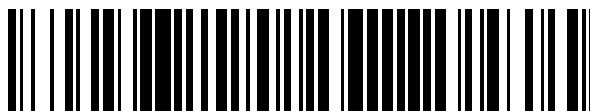


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 725 427**

51 Int. Cl.:

H04S 3/00 (2006.01)

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **17.07.2014** **PCT/EP2014/065395**

87 Fecha y número de publicación internacional: **29.01.2015** **WO15011014**

96 Fecha de presentación y número de la solicitud europea: **17.07.2014** **E 14741278 (7)**

97 Fecha y número de publicación de la concesión europea: **13.02.2019** **EP 3025515**

54 Título: **Decorrelacionador multicanal, decodificador de audio multicanal, procedimientos y programa informático que utilizan una premezcla de señales de entrada del decorrelacionador**

30 Prioridad:

22.07.2013 EP 13177374

18.10.2013 EP 13189339

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

24.09.2019

73 Titular/es:

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

**Hansastraße 27C
80686 München, DE**

72 Inventor/es:

**DISCH, SASCHA;
FUCHS, HARALD;
HELLMUTH, OLIVER;
HERRE, JÜRGEN;
MURTAZA, ADRIAN;
PAULUS, JOUNI;
RIDDERBUSCH, FALKO y
TERENTIV, LEON**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 725 427 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

Decorrelacionador multicanal, decodificador de audio multicanal, procedimientos y programa informático que utilizan una premezcla de señales de entrada del decorrelacionador

5

Campo técnico

10 **[0001]** Las formas de realización según la invención se refieren a un decorrelacionador multicanal para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador.

[0002] Otras formas de realización según la invención se refieren a un decodificador de audio multicanal para proporcionar al menos dos señales de audio de salida en base a una representación codificada.

15 **[0003]** Otras formas de realización según la invención están relacionadas con un procedimiento para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador.

20 **[0004]** Algunas formas de realización según la invención se refieren a un procedimiento para proporcionar al menos dos señales de audio de salida en base a una representación codificada.

[0005] Algunas formas de realización según la invención se refieren a un programa informático para llevar a cabo uno de dichos procedimientos.

25 **[0006]** En general, algunas formas de realización según la invención se refieren a un concepto de decorrelación para sistemas de codificación de objetos de audio paramétrica de mezcla descendente/mezcla ascendente multicanal.

Antecedentes de la invención

30

[0007] En los últimos años la demanda de almacenamiento y transmisión de contenidos de audio ha crecido constantemente. Además, los requisitos de calidad para el almacenamiento y la transmisión de contenidos de audio también han aumentado constantemente. Por consiguiente, los conceptos para la codificación y decodificación del contenido de audio han aumentado.

35

[0008] Por ejemplo, se ha desarrollado la así llamada "codificación avanzada de audio" (AAC, por sus siglas en inglés), que se describe, por ejemplo, en la norma internacional ISO/IEC 13818-7:2003. Además, se han creado algunas extensiones espaciales, como por ejemplo, el concepto así llamado "sonido envolvente de MPEG", que se describe, por ejemplo, en la norma internacional ISO/IEC 23003-1:2007. Además, mejoras adicionales para la codificación y decodificación de información espacial de señales de audio se describen en la norma internacional ISO/IEC 23003-2:2010, que se refiere a la así llamada "codificación de objetos de audio espacial".

40 **[0009]** Además, un concepto de codificación/decodificación de audio conmutable que proporciona la posibilidad de codificar tanto las señales de audio generales como las señales de voz con buena eficacia de codificación y manipular las señales de audio multicanal se define en la norma internacional ISO/IEC 23003-3:2012, que describe el concepto así llamado de "codificación de audio y voz unificado".

45 **[0010]** Además, otros conceptos convencionales se describen en las referencias, que se mencionan al final de la presente descripción.

50

[0011] El documento WO 2008/131903 A1 describe un aparato para sintetizar una señal de salida renderizada que tiene un primer canal de audio y un segundo canal de audio. El aparato incluye una fase de decorrelacionador para generar una señal de decorrelacionador basada en una señal de mezcla descendente, y un combinador para realizar una combinación ponderada de la señal de mezcla descendente y la señal decorrelacionada basada en información de objeto de audio paramétrica, información de mezcla descendente e información de renderización de destino. El combinador combina maxtrixing con decorrelación para una reproducción de escena estéreo de alta calidad de una serie de objetos de audio individuales mediante el uso de una mezcla descendente multicanal.

60 **[0012]** Sin embargo, existe un deseo de proporcionar un concepto aún más avanzado para una codificación y decodificación eficaz de escenas de audio tridimensionales.

Resumen de la invención

65 **[0013]** Formas de realización según la presente invención crean un decorrelacionador multicanal según la

reivindicación 1 o la reivindicación 29 o la reivindicación 30.

[0014] Formas de realización según la invención crean un decodificador de audio multicanal según la reivindicación 31 o la reivindicación 32 o la reivindicación 33.

5

[0015] Otras formas de realización según la invención crean procedimientos según las reivindicaciones 26, 34, 35, 36, 37 y 38.

[0016] Otras formas de realización según la invención crean programas informáticos según las reivindicaciones 28 y 39.

[0017] Una forma de realización según la reivindicación 1 crea un decorrelacionador multicanal para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador. El decorrelacionador multicanal está configurado para premezclar un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K, en el que $K < N$. El decorrelacionador multicanal está configurado para proporcionar un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K. El decorrelacionador multicanal está configurado además para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto de señales de salida del decorrelacionador N', en el que $N' > K'$.

[0018] Esta forma de realización según la invención se basa en la idea de que una complejidad de la decorrelación puede ser reducida premezclando el primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K, en el que el segundo conjunto de señales de entrada del decorrelacionador K comprende menos señales que el primer conjunto de señales de entrada del decorrelacionador N. Por consiguiente, la funcionalidad fundamental del decorrelacionador se realiza solo en las señales K (las señales de entrada del decorrelacionador K del segundo conjunto) de tal modo que, por ejemplo, solo se requieren decorrelacionadores K (individuales) (o decorrelaciones individuales) (y no decorrelacionadores N). Además, para proporcionar señales de salida del decorrelacionador N', se realiza una mezcla ascendente, en la que el primer conjunto de señales de salida del decorrelacionador K' es mezclado de manera ascendente en el segundo conjunto de señales de salida del decorrelacionador N'. Por consiguiente, es posible obtener un número comparativamente grande de señales decorrelacionadas (a saber, señales N' del segundo conjunto de señales de salida del decorrelacionador) en base a un número comparativamente grande de señales de entrada del decorrelacionador (a saber, señales N del primer conjunto de señales de entrada del decorrelacionador), en donde una funcionalidad núcleo de decorrelación se realiza en base solo a señales K (por ejemplo mediante el uso solo de decorrelacionadores individuales K). Así se logra una ganancia significativa en la eficacia de decorrelación, que ayuda a ahorrar potencia de procesamiento y recursos (por ejemplo, energía).

En una forma de realización preferida, el número K de señales del segundo conjunto de señales de entrada del decorrelacionador es igual al número K' de señales del primer conjunto de señales de salida del decorrelacionador. Por consiguiente, puede haber, por ejemplo, K decorrelacionadores individuales, cada uno de los cuales recibe una señal de entrada del decorrelacionador (del segundo conjunto de señales de entrada del decorrelacionador) del premezclado, y cada uno de los cuales proporciona una señal de salida del decorrelacionador (del primer conjunto de señales de salida del decorrelacionador) al mezclado ascendente. Así se pueden usar decorrelacionadores individuales simples, cada uno de los cuales proporciona una señal de salida en base a una señal de entrada.

En otra forma de realización preferida, el número N de señales del primer conjunto de señales de entrada del decorrelacionador puede ser igual al número N' de señales del segundo conjunto de señales de salida del decorrelacionador. Así, el número de señales recibidas por el decorrelacionador multicanal es igual al número de señales proporcionadas por el decorrelacionador multicanal, de tal modo que el decorrelacionador multicanal aparece, desde fuera, como un banco de N decorrelacionadores independientes (en donde, sin embargo, el resultado de la decorrelación puede comprender algunas imperfecciones debido al uso solo de las K señales de entrada para el núcleo del decorrelacionador). Por consiguiente, el decorrelacionador multicanal se puede usar como un reemplazo idéntico para decorrelacionadores convencionales que tienen un número igual de señales de entrada y señales de salida. Además, debería señalarse que el mezclado ascendente puede derivarse, por ejemplo, del premezclado en una configuración de este tipo con moderado esfuerzo.

En una forma de realización preferida, el número N de señales del primer conjunto de señales de entrada del decorrelacionador puede ser mayor que, o igual a, 3, y el número N' de señales del segundo conjunto de señales de salida del decorrelacionador también puede ser mayor que, o igual a, 3. En un caso como éste, el decorrelacionador multicanal puede proporcionar una eficacia particular.

[0019] En una forma de realización preferida, el decorrelacionador multicanal puede ser configurado para premezclar el primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K mediante el uso de una matriz de premezclado (es decir, mediante el uso de una funcionalidad de premezclado lineal). En este caso, el decorrelacionador multicanal puede ser configurado para obtener el primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K (por ejemplo, mediante el uso de decorrelacionadores individuales). El

decorrelacionador multicanal también puede ser configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en el segundo conjunto de señales de salida del decorrelacionador N' mediante el uso de una matriz de post-mezclado, es decir, mediante el uso de una función de post-mezclado lineal. Por consiguiente, las distorsiones se pueden mantener pequeñas. Además el premezclado y el post-mezclado (también denominado mezclado ascendente) pueden ser realizados de una manera computacionalmente eficaz. En una forma de realización preferida, el decorrelacionador multicanal puede ser configurado para seleccionar la matriz de premezclado en dependencia de las posiciones espaciales a las cuales están asociadas las señales de canales del primer conjunto de señales de entrada del decorrelacionador N. Por consiguiente, las dependencias (o correlaciones) espaciales pueden ser consideradas en el proceso de premezclado, que ayuda a evitar una degradación excesiva debido al proceso de premezclado realizado en el decorrelacionador multicanal.

[0020] En una forma de realización preferida, como se define en la reivindicación 29, el decorrelacionador multicanal está configurado para seleccionar la matriz de premezclado en dependencia de las características de correlación o características de covarianza de las señales de canales del primer conjunto de señales de entrada del decorrelacionador N. Una funcionalidad como ésta también puede ayudar a evitar distorsiones excesivas debido al premezclado realizado por el decorrelacionador multicanal. Por ejemplo, las señales de entrada del decorrelacionador (del primer conjunto de señales de entrada del decorrelacionador), que están estrechamente relacionadas (es decir, comprenden una alta correlación cruzada o una alta covarianza cruzada) pueden ser combinadas, por ejemplo, en una sola señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador, y pueden ser procesadas, en consecuencia, por ejemplo, por un decorrelacionador individual común (del núcleo del decorrelacionador). Así se puede evitar que señales de entrada del decorrelacionador (del primer conjunto de señales de entrada del decorrelacionador) sustancialmente diferentes sean premezcladas (o mezcladas de manera descendente) en una sola señal de entrada del decorrelacionador (del segundo conjunto de señales de entrada del decorrelacionador), que es entrada en el núcleo del decorrelacionador, ya que esto resultará típicamente en señales de salida del decorrelacionador inapropiadas (las que perturbarían, por ejemplo, una percepción espacial cuando se usan para llevar señales de audio a características de correlación cruzada o características de covarianza cruzadas deseadas). Por consiguiente, el decorrelacionador multicanal puede decidir, de una manera inteligente, qué señales deberían ser combinadas en el proceso de premezclado (o mezclado descendente) para permitir un buen compromiso entre la eficacia de la decorrelación y la calidad del audio. En una forma de realización preferida, el decorrelacionador multicanal está configurado para determinar la matriz de premezclado de tal modo que un producto de matriz entre la matriz de premezclado y un hermítico de ésta está bien acondicionado con respecto a una operación de inversión. Por consiguiente, la matriz de premezclado puede ser elegida de tal modo que se puede determinar una matriz de post-mezclado sin problemas numéricos. En una forma de realización preferida, el decorrelacionador multicanal está configurado para obtener la matriz de post-mezclado en base a la matriz de premezclado mediante el uso de algunas operaciones de multiplicación de matrices e inversión de matrices. De esta manera, la matriz de post-mezclado se puede obtener de forma eficaz, de tal modo que la matriz de post-mezclado está bien adaptada al proceso de premezclado.

[0021] En una forma de realización preferida, el decorrelacionador multicanal está configurado para recibir una información acerca de una configuración de representación asociada con las señales de canales del primer conjunto de señales de entrada del decorrelacionador N. En este caso, el decorrelacionador multicanal está configurado para seleccionar una matriz de premezclado en dependencia de la información acerca de la configuración de representación. Por consiguiente, la matriz de premezclado puede ser seleccionada de una manera que está bien adaptada a la configuración de representación, de tal modo que se puede obtener una buena calidad de audio.

[0022] En una forma de realización preferida, el decorrelacionador multicanal está configurado para combinar señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones espacialmente adyacentes de una escena de audio cuando se realiza el premezclado. Así, el hecho de que las señales de canales asociadas con posiciones espacialmente adyacentes de una escena de audio son típicamente similares es explotado cuando se realiza el premezclado. En consecuencia, señales de audio similares pueden ser combinadas en el premezclado y procesadas mediante el uso del mismo decorrelacionador individual en el núcleo del decorrelacionador. Por consiguiente se pueden evitar las degradaciones inaceptables del contenido del audio.

[0023] En una forma de realización preferida, el decorrelacionador multicanal está configurado para combinar señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes de una escena de audio cuando se realiza el premezclado. Este concepto se basa en el hallazgo de que las señales de audio de posiciones vertical y espacialmente adyacentes de la escena de audio son típicamente similares. Además, la percepción humana no es particularmente sensible con respecto a diferencias entre señales asociadas con posiciones vertical y espacialmente adyacentes de la escena de audio. Por consiguiente, se ha encontrado que combinando señales de audio asociadas con posiciones vertical y espacialmente adyacentes de la escena de audio no resulta en una degradación sustancial de una impresión auditiva obtenida en base a las señales de audio decorrelacionadas.

[0024] En una forma de realización preferida, el decorrelacionador multicanal puede ser configurado para combinar señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con un par horizontal de posiciones espaciales que comprenden una posición del lado izquierdo y una posición del lado derecho. Se ha encontrado que las señales de canales que están asociadas con un par horizontal de posiciones espaciales que comprenden una posición del lado izquierdo y una posición del lado derecho típicamente también están un poco relacionadas ya que las señales de canales asociadas con un par horizontal de posiciones espaciales se usan típicamente para obtener una impresión espacial. Por consiguiente, se ha encontrado que es una solución razonable combinar señales de canales asociadas con un par horizontal de posiciones espaciales, por ejemplo, si no es suficiente combinar señales de canales asociadas con posiciones vertical y espacialmente adyacentes de la escena de audio, debido a que la combinación de señales de canales asociadas con un par horizontal de posiciones espaciales típicamente no resulta en una degradación excesiva de una impresión auditiva.

[0025] En una forma de realización preferida, el decorrelacionador multicanal está configurado para combinar al menos cuatro señales de canales del primer conjunto de señales de entrada del decorrelacionador N, en donde al menos dos de al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado izquierdo de una escena de audio, y en donde al menos dos de al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado derecho de una escena de audio. Por consiguiente, cuatro o más señales de canales son combinadas, de tal modo que se puede obtener una decorrelación eficaz sin comprender significativamente una impresión auditiva.

[0026] En una forma de realización preferida, al menos dos señales de canales del lado izquierdo (es decir, señales de canales asociadas con posiciones espaciales en el lado izquierdo de la escena de audio) que van a ser combinadas están asociadas con posiciones espaciales que son simétricas, con respecto a un plano central de la escena de audio, con relación a las posiciones espaciales asociadas con al menos dos señales de canales del lado derecho que van a ser combinadas (es decir, señales de canales asociadas con posiciones espaciales en el lado derecho de la escena de audio). Se ha encontrado que una combinación de señales de canales asociadas con posiciones espaciales "simétricas" típicamente lleva consigo buenos resultados, ya que las señales asociadas con tales posiciones espaciales "simétricas" están típicamente de algún modo relacionadas, lo que es ventajoso para realizar la decorrelación común (combinada).

[0027] En una forma de realización preferida, el decorrelacionador multicanal está configurado para recibir una información de complejidad que describe un número K de señales de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador. En este caso, el decorrelacionador multicanal puede estar configurado para seleccionar una matriz de premezclado en dependencia de la información de complejidad. Por consiguiente, el decorrelacionador multicanal puede ser adaptado de forma flexible a diferentes requisitos de complejidad. Así, es posible variar un compromiso entre la calidad del audio y la complejidad.

[0028] En una forma de realización preferida, el decorrelacionador multicanal está configurado para aumentar gradualmente (por ejemplo, escalonadamente) un número de señales de entrada del decorrelacionador del primer conjunto de señales de entrada del decorrelacionador que son combinadas entre sí para obtener las señales de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador con un valor decreciente de la información de complejidad. Por consiguiente, es posible combinar más y más señales de entrada del decorrelacionador del primer conjunto de señales de entrada del decorrelacionador (por ejemplo, en una sola señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador) si se desea disminuir la complejidad, lo que permite variar la complejidad con poco esfuerzo.

[0029] En una forma de realización preferida, el decorrelacionador multicanal está configurado para combinar solo señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes de una escena de audio cuando se realiza el premezclado para un primer valor de la información de complejidad. Sin embargo, el decorrelacionador multicanal puede ser configurado (también) para combinar al menos dos señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en el lado izquierdo de la escena de audio y al menos dos señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en el lado derecho de la escena de audio para obtener una señal dada del segundo conjunto de señales de entrada del decorrelacionador cuando se realiza el premezclado para un segundo valor de la información de complejidad. En otras palabras, para el primer valor de la información de complejidad, no se puede realizar una combinación de señales de canales de diferentes lados de la escena de audio, que resulta en una calidad particularmente buena de las señales de audio (y de una impresión auditiva, que se puede obtener en base a las señales de audio decorrelacionadas). En cambio, si se requiere una menor complejidad, también se puede realizar una combinación horizontal además de la combinación vertical. Se ha encontrado que éste es un concepto razonable para un ajuste escalonado de la complejidad, en donde se encuentra una degradación un poco mayor de una impresión auditiva para la complejidad reducida.

[0030] En una forma de realización preferida, el decorrelacionador multicanal está configurado para combinar al menos cuatro señales de canales del primer conjunto de señales de entrada del decorrelacionador N, en donde al menos dos de al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado izquierdo de una escena de audio, y en donde al menos dos de al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado derecho de la escena de audio cuando se realiza el premezclado para un segundo valor de la información de complejidad. Este concepto se basa en el hallazgo de que una complejidad computacional comparativamente baja se puede obtener combinando al menos dos señales de canales asociadas con posiciones espaciales en un lado izquierdo de la escena de audio y al menos dos señales de canales asociadas con posiciones espaciales en un lado derecho de la escena de audio, aún si las señales de canales no son verticalmente adyacentes (o al menos no perfectamente verticalmente adyacentes).

[0031] En una forma de realización preferida, el decorrelacionador multicanal está configurado para combinar al menos dos señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en un lado izquierdo de la escena de audio, para obtener una primera señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador, y para combinar al menos dos señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en un lado derecho de la escena de audio, para obtener una segunda señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador para un primer valor de la información de complejidad. Además, el decorrelacionador multicanal está configurado preferentemente para combinar al menos dos señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en el lado izquierdo de la escena de audio y al menos dos señales de canales del primer conjunto de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en el lado derecho de la escena de audio, para obtener una señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador para un segundo valor de la información de complejidad. En este caso, un número de señales de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador es mayor para el primer valor de la información de complejidad que para el segundo valor de la información de complejidad. En otras palabras, cuatro señales de canales, que se usan para obtener dos señales de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador para el primer valor de la información de complejidad se pueden usar para obtener una sola señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador para el segundo valor de la información de complejidad. Así, las señales que sirven como señales de entrada para dos decorrelacionadores individuales para el primer valor de la información de complejidad son combinados para servir como señales de entrada para un solo decorrelacionador individual para el segundo valor de la información de complejidad. Así, se puede obtener una reducción eficaz del número de decorrelacionadores individuales (o del número de señales de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador) para un valor reducido de la información de complejidad.

[0032] Una forma de realización según la invención crea un decodificador de audio multicanal para proporcionar al menos dos señales de audio de salida en base a una representación codificada. El decodificador de audio multicanal comprende un decorrelacionador multicanal, como se explica en esta invención.

[0033] Esta forma de realización se basa en el hallazgo de que el decorrelacionador de audio multicanal es muy adecuado para la aplicación en un decodificador de audio multicanal.

[0034] En una forma de realización preferida, el decodificador de audio multicanal está configurado para representar una pluralidad de señales de audio decodificadas, que se obtienen en base a la representación codificada, en dependencia de uno o más parámetros de representación, para obtener una pluralidad de señales de audio renderizadas. El decodificador de audio multicanal está configurado para derivar una o más señales de audio decorrelacionadas de las señales de audio renderizadas mediante el uso del decorrelacionador multicanal, en el que las señales de audio renderizadas constituyen el primer conjunto de señales de entrada del decorrelacionador, y en el que el segundo conjunto de señales de salida del decorrelacionador constituyen las señales de audio decorrelacionadas. El decodificador de audio multicanal está configurado para combinar las señales de audio renderizadas, o una versión en escala de éstas, con una o más señales de audio decorrelacionadas (del segundo conjunto de señales de salida del decorrelacionador), para obtener las señales de audio de salida. Esta forma de realización según la invención se basa en el hallazgo de que el decorrelacionador multicanal descrito en esta invención es muy adecuado para un procesamiento de post-renderización, en donde un número de señales de audio renderizadas comparativamente grande es introducido en el decorrelacionador multicanal, y en donde un número de señales decorrelacionadas comparativamente grande es combinado, a continuación, con las señales de audio renderizadas. Además, se ha encontrado que las imperfecciones causadas por el uso de un número de decorrelacionadores comparativamente pequeño (reducción de complejidad en el decorrelacionador multicanal) típicamente no resulta en una degradación severa de una calidad de las señales de audio de salida producidas por el decodificador multicanal.

En una forma de realización preferida, el decodificador de audio multicanal está configurado para seleccionar una matriz de premezclado para el uso por el decorrelacionador multicanal en dependencia de una información de

control incluida en la representación codificada. Por consiguiente, aún es posible que un codificador de audio controle la calidad de la decorrelación, de tal modo que la calidad de la decorrelación pueda ser bien adaptada al contenido de audio específico, que lleva consigo una buena compensación entre la calidad del audio y la complejidad de la decorrelación.

5

[0035] En una forma de realización preferida, como se define en la reivindicación 31, el decodificador de audio multicanal está configurado para seleccionar una matriz de premezclado para el uso por el decorrelacionador multicanal en dependencia de una configuración de salida que describe una asignación de señales de audio de salida con posiciones espaciales de la escena de audio. Por consiguiente, el decorrelacionador multicanal puede ser adaptado a un escenario de representación específico, que ayuda a evitar la degradación sustancial de la calidad del audio por la decorrelación eficaz.

10

[0036] En una forma de realización preferida, como se define en la reivindicación 32, el decodificador de audio multicanal está configurado para seleccionar entre tres o más matrices de premezclado diferentes para el uso por el decorrelacionador multicanal en dependencia de una información de control incluida en la representación codificada para una representación de salida dada. En este caso, cada una de las tres o más matrices de premezclado diferentes están asociadas con un número de señales diferente del segundo conjunto de señales de entrada del decorrelacionador K. Así, la complejidad de la decorrelación se puede ajustar en un amplio rango. En una forma de realización preferida, el decodificador de audio multicanal está configurado para seleccionar una matriz de premezclado (M_{pre}) para el uso por el decorrelacionador multicanal en dependencia de una matriz de mezclado (Dconv, Dvisual) que es usada por un convertidor de formato o renderizador que recibe al menos dos señales de audio de salida.

15

20

[0037] En otra forma de realización, como se define en la reivindicación 33, el decodificador de audio multicanal está configurado para seleccionar la matriz de premezclado (M_{pre}) para el uso por el decorrelacionador multicanal para que sea igual a una matriz de mezclado (Dconv, Dvisual) que es usada por un convertidor de formato o renderizador que recibe al menos dos señales de audio de salida.

25

[0038] Otra forma de realización según la reivindicación 26 crea un procedimiento para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador. El procedimiento comprende el premezclado de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K, en donde $K < N$. El procedimiento comprende también el suministro de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K. Además, el procedimiento comprende el mezclado ascendente del primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto de señales de salida del decorrelacionador N', en donde $N' > K'$. Este procedimiento se basa en las mismas ideas que las presentadas para el decorrelacionador multicanal descrito en lo anterior.

30

35

[0039] Además, las formas de realización según las reivindicaciones 28 y 39 crean un programa informático para llevar a cabo dichos procedimientos.

40

[0040] Además, debería señalarse que los procedimientos descritos en lo anterior pueden ser suplementados por cualquiera de las características y funcionalidad descritas con respecto a los aparatos como se mencionaron en lo anterior.

45

Breve descripción de las figuras

[0041] Las formas de realización según la presente invención se describirán subsiguientemente haciendo referencia a las figuras adjuntas, en las que:

50

la fig. 1 muestra un diagrama de bloque esquemático de un decodificador de audio multicanal, según una forma de realización de la presente invención;

la fig. 2 muestra un diagrama de bloque esquemático de un codificador de audio multicanal, según una forma de realización de la presente invención;

55

la fig. 3 muestra un diagrama de flujo de un procedimiento para proporcionar al menos dos señales de audio de salida en base a una representación codificada, según una forma de realización de la invención;

la fig. 4 muestra un diagrama de flujo de un procedimiento para proporcionar una representación codificada en base al menos a dos señales de audio de entrada, según una forma de realización de la presente invención;

60

la fig. 5 muestra una representación esquemática de una representación de audio codificada, según una forma de realización de la presente invención;

la fig. 6 muestra un diagrama de bloque esquemático de un decorrelacionador multicanal, según una forma de realización de la presente invención;

la fig. 7 muestra un diagrama de bloque esquemático de un decodificador de audio multicanal, según una forma de realización de la presente invención;

65

la fig. 8 muestra un diagrama de bloque esquemático de un codificador de audio multicanal, según una forma de

realización de la presente invención;

la fig. 9 muestra un diagrama de flujo de un procedimiento para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador, según una forma de realización de la presente invención;

5 la fig. 10 muestra un diagrama de flujo de un procedimiento para proporcionar al menos dos señales de audio de salida en base a una representación codificada, según una forma de realización de la presente invención;

la fig. 11 muestra un diagrama de flujo de un procedimiento para proporcionar una representación codificada en base al menos a dos señales de audio de entrada, según una forma de realización de la presente invención;

10 la fig. 12 muestra una representación esquemática de una representación codificada, según una forma de realización de la presente invención;

la fig. 13 muestra una representación esquemática que proporciona una visión de conjunto de un concepto de mezcla descendente/mezcla ascendente paramétrico basado en MMSE;

la fig. 14 muestra una representación geométrica de un principio de ortogonalidad en un espacio tridimensional;

15 la fig. 15 muestra un diagrama de bloque esquemático de un sistema de reconstrucción paramétrico con decorrelación aplicada a la salida representada, según una forma de realización de la presente invención;

la fig. 16 muestra un diagrama de bloque esquemático de una unidad de decorrelación;

la fig. 17 muestra un diagrama de bloque esquemático de una unidad de decorrelación de complejidad reducida, según una forma de realización de la presente invención;

20 la fig. 18 muestra una representación en forma de tabla de posiciones de altavoces, según una forma de realización de la presente invención;

las figs. 19A a 19G muestran representaciones en forma de tabla de coeficientes de premezclado para $N = 22$ y K entre 5 y 11;

las figs. 20A a 20D muestran representaciones en forma de tabla de coeficientes de premezclado para $N = 10$ y K entre 2 y 5;

25 las figs. 21A a 21C muestran representaciones en forma de tabla de coeficientes de premezclado para $N = 8$ y K entre 2 y 4;

las figs. 21D a 21F muestran representaciones en forma de tabla de coeficientes de premezclado para $N = 7$ y K entre 2 y 4;

30 las figs. 22A y 22B muestran representaciones en forma de tabla de coeficientes de premezclado para $N = 5$ y $K = 2$ o $K = 3$;

la fig. 23 muestra una representación en forma de tabla de coeficientes de premezclado para $N = 2$ y $K = 1$;

la fig. 24 muestra una representación en forma de tabla de grupos de señales de canales;

la fig. 25 muestra una representación sintáctica de parámetros adicionales, que pueden estar incluidos en la sintaxis de SAOCSpecificConfig() o, de forma equivalente, SAOC3DSpecificConfig();

35 la fig. 26 muestra una representación en forma de tabla de diferentes valores para la variable de flujo de bits bsDecorrelationMethod;

la fig. 27 muestra una representación en forma de tabla de un número de decorrelacionadores para diferentes niveles de decorrelación y configuraciones de salida, indicados por la variable de flujo de bits bsDecorrelationLevel;

40 la fig. 28 muestra, en la forma de un diagrama de bloque esquemático, una visión de conjunto de un codificador de audio 3D;

la fig. 29 muestra, en la forma de un diagrama de bloque esquemático, una visión de conjunto de un decodificador de audio 3D; y

la fig. 30 muestra un diagrama de bloque esquemático de una estructura de un convertidor de formato.

45 La fig. 31 muestra un diagrama de bloque esquemático de un procesador de mezcla descendente, según una forma de realización de la presente invención;

la fig. 32 muestra una tabla que representa modos de decodificación para un número diferente de objetos de mezcla descendente de SAOC; y

la fig. 33 muestra una representación de sintaxis de un elemento de flujo de bits "SAOC3DSpecificConfig".

50 Descripción detallada de las formas de realización

1. Decodificador de audio multicanal según la figura 1

55 [0042] La figura 1 muestra un diagrama de bloque esquemático de un decodificador de audio multicanal 100, según una forma de realización de la presente invención.

[0043] El decodificador de audio multicanal 100 está configurado para recibir una representación codificada 110 y para proporcionar, en base a ésta, al menos dos señales de audio de salida 112, 114.

60 [0044] El decodificador de audio multicanal 100 comprende preferentemente un decodificador 120 que está configurado para proporcionar señales de audio decodificadas 122 en base a la representación codificada 110. Además, el decodificador de audio multicanal 100 comprende un renderizador 130, que está configurado para renderizar una pluralidad de señales de audio decodificadas 122, que se obtienen en base a la representación codificada 110 (por ejemplo, por el decodificador 120) en dependencia de uno o más parámetros de representación
65 132, para obtener una pluralidad de señales de audio renderizadas 134, 136. Además, el decodificador de audio

multicanal 100 comprende un decorrelacionador 140, que está configurado para derivar una o más señales de audio decorrelacionadas 142, 144 de las señales de audio renderizadas 134, 136. Además, el decodificador de audio multicanal 100 comprende un combinador 150, que está configurado para combinar las señales de audio renderizadas 134, 136, o una versión en escala de éstas, con una o más señales de audio decorrelacionadas 142, 144 para obtener las señales de audio de salida 112, 114.

[0045] Sin embargo, debería señalarse que puede ser posible una estructura de hardware diferente del decodificador de audio multicanal 100, siempre que estén dadas las funcionalidades descritas más arriba.

10 **[0046]** Con respecto a la funcionalidad del decodificador de audio multicanal 100, debería señalarse que las señales de audio decorrelacionadas 142, 144 se derivan de las señales de audio renderizadas 134, 136, y que las señales de audio decorrelacionadas 142, 144 son combinadas con las señales de audio renderizadas 134, 136 para obtener las señales de audio de salida 112, 114. Derivando las señales de audio decorrelacionadas 142, 144 de las
15 de señales de audio renderizadas 134, 136, se puede lograr un procesamiento particularmente eficaz, ya que el número de señales de audio decorrelacionadas 134, 136 es típicamente independiente del número de señales de audio decodificadas 122 que son introducidas en el renderizador 130. Así, el esfuerzo de decorrelación es típicamente independiente del número de señales de audio decodificadas 122, que mejora la eficacia de implementación. Además, aplicando la decorrelación después de la representación evita la introducción de artefactos, que podría ser causada por el renderizador cuando combina múltiples señales decorrelacionadas en el caso de que la
20 decorrelación sea aplicada antes de la renderización. Además, las características de las señales de audio renderizadas pueden ser consideradas en la decorrelación realizada por el decorrelacionador 140, que típicamente resulta en señales de audio de salida de buena calidad.

[0047] Además, debería señalarse que el decodificador de audio multicanal 100 puede ser suplementado por
25 cualquiera de las características y funcionalidades descritas en esta invención. En particular, debería señalarse que las mejoras individuales como se describen en esta invención pueden ser introducidas en el decodificador de audio multicanal 100 para de este modo mejorar incluso la eficacia del procesamiento y/o la calidad de las señales de audio de salida.

30 2. Codificador de audio multicanal según la figura 2

[0048] La figura 2 muestra un diagrama de bloque esquemático de un codificador de audio multicanal 200, según una forma de realización de la presente invención. El codificador de audio multicanal 200 está configurado para recibir dos o más señales de audio de entrada 210, 212, y para proporcionar, en base a éstas, una
35 representación codificada 214. El codificador de audio multicanal comprende un proporcionador de señales de mezcla descendente 220, que está configurado para proporcionar una o más señales de mezcla descendente 222 en base al menos a dos señales de audio de entrada 210, 212. Además, el codificador de audio multicanal 200 comprende un proporcionador de parámetros 230, que está configurado para proporcionar uno o más parámetros 232 que describen una relación (por ejemplo, una correlación cruzada, una covarianza cruzada, una diferencia de
40 nivel o similar) entre al menos dos señales de audio de entrada 210, 212.

[0049] Además, el codificador de audio multicanal 200 también comprende un proporcionador de parámetros del procedimiento de decorrelación 240, que está configurado para proporcionar un parámetro del procedimiento de decorrelación 242 que describe qué modo de decorrelación de una pluralidad de modos de decorrelación debería
45 usarse en el lado de un decodificador de audio. Una o más señales de mezcla descendente 222, uno o más parámetros 232 y el parámetro del procedimiento de decorrelación 242 están incluidos, por ejemplo, en una forma codificada, en la representación codificada 214.

[0050] Sin embargo, debería señalarse que la estructura de hardware del codificador de audio multicanal 200
50 puede ser diferente, siempre que se cumplan las funcionalidades como se describe en lo anterior. En otras palabras, la distribución de las funcionalidades del codificador de audio multicanal 200 a bloques individuales (por ejemplo, al proporcionador de señales de mezcla descendente 220, al proporcionador de parámetros 230 y al proporcionador de parámetros del procedimiento de decorrelación 240) debería ser considerada solo como un ejemplo.

55 **[0051]** Con respecto a la funcionalidad del codificador de audio multicanal 200, debería señalarse que una o más señales de mezcla descendente 222 y uno o más parámetros 232 se proporcionan de una manera convencional, por ejemplo, como en un codificador de audio multicanal SAOC o en un codificador de audio multicanal USAC. Sin embargo, el parámetro del procedimiento de decorrelación 242, que también es proporcionado por el codificador de audio multicanal 200 y está incluido en la representación codificada 214, se puede usar para
60 adaptar un modo de decorrelación a las señales de audio de entrada 210, 212 o a una calidad de reproducción deseada. Por consiguiente, el modo de decorrelación puede ser adaptado a diferentes tipos de contenido de audio. Por ejemplo, se pueden elegir diferentes modos de decorrelación para tipos de contenidos de audio en el que las señales de audio de entrada 210, 212 están fuertemente correlacionadas y para tipos de contenido de audio en el que las señales de audio de entrada 210, 212 son independientes. Además, diferentes modos de decorrelación
65 pueden ser señalizados, por ejemplo, por el parámetro de modo de decorrelación 242 para tipos de contenidos de

audio en donde es particularmente importante una percepción espacial y para tipos de contenido de audio en donde es menos importante una impresión espacial o incluso es de importancia subordinada (por ejemplo, cuando se compara con una reproducción de canales individuales). Por consiguiente, un decodificador de audio multicanal, que recibe la representación codificada 214, puede ser controlado por el codificador de audio multicanal 200, y puede ser
5 ajustado a un modo de decodificación que lleva consigo un compromiso lo mejor posible entre la complejidad de decodificación y la calidad de reproducción.

[0052] Además, debería señalarse que el codificador de audio multicanal 200 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención. Debería señalarse que las posibles
10 características adicionales y mejoras descritas en esta invención pueden ser agregadas al codificador de audio multicanal 200 de forma individual o en combinación, para mejorar (o aumentar) de este modo el codificador de audio multicanal 200.

3. Procedimiento para proporcionar al menos dos señales de audio de salida según la figura 3

[0053] La figura 3 muestra un diagrama de flujo de un procedimiento 300 para proporcionar al menos dos
15 señales de audio de salida en base a una representación codificada. El procedimiento comprende la renderización 310 de una pluralidad de señales de audio decodificadas, que se obtienen en base a una representación codificada 312, en dependencia de uno o más parámetros de representación, para obtener una pluralidad de señales de audio
20 renderizadas. El procedimiento 300 comprende también la derivación 320 de una o más señales de audio decorrelacionadas de las señales de audio renderizadas. El procedimiento 300 comprende también la combinación 330 de las señales de audio renderizadas, o una versión en escala de éstas, con una o más señales de audio decorrelacionadas, para obtener las señales de audio de salida 332.

[0054] Debería señalarse que el procedimiento 300 se basa en las mismas consideraciones que el
25 decodificador de audio multicanal 100 según la figura 1. Además, debería señalarse que el procedimiento 300 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención (ya sea individualmente o en combinación). Por ejemplo, el procedimiento 300 puede ser suplementado por cualquiera de las características y funcionalidades descritas con respecto a los decodificadores de audio multicanal descritos en
30 esta invención.

4. Procedimiento para proporcionar una representación codificada según la figura 4

[0055] La figura 4 muestra un diagrama de flujo de un procedimiento 400 para proporcionar una
35 representación codificada en base al menos a dos señales de audio de entrada. El procedimiento 400 comprende el suministro 410 de una o más señales de mezcla descendente en base al menos a dos señales de audio de entrada 412. El procedimiento 400 comprende además el suministro 420 de uno o más parámetros que describen una relación entre al menos dos señales de audio de entrada 412 y el suministro 430 de un parámetro del procedimiento de decorrelación que describe qué modo de decorrelación de una pluralidad de modos de decorrelación debería
40 usarse en el lado de un decodificador de audio. Por consiguiente, se proporciona una representación codificada 432, que incluye preferentemente una representación codificada de una o más señales de mezcla descendente, uno o más parámetros que describen una relación entre al menos dos señales de audio de entrada, y el parámetro del procedimiento de decorrelación.

[0056] Debería señalarse que el procedimiento 400 se basa en las mismas consideraciones que las indicadas
45 para el codificador de audio multicanal 200 según la figura 2, de tal modo que las explicaciones dadas más arriba también se aplican.

[0057] Además, debería señalarse que el orden de las etapas 410, 420, 430 puede variarse de forma flexible,
50 y que las etapas 410, 420, 430 también pueden ser realizadas en la medida de lo posible en un ambiente de ejecución para el procedimiento 400. Además, debería señalarse que el procedimiento 400 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención, ya sea individualmente o en combinación. Por ejemplo, el procedimiento 400 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención con respecto a los codificadores multicanal de audio. Sin embargo,
55 también es posible introducir características y funcionalidades que corresponden a las características y funcionalidades de los decodificadores de audio multicanal descritas en esta invención, que reciben la representación codificada 432.

5. Representación de audio codificada según la figura 5

[0058] La figura 5 muestra una representación esquemática de una representación de audio codificada 500
60 según una forma de realización de la presente invención.

[0059] La representación de audio codificada 500 comprende una representación codificada 510 de una
65 señal de mezcla descendente, una representación codificada 520 de uno o más parámetros que describen una

relación entre al menos dos señales de audio. Además, la representación de audio codificada 500 comprende también un parámetro del procedimiento de decorrelación codificado 530 que describe qué modo de decorrelación de una pluralidad de modos de decorrelación debería usarse en el lado de un decodificador de audio. Por consiguiente, la representación de audio codificada permite señalar un modo de decorrelación de un codificador de audio a un decodificador de audio. Por consiguiente, es posible obtener un modo de decorrelación que está bien adaptado a las características del contenido de audio (que se describe, por ejemplo, por la representación codificada 510 de una o más señales de mezcla descendente y por la representación codificada 520 de uno o más parámetros que describen una relación entre al menos dos señales de audio (por ejemplo, al menos dos señales de audio que han sido mezcladas de manera descendente en la representación codificada 510 de una o más señales de mezcla descendente)). Así, la representación de audio codificada 500 permite una representación de un contenido de audio representado por la representación de audio codificada 500 con una impresión espacial auditiva particularmente buena y/o una compensación particularmente buena entre la impresión espacial auditiva y la complejidad de decodificación.

- 15 **[0060]** Además, debería señalarse que la representación codificada 500 puede ser suplementada por cualquiera de las características y funcionalidades descritas con respecto a los codificadores de audio multicanal y los decodificadores de audio multicanal, ya sea individualmente o en combinación.

6. Decorrelacionador multicanal según la figura 6

- 20 **[0061]** La figura 6 muestra un diagrama de bloque esquemático de un decorrelacionador multicanal 600, según una forma de realización de la presente invención.

- [0062]** El decorrelacionador multicanal 600 está configurado para recibir un primer conjunto de señales de entrada del decorrelacionador N 610a a 610n y proporcionar, en base a éstas, un segundo conjunto de señales de salida del decorrelacionador N' 612a a 612n'. En otras palabras, el decorrelacionador multicanal 600 está configurado para proporcionar una pluralidad de señales decorrelacionadas (al menos aproximadamente) 612a a 612n' en base a las señales de entrada del decorrelacionador 610a a 610n.

- 30 **[0063]** El decorrelacionador multicanal 600 comprende un premezclador 620, que está configurado para premezclar el primer conjunto de señales de entrada del decorrelacionador N 610a a 610n en un segundo conjunto de señales de entrada del decorrelacionador K 622a a 622k, en donde K es menor que N (siendo K y N números enteros). El decorrelacionador multicanal 600 comprende también una decorrelación (o núcleo de decorrelacionador) 630, que está configurada para proporcionar un primer conjunto de señales de salida del decorrelacionador K' 632a a 632k' en base al segundo conjunto de señales de entrada del decorrelacionador K 622a a 622k. Además, el decorrelacionador multicanal comprende un post-mezclador 640, que está configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' 632a a 632k' en un segundo conjunto de señales de salida del decorrelacionador N' 612a a 612n', en donde N' es mayor que K' (siendo N' y K' números enteros).

- 40 **[0064]** Sin embargo, debería señalarse que la estructura dada del decorrelacionador multicanal 600 debería ser considerada como un ejemplo solamente, y que no es necesario subdividir el decorrelacionador multicanal 600 en bloques funcionales (por ejemplo, en el premezclador 620, la decorrelación o el núcleo del decorrelacionador 630 y el post-mezclador 640) siempre que se proporcione la funcionalidad descrita en esta invención.

- 45 **[0065]** Con respecto a la funcionalidad del decorrelacionador multicanal 600, debería señalarse además que el concepto de realizar un premezclado, para derivar el segundo conjunto de señales de entrada del decorrelacionador K del primer conjunto de señales de entrada del decorrelacionador N, y de realizar la decorrelación en base al segundo conjunto de señales de entrada del decorrelacionador K (premezcladas o "mezcladas de manera descendente") lleva consigo una reducción de una complejidad cuando se compara con un concepto en el cual se aplica la decorrelación real, por ejemplo, directamente a señales de entrada del decorrelacionador N. Además, el segundo conjunto de señales de salida del decorrelacionador N' (mezclado ascendente) se obtiene en base al primer conjunto de señales de salida del decorrelacionador (original), que son el resultado de la decorrelación real, en base a un post-mezclado, que puede ser realizado por el mezclador ascendente 640. Así, el decorrelacionador multicanal 600 recibe efectivamente (cuando se ve desde fuera) señales de entrada del decorrelacionador N y proporciona, en base a éstas, señales de salida del decorrelacionador N', mientras que el núcleo del decorrelacionador real 630 solo opera en un número de señales más pequeño (a saber señales de entrada del decorrelacionador mezcladas de manera descendente K 622a a 622k del segundo conjunto de señales de entrada del decorrelacionador K). Así, la complejidad del decorrelacionador multicanal 600 puede ser sustancialmente reducida, cuando se compara con los decorrelacionadores convencionales, realizando un mezclado descendente o "premezclado" (que puede ser preferentemente un premezclado lineal sin una funcionalidad de decorrelación) en un lado de entrada de la decorrelación (o núcleo del decorrelacionador) 630 y realizando el mezclado ascendente o "post-mezclado" (por ejemplo, un mezclado de manera ascendente lineal sin una funcionalidad de decorrelación adicional) en base a las señales de salida (originales) 632a a 632k' de la decorrelación (núcleo del decorrelacionador) 630.

[0066] Además, debería señalarse que el decorrelacionador multicanal 600 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención con respecto a la decorrelación multicanal y también con respecto a los decodificadores de audio multicanal. Debería señalarse que las características descritas en esta invención se pueden agregar al decorrelacionador multicanal 600 ya sea individualmente o en combinación, para mejorar o aumentar de este modo el decorrelacionador multicanal 600.

[0067] Debería señalarse que un decorrelacionador multicanal sin reducción de complejidad puede derivarse del decorrelacionador multicanal descrito en lo anterior para $K=N$ (y posiblemente $K'=N'$ o incluso $K=N=K'=N'$).

7. Decodificador de audio multicanal según la figura 7

[0068] La figura 7 muestra un diagrama de bloque esquemático de un decodificador de audio multicanal 700, según una forma de realización de la invención.

[0069] El decodificador de audio multicanal 700 está configurado para recibir una representación codificada 710 y para proporcionar, en base a ésta, al menos dos señales de salida 712, 714. El decodificador de audio multicanal 700 comprende un decorrelacionador multicanal 720, que puede ser sustancialmente idéntico al decorrelacionador multicanal 600 según la figura 6. Además, el decodificador de audio multicanal 700 puede comprender cualquiera de las características y funcionalidades de un decodificador de audio multicanal que son conocidas por el experto en la técnica o que se describen en esta invención con respecto a otros decodificadores de audio multicanal.

[0070] Además, debería señalarse que el decodificador de audio multicanal 700 comprende una eficacia particularmente alta cuando se compara con decodificadores convencionales de audio multicanal, ya que el decodificador de audio multicanal 700 usa el decorrelacionador multicanal de alta eficacia 720.

8. Codificador de audio multicanal según la figura 8

[0071] La figura 8 muestra un diagrama de bloque esquemático de un codificador de audio multicanal 800 según una forma de realización de la presente invención. El codificador de audio multicanal 800 está configurado para recibir al menos dos señales de audio de entrada 810, 812 y para proporcionar, en base a éstas, una representación codificada 814 de un contenido de audio representado por las señales de audio de entrada 810, 812.

[0072] El codificador de audio multicanal 800 comprende un proporcionador de señales de mezcla descendente 820, que está configurado para proporcionar una o más señales de mezcla descendente 822 en base al menos a dos señales de audio de entrada 810, 812. El codificador de audio multicanal 800 comprende también un proporcionador de parámetros 830 que está configurado para proporcionar uno o más parámetros 832 (por ejemplo, parámetros de correlación cruzada o parámetros de covarianza cruzada, o parámetros de correlación interobjetos y/o parámetros de diferencia de nivel de objetos) en base a las señales de audio de entrada 810, 812. Además, el codificador de audio multicanal 800 comprende un parámetro de complejidad de decorrelación 840 que está configurado para proporcionar un parámetro de complejidad de decorrelación 842 que describe una complejidad de una decorrelación que va a ser usada en el lado de un decodificador de audio (que recibe la representación codificada 814). Una o más señales de mezcla descendente 822, uno o más parámetros 832 y el parámetro de complejidad de decorrelación 842 están incluidos en la representación codificada 814, preferentemente de una forma codificada.

[0073] Sin embargo, debería señalarse que la estructura interna del codificador de audio multicanal 800 (por ejemplo, la presencia del proporcionador de señales de mezcla descendente 820, del proporcionador de parámetros 830 y del proporcionador de parámetros de complejidad de decorrelación 840) debería ser considerado como un ejemplo solamente. Diferentes estructuras son posibles, siempre que se logre la funcionalidad descrita en esta invención.

[0074] Con respecto a la funcionalidad del codificador de audio multicanal 800, debería señalarse que el codificador multicanal proporciona una representación codificada 814, en donde una o más señales de mezcla descendente 822 y uno o más parámetros 832 pueden ser similares a, o iguales a, señales de mezcla descendente y parámetros proporcionados por codificadores de audio convencionales (como, por ejemplo, codificadores de audio SAOC convencionales o codificadores de audio USAC). Sin embargo, el codificador de audio multicanal 800 también está configurado para proporcionar el parámetro de complejidad de decorrelación 842, que permite determinar una complejidad de decorrelación que se aplica en el lado de un decodificador de audio. Por consiguiente, la complejidad de decorrelación puede ser adaptada a un contenido de audio que es codificado actualmente. Por ejemplo, es posible señalar una complejidad de decorrelación deseada, que corresponde a una calidad de audio alcanzable, en dependencia de un conocimiento del lado del codificador con respecto a las características de las señales de audio de entrada. Por ejemplo, si se encuentra que las características espaciales son importantes para una señal de audio, se puede señalar una mayor complejidad de decorrelación, usando el parámetro de complejidad de decorrelación

842, cuando se compara con un caso en el cual las características espaciales no son tan importantes. Alternativamente, el uso de una alta complejidad de decorrelación puede ser señalizado mediante el uso del parámetro de complejidad de decorrelación 842, si se encuentra que un pasaje del contenido de audio o todo el contenido de audio es tal que se requiere una alta complejidad de decorrelación en un lado de un decodificador de audio por otras razones.

[0075] Para resumir, el codificador de audio multicanal 800 proporciona la posibilidad de controlar un decodificador de audio multicanal, para usar una complejidad de decorrelación que está adaptada a las características de señales o las características de reproducción deseadas que pueden ser ajustadas por el codificador de audio multicanal 800.

[0076] Además, debería señalarse que el codificador de audio multicanal 800 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención con respecto a un codificador de audio multicanal, ya sea individualmente o en combinación. Por ejemplo, algunas o todas las características descritas en esta invención con respecto a codificadores de audio multicanal pueden ser agregadas al codificador de audio multicanal 800. Además, el codificador de audio multicanal 800 puede ser adaptado para cooperación con los decodificadores de audio multicanal descritos en esta invención.

9. Procedimiento para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador, según la figura 9

[0077] La figura 9 muestra un diagrama de flujo de un procedimiento 900 para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador.

[0078] El procedimiento 900 comprende el premezclado 910 de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador k, en donde K es menor que N. El procedimiento 900 comprende también el suministro 920 de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K. Por ejemplo, el primer conjunto de señales de salida del decorrelacionador K' puede ser proporcionado en base al segundo conjunto de señales de entrada del decorrelacionador K mediante el uso de una decorrelación, lo que se puede realizar, por ejemplo, mediante el uso de un núcleo de decorrelacionador o mediante el uso de un algoritmo de decorrelación. El procedimiento 900 comprende además el post-mezclado 930 del primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto a señales de salida del decorrelacionador N', en donde N' es mayor que K' (siendo N' y K' números enteros). Por consiguiente, el segundo conjunto de señales de salida del decorrelacionador N', que son la salida del procedimiento 900, puede ser proporcionado en base al primer conjunto de señales de entrada del decorrelacionador N, que son la entrada al procedimiento 900.

[0079] Debería señalarse que el procedimiento 900 se basa en las mismas consideraciones que las indicadas para el decorrelacionador multicanal descrito en lo anterior. Además, debería señalarse que el procedimiento 900 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención con respecto al decorrelacionador multicanal (y también con respecto al codificador de audio multicanal, si es aplicable), ya sea individualmente o en combinación.

10. Procedimiento para proporcionar al menos dos señales de audio de salida en base a una representación codificada, según la figura 10

[0080] La figura 10 muestra un diagrama de flujo de un procedimiento 1000 para proporcionar al menos dos señales de audio de salida en base a una representación codificada.

[0081] El procedimiento 1000 comprende el suministro 1010 de al menos dos señales de audio de salida 1014, 1016 en base a una representación codificada 1012. El procedimiento 1000 comprende el suministro 1020 de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador según el procedimiento 900 según la figura 9.

[0082] Debería señalarse que el procedimiento 1000 se basa en las mismas consideraciones que las indicadas para el decodificador de audio multicanal 700 según la figura 7.

[0083] Además, debería señalarse que el procedimiento 1000 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención con respecto a los decodificadores multicanal, ya sea individualmente o en combinación.

11. Procedimiento para proporcionar una representación codificada en base al menos a dos señales de audio de entrada, según la figura 11

[0084] La figura 11 muestra un diagrama de flujo de un procedimiento 1100 para proporcionar una

representación codificada en base al menos a dos señales de audio de entrada.

[0085] El procedimiento 1100 comprende el suministro 1110 de una o más señales de mezcla descendente en base al menos a dos señales de audio de entrada 1112, 1114. El procedimiento 1100 comprende también el suministro 1120 de uno o más parámetros que describen una relación entre al menos dos señales de audio de entrada 1112, 1114. Además, el procedimiento 1100 comprende el suministro 1130 de un parámetro de complejidad de decorrelación que describe una complejidad de una decorrelación que va a ser usada en el lado de un decodificador de audio. Por consiguiente, se proporciona una representación codificada 1132 en base al menos a dos señales de audio de entrada 1112, 1114, en el que la representación codificada comprende típicamente una o más señales de mezcla descendente, uno o más parámetros que describen una relación entre al menos dos señales de audio de entrada y el parámetro de complejidad de decorrelación en una forma codificada.

[0086] Debería señalarse que las etapas 1110, 1120, 1130 pueden ser realizadas en paralelo o en un orden diferente en algunas formas de realización según la invención. Además, debería señalarse que el procedimiento 1100 se basa en las mismas consideraciones que las indicadas para el codificador de audio multicanal 800 según la figura 8, y que el procedimiento 1100 puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención con respecto al codificador de audio multicanal, o bien en combinación o individualmente. Además, debería señalarse que el procedimiento 1100 puede ser adaptado para adaptar el decodificador de audio multicanal y el procedimiento para proporcionar al menos dos señales de audio de salida descritas en esta invención.

12. Representación de audio codificada según la figura 12

[0087] La figura 12 muestra una representación esquemática de una representación de audio codificada, según una forma de realización de la presente invención. La representación de audio codificada 1200 comprende una representación codificada 1210 de una señal de mezcla descendente, una representación codificada 1220 de uno o más parámetros que describen una relación entre al menos dos señales de audio de entrada, y un parámetro de complejidad de decorrelación codificado 1230 que describe una complejidad de una decorrelación que va a ser usada en el lado de un decodificador de audio. Por consiguiente, la representación de audio codificada 1200 permite ajustar la complejidad de decorrelación usada por un decodificador de audio multicanal, que lleva consigo una eficacia de decodificación mejorada, y una posible calidad de audio mejorada, o una compensación mejorada entre la eficacia de codificación y la calidad de audio. Además, debería señalarse que la representación de audio codificada 1200 puede ser proporcionada por el codificador de audio multicanal como se describe en esta invención, y puede ser usada por el decodificador de audio multicanal como se describe en esta invención. Por consiguiente, la representación de audio codificada 1200 puede ser suplementada por cualquiera de las características descritas con respecto a los codificadores de audio multicanal y con respecto a los decodificadores de audio multicanal.

13. Notación y consideraciones subyacentes

[0088] Recientemente se han propuesto técnicas paramétricas para la transmisión eficaz de velocidades de transmisión de bits/almacenamiento de escenas de audio que contienen múltiples objetos de audio en el campo de la codificación de audio (véanse, por ejemplo, las referencias [BCC], [JSC], [SAOC], [SAOC1], [SAOC2]) y separación de fuentes informada (véanse, por ejemplo, las referencias [ISS1], [ISS2], [ISS3], [ISS4], [ISS5], [ISS6]). Estas técnicas tienen como objetivo reconstruir una escena de audio de salida deseada o un objeto de fuente de audio en base a información lateral adicional que describe la escena de audio transmitida/almacenada y/u objetos de fuente en la escena de audio. Esta reconstrucción se produce en el decodificador mediante el uso de un esquema de separación de fuentes informadas paramétrica. Además, también se hace referencia al concepto así llamado "sonido envolvente de MPEG", que se describe, por ejemplo, en la norma internacional ISO/IEC 23003-1:2007. Además, también se hace referencia a la así llamada "codificación de objetos de audio espacial" que se describe en la norma internacional ISO/IEC 23003-2:2010. Además, se hace referencia al concepto así llamado de "codificación de audio y voz unificada", que se describe en la norma internacional ISO/IEC 23003-3:2012. Los conceptos de estas normas se pueden usar en formas de realización según la invención, por ejemplo, en los codificadores de audio multicanal mencionados en esta invención y los decodificadores de audio multicanal mencionados en esta invención, en donde se pueden requerir algunas adaptaciones.

[0089] A continuación se describirá alguna información de antecedentes. En particular, se proporcionará una vista general de esquemas de separación paramétrica, mediante el uso del ejemplo de la tecnología de codificación de objetos de audio espacial de (SAOC) de MPEG (véase, por ejemplo, la referencia [SAOC]). Se consideran las propiedades matemáticas de este procedimiento.

13.1. Notación y definiciones

[0090] La siguiente notación matemática se aplica al documento actual:

N_{Objetos} número de señales de objetos de audio

- N_{DmxCh} número de canales de mezcla descendente (DwmixCh) (procesados)
- $N_{UpmixCh}$ número de canales de mezcla ascendente (UpmixCh) (salida)
- $N_{Muestras}$ número de muestras de datos procesados
- $\mathbf{D}$ matriz de mezcla descendente, tamaño $N_{DmxCh} \times N_{Objetos}$
- 5 $\mathbf{X}$ señal de objeto de audio de entrada, tamaño $N_{Objetos} \times N_{Muestras}$
- $\mathbf{E}_X$ matriz de covarianza del objeto, tamaño $N_{Objetos} \times N_{Objetos}$ definida como $\mathbf{E}_X = \mathbf{X}\mathbf{X}^H$
- $\mathbf{Y}$ señal de audio de mezcla descendente, tamaño $N_{DmxCh} \times N_{Muestras}$ definida como $\mathbf{Y} = \mathbf{D}\mathbf{X}$
- $\mathbf{E}_Y$ matriz de covarianza de las señales de mezcla descendente, tamaño $N_{DmxCh} \times N_{DmxCh}$ definida como $\mathbf{E}_Y = \mathbf{Y}\mathbf{Y}^H$
- 10 $\mathbf{G}$ matriz de estimación de fuente paramétrica, tamaño $N_{Objetos} \times N_{DmxCh}$ que se aproxima a $\mathbf{E}_X \mathbf{D}^H (\mathbf{D} \mathbf{E}_X \mathbf{D}^H)^{-1}$
- $\hat{\mathbf{X}}$ señal de objeto paramétricamente reconstruida, tamaño $N_{Objetos} \times N_{Muestras}$ que se aproxima a $\mathbf{X}$ y definida como $\hat{\mathbf{X}} = \mathbf{G}\mathbf{Y}$
- $\mathbf{R}$ matriz de representación (especificada en el lado del decodificador), tamaño $N_{UpmixCh} \times N_{Objetos}$
- 15 $\mathbf{Z}$ señal de escena de salida visualizada ideal, tamaño $N_{UpmixCh} \times N_{Muestras}$ definida como $\mathbf{Z} = \mathbf{R}\hat{\mathbf{X}}$
- $\hat{\mathbf{Z}}$ salida paramétrica visualizada, tamaño $N_{UpmixCh} \times N_{Muestras}$ definida como $\hat{\mathbf{Z}} = \mathbf{R}\hat{\mathbf{X}}$
- $\mathbf{C}$ matriz de covarianza de salida ideal, tamaño $N_{UpmixCh} \times N_{UpmixCh}$ definida como $\mathbf{C} = \mathbf{R}\mathbf{E}_X \mathbf{R}^H$
- $\mathbf{W}$ salidas del decorrelacionador, tamaño $N_{UpmixCh} \times N_{Muestras}$
- $\mathbf{S}$ señal combinada, tamaño $2N_{UpmixCh} \times N_{Muestras}$
- 20 $\mathbf{E}_S$ matriz de covarianza de señal combinada, tamaño $2N_{UpmixCh} \times 2N_{UpmixCh}$ definida como $\mathbf{E}_S = \mathbf{S}\mathbf{S}^H$ $\mathbf{S} = \begin{bmatrix} \hat{\mathbf{Z}} \\ \mathbf{W} \end{bmatrix}$
- $\tilde{\mathbf{Z}}$ salida final, tamaño $N_{UpmixCh} \times N_{Muestras}$
- $(\cdot)^H$ operador autoadjunto (hermítico) que representa la trasposición conjugada compleja de $(\cdot)$. También se puede usar la notación $(\cdot)^*$.
- 25 $F_{decorr}(\cdot)$ función del decorrelacionador
- ϵ es una constante aditiva para evitar la división por cero
- $\mathbf{H} = \text{matdiag}(\mathbf{M})$ es una matriz que contiene los elementos de la diagonal principal de la matriz $\mathbf{M}$ en la diagonal principal y valores cero en las posiciones fuera de la diagonal.
- 30 **[0091]** Sin perder la generalidad, para mejorar la legibilidad de las ecuaciones, para todas las variables introducidas los índices que denotan dependencia de tiempo y frecuencia se omiten en este documento.

13.2. Sistemas de separación paramétrica

- 35 **[0092]** Los sistemas de separación paramétrica generales tienen como objetivo calcular un número de fuentes de audio de una mezcla de señales (mezcla descendente) mediante el uso de información de parámetros auxiliar (como, por ejemplo, valores de correlación intercanales, valores de diferencia de nivel intercanales, valores de correlación interobjetos y/o información de diferencia de nivel de objetos). Una solución típica de esta tarea se basa en la aplicación de los algoritmos de cálculo del mínimo error cuadrático medio (MMSE). La tecnología SAOC
- 40 es un ejemplo de tales sistemas de codificación/decodificación de audio paramétricos.

[0093] La figura 13 muestra el principio general de la arquitectura codificador/decodificador de SAOC. En otras palabras, la figura 13 muestra, en la forma de un diagrama de bloque esquemático, una vista general del concepto de mezcla descendente/mezcla ascendente paramétrica basada en MMSE.

[0094] Un codificador 1310 recibe una pluralidad de señales de objetos 1312a, 1312b a 1312n. Además, el codificador 1310 recibe también parámetros de mezclado D , 1314, que pueden ser, por ejemplo, parámetros de mezcla descendente. El codificador 1310 proporciona, en base a estos, una o más señales de mezcla descendente 1316a, 1316b, y así sucesivamente. Además, el codificador proporciona una información lateral 1318. Una o más
5 señales de mezcla descendente y la información lateral pueden ser proporcionadas, por ejemplo, en una forma codificada.

[0095] El codificador 1310 comprende un mezclador 1320, que está configurado típicamente para recibir las señales de objetos 1312a a 1312n y para combinar (por ejemplo mezclar de manera descendente) las señales de
10 objetos 1312a a 1312n en una o más señales de mezcla descendente 1316a, 1316b en dependencia de los parámetros de mezclado 1314. Además, el codificador comprende un estimador de información lateral 1330, que está configurado para derivar la información lateral 1318 de las señales de objetos 1312a a 1312n. Por ejemplo, el estimador de información lateral 1330 puede ser configurado para derivar la información lateral 1318 de tal modo que la información lateral describe una relación entre señales de objetos, por ejemplo, una correlación cruzada entre
15 las señales de objetos (que puede ser denominada "correlación interobjetos" IOC) y/o una información que describe diferencias de nivel entre las señales de objetos (que pueden ser denominadas "información de diferencia de nivel de objetos" OLD).

[0096] Una o más señales de mezcla descendente 1316a, 1316b y la información lateral 1318 pueden ser
20 almacenadas y/o transmitidas a un decodificador 1350, que es indicado con el número de referencia 1340.

[0097] El decodificador 1350 recibe una o más señales de objeto 1316a, 1316b y la información lateral 1318 (por ejemplo, en una forma codificada) y proporciona, en base a éstas, una pluralidad de señales de audio de salida 1352a a 1352n. El decodificador 1350 también puede recibir una información de interacción del usuario 1354, que
25 puede comprender uno o más parámetros de renderización R (que pueden definir una matriz de representación). El decodificador 1350 comprende un separador de objetos paramétrico 1360, un procesador de información lateral 1370 y un renderizador 1380. El procesador de información lateral 1370 recibe la información lateral 1318 y proporciona, en base a ésta, una información de control 1372 para el separador de objetos paramétrico 1360. El separador de objetos paramétrico 1360 proporciona una pluralidad de señales de objetos 1362a a 1362n en base a
30 las señales de mezcla descendente 1360a, 1360b y la información de control 1372, que se deriva de la información lateral 1318 por el procesador de información lateral 1370. Por ejemplo, el separador de objetos puede realizar una decodificación de las señales de mezcla descendente codificadas y una separación de objetos. El renderizador 1380 representa las señales de objetos reconstruidas 1362a a 1362n, para obtener de este modo las señales de audio de salida 1352a a 1352n.

[0098] A continuación, se comentará la funcionalidad del concepto de mezcla descendente/mezcla ascendente de parámetros basado en MMSE.

[0099] El procesamiento general de mezcla descendente/mezcla ascendente paramétrica es llevado a cabo
40 de una manera selectiva de tiempo/frecuencia y puede ser descrito como una secuencia de las siguientes etapas:

- El "codificador" 1310 se proporciona con "objetos de audio" de entrada X y "parámetros de mezclado" D . El "mezclador" 1320 mezcla descendente los "objetos de audio" X en un número de "señales de mezcla descendente" Y mediante el uso de "parámetros de mezclado" D (por ejemplo, ganancias de mezcla descendente). El "estimador de información lateral" extrae la información lateral 1318 que describe características de los "objetos de audio" de
45 entrada X (por ejemplo, propiedades de covarianza).

- Las "señales de mezcla descendente" Y y la información lateral son transmitidas o almacenadas. Estas señales de audio de mezcla descendente pueden ser comprimidas adicionalmente mediante el uso de codificadores de audio
50 (tales como MPEG-1/2 Capas II o III, MPEG-2/4 Codificación avanzada de audio (AAC), MPEG Codificación unificada de voz y audio (USAC), etc.). La información lateral también puede ser representada y codificada eficazmente (por ejemplo, como relaciones codificadas sin pérdida de las potencias de objetos y los coeficientes de correlación de objetos).

- El "decodificador" 1350 restaura los "objetos de audio" originales de las "señales de mezcla descendente" decodificadas mediante el uso de la información lateral transmitida 1318. El "procesador de información lateral" 1370 estima los coeficientes de no mezclado 1372 que van a ser aplicados en las "señales de mezcla descendente" dentro del "separador de objetos paramétrico" 1360 para obtener la reconstrucción de objetos paramétrica de X . Los "objetos de audio" reconstruidos 1362a a 1362n son renderizados a una escena objetivo (multicanal), representada
60 por los canales de salida $\hat{Z}$, mediante la aplicación de "parámetros de representación" R , 1354.

[0100] Además, debería señalarse que las funcionalidades descritas con respecto al codificador 1310 y al decodificador 1350 se pueden usar en los otros codificadores de audio y decodificadores de audio descritos también en esta invención.

13.3. Principio de ortogonalidad de la estimación del mínimo error cuadrático medio

[0101] El principio de ortogonalidad es una propiedad principal de los estimadores del MMSE. Considerar dos espacios de Hilbert W y V , con V generado por un conjunto de vectores y_i , y un vector $x \in W$. Si se desea hallar un estimado $\hat{x} \in V$ que se aproximará a x como una combinación lineal de los vectores $y_i \in V$, mientras minimiza el error cuadrático medio, entonces el vector de error será ortogonal en el espacio generado por los vectores y_i :

$$(x - \hat{x}) y^H = 0,$$

10 **[0102]** Como consecuencia, el error de estimación y el estimado propiamente dicho son ortogonales:

$$(x - \hat{x}) \hat{x}^H = 0.$$

[0103] Geométricamente se podría visualizar esto por los ejemplos mostrados en la figura 14.

15 **[0104]** La figura 14 muestra una representación geométrica del principio de ortogonalidad en el espacio tridimensional. Como se puede ver, un espacio vectorial es generado por los vectores y_1, y_2 . Un vector x es igual a una suma de un vector $\hat{x}$ y un vector de diferencia (o vector de error) e . Como se puede ver, el vector de error y es ortogonal con respecto al espacio vectorial (o plano) V generado por los vectores y_1 e y_2 . Por consiguiente, el vector $\hat{x}$ puede ser considerado como una mejor aproximación de x dentro del espacio vectorial V .

20

13.4. Error de reconstrucción paramétrica

[0105] Definiendo una matriz que comprende señales N : X y denotando el error de estimación con X_{Error} , se pueden formular las siguientes identidades. La señal original puede ser representada como una suma de la reconstrucción paramétrica $\hat{X}$ y el error de reconstrucción X_{Error} como:

$$X = \hat{X} + X_{Error}$$

[0106] Debido al principio de ortogonalidad, la matriz de covarianza de las señales originales $E_X = XX^H$ puede ser formulada como una suma de la matriz de covarianza de las señales reconstruidas $\hat{X}\hat{X}^H$ y la matriz de covarianza de los errores de estimación $X_{Error}X_{Error}^H$ como:

$$\begin{aligned} E_X = XX^H &= (\hat{X} + X_{Error})(\hat{X} + X_{Error})^H = \hat{X}\hat{X}^H + X_{Error}X_{Error}^H + \hat{X}X_{Error}^H + X_{Error}\hat{X}^H = \\ &= \hat{X}\hat{X}^H + X_{Error}X_{Error}^H. \end{aligned}$$

35 **[0107]** Cuando los objetos de entrada X no están en el espacio generado por los canales de mezcla descendente (por ejemplo, el número de canales de mezcla descendente es menor que el número de señales de entrada) y los objetos de entrada no pueden ser representados como combinaciones lineales de los canales de mezcla descendente, los algoritmos basados en el MMSE introducen la inexactitud de reconstrucción $X_{Error}X_{Error}^H$.

40 13.5. Correlación interobjetos

[0108] En el sistema auditivo, la covarianza cruzada (coherencia/correlación) está estrechamente relacionada con la percepción de envolvente, de estar rodeado por el sonido, y con el ancho percibido de una fuente de sonido. Por ejemplo en los sistemas basados en SAOC los parámetros de correlación interobjetos (IOC) se usan para la

45 caracterización de esta propiedad:

$$IOC(i, j) = \frac{E_X(i, j)}{\sqrt{E_X(i, i)E_X(j, j)}}$$

[0109] Consideremos un ejemplo de reproducción de una fuente de sonido mediante el uso de dos señales de audio. Si el valor de IOC es cercano a uno, el sonido es percibido como una fuente puntual bien localizada. Si el valor de IOC es cercano a cero, el ancho percibido de la fuente de sonido aumenta y para casos extremos incluso puede ser percibido como dos fuentes distintas [Blauert, Capítulo 3].

13.6. Compensación de la inexactitud de reconstrucción

[0110] En el caso de una reconstrucción paramétrica imperfecta, la señal de salida puede presentar una energía menor en comparación con los objetos originales. El error en los elementos diagonales de la matriz de covarianza puede resultar en diferencias de nivel audibles y error en los elementos fuera de la diagonal en una imagen de sonido espacial distorsionada (comparada con la salida de referencia ideal). El procedimiento propuesto tiene el propósito de resolver este problema.

[0111] En el sonido envolvente de MPEG (MPS), por ejemplo, este tema es tratado solo para algunos escenarios de procesamiento basados en canales específicos, a saber, para configuraciones de salida estática limitada y de mezcla descendente mono/estéreo (por ejemplo, mono, estéreo, 5.1, 7.1, etc.). En tecnologías orientadas a objetos, como SAOC, que también usa la mezcla descendente mono/estéreo este problema es tratado aplicando el renderizado post-procesamiento de MPS para la configuración de salida 5.1 solamente.

[0112] Las soluciones existentes están limitadas a configuraciones de salida estándar y a un número fijo de canales de entrada/salida. A saber, son entendidos como una aplicación consecuente de varios bloques implementando solo procedimientos de decorrelación de canales "mono a estéreo" (o "estéreo a tres").

[0113] Por lo tanto, se desea una solución general (por ejemplo, procedimiento de corrección de propiedades de correlación y nivel de energía) para la compensación de la inexactitud de la reconstrucción paramétrica, que puede ser aplicada para un número flexible de canales de mezcla descendente/salida y ajustes de configuraciones de salida arbitrarios.

13.7. Conclusiones

[0114] Para concluir, se ha proporcionado una vista general de la notación. Además, se ha descrito un sistema de separación paramétrica en el cual se basan las formas de realización según la invención. Además, se ha indicado que el principio de ortogonalidad se aplica a una estimación del mínimo error cuadrático medio. Además, se ha proporcionado una ecuación para la computación de una matriz de covarianza EX que se aplica en la presencia de un error de reconstrucción X_{error} . Además, se ha proporcionado la relación entre los así llamados valores de correlación interobjetos y los elementos de una matriz de covarianza EX, que puede ser aplicada, por ejemplo, en formas de realización según la invención para derivar características de covarianza deseadas (o características de correlación) de los valores de correlación interobjetos (que pueden ser incluidos en la información lateral paramétrica), y posiblemente forman las diferencias de nivel de los objetos. Además, se ha indicado que las características de las señales de objetos reconstruidas pueden diferir de las características deseadas debido a una reconstrucción imperfecta. Además, se ha indicado que las soluciones existentes para tratar con el problema están limitadas a algunas configuraciones de salida específicas y se basan en una combinación específica de bloques estándar, que hace que las soluciones convencionales sean inflexibles.

14. Forma de realización de acuerdo con la figura 15

14.1. Vista general del concepto

[0115] Las formas de realización según la invención extienden los procedimientos de reconstrucción paramétrica de MMSE usados en los esquemas de separación de audio paramétricos con una solución de decorrelación para un número arbitrario de canales de mezcla descendente/mezcla ascendente. Las formas de realización según la invención, como, por ejemplo, el aparato de la invención y el procedimiento de la invención, pueden compensar la pérdida de energía durante una reconstrucción paramétrica y restaurar las propiedades de correlación de los objetos estimados.

[0116] La figura 15 proporciona una vista general del concepto paramétrico de mezcla descendente/mezcla ascendente con un camino de decorrelación integrado. En otras palabras, la figura 15 muestra, en la forma de un diagrama de bloque esquemático, un sistema de reconstrucción paramétrica con decorrelación aplicada en la salida renderizada.

[0117] El sistema según la figura 15 comprende un codificador 1510, que es sustancialmente idéntico al codificador 1310 según la figura 13. El codificador 1510 recibe una pluralidad de señales de objetos 1512a a 1512n, y proporciona en base a éstas, una o más señales de mezcla descendente 1516a, 1516b, así como una información lateral 1518. Las señales de mezcla descendente 1516a, 1516b pueden ser sustancialmente idénticas a las señales de mezcla descendente 1316a, 1316b y pueden ser designadas con Y. La información lateral 1518 puede ser sustancialmente idéntica a la información lateral 1318. Sin embargo, la información lateral puede comprender, por ejemplo, un parámetro de modo de decorrelación o un parámetro de procedimiento de decorrelación, o un parámetro de complejidad de decorrelación. Además, el codificador 1510 puede recibir parámetros de mezclado 1514.

[0118] El sistema de reconstrucción paramétrica comprende también una transmisión y/o un almacenamiento

de una o más señales de mezcla descendente 1516a, 1516b y de la información lateral 1518, en el que la transmisión y/o el almacenamiento es designado con 1540, y en donde una o más señales de mezcla descendente 1516a, 1516b y la información lateral 1518 (que puede incluir la información lateral paramétrica) pueden ser codificadas.

5

[0119] Además, el sistema de reconstrucción paramétrica según la figura 15 comprende un decodificador 1550, que está configurado para recibir una o más señales 1516a, 1516b de mezcla descendente (posiblemente codificadas) transmitidas o almacenadas y la información lateral 1518 (posiblemente codificada) transmitida o almacenada y proporcionar, con base en éstas, señales 1552a a 1552n de salida de audio. El decodificador 1550 (que puede ser considerado como un decodificador de audio multicanal) comprende un separador 1560 de objetos paramétrico y un procesador 1570 de información lateral. Además, el decodificador 1550 comprende un renderizador 1580, un decorrelacionador 1590 y un mezclador 1598.

10

[0120] El separador de objetos paramétrico 1560 está configurado para recibir una o más señales 1516a, 1516b de mezcla descendente y una información de control 1572, que es proporcionada por el procesador 1570 de información lateral con base en la información lateral 1518, y para proporcionar, con base en ésta, señales 1562a a 1562n de objetos, que también están designadas con $\hat{x}$, y que pueden ser consideradas como señales de audio decodificadas. La información 1572 de control puede comprender, por ejemplo, coeficientes de desmezclado que van a ser aplicados a las señales de mezcla descendente (por ejemplo, las señales de mezcla descendente decodificadas derivadas de las señales 1516a, 1516b de mezcla descendente codificadas) dentro del separador de objetos paramétrico para obtener señales de objetos reconstruidas (por ejemplo, las señales 1562a a 1562n de audio decodificadas). El renderizador 1580 renderiza las señales 1562a a 1562n de audio decodificadas (que pueden ser señales de objetos reconstruidas, y que pueden corresponder, por ejemplo, a las señales 1512a a 1512n de objetos de entrada), para obtener de este modo una pluralidad de señales 1582a a 1582n de audio renderizadas. Por ejemplo, el renderizador 1580 puede considerar la renderización de los parámetros R, que pueden ser proporcionados, por ejemplo, por la interacción del usuario y que pueden definir, por ejemplo, una matriz de renderizado. Sin embargo, alternativamente, los parámetros de renderizado pueden ser tomados de la representación codificada (que puede incluir las señales 1516a, 1516b de mezcla descendente codificadas y la información 1518 lateral codificada).

30

[0121] El decorrelacionador 1590 está configurado para recibir las señales 1582a a 1582n de audio renderizadas y para proporcionar, con base en éstas, señales 1592a a 1592n de audio decorrelacionadas, que también son designadas con W . El mezclador 1598 recibe las señales 1582a a 1582n de audio renderizadas y las señales 1592a a 1592n de audio decorrelacionadas, y combina las señales 1582a a 1582n de audio renderizadas y las señales 1592a a 1592n de audio decorrelacionadas, para obtener de este modo las señales 1552a a 1552n de audio de salida. El mezclador 1598 también puede usar información de control 1574 que es derivada por el procesador 1570 de información lateral de la información lateral codificada 1518, como se describirá más adelante.

35

14.2. Función del decorrelacionador

40

[0122] A continuación se describirán algunos detalles con respecto al decorrelacionador 1590. Sin embargo, debería señalarse que se pueden usar diferentes conceptos de decorrelacionador, algunos de los cuales se describirán más abajo.

45

[0123] En una forma de realización, la función del decorrelacionador $w = F_{decorr}(\hat{z})$ proporciona una señal de salida w que es ortogonal a la señal de entrada $\hat{z}$ ($E\{w\hat{z}^H\} = 0$). La señal de salida w tiene propiedades envolventes espectrales y temporales iguales (a la señal de entrada $\hat{z}$) (o al menos propiedades similares). Además, la señal w es percibida de forma similar y tiene la misma calidad subjetiva (o similar) que la señal de entrada $\hat{z}$ (véase, por ejemplo, [SAOC2]).

50

[0124] En el caso de múltiples señales de entrada, es beneficioso si la función de decorrelación produce múltiples salidas que son mutuamente ortogonales (es decir, $W_i = F_{decorr}(\hat{Z}_i)$, de tal modo que $W_i \hat{Z}_j^H = 0$ para todos i y j , y $W_i W_j^H = 0$ para $i \neq j$).

55

[0125] La especificación exacta para la implementación de la función de decorrelacionador está fuera del alcance de esta descripción. Por ejemplo, el banco de varios decorrelacionadores basados en filtros de respuesta infinita al impulso (IIR) especificados en el estándar de sonido envolvente de MPEG puede ser utilizado para propósitos de decorrelación [MPS].

60

[0126] Los decorrelacionadores genéricos presentados en esta descripción son supuestos como ideales. Esto implica que (además de los requisitos perceptuales) la salida de cada decorrelacionador es ortogonal en su

entrada y en la salida de todos los otros decorrelacionadores. Por lo tanto, para la entrada dada $\hat{\mathbf{z}}$ con covarianza $\mathbf{E}_{\hat{\mathbf{z}}} = \hat{\mathbf{z}}\hat{\mathbf{z}}^H$ y salida $\mathbf{w} = F_{decorr}(\hat{\mathbf{z}})$ valen las siguientes propiedades de matrices de covarianza:

$$\mathbf{E}_w(i, i) = \mathbf{E}_{\hat{\mathbf{z}}}(i, i), \mathbf{E}_w(i, j) = 0, \text{ for } i \neq j, \hat{\mathbf{z}}\mathbf{w}^H = \mathbf{w}\hat{\mathbf{z}}^H = \mathbf{0}.$$

5

[0127] De estas relaciones, se deduce que:

$$(\hat{\mathbf{z}} + \mathbf{w})(\hat{\mathbf{z}} + \mathbf{w})^H = \mathbf{E}_{\hat{\mathbf{z}}} + \hat{\mathbf{z}}\mathbf{w}^H + \mathbf{w}\hat{\mathbf{z}}^H + \mathbf{E}_w = \mathbf{E}_{\hat{\mathbf{z}}} + \mathbf{E}_w$$

10 [0128] La salida del decorrelacionador $\mathbf{w}$ puede ser usada para compensar la inexactitud de predicción en un estimador de MMSE (recordando que el error de predicción es ortogonal con respecto a las señales predichas) mediante el uso de las señales predichas como las entradas.

[0129] Se debería señalar además que los errores de predicción no son en el caso general ortogonales entre sí. Así, un objetivo del concepto de la invención (por ejemplo, el procedimiento) es crear una mezcla de la señal "seca" (es decir, la entrada del decorrelacionador) (por ejemplo, señales 1582a a 1582n de audio renderizadas) y una señal "húmeda" (es decir, salida del decorrelacionador) (por ejemplo, señales 1592a a 1592n de audio decorrelacionadas), de tal modo que la matriz de covarianza de la mezcla resultante (por ejemplo, señales 1552a a 1552n de audio de salida) se vuelve similar a la matriz de covarianza de la salida deseada.

20

[0130] Además, debería señalarse que se puede usar una reducción de complejidad para la unidad de decorrelación, que será descrita en detalle más abajo, y que puede llevar consigo algunas imperfecciones de la señal decorrelacionada, que pueden ser, sin embargo, aceptables.

25 14.3. Corrección de la covarianza de salida mediante el uso de señales decorrelacionadas

[0131] A continuación, se describirá un concepto para ajustar las características de covarianza de las señales 1552a a 1552n de audio de salida para obtener una impresión auditiva razonablemente buena.

30 [0132] El procedimiento propuesto para la corrección del error de covarianza de salida compone la señal de salida $\tilde{\mathbf{z}}$ (por ejemplo, las señales 1552a a 1552n de audio de salida) como una suma ponderada de la señal paraméricamente reconstruida $\hat{\mathbf{z}}$ (por ejemplo, las señales 1582a a 1582n de audio renderizadas) y su parte decorrelacionada $\mathbf{w}$. Esta suma puede ser representada del siguiente modo:

$$\tilde{\mathbf{z}} = \mathbf{P}\hat{\mathbf{z}} + \mathbf{M}\mathbf{w}.$$

35

[0133] Las matrices de mezclado $\mathbf{P}$ aplicadas a la señal directa $\hat{\mathbf{z}}$ y $\mathbf{M}$ aplicadas a la señal decorrelacionada $\mathbf{w}$ tienen la siguiente estructura (con $N = N_{UpmixCh}$, en donde $N_{UpmixCh}$ designa un número de señales de audio renderizadas, que puede ser igual a un número de señales de audio de salida):

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,N} \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,N} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \cdots & m_{N,N} \end{bmatrix}$$

40

[0134] Aplicar la notación para la matriz combinada $\mathbf{F} = [\mathbf{P} \quad \mathbf{M}]$ y la señal da:

$$\tilde{\mathbf{z}} = \mathbf{F}\mathbf{s} \quad \mathbf{s} = \begin{bmatrix} \hat{\mathbf{z}} \\ \mathbf{w} \end{bmatrix}$$

45

[0135] Al usar esta representación, la matriz de covarianza $\mathbf{E}_{\tilde{\mathbf{z}}}$ de la señal de salida $\tilde{\mathbf{z}}$ es definida como:

$$\mathbf{E}_{\tilde{Z}} = \mathbf{F}\mathbf{E}_S\mathbf{F}^H$$

[0136] La covarianza objetivo $\mathbf{C}$ de la escena de salida renderizada creada idealmente es definida como:

5

$$\mathbf{C} = \mathbf{R}\mathbf{E}_X\mathbf{R}^H$$

[0137] La matriz de mezclado $\mathbf{F}$ es computada de tal modo que la matriz de covarianza de la salida final se aproxima, o es igual, a la covarianza objetivo $\mathbf{C}$ como:

10

$$\mathbf{E}_{\tilde{Z}} \approx \mathbf{C}$$

[0138] La matriz de mezclado $\mathbf{F}$ es computada, por ejemplo, como una función de cantidades conocidas $\mathbf{F} = \mathbf{F}(\mathbf{E}_S, \mathbf{E}_X, \mathbf{R})$ como:

15

$$\mathbf{F} = (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)\mathbf{H}(\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H)$$

[0139] en el que las matrices $\mathbf{U}$, $\mathbf{T}$ y $\mathbf{V}$, $\mathbf{Q}$ pueden ser determinadas, por ejemplo, mediante el uso de la descomposición de valor singular (SVD) de las matrices de covarianza $\mathbf{E}_S$ y $\mathbf{C}$ dando:

20

$$\mathbf{C} = \mathbf{U}\mathbf{T}\mathbf{U}^H, \mathbf{E}_S = \mathbf{V}\mathbf{Q}\mathbf{V}^H.$$

[0140] La matriz prototipo $\mathbf{H}$ puede ser elegida según los factores de ponderación deseados para los caminos de señales directas y decorrelacionadas.

25

[0141] Por ejemplo, una matriz prototipo posible $\mathbf{H}$ puede ser determinada como:

$$\mathbf{H} = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 & b_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 & 0 & b_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{N,N} & 0 & 0 & \dots & b_{N,N} \end{bmatrix}, \text{ en donde } a_{i,i}^2 + b_{i,i}^2 = 1.$$

30

[0142] A continuación, se proporcionarán algunas derivaciones matemáticas para la estructura $\mathbf{F}$ de la matriz general.

[0143] En otras palabras, la derivación de la matriz de mezclado $\mathbf{F}$ para una solución general se describirá a continuación.

35

[0144] Las matrices de covarianza $\mathbf{E}_S$ y $\mathbf{C}$ pueden ser expresadas mediante el uso, por ejemplo, de la descomposición del valor singular (SVD) como:

40

$$\mathbf{E}_S = \mathbf{V}\mathbf{Q}\mathbf{V}^H, \mathbf{C} = \mathbf{U}\mathbf{T}\mathbf{U}^H.$$

siendo $\mathbf{T}$ y $\mathbf{Q}$ matrices diagonales con los valores singulares de $\mathbf{C}$ y $\mathbf{E}_S$ respectivamente, y siendo $\mathbf{U}$ y $\mathbf{V}$ matrices unitarias que contienen los vectores singulares correspondientes.

45

[0145] Cabe destacar que la aplicación de la triangulación de Schur o la descomposición de Eigenvalue (en lugar de SVD) lleva a resultados similares (o incluso resultados idénticos, si las matrices diagonales $\mathbf{Q}$ y $\mathbf{T}$ están restringidas a valores positivos).

[0146] Aplicando esta descomposición al requisito $\mathbf{E}_z \approx \mathbf{C}$, da (al menos aproximadamente):

$$\begin{aligned} \mathbf{C} &= \mathbf{F} \mathbf{E}_s \mathbf{F}^H \\ \mathbf{U} \mathbf{T} \mathbf{U}^H &= \mathbf{F} \mathbf{V} \mathbf{Q} \mathbf{V}^H \mathbf{F}^H, \\ (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) &= \mathbf{F} (\mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H) (\mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H) \mathbf{F}^H, \\ (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) &= (\mathbf{F} \mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H) (\mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H \mathbf{F}^H), \\ (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H)^H &= (\mathbf{F} \mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H) (\mathbf{F} \mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H)^H. \end{aligned}$$

10 [0147] Para cuidar la dimensionalidad de las matrices de covarianza, se necesita la regularización en algunos casos. Por ejemplo, se puede aplicar una matriz prototipo $\mathbf{H}$ de tamaño $N_{UpmixCh} \times 2N_{UpmixCh}$, con la propiedad de que $\mathbf{H} \mathbf{H}^H = \mathbf{I}_{N_{UpmixCh}}$:

$$\begin{aligned} (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) \mathbf{H} \mathbf{H}^H (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) &= \mathbf{F} (\mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H) (\mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H) \mathbf{F}^H, \\ (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) \mathbf{H} &= \mathbf{F} (\mathbf{V} \sqrt{\mathbf{Q}} \mathbf{V}^H). \end{aligned}$$

[0148] Se deduce que la matriz de mezclado $\mathbf{F}$ puede ser determinada como:

$$\mathbf{F} = (\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H) \mathbf{H} (\mathbf{V} \sqrt{\mathbf{Q}^{-1}} \mathbf{V}^H)$$

20 [0149] La matriz prototipo $\mathbf{H}$ es elegida según los factores de ponderación deseados para los caminos de señales directas y decorrelacionadas. Por ejemplo, una matriz prototipo posible $\mathbf{H}$ puede ser determinada como:

$$\mathbf{H} = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 & b_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 & 0 & b_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{N,N} & 0 & 0 & \dots & b_{N,N} \end{bmatrix}, \text{ en donde } a_{i,i}^2 + b_{i,i}^2 = 1.$$

[0150] Dependiendo de la condición de la matriz de covarianza $\mathbf{E}_s$ de las señales combinadas, la última ecuación puede necesitar incluir una cierta regularización, pero por lo demás debería ser numéricamente estable.

[0151] Para concluir, se ha descrito un concepto para derivar las señales de audio de salida (representadas por la matriz $\hat{\mathbf{Z}}$, o de forma equivalente, por el vector $\hat{\mathbf{z}}$) con base en las señales de audio renderizadas (representadas por la matriz $\hat{\mathbf{Z}}$, o de forma equivalente, el vector $\hat{\mathbf{z}}$) y las señales de audio decorrelacionadas (representadas por la matriz $\mathbf{W}$, o de forma equivalente, el vector $\mathbf{w}$). Como se puede ver, dos matrices de mezclado $\mathbf{P}$ y $\mathbf{M}$ de estructura de matriz general son determinadas comúnmente. Por ejemplo, una matriz combinada $\mathbf{F}$, como se definió anteriormente, puede ser determinada, de tal modo que una matriz de covarianza $\mathbf{E}_z$ de las señales 1552a a 1562n de audio de salida se aproxima, o es igual, a una covarianza deseada (también denominada covarianza objetivo) $\mathbf{C}$. La matriz de covarianza matriz deseada $\mathbf{C}$ puede ser derivada, por ejemplo, con base en el conocimiento de la matriz de renderizado $\mathbf{R}$ (que puede ser proporcionada por la interacción del usuario, por ejemplo) y con base en el conocimiento de la matriz de covarianza del objeto $\mathbf{E}_x$, que puede ser derivada, por ejemplo, con base en la información 1518 lateral codificada. Por ejemplo, la matriz de covarianza del objeto $\mathbf{E}_x$ puede ser derivada mediante el uso de los valores de correlación entre objetos IOC, que se describieron con anterioridad, y que pueden ser incluidos en la información 1518 lateral codificada. Así, la matriz de covarianza objetivo $\mathbf{C}$ puede ser proporcionada, por ejemplo, por el procesador 1570 de información lateral como la información 1574, o como parte de la información 1574.

45 [0152] Sin embargo, alternativamente, el procesador 1570 de información lateral también puede proporcionar directamente la matriz de mezclado $\mathbf{F}$ como la información 1574 al mezclador 1598.

[0153] Además, se ha descrito una regla de computación para la matriz de mezclado $\mathbf{F}$ que usa una descomposición de valor singular. Sin embargo, debería señalarse que existen algunos grados de libertad, ya que se pueden elegir las entradas $a_{i,i}$ y $b_{i,i}$ de la matriz prototipo $\mathbf{H}$. Preferentemente, las entradas de la matriz prototipo $\mathbf{H}$ son elegidas para estar entre 0 y 1. Si se eligen valores $a_{i,i}$ más cercanos a uno, habrá un mezclado significativo de señales de audio de salida renderizadas, mientras que el impacto de las señales de audio decorrelacionadas es comparativamente pequeño, lo que puede ser conveniente en algunas situaciones. Sin embargo, en algunas otras situaciones puede ser más conveniente tener un impacto comparativamente grande de las señales de audio decorrelacionadas, mientras que hay solo un débil mezclado entre las señales de audio renderizadas. En este caso, los valores $b_{i,i}$ son elegidos típicamente mayores que $a_{i,i}$. Así, el decodificador 1550 puede ser adaptado a los requisitos eligiendo apropiadamente las entradas de la matriz prototipo $\mathbf{H}$.

14.4. Procedimientos simplificados para la corrección de la covarianza de salida

- [0154]** En esta sección se describen dos estructuras alternativas para la matriz de mezclado $\mathbf{F}$ mencionada más arriba junto con algoritmos ilustrativos para determinar sus valores. Las dos alternativas están diseñadas para diferente contenido de entrada (por ejemplo, contenido de audio):
- Procedimiento de ajuste de covarianza para contenido altamente correlacionado (por ejemplo, entrada basada en canales con alta correlación entre diferentes pares de canales).
 - Procedimiento de compensación de energía para señales de entrada independientes (por ejemplo, entrada basada en objetos, que se suponen usualmente independientes).

14.4.1. Procedimiento de ajuste de covarianza (A)

- [0155]** Teniendo en cuenta que la señal $\hat{\mathbf{z}}$ (por ejemplo, las señales 1582a a 1582n de audio renderizadas) son ya óptimas en el sentido de MMSE, por lo general no es aconsejable modificar las reconstrucciones paramétricas $\hat{\mathbf{z}}$ (por ejemplo, las señales 1552a a 1552n de audio de salida) para mejorar las propiedades de covarianza de la salida $\tilde{\mathbf{z}}$ debido a que esto puede afectar a la calidad de separación.
- [0156]** Si solo se manipula la mezcla de las señales decorrelacionadas $\mathbf{W}$, la matriz de mezclado $\mathbf{P}$ puede ser reducida a una matriz de identidad (o un múltiplo de ésta). Así, este procedimiento simplificado puede ser descrito ajustando:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,N} \\ m_{2,2} & m_{2,2} & \dots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \dots & m_{N,N} \end{bmatrix}$$

- [0157]** La salida final del sistema puede ser representada como:

$$\tilde{\mathbf{z}} = \hat{\mathbf{z}} + \mathbf{M} \mathbf{W}$$

- [0158]** En consecuencia, la covarianza de salida final del sistema puede ser representada como:

$$\mathbf{E}_{\tilde{\mathbf{z}}} = \mathbf{E}_{\hat{\mathbf{z}}} + \mathbf{M} \mathbf{E}_{\mathbf{W}} \mathbf{M}^H$$

- [0159]** La diferencia Δ_E entre la matriz de covarianza de salida ideal (o deseada) $\mathbf{C}$ y la matriz de covarianza $\mathbf{E}_{\hat{\mathbf{z}}}$ de la reconstrucción paramétrica renderizada (por ejemplo, de las señales de audio renderizadas) está dada por:

$$\Delta_E = \mathbf{C} - \mathbf{E}_{\hat{\mathbf{z}}}$$

- [0160]** Por lo tanto, la matriz de mezclado $\mathbf{M}$ es determinada de tal modo que:

$$\Delta_E \approx \mathbf{M} \mathbf{E}_{\mathbf{W}} \mathbf{M}^H$$

[0161] La matriz de mezclado $\mathbf{M}$ es computada de tal modo que la matriz de covarianza de las señales decorrelacionadas mixtas $\mathbf{M}\mathbf{W}$ es igual a, o se aproxima a, la diferencia de covarianza entre la covarianza deseada y la covarianza de las señales secas (por ejemplo, de las señales de audio renderizadas). En consecuencia, la covarianza de la salida final será aproximada a la covarianza objetivo $\mathbf{E}_z \approx \mathbf{C}$:

5

$$\mathbf{M} = (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H),$$

en la que las matrices $\mathbf{U}$, $\mathbf{T}$ y $\mathbf{V}$, $\mathbf{Q}$ pueden ser determinadas, por ejemplo, mediante el uso de la descomposición de valor singular (SVD) de las matrices de covarianza $\mathbf{\Lambda}_E$ y $\mathbf{E}_W$ dando:

10

$$\mathbf{\Lambda}_E = \mathbf{U}\mathbf{T}\mathbf{U}^H, \mathbf{E}_W = \mathbf{V}\mathbf{Q}\mathbf{V}^H.$$

[0162] Esta estrategia asegura una buena reconstrucción de la correlación cruzada maximizando el uso de la salida seca (por ejemplo, de las señales de audio renderizadas 1582a a 1582n) y utiliza la libertad de mezclado de las señales decorrelacionadas solamente. En otras palabras, no hay mezclado entre diferentes señales de audio renderizadas permitido cuando se combinan las señales de audio renderizadas (o una versión en escala de éstas) con una o más señales de audio decorrelacionadas. Sin embargo, se permite que una señal decorrelacionada dada sea combinada, con una escala igual o diferente, con una pluralidad de señales de audio renderizadas, o una versión en escala de éstas, para ajustar las características de correlación cruzada o las características de covarianza cruzada de las señales de audio de salida. La combinación es definida, por ejemplo, por la matriz $\mathbf{M}$ como se ha definido en esta invención

[0163] A continuación se proporcionarán algunas derivaciones matemáticas para la estructura $\mathbf{F}$ de la matriz restringida.

25 **[0164]** En otras palabras, se explicará la derivación de la matriz de mezclado $\mathbf{M}$ para el procedimiento simplificado "A".

[0165] Las matrices de covarianza $\mathbf{\Lambda}_E$ y $\mathbf{E}_W$ pueden ser expresadas mediante el uso, por ejemplo, de la descomposición de valor singular (SVD) como:

30

$$\mathbf{\Lambda}_E = \mathbf{U}\mathbf{T}\mathbf{U}^H, \mathbf{E}_W = \mathbf{V}\mathbf{Q}\mathbf{V}^H.$$

siendo $\mathbf{T}$ y $\mathbf{Q}$ matrices diagonales con los valores singulares de $\mathbf{\Lambda}_E$ y $\mathbf{E}_W$ respectivamente, y siendo $\mathbf{U}$ y $\mathbf{V}$ matrices unitarias que contienen los vectores singulares correspondientes.

35

[0166] Cabe destacar que la aplicación de la triangulación de Schur o la descomposición de Eigenvalue (en lugar de SVD) lleva a resultados similares (o incluso resultados idénticos si las matrices diagonales $\mathbf{Q}$ y $\mathbf{T}$ están restringidas a valores positivos).

40 **[0167]** Mediante la aplicación de esta descomposición al requisito $\mathbf{E}_z \approx \mathbf{C}$, da (al menos aproximadamente)

$$\begin{aligned} \mathbf{\Lambda}_E &= \mathbf{M}\mathbf{E}_W\mathbf{M}^H \\ \mathbf{U}\mathbf{T}\mathbf{U}^H &= \mathbf{M}\mathbf{V}\mathbf{Q}\mathbf{V}^H\mathbf{M}^H, \\ (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) &= \mathbf{M}(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)\mathbf{M}^H, \\ (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) &= (\mathbf{M}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H\mathbf{M}^H), \\ (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)^H &= (\mathbf{M}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{M}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)^H, \\ (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) &= \mathbf{M}(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H). \end{aligned}$$

[0168] Observando que ambos lados de la ecuación representan un cuadrado de una matriz, dejamos el cuadrado, y resolvemos la matriz completa $\mathbf{M}$.

50

[0169] Se deduce que la matriz de mezclado $\mathbf{M}$ puede ser determinada como:

$$\mathbf{M} = (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H)$$

[0170] Este procedimiento puede ser derivado del procedimiento general mediante el ajuste de la matriz prototipo $\mathbf{H}$ del siguiente modo:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix}$$

[0171] Dependiendo de la condición de la matriz de covarianza $\mathbf{E}_w$ de las señales húmedas, la última ecuación puede necesitar incluir una cierta regularización, pero por lo demás debería ser numéricamente estable.

14.4.2. Procedimiento de compensación de energía (B)

[0172] En ocasiones (dependiendo del escenario de aplicación) no se desea permitir el mezclado de las reconstrucciones paramétricas (por ejemplo, de las señales de audio renderizadas) o de las señales decorrelacionadas, sino mezclar individualmente cada señal reconstruida paramétricamente (por ejemplo, señal de audio renderizada) con su propia señal decorrelacionada.

[0173] Para lograr este requisito, debería introducirse una restricción adicional al procedimiento simplificado "A". Ahora, se requiere que la matriz de mezclado $\mathbf{M}$ de las señales húmedas (señales decorrelacionadas) tenga una forma diagonal:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} m_{1,1} & 0 & \dots & 0 \\ 0 & m_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_{N,N} \end{bmatrix}$$

[0174] El objetivo principal de esta estrategia es usar señales decorrelacionadas para compensar la pérdida de energía en la reconstrucción paramétrica (por ejemplo, señal de audio renderizada), mientras se ignora la modificación fuera de la diagonal de la matriz de covarianza de la señal de salida, es decir, no hay una manipulación directa de las correlaciones cruzadas. Por lo tanto, no se introduce una pérdida cruzada entre los objetos/canales de salida (por ejemplo, entre las señales de audio renderizadas) en la aplicación de las señales decorrelacionadas.

[0175] Como resultado, solo se puede alcanzar la diagonal principal de la matriz de covarianza objetivo (o la matriz de covarianza deseada) y las diagonales secundarias se encuentran a merced de la exactitud de la reconstrucción paramétrica y las señales decorrelacionadas agregadas. Este procedimiento es muy adecuado para las aplicaciones basadas solo en el objeto, en el cual las señales pueden ser consideradas como decorrelacionadas.

[0176] La salida final del procedimiento (por ejemplo, las señales de audio de salida) está dada por $\hat{\mathbf{Z}} = \hat{\mathbf{Z}} + \mathbf{M} \mathbf{W}$ con una matriz diagonal $\mathbf{M}$ computada de tal modo que las entradas de la matriz de covarianza correspondientes a las energías de las señales reconstruidas $\mathbf{E}_{\hat{\mathbf{Z}}}(i, i)$ son iguales a las energías deseadas:

$$\mathbf{E}_{\hat{\mathbf{Z}}}(i, i) = \mathbf{C}(i, i)$$

[0177] $\mathbf{C}$ puede ser determinada como se ha explicado anteriormente para el caso general.

[0178] Por ejemplo, la matriz de mezclado $\mathbf{M}$ puede ser derivada directamente dividiendo las energías deseadas de las señales de compensación (diferencias entre las energías deseadas (que pueden ser descritas por los elementos diagonales de la matriz de covarianza cruzada $\mathbf{C}$) y las energías de las reconstrucciones paramétricas (que pueden ser determinadas por el decodificador de audio)) con las energías de las señales decorrelacionadas (que pueden ser determinadas por el decodificador de audio):

$$\mathbf{M}(i, j) = \begin{cases} \sqrt{\min \left(\lambda_{Dec}, \max \left(0, \frac{\mathbf{C}(i, i) - \mathbf{E}_{\hat{z}}(i, i)}{\max(\mathbf{E}_w(i, i), \varepsilon)} \right) \right)} & i = j, \\ 0 & i \neq j. \end{cases}$$

en donde λ_{Dec} es un umbral no negativo usado para limitar la cantidad de componentes decorrelacionados
5 agregados a las señales de salida (por ejemplo, $\lambda_{Dec} = 4$).

[0179] Debería señalarse que las energías pueden ser reconstruidas paramétricamente (por ejemplo, mediante el uso de OLDs, IOC's y coeficientes de renderización) o pueden ser computadas realmente por el decodificador (que es típicamente más costoso computacionalmente).

10

[0180] Este procedimiento puede ser derivado del procedimiento general mediante el ajuste de la matriz prototipo $\mathbf{H}$ del siguiente modo:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix}$$

15

[0181] Este procedimiento maximiza el uso de las salidas renderizadas secas explícitamente. El procedimiento es equivalente a la simplificación "A" cuando las matrices de covarianza no tienen entradas fuera de las diagonales.

20 **[0182]**

Este procedimiento tiene una complejidad computacional reducida.

[0183] Sin embargo, debería señalarse que el procedimiento de compensación de energía no necesariamente implica que los términos de correlación cruzada no están modificados. Esto se cumple solo si usamos decorrelacionadores ideales y ninguna reducción de la complejidad para la unidad de decorrelación. La idea del procedimiento es recuperar la energía e ignorar las modificaciones en los términos cruzados (los cambios en los términos cruzados no modificarán sustancialmente las propiedades de correlación y no afectarán a la impresión espacial general).

25

14.5. Requisitos para la matriz de mezclado $\mathbf{F}$

30

[0184] A continuación se explicará que la matriz de mezclado $\mathbf{F}$, una derivación de la cual ha sido descrita en las secciones 14.3 y 14.4, cumple con los requisitos para evitar degradaciones.

[0185] Para evitar degradaciones en la salida, cualquier procedimiento para compensar los errores de la reconstrucción paramétrica debería producir un resultado con la siguiente propiedad: si la matriz de renderización es igual a la matriz de mezcla descendente, entonces los canales de salida deberían ser iguales (o al menos aproximados) a los canales de mezcla descendente. El modelo propuesto cumple con esta propiedad. Si la matriz de renderizado es igual a la matriz de mezcla descendente $\mathbf{R} = \mathbf{D}$, la reconstrucción paramétrica está dada por:

35

$$\hat{\mathbf{Z}} = \mathbf{R}\hat{\mathbf{X}} = \mathbf{D}\hat{\mathbf{X}} = \mathbf{D}\mathbf{G}\mathbf{Y} = \mathbf{D}\mathbf{E}\mathbf{D}^H (\mathbf{D}\mathbf{E}\mathbf{D}^H)^{-1} \mathbf{Y} \approx \mathbf{Y},$$

40

y la matriz de covarianza deseada será:

$$\mathbf{C} = \mathbf{R}\mathbf{E}_x \mathbf{R}^H = \mathbf{D}\mathbf{E}_x \mathbf{D}^H = \mathbf{E}_y.$$

45

[0186] Por lo tanto la ecuación que hay que resolver para obtener la matriz de mezclado $\mathbf{F}$ es:

$$\mathbf{E}_y = \mathbf{F} \begin{bmatrix} \mathbf{E}_y & \mathbf{0}_{N_{UpmixCh}} \\ \mathbf{0}_{N_{UpmixCh}} & \mathbf{E}_w \end{bmatrix} \mathbf{F}^H$$

en donde $\mathbf{0}_{N_{UpmixCh}}$ es una matriz cuadrada de tamaño $N_{UpmixCh} \times N_{UpmixCh}$ de ceros. Al resolver la ecuación anterior para $\mathbf{F}$, se puede obtener:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \end{bmatrix}$$

5

[0187] Esto significa que las señales decorrelacionadas tendrán un peso cero en la suma, y la salida final estará dada por las señales secas, que son idénticas a las señales de mezcla descendente:

$$\tilde{\mathbf{Z}} = \mathbf{P} \hat{\mathbf{Z}} + \mathbf{M} \mathbf{W} = \hat{\mathbf{Z}} \approx \mathbf{Y}.$$

10

[0188] Como resultado, se cumple el requisito dado para que la salida del sistema sea igual a la señal de mezcla descendente en este escenario de renderización.

14.6. Estimación de la matriz de covarianza de señales $\mathbf{E}_S$

15

[0189] Para obtener la matriz de mezclado $\mathbf{F}$ se requiere, o al menos es deseable, el conocimiento de la matriz de covarianza $\mathbf{E}_S$ de las señales combinadas $\mathbf{S}$.

[0190] En principio, es posible estimar la matriz de covarianza $\mathbf{E}_S$ directamente de las señales disponibles (a saber, de la reconstrucción paramétrica $\hat{\mathbf{Z}}$ y la salida del decorrelacionador $\mathbf{W}$). Aunque esta propuesta puede llevar a resultados más exactos, puede no ser práctica debido a la complejidad computacional asociada. Los procedimientos propuestos usan aproximaciones paramétricas de la matriz de covarianza $\mathbf{E}_S$.

25

[0191] La estructura general de la matriz de covarianza $\mathbf{E}_S$ puede ser representada como:

$$\mathbf{E}_S = \begin{bmatrix} \mathbf{E}_{\hat{\mathbf{Z}}} & \mathbf{E}_{\hat{\mathbf{Z}}\mathbf{W}}^H \\ \mathbf{E}_{\hat{\mathbf{Z}}\mathbf{W}} & \mathbf{E}_{\mathbf{W}} \end{bmatrix}$$

en el que la matriz $\mathbf{E}_{\hat{\mathbf{Z}}\mathbf{W}}$ es de covarianza cruzada entre las señales directas $\hat{\mathbf{Z}}$ y decorrelacionadas $\mathbf{W}$.

30

[0192] Suponiendo que los decorrelacionadores son ideales (es decir, que conservan la energía, siendo las salidas ortogonales con respecto a las entradas, y siendo todas las salidas mutuamente ortogonales), la matriz de covarianza $\mathbf{E}_S$ puede ser expresada mediante el uso de la forma simplificada como:

$$\mathbf{E}_S = \begin{bmatrix} \mathbf{E}_{\hat{\mathbf{Z}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{\mathbf{W}} \end{bmatrix}$$

35

[0193] La matriz de covarianza $\mathbf{E}_{\hat{\mathbf{Z}}}$ de la señal paramétricamente reconstruida $\hat{\mathbf{Z}}$ puede ser determinada paramétricamente como:

$$\mathbf{E}_{\hat{\mathbf{Z}}} = \mathbf{R} \mathbf{E}_{\hat{\mathbf{x}}} \mathbf{R}^H = \mathbf{R} \mathbf{G} \mathbf{D} \mathbf{E}_x \mathbf{D}^H \mathbf{G}^H \mathbf{R}^H.$$

40

[0194] La matriz de covarianza $\mathbf{E}_{\mathbf{W}}$ de la señal decorrelacionada $\mathbf{W}$ se supone que cumple con la propiedad de ortogonalidad mutua y que contiene solo los elementos diagonales de $\mathbf{E}_{\hat{\mathbf{Z}}}$ del siguiente modo:

$$\mathbf{E}_w(i, j) = \begin{cases} \mathbf{E}_{\hat{z}}(i, i) & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

[0195] Si no se cumple la suposición de ortogonalidad mutua y/o conservación de la energía (por ejemplo, en el caso en que el número de decorrelacionadores disponible es más pequeño que el número de señales que van a

5 ser decorrelacionadas), entonces la matriz de covarianza $\mathbf{E}_w$ puede ser estimada como:

$$\mathbf{E}_w = \mathbf{M}_{post} \left[\text{matdiag}(\mathbf{M}_{pre} \mathbf{E}_{\hat{z}} \mathbf{M}_{pre}^H) \right] \mathbf{M}_{post}^H$$

15. Reducción de la complejidad para la unidad de decorrelación

10

[0196] A continuación se describirá cómo la complejidad de los decorrelacionadores usados en las formas de realización según la presente invención puede ser reducida.

[0197]

Debería señalarse que la implementación de la función del decorrelacionador es frecuentemente computacionalmente compleja. En algunas aplicaciones (por ejemplo, soluciones de decodificador portables) el número de decorrelacionadores puede necesitar introducirse debido a las fuentes computacionales restringidas. Esta sección proporciona una descripción de medios para la reducción de la complejidad de la unidad de decorrelacionador controlando el número de decorrelacionadores (o decorrelaciones) aplicados. La interfaz de la unidad de decorrelación se ilustra en las figuras 16 y 17.

20

[0198] La figura 16 muestra un diagrama de bloque esquemático de una unidad de decorrelación simple (convencional). La unidad 1600 de decorrelación según la figura 6 está configurada para recibir $\mathbf{N}$ señales 1610a a 1610n de entrada del decorrelacionador, como, por ejemplo, señales de audio renderizadas $\hat{\mathbf{z}}$. Además, La unidad 1600 de decorrelación proporciona señales 1612a a 1612n de salida del decorrelacionador $\mathbf{N}$. La unidad 1600 de decorrelación puede comprender, por ejemplo, decorrelacionadores individuales $\mathbf{N}$ (o funciones de decorrelación) 1620a a 1620n. Por ejemplo, cada uno de los decorrelacionadores 1620a a 1620n individuales puede proporcionar una de las señales 1612a a 1612n de salida del decorrelacionador con base en una asociada de las señales 1610a a 1610n de entrada del decorrelacionador. Por consiguiente, los decorrelacionadores 1620a a 1620n individuales $\mathbf{N}$, o funciones de decorrelación, pueden ser requeridos para proporcionar las señales decorrelacionadas $\mathbf{N}$ 1612a a 1612n con base en las señales 1610a a 1610n de entrada del decorrelacionador $\mathbf{N}$.

30

[0199] Sin embargo, la figura 17 muestra un diagrama de bloque esquemático de una unidad 1700 de decorrelación de complejidad reducida. La unidad 1700 de decorrelación de complejidad reducida está configurada para recibir señales 1710a a 1710n de entrada del decorrelacionador $\mathbf{N}$ y para proporcionar, con base en éstas, señales 1712a a 1712n de salida del decorrelacionador $\mathbf{N}$. Por ejemplo, las señales 1710a a 1710n de entrada del decorrelacionador pueden ser señales de audio renderizadas $\hat{\mathbf{z}}$, y las señales 1712a a 1712n de salida del decorrelacionador pueden ser señales de audio decorrelacionadas $\mathbf{W}$.

35

[0200] El decorrelacionador 1700 comprende un pre-mezclador (o de forma equivalente, una funcionalidad de premezclado) 1720 que está configurada para recibir el primer conjunto de señales de entrada del decorrelacionador $\mathbf{N}$ 1710a a 1710n y para proporcionar, con base en éstas, un segundo conjunto de señales 1722a a 1722k de entrada del decorrelacionador $\mathbf{K}$. Por ejemplo, el pre-mezclador 1720 puede realizar un así llamado "premezclado" o "mezclado descendente" para derivar el segundo conjunto de señales 1722a a 1722k de entrada del decorrelacionador $\mathbf{K}$ con base en el primer conjunto de señales 1710a a 1710n de entrada del decorrelacionador $\mathbf{N}$.

45

Por ejemplo, las señales $\mathbf{K}$ del segundo conjunto de señales 1722a a 1722k de entrada del decorrelacionador $\mathbf{K}$ pueden ser representadas mediante el uso de una matriz $\hat{\mathbf{z}}_{mix}$. La unidad de decorrelación (o, de forma equivalente, el decorrelacionador multicanal) 1700 también comprende un núcleo de decorrelacionador 1730, que está configurado para recibir las señales $\mathbf{K}$ del segundo conjunto de señales 1722a a 1722k de entrada del decorrelacionador $\mathbf{K}$, y para proporcionar, con base en éstas, las señales de salida del decorrelacionador $\mathbf{K}$ que constituyen un primer conjunto de señales 1732a a 1732k de salida del decorrelacionador. Por ejemplo, el núcleo del decorrelacionador 1730 puede comprender decorrelacionadores individuales $\mathbf{K}$ (o funciones de decorrelación), en donde cada uno de los decorrelacionadores individuales (o funciones de decorrelación) proporciona una de las señales de salida del decorrelacionador del primer conjunto de señales 1732a a 1732k de salida del decorrelacionador $\mathbf{K}$ con base en una señal de entrada del decorrelacionador correspondiente del segundo conjunto de señales 1722a a 1722k de entrada del decorrelacionador $\mathbf{K}$. Alternativamente, un decorrelacionador dado, o función de decorrelación, puede ser aplicada $\mathbf{K}$ veces, de tal modo que cada una de las señales de salida del decorrelacionador del primer conjunto de señales 1732a a 1732k de salida del decorrelacionador $\mathbf{K}$ se basa en una sola de las señales de entrada del decorrelacionador del segundo conjunto de señales 1722a a 1722k de entrada del

50

55

decorrelacionador K.

[0201] La unidad de decorrelación 1700 comprende también un post-mezclador 1740, que está configurado para recibir las señales 1732a a 1732k de salida del decorrelacionador K del primer conjunto de señales de salida del decorrelacionador y para proporcionar, con base en éstas, las N señales 1712a a 1712n del segundo conjunto de señales de salida del decorrelacionador (que constituyen las señales de salida del decorrelacionador "externas").

[0202] Debería señalarse que el pre-mezclador 1720 puede realizar preferentemente una operación de mezclado lineal, que puede ser descrita por una matriz de premezclado $\mathbf{M}_{pre}$. Además, el post-mezclador 1740 realiza preferentemente una operación de mezclado lineal (o mezclado ascendente), que puede ser representada por una matriz de post-mezclado $\mathbf{M}_{post}$, para derivar las señales de salida del decorrelacionador N 1712a a 1712n del segundo conjunto de señales de salida del decorrelacionador del primer conjunto de señales 1732a a 1732k de salida del decorrelacionador K (es decir, de las señales de salida del núcleo del decorrelacionador 1730).

[0203] La idea principal del procedimiento propuesto y del aparato es reducir el número de señales de entrada a los decorrelacionadores (o al núcleo del decorrelacionador) de N a K por:

- Premezclado de las señales (por ejemplo, las señales de audio renderizadas) para bajar el número de canales con:

$$\hat{\mathbf{Z}}_{mix} = \mathbf{M}_{pre} \hat{\mathbf{Z}}$$

- Aplicación de la decorrelación mediante el uso de los decorrelacionadores K disponibles (por ejemplo, del núcleo del decorrelacionador) con:

$$\hat{\mathbf{Z}}_{mix}^{dec} = Decorr(\hat{\mathbf{Z}}_{mix})$$

- Mezclado de manera ascendente de las señales decorrelacionadas de nuevo a los canales N con:

$$\mathbf{W} = \mathbf{M}_{post} \hat{\mathbf{Z}}_{mix}^{dec}$$

[0204] La matriz de premezclado $\mathbf{M}_{pre}$ puede ser construida con base en la información de mezcla descendente/renderización/correlación/etc., de tal modo que el producto de la matriz $(\mathbf{M}_{pre} \mathbf{M}_{pre}^H)$ se vuelve bien acondicionado (con respecto a la operación de inversión). La matriz de post-mezclado puede ser computada como:

$$\mathbf{M}_{post} \approx \mathbf{M}_{pre}^H (\mathbf{M}_{pre} \mathbf{M}_{pre}^H)^{-1}$$

[0205] Aunque la matriz de covarianza de las señales decorrelacionadas intermedias $\hat{\mathbf{S}}$ (o $\hat{\mathbf{Z}}_{mix}^{dec}$) es diagonal (suponiendo decorrelacionadores ideales), la matriz de covarianza de las señales decorrelacionadas finales $\mathbf{W}$ probablemente ya no sea diagonal cuando usa este tipo de procesamiento. Por lo tanto, la matriz de covarianza puede ser estimada mediante el uso de las matrices de mezclado como:

$$\mathbf{E}_W = \mathbf{M}_{post} \left[matdiag(\mathbf{M}_{pre} \mathbf{E}_{\hat{\mathbf{Z}}} \mathbf{M}_{pre}^H) \right] \mathbf{M}_{post}^H$$

[0206] El número de decorrelacionadores (o decorrelaciones) usados, K , no es especificado y depende de la complejidad computacional deseada y de los decorrelacionadores disponibles. Su valor puede variar de N (complejidad computacional más alta) hacia abajo hasta 1 (complejidad computacional más baja).

[0207] El número de señales de entrada a la unidad de decorrelacionador, N , es arbitrario y el procedimiento propuesto soporta cualquier número de señales de entrada, independientemente de la configuración de representación del sistema.

[0208] Por ejemplo, en aplicaciones que usan un contenido de audio 3D, con alto número de canales de salida, dependiendo de la configuración de salida una expresión posible para la matriz de premezclado $\mathbf{M}_{pre}$ se describe más abajo.

[0209] A continuación, se describirá cómo el premezclado, que se realiza por el pre-mezclador 1720 (y, por consiguiente, el post-mezclado, que se lleva a cabo por el post-mezclador 1740) es ajustado si la unidad de decorrelación 1700 se usa en un decodificador de audio multicanal, en el que las señales 1710a a 1710n de entrada del decorrelacionador del primer conjunto de señales de entrada del decorrelacionador están asociadas con diferentes posiciones espaciales de una escena de audio.

[0210] Para este propósito, la figura 18 muestra una representación en forma de tabla de posiciones de altavoz, que se usan para diferentes formatos de salida.

[0211] En la tabla 1800 de la figura 18, una primera columna 1810 describe un número de índice de altavoz. Una segunda columna 1820 describe una marca de altavoz. Una tercera columna 1830 describe una posición azimutal del altavoz respectivo y una cuarta columna 1832 describe una tolerancia azimutal de la posición del altavoz. Una quinta columna 1840 describe una elevación de una posición del altavoz respectivo y una sexta columna 1842 describe una tolerancia de elevación correspondiente. Una séptima columna 1850 indica qué altavoces se usan para el formato de salida O-2.0. Una octava columna 1860 muestra qué altavoces se usan para el formato de salida O-5.1. Una novena columna 1864 muestra qué altavoces se usan para el formato de salida O-7.1. Una décima columna 1870 muestra qué altavoces se usan para el formato de salida O-8.1, una undécima columna 1880 muestra qué altavoces se usan para el formato de salida O-10.1, y una duodécima segunda columna 1890 muestra qué altavoces se usan para el formato de salida O-22.2. Como se puede observar, dos altavoces se usan para el formato de salida O-2.0, seis altavoces se usan para el formato de salida O-5.1, ocho altavoces se usan para el formato de salida O-7.1, nueve altavoces se usan para el formato de salida O-8.1, 11 altavoces se usan para el formato de salida O-10.1, y 24 altavoces se usan para el formato de salida O-22.2.

[0212] Sin embargo, debería señalarse que se usa un altavoz de efecto de baja frecuencia para formatos de salida O-5.1, O-7.1, O-8.1 y O-10.1, y que se usan dos altavoces de efecto de baja frecuencia (LFE1, LFE2) para el formato de salida O-22.2. Además, debería señalarse que, en una forma de realización preferida, una señal de audio renderizada (por ejemplo, una de las señales 1582a a 1582n de audio renderizadas) está asociada con cada uno de los altavoces, excepto por uno o más altavoces de efecto de baja frecuencia. Por consiguiente, dos señales de audio renderizadas están asociadas con los dos altavoces usados según el formato O-2.0, cinco señales de audio renderizadas están asociadas con los cinco altavoces que no son de efecto de baja frecuencia si se usa el formato O-5.1, siete señales de audio renderizadas están asociadas con siete altavoces que no son de efecto de baja frecuencia si se usa el formato O-7.1, ocho señales de audio renderizadas están asociadas con los ocho altavoces que no son de baja frecuencia si se usa el formato O-8.1, diez señales de audio renderizadas están asociadas con los diez altavoces que no son de baja frecuencia si se usa el formato O-10.1, y 22 señales de audio renderizadas están asociadas con los 22 altavoces que no son de efecto de baja frecuencia si se usa el formato O-22.2.

[0213] Sin embargo, frecuentemente es conveniente usar un número más pequeño de decorrelacionadores (de núcleo de decorrelacionador) (individuales), como se ha mencionado más arriba. A continuación se describirá cómo se puede reducir el número de decorrelacionadores de forma flexible cuando el formato de salida O-22.2 es usado por un decodificador de audio multicanal, de tal modo que hay 22 señales 1582a a 1582n de audio renderizadas (que pueden ser representadas por una matriz $\hat{Z}$, o por un vector $\hat{z}$).

[0214] Las figuras 19a a 19g representan diferentes opciones para el premezclado de las señales 1582a a 1582n de audio renderizadas bajo la suposición de que hay $N = 22$ señales de audio renderizadas. Por ejemplo, la figura 19a muestra una representación en forma de tabla de entradas de una matriz de premezclado M_{pre} . Las filas, marcadas con 1 a 11 en la figura 19a, representan las filas de la matriz de premezclado M_{pre} , y las columnas, marcadas con 1 a 22 están asociadas con columnas de la matriz de premezclado M_{pre} . Además, debería señalarse que cada fila de la matriz de premezclado M_{pre} está asociada con una de las señales de entrada del decorrelacionador K 1722a a 1722k del segundo conjunto de señales de entrada del decorrelacionador (es decir, con las señales de entrada del núcleo de decorrelacionador). Además, cada columna de la matriz de premezclado M_{pre} está asociada con una de las señales de entrada del decorrelacionador N 1710a a 1710n del primer conjunto de señales de entrada del decorrelacionador, y en consecuencia con una de las señales 1582a a 1582n de audio renderizadas (ya que las señales 1710a a 1710n de entrada del decorrelacionador del primer conjunto de señales de entrada del decorrelacionador son típicamente idénticas a las señales 1582 a 1582n de audio renderizadas en una forma de realización). Por consiguiente, cada columna de la matriz de premezclado M_{pre} está asociada con un altavoz específico y, en consecuencia, como los altavoces están asociados con posiciones espaciales, con una posición espacial específica. Una fila 1910 indica a qué altavoz (y, en consecuencia, a qué posición espacial) están asociadas las columnas de la matriz de premezclado M_{pre} (en el que las marcas de los altavoces están definidas en la columna 1820 de la tabla 1800).

[0215] A continuación, se describe con más detalle la funcionalidad definida por el premezclado M_{pre} de la figura 19a. Como se puede observar, las señales de audio renderizadas asociadas con los altavoces (o, de forma equivalente, las posiciones de los altavoces) "CH_M_000" y "CH_L_000" están combinadas, para obtener una

primera señal de entrada de decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador (es decir, una primera señal de entrada de decorrelacionador mezclada hacia abajo), que está indicada por los valores "1" en la primera y la segunda columnas de la primera fila de la matriz de premezclado M_{pre} . Similarmente, las señales de audio renderizadas asociadas con altavoces (o, de forma equivalente, posiciones de altavoces)

5 "CH_U_000" y "CH_T_000" están combinadas para obtener una segunda señal de entrada de decorrelacionador mezclada hacia abajo (es decir, una segunda señal de entrada de decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador). Además, se puede observar que la matriz de premezclado M_{pre} de la figura 19a define once combinaciones de dos señales de audio renderizadas cada una, de tal modo que once

10 señales de entrada del decorrelacionador mezcladas de manera descendente son derivadas de 22 señales de audio renderizadas. También se puede observar que cuatro señales centrales son combinadas para obtener dos señales de entrada del decorrelacionador mezcladas de manera descendente (comparar las columnas 1 a 4 y las filas 1 y 2 de la matriz de premezclado). Además, se puede observar que las otras señales de entrada del decorrelacionador mezcladas de manera descendente son obtenidas cada una combinando dos señales de audio asociadas con el mismo lado de la escena de audio. Por ejemplo, una tercera señal de entrada de decorrelacionador de mezclado

15 descendente, representada por la tercera fila de la matriz de premezclado, es obtenida combinando señales de audio renderizadas asociadas con una posición azimutal de +135 ("CH_M_L135"; "CH_U_L135"). Además, se puede observar que una cuarta señal de entrada de decorrelacionador (representada por una cuarta fila de la matriz de premezcla) es obtenida combinando señales de audio renderizadas asociadas con una posición azimutal de -135° ("CH_M_R135"; "CH_U_R135"). Por consiguiente, cada una de las señales de entrada del decorrelacionador

20 mezcladas de manera descendente es obtenida combinando dos señales de audio renderizadas asociadas con la misma posición azimutal (o similar) (o, de forma equivalente, la posición horizontal), en donde hay típicamente una combinación de señales asociadas con diferente elevación (o, de forma equivalente, la posición vertical).

[0216] Haciendo referencia ahora a la figura 19b, que muestra coeficientes de premezclado (entradas de la

25 matriz de premezclado M_{pre}) para $N = 22$ y $K = 10$. La estructura de la tabla de la figura 19b es idéntica a la estructura de la tabla de la figura 19a. Sin embargo, como se puede observar, la matriz de premezclado M_{pre} según la figura 19b difiere de la matriz de premezclado M_{pre} de la figura 19a en cuanto a que la primera fila describe la combinación de cuatro señales de audio renderizadas que tienen IDs de canales (o posiciones) "CH_M_000", "CH_L_000", "CH_U_000" y "CH_T_000". En otras palabras, cuatro señales de audio renderizadas asociadas con

30 posiciones verticalmente adyacentes están combinadas en el premezclado para reducir el número de decorrelacionadores requeridos (diez decorrelacionadores en lugar de once decorrelacionadores para la matriz según la figura 19a).

[0217] Haciendo referencia ahora a la figura 19c, que muestra coeficientes de premezclado (entradas de la

35 matriz de premezclado M_{pre}) para $N = 22$ y $K = 9$, se puede observar que la matriz de premezclado M_{pre} según la figura 19c comprende nueve filas. Además, se puede observar de la segunda fila de la matriz de premezclado M_{pre} de la figura 19c que las señales de audio renderizadas asociadas con las IDs de los canales (o las posiciones) "CH_M_L135", "CH_U_L135", "CH_M_R135" y "CH_U_R135" están combinadas (en un pre-mezclador configurado según la matriz de premezclado de la figura 19c) para obtener una segunda señal de entrada de decorrelacionador

40 de mezclado descendente (señal de entrada de decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador). Como se puede observar, las señales de audio renderizadas que han sido combinadas en señales de entrada del decorrelacionador separadas mezcladas de manera descendente por las matrices de premezclado según las figuras 19a y 19b son mezcladas de manera descendente en una señal de entrada de decorrelacionador de mezclado descendente común según la figura 19c. Además, debería señalarse que las

45 señales de audio renderizadas que tienen las IDs de canales "CH_M_L135" y "CH_U_L135" están asociadas con posiciones horizontales idénticas (o posiciones azimutales) en el mismo lado de la escena de audio y posiciones verticales espacialmente adyacentes (o elevaciones), y que las señales de audio renderizadas que tienen las IDs de canales "CH_M_R135" y "CH_U_R135" están asociadas con posiciones horizontales idénticas (o posiciones azimutales) en un segundo lado de la escena de audio y posiciones verticales espacialmente adyacentes (o elevaciones). Además, se puede decir que las señales de audio renderizadas que tienen las IDs de canales

50 "CH_M_L135", "CH_U_L135", "CH_M_R135" y "CH_U_R135" están asociadas con un par horizontal (o incluso un cuarteto horizontal) de posiciones espaciales que comprenden una posición del lado izquierdo y una posición del lado derecho. En otras palabras, se puede observar en la segunda fila de la matriz de premezclado M_{pre} de la figura 19c que dos de las cuatro señales de audio renderizadas, que están combinadas para ser decorrelacionadas

55 mediante el uso de un solo decorrelacionador dado, están asociadas con posiciones espaciales en un lado izquierdo de una escena de audio, y que dos de las cuatro señales de audio renderizadas que están combinadas para ser decorrelacionadas mediante el uso del mismo decorrelacionador dado, están asociadas con posiciones espaciales en un lado derecho de la escena de audio. Además, se puede observar que las señales de audio renderizadas del lado izquierdo (de dichas cuatro señales de audio renderizadas) están asociadas con posiciones espaciales que son

60 simétricas, con respecto a un plano central de la escena de audio, con las posiciones espaciales asociadas con las señales de audio renderizadas del lado derecho (de dichas cuatro señales de audio renderizadas), de tal modo que un cuarteto "simétrico" de señales de audio renderizadas es combinado por el premezclado para ser decorrelacionado mediante el uso de un solo decorrelacionador (individual).

65 **[0218]** Haciendo referencia a las figuras 19d, 19e, 19f y 19g, se puede observar que más y más señales de

audio renderizadas son combinadas con un número decreciente de decorrelacionadores (individuales) (es decir, con K decreciente). Como se puede observar en las figuras 19a a 19g, típicamente las señales de audio renderizadas que son mezcladas de manera descendente en dos señales de entrada del decorrelacionador separadas de aquellas mezcladas de manera descendente son combinadas cuando disminuye el número de decorrelacionadores por 1. Además, se puede observar que típicamente son combinadas las señales de audio renderizadas, que están asociadas con un “cuarteto simétrico” de posiciones espaciales, en donde, para un número comparativamente alto de decorrelacionadores, solo señales de audio renderizadas asociadas con posiciones horizontales iguales o al menos similares (o posiciones azimutales) son combinadas, mientras que para un número comparativamente más bajo de decorrelacionadores, las señales de audio renderizadas asociadas con posiciones espaciales en lados opuestos de la escena de audio también son combinadas.

[0219] Haciendo referencia ahora a las figuras 20a a 20d, 21a a 21c, 22a a 22b y 23, debería señalarse que también se pueden aplicar conceptos similares para un número diferente de señales de audio renderizadas.

[0220] Por ejemplo, las figuras 20a a 20d describen entradas de la matriz de premezclado M_{pre} para $N=10$ y para K entre 2 y 5.

[0221] Similarmente, las figuras 21a a 21c describen entradas de la matriz de premezclado M_{pre} para $N=8$ y K entre 2 y 4.

[0222] Similarmente, las figuras 21d a 21f describen entradas de la matriz de premezclado M_{pre} para $N=7$ y K entre 2 y 4.

[0223] Las figuras 22a y 22b muestran entradas de la matriz de premezclado para $N=5$ y $K=2$ y $K=3$.

[0224] Finalmente, la figura 23 muestra entradas de la matriz de premezclado para $N=2$ y $K=1$.

[0225] Para resumir, las matrices de premezclado según las figuras 19 a 23 se pueden usar, por ejemplo, de una manera conmutable, en un decorrelacionador multicanal que es parte de un decodificador de audio multicanal. La conmutación entre las matrices de premezclado se puede realizar, por ejemplo, dependiendo de una configuración de salida deseada (que determina típicamente un número N de señales de audio renderizadas) y también dependiendo de una complejidad deseada de la decorrelación (que determina el parámetro K, y que puede ser ajustada, por ejemplo, dependiendo de una información de complejidad incluida en una representación codificada de un contenido de audio).

[0226] Haciendo referencia ahora a la figura 24, se describirá ahora con más detalle la reducción de la complejidad para el formato de salida 22.2. Como ya se ha indicado más arriba, una solución posible para construir la matriz de premezclado y la matriz de post-mezclado es usar la información espacial del diseño de reproducción para seleccionar los canales que van a ser mezclados entre sí y computar los coeficientes de mezclado. Con base en su posición, los altavoces geoméricamente relacionados (y, por ejemplo, las señales de audio renderizadas asociadas con estos) están agrupados juntos, tomando pares verticales y horizontales, como se describe en la tabla de la figura 24. En otras palabras, la figura 24 muestra, en forma de una tabla, un agrupamiento de posiciones de altavoces, que puede estar asociado con señales de audio renderizadas. Por ejemplo, una primera fila 2410 describe un primer grupo de posiciones de altavoces, que están en un centro de una escena de audio. Una segunda fila 2412 representa un segundo grupo de posiciones de altavoces, que están relacionados espacialmente. Las posiciones de altavoces “CH_M_L135” y “CH_U_L135” están asociadas con posiciones azimutales idénticas (o posiciones horizontales equivalentes) y posiciones de elevación adyacentes (o de forma equivalente, posiciones verticalmente adyacentes). Similarmente, las posiciones “CH_M_R135” y “CH_U_R135” comprenden ángulos de azimut idénticos (o, de forma equivalente, posición horizontal idéntica) y elevación idéntica (o, de forma equivalente, posición verticalmente adyacente). Además, las posiciones “CH_M_L135”, “CH_U_L135”, “CH_M_R135” y “CH_U_R135” forman un cuarteto de posiciones, en el que las posiciones “CH_M_L135” y “CH_U_L135” son simétricas con respecto a las posiciones “CH_M_R135” y “CH_U_R135” con respecto a un plano central de la escena de audio. Además, las posiciones “CH_M_180” y “CH_U_180” comprenden también una posición azimutal idéntica (o, de forma equivalente, una posición horizontal idéntica) y una elevación similar (o, de forma equivalente, una posición vertical adyacente).

[0227] Una tercera fila 2414 representa un tercer grupo de posiciones. Debería señalarse que las posiciones “CH_M_L030” y “CH_L_L045” son posiciones espacialmente adyacentes y comprenden un ángulo de azimut similar (o, de forma equivalente, una posición horizontal similar) y una elevación similar (o, de forma equivalente, una posición vertical similar). Lo mismo vale para posiciones “CH_M_R030” y “CH_L_R045”. Además, las posiciones del tercer grupo de posiciones forman un cuarteto de posiciones, en el que las posiciones “CH_M_L030” y “CH_L_L045” son espacialmente adyacentes y simétricas con respecto a un plano central de la escena de audio, a posiciones “CH_M_R030” y “CH_L_R045”.

[0228] Una cuarta fila 2416 representa cuatro posiciones adicionales, que tienen características similares

cuando se comparan con las primeras cuatro posiciones de la segunda fila, y que forman un cuarteto simétrico de posiciones.

[0229] Una quinta fila 2418 representa otro cuarteto de posiciones simétricas "CH_M_L060", "CH_U_L045",
5 "CH_M_R060" y "CH_U_R045".

[0230] Además, debería señalarse que las señales de audio renderizadas asociadas con las posiciones de los diferentes grupos de posiciones pueden ser combinadas más y más con un número decreciente de decorrelacionadores. Por ejemplo, en presencia de once decorrelacionadores individuales en un decorrelacionador
10 multicanal, las señales de audio renderizadas asociadas con posiciones en la primera y la segunda columna pueden ser combinadas para cada grupo. Además, las señales de audio renderizadas asociadas con las posiciones representadas en una tercera y una cuarta columna pueden ser combinadas para cada grupo. Además, las señales de audio renderizadas asociadas con las posiciones mostradas en la quinta y la sexta columnas pueden ser combinadas para el segundo grupo. Por consiguiente, se pueden obtener once señales de mezcla descendente de
15 entrada del decorrelacionador (que son introducidas en los decorrelacionadores individuales). Sin embargo, si se desea tener menos decorrelacionadores individuales, las señales de audio renderizadas asociadas con las posiciones mostradas en las columnas 1 a 4 pueden ser combinadas para uno o más de los grupos. También, las señales de audio renderizadas asociadas con todas las posiciones del segundo grupo pueden ser combinadas, si se desea reducir adicionalmente un número de decorrelacionadores individuales.

[0231] Para resumir, las señales suministradas al diseño de salida (por ejemplo, a los altavoces) tienen dependencias horizontales y verticales, que deberían ser conservadas durante el proceso de decorrelación. Por lo tanto, los coeficientes de mezclado son computados de tal modo que los canales que corresponden a diferentes grupos de altavoces no están mezclados entre sí.

[0232] Dependiendo del número de decorrelacionadores disponibles, o el nivel deseado de decorrelación, en cada grupo primero se mezclan juntos los pares verticales (entre la capa media y la capa superior o entre la capa media y la capa inferior). Segundo, los pares horizontales (entre izquierdo y derecho) o los pares verticales remanentes se mezclan entre sí. Por ejemplo, en el grupo tres, primero se mezclan entre sí los canales en el par
30 vertical izquierdo ("CH_M_L030" y "CH_L_L045"), y en el par vertical derecho ("CH_M_R030" y "CH_L_R045"), reduciendo de esta manera el número de decorrelacionadores requerido para este grupo de cuatro a dos. Si se desea reducir aún más el número de decorrelacionadores, el par horizontal obtenido es mezclado de manera descendente a un solo canal, y el número de decorrelacionadores requeridos para este grupo es reducido de cuatro a uno.

[0233] Con base en las reglas de mezclado presentadas, las tablas mencionadas más arriba (por ejemplo, mostradas en las figuras 19 a 23) se derivan para diferentes niveles de decorrelación deseada (o para diferentes niveles de complejidad de decorrelación deseada).

40 16. Compatibilidad con un convertidor de formato/renderizador externo secundario

[0234] En el caso en el que el decodificador SAOC (o, de forma más general, el decodificador de audio multicanal) se usa junto con un renderizador secundario/convertidor de formato externos, se pueden usar los siguientes cambios al concepto (procedimiento o aparato) propuesto:

45 - la matriz de renderización interna $\mathbf{R}$ (por ejemplo, del renderizador) es fijada con la identidad $\mathbf{R} = \mathbf{I}_{N_{\text{Objects}}}$ (cuando se usa un renderizador externo) o inicializada con los coeficientes de mezclado derivados de una configuración intermedia de representación (cuando se usa un convertidor de formato externo).

- el número de decorrelacionadores se reduce mediante el uso del procedimiento descrito en la sección 15 con la matriz de premezclado $\mathbf{M}_{\text{pre}}$ computada con base en la información de retroalimentación recibida del

50 renderizador/convertidor de formato (por ejemplo, $\mathbf{M}_{\text{pre}} = \mathbf{D}_{\text{convert}}$ en donde $\mathbf{D}_{\text{convert}}$ es la matriz de mezcla descendente usada dentro del convertidor de formato). Los canales que serán mezclados entre sí fuera del decodificador SAOC, son premezclados entre sí y suministrados en el mismo decorrelacionador dentro del decodificador SAOC.

55 [0235] Mediante el uso de un convertidor de formato externo, el renderizador interno SAOC prrenderizará a una configuración intermedia (por ejemplo, la configuración con el mayor número de altavoces).

[0236] Para concluir, en algunas formas de realización se usa una información acerca de cuáles de las señales de audio de salida son mezcladas entre sí en un renderizador o convertidor de formato externo para
60 determinar la matriz de premezclado $\mathbf{M}_{\text{pre}}$, de tal modo que la matriz de premezclado define una combinación de tales señales de entrada del decorrelacionador (del primer conjunto de señales de entrada del decorrelacionador) que son combinadas realmente en el renderizador externo. Así, la información recibida del renderizador/convertidor

de formato externo (que recibe las señales de audio de salida del decodificador multicanal) se usa para seleccionar o ajustar la matriz de premezclado (por ejemplo, cuando la matriz de renderización interna del decodificador de audio multicanal es fijada en la identidad, o inicializada con los coeficientes de mezclado derivados de una configuración de representación intermedia), y el renderizador/convertidor de formato externo es conectado para recibir las señales de audio de salida como se ha mencionado más arriba con respecto al decodificador de audio multicanal.

17. Flujo de bits

- 10 **[0237]** A continuación se describirá qué información de señalización adicional puede ser usada en un flujo de bits (o, de forma equivalente, en una representación codificada del contenido de audio). En formas de realización según la invención, el procedimiento de decorrelación puede ser señalizado en el flujo de bits para asegurar un nivel de calidad deseado. De esta manera, el usuario (o un codificador de audio) tiene más flexibilidad para seleccionar el procedimiento basado en el contenido. Para este propósito, la sintaxis de flujo de bits SAOC de MPEG puede ser extendida, por ejemplo, con dos bits para especificar el procedimiento de decorrelación usado y/o dos bits para especificar la configuración (o complejidad).

- [0238]** La figura 25 muestra una representación de sintaxis de elementos de flujo de bits "bsDecorrelationMethod" y "bsDecorrelationLevel", que puede ser agregada, por ejemplo, a una porción de flujo de bits "SAOCSpecificConfig()" o "SAOC3DSpecificConfig()". Como se puede observar en la figura 25, se pueden usar dos bits para el elemento de flujo de bits "bsDecorrelationMethod", y dos bits se pueden usar para el elemento de flujo de bits "bsDecorrelationLevel".

- [0239]** La figura 26 muestra, en forma de una tabla, una asociación entre valores del flujo de bits variable "bsDecorrelationMethod" y los diferentes procedimientos de decorrelación. Por ejemplo, tres diferentes procedimientos de decorrelación pueden ser señalizados por diferentes valores de dicho flujo de bits variable. Por ejemplo, una corrección de la covarianza de salida mediante el uso de señales decorrelacionadas, como se describe, por ejemplo, en la sección 14.3, puede ser señalizada como una de las opciones. Como otra opción, se puede señalar un procedimiento de ajuste de covarianza, por ejemplo, como se describe en la sección 14.4.1. Como otra opción, se puede señalar un procedimiento de compensación de energía, por ejemplo, como se describe en la sección 14.4.2. Por consiguiente, tres diferentes procedimientos para la reconstrucción de características de señales de las señales de audio de salida con base en las señales de audio renderizadas y las señales de audio decorrelacionadas pueden ser seleccionados dependiendo de un flujo de bits variable.

- 35 **[0240]** El modo de compensación de energía usa el procedimiento descrito en la sección 14.4.2, el modo de ajuste de covarianza limitada usa el procedimiento descrito en la sección 14.4.1, y el modo de ajuste de covarianza general usa el procedimiento descrito en la sección 14.3.

- [0241]** Haciendo referencia ahora a la figura 27, que muestra, a modo de tabla representativa, cómo diferentes niveles de decorrelación pueden ser señalizados por el flujo de bits variable "bsDecorrelationLevel", se describirá ahora un procedimiento para seleccionar la complejidad de decorrelación. En otras palabras, dicha variable puede ser evaluada por un decodificador de audio multicanal que comprende el decorrelacionador multicanal descrito en lo anterior para decidir qué complejidad de decorrelación se usa. Por ejemplo, dicho parámetro de flujo de bits puede señalar diferentes "niveles" de decorrelación que pueden ser designados con los valores: 0, 1, 2 y 3.

- [0242]** Un ejemplo de configuraciones de decorrelación (que pueden ser designados, por ejemplo, como "niveles" de decorrelación) se da en la tabla de la figura 27. La figura 27 muestra una representación en forma de tabla de un número de decorrelacionadores para diferentes "niveles" (por ejemplo, niveles de decorrelación) y configuraciones de salida. En otras palabras, la figura 27 muestra el número K de señales de entrada del decorrelacionador (del segundo conjunto de señales de entrada del decorrelacionador), que es usado por el decorrelacionador multicanal. Como se puede ver en la tabla de la figura 27, un número de decorrelacionadores (individuales) usado en el decorrelacionador multicanal es conmutado entre 11, 9, 7 y 5 para una configuración de salida 22.2, dependiendo de la cual un "nivel de decorrelación" es señalizado por el parámetro de flujo de bits "bsDecorrelationLevel". Para una configuración de salida 10.1, se realiza una selección entre 10, 5, 3 y 2 decorrelacionadores individuales, para una configuración 8.1, se realiza una selección entre 8, 4, 3 o 2 decorrelacionadores individuales, y para una configuración de salida 7.1, se realiza una selección entre 7, 4, 3 y 2 decorrelacionadores dependiendo del "nivel de decorrelación" señalizado por dicho parámetro de flujo de bits. En la configuración de salida 5.1, hay solo tres opciones válidas para los números de decorrelacionadores individuales, a saber 5, 3, o 2. Para la configuración de salida 2.1, hay solo una elección entre dos decorrelacionadores individuales (nivel de decorrelación 0) y un decorrelacionador individual (nivel de decorrelación 1).

- [0243]** Para resumir, el procedimiento de decorrelación puede determinarse en el lado del decodificador basado en la potencia computacional y un número de decorrelacionadores disponibles. Además, se puede realizar la selección del número de decorrelacionadores en el lado del codificador y se puede señalar mediante el uso de un

parámetro de flujo de bits.

[0244] Por consiguiente, tanto el procedimiento de cómo se aplican las señales de audio decorrelacionadas para obtener las señales de audio de salida, como la complejidad para la provisión de las señales decorrelacionadas pueden ser controlados desde el lado de un codificador de audio mediante el uso de los parámetros de flujo de bits mostrados en la figura 25 y definidos con mayor detalle en las figuras 26 y 27.

18. Campos de aplicación para el procesamiento de la invención

[0245] Debería señalarse que uno de los propósitos de los procedimientos introducidos es restaurar pistas de audio, que son de mayor importancia para la percepción humana de una escena de audio. Las formas de realización según la invención mejoran una exactitud de reconstrucción de nivel de energía y propiedades de correlación y por lo tanto aumentan la calidad de audio perceptual de la señal de salida final. Las formas de realización según la invención pueden ser aplicadas para un número arbitrario de canales de mezcla descendente/mezcla ascendente. Además, los procedimientos y aparatos descritos en esta invención pueden ser combinados con algoritmos de separación de fuentes paramétricas existentes. Las formas de realización según la invención permiten controlar la complejidad computacional del sistema fijando restricciones al número de funciones de decorrelacionador aplicadas. Las formas de realización según la invención pueden llevar a una simplificación de los algoritmos de construcción paramétrica basados en objetos como SAOC mediante la eliminación de una etapa de transcodificación de MPS.

19. Entorno de codificación/decodificación

[0246] A continuación se describirá un entorno de codificación/decodificación de audio en el cual se pueden aplicar conceptos según la presente invención.

[0247] Un sistema códec de audio 3D, en el cual se pueden usar los conceptos según la presente invención, se basa en un códec USAC de MPEG-D para la codificación de señales de canales y objetos para aumentar la eficacia para la codificación de una gran cantidad de objetos. Se ha adaptado la tecnología MPEG-SAOC. Tres tipos de renderizadores realizan las tareas de renderizar objetos a canales, renderizar canales a auriculares o renderizar canales a diferentes configuraciones de altavoces. Cuando las señales de objetos son transmitidas explícitamente o son codificadas paramétricamente mediante el uso de SAOC, la información de metadatos de objetos correspondiente es comprimida y multiplexada en la corriente de audio 3D.

[0248] Las figuras 28, 29 y 30 muestran los diferentes bloques algorítmicos del sistema de audio 3D.

[0249] La figura 28 muestra un diagrama de bloque esquemático de un codificador de audio de este tipo, y la figura 29 muestra un diagrama de bloque esquemático de un decodificador de audio de este tipo. En otras palabras, las figuras 28 y 29 muestran los diferentes bloques de algoritmos del sistema de audio 3D.

[0250] Haciendo referencia ahora a la figura 28, que muestra un diagrama de bloque esquemático de un codificador 2900 de audio 3D, se explicarán algunos detalles. El codificador 2900 comprende un prrenderizador/mezclador 2910 opcional, que recibe una o más señales 2912 de canales y una o más señales 2914 de objetos y proporciona, con base en éstas, una o más señales 2916 de canales así como una o más señales 2918, 2920 de objetos. El codificador 2930 de audio comprende también un codificador USAC y opcionalmente un codificador 2940 SAOC. El codificador 2940 SAOC está configurado para proporcionar uno o más canales 2942 de transporte SAOC y una información 2944 lateral SAOC con base en uno o más objetos 2920 proporcionados al codificador SAOC. Además, el codificador 2930 USAC está configurado para recibir las señales 2916 de canales que comprenden canales y objetos 2910 prrenderizados del prrenderizador/mezclador, para recibir una o más señales 2918 de objetos del prrenderizador/mezclador 2910, y para recibir una o más señales 2942 de transporte SAOC e información 2944 lateral SAOC, y proporciona, con base en éstas, una representación 2932 codificada. Además, el codificador 2900 de audio comprende también un codificador 2950 de metadatos de objetos que está configurado para recibir metadatos 2952 de objetos (que pueden ser evaluados por el prrenderizador/mezclador 2910) y para codificar los metadatos de objetos para obtener metadatos 2954 de objetos codificados. Los metadatos codificados también son recibidos por el codificador 2930 USAC y usados para proporcionar la representación 2932 codificada.

[0251] Algunos detalles con respecto a los componentes individuales del codificador 2900 de audio serán descritos más abajo.

[0252] Haciendo referencia ahora a la figura 29, se describirá un decodificador 3000 de audio. El decodificador 3000 de audio está configurado para recibir una representación codificada 3010 y para proporcionar, con base en ésta, una señal 3012 de altavoz multicanal, señales 3014 de auriculares y/o señales 3016 de altavoces en un formato alternativo (por ejemplo, en un formato 5.1). El decodificador 3000 de audio comprende un decodificador 3020 USAC, que proporciona una o más señales 3022 de canales, una o más señales 3024 de objetos prrenderizadas, una o más señales 3026 de objetos, una o más señales 3028 de transporte SAOC, una

información 3030 lateral SAOC y una información 3032 de metadatos de objetos comprimida con base en la representación 3010 codificada. El decodificador 3000 de audio también comprende un renderizador 3040 de objetos, que está configurado para proporcionar una o más señales 3042 de objetos renderizados con base en una o más señales 3026 de objetos y una información 3044 de metadatos de objetos, en el que la información 3044 de metadatos de objetos es proporcionada por un decodificador 3050 de metadatos de objetos con base en la información 3032 de metadatos de objetos comprimida. El decodificador 3000 de audio comprende también, opcionalmente, un decodificador 3060 SAOC, que está configurado para recibir el canal 3028 de transporte SAOC y la información 3030 lateral SAOC, y para proporcionar, con base en éstas, una o más señales 3062 de objetos renderizados. El decodificador 3000 de audio comprende también un mezclador 3070, que está configurado para recibir las señales 3022 de canales, las señales 3024 de objetos prerrenderizados, las señales 3042 de objetos renderizados y las señales 3062 de objetos renderizados, y para proporcionar, con base en éstas, una pluralidad de señales de canales 3072 mezcladas, que pueden constituir, por ejemplo, las señales 3012 de altavoces multicanal. El decodificador 3000 de audio puede comprender también, por ejemplo, un renderizador 3080 binaural, que está configurado para recibir las señales 3072 de canales mezcladas y para proporcionar, con base en éstas, las señales 3014 de auriculares. Además, el decodificador 3000 de audio puede comprender una conversión 3090 de formato, que está configurada para recibir las señales de canales 3072 mezcladas y una información 3092 de diseño de reproducción y para proporcionar, con base en éstas, una señal 3016 de altavoz para una configuración de altavoz alternativa.

20 **[0253]** A continuación, se describirán algunos detalles con respecto a los componentes del codificador 2900 de audio y del decodificador 3000 de audio.

19.1. Prerrenderizador/mezclador

25 **[0254]** El prerrenderizador/mezclador 2910 puede ser usado opcionalmente para convertir una escena de canal más entrada de objeto en una escena de canal antes de la codificación. Funcionalmente, puede ser, por ejemplo, idéntica al renderizador/mezclador de objetos descrito más abajo.

30 **[0255]** La prerrenderización de objetos puede asegurar, por ejemplo, una entropía de señales determinística en la entrada del codificador que es básicamente independiente del número de señales de objetos simultáneamente activos.

[0256] Con la prerrenderización de objetos, no se requiere una transmisión de metadatos de objetos.

35 **[0257]** Las señales de objetos discretas son renderizadas al diseño del canal para el cual el codificador está configurado, los pesos de los objetos para cada canal son obtenidos de los metadatos de objetos 1952 asociados (OAM).

19.2. Códec núcleo de USAC

40 **[0258]** El códec núcleo 2930, 3020 para señales de canales de altavoces, señales de objetos discretas, señales de objetos de mezcla descendente y señales prerrenderizadas se basa en la tecnología MPEG-D USAC. Maneja la decodificación de la multitud de señales creando información de mapeo de canales y objetos basada en la información geométrica y semántica del canal de entrada y de la asignación de objetos. Esta información de mapeo describe cómo los canales de entrada y objetos son mapeados a los elementos de canales USAC (CPes, SCEs, LFEs) y la información correspondiente es transmitida al decodificador.

50 **[0259]** Todas las cargas útiles adicionales como los datos SAOC o los metadatos de objetos han sido pasados a través de los elementos de extensión y han sido considerados en el control de velocidad de los codificadores. La decodificación de objetos es posible de diferentes maneras, dependiendo de los requisitos de velocidad/distorsión y los requisitos de interactividad para el renderizador. Las siguientes variantes de codificación de objetos son posibles:

• Objetos prerrenderizados: las señales de objetos son prerrenderizadas y mezcladas con las señales de canales 22.2 antes de la codificación. La subsiguiente cadena de codificación ve 22.2 señales de canales.

• Formas de onda de objetos discretos: los objetos son aplicados como formas de onda monofónicas al codificador. El codificador usa elementos de canales únicos SCEs para transmitir los objetos además de las señales de canales. Los objetos decodificados son renderizados y mezclados en el lado del receptor. La información de metadatos de objetos comprimida es transmitida también al receptor/renderizador.

65 • Formas de onda de objetos paramétricos: las propiedades de objetos y su relación entre sí son descritas por medio de parámetros SAOC. La mezcla descendente de las señales de objetos es codificada con USAC. La información paramétrica es transmitida conjuntamente. El número de canales de mezcla descendente es elegido dependiendo del número de objetos y la velocidad de datos general. La información de metadatos de objetos comprimida es

transmitida al renderizador SAOC.

19.3. SAOC

- 5 **[0260]** El codificador 2940 SAOC y el decodificador 3060 SAOC para señales de objetos se basan en la tecnología SAOC de MPEG. El sistema es capaz de recrear, modificar y renderizar un número de objetos de audio con base en un número más pequeño de canales transmitidos y datos paramétricos adicionales (diferencias de niveles de objetos OLDs, correlaciones entre objetos IOCs, ganancias de mezcla descendente DMGs). Los datos paramétricos adicionales presentan una velocidad de datos significativamente menor que la requerida para transmitir
- 10 todos los objetos individualmente, haciendo que la decodificación sea muy eficaz. El codificador SAOC toma como entrada las señales de objetos/canales como formas de onda monofónicas y da salida a la información paramétrica (que está empaquetada en el flujo de bits de audio 3D 2932, 3010) y los canales de transporte SAOC (que son codificados mediante el uso de elementos de canales únicos y transmitidos). El decodificador 3000 SAOC reconstruye las señales de objetos/canales de los canales 3028 de transporte SAOC decodificados y la información
- 15 3030 paramétrica, y genera la escena de audio de salida con base en el diseño de reproducción, la información de metadatos de objetos descomprimida y opcionalmente la información de interacción del usuario.

19.4. Códec de metadatos de objetos

- 20 **[0261]** Para cada objeto, los metadatos asociados que especifican la posición geométrica y el volumen del objeto en espacio 3D son codificados eficazmente por cuantificación de las propiedades de los objetos en tiempo y espacio. Los metadatos 2954 3032 de objetos comprimidos cOAM, son transmitidos al receptor como información lateral.

25 19.5. Renderizador/mezclador de objetos

- [0262]** El renderizador de objetos utiliza los metadatos 3044 de objetos descomprimidos OAM para generar formas de ondas de objetos según el formato de reproducción dado. Cada objeto es renderizado a determinados canales de salida según sus metadatos. La salida de este bloque resulta de la suma de los resultados parciales.
- 30 **[0263]** Si se decodifican tanto el contenido basado en los canales como los objetos discretos/paramétricos, las formas de onda basadas en canales y las formas de onda de objetos renderizados son mezcladas antes de la salida de las formas de ondas resultantes (o antes de suministrarlas a un módulo post-procesador como el renderizador binaural o el módulo del renderizador de altavoces).

35 19.6. Renderizador binaural

- [0264]** El módulo 3080 renderizador binaural produce una mezcla descendente binaural del material de audio multicanal, de tal modo que cada canal de entrada es representado por una fuente de sonido virtual. El
- 40 procesamiento es conducido en forma de marcos en el dominio QMF. La binauralización se basa en respuestas de impulsos de ambientes binaurales medidos.

19.7. Renderizador de altavoces/conversión de formato

- 45 **[0265]** El renderizador 3090 de altavoces convierte entre la configuración de canal transmitida y el formato de reproducción deseado. Por eso es denominado "convertidor de formato" a continuación. El convertidor de formato realiza conversiones a números más bajos de canales de salida, es decir, crea mezclas descendentes. El sistema genera automáticamente matrices de mezcla descendente optimizadas para la combinación dada de formatos de entrada y salida y aplica estas matrices en un proceso de mezcla descendente. El convertidor de formato permite
- 50 configuraciones de altavoces estándar así como configuraciones al azar con posiciones de altavoces no estándar.

[0266] La figura 30 muestra un diagrama de bloque esquemático de un convertidor de formato. En otras palabras, la figura 30 muestra la estructura del convertidor de formato.

- 55 **[0267]** Como se puede ver, el convertidor 3100 de formato recibe las señales 3110 de salida del mezclador, por ejemplo, las señales 3072 de canales mezcladas, y proporciona señales 3112 de altavoces, por ejemplo, las señales 3016 de altavoces. El convertidor 3120 de formato comprende un proceso de mezcla descendente en el dominio QMF y un configurador 3130 de mezcla descendente, en el que el configurador de mezcla descendente proporciona información de configuración para el proceso 3020 de mezcla descendente con base en una
- 60 información 3032 de diseño de salida del mezclador y una información 3034 de diseño de reproducción.

19.8. Observaciones generales

- [0268]** Además, debería señalarse que los conceptos descritos en esta invención, por ejemplo, el
- 65 decodificador 100 de audio, el codificador 200 de audio, el decorrelacionador 600 multicanal, el decodificador 700 de

audio multicanal, el codificador 800 de audio o el decodificador 1550 de audio se pueden usar dentro del codificador 2900 de audio y/o dentro del decodificador 3000 de audio. Por ejemplo, los codificadores/decodificadores de audio mencionados más arriba se pueden usar como parte del codificador 2940 SAOC y/o como parte del decodificador 3060 SAOC. Sin embargo, los conceptos mencionados más arriba también pueden ser usados en otras posiciones del decodificador 3000 de audio 3D y/o del codificador 2900 de audio.

[0269] Naturalmente, los procedimientos mencionados más arriba también pueden ser usados en conceptos para la codificación o decodificación de información de audio según las figuras 28 y 29.

10 20. Formas de realización adicionales

20.1 Introducción

[0270] A continuación se describirán otras formas de realización según la presente invención.

[0271] La figura 31 muestra un diagrama de bloque esquemático de un procesador de mezcla descendente, según una forma de realización de la presente invención.

[0272] El procesador 3100 de mezcla descendente comprende un desmezclador 3110, un renderizador 3120, un combinador 3130 y un decorrelacionador 3140 multicanal. El renderizador proporciona señales de audio renderizadas Y_{dry} al combinador 3130 y al decorrelacionador 3140 multicanal. El decorrelacionador multicanal comprende un pre-mezclador 3150, que recibe las señales de audio renderizadas (que pueden ser consideradas como un primer conjunto de señales de entrada del decorrelacionador) y proporciona, con base en éstas, un segundo conjunto de señales de entrada del decorrelacionador premezcladas a un núcleo de decorrelacionador 3160. El núcleo de decorrelacionador proporciona un primer conjunto de señales de salida del decorrelacionador con base en el segundo conjunto de señales de entrada del decorrelacionador para el uso por un post-mezclador 3170. El post-mezclador post-mezcla (o mezcla ascendente) las señales de salida del decorrelacionador proporcionadas por el núcleo del decorrelacionador 3160, para obtener un segundo conjunto de señales de salida del decorrelacionador post-mezclado, que se proporciona al combinador 3130.

[0273] El renderizador 3130 puede aplicar, por ejemplo, una matriz R para la renderización, el pre-mezclador puede aplicar, por ejemplo, una matriz M_{pre} para el premezclado, el post-mezclador puede aplicar, por ejemplo, una matriz M_{post} para el post-mezclado, y el combinador puede aplicar, por ejemplo, una matriz P para la combinación.

[0274] Debería señalarse que el procesador de mezcla descendente 3100, o componentes individuales o funcionalidades del mismo, se pueden usar en los decodificadores de audio descritos en esta invención. Además, debería señalarse que el procesador de mezcla descendente puede ser suplementado por cualquiera de las características y funcionalidades descritas en esta invención.

40 20.2 Procesamiento en 3D SAOC

[0275] Se aplica el banco de filtro híbrido descrito en ISO/IEC 23003-1:2007. La descuantificación de los parámetros DMG, OLD, IOC sigue las mismas reglas que las definidas en 7.1.2 de ISO/IEC 23003-2:2010.

45 20.2.1 Señales y parámetros

[0276] Las señales de audio son definidas para cada intervalo de tiempo n y cada sub-banda híbrida k . Los parámetros 3D SAOC correspondientes son definidos para cada parámetro de intervalo de tiempo l y procesamiento por m . El mapeo subsiguiente entre el híbrido y el dominio del parámetro es especificado por la Tabla A.31 de ISO/IEC 23003-1:2007. Por lo tanto, todos los cálculos se realizan con respecto a determinados índices de tiempo/banda y las dimensionalidades correspondientes están implícitas para cada variable introducida.

[0277] Los datos disponibles en el decodificador 3D SAOC consisten en la señal de mezcla descendente multicanal X , la matriz de covarianza E , la matriz de renderización R y la matriz de mezcla descendente D .

55 20.2.1.1 Parámetros de objetos

[0278] La matriz de covarianza E de tamaño $N \times N$ con elementos $e_{i,j}$ representa una aproximación de la matriz de covarianza de señales original $E \approx SS^*$ y es obtenida de los parámetros OLD y IOC como:

$$e_{i,j} = \sqrt{OLD_i OLD_j} IOC_{i,j}.$$

[0279] Aquí, los parámetros de objetos descuantificados son obtenidos como:

$$OLD_i = \mathbf{D}_{OLD}(i, l, m), \quad IOC_{i,j} = \mathbf{D}_{IOC}(i, j, l, m).$$

20.2.1.3 Matriz de mezcla descendente

[0280] La matriz de mezcla descendente $\mathbf{D}$ aplicada a las señales de audio de entrada $\mathbf{S}$ determina la señal de mezcla descendente como $\mathbf{X} = \mathbf{D} \mathbf{S}$. La matriz de mezcla descendente $\mathbf{D}$ de tamaño $N_{dmx} \times N$ es obtenida como:

$$\mathbf{D} = \mathbf{D}_{dmx} \mathbf{D}_{premix}.$$

10 [0281] La matriz $\mathbf{D}_{dmx}$ y la matriz $\mathbf{D}_{premix}$ tienen diferentes tamaños dependiendo del modo de procesamiento. La matriz $\mathbf{D}_{dmx}$ es obtenida de los parámetros DMG como:

$$15 \quad d_{i,j} = \begin{cases} 0 & , \text{si ningún dato de DMG para } (ij) \text{ está presente en el flujo de bits} \\ 10^{0.05 DMG_{i,j}} & , \text{de otro modo} \end{cases}$$

20 [0282] Aquí, los parámetros de mezcla descendente descuantificados son obtenidos como:

$$DMG_{i,j} = \mathbf{D}_{DMG}(i, j, l).$$

20.2.1.3.1 Modo directo

25 [0283] En el caso de modo directo, no se usa un premezclado. La matriz $\mathbf{D}_{premix}$ tiene el tamaño $N \times N$ y está dada por: $\mathbf{D}_{premix} = \mathbf{I}$. La matriz $\mathbf{D}_{dmx}$ tiene el tamaño $N_{dmx} \times N$ y es obtenida a partir de los parámetros DMG según 20.2.1.3.

20.2.1.3.2 Modo de premezclado

30 [0284] En el caso del modo de premezclado la matriz $\mathbf{D}_{premix}$ tiene el tamaño $(N_{ch} + N_{premix}) \times N$ y está dada por:

$$35 \quad \mathbf{D}_{premix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix} \quad \text{matriz de premezclado } \mathbf{A} \text{ de tamaño } N_{premix} \times N_{obj} \text{ es recibida como una}$$

entrada al decodificador 3D SAOC, del renderizador de objetos.

[0285] La matriz $\mathbf{D}_{dmx}$ tiene el tamaño $N_{dmx} \times (N_{ch} + N_{premix})$ y es obtenida a partir de los parámetros DMG según 20.2.1.3

2.2.1.2 Matriz de renderización

[0286] La matriz de renderización $\mathbf{R}$ aplicada a las señales de audio de entrada $\mathbf{S}$ determina la salida renderizada objetivo como $\mathbf{Y} = \mathbf{R} \mathbf{S}$. La matriz de renderización $\mathbf{R}$ de tamaño $N_{out} \times N$ está dada por:

$$45 \quad \mathbf{R} = \begin{pmatrix} \mathbf{R}_{ch} & \mathbf{R}_{obj} \end{pmatrix},$$

en donde $\mathbf{R}_{ch}$ de tamaño $N_{out} \times N_{ch}$ representa la matriz de renderización asociada con los canales de entrada y $\mathbf{R}_{obj}$ de tamaño $N_{out} \times N_{obj}$ representa la matriz de renderización asociada con los objetos de entrada.

20.2.1.4 La matriz de covarianza de salida objetivo

[0287] La matriz de covarianza $\mathbf{C}$ de tamaño $N_{out} \times N_{out}$ con elementos $c_{i,j}$ representa una aproximación de

la matriz de covarianza de señal de salida objetivo $\mathbf{C} \approx \mathbf{Y} \mathbf{Y}^*$ y es obtenida a partir de la matriz de covarianza $\mathbf{E}$ y la matriz de renderización $\mathbf{R}$:

$$\mathbf{C} = \mathbf{R} \mathbf{E} \mathbf{R}^*.$$

5

20.2.2 Decodificación

[0288] Se describe el procedimiento para obtener una señal de salida mediante el uso de parámetros 3D SAOC e información de renderización. El decodificador 3D SAOC manda, por ejemplo, y consiste en el procesador de parámetros 3D SAOC y el procesador de mezcla descendente 3D SAOC.

20.2.2.1 Procesador de mezcla descendente

[0289] La señal de salida del procesador de mezcla descendente (representada en el dominio QMF híbrido) es suministrada en el banco de filtro de síntesis correspondiente como se describe en ISO/IEC 23003–1:2007 dando la salida final del decodificador 3D SAOC. Una estructura detallada del procesador de mezcla descendente se ilustra en la figura 31

[0290] La señal de salida $\hat{\mathbf{Y}}$ es computada a partir de la señal de mezcla descendente multicanal $\mathbf{X}$ y la señal multicanal decorrelacionada $\mathbf{X}_d$ como:

$$\hat{\mathbf{Y}} = \mathbf{P}_{\text{dry}} \mathbf{R} \mathbf{U} \mathbf{X} + \mathbf{P}_{\text{wet}} \mathbf{M}_{\text{post}} \mathbf{X}_d,$$

en donde $\mathbf{U}$ representa la matriz de desmezclado paramétrica y se define en 20.2.2.1.1 y 20.2.2.1.2.

25

[0291] La señal multicanal decorrelacionada $\mathbf{X}_d$ es computada según 20.2.3.

$$\mathbf{X}_d = \text{decorrFunc}(\mathbf{M}_{\text{pre}} \mathbf{Y}_{\text{dry}})$$

[0292] La matriz de mezclado $\mathbf{P} = \begin{pmatrix} \mathbf{P}_{\text{dry}} & \mathbf{P}_{\text{wet}} \end{pmatrix}$ es descrita en 20.2.3. Las matrices $\mathbf{M}_{\text{pre}}$ para diferente configuración de salida se dan en las figuras 19 a 23 y las matrices $\mathbf{M}_{\text{post}}$ se obtienen mediante el uso de la siguiente ecuación:

$$\mathbf{M}_{\text{post}} = \mathbf{M}_{\text{pre}}^* (\mathbf{M}_{\text{pre}} \mathbf{M}_{\text{pre}}^*)^{-1}$$

[0293] El modo de decodificación es controlado por el elemento de flujo de bits bsNumSaocDmxObjetos, como se muestra en la figura 32.

20.2.2.1.1 Modo de decodificación combinado

[0294] En el caso del modo de decodificación combinado la matriz de desmezclado paramétrica $\mathbf{U}$ está dada por:

$$\mathbf{U} = \mathbf{E} \mathbf{D}^* \mathbf{J}$$

[0295] La matriz $\mathbf{J}$ de tamaño $N_{\text{dmx}} \times N_{\text{dmx}}$ está dada por $\mathbf{J} \approx \mathbf{\Lambda}^{-1}$ con $\mathbf{\Lambda} = \mathbf{D} \mathbf{E} \mathbf{D}^*$.

20.2.2.1.2 Modo de decodificación independiente

[0296] En el caso del modo de decodificación independiente la matriz de desmezclado $\mathbf{U}$ está dada por:

50

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_{\text{ch}} & 0 \\ 0 & \mathbf{U}_{\text{obj}} \end{pmatrix},$$

en donde: $\mathbf{U}_{\text{ch}} = \mathbf{E}_{\text{ch}} \mathbf{D}_{\text{ch}}^* \mathbf{J}_{\text{ch}}$ y $\mathbf{U}_{\text{obj}} = \mathbf{E}_{\text{obj}} \mathbf{D}_{\text{obj}}^* \mathbf{J}_{\text{obj}}$.

[0297] La matriz de covarianza basada en canales $\mathbf{E}_{ch}$ de tamaño $N_{ch} \times N_{ch}$ y la matriz de covarianza basada en objetos $\mathbf{E}_{obj}$ de tamaño $N_{obj} \times N_{obj}$ son obtenidas a partir de la matriz de covarianza $\mathbf{E}$ seleccionando solo los bloques diagonales correspondientes:

5

$$\mathbf{E} = \begin{pmatrix} \mathbf{E}_{ch} & \mathbf{E}_{ch,obj} \\ \mathbf{E}_{obj,ch} & \mathbf{E}_{obj} \end{pmatrix}$$

en el que la matriz $\mathbf{E}_{ch,obj} = (\mathbf{E}_{obj,ch})^*$ representa la matriz de covarianza cruzada entre los canales de entrada y los objetos de entrada y no se requiere que sean calculados.

10

[0298] La matriz de mezcla descendente basada en canales $\mathbf{D}_{ch}$ de tamaño $N_{ch}^{dm \times} \times N_{ch}$ y la matriz de mezcla descendente basada en objetos $\mathbf{D}_{obj}$ de tamaño $N_{obj}^{dm \times} \times N_{obj}$ son obtenidas a partir de la matriz de mezcla descendente $\mathbf{D}$ seleccionando solo los bloques diagonales correspondientes:

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{ch} & 0 \\ 0 & \mathbf{D}_{obj} \end{pmatrix}.$$

15

[0299] La matriz $\mathbf{J}_{ch} \approx (\mathbf{D}_{ch} \mathbf{E}_{ch} \mathbf{D}_{ch}^*)^{-1}$ de tamaño $N_{ch}^{dm \times} \times N_{ch}^{dm \times}$ es derivada por consiguiente a 20.2.2.1.4 para $\mathbf{\Lambda} = \mathbf{D}_{ch} \mathbf{E}_{ch} \mathbf{D}_{ch}^*$.

[0300] La matriz $\mathbf{J}_{obj} \approx (\mathbf{D}_{obj} \mathbf{E}_{obj} \mathbf{D}_{obj}^*)^{-1}$ de tamaño $N_{obj}^{dm \times} \times N_{obj}^{dm \times}$ es derivada por consiguiente a 20.2.2.1.4 para $\mathbf{\Lambda} = \mathbf{D}_{obj} \mathbf{E}_{obj} \mathbf{D}_{obj}^*$.

20.2.2.1.4 Cálculo de la matriz $\mathbf{J}$

[0301] La matriz $\mathbf{J} \approx \mathbf{\Lambda}^{-1}$ se calcula mediante el uso de la siguiente ecuación:

$$\mathbf{J} = \mathbf{V} \mathbf{\Lambda}^{inv} \mathbf{V}^*.$$

[0302] Aquí el vector singular $\mathbf{V}$ de la matriz $\mathbf{\Lambda}$ se obtiene mediante el uso de la siguiente ecuación característica:

30

$$\mathbf{V} \mathbf{\Lambda} \mathbf{V}^* = \mathbf{\Lambda}.$$

[0303] La inversa regularizada $\mathbf{\Lambda}^{inv}$ de la matriz de valor singular diagonal $\mathbf{\Lambda}$ es computada como:

$$\lambda_{i,j}^{inv} = \begin{cases} \frac{1}{\lambda_{i,j}}, & \text{si } i = j \text{ y } \lambda_{i,j} \geq T_{reg}^{\Lambda} \\ 0, & \text{de otro modo} \end{cases}$$

35

[0304] La regularización relativa escalar T_{reg}^{Λ} es determinada mediante el uso del umbral absoluto T_{reg} y el valor máximo de $\mathbf{\Lambda}$ como:

$$T_{reg}^{\Lambda} = \max(\lambda_{i,i}) T_{reg}, \quad T_{reg} = 10^{-2}.$$

40 20.2.3. Decorrelación

[0305] Las señales decorrelacionadas $\mathbf{X}_d$ son creadas a partir del decorrelacionador descrito en 6.6.2 de ISO/IEC 23003-1:2007, con bsDecorrConfig == 0 y un índice de decorrelacionador, X , según las tablas en las figuras 19 a 24. Por lo tanto, la $decorrFunc()$ denota el proceso de decorrelación:

$$\mathbf{X}_d = decorrFunc(\mathbf{M}_{pre} \mathbf{Y}_{dry})$$

20.2.4. Matriz de mezclado $\mathbf{P}$

5 **[0306]** El cálculo de la matriz de mezclado $\mathbf{P} = \begin{pmatrix} \mathbf{P}_{dry} & \mathbf{P}_{wet} \end{pmatrix}$ es controlado por el elemento de flujo de bits bsDecorrelaciónMétodo. La matriz $\mathbf{P}$ tiene el tamaño $N_{out} \times 2N_{out}$ y las $\mathbf{P}_{dry}$ y $\mathbf{P}_{wet}$ tienen ambas el tamaño $N_{out} \times N_{out}$.

20.2.4.1 Modo de compensación de energía

10 **[0307]** El modo de compensación de energía usa señales decorrelacionadas para compensar la pérdida de energía en la reconstrucción paramétrica. Las matrices de mezclado $\mathbf{P}_{dry}$ y $\mathbf{P}_{wet}$ están dadas por:

$$\mathbf{P}_{dry} = \mathbf{I}$$

15

$$p_{i,j}^{wet} = \begin{cases} \sqrt{\min\left(\lambda_{Dec}, \max\left(0, \frac{\mathbf{C}(i,i) - \mathbf{E}_Y^{dry}(i,i)}{\max(\epsilon, \mathbf{E}_Y^{wet}(i,i))}\right)\right)} & i = j, \\ 0 & i \neq j. \end{cases}$$

en donde $\lambda_{Dec} = 4$ es una constante usada para limitar la cantidad de componente decorrelacionado agregado a las señales de salida.

20

20.2.4.2 Modo de ajuste de covarianza limitada

[0308] El modo de ajuste de covarianza limitada asegura que la matriz de covarianza de las señales decorrelacionadas mezcladas $\mathbf{P}_{wet} \mathbf{Y}_{dry}$ se aproxima a la diferencia de la matriz de covarianza Δ_E : $\mathbf{P}_{wet} \mathbf{E}_Y^{wet} \mathbf{P}_{wet}^* \approx \Delta_E$. Las matrices de mezclado $\mathbf{P}_{dry}$ y $\mathbf{P}_{wet}$ son definidas mediante el uso de las siguientes ecuaciones:

25

$$\mathbf{P}_{dry} = \mathbf{I},$$

$$\mathbf{P}_{wet} = \left(\mathbf{V}_1 \sqrt{\mathbf{Q}_1} \mathbf{V}_1^* \right) \left(\mathbf{V}_2 \sqrt{\mathbf{Q}_2^{inv}} \mathbf{V}_2^* \right),$$

30

en el que la inversa regularizada $\mathbf{Q}_2^{inv}$ de la matriz de valor singular diagonal $\mathbf{Q}_2$ es computada como:

$$\mathbf{Q}_2^{inv}(i,j) = \begin{cases} \frac{1}{\mathbf{Q}_2(i,j)}, & \text{si } i = j \text{ y } \mathbf{Q}_2(i,j) \geq T_{reg}^A, \\ 0, & \text{de otro modo,} \end{cases}$$

35 **[0309]** La regularización relativa escalar T_{reg}^A es determinada mediante el uso del umbral absoluto T_{reg} y valor máximo de $\mathbf{Q}_2^{inv}$ como:
 $T_{reg}^A = \max(\mathbf{Q}_2^{inv}(i,i)) T_{reg}, T_{reg} = 10^{-2}$.

[0310] La matriz Δ_E es descompuesta mediante el uso de la descomposición de valor singular como:

40

$$\Delta_E = \mathbf{V}_1 \mathbf{Q}_1 \mathbf{V}_1^*$$

[0311] La matriz de covarianza de las señales decorrelacionadas $\mathbf{E}_Y^{\text{wet}}$ también se expresa mediante el uso de la descomposición de valor singular:

$$\mathbf{E}_Y^{\text{wet}} = \mathbf{V}_2 \mathbf{Q}_2 \mathbf{V}_2^*$$

5

20.2.4.3. Modo de ajuste de covarianza general

[0312] El modo de ajuste de covarianza general asegura que la matriz de covarianza de las señales de salida finales $\hat{\mathbf{Y}}$ ($\mathbf{E}_{\hat{\mathbf{Y}}} = \hat{\mathbf{Y}} \hat{\mathbf{Y}}^*$) se aproxima a la matriz de covarianza objetivo: $\mathbf{E}_{\hat{\mathbf{Y}}} \approx \mathbf{C}$. La matriz de mezclado $\mathbf{P}$ se define mediante el uso de la siguiente ecuación:

$$\mathbf{P} = \left(\mathbf{V}_1 \sqrt{\mathbf{Q}_1} \mathbf{V}_1^* \right) \mathbf{H} \left(\mathbf{V}_2 \sqrt{\mathbf{Q}_2^{\text{inv}}} \mathbf{V}_2^* \right),$$

en el que la inversa regularizada $\mathbf{Q}_2^{\text{inv}}$ de la matriz de valor singular diagonal $\mathbf{Q}_2$ es computada como:

$$\mathbf{Q}_2^{\text{inv}}(i, j) = \begin{cases} \frac{1}{\mathbf{Q}_2(i, j)}, & \text{si } i = j \text{ y } \mathbf{Q}_2(i, j) \geq T_{\text{reg}}^{\Lambda}, \\ 0, & \text{de otro modo,} \end{cases}$$

15

[0313] La regularización relativa escalar T_{reg}^{Λ} se determina mediante el uso del umbral absoluto T_{reg} y el valor máximo de $\mathbf{Q}_2^{\text{inv}}$ como:

$$T_{\text{reg}}^{\Lambda} = \max(\mathbf{Q}_2^{\text{inv}}(i, i)) T_{\text{reg}}, \quad T_{\text{reg}} = 10^{-2}.$$

20

[0314] La matriz de covarianza objetivo $\mathbf{C}$ es descompuesta mediante el uso de la descomposición de valor singular como:

$$\mathbf{C} = \mathbf{V}_1 \mathbf{Q}_1 \mathbf{V}_1^*$$

25

[0315] La matriz de covarianza de las señales combinadas $\mathbf{E}_Y^{\text{com}}$ también se expresa mediante el uso de la descomposición de valor singular:

$$\mathbf{E}_Y^{\text{com}} = \mathbf{V}_2 \mathbf{Q}_2 \mathbf{V}_2^*$$

30

[0316] La matriz $\mathbf{H}$ representa una matriz de factor de ponderación prototipo de tamaño $(N_{\text{out}} \times 2 N_{\text{out}})$ y está dada por la siguiente ecuación:

$$\mathbf{H} = \begin{pmatrix} 1/\sqrt{2} & 0 & \dots & 0 & 1/\sqrt{2} & 0 & \dots & 0 \\ 0 & 1/\sqrt{2} & \dots & 0 & 0 & 1/\sqrt{2} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1/\sqrt{2} & 0 & 0 & \dots & 1/\sqrt{2} \end{pmatrix}.$$

35

20.2.4.4 Matrices de covarianza introducidas

[0317] La matriz Δ_E representa la diferencia entre la matriz de covarianza de salida objetivo $\mathbf{C}$ y la matriz de covarianza $\mathbf{E}_Y^{\text{dry}}$ de las señales paramétricamente reconstruidas y está dada por:

$$\Delta_E = \mathbf{C} - \mathbf{E}_Y^{\text{dry}}.$$

40

[0318] La matriz $\mathbf{E}_Y^{dry}$ representa la matriz de covarianza de las señales paramétricamente estimadas $\mathbf{E}_Y^{dry} \approx \mathbf{Y}_{dry} \mathbf{Y}_{dry}^*$ y se define mediante el uso de la siguiente ecuación:

$$\mathbf{E}_Y^{dry} = \mathbf{R} \mathbf{U} \mathbf{U}^* \mathbf{R}^*$$

5

[0319] La matriz $\mathbf{E}_Y^{wet}$ representa la matriz de covarianza de las señales decorrelacionadas $\mathbf{E}_Y^{wet} \approx \mathbf{Y}_{wet} \mathbf{Y}_{wet}^*$ y se define mediante el uso de la siguiente ecuación:

$$\mathbf{E}_Y^{wet} = \mathbf{M}_{post} \left[\text{matdiag}(\mathbf{M}_{pre} \mathbf{E}_Y^{dry} \mathbf{M}_{pre}^*) \right] \mathbf{M}_{post}^*$$

10

[0320] Considerando la señal $\mathbf{Y}_{com}$ que consiste en la combinación de las señales paramétricas estimadas y decorrelacionadas:

$$\mathbf{Y}_{com} = \begin{pmatrix} \mathbf{Y}_{dry} \\ \mathbf{Y}_{wet} \end{pmatrix}$$

15

la matriz de covarianza de $\mathbf{Y}_{com}$ es definida por la siguiente ecuación:

$$\mathbf{E}_Y^{com} = \begin{pmatrix} \mathbf{E}_Y^{dry} & 0 \\ 0 & \mathbf{E}_Y^{wet} \end{pmatrix}$$

20 21. Alternativas de implementación

[0321] Aunque se han descrito algunos aspectos en el contexto de un aparato, está claro que estos aspectos también representan una descripción del procedimiento correspondiente, en donde un bloque o dispositivo corresponde a una etapa del procedimiento o una característica de una etapa del procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa del procedimiento también representan una descripción de un bloque correspondiente o elemento o característica de un aparato correspondiente. Algunas o todas las etapas del procedimiento pueden ser ejecutadas por (o mediante el uso de) un aparato de hardware como, por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas formas de realización, una o más de las etapas del procedimiento más importantes pueden ser ejecutadas por un aparato de este tipo.

30

[0322] La señal de audio codificada de la invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión alámbrico tal como Internet.

35 [0323]

Dependiendo de ciertos requisitos de implementación, formas de realización de la invención pueden ser implementadas en hardware o en software. La implementación puede ser realizada mediante el uso de un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blu-Ray, un CD, una memoria ROM, PROM, EPROM, EEPROM o una memoria FLASH, que tienen señales de control que se pueden leer electrónicamente almacenadas en éstas, que cooperan (o son capaces de cooperar) con un sistema informático programable de tal modo que se realiza el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador. Algunas formas de realización según la invención comprenden un soporte de datos que tiene señales de control que se pueden leer electrónicamente, que son capaces de cooperar con un sistema informático programable, de tal modo que se realiza uno de los procedimientos descritos en esta invención. En general, las formas de realización de la presente invención pueden ser implementadas como un producto de un programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los procedimientos cuando el producto del programa informático se ejecuta en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un soporte que se puede leer en una máquina. Otras formas de realización comprenden el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención, almacenado en un soporte que se puede leer en una máquina. En otras palabras, una forma de realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los procedimientos descritos en esta invención, cuando el programa informático se ejecuta en un ordenador.

50

[0324] Una forma de realización adicional es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención. El soporte de datos, el medio de almacenamiento digital o el medio grabado son típicamente tangibles y/o no transitorios.

[0325] Una forma de realización adicional es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención. El flujo de datos o la secuencia de señales pueden ser configurados, por ejemplo, para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet. Una forma de realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo de lógica programable, configurado o adaptado para realizar uno de los procedimientos descritos en esta invención. Una forma de realización adicional comprende un ordenador que tiene instalado el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención.

[0326] Una forma de realización adicional según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, de forma electrónica u óptica) un programa informático para llevar a cabo uno de los procedimientos descritos en esta invención a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa informático al receptor. En algunas formas de realización, se puede usar un dispositivo de lógica programable (por ejemplo, una matriz de puertas programable de campo) para llevar a cabo algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas formas de realización, una matriz de puertas programable de campo puede cooperar con un microprocesador para llevar a cabo uno de los procedimientos descritos en esta invención. En general, los procedimientos son realizados preferentemente por cualquier aparato de hardware. Las formas de realización anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las configuraciones y los detalles descritos en esta invención resultarán evidentes para otros expertos en la técnica. Se pretende, por lo tanto, estar limitado solamente por el alcance de las reivindicaciones de patente adjuntas y no por los detalles específicos presentados a modo de descripción y explicación de las formas de realización en esta invención.

Referencias

[0327]

[BCC] C. Faller y F. Baumgarte, "Binaural Cue Codification – Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003.

[Blauert] J. Blauert, "Spatial Hearing – The Psychophysics of Human Sound Localization", Revised Edition, The MIT Press, Londres, 1997.

[JSC] C. Faller, "Parametric Joint-Codification of Audio Sources", 120th AES Convention, París, 2006.

[ISS1] M. Parvaix y L. Girin: "Informed Source Separation of underdetermined instantaneous Stereo Mixtures using Source Index Embedding", IEEE ICASSP, 2010.

[ISS2] M. Parvaix, L. Girin, J.-M. Brossier: "A watermarking-based method for informed source separation of audio signals with a single sensor", IEEE Transactions on Audio, Speech and Language Processing, 2010.

[ISS3] A. Liutkus y J. Pintel y R. Badeau y L. Girin y G. Richard: "Informed source separation through spectrogram codification and data embedding", Signal Processing Journal, 2011.

[ISS4] A. Ozerov, A. Liutkus, R. Badeau, G. Richard: "Informed source separation: source codification meets source separation", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011.

[ISS5] S. Zhang y L. Girin: "An Informed Source Separation System for Speech Signals", INTERSPEECH, 2011.

[ISS6] L. Girin y J. Pintel: "Informed Audio Source Separation from Compressed Linear Stereo Mixtures", AES 42nd International Conference: Semantic Audio, 2011.

[MPS] ISO/IEC, "Information technology – MPEG audio technologies – Part 1: MPEG Surround," ISO/IEC JTC1/SC29/WG11 (MPEG) international Standard 23003-1:2006.

[OCD] J. Vilkamo, T. Bäckström, y A. Kuntz. "Optimized covariance domain framework for time-frequency processing of spatial audio", Journal of the Audio Engineering Society, 2013. en prensa.

[SAOC1] J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC To SAOC – Recent Developments in Parametric Codification of Spatial Audio", 22nd Regional UK AES Conference, Cambridge, UK, Abril 2007.

[SAOC2] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens,
5 E. Schuijers y W. Oomen: " Spatial Audio Object Codification (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Codification", 124th AES Convention, Ámsterdam 2008.

[SAOC] ISO/IEC, "MPEG audio technologies – Part 2: Spatial Audio Object Codification (SAOC)," ISO/IEC JTC1/SC29/WG11 (MPEG) International Standard 23003–2.

10

Patente internacional n.º WO/2006/026452, "MULTICHANNEL DECORRELATION IN SPATIAL AUDIO CODING"(decorrelación multicanal en codificación de audio especial) publicada el 9 de marzo de 2006.

REIVINDICACIONES

1. Un decorrelacionador multicanal (140; 600; 1590; 1700) para proporcionar una pluralidad de señales decorrelacionadas (142, 144; 612a- 612n'; 1592a-1592n; 1712a-1712n) en base a una pluralidad de señales de entrada del decorrelacionador (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n) que comprende señales de canal de audio, en el que el decorrelacionador multicanal está configurado para premezclar un primer conjunto $\hat{\mathbf{Z}}$ (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n; $\hat{\mathbf{Z}}$) de señales de entrada del decorrelacionador N en un segundo conjunto $\hat{\mathbf{Z}}_{mix}$ (622a-622k; 1722a-1722k; $\hat{\mathbf{Z}}_{mix}$) de señales de entrada del decorrelacionador K, en el que $K < N$; en el que el decorrelacionador multicanal está configurado para proporcionar un primer conjunto (632a-632k'; 1732a-1732k) de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) de señales de salida del decorrelacionador N', en el que $N' > K'$,
- 15 en el que el decorrelacionador multicanal está configurado para premezclar el primer conjunto $\hat{\mathbf{Z}}$ de señales de entrada del decorrelacionador N en el segundo conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de entrada del decorrelacionador K mediante el uso de una matriz premezclada $\mathbf{M}_{pre}$ de acuerdo a:

$$\hat{\mathbf{Z}}_{mix} = \mathbf{M}_{pre} \hat{\mathbf{Z}}$$

- en el que el decorrelacionador multicanal está configurado para obtener el primer conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de salida del decorrelacionador K' en base al segundo conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de entrada del decorrelacionador K, y en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de salida del decorrelacionador K' en el segundo conjunto W de señales de salida del decorrelacionador N' mediante el uso de una matriz de post-mezclado $\mathbf{M}_{post}$ según:

$$\mathbf{W} = \mathbf{M}_{post} \hat{\mathbf{Z}}_{mix}^{dec}$$

- 30 en el que el decorrelacionador multicanal está configurado para seleccionar la matriz de premezclado $\mathbf{M}_{pre}$ en dependencia de las posiciones espaciales a las cuales están asociadas las señales de canales del primer conjunto $\hat{\mathbf{Z}}$ de señales de entrada del decorrelacionador N.
2. El decorrelacionador multicanal según la reivindicación 1, en el que $K=K'$.
3. El decorrelacionador multicanal según la reivindicación 1 o la reivindicación 2, en el que $N=N'$.
4. El decorrelacionador multicanal según una de las reivindicaciones 1 a 3, en el que $N \geq 3$ y $N' \geq 3$.
5. El decorrelacionador multicanal según una de las reivindicaciones 1 a 4, en el que el decorrelacionador multicanal está configurado para seleccionar la matriz de premezclado $\mathbf{M}_{pre}$ en dependencia de las características de correlación o las características de covarianza de las señales de canales del primer conjunto $\hat{\mathbf{Z}}$ de señales de entrada del decorrelacionador N.
6. El decorrelacionador multicanal según una de las reivindicaciones 1 a 5, en el que el decorrelacionador multicanal está configurado para determinar la matriz de premezclado de tal modo que un producto de la matriz:

$$(\mathbf{M}_{pre} \mathbf{M}_{pre}^H)$$

- está bien acondicionado con respecto a una operación de inversión.

$$\hat{\mathbf{Z}}_{mix} = \mathbf{M}_{pre} \hat{\mathbf{Z}}$$

7. El decorrelacionador multicanal según una de las reivindicaciones 1 a 6, en el que el

decorrelacionador multicanal está configurado para obtener la matriz de post-mezclado $\mathbf{M}_{\text{post}}$ según:

$$\mathbf{M}_{\text{post}} = \mathbf{M}_{\text{pre}}^H (\mathbf{M}_{\text{pre}} \mathbf{M}_{\text{pre}}^H)^{-1}.$$

- 5 8. El decorrelacionador multicanal según una de las reivindicaciones 1 a 7, en el que el decorrelacionador multicanal está configurado para recibir una información sobre una configuración de renderización asociada con las señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N, y en el que el decorrelacionador multicanal está configurado para seleccionar una matriz de premezclado ($\mathbf{M}_{\text{pre}}$) en dependencia de la información sobre la configuración de renderización.
- 10 9. El decorrelacionador multicanal según una de las reivindicaciones 1 a 8, en el que el decorrelacionador multicanal está configurado para combinar señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N que están asociadas con posiciones espacialmente adyacentes de una escena de audio cuando se realiza el premezclado.
- 15 10. El decorrelacionador multicanal según la reivindicación 9, en el que el decorrelacionador multicanal está configurado para combinar señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes de la escena de audio cuando se realiza el premezclado.
- 20 11. El decorrelacionador multicanal según una de las reivindicaciones 1 a 10, en el que el decorrelacionador multicanal está configurado para combinar señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N que están asociadas con un par horizontal de posiciones espaciales que comprenden una posición del lado izquierdo y una posición del lado derecho.
- 25 12. El decorrelacionador multicanal según una de las reivindicaciones 1 a 11, en el que el decorrelacionador multicanal está configurado para combinar al menos cuatro señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N, en el que al menos dos de dichas al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado izquierdo de una escena de audio, y en el que al menos dos de dichas al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado derecho de la escena de audio.
- 30 13. El decorrelacionador multicanal según la reivindicación 12, en el que al menos dos señales de canales del lado izquierdo que van a ser combinadas están asociadas con posiciones espaciales que son simétricas, con respecto a un plano central de la escena de audio, a las posiciones espaciales asociadas con al menos dos señales de canales del lado derecho que van a ser combinadas.
- 35 14. El decorrelacionador multicanal según una de las reivindicaciones 1 a 13, en el que el decorrelacionador multicanal está configurado para recibir una información de complejidad que describe un número de señales de entrada del decorrelacionador K del segundo conjunto de señales de entrada del decorrelacionador, y en el que el decorrelacionador multicanal está configurado para seleccionar una matriz de premezclado ($\mathbf{M}_{\text{pre}}$) en dependencia de la información de complejidad.
- 40 15. El decorrelacionador multicanal según la reivindicación 14, en el que el decorrelacionador multicanal está configurado para aumentar etapa a etapa un número de señales de entrada del decorrelacionador del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador que están combinadas para obtener las señales de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador con un valor decreciente de la información de complejidad.
- 45 16. El decorrelacionador multicanal según la reivindicación 14 o 15, en el que el decorrelacionador multicanal está configurado para combinar solo señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes de una escena de audio cuando se realiza el premezclado para un primer valor de la información de complejidad, y en el que el decorrelacionador multicanal está configurado para combinar al menos dos señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en un lado izquierdo de la escena de audio y al menos dos señales de canales del primer conjunto ($\tilde{\mathbf{z}}$) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en un lado derecho de la escena de audio para obtener una señal dada del segundo

conjunto de señales de entrada del decorrelacionador cuando se realiza el premezclado para un segundo valor de la información de complejidad.

17. El decorrelacionador multicanal según una de las reivindicaciones 14 a 16, en el que el decorrelacionador multicanal está configurado para combinar al menos cuatro señales de canales del primer conjunto (Z) de señales de entrada del decorrelacionador N, en el que al menos dos de dichas al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado izquierdo de una escena de audio, y en el que al menos dos de dichas al menos cuatro señales de canales están asociadas con posiciones espaciales en un lado derecho de una escena de audio, para obtener una señal dada del segundo conjunto de señales de entrada del decorrelacionador cuando se realiza el premezclado para un segundo valor de la información de complejidad.

18. El decorrelacionador multicanal según una de las reivindicaciones 14 a 17, en el que el decorrelacionador multicanal está configurado para combinar al menos dos señales de canales del primer conjunto (Z) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en un lado izquierdo de la escena de audio, para obtener una primera señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador, y para combinar al menos dos señales de canales del primer conjunto (Z) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en un lado derecho de la escena de audio, para obtener una segunda señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador para un primer valor de la información de complejidad, y en el que el decorrelacionador multicanal está configurado para combinar al menos dos señales de canales del primer conjunto (Z) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes del lado izquierdo de la escena de audio y al menos dos señales de canales del primer conjunto (Z) de señales de entrada del decorrelacionador N que están asociadas con posiciones vertical y espacialmente adyacentes en el lado derecho de la escena de audio, para obtener una señal de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador para un segundo valor de la información de complejidad, en el que un número de señales de entrada del decorrelacionador del segundo conjunto de señales de entrada del decorrelacionador es mayor para el primer valor de la información de complejidad que para el segundo valor de la información de complejidad.

19. Un decodificador de audio multicanal (100; 1550) para proporcionar al menos dos señales de audio de salida (112, 114; 1552a-1552n) en base a una representación codificada (110; 1516a, 1516b, 1518), en el que el decodificador de audio multicanal comprende un decorrelacionador multicanal (140; 600; 1590; 1700) según una de las reivindicaciones 1 a 18.

20. El decodificador de audio multicanal según la reivindicación 19, en el que el decodificador de audio multicanal está configurado para renderizar una pluralidad de señales de audio decodificadas (122; 1562a-1562n), que son obtenidas en base a la representación codificada, en dependencia de uno o más parámetros de renderización (132), para obtener una pluralidad de señales de audio renderizadas (134, 136; 1582a-1582n), y en el que el decodificador de audio multicanal está configurado para derivar una o más señales de audio decorrelacionadas (142, 144; 1592a-1592n) de las señales de audio renderizadas mediante el uso del decorrelacionador multicanal, en el que las señales de audio renderizadas constituyen el primer conjunto de señales de entrada del decorrelacionador, y en el que el segundo conjunto de señales de salida del decorrelacionador constituyen las señales de audio decorrelacionadas, y en el que el decodificador de audio multicanal está configurado para combinar (150; 1598) las señales de audio renderizadas, o una versión en escala de éstas con una o más señales de audio decorrelacionadas, para obtener las señales de audio de salida.

21. El decodificador de audio multicanal según la reivindicación 19 o la reivindicación 20, en el que el decodificador de audio multicanal está configurado para seleccionar la matriz de premezclado (M_{pre}) para su uso por el decorrelacionador multicanal en dependencia de una información de control incluida en la representación codificada.

22. El decodificador de audio multicanal según una de las reivindicaciones 19 a 21, en el que el decodificador de audio multicanal está configurado para seleccionar la matriz de premezclado (M_{pre}) para su uso por el decorrelacionador multicanal en dependencia de una configuración de salida que describe una asignación de las señales de audio de salida con posiciones espaciales de una escena de audio.

23. El decodificador de audio multicanal según una de las reivindicaciones 19 a 22, en el que el decodificador de audio multicanal está configurado para seleccionar entre tres o más matrices de premezclado

diferentes ($\mathbf{M}_{pre}$) para su uso por el decorrelacionador multicanal en dependencia de una información de control incluida en la representación codificada para una configuración de salida dada, en donde cada una de las tres o más matrices de premezclado diferentes está asociada con un número de señales diferente del segundo conjunto de señales de entrada del decorrelacionador K.

24. El decodificador de audio multicanal según una de las reivindicaciones 19 a 23, en el que el decodificador de audio multicanal está configurado para seleccionar la matriz de premezclado ($\mathbf{M}_{pre}$) para su uso por el decorrelacionador multicanal en dependencia de una matriz de mezclado (Dconv, Drender) que es usada por un convertidor de formato o renderizador que recibe al menos dos señales de audio de salida.

25. El decodificador de audio multicanal según la reivindicación 24, en el que el decodificador de audio multicanal está configurado para seleccionar la matriz de premezclado ($\mathbf{M}_{pre}$) para ser usada por el decorrelacionador multicanal para que sea igual a una matriz de mezclado (Dconv, Drender) que es usada por un convertidor de formato o renderizador que recibe al menos dos señales de audio de salida.

26. Un procedimiento (900) para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador que comprenden señales de canal de audio, comprendiendo el procedimiento:

premezclado (910) de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K, en el que $K < N$;
suministro (920) de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y
mezclado de manera ascendente (930) del primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto de señales de salida del decorrelacionador N', en el que $N' > K'$,

en el que el primer conjunto $\hat{\mathbf{Z}}$ de señales de entrada del decorrelacionador N en el segundo conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de entrada del decorrelacionador K mediante el uso de una matriz premezclada $\mathbf{M}_{pre}$ según,

$$\hat{\mathbf{Z}}_{mix} = \mathbf{M}_{pre} \hat{\mathbf{Z}}$$

en el que el primer conjunto $\hat{\mathbf{Z}}_{mix}^{dec}$ de señales de salida del decorrelacionador K' en base al segundo conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de entrada del decorrelacionador K, y

en el que el primer conjunto $\hat{\mathbf{Z}}_{mix}^{dec}$ de señales de salida del decorrelacionador K' está mezclado de manera ascendente en el segundo conjunto W de señales de salida del decorrelacionador N' mediante el uso de una matriz de post-mezclado $\mathbf{M}_{post}$ según,

$$\mathbf{W} = \mathbf{M}_{post} \hat{\mathbf{Z}}_{mix}^{dec}$$

en el que la matriz de premezclado $\mathbf{M}_{pre}$ se selecciona en dependencia de las posiciones espaciales a las cuales están asociadas las señales de canales del primer conjunto $\hat{\mathbf{Z}}$ de señales de entrada del decorrelacionador N.

27. Un procedimiento (1000) para proporcionar al menos dos señales de audio de salida en base a una representación codificada, en el que el procedimiento comprende el suministro (1020) de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador según la reivindicación 26.

28. Un programa informático para llevar a cabo el procedimiento según la reivindicación 26 o la reivindicación 27, cuando el programa informático se ejecuta en un ordenador.

29. Un decorrelacionador multicanal (140; 600; 1590; 1700) para proporcionar una pluralidad de señales decorrelacionadas (142, 144; 612a- 612n; 1592a-1592n; 1712a-1712n) en base a una pluralidad de señales de entrada del decorrelacionador (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n) que comprende señales de canal de audio, en el que el decorrelacionador multicanal es configurado para premezclar un primer conjunto $\hat{\mathbf{Z}}$ (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n; $\hat{\mathbf{Z}}$) de señales de entrada del decorrelacionador N en un segundo conjunto $\hat{\mathbf{Z}}_{mix}$ (622a-622k; 1722a-1722k; $\hat{\mathbf{Z}}_{mix}$) de señales de entrada del decorrelacionador K, en el que $K < N$;

en el que el decorrelacionador multicanal está configurado para proporcionar un primer conjunto (632a-632k; 1732a-1732k) de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y

en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) de señales de salida del decorrelacionador N', en el que N'>K';

en el que el decorrelacionador multicanal está configurado para premezclar el primer conjunto $\hat{\mathbf{Z}}$ de señales de entrada del decorrelacionador N en el segundo conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de entrada del decorrelacionador K mediante el uso de una matriz premezclada $\mathbf{M}_{pre}$ según

$$\hat{\mathbf{Z}}_{mix} = \mathbf{M}_{pre} \hat{\mathbf{Z}}$$

en el que el decorrelacionador multicanal está configurado para obtener el primer conjunto $\hat{\mathbf{Z}}_{mix}^{dec}$ de señales de salida del decorrelacionador K' en base al segundo conjunto $\hat{\mathbf{Z}}_{mix}$ de señales de entrada del decorrelacionador K, y en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto $\hat{\mathbf{Z}}_{mix}^{dec}$ de señales de salida del decorrelacionador K' en el segundo conjunto W de señales de salida del decorrelacionador N' mediante el uso de una matriz de post-mezclado $\mathbf{M}_{post}$ según,

$$\mathbf{W} = \mathbf{M}_{post} \hat{\mathbf{Z}}_{mix}^{dec}$$

en el que el decorrelacionador multicanal está configurado para seleccionar la matriz de premezclado $\mathbf{M}_{pre}$ en dependencia de las características de correlación o las características de covarianza de las señales de canal del primer conjunto $\hat{\mathbf{Z}}$ de las señales de entrada del decorrelacionador N.

30. Un decorrelacionador multicanal (140; 600; 1590; 1700) para proporcionar una pluralidad de señales decorrelacionadas (142, 144; 612a- 612n'; 1592a-1592n; 1712a-1712n) en base a una pluralidad de señales de entrada del decorrelacionador (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n),

25 en el que el decorrelacionador multicanal está configurado para premezclar un primer conjunto (134, 136; 610a-610n; 1582a-1582n; 1710a- 1710n; $\hat{\mathbf{Z}}$) de señales de entrada del decorrelacionador N en un segundo conjunto (622a-622k; 1722a-1722k; $\hat{\mathbf{Z}}_{mix}$) de señales de entrada del decorrelacionador K, en el que K<N;

en el que el decorrelacionador multicanal está configurado para proporcionar un primer conjunto (632a-632k'; 1732a-1732k) de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y

30 en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) de señales de salida del decorrelacionador N', en el que N'>K';

en el que el decorrelacionador multicanal está configurado para recibir una información sobre una configuración de renderización asociada con las señales de canal del primer conjunto ($\hat{\mathbf{Z}}$) de señales de entrada del decorrelacionador N, y en el que el decorrelacionador multicanal está configurado para seleccionar una matriz de premezclado ($\mathbf{M}_{pre}$) en dependencia de la información sobre la configuración de renderización.

31. Un decodificador de audio multicanal (100; 1550) para proporcionar al menos dos señales de audio de salida (112, 114; 1552a-1552n) en base a una representación codificada (110; 1516a, 1516b, 1518),

45 en el que el decodificador de audio multicanal comprende un decorrelacionador multicanal (140; 600; 1590; 1700) para proporcionar una pluralidad de señales decorrelacionadas (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) en base a una pluralidad de señales de entrada del decorrelacionador (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n),

en el que el decorrelacionador multicanal está configurado para premezclar un primer conjunto (134, 136; 610a-610n; 1582a-1582n; 1710a- 1710n; $\hat{\mathbf{Z}}$) de señales de entrada del decorrelacionador N en un segundo conjunto (622a-622k; 1722a-1722k; $\hat{\mathbf{Z}}_{mix}$) de señales de entrada del decorrelacionador K, en el que K<N;

50 en el que el decorrelacionador multicanal está configurado para proporcionar un primer conjunto (632a-632k'; 1732a-1732k) de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y

en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) de señales de salida del decorrelacionador N', en el que N'>K';

55 en el que el decodificador de audio multicanal está configurado para seleccionar una matriz de premezclado ($\mathbf{M}_{pre}$) para el uso por el decorrelacionador multicanal en dependencia de una configuración de salida que describe una

asignación de señales de audio de salida con posiciones espaciales de la escena de audio.

32. Un decodificador de audio multicanal (100; 1550) para proporcionar al menos dos señales de audio de salida (112, 114; 1552a-1552n) en base a una representación codificada (110; 1516a, 1516b, 1518),
- 5 en el que el decodificador de audio multicanal comprende un decorrelacionador multicanal (140; 600; 1590; 1700) para proporcionar una pluralidad de señales decorrelacionadas (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) en base a una pluralidad de señales de entrada del decorrelacionador (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n),
- 10 en el que el decorrelacionador multicanal está configurado para premezclar un primer conjunto (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n; $\hat{Z}$) de señales de entrada del decorrelacionador N en un segundo conjunto (622a-622k; 1722a-1722k; $\hat{Z}_{mix}$) de señales de entrada del decorrelacionador K, en el que $K < N$;
- en el que el decorrelacionador multicanal está configurado para proporcionar un primer conjunto (632a-632k'; 1732a-1732k) de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del
- 15 decorrelacionador K; y
- en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) de señales de salida del decorrelacionador N', en el que $N' > K'$;
- 20 en el que el decodificador de audio multicanal está configurado para seleccionar entre tres o más matrices de premezclado diferentes ($\mathbf{M}_{pre}$) para el uso por el decorrelacionador multicanal en dependencia de una información de control incluida en la representación codificada para una configuración de salida dada, en el que cada una de las tres o más matrices de premezclado diferentes están asociadas con un número de señales diferente del segundo conjunto de señales de entrada del decorrelacionador K.
- 25 33. Un decodificador de audio multicanal (100; 1550) para proporcionar al menos dos señales de audio de salida (112, 114; 1552a-1552n) en base a una representación codificada (110; 1516a, 1516b, 1518),
- en el que el decodificador de audio multicanal comprende un decorrelacionador multicanal (140; 600; 1590; 1700) para proporcionar una pluralidad de señales decorrelacionadas (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n)
- 30 en base a una pluralidad de señales de entrada del decorrelacionador (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n),
- en el que el decorrelacionador multicanal está configurado para premezclar un primer conjunto (134, 136; 610a-610n; 1582a-1582n; 1710a-1710n; $\hat{Z}$) de señales de entrada del decorrelacionador N en un segundo conjunto (622a-622k; 1722a-1722k; $\hat{Z}_{mix}$) de señales de entrada del decorrelacionador K, en el que $K < N$;
- 35 en el que el decorrelacionador multicanal está configurado para proporcionar un primer conjunto (632a-632k'; 1732a-1732k) de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y
- en el que el decorrelacionador multicanal está configurado para mezclar de manera ascendente el primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto (142, 144; 612a-612n'; 1592a-1592n; 1712a-1712n) de señales de salida del decorrelacionador N', en el que $N' > K'$;
- 40 en el que el decodificador de audio multicanal está configurado para seleccionar una matriz de premezclado ($\mathbf{M}_{pre}$) para el uso por el decorrelacionador multicanal en dependencia de una matriz de mezclado (Dconv, Dvisual) que es usada por un convertidor de formato o renderizador que recibe al menos dos señales de audio de salida.
- 45 34. Un procedimiento (900) para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador que comprenden señales de canal de audio, comprendiendo el procedimiento:
- premezclado (910) de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de
- 50 señales de entrada del decorrelacionador K, en el que $K < N$;
- suministro (920) de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y
- mezclado de manera ascendente (930) del primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto de señales de salida del decorrelacionador N', en el que $N' > K'$;
- 55 en el que el primer conjunto $\hat{Z}$ de señales de entrada del decorrelacionador N se premezcla en el segundo conjunto $\hat{Z}_{mix}$ de señales de entrada del decorrelacionador K mediante el uso de una matriz de premezclado $\mathbf{M}_{pre}$ según

$$\hat{Z}_{mix} = \mathbf{M}_{pre} \hat{Z}$$

en el que el primer conjunto $\hat{\mathbf{Z}}_{mix}^{dec}$ de las señales de salida del decorrelacionador K' se obtiene en base a un segundo conjunto $\hat{\mathbf{Z}}_{mix}$ de las señales de entrada del decorrelacionador K, y

en el que el primer conjunto $\hat{\mathbf{Z}}_{mix}^{dec}$ de las señales de salida del decorrelacionador K' se mezcla de manera ascendente en el segundo conjunto $\mathbf{W}$ de las señales de salida del decorrelacionador N' mediante el uso de una matriz de post-mezclado $\mathbf{M}_{post}$ según

$$\mathbf{W} = \mathbf{M}_{post} \hat{\mathbf{Z}}_{mix}^{dec},$$

en el que la matriz de premezclado $\mathbf{M}_{pre}$ se selecciona en dependencia de características de correlación o características de covarianza de las señales de canal del primer conjunto $\hat{\mathbf{Z}}$ de señales de entrada del decorrelacionador N.

35. Un procedimiento (900) para proporcionar una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador que comprenden señales de canal de audio, comprendiendo el procedimiento:

premezclado (910) de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K, en el que $K < N$;
suministro (920) de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y
mezclado (930) de manera ascendente del primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto de señales de salida del decorrelacionador N', en el que $N' > K'$;
en el que el procedimiento comprende la recepción de información acerca de una configuración de renderización asociada con las señales de canales del primer conjunto ($\hat{\mathbf{Z}}$) de señales de entrada del decorrelacionador N, y en el que una matriz de premezclado ($\mathbf{M}_{pre}$) se selecciona en dependencia de la información acerca de la configuración de renderización.

36. Un procedimiento (1000) para proporcionar al menos dos señales de audio de salida en base a una representación codificada, en el que el procedimiento comprende el suministro (1020) de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador, en el que el suministro de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador comprende:

premezclado (910) de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K, en el que $K < N$;
suministro (920) de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y
mezclado de manera ascendente (930) del primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto de señales de salida del decorrelacionador N', en el que $N' > K'$;
en el que se selecciona una matriz de premezclado ($\mathbf{M}_{pre}$) para el uso por el decorrelacionador multicanal en dependencia de una configuración de salida que describe una asignación de señales de audio de salida con posiciones espaciales de la escena de audio.

37. Un procedimiento (1000) para proporcionar al menos dos señales de audio de salida en base a una representación codificada, en el que el procedimiento comprende el suministro (1020) de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador,

en el que el suministro de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador comprende:

premezclado (910) de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de señales de entrada del decorrelacionador K, en el que $K < N$;
suministro (920) de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto de señales de entrada del decorrelacionador K; y
mezclado de manera ascendente (930) del primer conjunto de señales de salida del decorrelacionador K' en un segundo conjunto de señales de salida del decorrelacionador N', en el que $N' > K'$;
en el que el procedimiento comprende seleccionar entre tres o más matrices de premezclado ($\mathbf{M}_{pre}$) diferentes para el uso por el decorrelacionador multicanal en dependencia de una información de control incluida en la representación codificada para una configuración de salida dada, en donde cada una de las tres o más matrices de premezclado diferentes están asociadas con un número de señales diferente del segundo conjunto de señales de entrada del decorrelacionador K.

38. Un procedimiento (1000) para proporcionar al menos dos señales de audio de salida en base a una representación codificada, en el que el procedimiento comprende el suministro (1020) de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador,

- 5 en el que el suministro de una pluralidad de señales decorrelacionadas en base a una pluralidad de señales de entrada del decorrelacionador comprende:
premezclado (910) de un primer conjunto de señales de entrada del decorrelacionador N en un segundo conjunto de
señales de entrada del decorrelacionador K, en el que $K < N$;
suministro (920) de un primer conjunto de señales de salida del decorrelacionador K' en base al segundo conjunto
10 de señales de entrada del decorrelacionador K; y
mezclado de manera ascendente (930) del primer conjunto de señales de salida del decorrelacionador K' en un
segundo conjunto de señales de salida del decorrelacionador N', en el que $N' > K'$;
en el que la matriz de premezclado (M_{pre}) para el uso por el decorrelacionador multicanal se selecciona en
dependencia de una matriz de mezclado (Dconv, Dvisual) que es usada por un convertidor de formato o
15 renderizador que recibe al menos dos señales de audio de salida.

39. Un programa informático para llevar a cabo el procedimiento de cualquiera de las reivindicaciones 34 a 38 cuando el programa informático se ejecuta en un ordenador.

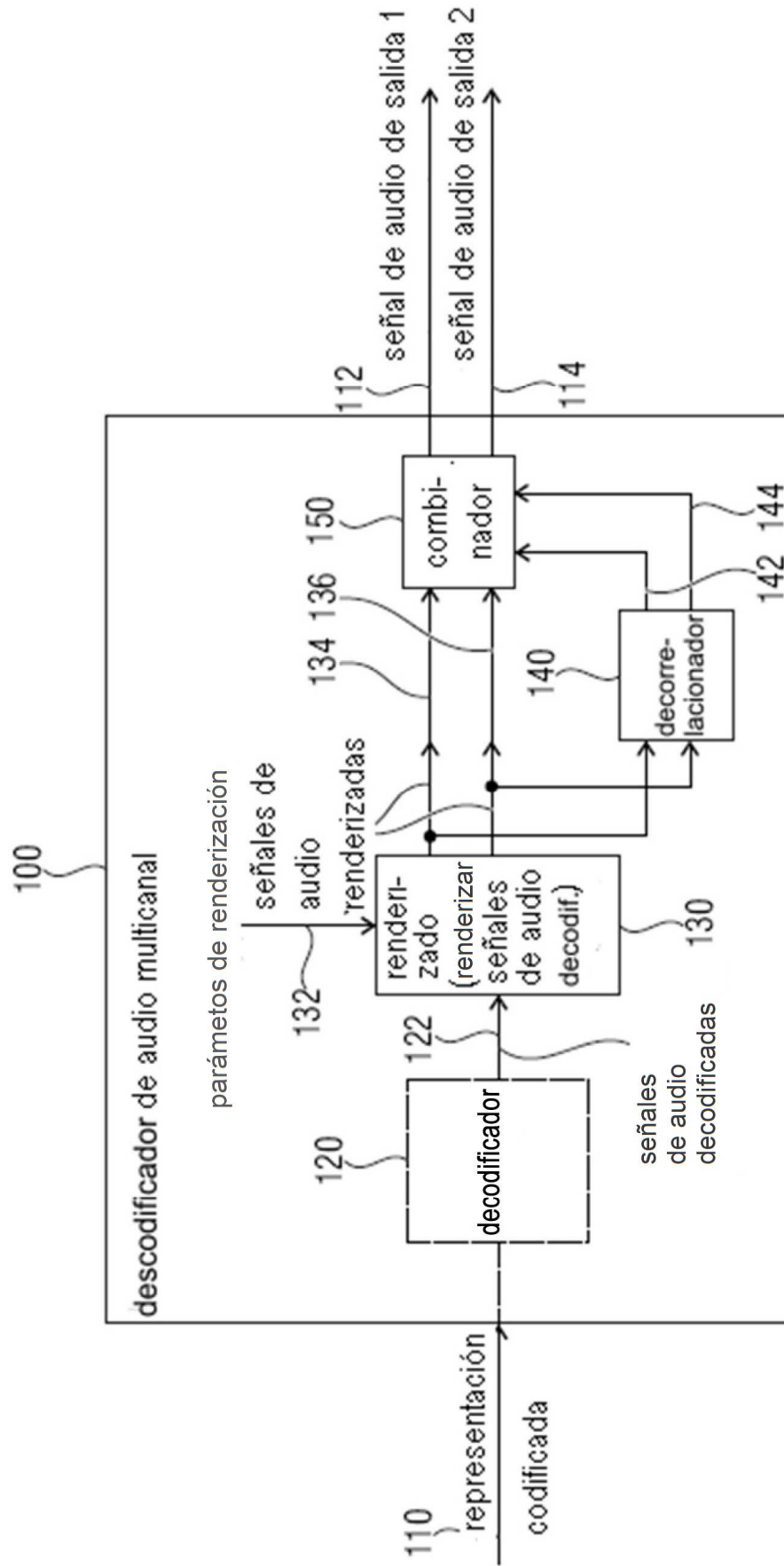


FIG 1

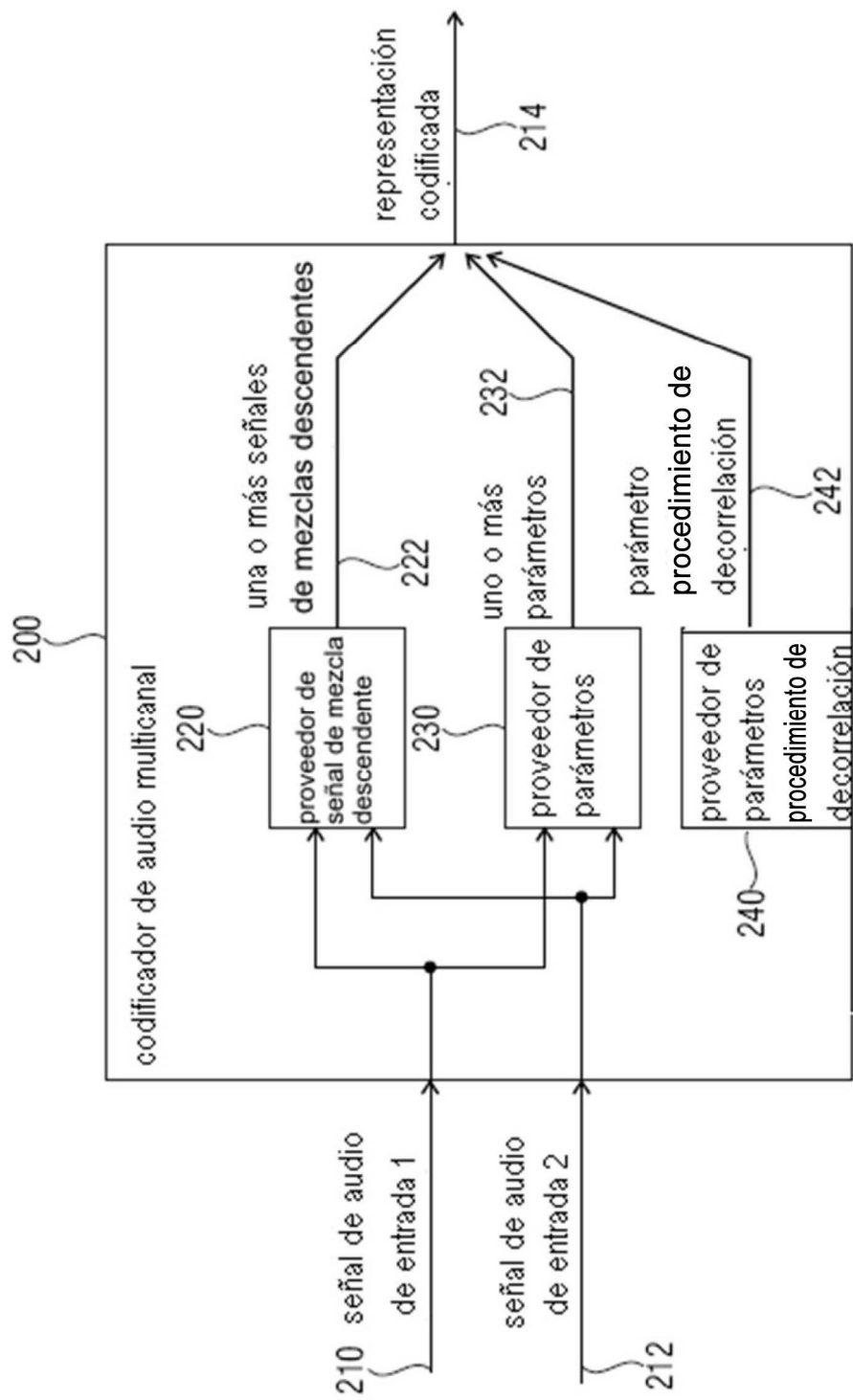


FIG 2

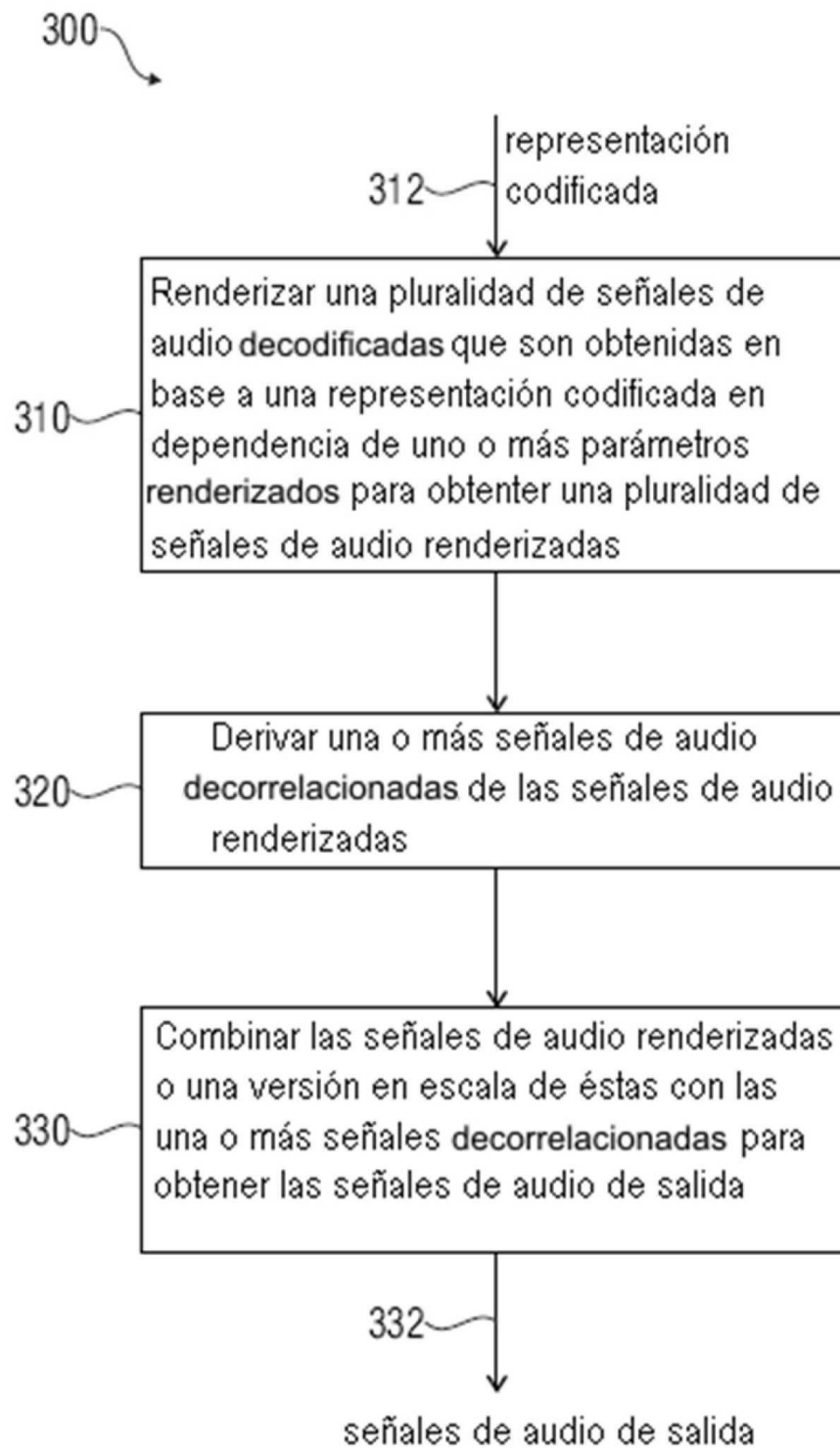


FIG 3

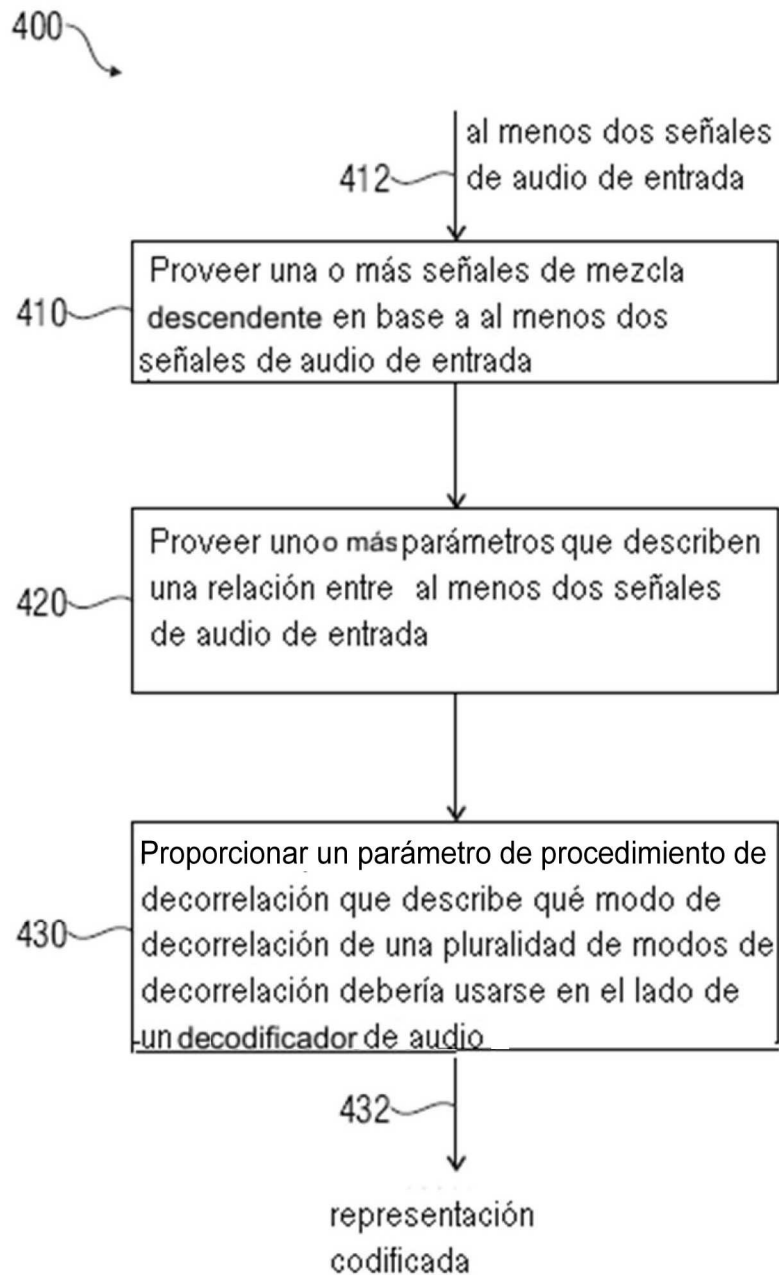


FIG 4

500

representación de audio codificada

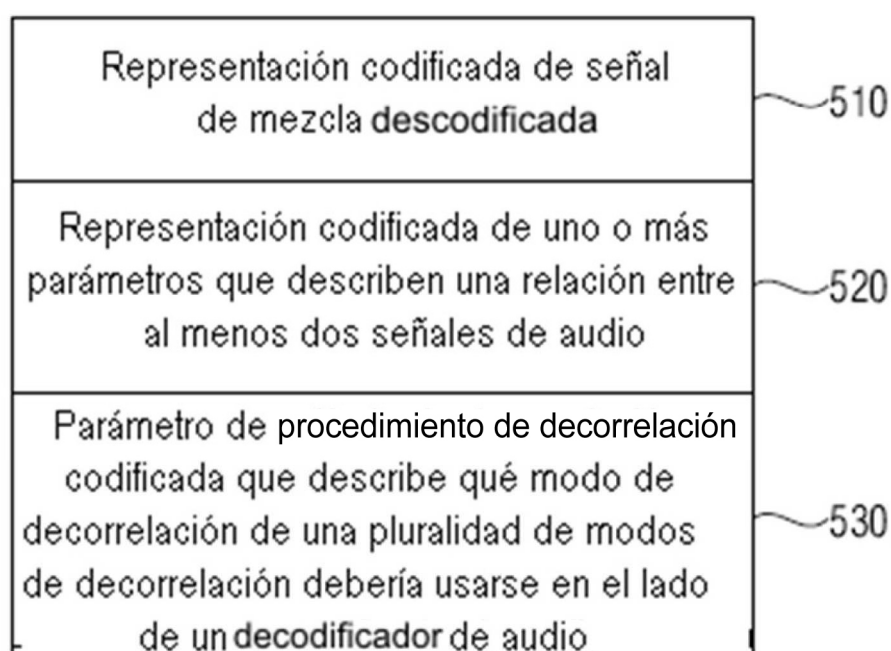


FIG 5

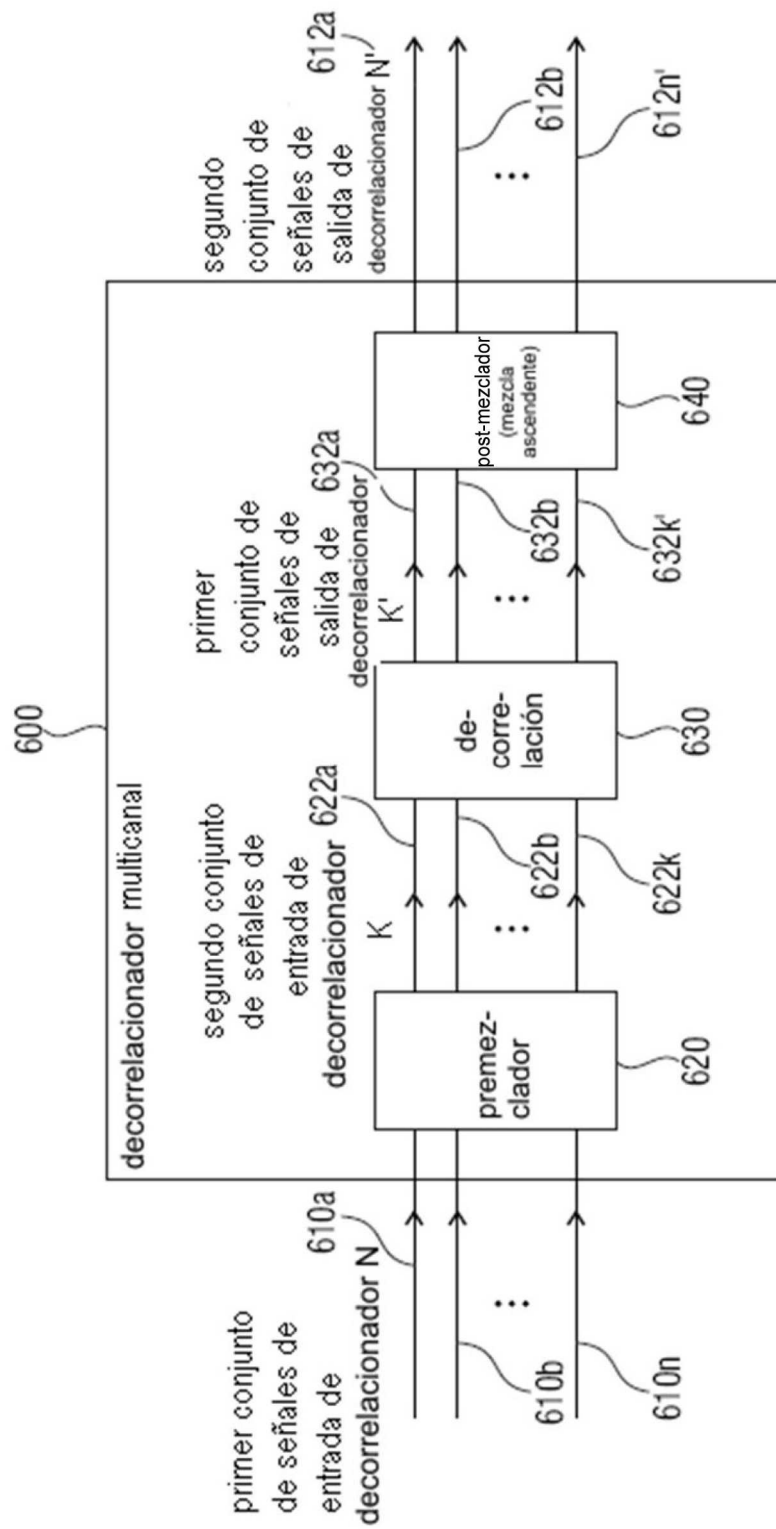


FIG 6

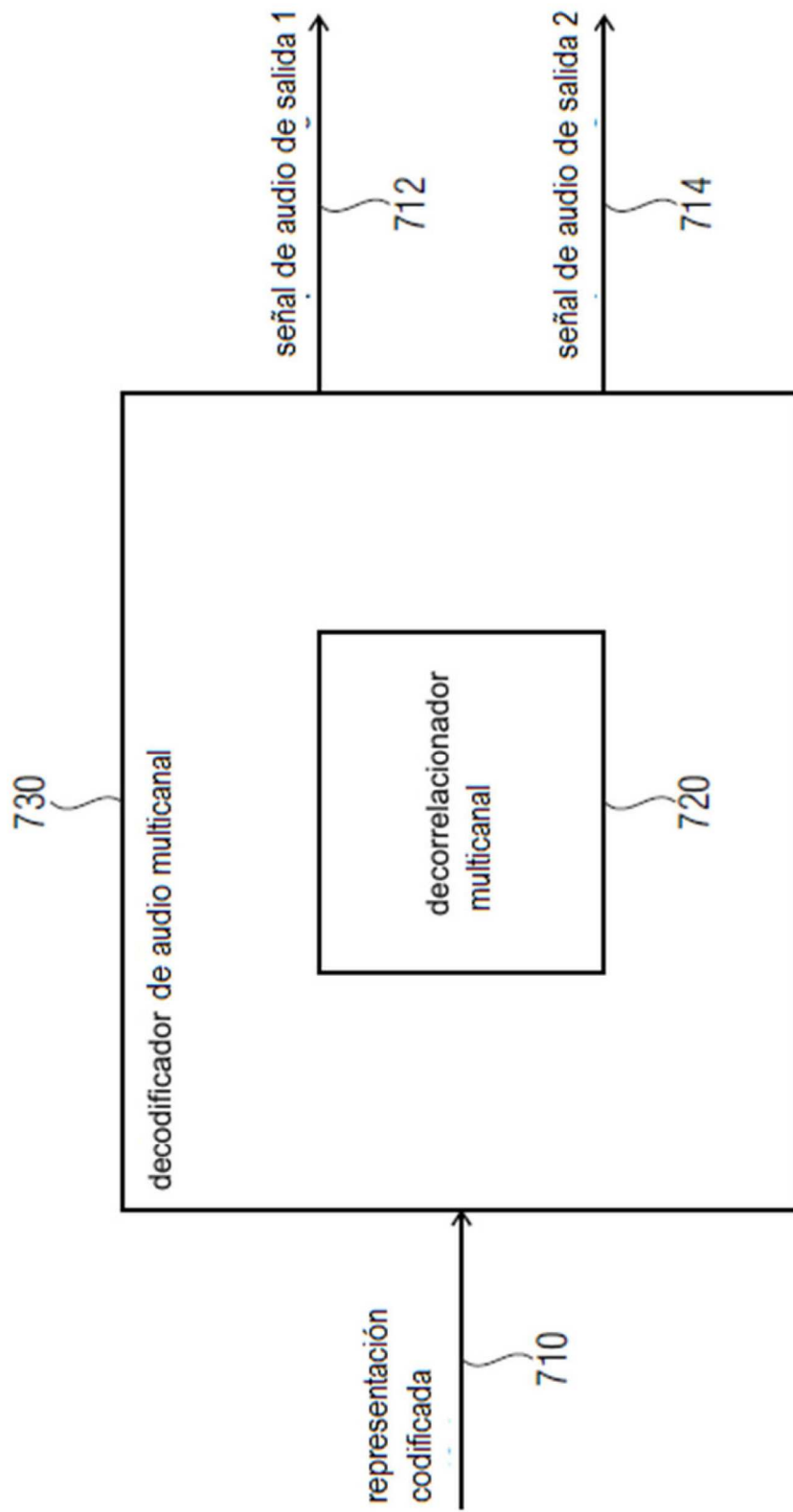


FIG 7

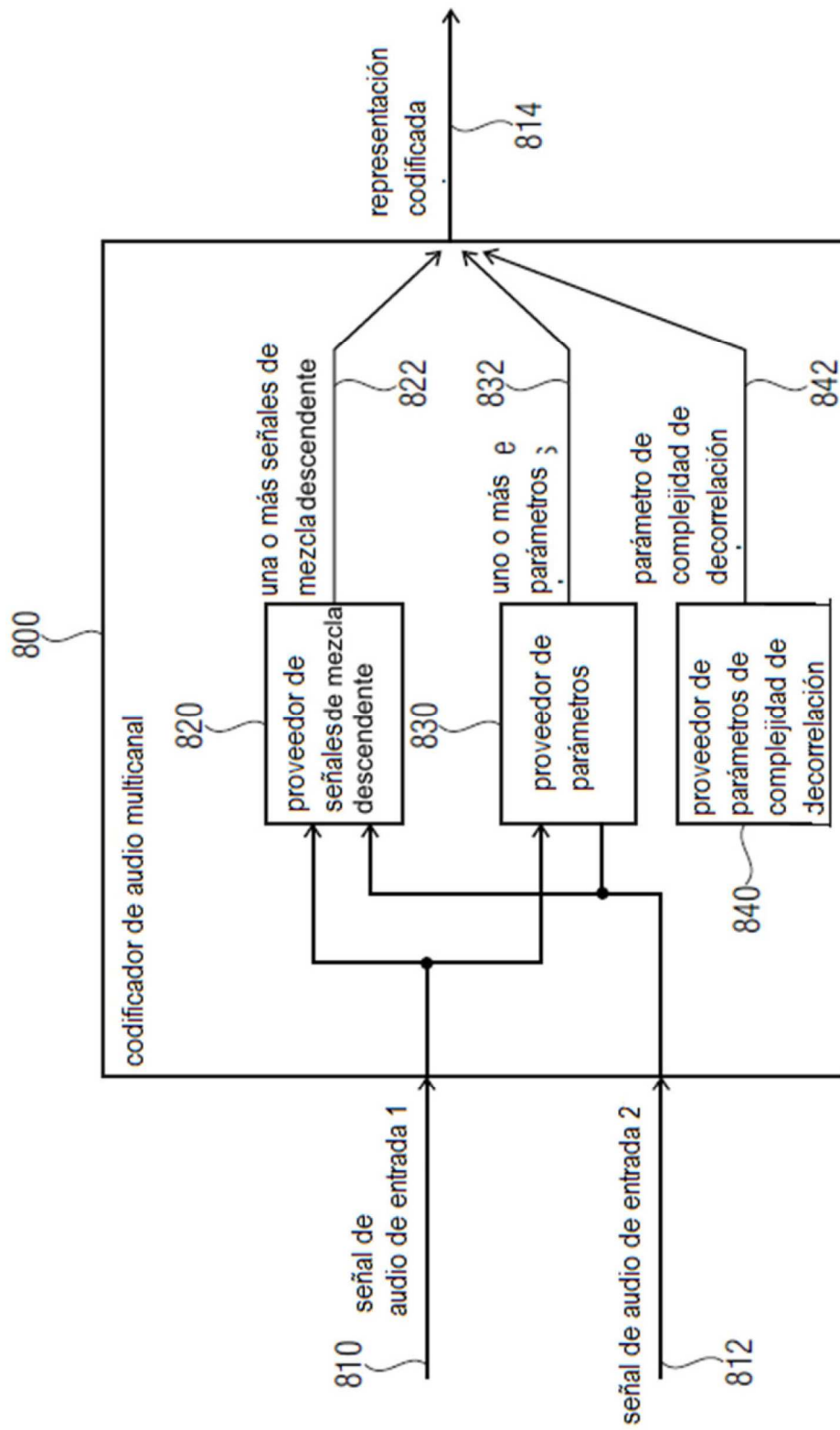


FIG 8

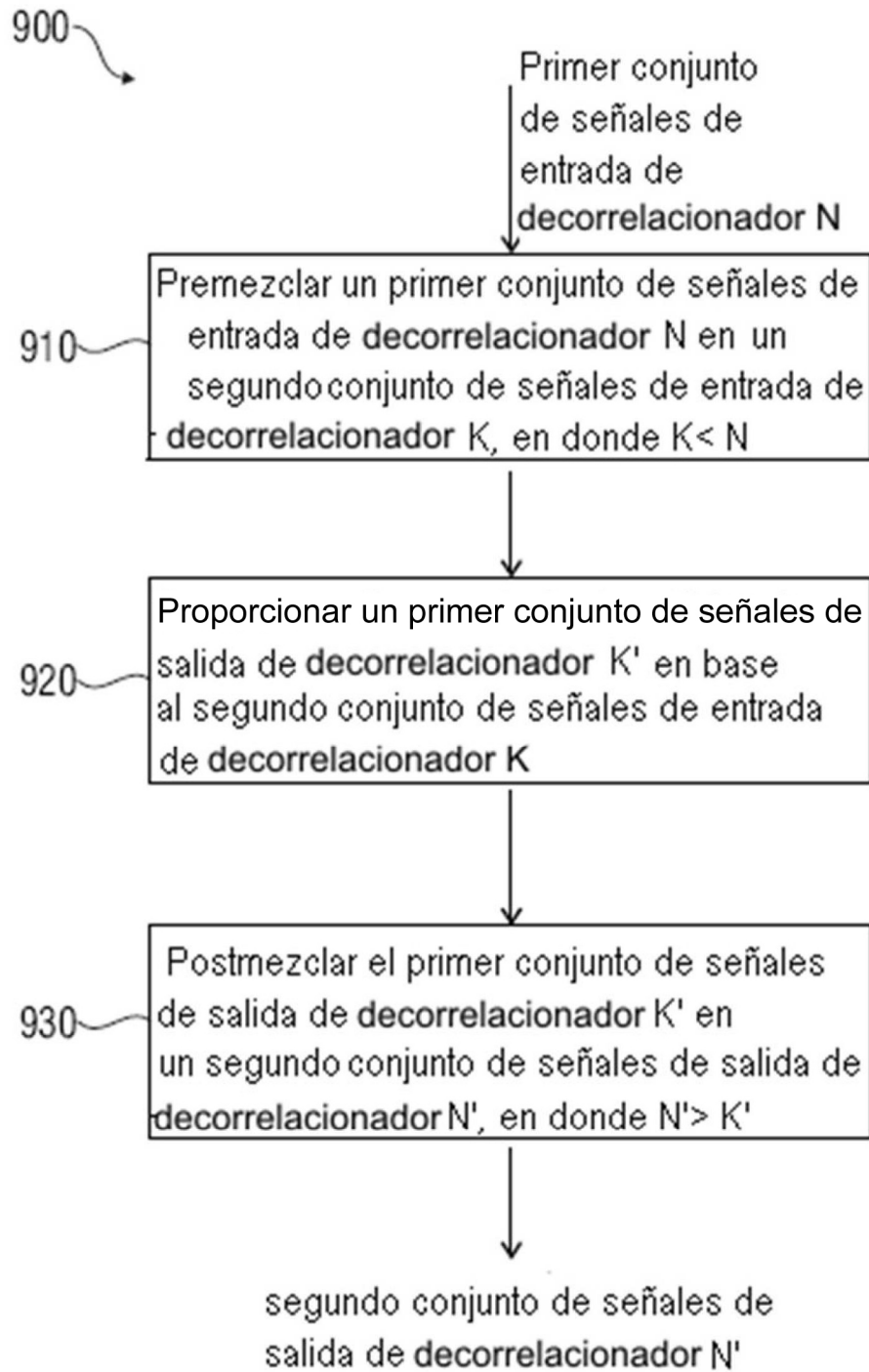


FIG 9

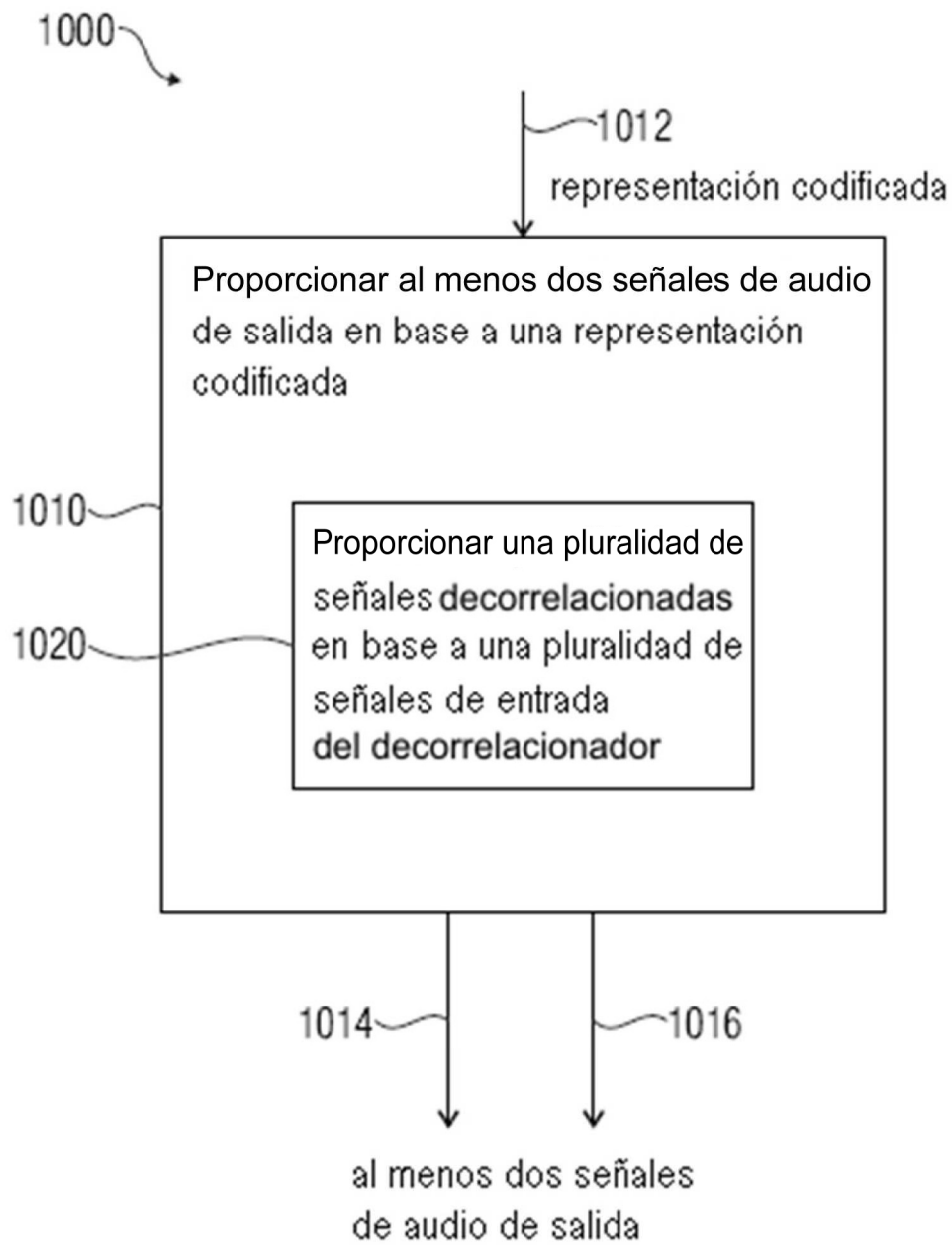


FIG 10

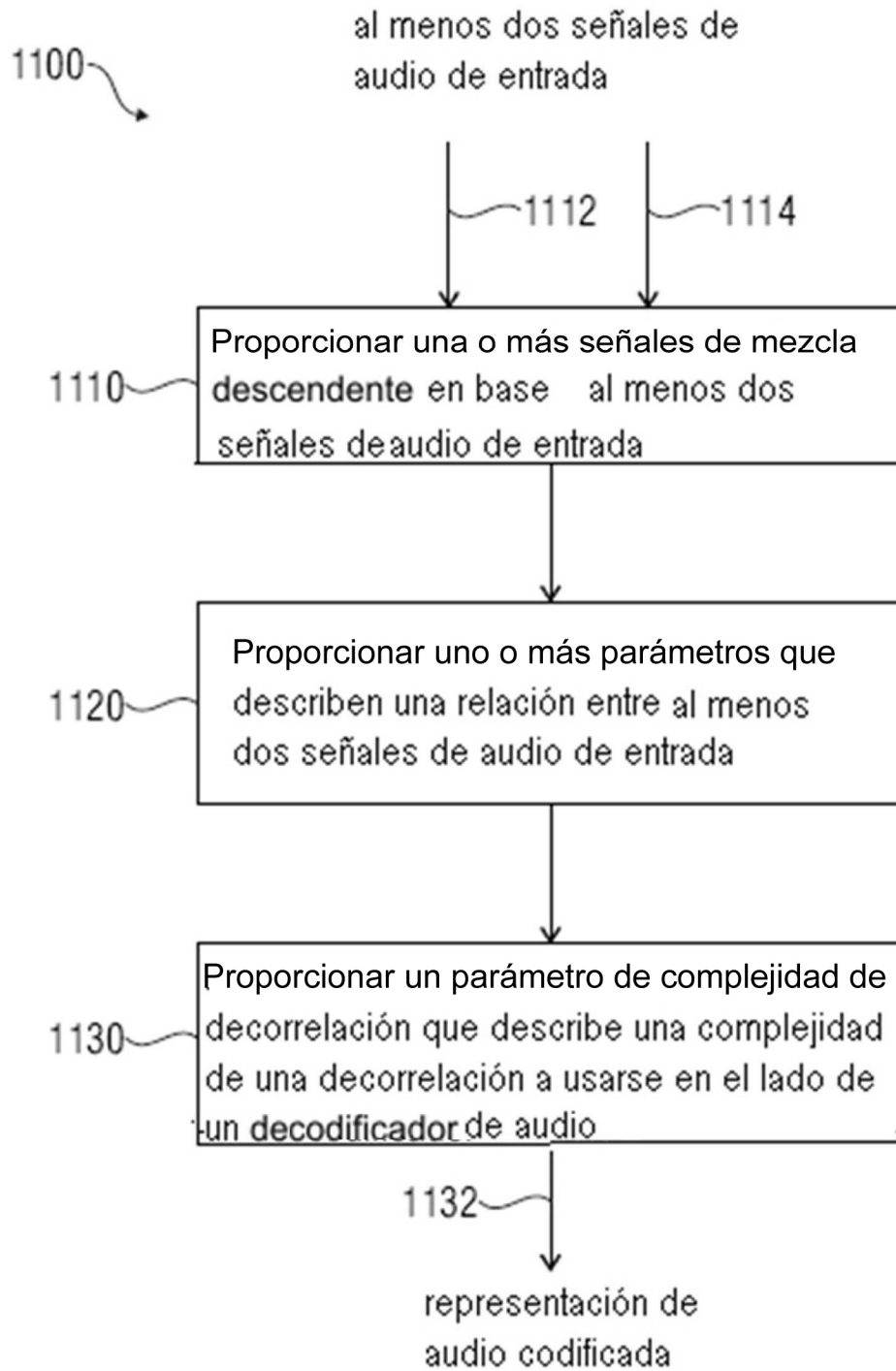


FIG 11

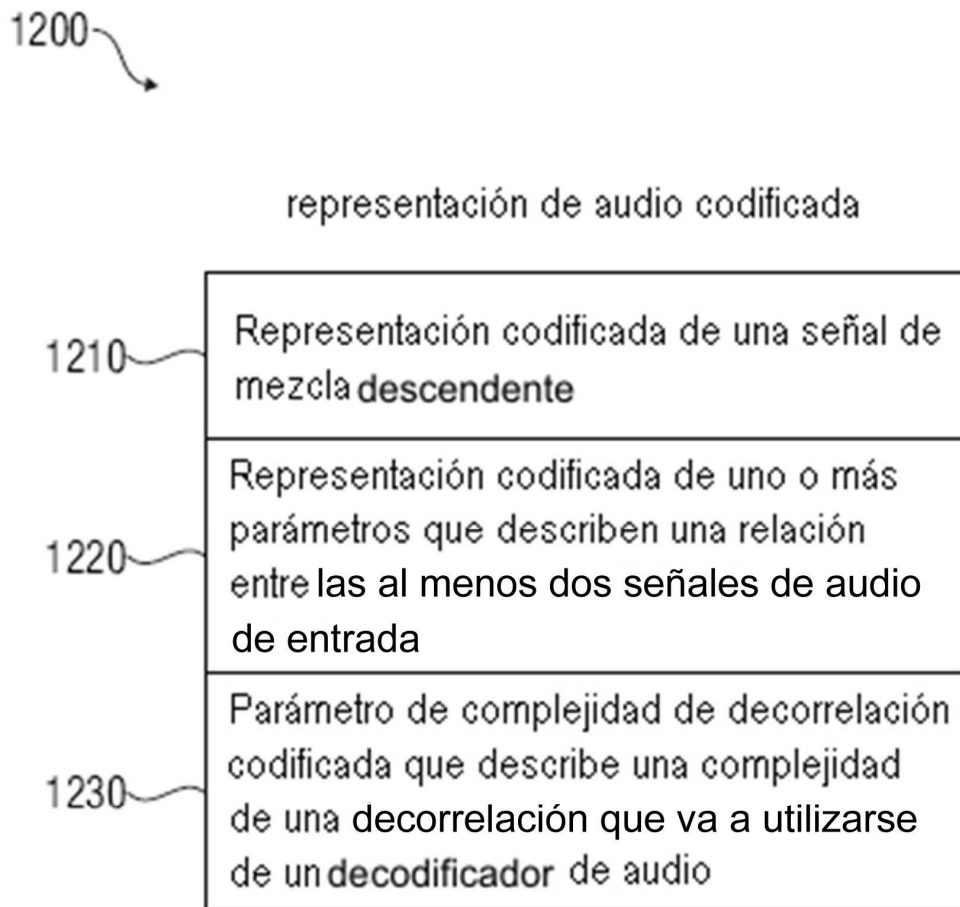
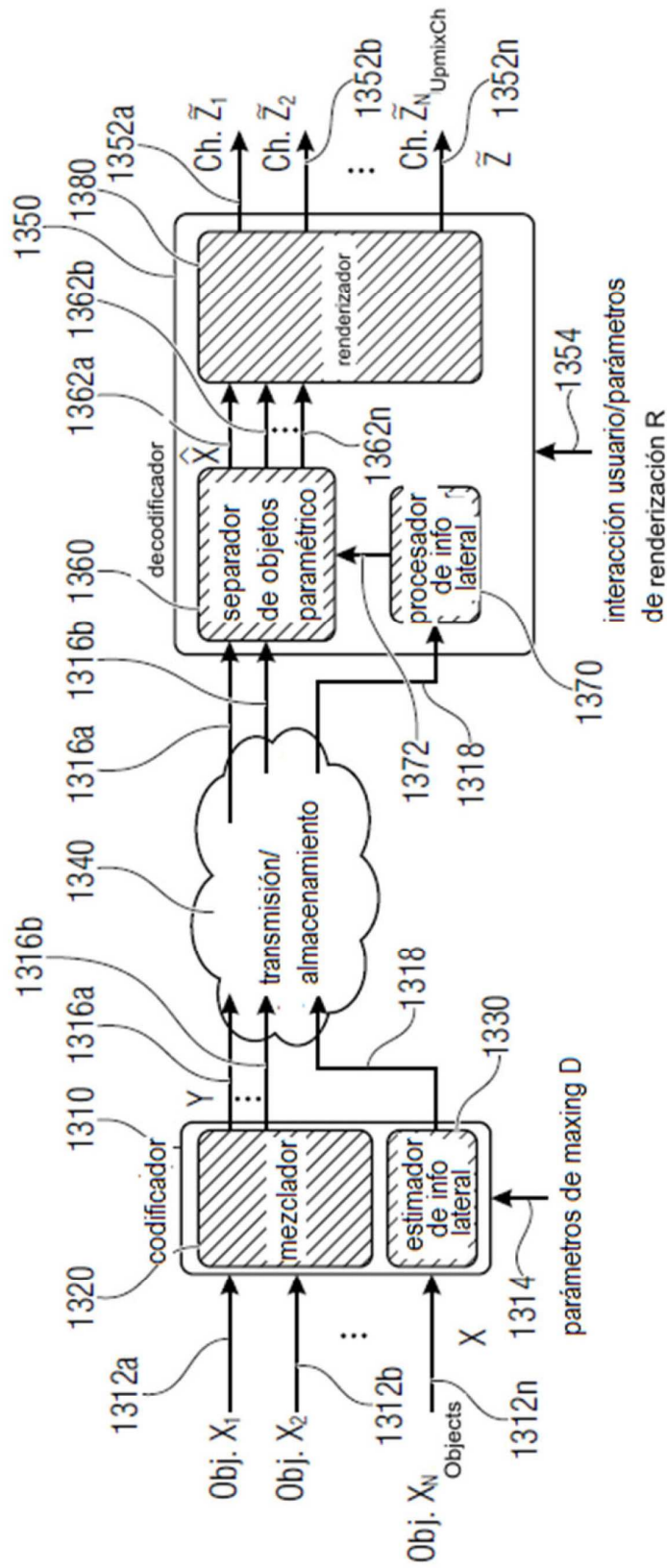
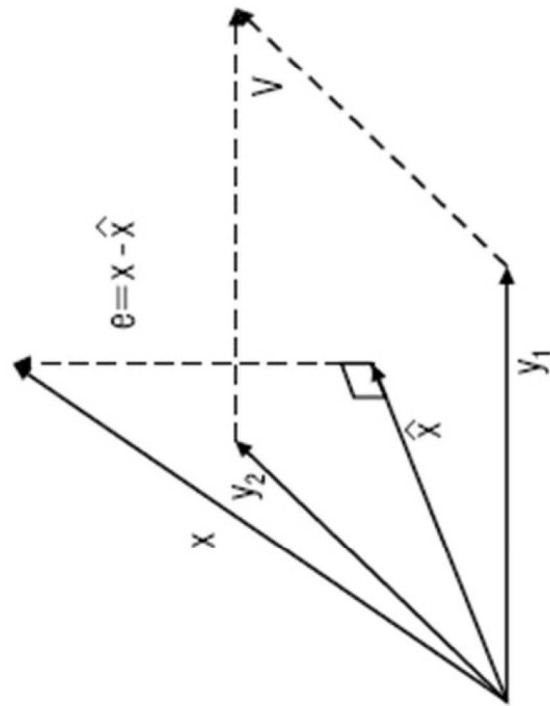


FIG 12



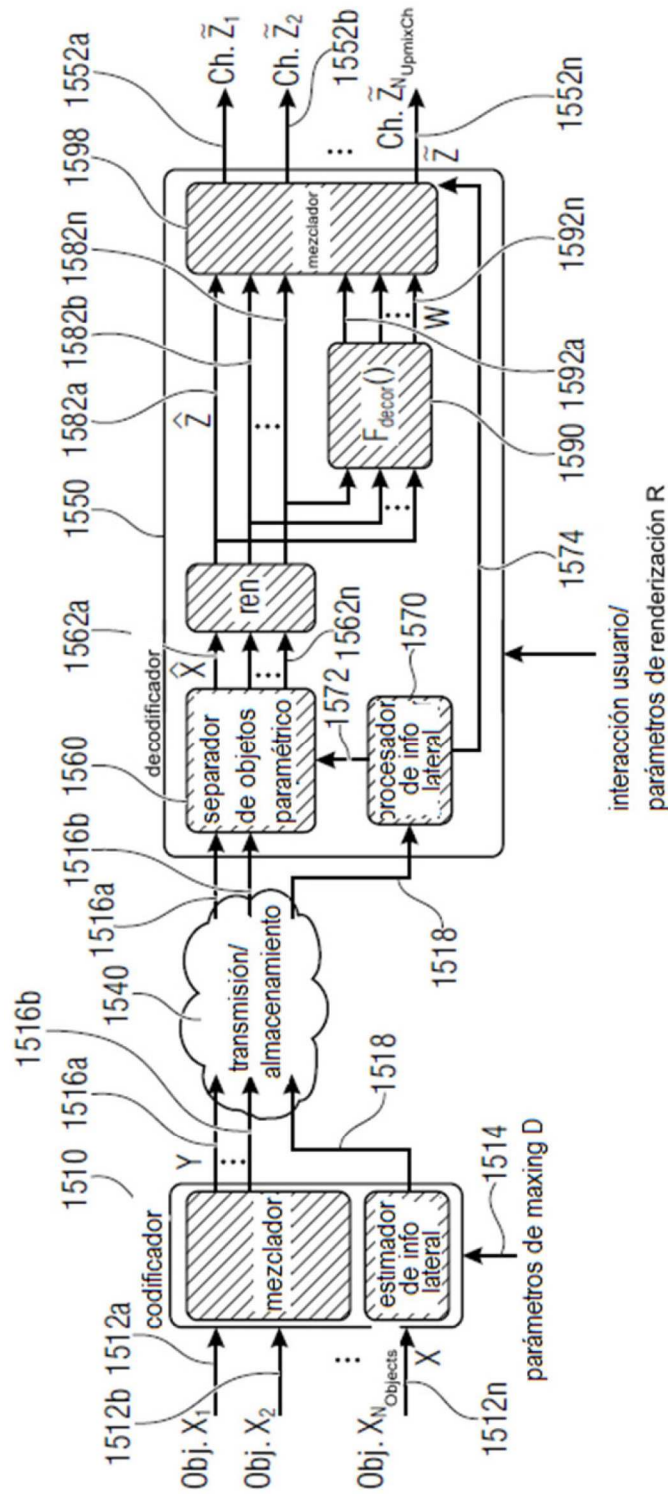
VISTA GENERAL DE UN CONCEPTO DE MEZCLA DESCENDENTE/MEZCLA ASCENDENTE PARAMÉTRICA BASADO EN MMSE

FIG 13



Representación geométrica del principio de ortogonalidad en un espacio tridimensional

FIG 14



SISTEMA DE RECONSTRUCCION PARAMÉTRICA CON DECORRELACIÓN APLICADA A LA SALIDA RENDERIZADA

FIG 15

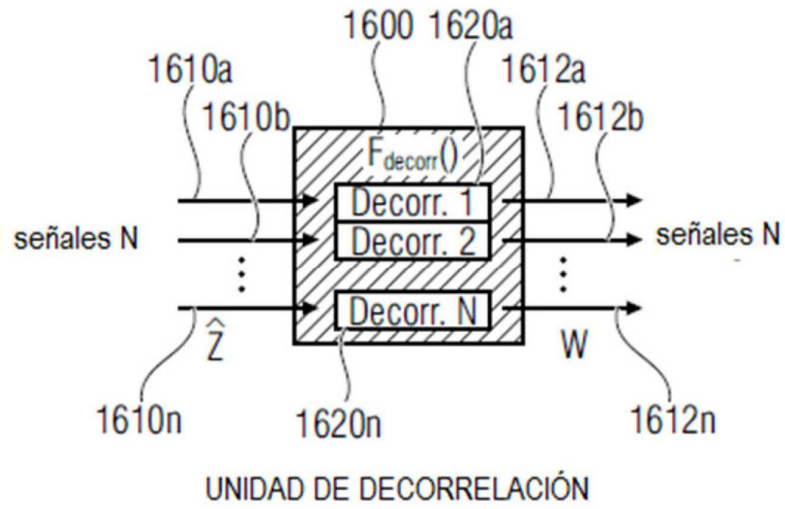


FIG 16

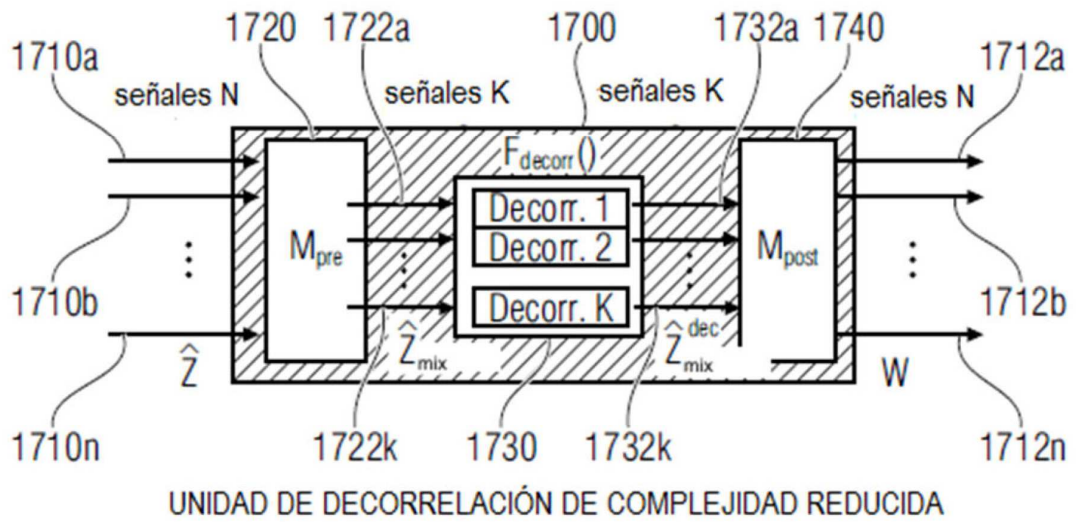


FIG 17

Posiciones de altavoz y formatos de salida

No.	Marca LS	Az. °	Az. Tol. °	El. °	El. Tol. °	formatos de salida					
						0-2.0	0-5.1	0-7.1	0-8.1	0-10.1	0-22.2
1	CH_M_000	0	±2	0	±2		3	3		3	3
2	CH_M_L030	30	±2	0	±2	1	1	1	1	1	7
3	CH_M_R030	-30	±2	0	±2	2	2	2	2	2	8
4	CH_M_L060	60	±2	0	±2						1
5	CH_M_R060	-60	±2	0	±2						2
6	CH_M_L090	90	±5	0	±2						11
7	CH_M_R090	-90	±5	0	±2						12
8	CH_M_L110	110	±5	0	±2		5	5	5	5	
9	CH_M_R110	-110	±5	0	±2		6	6	6	6	
10	CH_M_L135	135	±5	0	±2						5
11	CH_M_R135	-135	±5	0	±2						6
12	CH_M_180	180	±5	0	±2						9
13	CH_U_000	0	±2	35	±10				3		15
14	CH_U_L045	45	±5	35	±10						13
15	CH_U_R045	-45	±5	35	±10						14
16	CH_U_L030	30	±5	35	±10			7	7	7	
17	CH_U_R030	-30	±5	35	±10			8	8	8	
18	CH_U_L090	90	±5	35	±10						19
19	CH_U_R090	-90	±5	35	±10						20
20	CH_U_L110	110	±5	35	±10					9	
21	CH_U_R110	-110	±5	35	±10					10	
22	CH_U_L135	135	±5	35	±10						17
23	CH_U_R135	-135	±5	35	±10						18
24	CH_U_180	180	±5	35	±10						21
25	CH_T_000	0	±2	90	±10					11	16
26	CH_L_000	0	±2	-15	+5-25				9		22
27	CH_L_L045	45	±5	-15	+5-25						23
28	CH_L_R045	-45	±5	-15	+5-25						24
29	CH_LFE1	45	±15	-15	±15		4	4	4	4	4
30	CH_LFE2	-45	±15	-15	±15						10

1810 1820 1830 1832 1840 1842 1850 1860 1864 1870 1880 1890

FIG 18

Reducción de complejidad para formato de salida 22,2

- Coeficientes de premezclado para N=22 y K=11, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1$.

Ch. ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
$M_{\text{pre}}^{i,j}$	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

FIG 19A

- Coeficientes de premezclado para N=22 y K=10, $\text{cond}(M_{\text{pre}}^H) = 2$.

Ch. ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
CH_M_000	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_L_000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_T_000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_L135	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_L135	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_R135	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_R135	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_180	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_180	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_L030	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CH_L_L045	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
CH_M_R030	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
CH_L_R045	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CH_M_L090	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
CH_U_L090	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
CH_M_R090	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
CH_U_R090	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
CH_M_L060	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
CH_U_L045	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
CH_M_R060	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
CH_U_R045	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

FIG 19B

- Coeficientes de premezclado para N=22 y K=9, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

FIG 19C

- Coeficientes de premezclado para N=22 y K=8, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	$M_{\text{pre}}^{i,j}$	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

FIG 19D

- Coeficientes de premezclado para $N=22$ y $K=7$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
CH_M_000	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_L_000	2	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_000	3	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_T_000	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_L135	5	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
CH_U_L135	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CH_M_R135	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_R135	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_180	9	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_180	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_L030	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_L_L045	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_R030	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_L_R045	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_L090	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_L090	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_R090	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_R090	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_L060	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_L045	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_M_R060	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH_U_R045	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIG 19E

. Coeficientes de premezclado para N=22 y K=6 , $\text{cond}(M_{\text{pre}}^H) = 2$.

Ch. ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
$M_{\text{pre}}^{i,j}$	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

FIG 19F

- Coeficientes de premezclado para N=22 y K=5, $\text{cond}(M_{\text{pre}}^H M_{\text{pre}}^H) = 1,5$

Ch. ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
M_{pre}^{ij}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045

FIG 19G

Reducción de complejidad para formato de salida 10.1

- Coeficientes de premezclado para N=10 y K=5, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_U_L110	CH_M_R110	CH_U_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	0	0	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	0	0
4	0	0	0	0	0	0	1	1	0	0
5	0	0	0	0	0	0	0	0	1	1

FIG 20A

- Coeficientes de premezclado para N=10 y K=4, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_U_L110	CH_M_R110	CH_U_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	0	0	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	0	0
4	0	0	0	0	0	0	1	1	1	1

FIG 20B

- Coeficientes de premezclado para $N=10$ y $K=3$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_U_L110	CH_M_R110	CH_U_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	0	0	0	0	0	0
2	0	0	0	0	1	1	0	0	0	0
3	0	0	0	0	0	0	1	1	1	1

FIG 20C

- Coeficientes de premezclado para $N=10$ y $K=2$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1,5$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_U_L110	CH_M_R110	CH_U_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	0	0	0	0
2	0	0	0	0	0	0	1	1	1	1

FIG 20D

Reducción de complejidad para formato de salida 8.1

- Coeficientes de premezclado para $N=8$ y $K=4$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_M_R110
M_{pre}^{ij}	1	2	3	4	5	6	7	8
1	1	1	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	1

FIG 21A

- Coeficientes de premezclado para $N=8$ y $K=3$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_M_R110
M_{pre}^{ij}	1	2	3	4	5	6	7	8
1	1	1	1	1	0	0	0	0
2	0	0	0	0	1	1	0	0
3	0	0	0	0	0	0	1	1

FIG 21B

- Coeficientes de premezclado para $N=8$ y $K=2$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 3$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	0	0
3	0	0	0	0	0	0	1	1

FIG 21C

Reducción de complejidad para formato de salida 7.1

- Coeficientes de premezclado para $N=7$ y $K=4$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7
1	1	1	0	0	0	0	0
2	0	0	1	1	0	0	0
3	0	0	0	0	1	0	0
4	0	0	0	0	0	1	1

FIG 21D

- Coeficientes de premezclado para $N=7$ y $K=3$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7
1	1	1	1	1	0	0	0
2	0	0	0	0	1	0	0
3	0	0	0	0	0	1	1

FIG 21E

- Coeficientes de premezclado para $N=7$ y $K=2$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2,5$

Ch. ID	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7
1	1	1	1	1	1	0	0
2	0	0	0	0	0	1	1

FIG 21F

Reducción de complejidad para formato de salida 5.1

- Coeficientes de premezclado para $N=5$ y $K=3$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	CH M L030	CH M R030	CH M 000	CH M L110	CH M R110
M_{pre}^{ij}	1	2	3	4	5
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	1

FIG 22A

- Coeficientes de premezclado para $N=5$ y $K=2$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^{H^*}) = 1,5$

Ch. ID	CH M L030	CH M R030	CH M 000	CH M L110	CH M R110
M_{pre}^{ij}	1	2	3	4	5
1	1	1	1	0	0
2	0	0	0	1	1

FIG 22B

Reducción de complejidad para formato de salida 2.0

- Coeficientes de premezclado para $N=2$ y $K=1$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

Ch. ID	CH_M_L030	CH_M_R030
$M_{\text{pre}}^{i,j}$	1	2
1	1	1

FIG 23

2410	Grupo 1	CH_M_000	CH_L_000	CH_U_000	CH_T_000		
		(Ch. ID. 3)	(Ch. ID. 22)	(Ch. ID. 15)	(Ch. ID. 16)		
2412	Grupo 2	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180
		(Ch. ID. 5)	(Ch. ID. 17)	(Ch. ID. 6)	(Ch. ID. 18)	(Ch. ID. 9)	(Ch. ID. 21)
2414	Grupo 3	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045		
		(Ch. ID. 7)	(Ch. ID. 23)	(Ch. ID. 8)	(Ch. ID. 24)		
2416	Grupo 4	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090		
		(Ch. ID. 11)	(Ch. ID. 19)	(Ch. ID. 12)	(Ch. ID. 20)		
2418	Grupo 5	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045		
		(Ch. ID. 1)	(Ch. ID. 13)	(Ch. ID. 2)	(Ch. ID. 14)		

FIG 24

bsDecorrelationMethod;	2	uimsbf
bsDecorrelationLevel;	2	uimsbf

FIG 25

bsDecorrelationMethod

Indica el modo operativo de decorrelación del decodificador de acuerdo con:

bsDecorrelationMethod

bsDecorrelationMethod	Significado
0	Modo de compensación de energía
1	Modo de ajuste de covarianza limitada
2	Modo de ajuste de covarianza general
3	N/A

FIG 26

bsDecorrelationLevel

Define el nivel de decorrelación de acuerdo con:

bsDecorrelationLevel

bsDecorrelationLevel	22.2	10.1	8.1	7.1	5.1	2.1
0	11	10	8	7	5	2
1	9	5	4	4	3	1
2	7	3	3	3	2	0
3	5	2	2	2	0	0

FIG 27

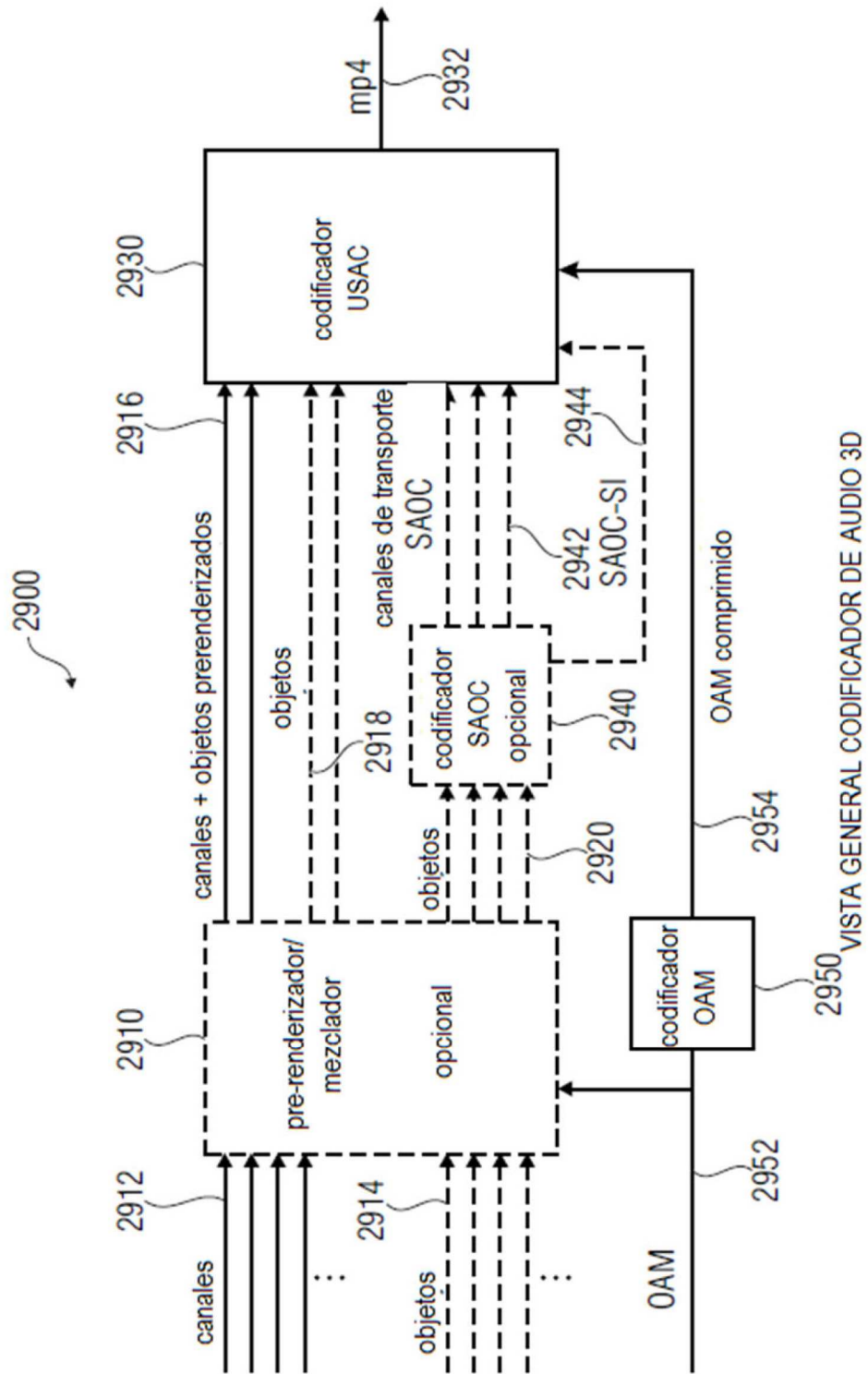
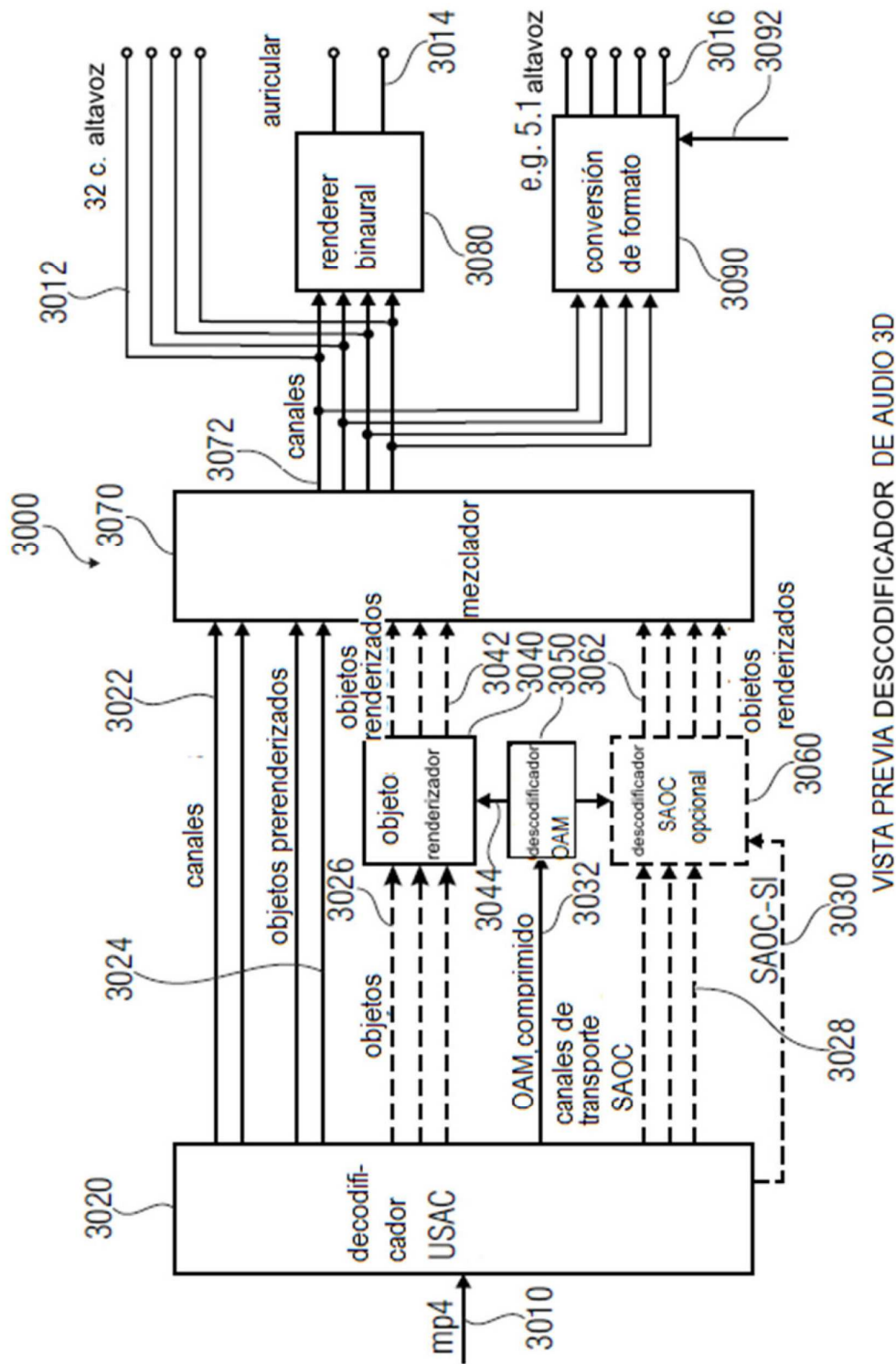


FIG 28



VISTA PREVIA DESCODIFICADOR DE AUDIO 3D

FIG 29

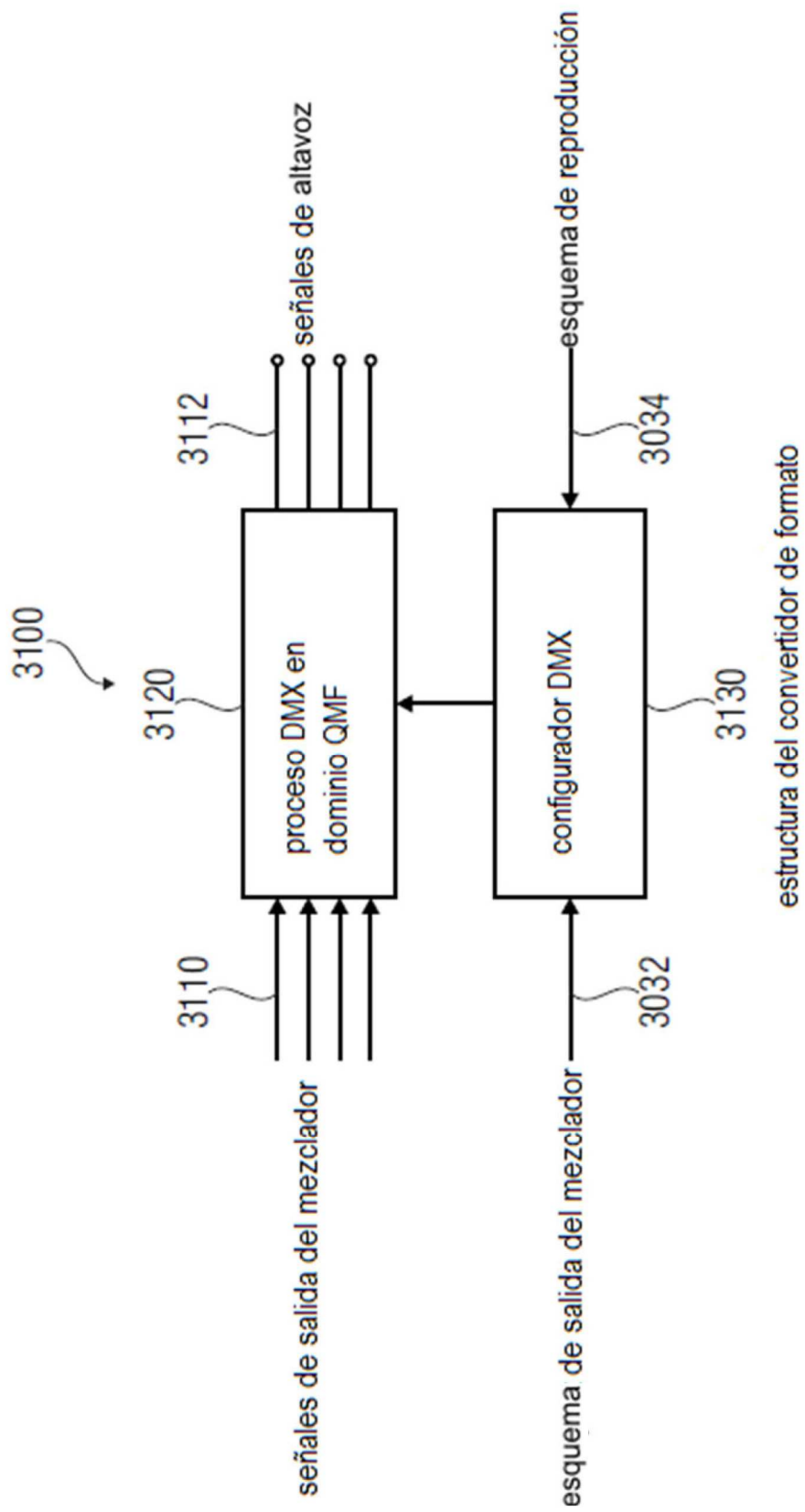


FIG 30

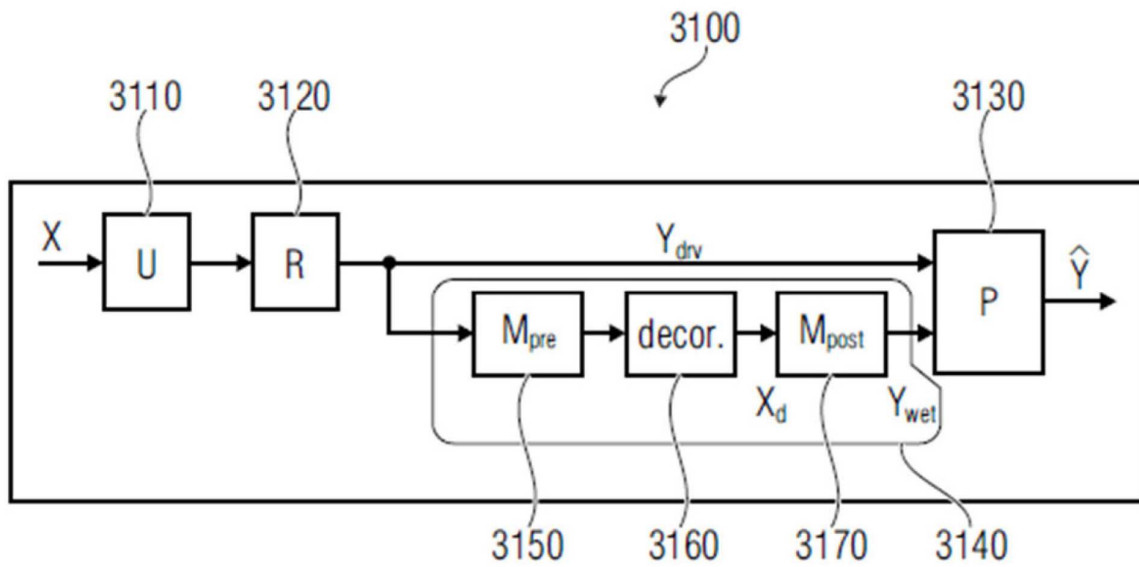


FIG 31

bsNumSaocDmxObjects	modo de decodif.	significado
0	combinado	Las señales basadas en canales de entrada y las señales basadas en objetos de entrada son mezclados de manera descendente juntos en los canales N_{obj}
≥ 1	independiente	Las señales basadas en canales de entrada son mezclados hacia abajo en canales N_{ch} Las señales basadas en objetos de entrada son mezclados de manera descendente juntos en los canales N_{obj}

FIG 32

- Sintaxis de SAOC3DSpecificConfig()

Sintaxis	No. de bits	Mnemonic
SAOC3DSpecificConfig() {		
bsSamplingFrequencyIndex;	4	uimsbf
if (bsSamplingFrequencyIndex == 15) {		
bsSamplingFrequency;	24	uimsbf
}		
bsFreqRes;	3	uimsbf
bsFrameLenth;	7	uimsbf
bsNumSaocDmxChannels;	5	uimsbf
bsNumSaocDmxObjects;	5	uimsbf
bsDecorrelationMethod;	2	uimsbf
NumInputSignals = 0;		
if (bsNumSaocDmxChannels < 0) {		
bsNumSaocChannels;	6	uimsbf
bsNumSaocLFES;	2	uimsbf
NumInputSigns += bsNumSaocChannels;		
}		
bsNumSaocObjects;	8	uimsbf
bsDecorrelationLevel;	2	uimsbf
NumInputSignals += bsNumSaocObjects;		
for (i=0; i<bsNumSaocChannels; i++) {		
bsRelatedTo[i][j] = 1;		
for(j=i+1; j< bsNumSaocChannels; j++) {		
bsRelatedTo[i][j];	1	uimsbf
bsRelatedTo[j][i] = bsRelatedTo[i][j];		
}		
}		

FIG 33A

FIG 33A-1

FIG 33A-2

FIG 33A-1

<pre> for (i=bsNumSaocChannels; i<NumInputSignals; i++) { for (j=0; j<bsNumSaocChannels; j++) { bsRelatedTo[i][j] = 0; bsRelatedTo[j][i] = 0; } } for (i=bsNumSaocChannels; i<NumInputSignals; i++) { bsRelatedTo[i][i] = 1; for (j=i+1; j<NumInputSignals; j++) { bsRelatedTo[i][j]; bsRelatedTo[j][i] = bsRelatedTo[i][j]; } } bsOneIOC; bsSaocDmxMethod; if (bsSaocDmxMethod == 15) { bsNumPremixedChannels; } bsDualMode; if (bsDualMode) { bsBandsLow; bsBandsHigh = numBands; } else { bsBandsLow = numBands; </pre>		
	1	uimsbf
	1	uimsbf
	4	uimsbf
	5	uimsbf
	1	uimsbf
	5	uimsbf
		Note 1

FIG 33A	FIG 33A-1
	FIG 33A-2

FIG 33A-2

}		
bsDcuFlag;	1	uimsbf
if (bsDcuFlag == 1) {		
bsDcuMandatory;	1	uimsbf
bsDcuDynamic;	1	uimsbf
if (bsDcuDynamic == 0) {		
bsDcuMode;	1	uimsbf
bsDcuParam;	4	uimsbf
}		
} else {		
bsDcuMandatory = 0;		
bsDcuDynamic = 0;		
bsDcuMode = 0;		
bsDcuParam = 0;		
}		
bsSaocReserved;	3	uimsbf
ByteAlign();		
SAOC3DExtensionConfig();		
}		
Nota 1: numBands se define en la Tabla 33 en ISO/IEC 23003-2:2010.		

FIG 33B