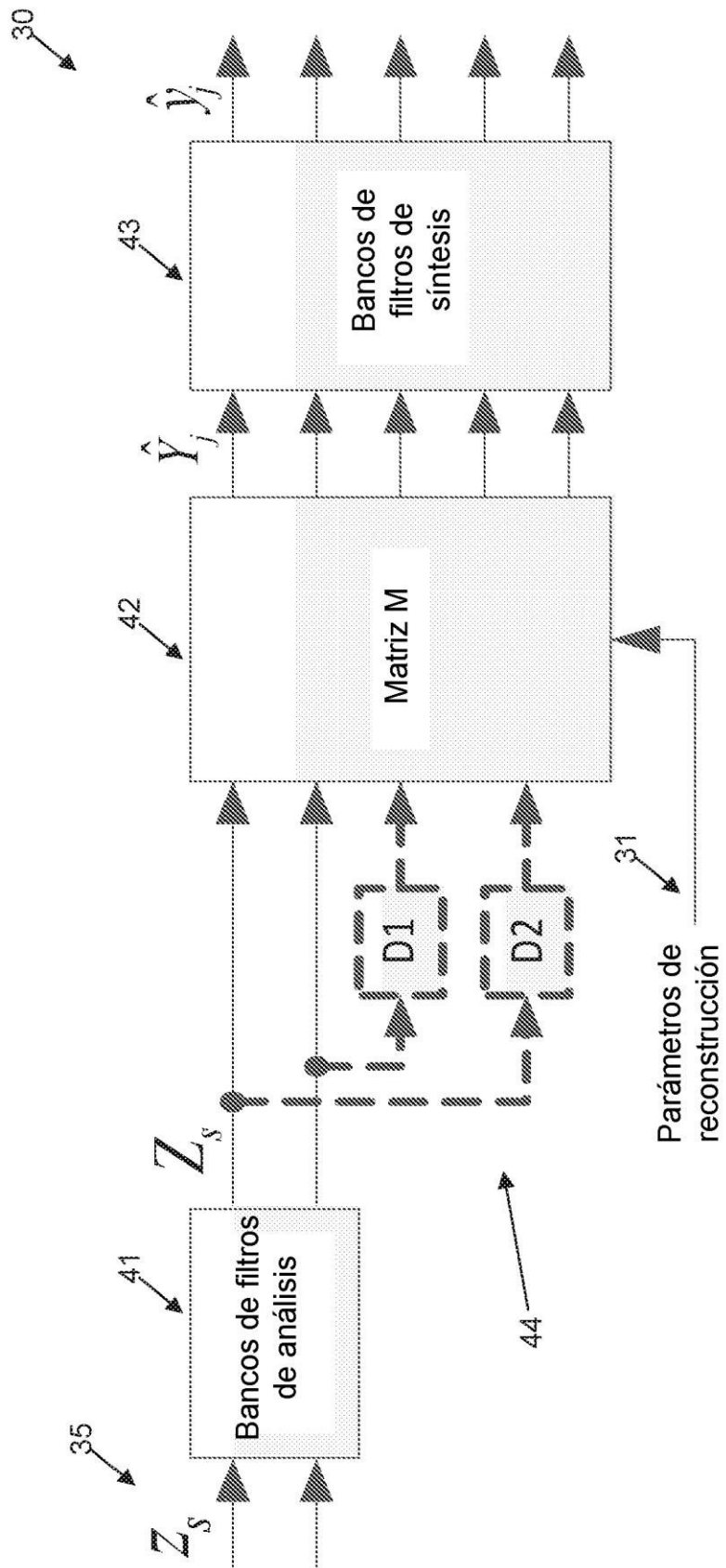


FIG. 3

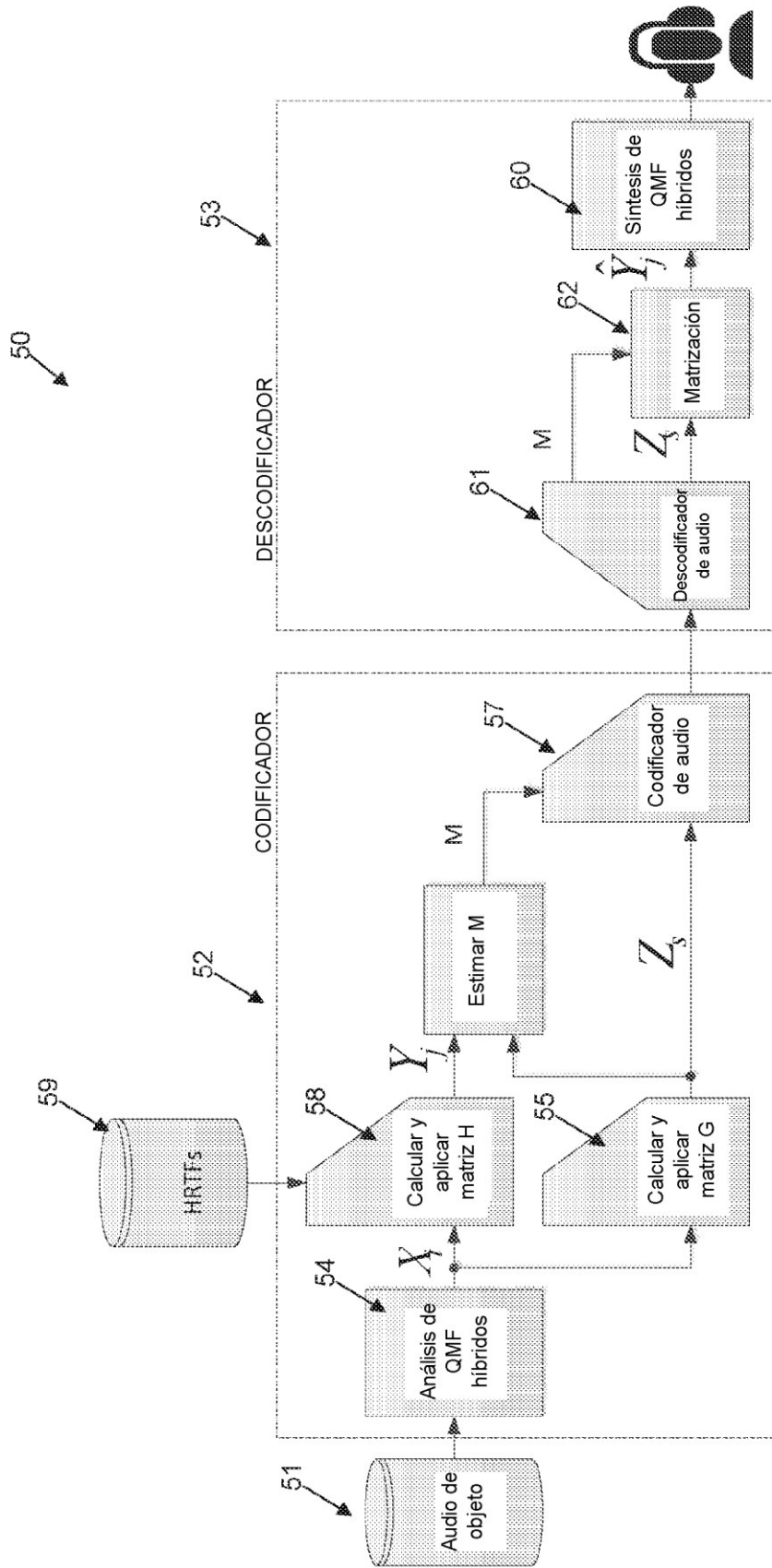


FIG. 4

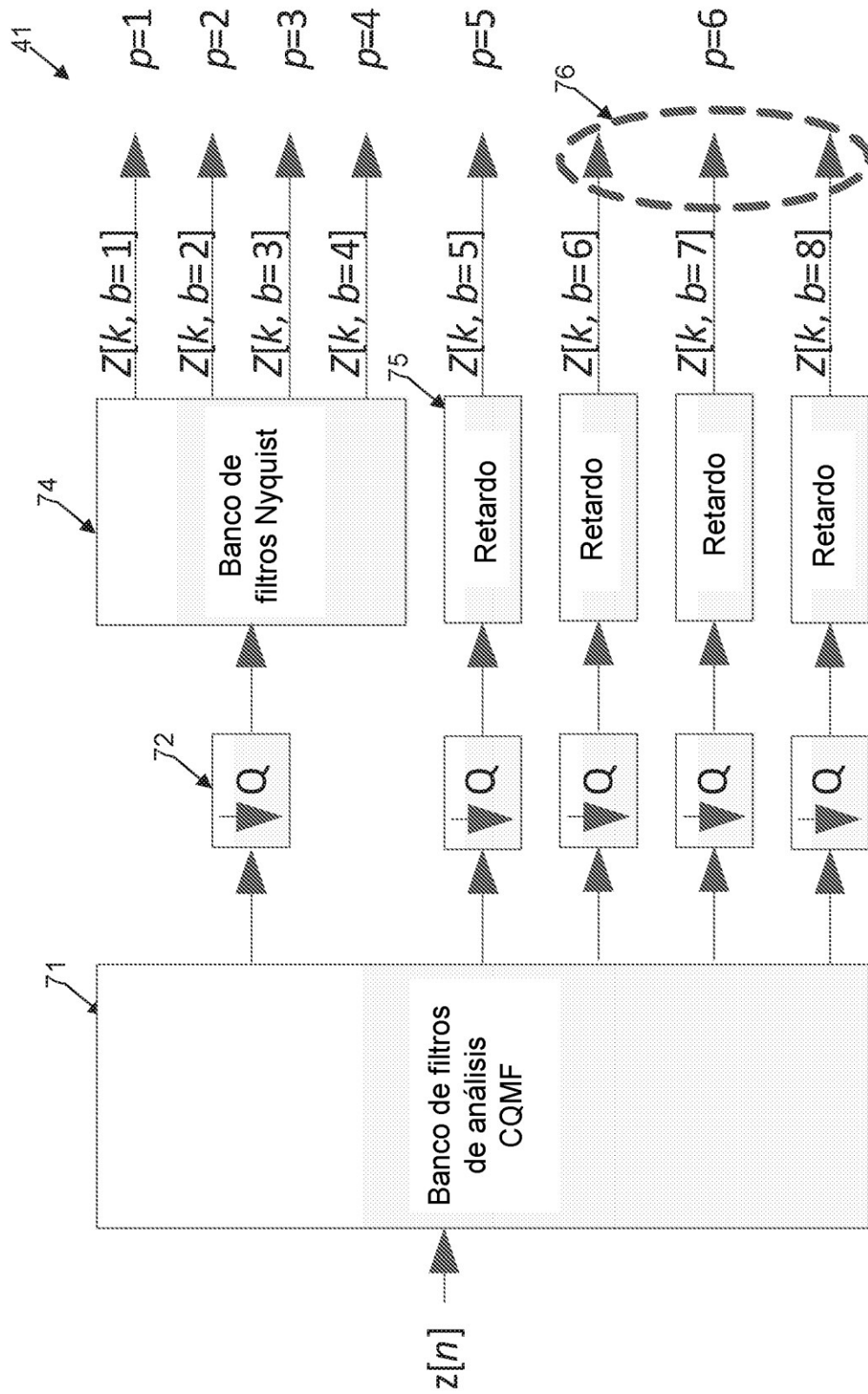


FIG. 5

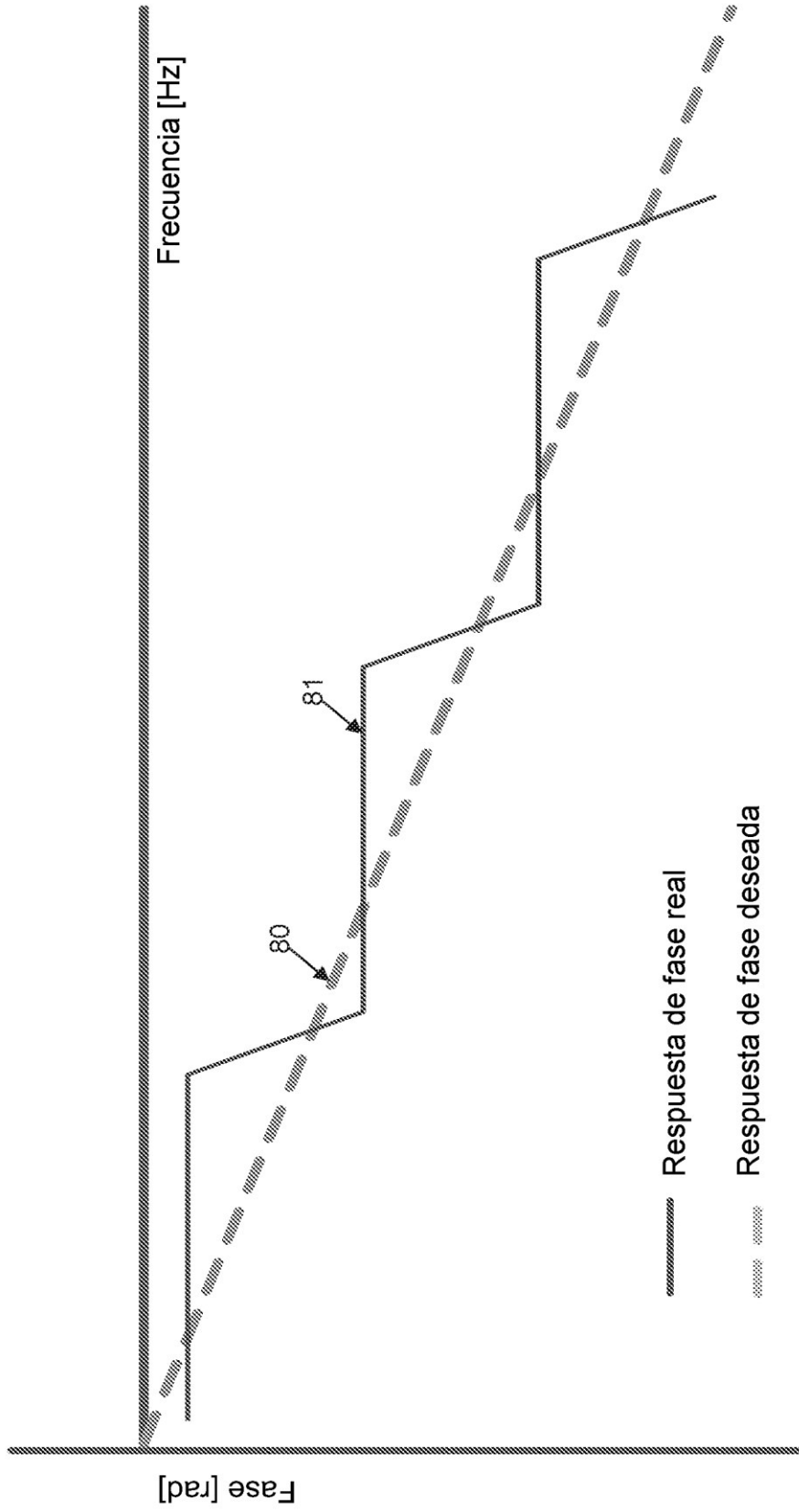


Fig. 6

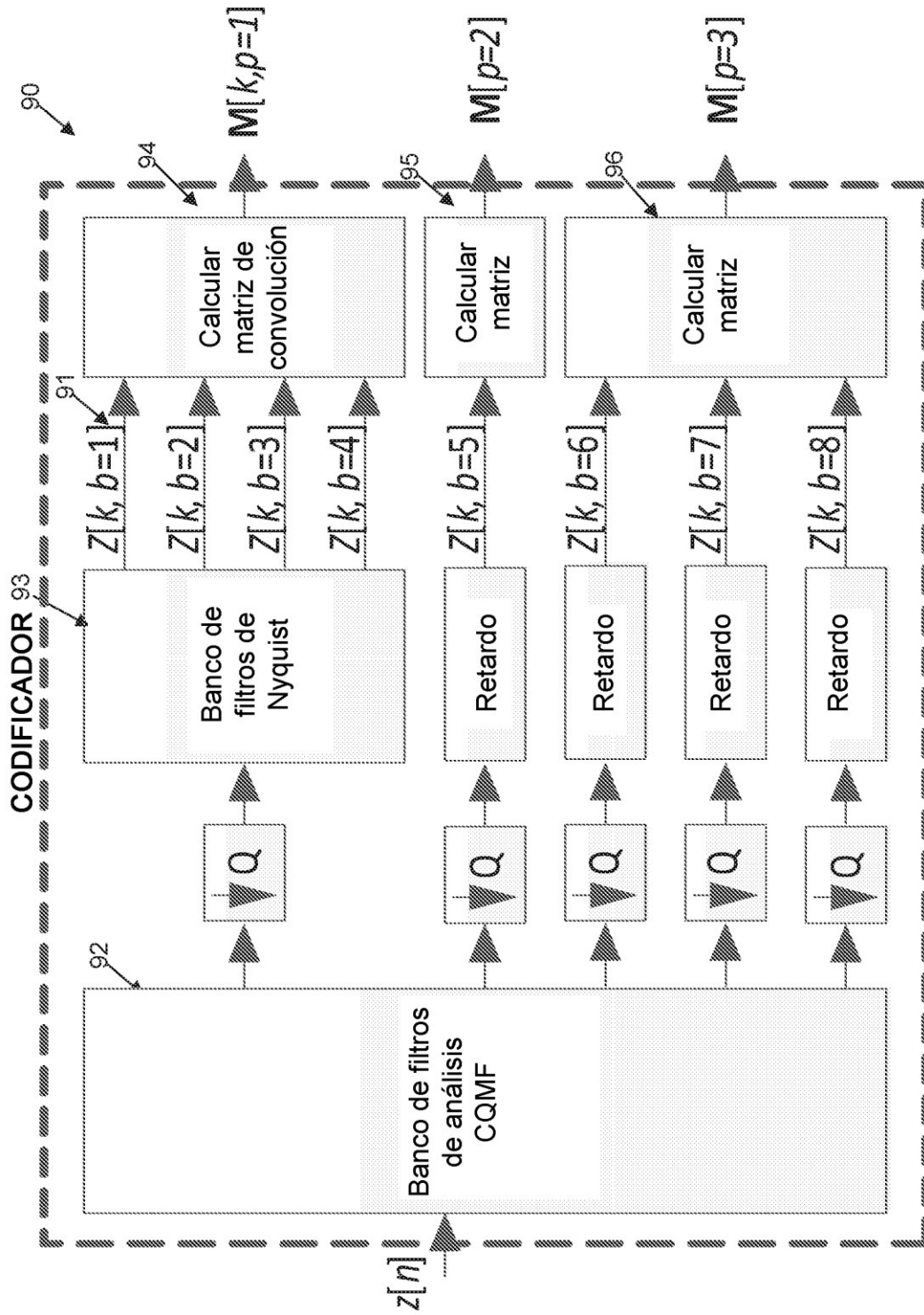


Fig. 7

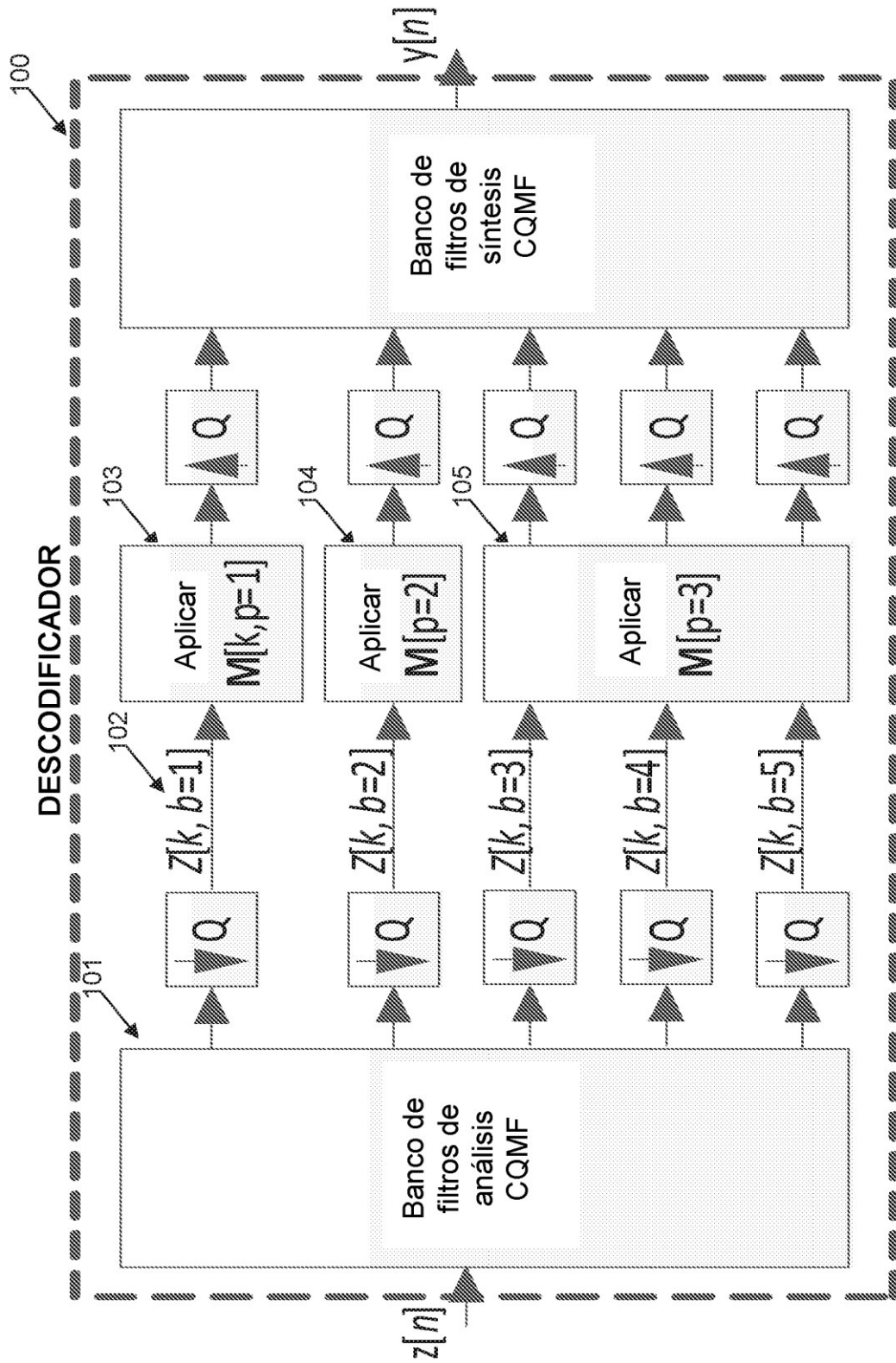


Fig. 8

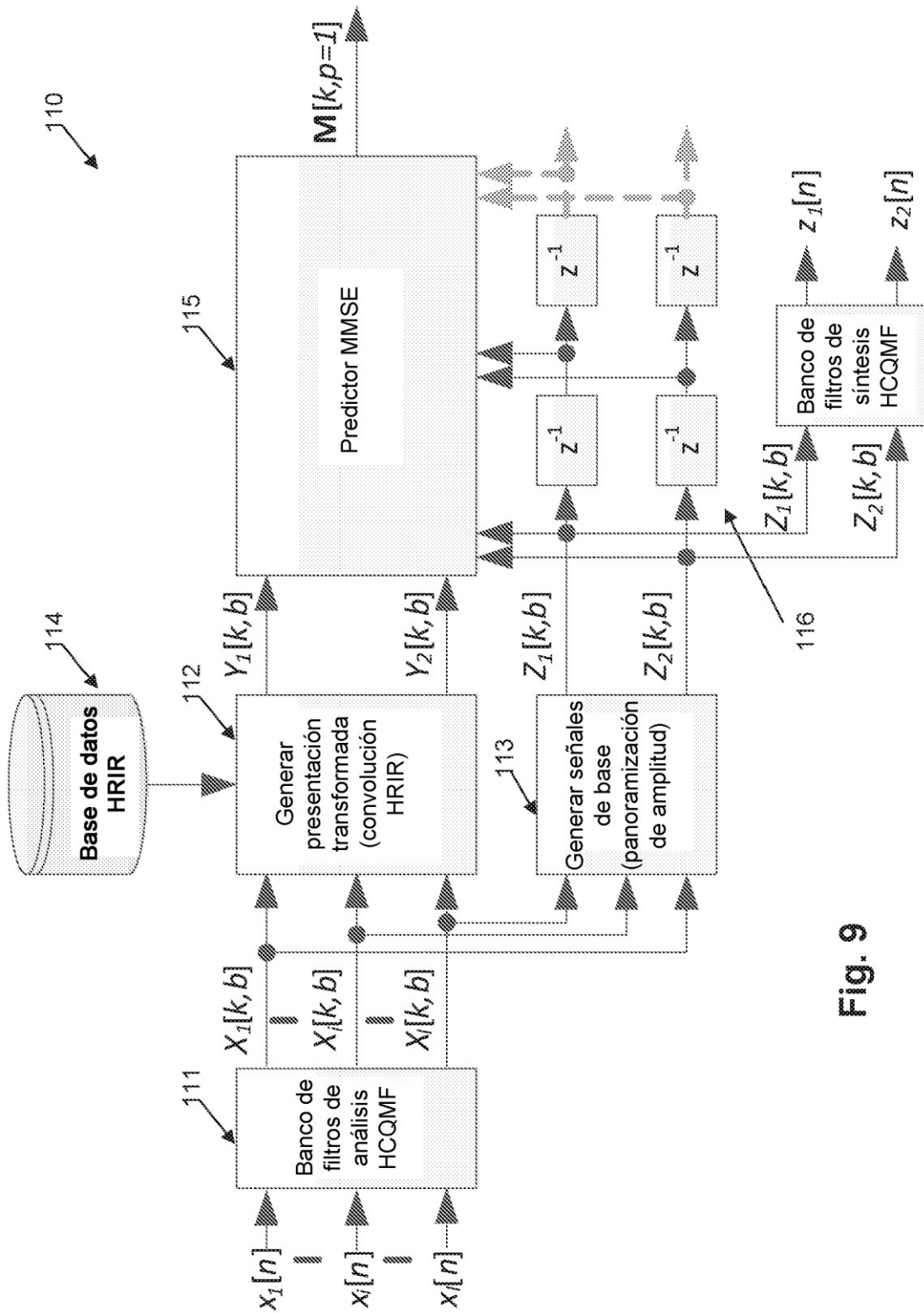


Fig. 9

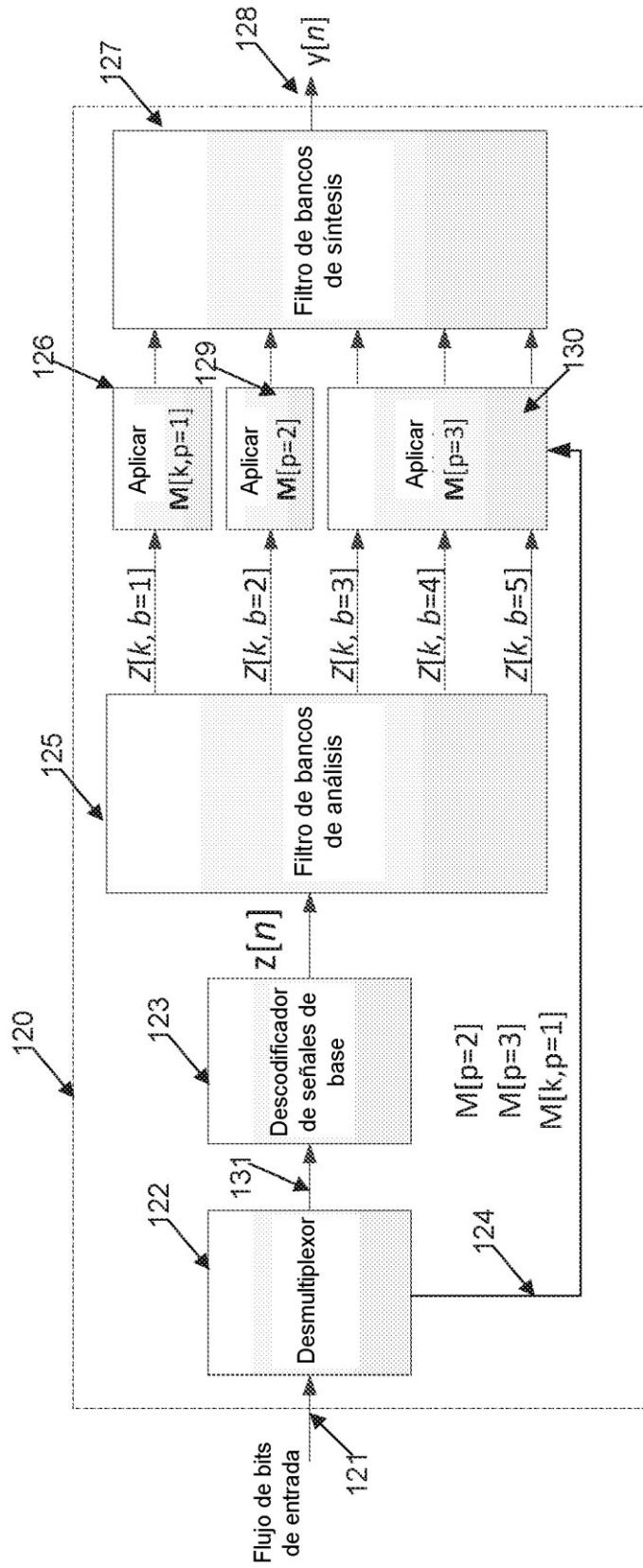


Fig. 10