

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 426 677**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 19/20 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.06.2010** **E 10727721 (2)**

97 Fecha y número de publicación de la concesión europea: **05.06.2013** **EP 2446435**

54 Título: **Descodificador de señal de audio, procedimiento para descodificar una señal de audio y programa de ordenador que utiliza etapas de procesamiento de objetos de audio en cascada**

30 Prioridad:

24.06.2009 US 220042 P

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

24.10.2013

73 Titular/es:

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

72 Inventor/es:

**HELLMUTH, OLIVER;
FALCH, CORNELIA;
HERRE, JÜRGEN;
HILPERT, JOHANNES;
RIDDERBUSCH, FALKO y
TERENTIEV, LEONID**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 426 677 T3

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

DESCRIPCIÓN

Descodificador de señal de audio, procedimiento para descodificar una señal de audio y programa de ordenador que utiliza etapas de procesamiento de objetos de audio en cascada

Sector técnico

5 **[0001]** Las realizaciones según la invención se refieren a un descodificador de señal de audio para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto.

10 **[0002]** Otras realizaciones según la invención se refieren a un procedimiento para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto.

[0003] Otras realizaciones según la invención se refieren a un programa de ordenador.

[0004] Algunas realizaciones según la invención se refieren a un sistema SAOC Karaoke/Solo mejorado.

Antecedentes de la invención

15 **[0005]** En los sistemas de audio modernos, se desea transferir y almacenar información de audio de una manera eficiente en velocidad de bits. Además, a menudo se desea reproducir un contenido de audio utilizando una pluralidad de dos o incluso más altavoces, que se distribuyen espacialmente en una habitación. En estos casos, se desea explotar las capacidades de una disposición como esta de múltiples altavoces para permitir a un usuario identificar espacialmente diferentes contenidos de audio o diferentes partes de un solo contenido de audio. Esto se puede conseguir mediante la distribución de forma individual de los diferentes contenidos de audio a altavoces diferentes.

20 **[0006]** En otras palabras, en la técnica de procesamiento de audio, transmisión de audio y almacenamiento de audio, existe una creciente demanda de gestionar contenidos de varios canales con el fin de mejorar la impresión de audición. El uso de contenido de audio multi-canal trae consigo mejoras significativas para el usuario. Por ejemplo, se puede obtener una impresión auditiva de 3 dimensiones, lo que trae consigo una mejora de la satisfacción del usuario en aplicaciones de entretenimiento. Sin embargo, los contenidos de audio multicanal también son útiles en entornos profesionales, por ejemplo, en aplicaciones de conferencia de teléfono, debido a que se puede mejorar la inteligibilidad de los conferenciantes mediante el uso de una reproducción de audio de múltiples canales.

25 **[0007]** Sin embargo, también es deseable tener un buen compromiso entre la calidad de audio y los requisitos de velocidad de bits con el fin de evitar una carga excesiva de los recursos causada por aplicaciones de múltiples canales.

30 **[0008]** Recientemente, se han propuesto las técnicas paramétricas relativas a la transmisión y / o almacenamiento con tasas de bits eficientes de escenas de audio que contienen múltiples objetos de audio, por ejemplo, Codificación de cola Binaural (Tipo I) (ver, por ejemplo la referencia [BCC]), Codificación de fuente conjunta (ver, por ejemplo, la referencia [JSC]), y Codificación de Objeto de Audio Espacial MPEG (SAOC) (ver, por ejemplo, las referencias [SAOC1], [SAOC2]).

35 **[0009]** Estas técnicas tienen por objeto reconstruir perceptivamente la escena de audio de salida deseada en lugar de hacerlo por coincidencia de forma de onda.

40 **[0010]** La figura 8 muestra una vista general del sistema de dicho sistema (aquí: MPEG SAOC). El Sistema MPEG SAOC 800 mostrado en la figura 8 comprende un codificador SAOC 810 y un descodificador SAOC 820. El codificador SAOC 810 recibe una pluralidad de señales de objeto x_1 a x_N , que pueden ser representadas, por ejemplo, como señales en el dominio del tiempo o como señales en el dominio de tiempo-frecuencia (por ejemplo, en la forma de un conjunto de coeficientes de transformada de una transformada de tipo Fourier o en forma de señales de sub-banda QMF). El codificador SAOC 810 típicamente también recibe coeficientes de mezcla hacia abajo d_1 a d_N , que están asociados con las señales de objeto x_1 a x_N . Unos conjuntos separados de coeficientes de mezcla hacia abajo pueden estar disponibles para cada canal de la señal de mezcla hacia abajo. El codificador SAOC 810 está típicamente configurado para obtener un canal de la señal de mezcla hacia abajo mediante la combinación de las señales de objeto x_1 a x_N de acuerdo con los coeficientes de mezcla hacia abajo asociados d_1 a d_N . Típicamente, hay menos canales de mezcla hacia abajo que señales de objeto x_1 a x_N . Con la finalidad de permitir (al menos de manera aproximada) una separación (o tratamiento por separado) de las señales de objeto en el lado del descodificador SAOC 820, el codificador SAOC 810 proporciona tanto la o las varias señales de mezcla hacia abajo (designadas como canales de mezcla hacia abajo) 812 y una información lateral 814. La información lateral 814 describe características de las señales de objeto x_1 a x_N , con la finalidad de permitir un procesamiento específico de objeto en el lado del descodificador.

[0011] El descodificador SAOC 820 está configurado para recibir tanto la una o varias señales de mezcla hacia abajo 812 como la información lateral 814. También, el descodificador SAOC 820 está típicamente configurado para recibir una información de interacción de usuario y/o una información de control de usuario 822, que describe una configuración de representación deseada. Por ejemplo, la información de interacción de usuario/ información de control de usuario 822 puede describir una configuración de altavoces y la colocación espacial deseada de los objetos suministrados por las señales de objeto x_1 a x_N .

[0012] El descodificador SAOC 820 está configurado para proporcionar, por ejemplo, una pluralidad de señales de canal de mezcla hacia arriba descodificadas y_1 a y_M . Las señales de canal de mezcla hacia arriba pueden por ejemplo estar asociadaa con altavoces individuales de una disposición de representación de múltiples altavoces. El descodificador SAOC 820 puede, por ejemplo, comprender un separador de objetos 820a, que está configurado para reconstruir, al menos de manera aproximada, las señales de objeto x_1 a x_N a partir de la o las varias señales de mezcla hacia abajo 812 y la información lateral 814, obteniéndose de este modo señales de objeto reconstruidas 820b. Sin embargo, las señales de objeto reconstruidas 820b pueden desviarse algo de las señales de objeto originales x_1 a x_N , por ejemplo, porque la información lateral 814 no es suficiente no es suficiente para una reconstrucción perfecta debido a las limitaciones de tasa de bits. El descodificador SAOC 820 puede comprender también un mezclador 820c, que puede estar configurado para recibir las señales de objeto reconstruidas 820b y la información de interacción de usuario/ información de control de usuario 822, y para proporcionar, a partir de esta, las señales de canal de mezcla hacia arriba y_1 a y_M . El mezclador $820c$ puede estar configurado para utilizar la información de interacción de usuario /información de control de usuario 822 para determinar la contribución de las señales de objeto reconstruidas individuales 820b a las señales de canal de mezcla hacia arriba y_1 a y_M . La información de interacción de usuario/ información de control de usuario 822 puede, por ejemplo, comprender parámetros de representación (también designados como coeficientes de representación), que determinan la contribución de las señales de objeto reconstruidas individuales 820b a las señales de canal de mezcla hacia arriba y_1 a y_M .

[0013] Sin embargo, cabe señalar que en muchas realizaciones, la separación de objetos, que está indicada por el separador de objetos 820a en la figura 8, y la mezcla, que se indica por el mezclador 820c en la figura 8, se realizan en una sola etapa. Para este propósito, pueden calcularse unos parámetros globales que describen un cartografiado directo de la o las varias señales de mezcla hacia abajo 812 en las señales de canal de mezcla hacia arriba y_1 a y_M . Estos parámetros se pueden calcular a partir de la información lateral 814 y la información de interacción de usuario/ información de control de usuario 822.

[0014] Con referencia ahora las figuras 9a, 9b y 9c, se describirán diferentes aparatos para obtener una representación de señal de mezcla hacia arriba a partir de una representación de señal de mezcla hacia abajo e información lateral relativa a objeto. La figura 9a muestra un diagrama de bloques esquemático de un sistema MPEG SAOC 900 que comprende un descodificador SAOC 920. El descodificador SAOC 920 comprende, como bloques funcionales separados, un decodificador de objeto 922 y un mezclador / representador 926. El descodificador de objeto 922 proporciona una pluralidad de señales de objeto reconstruidas 924 en función de la representación de señal de mezcla hacia abajo (por ejemplo, en la forma de uno o más señales de mezcla hacia abajo representadas en el dominio del tiempo o en el dominio del tiempo-frecuencia) e información lateral relativa a objeto (por ejemplo, en la forma de metadatos de objetos). El mezclador / representador 926 recibe las señales de objeto reconstruidas 924 asociadas con una pluralidad de N objetos y proporciona, a partir de esta, uno o más señales de canal de mezcla hacia arriba 928. En el descodificador SAOC 920, la extracción de las señales de objeto 924 se realiza de manera separada del mezclado/ representación que permite una funcionalidad de descodificación de separación del objeto de la funcionalidad mezclado/ representación pero trae consigo una complejidad computacional relativamente alta.

[0015] Haciendo referencia a continuación a la figura 9b, se discutirá brevemente otro sistema MPEG SAOC 930, que comprende un descodificador SAOC 950. El descodificador SAOC 950 proporciona una pluralidad de señales de canal de mezcla hacia arriba 958 en función de una representación de señal de mezcla hacia abajo (por ejemplo, en la forma de uno o más señales de mezcla hacia abajo) y una información lateral relativa a objeto (por ejemplo, en la forma de metadatos de objeto). El descodificador SAOC 950 comprende un descodificador de objeto y mezclador / representador combinado, que está configurado para obtener las señales de canal de mezcla hacia arriba 958 en un proceso de mezclado conjunto sin una separación de la descodificación del objeto y el mezclado/ representación, en el que los parámetros para dicho proceso de mezclado hacia arriba conjunto dependen de ambos, la información lateral relativa a objeto y la información de representación. El proceso de mezclado hacia arriba conjunto también depende de la información de mezcla hacia abajo, que se considera como parte de la información lateral relativa a objeto.

[0016] Para resumir lo anterior, la provisión de las señales de canal de mezcla hacia arriba 928, 958 se puede realizar en un proceso de una sola etapa o en un proceso de dos etapas.

[0017] Haciendo referencia a continuación a la figura 9c, se describirá un sistema MPEG SAOC 960. El sistema SAOC 960 comprende un transcodificador SAOC a MPEG Surround 980, en lugar de un descodificador SAOC.

- 5 **[0018]** El transcodificador SAOC a MPEG Surround comprende un transcodificador de información lateral 982, que está configurado para recibir la información lateral relativa a objeto (por ejemplo, en la forma de metadatos de objeto) y, de manera opcional, información acerca de la o las varias señales de mezcla hacia abajo y la información de representación. El transcodificador de información lateral también está configurado para proporcionar una información lateral MPEG Surround 984 (por ejemplo, en la forma de una corriente de bits MPEG Surround) a partir de datos recibidos. En consecuencia, el transcodificador de información lateral 982 está configurado para transformar una información lateral relativa a objeto (paramétrica), que se alivia del codificador de objeto, en una relativa a canal información lateral (paramétrica) 984, teniendo en cuenta la información de representación y, de manera opcional, la información sobre el contenido de la o las varias señales de mezcla hacia abajo.
- 10 **[0019]** De manera opcional, el transcodificador SAOC a MPEG Surround 980 puede estar configurado para manipular la o las varias señales de mezcla hacia abajo, descritas, por ejemplo, por la representación de señal de mezcla hacia abajo, para obtener una representación manipulada de la señal de mezcla hacia abajo 988. Sin embargo, se puede omitir el manipulador de señal de mezcla hacia abajo 986, de modo que la representación de señal de mezcla hacia abajo de salida 988 del transcodificador SAOC a MPEG Surround 980 es idéntica a la representación de señal de mezcla hacia abajo de entrada del transcodificador SAOC a MPEG Surround. Podría utilizarse el manipulador de señal de mezcla hacia abajo 986, por ejemplo, si la información lateral relativa a canal MPEG Surround 984 no permitiera proporcionar una impresión de audición deseada a partir de la representación de señal de mezcla hacia abajo de entrada del transcodificador SAOC a MPEG Surround 980, que puede ser el caso en algunas constelaciones de representación.
- 15 **[0020]** En consecuencia, el transcodificador SAOC a MPEG Surround 980 proporciona la representación de señal de mezcla hacia abajo 988 y la corriente de bits MPEG Surround 984 de modo que una pluralidad de señales de canal de mezcla hacia arriba, que representan los objetos de audio de acuerdo con la información de representación de entrada al transcodificador SAOC a MPEG Surround 980 pueden ser generadas empleando un Descodificador MPEG Surround que recibe la corriente de bits MPEG Surround 984 y la representación de señal de mezcla hacia abajo 988.
- 20 **[0021]** Para resumir, se pueden utilizar los diferentes conceptos anteriores para descodificar señales de audio codificadas SAOC. En algunos casos, se utiliza un descodificador SAOC, que proporciona señales de canal de mezcla hacia arriba (por ejemplo, señales de canal de mezcla hacia arriba 928, 958) en función de la representación de señal de mezcla hacia abajo y la información lateral relativa a objeto paramétrica. Se pueden ver ejemplos de este concepto en las figuras 9a y 9b. Como alternativa, la información de audio codificada mediante SAOC puede ser transcodificada para obtener una representación de señal de mezcla hacia abajo (por ejemplo, una representación de señal de mezcla hacia abajo 988) y una información lateral relativa a canal (por ejemplo, la Corriente de bits MPEG Surround relativa a canal 984), que puede ser usada por un descodificador MPEG Surround para proporcionar las señales de canal de mezcla hacia arriba deseadas.
- 25 **[0022]** En el sistema MPEG SAOC 800, cuya vista general se ofrece en la figura 8, el procesamiento general se lleva a cabo de una manera selectiva según la frecuencia y puede ser descrito de la siguiente manera dentro de cada banda de frecuencia:
- 30 • N señales de audio de objeto de entrada x_1 a x_N se mezclan hacia abajo como parte del procesamiento de codificación SAOC. Para un mezclado hacia abajo mono, los coeficientes de mezcla hacia abajo se designan por d_1 a d_N . Además, el codificador SAOC 810 extrae información lateral 814 que describe las características de los objetos de audio de entrada. Para MPEG SAOC, las potencias de relaciones de objeto con respecto a cada una de las demás son la forma más básica de dicha información lateral.
 - 35 • La señal de mezcla hacia abajo (o señales) 812 e información lateral 814 se transmiten y/o almacenan. Con esta finalidad, la señal de audio de mezcla hacia abajo puede ser comprimida empleando los codificadores de audio de percepción bien conocidos, tales como MPEG-1 Layer II o III (también conocidos como ".mp3"), MPEG Advanced Audio Coding (AAC), o cualquier otro codificador de audio.
 - 40 • En el extremo receptor, el descodificador SAOC 820 trata conceptualmente de restaurar la señal de objeto original ("separación de objeto") empleando la información lateral transmitida 814 (y, naturalmente, la o las varias señales de mezcla hacia abajo 812). Estas señales de objeto aproximadas (también designadas como señales de objeto reconstruidas 820b) se mezclan entonces en una escena objetivo representada por M canales de salida de audio (que pueden, por ejemplo, ser representados por las señales de canal de mezcla hacia arriba y_1 a y_M) empleando una matriz de representación. Para una salida mono, los coeficientes de matriz de representación están dados por r_1 a r_N .
 - 45 • Efectivamente, la separación de las señales de objetos rara vez se ejecuta (o incluso nunca), puesto que tanto la etapa de separación (indicada por el separador de objetos 820a) y la etapa de mezclado (indicada por el mezclador 820c) se combinan en una sola etapa de transcodificación, que a menudo resulta en una enorme reducción en la complejidad de cálculo.
- 50
- 55

[0023] Se ha comprobado que este sistema es tremendamente eficiente, tanto en términos de tasa de bits de transmisión (sólo es necesario transmitir unos canales de mezclado hacia abajo y alguna información lateral en lugar de N señales de audio de objetos discretos o un sistema discreto) y de complejidad de cálculo (la complejidad de tratamiento se refiere principalmente al número de canales de salida en lugar del número de objetos de audio). Otras ventajas para el usuario en el extremo receptor incluyen la libertad de elegir una configuración de representación de su elección (mono, estéreo, Surround, reproducción de auriculares virtualizada, y así sucesivamente) y la característica de la interactividad del usuario: la matriz de representación, y por lo tanto la escena de salida, se pueden ajustar y modificar interactivamente por el usuario de acuerdo con su voluntad, la preferencia personal u otros criterios. Por ejemplo, es posible localizar los oradores de un grupo conjuntamente en una zona espacial para maximizar la discriminación de otros oradores restantes. Esta interactividad se consigue proporcionando una interfaz decodificadora de usuario.

[0024] Para cada objeto de sonido transmitido, se pueden ajustar su nivel relativo y (para representación no-mono) la posición espacial de la representación. Esto puede suceder en tiempo real a medida que el usuario cambia la posición de los reguladores de la interfaz gráfica de usuario asociada (GUI) (por ejemplo: nivel de objeto = 5 dB, posición del objeto = -30deg).

[0025] Sin embargo, se ha encontrado que es difícil manejar objetos de audio de diferentes tipos de objetos de audio en un sistema de este. En particular, se ha encontrado que es difícil procesar objetos de audio de diferentes tipos de objetos de audio, por ejemplo, objetos de audio que tienen información lateral diferente asociada, si no está predeterminado el número total de objetos de audio a procesar.

[0026] En vista de esta situación, es un objetivo de la presente invención crear un concepto, que permita una decodificación eficiente desde el punto de vista del cálculo y flexible de una señal de audio que comprende una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto, en el que la información paramétrica relativa a un objeto describe objetos de audio de dos o más diferentes tipos de objetos de audio.

Resumen de la invención

[0027] Este objetivo se consigue con un decodificador de señal de audio para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto, un procedimiento para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto, y un programa de ordenador, tal como se define en las reivindicaciones independientes.

[0028] Una realización según la invención crea un decodificador de señal de audio para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto. El decodificador de señal de audio comprende un separador de objetos configurado para descomponer la representación de señal de mezcla hacia abajo, para proporcionar una primera información de audio que describe un primer conjunto de uno o más objetos de audio de un primer tipo de objeto de audio, y una segunda información de audio que describe un segundo conjunto de uno o más objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y empleando al menos una parte de la información paramétrica relativa a un objeto ^^

en el que la segunda información de audio es una información de audio que describe los objetos de audio del segundo tipo de objeto de audio de una manera combinada. El decodificador de señal de audio también comprende un procesador de señal de audio configurado para recibir la segunda información de audio y para procesar la segunda información de audio en función de la información paramétrica relativa a un objeto, para obtener una versión procesada de la segunda información de audio. El decodificador de señal de audio también comprende un combinador de señal de audio configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio para obtener la representación de señal de mezcla hacia arriba.

[0029] Es una idea clave de la presente invención que un procesamiento eficiente de diferentes tipos de objetos de audio se puede obtener con una estructura en cascada, que permite una separación de los diferentes tipos de objetos de audio empleando al menos una parte de la información paramétrica relativa a un objeto en una primera etapa de procesamiento realizada por el separador de objetos, y que permite un procesamiento espacial adicional en una segunda etapa de procesamiento realizada en función de al menos una parte de la información paramétrica relativa a un objeto por el procesador de señal de audio.

[0030] Se ha constatado que extraer una segunda información de audio, que comprende objetos de audio del segundo tipo de objeto de audio, desde una representación de señal de mezcla hacia abajo puede ser realizada con una complejidad moderada incluso si hay un gran número de objetos de audio del segundo tipo de objeto de audio. Además, se ha encontrado que un procesamiento espacial de los objetos de audio del segundo tipo de audio puede ser realizada eficientemente una vez que la segunda información de audio se ha separado de la primera información de audio que describe los objetos de audio del primer tipo de objeto de audio.

5 [0031] Adicionalmente, se ha encontrado que el algoritmo de procesamiento realizada por el separador de objetos para procesar la primera información de audio y la segunda información de audio puede ser realizada con complejidad comparativamente pequeña si el procesamiento individual por objeto de los objetos de audio del segundo tipo de objeto de audio se pospone al procesador de señal de audio y no se realiza al mismo tiempo que la separación de la primera información de audio y la segunda información de audio.

10 [0032] Por lo tanto, el descodificador de señal de audio se vuelve a configurar para proporcionar la representación de señal de mezcla hacia arriba en función de la representación de señal de mezcla hacia abajo, la información paramétrica relativa a un objeto y una información residual asociada con un subconjunto de objetos de audio representados por la representación de señal de mezcla hacia abajo. En este caso, el separador de objetos está configurado para descomponer la representación de señal de mezcla hacia abajo para proporcionar la primera información de audio que describe el primer conjunto de uno o más objetos de audio (por ejemplo, objetos de primer plano FGO) del primer tipo de objeto de audio al cual la información residual está asociada y describiendo la segunda información de audio el segundo conjunto de uno o más objetos de audio (por ejemplo, objetos de segundo plano FGO) del segundo tipo de objeto de audio que no tiene información residual asociada en función de la representación de señal de mezcla hacia abajo y empleando al menos parte de la información paramétrica relativa a un objeto y la información residual.

20 [0033] El procesador de señal de audio está configurado para recibir la segunda información de audio y para procesar la segunda información de audio en función de (al menos una parte de) la información paramétrica relativa a un objeto, empleando información paramétrica relativa a un objeto asociado con más de dos objetos de audio del segundo tipo de objeto de audio. En consecuencia, se realiza un procesamiento individual por objeto por el procesador de audio, mientras que este procesamiento individual por objeto no se realiza para objetos de audio del segundo tipo de objeto de audio por el separador de objetos.

25 [0034] Esto se basa en el hallazgo de que con una separación especialmente precisa entre la primera información de audio que describe el primer conjunto de objetos de audio del primer tipo de objeto de audio y describiendo la segunda información de audio el segundo conjunto de objetos de audio del segundo tipo de objeto de audio se puede obtener empleando una información residual además a la información paramétrica relativa a un objeto. Se ha encontrado que el mero uso de la información paramétrica relativa a un objeto daría como resultado distorsiones en muchos casos, que se podrían reducir significativamente o incluso enteramente eliminadas por la utilización de información residual. La información residual describe una distorsión residual, que se prevé que permanezca si un objeto de audio del primer tipo de objeto de audio se aísla simplemente empleando la información paramétrica relativa a un objeto. La información residual es estimada típicamente por un codificador de señal de audio. Aplicando la información residual, se puede mejorar la separación entre los objetos de audio del primer tipo de objeto de audio y los objetos de audio del segundo tipo de objeto de audio.

35 [0035] Esto permite obtener la primera información de audio y la segunda información de audio con separación especialmente buena entre los objetos de audio del primer tipo de objeto de audio y los objetos de audio del segundo tipo de objeto de audio, los cuales, a su vez, permite conseguir un procesamiento espacial de calidad elevada de los objetos de audio del segundo tipo de objeto de audio cuando se procesa la segunda información de audio en el procesador de señal de audio.

40 [0036] Según una realización preferida, el separador de objetos está por lo tanto configurado para proporcionar la primera información de audio de modo que objetos de audio del primer tipo de objeto de audio se enfatice por encima de objetos de audio del segundo tipo de objeto de audio en la primera información de audio. El separador de objetos también está configurado para proporcionar la segunda información de audio de modo que objetos de audio del segundo tipo de objeto de audio se enfatice por encima de objetos de audio del primer tipo de objeto de audio en el segunda información de audio.

45 [0037] Según una realización preferida, el descodificador de señal de audio está configurado para realizar un procesamiento de dos etapas, de modo que un procesamiento de la segunda información de audio en el procesador de señal de audio se realiza de manera subsiguiente a una separación entre la primera información de audio que describe el primer conjunto de uno o más objetos de audio del primer tipo de objeto de audio y describiendo la segunda información de audio el segundo conjunto de uno o más objetos de audio del segundo tipo de objeto de audio.

50 [0038] Según una realización preferida, el procesador de señal de audio está configurado para procesar la segunda información de audio en función de la información paramétrica relativa a un objeto asociado con los objetos de audio del segundo tipo de objeto de audio e independiente de la información paramétrica relativa a un objeto asociado con los objetos de audio del primer tipo de objeto de audio. En consecuencia, se puede obtener un procesamiento separado de los objetos de audio del primer tipo de objeto de audio y los objetos de audio del segundo tipo de objeto de audio.

55 [0039] Según una realización preferida, el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio empleando una combinación lineal de uno o más canales de mezcla hacia abajo y uno o más canales residuales. En este caso, el separador de objetos está configurado para

5 obtener parámetros de combinación para realizar la combinación lineal en función de parámetros de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio y en función de coeficientes de predicción de canal de los objetos de audio del primer tipo de objeto de audio. El cálculo de los coeficientes de predicción de canal de los objetos de audio del primer tipo de objeto de audio pueden, por ejemplo, tener en cuenta los objetos de audio del segundo tipo de objeto de audio como un objeto de audio único y común. En consecuencia, se puede realizar un proceso de separación con complejidad de cálculo suficientemente pequeña, que puede, por ejemplo, ser casi independiente del número de objetos de audio del segundo tipo de objeto de audio.

10 **[0040]** Según una realización preferida, el separador de objetos está configurado para aplicar una matriz de representación a la primera información de audio para cartografiar señales de objeto de la primera información de audio sobre canales de audio de la representación de señal de audio de mezcla hacia arriba. Esto se puede hacer, porque el separador de objetos puede ser capaz de extraer señales de audio separadas individualmente que representan los objetos de audio del primer tipo de objeto de audio. En consecuencia, posible cartografiar las señales de objeto de la primera información de audio directamente en los canales de audio de la representación de señal de audio de mezcla hacia arriba.

15 **[0041]** Según una realización preferida, el procesador de audio está configurado para realizar a procesado estéreo de la segunda información de audio en función de una información de representación, una información de covarianza relativa a objeto y una información de mezcla hacia abajo, para obtener canales de audio de la representación de señal de audio de mezcla hacia arriba.

20 **[0042]** En consecuencia, el procesado estéreo de los objetos de audio del segundo tipo de objeto de audio se ha separado de la separación entre los objetos de audio del primer tipo de objeto de audio y los objetos de audio del segundo tipo de objeto de audio. Por lo tanto, la separación eficiente entre objetos de audio del primer tipo de objeto de audio y objetos de audio del segundo tipo de objeto de audio no se ve afectada (o degradada) por el procesado estéreo, que conduce típicamente a una distribución de objetos de audio sobre una pluralidad de canales de audio sin proporcionar el grado elevado de separación de objetos, que se puede obtener en el separador de objetos, por ejemplo, empleando la información residual.

25 **[0043]** Según otra realización preferida, el procesador de señal de audio está configurado para realizar un postprocesado de la segunda información de audio en función de una información de representación, una información de covarianza relativa a objeto y una información de mezcla hacia abajo. Esta forma de postprocesado permite una colocación espacial de los objetos de audio del segundo tipo de objeto de audio en una escena de audio. Sin embargo, debido al concepto en cascada, la complejidad de cálculo del procesador de señal de audio puede mantenerse suficientemente pequeña, puesto que el procesador de audio no necesita considerar la información paramétrica relativa a objeto asociada con los objetos de audio del primer tipo de objeto de audio.

30 **[0044]** Además, diferentes tipos de procesado pueden ser realizados por el procesador de audio, tal como, por ejemplo, un procesado mono a binaural, un procesado mono a estéreo, un procesado estéreo a binaural o un procesado estéreo a estéreo.

35 **[0045]** Según una realización preferida, el separador de objetos está configurado para procesar objetos de audio del segundo tipo de objeto de audio, que no tiene información residual asociada, como un objeto de audio único. Además, el procesador de señal de audio está configurado para considerar parámetros de representación específicos del objeto para ajustar contribuciones de los objetos del segundo tipo de objeto de audio a la representación de señal de mezcla hacia arriba. Por lo tanto, los objetos de audio del segundo tipo de objeto de audio se consideran como un objeto de audio único por el separador de objetos, que reduce significativamente la complejidad del separador de objetos y también permite tener una información residual única, que es independiente a partir de los parámetros de representación asociados con los objetos de audio del segundo tipo de objeto de audio.

40 **[0046]** Según una realización preferida, el separador de objetos está configurado para obtener un valor de diferencia de niveles de objetos común para una pluralidad de objetos de audio del segundo tipo de objeto de audio. El separador de objetos está configurado para utilizar el valor de diferencia de niveles de objetos común para un cálculo de coeficientes de predicción de canal. Además, el separador de objetos está configurado para utilizar los coeficientes de predicción de canal para obtener uno o dos canales de audio que representan la segunda información de audio. Para obtener un valor de diferencia de niveles de objetos común, los objetos de audio del segundo tipo de objeto de audio pueden ser manejados eficientemente como un objeto de audio único por el separador de objetos.

45 **[0047]** Según una realización preferida, el separador de objetos está configurado para obtener a valor de diferencia de niveles de objetos común para una pluralidad de objetos de audio del segundo tipo de objeto de audio y el separador de objetos está configurado para utilizar el valor de diferencia de niveles de objetos común para un cálculo de entradas de una matriz de cartografiado en modo energía. El separador de objetos está configurado para utilizar la matriz de cartografiado en modo energía para obtener el o los más canales de audio que representan la segunda información de audio. De nuevo, el valor de diferencia de niveles de objetos común permite un tratamiento común eficiente desde el punto de vista del cálculo de los objetos de audio del segundo tipo de objeto de audio por el separador de objetos.

5 **[0048]** Según una realización preferida, el separador de objetos está configurado para obtener selectivamente un valor de correlación entre objetos común asociado con los objetos de audio del segundo tipo de objeto de audio en función de la información paramétrica relativa a un objeto si se comprueba que hay dos objetos de audio del segundo tipo de objeto de audio y para ajustar el valor de correlación entre objetos asociado con los objetos de audio del segundo tipo de objeto de audio a cero si se comprueba que hay más o menos que dos objetos de audio del segundo tipo de objeto de audio.

10 **[0049]** El separador de objetos está configurado para utilizar el valor de correlación entre objetos común asociado con los objetos de audio del segundo tipo de objeto de audio para obtener el o los más canales de audio que representan la segunda información de audio. Empleando este enfoque, se explota el valor de correlación entre objetos si se puede obtener con gran eficiencia de cálculo, es decir si hay dos objetos de audio del segundo tipo de objeto de audio. De otro modo, sería costoso desde el punto de vista del cálculo obtener valores de correlación entre objetos. En consecuencia, se ha encontrado es un buen compromiso en términos de impresión de audición y de complejidad de cálculo ajustar el valor de correlación entre objetos asociado con los objetos de audio del segundo tipo de objeto de audio a cero si hay más o menos que dos objetos de audio del segundo tipo de objeto.

15 **[0050]** Según una realización preferida, el procesador de señal de audio está configurado para representar la segunda información de audio en función de (al menos una parte de) la información paramétrica relativa a un objeto, para obtener una representación parametrizada de los objetos de audio del segundo tipo de objeto de audio como una versión procesada de la segunda información de audio. En este caso, la representación puede hacerse independiente de los objetos de audio del primer tipo de objeto de audio.

20 **[0051]** Según una realización preferida, el separador de objetos está configurado para proporcionar la segunda información de audio de modo que la segunda información de audio describa más de dos objetos de audio del segundo tipo de objeto de audio. Algunas realizaciones según la invención permiten un ajuste flexible del número de objetos de audio del segundo tipo de objeto de audio, que se facilita significativamente por la estructura en cascada del procesamiento.

25 **[0052]** Según una realización preferida, el separador de objetos está configurado para obtener, como la segunda información de audio, una representación de señal de audio de un canal o una representación de señal de audio de dos canales que representa más de dos objetos de audio del segundo tipo de objeto de audio. La extracción de uno o dos señal de canales de audio puede ser realizada por el separador de objetos con baja complejidad de cálculo. En particular, la complejidad del separador de objetos puede mantenerse significativamente menor en comparación con un caso en el que el separador de objetos necesitaría tratar con más de dos objetos de audio del segundo tipo de objeto de audio. Sin embargo, se ha encontrado que es una representación eficiente desde el punto de vista del cálculo de los objetos de audio del segundo tipo de objeto de audio utilizar uno o dos canales de una señal de audio.

30 **[0053]** Según una realización preferida, el descodificador de audio está configurado para extraer una información de número total de objetos y una información de número de objetos de primer plano a partir de una información de configuración relacionada con la información paramétrica relativa a un objeto. El descodificador de audio también está configurado para determinar un número de objetos de audio del segundo tipo de objeto de audio por formar una diferencia entre la información de número total de objetos y la información de número de objetos de primer plano. En consecuencia, se logra la señalización eficiente del número de objetos de audio del segundo tipo de objeto de audio. Además, este concepto proporciona un elevado grado de flexibilidad con respecto al número de objetos de audio del segundo tipo de objeto de audio.

35 **[0054]** Según una realización preferida, el separador de objetos está configurado para utilizar información paramétrica relativa a un objeto asociado con N_{eao} objetos de audio del primer tipo de objeto de audio para obtener, como la primera información de audio, N_{eao} señales de audio que representan (preferentemente, individualmente) los N_{eao} objetos de audio del primer tipo de objeto de audio, y para obtener, como la segunda información de audio, una o dos señales de audio que representan los $N - N_{eao}$ objetos de audio del segundo tipo de objeto de audio, procesar los $N - N_{eao}$ objetos de audio del segundo tipo de objeto de audio como un objeto de audio de un único canal o de dos canales. El procesador de señal de audio está configurado para representar individualmente los $N - N_{eao}$ objetos de audio representados por la única o las dos señales de audio de la segunda información de audio empleando la información paramétrica relativa a un objeto asociado con los $N - N_{eao}$ objetos de audio del segundo tipo de objeto de audio. En consecuencia, la separación de objeto de audio entre los objetos de audio del primer tipo de objeto de audio y los objetos de audio del segundo tipo de objeto de audio se ha separado del procesamiento subsiguiente de los objetos de audio del segundo tipo de objeto de audio.

40 **[0055]** Una realización según la invención crea un procedimiento para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto, tal como se define en la reivindicación independiente 29.

45 **[0056]** Otra realización según la invención crea un programa de ordenador para realizar dicho procedimiento, tal como se define en la reivindicación independiente 30.

Breve Descripción de las figuras.

[0057] Algunas realizaciones según la invención se describirán a continuación con referencia a las figuras adjuntas, en las cuales:

- 5 La figura 1 muestra un diagrama de bloques esquemático de un descodificador de señal de audio, según una realización de la invención;
- La figura 2 muestra un diagrama de bloques esquemático de otro descodificador de señal de audio, según una realización de la invención;
- Las figuras 3a y 3b muestran unos diagramas de bloque esquemáticos de un procesador residual, que se puede usar como un separador de objetos en una realización de la invención;
- 10 Las figuras 4a a 4e muestran diagramas de bloque esquemáticos de señales de procesador de audio, que se pueden usar en un descodificador de señal de audio según una realización de la invención;
- La figura 4f muestra un diagrama de bloques de un modo de procesamiento de transcodificador SAOC;
- La figura 4g muestra un diagrama de bloques de un modo de procesamiento de descodificador SAOC;
- 15 La figura 5a muestra un diagrama de bloques esquemático de un descodificador de señal de audio, según una realización de la invención;
- La figura 5b muestra un diagrama de bloques esquemático de otro descodificador de señal de audio, según una realización de la invención;
- La figura 6a muestra una tabla que representa un diseño de prueba de audición descripción;
- La figura 6b muestra una tabla que representa sistemas bajo prueba;
- 20 La figura 6c muestra una tabla que representa los elementos de prueba de audición y matrices de representación;
- La figura 6d muestra una representación gráfica de valores MUSHRA promediados para una prueba de audición de representación de tipo Karaoke/Solo;
- La figura 6e muestra una representación gráfica de valores MUSHRA promediados para una prueba de audición de representación clásica;
- 25 La figura 7 muestra un diagrama de flujo de un procedimiento para proporcionar una representación de señal de mezcla hacia arriba, según una realización de la invención;
- La figura 8 muestra un diagrama de bloques esquemático de un sistema MPEG SAOC de referencia;
- La figura 9a muestra un diagrama de bloques esquemático de un sistema SAOC de referencia empleando un descodificador y mezclador separado;
- 30 La figura 9b muestra un diagrama de bloques esquemático de un sistema SAOC de referencia empleando un descodificador y mezclador integrado; y
- La figura 9c muestra un diagrama de bloques esquemático de un sistema SAOC de referencia empleando un transcodificador SAOC a MPEG.

Descripción detallada de las realizaciones

35 1. Descodificador de señal de audio según la figura 1

[0058] La figura 1 muestra un diagrama de bloques esquemático de un descodificador de señal de audio 100 según una realización de la invención.

- 40 **[0059]** El descodificador de señal de audio 100 está configurado para recibir una información paramétrica relativa a un objeto 110 y una representación de señal de mezcla hacia abajo 112. El descodificador de señal de audio 100 está configurado para proporcionar una representación de señal de mezcla hacia arriba 120 en función de la representación de señal de mezcla hacia abajo y la información paramétrica relativa a un objeto 110. El descodificador de señal de audio 100 comprende un separador de objetos 130, que está configurado para descomponer la representación de señal de mezcla hacia abajo 112 para proporcionar una primera información de audio 132 que describe un primer conjunto de uno o más objetos de audio de un primer tipo de objeto de audio y una
- 45 segunda información de audio 134 que describe un segundo conjunto de uno o más objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo 112 y empleando al menos una parte de la información paramétrica relativa a un objeto 110. El descodificador de señal de audio 100 también

comprende un procesador de señal de audio 140, que está configurado para recibir la segunda información de audio 134 y para procesar la segunda información de audio en función de al menos una parte de la información paramétrica relativa a un objeto 112, para obtener una versión procesada 142 de la segunda información de audio 134. El descodificador de señal de audio 100 también comprende un combinador de señal de audio 150 configurado para combinar la primera información de audio 132 con la versión procesada 142 de la segunda información de audio 134, para obtener la representación de señal de mezcla hacia arriba 120.

[0060] El descodificador de señal de audio 100 implementa un procesamiento en cascada de la representación de señal de mezcla hacia abajo, que representa objetos de audio del primer tipo de objeto de audio y objetos de audio del segundo tipo de objeto de audio de una manera combinada.

[0061] En una primera etapa de procesamiento, que es realizada por el separador de objetos 130, que describe la segunda información de audio un segundo conjunto de objetos de audio del segundo tipo de objeto de audio se ha separado de la primera información de audio 132 que describe un primer conjunto de objetos de audio de un primer tipo de objeto de audio empleando la información paramétrica relativa a un objeto 110. Sin embargo, la segunda información de audio 134 es típicamente una información de audio (por ejemplo, una señal de audio de un canal o una señal de audio de dos canales) que describe los objetos de audio del segundo tipo de objeto de audio de una manera combinada.

[0062] En la segunda etapa de procesamiento, el procesador de señal de audio 140 procesa la segunda información de audio 134 en función de la información paramétrica relativa a un objeto. En consecuencia, el procesador de señal de audio 140 es capaz de realizar un procesamiento individual por objeto o representar los objetos de audio del segundo tipo de objeto de audio, que se describen por la segunda información de audio 134, y no se realizan típicamente por el separador de objetos 130.

[0063] Por lo tanto, mientras que los objetos de audio del segundo tipo de objeto de audio preferentemente no se procesan de manera individual por objeto por el separador de objetos 130, los objetos de audio del segundo tipo de objeto de audio se procesan, en efecto, de manera individual por objeto (por ejemplo, representados de manera individual por objeto) en la segunda etapa de procesamiento, que es realizada por el procesador de señal de audio 140. Por lo tanto, la separación entre los objetos de audio del primer tipo de objeto de audio y los objetos de audio del segundo tipo de objeto de audio, que es realizada por el separador de objetos 130, se ha separado del procesamiento individual por objeto de los objetos de audio del segundo tipo de objeto de audio, que es realizada después por el procesador de señal de audio 140. En consecuencia, el procesamiento que es realizado por el separador de objetos 130 es sustancialmente independiente de un número de objetos de audio del segundo tipo de objeto de audio. Además, el formato (por ejemplo, señal de audio de un canal o la señal de audio de dos canales) de la segunda información de audio 134 es típicamente independiente del número de objetos de audio del segundo tipo de objeto de audio. Por lo tanto, el número de objetos de audio del segundo tipo de objeto de audio se puede variar sin tener que modificar la estructura del separador de objetos 130. Dicho de otro modo, los objetos de audio del segundo tipo de objeto de audio son tratados como un único objeto de audio (por ejemplo, de un canal o de dos canales) para el cual una información paramétrica común relativa a un objeto (por ejemplo, un valor de diferencia de niveles de objetos común asociado con uno o dos canales de audio) se obtiene por el separador de objetos 140.

[0064] En consecuencia, el descodificador de señal de audio 100 según la figura 1 es capaz de manejar un número variable de objetos de audio del segundo tipo de objeto de audio sin una modificación estructural del separador de objetos 130. Además, se pueden aplicar diferentes algoritmos de procesamiento de objetos de audio por el separador de objetos 130 y el procesador de señal de audio 140. En consecuencia, por ejemplo, es posible realizar una separación de objeto de audio empleando una información residual por el separador de objetos 130, que permite una separación especialmente buena de diferentes objetos de audio, empleando la información residual, que constituye una información lateral para mejorar la calidad de una separación de objetos. En contraste, el procesador de señal de audio 140 puede realizar un procesamiento individual por objeto sin emplear una información residual. Por ejemplo, el procesador de señal de audio 140 puede estar configurado para realizar un procesamiento de señal de audio de tipo codificación de objeto de audio espacial (SAOC de *spatial-audio-object-coding*) para representar los diferentes objetos de audio.

2. Descodificador de señal de audio según la figura 2

[0065] A continuación, se describirá un descodificador de señal de audio 200 según una realización de la invención. Un diagrama de bloques esquemático de este descodificador de señal de audio 200 se muestra en la figura 2.

[0066] El descodificador de audio 200 está configurado para recibir una señal de mezcla hacia abajo 210, una así llamada corriente de bits SAOC 212, una información de matriz de representación 214 y, de manera opcional, parámetros de función de transferencia relativos a cabecera (HRTF de *head-related-transfer-function*) 216. El descodificador de señal de audio 200 también está configurado para proporcionar una señal de mezcla hacia abajo de salida/MPS 220 y (de manera opcional) una corriente de bits MPS 222.

2.1. Señales de entrada y señales de salida del descodificador de señal de audio 200

[0067] A continuación, se describirán varios detalles relativos a señales de entrada y señales de salida del descodificador de audio 200.

5 [0068] La señal de mezcla hacia abajo 200 puede, por ejemplo, ser una señal de audio de un canal o una señal de audio de dos canales. La señal de mezcla hacia abajo 210 puede, por ejemplo, derivarse de una representación codificada de la señal de mezcla hacia abajo.

10 [0069] La corriente de bits codificada de objeto de audio espacial (corriente de bits SAOC) 212 puede, por ejemplo, comprender información paramétrica relativa a un objeto. Por ejemplo, la corriente de bits SAOC 212 puede comprender información de diferencia de niveles de objetos, por ejemplo, en la forma de parámetros de diferencia de nivel de objeto OLD, una información de correlación entre objetos, por ejemplo, en la forma de parámetros de correlación entre objetos IOC.

15 [0070] Además, la corriente de bits SAOC 212 puede comprender una información de mezcla hacia abajo que describe de que manera se han proporcionado las señales de mezcla hacia abajo a partir de una pluralidad de señales de objeto de audio empleando un proceso de mezcla hacia abajo. Por ejemplo, la corriente de bits SAOC puede comprender un parámetro de ganancia de mezcla hacia abajo DMG y (de manera opcional) parámetros de diferencia de niveles entre canales DCLD.

[0071] La información de matriz de representación 214 puede, por ejemplo, describir de que manera los diferentes objetos de audio se deberían representar por el descodificador de audio. Por ejemplo, la información de matriz de representación 214 puede describir una asignación de un objeto de audio a uno o más canales de la señal de mezcla hacia abajo de salida/MPS 220.

20 [0072] La información de parámetros de función de transferencia relativos a cabecera (HRTF) 216 opcional también puede describir una función de transferencia para derivar una señal de auriculares binaural.

25 [0073] La señal de mezcla hacia debajo de salida/MPEG-Surround (también designada brevemente por "señal de mezcla hacia abajo de salida/MPS") 220 representa uno o más canales de audio, por ejemplo, en la forma de una representación de señal de audio en el dominio del tiempo o una representación de señal de audio en el dominio de la frecuencia. Individualmente o en combinación con la corriente de bits opcional MPEG-Surround (Corriente de bits MPS) 222, que comprende parámetros MPEG-Surround que describen un cartografiado de la señal de mezcla hacia abajo de salida/MPS 220 sobre una pluralidad de canales de audio, se forma una representación de señal de mezcla hacia arriba.

2.2. Estructura y funcionalidad del descodificador de señal de audio 200

30 [0074] A continuación, se describirá con más detalle la estructura del descodificador de señal de audio 200, que puede cumplir la funcionalidad de un transcodificador SAOC o la funcionalidad de un descodificador SAOC.

35 [0075] El descodificador de señal de audio 200 comprende un procesador de mezcla hacia abajo 230, que está configurado para recibir la señal de mezcla hacia abajo 210 y para proporcionar, a partir de esta, la señal de mezcla hacia abajo de salida/MPS 220. El procesador de mezcla hacia abajo 230 también está configurado para recibir al menos una parte de la información de corriente de bits SAOC 212 y al menos una parte de la información de matriz de representación 214. Además, el procesador de mezcla hacia abajo 230 también puede recibir una información de parámetros SAOC procesada 240 desde un procesador de parámetros 250.

40 [0076] El procesador de parámetros 250 está configurado para recibir la información de corriente de bits SAOC 212, la información de matriz de representación 214 y, de manera opcional, la información de parámetros de función de transferencia relativos a cabecera 260, y para proporcionar, a partir de esta, la corriente de bits MPEG Surround 222 que lleva Parámetros MPEG Surround (si se necesitan los parámetros MPEG Surround, lo cual es cierto, por ejemplo, en el modo de operación de transcodificación). Además, el procesador de parámetros 250 proporciona la información SAOC procesada 240 (si esta información SAOC procesada se necesita).

45 [0077] A continuación, se describirá con más detalle la estructura y funcionalidad del procesador de mezcla hacia abajo 230.

50 [0078] El procesador de mezcla hacia abajo 230 comprende un procesador residual 260, que está configurado para recibir la señal de mezcla hacia abajo 210 y para proporcionar, a partir de esta, un primera señal de objeto de audio 262 que describe los así llamados objetos de audio mejorados (EAOs), que se pueden considerar como objetos de audio de un primer tipo de objeto de audio. La primera señal de objeto de audio puede comprender uno o más canales de audio y puede ser considerado como una primera información de audio. El procesador residual 260 también está configurado para proporcionar una segunda señal de objeto de audio 264, que describe objetos de audio de un segundo tipo de objeto de audio y puede ser considerado como una segunda información de audio. La segunda señal de objeto de audio 264 puede comprender uno o más canales y puede comprender típicamente uno o dos canales de audio que describen una pluralidad de objetos de audio. Típicamente, la segunda señal de objeto de audio puede describir incluso más de dos objetos de audio del segundo tipo de objeto de audio.

55

[0079] El procesador de mezcla hacia abajo 230 también comprende un preprocesador de mezcla hacia abajo SAOC 270, que está configurado para recibir la segunda señal de objeto de audio 264 y para proporcionar, a partir de esta, una versión procesada 272 de la segunda señal de objeto de audio 264, que se puede considerar como una versión procesada de la segunda información de audio.

5 **[0080]** El procesador de mezcla hacia abajo 230 también comprende un combinador de señal de audio 280, que está configurado para recibir la primera señal de objeto de audio 262 y la versión procesada 272 de la segunda señal de objeto de audio 264, y para proporcionar, a partir de esta, la señal de mezcla hacia abajo de salida/MPS 220, que se puede considerar, individualmente o junto con la corriente de bits MPEG-Surround correspondiente 222 (opcional), como una representación de señal de mezcla hacia arriba.

10 **[0081]** A continuación, se tratará con más detalle la funcionalidad de las unidades individuales del procesador de mezcla hacia abajo 230.

[0082] El procesador residual 260 está configurado para proporcionar de manera separada la primera señal de objeto de audio 262 y la segunda señal de objeto de audio 264. Para este propósito, el procesador residual 260 puede estar configurado para aplicar al menos una parte de la información de corriente de bits SAOC 212. Por ejemplo, el procesador residual 260 puede estar configurado para evaluar una información paramétrica relativa a un objeto asociado con los objetos de audio del primer tipo de objeto de audio, es decir los así llamados "objetos de audio mejorados" EAO. Además, el procesador residual 260 puede estar configurado para obtener una información general que describe los objetos de audio del segundo tipo de objeto de audio, por ejemplo, los así llamados comúnmente "objetos de audio no-mejorados". El procesador residual 260 también puede estar configurado para evaluar una información residual, que se proporciona en la información de corriente de bits SAOC 212, para una separación entre objetos de audio mejorados (objetos de audio del primer tipo de objeto de audio) y objetos de audio no-mejorados (objetos de audio del segundo tipo de objeto de audio). La información residual puede, por ejemplo, codificar una señal residual en el dominio del tiempo, que se aplica para obtener una separación especialmente limpia entre los objetos de audio mejorados y los objetos de audio no mejorados. Además, el procesador residual 260 puede, de manera opcional, evaluar al menos una parte de la información de matriz de representación 214, por ejemplo, con la finalidad de determinar una distribución de los objetos de audio mejorados a los canales de audio de la primera señal de objeto de audio 262.

[0083] El preprocesador de mezcla hacia abajo SAOC 270 comprende un redistribuidor de canal 274, que está configurado para recibir el o los más canales de audio de la segunda señal de objeto de audio 264 y para proporcionar, a partir de estos, uno o más (típicamente dos) canales de audio de la segunda señal de objeto de audio procesado 272. Además, el preprocesador de mezcla hacia abajo SAOC 270 comprende un proveedor de señal descorrelacionada 276, que está configurado para recibir el o los más canales de audio de la segunda señal de objeto de audio 264 y para proporcionar, a partir de estos, una o más señales descorrelacionadas 278a, 278b, que se añaden a las señales suministradas por el redistribuidor de canal 274 con la finalidad de obtener la versión procesada 272 de la segunda señal de objeto de audio 264.

[0084] Más detalles relativos al procesador de mezcla hacia abajo SAOC se discuten a continuación.

[0085] El combinador de señales de audio 280 combina la primera señal de objeto de audio 262 con la versión procesada 272 de la segunda señal de objeto de audio. Para este propósito, se puede realizar una combinación de canales. En consecuencia, se obtiene la señal de mezcla hacia abajo de salida/MPS 220.

40 **[0086]** El procesador de parámetros 250 está configurado para obtener los Parámetros MPEG-Surround (opcionales), que constituyen la corriente de bits MPEG-Surround 222 de la representación de señal de mezcla hacia arriba, a partir de la corriente de bits SAOC, considerando la información de matriz de representación 214 y, de manera opcional, la información de parámetros HRTF 216. Dicho de otro modo, el procesador de parámetros SAOC 252 está configurado para traducir la información de parámetros relativa a objeto, que está descrita por la información de corriente de bits SAOC 212, en una información paramétrica relativa a canal, que está descrita por la corriente de bits MPEG Surround 222.

[0087] A continuación, se dará una breve visión general de la estructura de la arquitectura transcodificador / decodificador SAOC que se muestra en la figura 2. La codificación espacial de objetos de audio (SAOC) es una técnica de codificación paramétrica de varios objetos. Está diseñada para transmitir un número de objetos de audio en una señal de audio (por ejemplo la señal de audio de mezcla hacia abajo 210) que comprende M canales. Junto con esta señal de mezcla hacia abajo compatible con versiones anteriores, se transmiten unos parámetros de objeto (por ejemplo, empleando la información de corriente de bits SAOC 212) que permite la reconstrucción y la manipulación de las señales de objeto originales. Un codificador SAOC (no mostrado aquí) produce una mezcla hacia abajo de las señales de objeto en su entrada y extrae estos parámetros de objeto. El número de objetos que pueden ser manejados está en principio no limitado. Los parámetros de objeto se cuantifican y codifican de manera eficiente en la corriente de bits SAOC 212. La señal de mezcla hacia abajo 210 puede ser comprimida y transmitida puede ser comprimida. Los parámetros de objeto, o información lateral SAOC, se transmite un canal lateral de baja velocidad de bits, por ejemplo, la parte de datos auxiliares del flujo de bits de mezcla hacia abajo.

[0088] En el lado del decodificador, se reconstruyen los objetos de entrada y se representan en un cierto número de canales de reproducción. La información de representación que contiene los niveles de reproducción y la posición de panorámica para cada objeto es suministrada por el usuario o se puede extraer de la corriente de bits SAOC (por ejemplo, como una información preestablecida). La información de representación puede ser variable en el tiempo. Los escenarios de salida pueden variar de canales mono a múltiples (por ejemplo, 5.1) y son independientes tanto del número de objetos de entrada como del número de canales de mezcla hacia abajo. Es posible la representación binaural de objetos, incluyendo el azimut y la elevación de las posiciones de los objetos virtuales. Una interfaz de efectos opcional permite la manipulación avanzada de señales de objetos, además de la modificación de nivel y panorámica.

[0089] Los propios objetos pueden ser señales mono, I estereofónicas, así como señales de múltiples canales (por ejemplo canales 5.1). Las configuraciones típicos de mezcla hacia abajo son mono y estéreo.

[0090] A continuación, se explicará la estructura básica del transcodificador/descodificador SAOC, que se muestra en la figura 2. El módulo transcodificador/descodificador SAOC aquí descrito puede actuar ya sea como un descodificador o como un transcodificador autónomo desde una corriente de bits SAOC a MPEG-Surround, dependiendo de la configuración de canal de salida prevista. En un primer modo de funcionamiento, la configuración de la señal de salida es mono, estéreo o binaural, y se utilizan dos canales de salida. En este primer caso, el módulo SAOC puede operar en un modo de decodificador, y la salida del módulo SAOC es una salida de código de pulso modulado (salida PCM). En el primer caso, no se necesita un decodificador MPEG. Más bien, la representación de señal de mezcla hacia arriba puede comprender solamente la señal de salida 220, mientras que se puede omitir la provisión de la corriente de bits MPEG Surround 222. En un segundo caso, la configuración de la señal de salida es una configuración de canales múltiples con más de dos canales de salida. El módulo SAOC puede funcionar en un modo transcodificador. La salida del módulo SAOC puede comprender tanto una señal de mezcla hacia abajo 220 y como una corriente de bits MPEG Surround 222 en este caso, tal como se muestra en la figura 2. En consecuencia, se necesita un descodificador MPEG Surround con la finalidad de obtener una representación de la señal de audio de salida final para por los altavoces.

[0091] La figura 2 muestra la estructura básica de la arquitectura transcodificador/descodificador SAOC. El procesador residual 216 extrae el objeto de audio ampliado de la señal de mezcla hacia debajo de entrada 210 empleando la información residual contenida en la corriente de bits SAOC 212. El preprocesador de mezcla hacia abajo 270 procesa los objetos de audio regulares (que son, por ejemplo, objetos de audio no mejorados, es decir, objetos de audio para los cuales no se transmite información residual en la corriente de bits SAOC 212). Los objetos de audio mejorados (representados por la primera señal de objeto de audio 262) y los objetos de audio regulares procesados (representados, por ejemplo, por la versión procesada 272 de la segunda señal de objeto de audio 264) se combinan con la señal de salida 220 por el modo descodificador SAOC o la señal de mezcla hacia abajo MPEG Surround 220 en el modo transcodificador SAOC. A continuación se ofrecen descripciones detalladas de los bloques de procesamiento.

3. Arquitectura y funcionalidad del procesador Residual y del procesador modo Energía

[0092] A continuación, se describirá detalles relativos a un procesador residual, que pueden, por ejemplo, hacerse cargo de la funcionalidad del separador de objetos 130 del descodificador de señal de audio 100 o del procesador residual 260 del descodificador de señal de audio 200. Para este propósito, las figuras 3a y 3b muestran diagramas de bloque esquemáticos de este procesador residual 300, que puede ocupar el lugar del separador de objetos 130 o del procesador residual 260. La figura 3a muestra menos detalles que la figura 3b. Sin embargo, al siguiente descripción se aplica al procesador residual 300 según la figura 3a y también al procesador residual 380 según la figura 3b.

[0093] El procesador residual 300 está configurado para recibir una señal de mezcla hacia abajo SAOC 310, que puede ser equivalente a la representación de señal de mezcla hacia abajo 112 de la figura 1 o la representación de señal de mezcla hacia abajo 210 de la figura 2. El procesador residual 300 está configurado para proporcionar, a partir de esta, una primera información de audio 320 que describe uno o más objetos de audio mejorados, que pueden, por ejemplo, ser equivalentes a la primera información de audio 132 o al primera señal de objeto de audio 262. También, el procesador residual 300 puede proporcionar una segunda información de audio 322 que describe uno o más otros objetos de audio (por ejemplo, objetos de audio no-mejorados, para los cuales no hay información residual disponible), en el que la segunda información de audio 322 puede ser equivalente a la segunda información de audio 134 o al segunda señal de objeto de audio 264.

[0094] El procesador residual 300 comprende una unidad 1-a-N/2-a-N (unidad OTN/TTN) 330, que recibe la señal de mezcla hacia abajo SAOC 310 y que también recibe datos y residuos SAOC d 332. La unidad 1-a-N/2-a-N 330 también proporciona una señal de objeto de audio mejorada 334, que describe los objetos de audio mejorados (EAO) contenidos en la señal de mezcla hacia abajo SAOC 310. Así mismo, la unidad 1-a-N/2-a-N 330 proporciona la segunda información de audio 322. El procesador residual 300 también comprende una unidad de representación 340, que recibe la señal de objeto de audio mejorada 334 y una información de matriz de representación 342 y proporciona, a partir de esta, la primera información de audio 320.

[0095] A continuación, se describirá con más detalle el procesamiento de objeto de audio ampliado (procesado EAO), que es realizado por el procesador residual 300,.

3.1. Introducción al funcionamiento del procesador residual 300

5 [0096] En cuanto a la funcionalidad del procesador residual 300, hay que señalar que la tecnología SAOC permite la manipulación individual de un número de objetos de audio en términos de sus niveles de amplificación / atenuación sin disminución significativa en la calidad de sonido resultante sólo de una manera muy limitada. Un escenario de aplicación especial "de tipo karaoke" requiere una supresión total (o casi total) de los objetos específicos, típicamente la voz principal, manteniendo sin dañar la calidad de percepción de la escena del sonido de fondo.

10 [0097] Un caso de aplicación típico contiene hasta cuatro señales de objetos de audio mejoradas (EAO), que pueden, por ejemplo, representar dos objetos estéreo independientes (por ejemplo, dos objetos estéreo independientes que se preparan para ser eliminados en el lado del decodificador).

15 [0098] Cabe señalar que los (uno o más) objetos de audio de calidad mejorada (o, más precisamente, las contribuciones de la señal de audio asociadas con los objetos de audio mejorados) se incluyen en la señal de mezcla hacia abajo SAOC 310. Por lo general, se mezclan las contribuciones de la señal de audio asociadas con los (uno o más) objetos de audio mejorados, por el procesamiento de mezcla hacia abajo realizado por el codificador de señal de audio, con contribuciones de señales de audio de otros objetos de audio, que no son objetos de audio mejorados. También, hay que destacar que las contribuciones de señales de audio de una pluralidad de objetos de audio mejorados por lo general también se solapan o mezclan por el procesamiento de mezcla hacia abajo realizado por el codificador de señal de audio.

20 3.2 Arquitectura SAOC de soporte de los objetos de audio mejorados

25 [0099] A continuación, se describen detalles sobre el procesador residual 300. El procesamiento de objetos de audio mejorados incorpora las unidades 1-a-N o 2-a-N, dependiendo del modo de mezcla hacia abajo SAOC. La unidad de procesamiento 1-a-N está dedicada a una señal de mezcla hacia abajo mono y la unidad de procesamiento 2-a-N está dedicada a una señal de mezcla hacia abajo estéreo 310. Estas dos unidades representan una modificación generalizada y mejorada de la caja 2-a-2 (TTT box) conocida a partir de ISO/IEC 23003-1: 2007. En el codificador se combinan las señales regular y EAO en la mezcla hacia abajo. Las unidades de procesamiento OTN-1/TTN-1 (que son unidades de procesamiento uno-a-N inversas o unidades de procesamiento 2-a-N inversas) se emplean para producir y codificar las señales residuales correspondientes.

30 [0100] Las señales EAO y regulares se recuperan de la mezcla hacia abajo 310 por las unidades OTN/TTN 330 empleando la información lateral SAOC y las señales residuales incorporadas. Las EAOs recuperadas (que están descritas por la señal de objeto de audio mejorada 334) se alimentan a la unidad de representación 340 que representa (o proporciona) el producto de la matriz de representación correspondiente (descrita por la información de matriz de representación 342) y la salida resultante de la unidad OTN/TTN. Los objetos de audio regulares (que están descritos por la segunda información de audio 322) se suministran al preprocesador de mezcla hacia abajo SAOC, por ejemplo, el preprocesador de mezcla hacia abajo SAOC 270, para un procesamiento posterior. Las figuras 3a y 3b muestran la estructura general del procesador residual, es decir, la arquitectura del procesador residual.

[0101] Las señales de salida del procesador residual 320,322 se calculan como

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ} \mathbf{X}_{res} ,$$

$$\mathbf{X}_{EAO} = \mathbf{A}_{EAO} \mathbf{M}_{EAO} \mathbf{X}_{res} ,$$

40 donde X_{OBJ} representa la señal de mezcla hacia abajo de los objetos de audio regulares (es decir no-EAOs) y X_{EAO} es la señal de salida EAO representada para el modo de descodificación SAOC o la señal de mezcla hacia abajo EAO correspondiente para el modo de transcodificación SAOC.

45 [0102] El procesador residual puede funcionar en modo predicción (empleando información residual) o modo energía (sin información residual). La señal de entrada extendida X_{res} se define en consecuencia:

$$\mathbf{X}_{res} = \begin{cases} \begin{pmatrix} \mathbf{X} \\ \text{---} \\ \mathbf{res} \end{pmatrix}, & \text{Para el modo predicción} \\ \mathbf{X}, & \text{Para el modo energía} \end{cases}$$

5 **[0103]** Aquí, X puede, por ejemplo, representar el o los más canales de la representación de señal de mezcla hacia abajo 310, que puede ser transportada en la corriente de bits que representa el contenido de audio multi-canal. res puede designar una o más señales residuales, que pueden ser descritas por la corriente de bits que representa el contenido de audio multi-canal.

[0104] El procesamiento OTN/TTN está representado por la matriz **M** y el procesador EAO por la matriz **A_{EAO}**.

[0105] La matriz de procesamiento OTN/TTN **M** se define según el modo de operación EAO (es decir predicción o energía) como

$$\mathbf{M} = \begin{cases} \mathbf{M}_{Prediction}, & \text{Para el modo predicción} \\ \mathbf{M}_{Energy}, & \text{Para el modo energía} \end{cases}$$

10 **[0106]** La matriz de procesamiento OTN/TTN **M** está representado como

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{OBJ} \\ \text{---} \\ \mathbf{M}_{EAO} \end{pmatrix},$$

donde la matriz **M_{OBJ}** se refiere a los objetos de audio regulares (es decir no-EAOs) y **M_{EAO}** a los objetos de audio mejorados (EAOs).

15 **[0107]** En algunas realizaciones, uno o más objetos de fondo multicanal (MBO) pueden ser tratados de la misma manera por el procesador residual 300.

20 **[0108]** Un Objeto de Fondo Multi-canal (MBO) es una mezcla hacia abajo mono o stereo MPS que es parte de la mezcla hacia abajo SAOC. En oposición al empleo de objetos SAOC individuales para cada canal en una señal multi-canal, se puede usar un MBO que permita a un SAOC manejar más eficientemente un objeto multi-canal. En el caso MBO, el encabezamiento SAOC se vuelve menor ya que los parámetros SAOC del MBO solamente se refieren a los canales de mezcla hacia abajo en lugar de a todos los canales de mezcla hacia arriba.

3.3 Otras definiciones

3.3.1 Dimensionalidad de señales y parámetros

[0109] A continuación, se discutirá brevemente la dimensionalidad de las señales y parámetros con el fin de proporcionar una comprensión de con qué frecuencia se realizan los diferentes cálculos.

25 **[0110]** Las señales de audio se definen para cada ranura de tiempo n, y cada subbanda híbrida (que puede ser una subbanda de frecuencia) k. Los parámetros SAOC correspondientes se definen para cada ranura de tiempo de parámetros 1 y banda de procesamiento m. Un cartografiado posterior entre el dominio híbrido y de parámetros se especifica mediante la tabla A.31 ISO/IEC 23003-1:2007. Por lo tanto, todos los cálculos se realizan con respecto a determinados índices tiempo / banda y las dimensionalidades correspondientes son implícitas para cada variable introducida.

30 **[0111]** Sin embargo, a continuación, los índices de tiempo y banda de frecuencia se omitirán para mantener una notación concisa.

3.3.2 Cálculo de la matriz **A_{EAO}**

35 **[0112]** La matriz de pre-representación EAO **A_{EAO}** se define según el número de canales de salida (es decir mono, stereo o binaural) como

$$\mathbf{A}_{EAO} = \begin{cases} \mathbf{A}_1^{EAO}, & \text{Para el caso mono} \\ \mathbf{A}_2^{EAO}, & \text{Para los casos restantes} \end{cases}$$

[0113] Las matrices \mathbf{A}_1^{EAO} de tamaño $1 \times N_{EAO}$ y \mathbf{A}_2^{EAO} de tamaño $2 \times N_{EAO}$ se definen como

$$\mathbf{A}_1^{EAO} = \mathbf{D}_{16}^{EAO} \mathbf{M}_{ren}^{EAO}, \quad \mathbf{D}_{16}^{EAO} = \begin{pmatrix} w_1^{EAO} & w_2^{EAO} & w_3^{EAO} & w_3^{EAO} & w_1^{EAO} & w_2^{EAO} \end{pmatrix},$$

$$\mathbf{A}_2^{EAO} = \mathbf{D}_{26}^{EAO} \mathbf{M}_{ren}^{EAO}, \quad \mathbf{D}_{26}^{EAO} = \begin{pmatrix} w_1^{EAO} & 0 & \frac{w_3^{EAO}}{\sqrt{2}} & \frac{w_3^{EAO}}{\sqrt{2}} & w_1^{EAO} & 0 \\ 0 & w_2^{EAO} & \frac{w_3^{EAO}}{\sqrt{2}} & \frac{w_3^{EAO}}{\sqrt{2}} & 0 & w_2^{EAO} \end{pmatrix},$$

5 donde la sub-matriz de representación \mathbf{M}_{ren}^{EAO} corresponde a la representación EAO (y describe un cartografiado deseado de objetos de audio mejorados en los canales de la representación de señal de mezcla hacia arriba).

[0114] Los valores w_i^{EAO} se calculan en función de información de representación asociada con los objetos de audio mejorados empleando los elementos EAO correspondientes y empleando las ecuaciones de la sección 4.2.2.1.

10 [0115] En caso de representación binaural la matriz \mathbf{A}_2^{EAO} se define por las ecuaciones dadas en la sección 4.1.2, para el cual la matriz de representación binaural objetivo correspondiente contiene solamente elementos relacionados con EAO.

3.4 Cálculo de los elementos OTN/TTN en el modo residual

15 [0116] A continuación, se tratará de que manera la señal de mezcla hacia abajo SAOC 310, que típicamente comprende uno o dos canales de audio, se cartografía sobre la señal de objeto de audio mejorado 334, que típicamente comprende uno o más canales de objeto de audio mejorados, y la segunda información de audio 322, que típicamente comprende uno o dos canales de objeto de audio regulares.

20 [0117] La funcionalidad de la unidad 1-a-N o unidad 2-a-N 330 puede, por ejemplo, implementarse empleando una multiplicación matriz vector, de modo que se obtiene un vector que describe tanto los canales de la señal de objeto de audio mejorado 334 como los canales de la segunda información de audio 322 multiplicando un vector que describe los canales de la señal de mezcla hacia abajo SAOC 310 y (de manera opcional) una o más señales residuales con una matriz $\mathbf{M}^{Predicción}$ o $\mathbf{M}^{Energía}$. En consecuencia, la determinación de la matriz $\mathbf{M}^{Predicción}$ o $\mathbf{M}^{Energía}$ es una etapa importante en la obtención de la primera información de audio 320 y la segunda información de audio 322 a partir de la mezcla hacia abajo SAOC 310.

25 [0118] En resumen, el proceso de mezcla hacia arriba OTN/TTN se presenta ya sea por una matriz $\mathbf{M}^{Predicción}$ para un modo de predicción o $\mathbf{M}^{Energía}$ para un modo energía.

30 [0119] El proceso de codificación/descodificación basado en energía está diseñado una codificación que no preserva la longitud de onda de la señal de mezcla hacia abajo. Por lo tanto la matriz de mezcla hacia arriba OTN/TTN para el modo de energía correspondiente no se basa en formas de onda específicas, si no que solamente describe una distribución de energías relativas de los objetos de audio de entrada, tal como se describirá con más detalle más abajo.

3.4.1 Modo de predicción

[0120] Para el modo de predicción la matriz $\mathbf{M}_{\text{Predicción}}$ se define como que explota la información de mezcla hacia abajo contenida en la matriz \mathbf{D}^{-1} y los datos CPC de la matriz \mathbf{C} :

$$\mathbf{M}_{\text{Predicción}} = \tilde{\mathbf{D}}^{-1} \mathbf{C}.$$

5 **[0121]** Con respecto a los varios modos SAOC, la matriz de mezcla hacia abajo extendida \mathbf{D} y la matriz CPC \mathbf{C} exhibe las dimensiones y estructuras siguientes:

3.4.1.1 Modos de mezcla hacia abajo estéreo (TTN):

10 **[0122]** Para los modos de mezcla hacia abajo estéreo (TTN) (por ejemplo, para el caso de una mezcla hacia abajo estéreo a partir de dos canales de objetos de audio regulares y NEAO canales de objetos de audio mejorado), la matriz de mezcla hacia abajo (extendida) \mathbf{D} y la matriz CPC \mathbf{C} se pueden obtener tal como se expone a continuación:

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \dots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \dots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \dots & -1 \end{array} \right),$$

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \dots & 1 \end{array} \right).$$

[0123] Con una mezcla hacia abajo estéreo, cada EAO j tiene dos CPCs $c_{j,0}$ y $c_{j,1}$ que dan la matriz \mathbf{C} .

[0124] Las señales de salida del procesador residual se calculan como

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Prediction} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix},$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Prediction} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}.$$

5 **[0125]** En consecuencia, se obtienen dos señales yL, yR (que están representadas por XOBJ), que representan uno o dos o incluso más de dos objetos de audio regulares (también designados como objetos de audio no extendidos). También, se obtienen las señales NEAO (representadas por XEAO) que representan NEAO objetos de audio mejorados. Estas señales se obtienen a partir de dos señales de mezcla hacia abajo SAOCs l0,r0 y NEAO señales residuales res0 a resNEAO-1, que se codificarán en la información lateral SAOC, por ejemplo, como una parte de la información paramétrica relativa a un objeto.

[0126] Debe destacarse que las señales yL y yR pueden ser equivalentes a la señal 322, y que las señales Y0,EAO a YNEAO-1, EAO (que están representadas por XEAO) pueden ser equivalentes a las señales 320.

10 **[0127]** La matriz AEAO es una matriz de representación. Las entradas de la matriz AEAO puede describir, por ejemplo, un cartografiado de objetos de audio mejorados con los canales de la señal de objeto de audio mejorada 334 (XEAO).

[0128] En consecuencia, una selección apropiada de la matriz AEAO puede permitir una integración opcional de la funcionalidad de la unidad de representación 340, de modo que la multiplicación del vector que describe los canales (l0,r0) de la señal de mezcla hacia abajo SAOC 310 y una o más señales residuales (res0,...,resNEAO-1) con la matriz

15 $\mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Prediction}$ pueden dar directamente como resultado una representación XEAO de la primera información de audio 320.

3.4.1.2 Modos de mezcla hacia abajo mono (OTN):

20 **[0129]** A continuación, se describirá la derivación de las señales de objeto de audio mejoradas 320 (o, como alternativa, de las señales de objeto de audio mejoradas 334) y de la señal de objeto de audio regular 322 para el caso en el que las señal de mezcla hacia abajo SAOC 310 comprenden solamente un canal de señal.

[0130] Para los modos de mezcla hacia abajo mono (OTN) (por ejemplo, una mezcla hacia abajo mono a partir de un canal de objeto de audio regular y NEAO canales de objeto de audio mejorados), la matriz de mezcla (extendida) hacia abajo D y la matriz CPC C se pueden obtener tal como se expone a continuación:

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \dots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \dots & -1 \end{array} \right),$$

$$\mathbf{C} = \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline c_{0,0} & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1,0} & 0 & \dots & 1 \end{array} \right).$$

[0131] Con una mezcla hacia abajo mono, se predice un EAO j mediante solamente un coeficiente c_j que da la matriz C. Todos los elementos de matriz c_j se obtienen, por ejemplo, a partir de los parámetros SAOC (por ejemplo, a partir de los datos SAOC 322) según las relaciones proporcionadas más abajo (sección 3.4.1.4).

[0132] Las señales de salida del procesador residual se calculan como

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Prediction} \left(\begin{array}{c} d_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array} \right),$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Prediction} \left(\begin{array}{c} d_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{array} \right).$$

[0133] La señal de salida XOBJ comprende, por ejemplo, un canal que describe los objetos de audio regulares (objetos de audio no mejorados). La señal de salida XEAO comprende, por ejemplo, uno, dos, o incluso más canales que describen los objetos de audio mejorados (preferentemente NEAO canales que describen los objetos de audio mejorados). De nuevo, dichas señales con equivalentesa las señales 320, 322.

5 3.4.1.3 Cálculo de la matriz de mezcla hacia abajo extendida inversa

[0134] La matriz \mathbf{D}^{-1} es la inversa de la matriz de mezcla hacia abajo extendida \mathbf{D} y \mathbf{C} implica los CPCs.

[0135] La matriz \mathbf{D}^{-1} es la inversa de la matriz de mezcla hacia abajo extendida \mathbf{D} y se puede calcular como

$$\tilde{\mathbf{D}}^{-1} = \frac{\tilde{d}_{i,j}}{den}.$$

10 **[0136]** Los elementos $d_{i,j}$ (por ejemplo, de la inversa \mathbf{D}^{-1} de la matriz de mezcla hacia abajo extendida \mathbf{D} de tamaño 6 3 6) se obtienen empleando los siguientes valores

$$\tilde{d}_{1,1} = 1 + \sum_{j=1}^4 n_j^2,$$

:

$$\bar{d}_{1,1} = 1 + \sum_{j=1}^4 n_j^2,$$

$$\bar{d}_{1,2} = - \left(\sum_{j=1}^4 m_j n_j \right),$$

$$\bar{d}_{1,3} = m_1 + m_1 n_2^2 + m_1 n_3^2 + m_1 n_4^2 - m_2 n_1 n_2 - m_3 n_1 n_3 - m_4 n_1 n_4,$$

$$\bar{d}_{1,4} = m_2 + m_2 n_1^2 + m_2 n_3^2 + m_2 n_4^2 - m_1 n_2 n_1 - m_3 n_2 n_3 - m_4 n_2 n_4,$$

$$\bar{d}_{1,5} = m_3 + m_3 n_1^2 + m_3 n_2^2 + m_3 n_4^2 - m_1 n_3 n_1 - m_2 n_3 n_2 - m_4 n_3 n_4,$$

$$\bar{d}_{1,6} = m_4 + m_4 n_1^2 + m_4 n_2^2 + m_4 n_3^2 - m_1 n_4 n_1 - m_2 n_4 n_2 - m_3 n_4 n_3,$$

$$\bar{d}_{2,2} = 1 + \sum_{j=1}^4 m_j^2,$$

$$\bar{d}_{2,3} = n_1 + n_1 m_2^2 + n_1 m_3^2 + n_1 m_4^2 - m_1 m_2 n_2 - m_1 m_3 n_3 - m_1 m_4 n_4,$$

$$\bar{d}_{2,4} = n_2 + n_2 m_1^2 + n_2 m_3^2 + n_2 m_4^2 - m_2 m_1 n_1 - m_2 m_3 n_3 - m_2 m_4 n_4,$$

$$\bar{d}_{2,5} = n_3 + n_3 m_1^2 + n_3 m_2^2 + n_3 m_4^2 - m_3 m_1 n_1 - m_3 m_2 n_2 - m_3 m_4 n_4,$$

$$\bar{d}_{2,6} = n_4 + n_4 m_1^2 + n_4 m_2^2 + n_4 m_3^2 - m_4 m_1 n_1 - m_4 m_2 n_2 - m_4 m_3 n_3,$$

$$\bar{d}_{3,3} = -1 - \sum_{j=2}^4 m_j^2 - \sum_{j=2}^4 n_j^2 - m_3^2 n_2^2 - m_4^2 n_2^2 - m_2^2 n_3^2 - m_4^2 n_3^2 - m_2^2 n_4^2 - m_3^2 n_4^2 + 2m_2 m_3 n_2 n_3 + 2m_2 m_4 n_2 n_4 + 2m_3 m_4 n_3 n_4$$

$$, \bar{d}_{3,4} = m_1 m_2 + n_1 n_2 + m_3^2 n_1 n_2 + m_4^2 n_1 n_2 + m_1 m_2 n_3^2 + m_1 m_2 n_4^2 - m_2 m_3 n_1 n_3 - m_1 m_3 n_2 n_3 - m_2 m_4 n_1 n_4 - m_1 m_4 n_2 n_4,$$

$$\bar{d}_{3,5} = m_1 m_3 + n_1 n_3 + m_2^2 n_1 n_3 + m_4^2 n_1 n_3 + m_1 m_3 n_2^2 + m_1 m_3 n_4^2 - m_2 m_3 n_1 n_2 - m_1 m_2 n_2 n_3 - m_3 m_4 n_1 n_4 - m_1 m_4 n_3 n_4,$$

$$\bar{d}_{3,6} = m_1 m_4 + n_1 n_4 + m_2^2 n_1 n_4 + m_3^2 n_1 n_4 + m_1 m_4 n_2^2 + m_1 m_4 n_3^2 - m_2 m_4 n_1 n_2 - m_3 m_4 n_1 n_3 - m_1 m_2 n_2 n_4 - m_1 m_3 n_3 n_4,$$

$$\bar{d}_{4,4} = -1 - \sum_{j=1}^4 m_j^2 - \sum_{j=1}^4 n_j^2 - m_2^2 n_1^2 - m_4^2 n_1^2 - m_1^2 n_2^2 - m_4^2 n_2^2 - m_1^2 n_3^2 - m_2^2 n_4^2 + 2m_1 m_2 n_1 n_2 + 2m_1 m_4 n_1 n_4 + 2m_2 m_4 n_2 n_4,$$

$$\bar{d}_{4,5} = m_2 m_3 + n_2 n_3 + m_1^2 n_2 n_3 + m_4^2 n_2 n_3 + m_2 m_3 n_1^2 + m_2 m_3 n_4^2 - m_1 m_3 n_1 n_2 - m_1 m_2 n_1 n_3 - m_3 m_4 n_2 n_4 - m_2 m_4 n_3 n_4,$$

$$\bar{d}_{4,6} = m_2 m_4 + n_2 n_4 + m_1^2 n_2 n_4 + m_3^2 n_2 n_4 + m_2 m_4 n_1^2 + m_2 m_4 n_3^2 - m_1 m_4 n_1 n_2 - m_3 m_4 n_2 n_3 - m_1 m_2 n_1 n_4 - m_2 m_3 n_3 n_4,$$

$$\bar{d}_{5,5} = -1 - \sum_{j=1}^4 m_j^2 - \sum_{j=3}^4 n_j^2 - m_2^2 n_1^2 - m_4^2 n_1^2 - m_1^2 n_2^2 - m_4^2 n_2^2 - m_1^2 n_4^2 - m_2^2 n_4^2 + 2m_1 m_2 n_1 n_2 + 2m_1 m_4 n_1 n_4 + 2m_2 m_4 n_2 n_4,$$

$$\bar{d}_{5,6} = m_3 m_4 + n_3 n_4 + m_1^2 n_3 n_4 + m_2^2 n_3 n_4 + m_3 m_4 n_1^2 + m_3 m_4 n_2^2 - m_1 m_4 n_1 n_3 - m_2 m_4 n_2 n_3 - m_1 m_3 n_1 n_4 - m_2 m_3 n_2 n_4,$$

$$\bar{d}_{6,6} = -1 - \sum_{j=1}^3 m_j^2 - \sum_{j=1}^3 n_j^2 - m_2^2 n_1^2 - m_3^2 n_1^2 - m_1^2 n_2^2 - m_3^2 n_2^2 - m_1^2 n_3^2 - m_2^2 n_3^2 + 2m_1 m_2 n_1 n_2 + 2m_1 m_3 n_1 n_3 + 2m_2 m_3 n_2 n_3,$$

$$\begin{aligned} den = 1 + \sum_{j=1}^4 m_j^2 + \sum_{j=1}^4 n_j^2 + m_2^2 n_1^2 + m_3^2 n_1^2 + m_4^2 n_1^2 + m_1^2 n_2^2 + m_3^2 n_2^2 + m_4^2 n_2^2 + m_1^2 n_3^2 + m_2^2 n_3^2 + m_4^2 n_3^2 + m_1^2 n_4^2 + m_2^2 n_4^2 + \\ + m_3^2 n_4^2 - 2m_1 m_2 n_1 n_2 - 2m_1 m_3 n_1 n_3 - 2m_2 m_3 n_2 n_3 - 2m_1 m_4 n_1 n_4 - 2m_2 m_4 n_2 n_4 - 2m_3 m_4 n_3 n_4. \end{aligned}$$

[0137] Los coeficientes m_j y n_j de la matriz de mezcla hacia abajo extendida \mathbf{D} indica los valores de mezcla hacia abajo para cada EAO j para los canales de mezcla hacia abajo derecho e izquierdo como

$$m_j = d_{0,EAO(j)}, \quad n_j = d_{1,EAO(j)}.$$

5

[0138] Los elementos $d_{i,j}$ de la matriz de mezcla hacia abajo \mathbf{D} se obtienen empleando la información de ganancia de mezcla hacia abajo DMG y la información de diferencia de niveles de canales de mezcla hacia abajo DCLD (opcional), que está incluida en la información SAOC 332, que está representado, por ejemplo, por la información paramétrica relativa a un objeto 110 o la información de corriente de bits SAOC 212.

10

[0139] Para el caso de mezcla hacia abajo estéreo la matriz de mezcla hacia abajo \mathbf{D} de tamaño $2 \times 3 \times N$ con elementos $d_{i,j}$ ($i = 0, 1; j = 0, \dots, N-1$) se obtiene a partir de los parámetros DMG y DCLD como

$$d_{0,j} = 10^{0.05 DMG_j} \sqrt{\frac{10^{0.1 DCLD_j}}{1 + 10^{0.1 DCLD_j}}}, \quad d_{1,j} = 10^{0.05 DMG_j} \sqrt{\frac{1}{1 + 10^{0.1 DCLD_j}}}.$$

[0140] Para el caso de mezcla hacia abajo mono la matriz de mezcla hacia abajo \mathbf{D} de tamaño $1 \times 3 \times N$ con elementos $d_{i,j}$ ($i = 0; j = 0, \dots, N-1$) se obtiene a partir de los parámetros DMG como

$$d_{0,j} = 10^{0.05 DMG_j}.$$

[0141] Aquí, se obtienen los parámetros de mezcla hacia abajo descuantizados DMG_j y $DCLD_j$, por ejemplo, a partir de la información lateral paramétrica 110 o a partir de la corriente de bits SAOC 212.

[0142] La función $EAO(j)$ determina un cartografiado entre índices de canales de objeto de audio de entrada y señales EAO:

$$EAO(j) = N - 1 - j, \quad j = 0, \dots, N_{EAO} - 1.$$

5

3.4.1.4 Cálculo de la matriz C

[0143] La matriz **C** implica los CPCs y se deriva de los transmitidas parámetros SAOC transmitidas (es decir los OLDs, IOCs, DMGs y DCLDs) como

$$c_{j,0} = (1 - \lambda) \tilde{c}_{j,0} + \lambda \gamma_{j,0}, \quad c_{j,1} = (1 - \lambda) \tilde{c}_{j,1} + \lambda \gamma_{j,1}.$$

10

[0144] Dicho de otro modo, los CPCs restringidos se obtienen de acuerdo con las ecuaciones de más arriba, que se pueden considerar como un algoritmo de restricción. Sin embargo, los CPCs restringidos también pueden derivarse de los valores $c_{j,0}, c_{j,1}$ empleando un enfoque de limitación diferente (algoritmo de restricción), o se pueden ajustar como iguales a los valores $c_{j,0}, c_{j,1}$.

15

[0145] Debe destacarse, que las entradas de matriz $c_{j,1}$ (y las cantidades intermedias a partir de las cuales se calculan las entradas de matriz $c_{j,1}$) en general solamente se requieren si la señal de mezcla hacia abajo es una señal de mezcla hacia abajo estéreo.

[0146] Los CPCs están restringidos por las siguientes funciones de limitación:

$$\gamma_{j,1} = \frac{m_j OLD_L + n_j e_{L,R} - \sum_{i=0}^{N_{EAO}-1} m_i e_{i,j}}{2 \left(OLD_L + \sum_{i=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_i m_k e_{i,k} \right)}, \quad \gamma_{j,2} = \frac{n_j OLD_R + m_j e_{L,R} - \sum_{i=0}^{N_{EAO}-1} n_i e_{i,j}}{2 \left(OLD_R + \sum_{i=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_i n_k e_{i,k} \right)},$$

20 con el factor de ponderación λ determinado como

$$\lambda = \left(\frac{P_{LoRo}^2}{P_{Lo} P_{Ro}} \right)^8.$$

[0147] Para un canal específico $EAO j = 0 \dots N_{EAO} - 1$ los CPCs sin restringir se estiman como

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j} P_{Ro} - P_{RoCo,j} P_{LoRo}}{P_{Lo} P_{Ro} - P_{LoRo}^2}, \quad \tilde{c}_{j,1} = \frac{P_{RoCo,j} P_{Lo} - P_{LoCo,j} P_{LoRo}}{P_{Lo} P_{Ro} - P_{LoRo}^2}.$$

[0148] Las cantidades de energía P_{Lo} , P_{Ro} , P_{LoRo} , $P_{LoCo,j}$ y $P_{RoCo,j}$ se calculan como

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k},$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k},$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k},$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j},$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}.$$

[0149] La matriz de covarianza $e_{i,j}$ se define de la siguiente manera:

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

5

La matriz de covarianza \mathbf{E} de tamaño $N \times N$ con elementos $e_{i,j}$ representa una aproximación de la matriz de covarianza de señales originales $\mathbf{E} \approx \mathbf{SS}^*$ y se obtiene a partir de los parámetros OLD y IOC como

[0150] Aquí, los parámetros de objeto descuantizados OLD_i , IOC_{ij} se obtienen, por ejemplo, a partir de la información lateral paramétrica 110 o a partir de la corriente de bits SAOC 212.

10

[0151] Además, $e_{L,R}$ puede, por ejemplo, obtenerse como

$$e_{L,R} = \sqrt{OLD_L OLD_R} IOC_{L,R}.$$

[0152] Los parámetros OLD_L , OLD_R y $IOC_{L,R}$ corresponden a los objetos (audio) regulares y se pueden obtener empleando la información de mezcla hacia abajo:

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i ,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & \text{En los casos restantes} \end{cases}$$

5 **[0153]** Como se puede ver, dos valores de diferencia de niveles de objetos comunes OLDL y OLDR se calculan para los objetos de audio regulares en el caso de una señal de mezcla hacia abajo estéreo (lo que implica preferentemente una señal de objeto de audio regular de dos canales). En contraste, solamente se calcula un valor de diferencia de niveles de objetos común OLDL para los objetos de audio regulares en el caso de una señal de mezcla hacia abajo de un canal (mono) (que preferentemente implica una señal de objeto de audio regular de un canal).

10 **[0154]** Como se puede ver, se obtiene el primero (en el caso de una señal de mezcla hacia abajo de dos canales) o único (en el caso de una señal de mezcla hacia abajo de un canal) valor de diferencia de niveles de objetos común OLDL mediante la suma de contribuciones de los objetos de audio regulares que tienen índice de objeto de audio (o índices) i al canal izquierdo (o único canal) de la señal de mezcla hacia abajo SAOC 310.

15 **[0155]** El segundo valor de diferencia de niveles de objetos común OLDR (que se utiliza en el caso de una señal de mezcla hacia abajo de dos canales) se obtiene sumando las contribuciones de los objetos de audio regulares que tienen el índice de objeto de audio (o índices) i al canal derecho de la señal de mezcla hacia abajo SAOC 310.

20 **[0156]** La contribución OLDL de los objetos de audio regulares (que tienen índices de objetos de audio i=0 a i=N-NEAO-1) en la señal de canal izquierdo (o única señal de canal) de la mezcla hacia abajo SAOC, se calcula la señal 710, por ejemplo, teniendo en cuenta la ganancia de mezcla hacia abajo $d_{0,i}$, que describe la ganancia de mezcla hacia abajo aplicada al objeto de audio regular que tiene el índice de objeto de audio i cuando se obtiene la señal de canal izquierdo de la señal de mezcla hacia abajo SAOC 310, y también el nivel de objeto del objeto de audio regular que tiene el objeto de audio i, que está representado por el valor OLDi.

25 **[0157]** De manera similar, el valor de diferencia de niveles de objetos común OLDR se obtiene empleando los coeficientes de mezcla hacia abajo $d_{1,i}$, que describen la ganancia de mezcla hacia abajo que se aplica al objeto de audio regular que tiene el índice de objeto de audio i cuando se forma la señal de canal derecho de la señal de mezcla hacia abajo SAOC 310, y la información de nivel OLDi asociada con el objeto de audio regular que tiene el índice de objeto de audio i.

30 **[0158]** Como se puede ver, las ecuaciones para el cálculo de las cantidades PLo, PRo, PLoRo, PLoCo,j y PRoCo,j no distinguen entre los objetos de audio regulares individuales, sino que solamente utilizan los valores de diferencia de nivel de objeto comunes OLDL, OLDR, considerando de este modo los objetos de audio regulares (que tienen índices de objeto de audio i) como un objeto de audio único.

[0159] También, el valor de correlación entre objetos IOCL,R, que está asociado con los objetos de audio regulares, se ajusta a 0 a menos que haya dos objetos de audio regulares.

[0160] La matriz de covarianza $e_{i,j}$ (y $e_{L,R}$) se define tal se indica a continuación:

35 **[0161]** 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 originales $\mathbf{E} \approx \mathbf{SS}^*$ y se obtiene a partir de los parámetros OLD y IOC como

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

[0162] Por ejemplo,

$$e_{L,R} = \sqrt{OLD_L OLD_R} IOC_{L,R},$$

en el que OLDL y OLDR y IOCL,R se calculan tal como se ha descrito más arriba.

[0163] Aquí, los parámetros de objeto descuantizados se obtienen como

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

5

donde DOLD y DIOC son matrices que comprenden parámetros de diferencia de niveles de objetos y parámetros de correlación entre objetos.

3.4.2. Modo Energía

[0164] A continuación, se describirá otro concepto, que puede ser utilizado para separar las señales de objeto de objeto de audio extendidas 320 y las señales de objeto de objeto de audio regulares (no extendidas) 322, y que se puede usar en combinación con una codificación de audio que no conserva las formas de onda de los canales de mezcla hacia abajo SAOC 310.

[0165] Dicho de otro modo, el proceso de codificación/descodificación basado en energía está diseñado para una codificación que no conserva las formas de onda de la señal de mezcla hacia abajo. Por lo tanto matriz de mezcla hacia arriba OTN/TTN para el modo de energía correspondiente no se basa en formas de onda específicas, sino que solamente describe la distribución de energías relativas de los objetos de audio de entrada.

[0166] También, el concepto aquí descrito, que se designa como un concepto "Modo Energía", se puede utilizar sin transmitir una información de señal residual. De nuevo, los objetos de audio regulares (objetos de audio no mejorados) se tratan como un objeto de audio de un único canal o de dos canales que tiene uno o dos valores de diferencia de niveles de objetos comunes OLDL, OLDR.

[0167] Para el Modo Energía la matriz $\mathbf{M}^{Energía}$ se define como que explota la información de mezcla hacia abajo y las OLDs, tal como se describirá a continuación.

3.4.2.1. Modo Energía para los Modos de mezcla hacia abajo estéreo (TTN)

[0168] En el caso estéreo (por ejemplo, una mezcla hacia abajo estéreo a partir de dos canales de objeto de audio

25 regulares y NEAO canales de objeto de audio mejorados), las matrices $\mathbf{M}_{OBJ}^{Energy}$ y $\mathbf{M}_{EAO}^{Energy}$ se obtienen a partir de las OLDs correspondientes a partir de

$$\mathbf{M}_{OBJ}^{Energy} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

[0169] Las señales de salida del procesador residual se calculan como

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energy} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix},$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Energy} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}.$$

5 [0170] Las señales y_L , y_R , que están representadas por la señal \mathbf{X}_{OBJ} , describen los objetos de audio regulares (y puede ser equivalente a la señal 322), y las señales $Y_{0,EAO}$ a $Y_{N_{EAO}-1,EAO}$, que se describen mediante la señal \mathbf{X}_{EAO} , describe los objetos de audio mejorados (y puede ser equivalente a la señal 334 o a la señal 320).

10 [0171] Si se desea una señal de mezcla hacia arriba para el caso de una señal de mezcla hacia abajo estéreo, se puede realizar un procesamiento 2-a-1, por ejemplo, por el preprocesador 270 a partir del canal de dos señales \mathbf{X}_{OBJ} .

3.4.2.2. Modo Energía para Modos de mezcla hacia abajo mono (OTN)

[0172] Para el caso mono (por ejemplo, una mezcla hacia abajo mono a partir de un canal de objeto de audio regular y N_{EAO} canales de objeto de audio mejorados), las matrices

$$\mathbf{M}_{OBJ}^{Energy}$$

15 y

$$\mathbf{M}_{EAO}^{Energy}$$

se obtienen a partir de los OLDs correspondientes según

$$\mathbf{M}_{OBJ}^{Energy} = \left(\frac{OLD_L}{\sqrt{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right),$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \frac{m_0^2 OLD_0}{\sqrt{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{\sqrt{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}.$$

[0173] Las señales de salida del procesador residual se calculan como

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energy} (d_0),$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Energy} (d_0).$$

5

[0174] Un único canal de objeto de audio regular 322 (representado por \mathbf{X}_{OBJ}) y N_{EAO} canales de objeto de audio mejorados 320 (representados por \mathbf{X}_{EAO}) se pueden obtener por aplicación de las matrices

$$\mathbf{M}_{OBJ}^{Energy}$$

y

$$\mathbf{M}_{EAO}^{Energy}$$

10

a una representación de un único canal de señal de mezcla hacia abajo SAOC 310 (representado aquí por d_0).

[0175] Si se desea un señal de mezcla hacia arriba (estéreo) de dos canales para el caso una señal de mezcla hacia abajo de un canal (mono), puede llevarse a cabo un procesamiento 1-a-2, por ejemplo, por el pre-procesador 270 a partir de la señal de un canal \mathbf{X}_{OBJ} .

15

4. Arquitectura y funcionamiento del preprocesador de mezcla hacia abajo SAOC

[0176] A continuación, se describirá el funcionamiento del preprocesador de mezcla hacia abajo SAOC 270 tanto para algunos modos de decodificación de operación como para algunos modos de operación de transcodificación.

4.1 Operación en los modos de decodificación

4.1.1 Introducción

5 **[0177]** A continuación, se describe un procedimiento para obtener una señal de salida empleando parámetros SAOC e información panorámica (o información de representación) asociada con cada objeto de audio. El descodificador SAOC 495 se representa en la figura 4g y consiste de el procesador de parámetros SAOC 496 y el procesador de mezcla hacia abajo 497.

10 **[0178]** Debe destacarse que el descodificador SAOC 494 puede ser utilizado para procesar los objetos de audio regulares, y puede por lo tanto recibir, como señal de mezcla hacia abajo 497a, la segunda señal de objeto de audio 264 o la señal de objeto de audio regular 322 o la segunda información de audio 134. En consecuencia, el procesador de mezcla hacia abajo 497 puede proporcionar, como sus señales de salida 497b, la versión procesada 272 de la segunda señal de objeto de audio 264 o la versión procesada 142 de la segunda información de audio 134. En consecuencia, el procesador de mezcla hacia abajo 497 puede asumir la función del preprocesador de mezcla hacia abajo SAOC 270, o la función del procesador de señal de audio 140.

15 **[0179]** El procesador de parámetros SAOC 496 puede asumir la función del procesador de parámetros SAOC 252 y por lo tanto proporcionar una información de mezcla hacia abajo 496a.

4.1.2 Procesador de mezcla hacia abajo

[0180] A continuación, se describirá con más detalle el procesador de mezcla hacia abajo, que es parte del procesador de señal de audio 140, y que se designa como un " pre-procesador de mezcla hacia abajo SAOC " 270 in la realización de la figura 2, y que se designa por 497 en el descodificador SAOC 495.

20 **[0181]** Para el modo decodificador del sistema SAOC, la señal de salida 142, 272, 497b del procesador de mezcla hacia abajo (representada en el dominio híbrido QMF) se introduce en el banco de filtros de síntesis correspondiente (que no se muestra en las figuras 1 y 2) tal como se describe en la norma ISO/IEC 23003-1: 2007 dando la señal PCM de salida final. Sin embargo, la señal de salida 142, 272, 497b del procesador de mezcla hacia abajo se combina en general con una o más señales de audio 132, 262 que representan los objetos de audio mejorados. Esta combinación puede llevarse a cabo antes del banco de filtros de síntesis correspondiente (de modo que una señal combinada que combina la salida del procesador de mezcla hacia abajo y la o las más señales que representan los objetos de audio mejorados se introducen en el banco de filtros de síntesis). Como alternativa, la señal de salida del procesador de mezcla hacia abajo puede ser combinada con una o más señales de audio que representan los objetos de audio mejorados solamente después del procesamiento del banco de filtros de síntesis. En consecuencia, la representación de señal de mezcla hacia arriba 120, 220 puede ser tanto una representación en el dominio QMF o una representación en el dominio PCM (o cualquier otro tipo de representación apropiada). El procesamiento de mezcla hacia abajo incorpora, por ejemplo, el procesamiento de mono, el procesamiento estéreo y, si es necesario, el procesamiento binaural posterior.

35 **[0182]** La señal de salida X del procesador de mezcla hacia abajo 270, 497 (también designada con 142, 272, 497b) se calcula a partir de la señal de mezcla hacia abajo mono X (también designada con 134, 264, 497a) y la señal de mezcla hacia abajo mono descorrelacionada X_d como

$$\hat{X} = GX + P_2 X_d.$$

[0183] La señal de mezcla hacia abajo mono descorrelacionada X_d se calcula como

$$X_d = \text{decorrFunc}(X).$$

40 **[0184]** Las señales descorrelacionadas X_d se crean a partir del descorrelacionador descrito en ISO/IEC 23003-1:2007, subapartado 6.6.2. Según este esquema, se debe utilizar la configuración `bsDecorrConfig == 0` con un índice de descorrelador, $X = 8$, de acuerdo con las tablas A.26 a A.29 de ISO/IEC 23003-1:2007. Por lo tanto, `decorrFunc()` denomina un proceso de descorrelación:

$$\mathbf{X}_d = \begin{pmatrix} x_{1d} \\ x_{2d} \end{pmatrix} = \begin{pmatrix} \text{decorrFunc}((1 \ 0)\mathbf{P}_1\mathbf{X}) \\ \text{decorrFunc}((0 \ 1)\mathbf{P}_1\mathbf{X}) \end{pmatrix}.$$

[0185] En caso de salida binaural los parámetros de mezcla hacia arriba \mathbf{G} y \mathbf{P}_2 derivados de los datos SAOC, la información de representación $\mathbf{M}_{\text{ren}}^{l,m}$ y los parámetros HRTF se aplican a la señal de mezcla hacia abajo \mathbf{X} (y \mathbf{X}_d) dando la salida binaural \mathbf{X} , ver la figura 2, referencia 270, donde se muestra la estructura básica del procesador de mezcla hacia abajo.

[0186] La matriz de representación binaural objetivo $\mathbf{A}_{l,m}$ de tamaño $2 \times N$ consiste en los elementos $a_{x,y}^{l,m}$. Cada elemento $a_{x,y}^{l,m}$ se deriva de los parámetros HRTF y matriz de representación $\mathbf{M}_{\text{ren}}^{l,m}$ con elementos $m_{y,j}^{l,m}$ por ejemplo, por el procesador de parámetros SAOC. La matriz de representación binaural objetivo $\mathbf{A}_{l,m}$ representa la relación entre todos los objetos de audio de entrada y de salida binaural y la deseada.

$$a_{y,1}^{l,m} = \sum_{i=0}^{N_{\text{HRTF}}-1} m_{y,i}^{l,m} H_{i,L}^m \exp\left(j \frac{\phi_i^m}{2}\right), \quad a_{y,2}^{l,m} = \sum_{i=0}^{N_{\text{HRTF}}-1} m_{y,i}^{l,m} H_{i,R}^m \exp\left(-j \frac{\phi_i^m}{2}\right).$$

[0187] Los parámetros HRTF se dan por $H_{i,L}^m$, $H_{i,R}^m$ y ϕ_i^m para cada banda de procesamiento m . Las posiciones espaciales para las que hay parámetros HRTF disponibles se caracterizan por el índice i . Estos parámetros se describen en ISO/IEC 23003-1:2007.

4.1.2.1 Visión general

[0188] A continuación, se dará una visión general sobre el procesamiento de mezcla hacia abajo haciendo referencia a las figuras 4a y 4b, que muestran una representación de bloques del procesamiento de mezcla hacia abajo, que puede ser realizado por el procesador de señal de audio 140 o por la combinación del procesador de parámetros SAOC 252 y el preprocesador de mezcla hacia abajo SAOC 270, o por la combinación del procesador de parámetros SAOC 496 y el procesador de mezcla hacia abajo 497.

[0189] Haciendo referencia a continuación a la figura 4a, el procesamiento de mezcla hacia abajo recibe una matriz de representación \mathbf{M} , una información de diferencia de niveles de objetos OLD, una información de correlación entre objetos IOC, una información de ganancia de mezcla hacia abajo DMG y (de manera opcional) una información de diferencia de niveles de canales de mezcla hacia abajo DCLD. El procesamiento de mezcla hacia abajo 400 según la figura 4a obtiene una matriz de representación \mathbf{A} a partir de la matriz de representación \mathbf{M} , por ejemplo, empleando un ajustador de parámetros y un cartografiado \mathbf{M} -a- \mathbf{A} . También se obtienen las entradas de una matriz de covarianza \mathbf{E} se obtienen en función de la información de diferencia de niveles de objetos OLD y la información de correlación entre objetos IOC, por ejemplo, tal como se ha tratado más arriba. De manera similar, las entradas de una matriz de mezcla hacia abajo \mathbf{D} se obtienen en función de la información de ganancia de mezcla hacia abajo DMG y la información de diferencia de niveles de canales de mezcla hacia abajo DCLD.

[0190] Las entradas f de una matriz de covarianza deseada \mathbf{F} se obtienen en función de la matriz de representación \mathbf{A} y la matriz de covarianza \mathbf{E} . También, se obtiene un valor escalar v en función de la matriz de covarianza \mathbf{E} y la matriz de mezcla hacia abajo \mathbf{D} (o en función de las entradas de esta).

[0191] Los valores de ganancia PL , PR para dos canales se obtienen en función de entradas de la matriz de covarianza deseada \mathbf{F} y el valor escalar v . También, se obtiene un valor de diferencia de fase entre canales ϕ_C en función de las entradas f de la matriz de covarianza deseada \mathbf{F} . También se obtiene un ángulo de rotación α en función de entradas f de la matriz de covarianza deseada \mathbf{F} , teniendo en cuenta, por ejemplo, una constante c .

Además, se obtiene un segundo ángulo de rotación β , por ejemplo, en función de las ganancias de canal PL, PR y el primer ángulo de rotación α . Las entradas de una matriz \mathbf{G} se obtienen, por ejemplo, en función de los dos valores de ganancias de canal PL,PR y también en función de la diferencia de fase entre canales ϕ_C y, de manera opcional, los ángulos de rotación α, β . De manera similar, las entradas de una matriz \mathbf{P}_2 se determinan en función de algunos o todos de dichos valores PL, PR, ϕ_C, α, β .

5 **[0192]** A continuación, se describirá como se puede obtener la matriz \mathbf{G} y/o \mathbf{P}_2 (o las entradas de esta), que puede ser aplicada por el procesador de mezcla hacia abajo tal como se ha tratado más arriba, para diferentes modos de procesamiento.

4.1.2.2 Modo de procesamiento mono a Binaural "x-1-b"

10 **[0193]** A continuación, se discutirá un modo de procesamiento en el que los objetos de audio regulares están representados por una única señal de mezcla hacia debajo de canal 134, 264, 322, 497a y en la cual se desea una representación binaural.

[0194] Los parámetros de mezcla hacia arriba $\mathbf{G}_{l,m}$ y $\mathbf{P}_{2,l,m}$ se calculan como

$$\mathbf{G}^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j\frac{\phi_C^{l,m}}{2}\right) \cos(\beta^{l,m} + \alpha^{l,m}) \\ P_R^{l,m} \exp\left(-j\frac{\phi_C^{l,m}}{2}\right) \cos(\beta^{l,m} - \alpha^{l,m}) \end{pmatrix},$$

$$\mathbf{P}_2^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j\frac{\phi_C^{l,m}}{2}\right) \sin(\beta^{l,m} + \alpha^{l,m}) \\ P_R^{l,m} \exp\left(-j\frac{\phi_C^{l,m}}{2}\right) \sin(\beta^{l,m} - \alpha^{l,m}) \end{pmatrix}.$$

15 **[0195]** Las ganancias

$$P_L^{l,m}$$

y

$$P_R^{l,m}$$

Las ganancias para los canales de salida izquierdo y derecho son

$$P_L^{l,m} = \sqrt{\max\left(\frac{f_{1,1}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}, \quad P_R^{l,m} = \sqrt{\max\left(\frac{f_{2,2}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}.$$

20

[0196] La matriz de covarianza deseada $F_{l,m}$ de tamaño 2×2 con elementos $f_{i,j}^{l,m}$ se da como

$$\mathbf{F}^{l,m} = \mathbf{A}^{l,m} \mathbf{E}^{l,m} (\mathbf{A}^{l,m})^*.$$

[0197] El escalar $v_{l,m}$ se calcula como

$$v^{l,m} = \mathbf{D}^l \mathbf{E}^{l,m} (\mathbf{D}^l)^* + \varepsilon^2.$$

[0198] La diferencia de fase $\phi_C^{l,m}$ entre canales se da como

$$\phi_C^{l,m} = \begin{cases} \arg(f_{1,2}^{l,m}), & 0 \leq m \leq 11, \quad \rho_C^{l,m} \geq 0.6, \\ 0, & \text{En los demás casos} \end{cases}$$

5

[0199] La coherencia entre canales $\rho_C^{l,m}$ se calcula como

$$\rho_C^{l,m} = \min \left(\frac{|f_{1,2}^{l,m}|}{\sqrt{\max(f_{1,1}^{l,m}, f_{2,2}^{l,m}, \varepsilon^2)}}, 1 \right).$$

[0200] Los ángulos de rotación $\alpha_{l,m}$ y $\beta_{l,m}$ se dan como

$$\alpha^{l,m} = \begin{cases} \frac{1}{2} \arccos \left(\rho_C^{l,m} \cos \left(\arg \left(f_{1,2}^{l,m} \right) \right) \right), & 0 \leq m \leq 11, \quad \rho_C^{l,m} < 0.6, \\ \frac{1}{2} \arccos \left(\rho_C^{l,m} \right), & \text{En los demás casos} \end{cases}$$

$$\beta^{l,m} = \arctan \left(\tan \left(\alpha^{l,m} \right) \frac{P_R^{l,m} - P_L^{l,m}}{P_L^{l,m} + P_R^{l,m} + \varepsilon} \right).$$

10 4.1.2.3 Modo de procesamiento Mono-a-Stereo "x-1-2"

[0201] A continuación, se describirá un modo de procesamiento en el que los objetos de audio regulares están representados por una señal de canal único 134, 264, 222, y en el que se desea una representación estéreo.

[0202] En caso de salida estéreo el modo de procesamiento "x-1-b" se puede aplicar sin necesidad de utilizar la información HRTF. Esto se puede hacer obteniendo todos los elementos $a_{x,y}^{l,m}$ de la matriz de representación **A**, que da como resultado:

$$a_{1,y}^{l,m} = m_{lf,y}^{l,m}, \quad a_{2,y}^{l,m} = m_{rf,y}^{l,m}.$$

5

4.1.2.4 Modo de Procesamiento Mono-a-Mono "x-1-1"

[0203] A continuación, se describirá un modo de procesamiento en el que los objetos de audio regulares están representados por una señal de canal 134, 264, 322, 497a y en el que se desea una representación de dos canales de los objetos de audio regulares.

10 [0204] En caso de salida mono el modo de procesamiento "x-1-2" se puede aplicar con las siguientes entradas:

$$a_{1,y}^{l,m} = m_{c,y}^{l,m}, \quad a_{2,y}^{l,m} = 0$$

4.1.2.5 Modo de Procesamiento Estéreo-a-binaural "x-2-b"

[0205] A continuación, se describirá un modo de procesamiento en el que los objetos de audio regulares están representados por una señal de dos canales 134, 264, 322, 497a, y en el que se desea una representación binaural de los objetos de audio regulares.

15

[0206] Los parámetros de mezcla hacia arriba $G_{l,m}$ y $\mathbf{P}_2^{l,m}$ se calculan como

$$\mathbf{G}^{l,m} = \begin{pmatrix} P_L^{l,m,1} \exp\left(j \frac{\phi^{l,m,1}}{2}\right) \cos(\beta^{l,m} + \alpha^{l,m}) & P_L^{l,m,2} \exp\left(j \frac{\phi^{l,m,2}}{2}\right) \cos(\beta^{l,m} + \alpha^{l,m}) \\ P_R^{l,m,1} \exp\left(-j \frac{\phi^{l,m,1}}{2}\right) \cos(\beta^{l,m} - \alpha^{l,m}) & P_R^{l,m,2} \exp\left(-j \frac{\phi^{l,m,2}}{2}\right) \cos(\beta^{l,m} - \alpha^{l,m}) \end{pmatrix},$$

$$\mathbf{P}_2^{l,m} = \begin{pmatrix} P_L^{l,m} \exp\left(j \frac{\arg(c_{1,2}^{l,m})}{2}\right) \sin(\beta^{l,m} + \alpha^{l,m}) \\ P_R^{l,m} \exp\left(-j \frac{\arg(c_{1,2}^{l,m})}{2}\right) \sin(\beta^{l,m} - \alpha^{l,m}) \end{pmatrix}.$$

[0207] Las ganancias correspondientes

$$P_L^{l,m,x}, \quad P_R^{l,m,x}$$

y

$$P_L^{l,m}, P_R^{l,m}$$

para los canales de salida izquierdo y derecho son

$$P_L^{l,m,x} = \sqrt{\max\left(\frac{f_{1,1}^{l,m,x}}{v^{l,m,x}}, \varepsilon^2\right)}, \quad P_R^{l,m,x} = \sqrt{\max\left(\frac{f_{2,2}^{l,m,x}}{v^{l,m,x}}, \varepsilon^2\right)},$$

$$P_L^{l,m} = \sqrt{\max\left(\frac{c_{1,1}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}, \quad P_R^{l,m} = \sqrt{\max\left(\frac{c_{2,2}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}.$$

- 5 [0208] La matriz de covarianza deseada $\mathbf{F}_{l,m,x}$ de tamaño 2×2 con elementos $f_{u,v}^{l,m,x}$ se da como

$$\mathbf{F}^{l,m,x} = \mathbf{A}^{l,m} \mathbf{E}^{l,m,x} (\mathbf{A}^{l,m})^*.$$

- [0209] La matriz de covarianza $\mathbf{C}_{l,m}$ de tamaño 2×2 con elementos de la señal binaural "seca" se estima como

$$\mathbf{C}^{l,m} = \tilde{\mathbf{G}}^{l,m} \mathbf{D}^l \mathbf{E}^{l,m} (\mathbf{D}^l)^* (\tilde{\mathbf{G}}^{l,m})^*,$$

- 10 donde

$$\tilde{\mathbf{G}}^{l,m} = \begin{pmatrix} P_L^{l,m,1} \exp\left(j \frac{\phi^{l,m,1}}{2}\right) & P_L^{l,m,2} \exp\left(j \frac{\phi^{l,m,2}}{2}\right) \\ P_R^{l,m,1} \exp\left(-j \frac{\phi^{l,m,1}}{2}\right) & P_R^{l,m,2} \exp\left(-j \frac{\phi^{l,m,2}}{2}\right) \end{pmatrix}.$$

- [0210] Los escalares correspondientes $v^{l,m,x}$ y $v^{l,m}$ se calculan como

$$v^{l,m,x} = \mathbf{D}^{l,x} \mathbf{E}^{l,m} (\mathbf{D}^{l,x})^* + \varepsilon^2, \quad v^{l,m} = (\mathbf{D}^{l,1} + \mathbf{D}^{l,2}) \mathbf{E}^{l,m} (\mathbf{D}^{l,1} + \mathbf{D}^{l,2})^* + \varepsilon^2.$$

[0211] La matriz de mezcla hacia abajo $D_{l,x}$ de tamaño $1 \times N$ con elementos $d_i^{l,x}$ se puede encontrar como

$$d_i^{l,1} = 10^{0.05 DMG_i^l} \sqrt{\frac{10^{0.1 DCLD_i^l}}{1 + 10^{0.1 DCLD_i^l}}}, \quad d_i^{l,2} = 10^{0.05 DMG_i^l} \sqrt{\frac{1}{1 + 10^{0.1 DCLD_i^l}}}.$$

[0212] La matriz de mezcla hacia abajo estéreo D' de tamaño $2 \times N$ con elementos $d_{x,i}^l$ se puede encontrar como

$$d_{x,i}^l = d_i^{l,x}.$$

5 [0213] La matriz $E_{l,m,x}$ con elementos $e_{i,j}^{l,m,x}$ se derivan de la siguiente relación

$$e_{i,j}^{l,m,x} = e_{i,j}^{l,m} \left(\frac{d_i^{l,x}}{d_i^{l,1} + d_i^{l,2}} \right) \left(\frac{d_j^{l,x}}{d_j^{l,1} + d_j^{l,2}} \right).$$

[0214] Las diferencias de fase entre canales $\phi_C^{l,m}$ se dan como

$$\phi^{l,m,x} = \begin{cases} \arg(f_{1,2}^{l,m,x}), & 0 \leq m \leq 11, \\ 0, & \text{otherwise.} \end{cases} \quad \rho_C^{l,m} > 0.6,$$

10 [0215] Los ICCs $\rho_C^{l,m}$ y $\rho_T^{l,m}$ se calculan como

$$\rho_T^{l,m} = \min \left(\frac{|f_{1,2}^{l,m}|}{\sqrt{\max(f_{1,1}^{l,m}, f_{2,2}^{l,m}, \varepsilon^2)}}, 1 \right), \quad \rho_C^{l,m} = \min \left(\frac{|c_{1,2}^{l,m}|}{\sqrt{\max(c_{1,1}^{l,m}, c_{2,2}^{l,m}, \varepsilon^2)}}, 1 \right).$$

[0216] Los ángulos de rotación $\alpha_{l,m}$ y $\beta_{l,m}$ se dan como

$$\alpha^{l,m} = \frac{1}{2} \left(\arccos(\rho_T^{l,m}) - \arccos(\rho_C^{l,m}) \right), \quad \beta^{l,m} = \arctan \left(\tan(\alpha^{l,m}) \frac{P_R^{l,m} - P_L^{l,m}}{P_L^{l,m} + P_R^{l,m}} \right).$$

4.1.2.6 Modo de Procesamiento Estereo-a-estereo "x-2-2"

[0217] A continuación, se describirá un modo de procesamiento en el que los objetos de audio regulares se describen mediante una señal (estéreo) de dos canales 134, 264, 322, 497a y en el que se desea una representación de 2-canales (estéreo).

5 [0218] En caso de salida estéreo, se aplica directamente el reprocesamiento estéreo, que se describirá a continuación en la Sección 4.2.2.3.

4.1.2.7 modo de procesamiento Estéreo-a-mono "x-2-1"

10 [0219] A continuación, se describirá un modo de procesamiento en el que los objetos de audio regulares están representados por una señal (estéreo) de dos canales 134, 264, 322, 497a, y en el que se desea una representación de un canal (mono).

[0220] En caso de salida mono, se aplica el reprocesamiento estéreo con una entrada de matriz de representación única activa, tal como se describe más abajo en la sección 4.2.2.3.

4.1.2.8 Conclusión

15 [0221] Haciendo referencia de nuevo a las figuras 4a y 4b, se ha descrito un procesamiento que se puede aplicar a una señal de un canal o de dos canales 134, 264, 322, 497a que representa los objetos de audio regulares subsecuentes a una separación entre los objetos de audio extendidos y los objetos de audio regulares. Las figuras 4a y 4b ilustran el procesamiento, en el que el procesamiento de las figuras 4a y 4b se diferencia en que se introduce un ajuste de parámetro opcional en las diferentes etapas del procesamiento.

4.2. Funcionamiento en los modos de transcodificación

20 4.2.1 Introducción

[0222] A continuación, se explica un procedimiento para combinar parámetros SAOC y mostrar información (o información de representación) asociada con cada objeto de audio (o, preferentemente, con cada objeto de audio regular) en una Corriente de bits MPEG Surround de acuerdo con la normativa (Corriente de bits MPS).

25 [0223] El transcodificador SAOC 490 se representa en la figura 4f y consta de un procesador de parámetros SAOC 491 y un procesador de mezcla hacia abajo 492 aplicado a una mezcla hacia abajo estéreo.

[0224] El transcodificador SAOC 490 puede, por ejemplo, asumir la funcionalidad del procesador de señal de audio 140. Como alternativa, el transcodificador SAOC 490 puede asumir la funcionalidad del preprocesador de mezcla hacia abajo SAOC 270 cuando se toma en combinación con el procesador de parámetros SAOC 252.

30 [0225] Por ejemplo, el procesador de parámetros SAOC 491 puede recibir una corriente de bits SAOC 491a, que es equivalente a la información paramétrica relativa a un objeto 110 o la corriente de bits SAOC 212. También, el procesador de parámetros SAOC 491 puede recibir una información de matriz de representación 491b, que puede incluirse en la información paramétrica relativa a un objeto 110, o que puede ser equivalente a la información de matriz de representación 214. El procesador de parámetros SAOC 491 puede proporcionar también información de procesamiento de mezcla hacia abajo 491c al procesador de mezcla hacia abajo 492, que puede ser equivalente a la información 240. Además, el procesador de parámetros SAOC 491 puede proporcionar una corriente de bits MPEG Surround (o corriente de bits de parámetros MPEG Surround) 491d, que comprende una información Surround paramétrica que es compatible con la norma MPEG Surround. La corriente de bits MPEG Surround 491 d puede, por ejemplo, formar parte de la versión procesada 142 de la segunda información de audio, o puede, por ejemplo formar parte de u ocupar el lugar de la corriente de bits MPS 222.

40 [0226] El procesador de mezcla hacia abajo 492 está configurado para recibir una señal de mezcla hacia abajo 492a, que es preferentemente una señal de mezcla hacia abajo de un canal o una señal de mezcla hacia abajo de dos canales, y que es preferentemente equivalente a la segunda información de audio 134, o a la segunda señal de objeto de audio 264, 322. El procesador de mezcla hacia abajo 492 puede proporcionar también una señal de mezcla hacia abajo MPEG Surround 492b, que es equivalente a (o parte de) la versión procesada 142 de la segunda información de audio 134, o equivalente a (o parte de) la versión procesada 272 de la segunda señal de objeto de audio 264.

[0227] Sin embargo, hay diferentes maneras de combinar la señal de mezcla hacia abajo MPEG Surround 492b con la señal de objeto de audio mejorado 132, 262. La combinación puede ser realizada en el dominio MPEG Surround.

50 [0228] Como alternativa, sin embargo, la representación MPEG Surround, que comprende la corriente de bits de parámetros MPEG Surround 491 d y la señal de mezcla hacia abajo MPEG Surround 492b, de los objetos de audio regulares puede ser convertida de vuelta a una representación de dominio de tiempo de varios canales o a una representación de dominio de frecuencia multicanal (que representa individualmente diferentes canales de audio)

por un decodificador MPEG Surround y puede ser subsecuentemente combinada con las señales de objeto de audio mejoradas.

5 **[0229]** Cabe señalar que los modos de transcodificación comprenden tanto uno o más modos de procesamiento de mezcla hacia abajo mono como uno o más modos de procesamiento de mezcla hacia abajo estéreo. Sin embargo, a continuación solamente se describirá el modo de procesamiento de mezcla hacia abajo estéreo, porque el procesamiento de las señales de objeto de audio regulares es más elaborada en el modo de Procesamiento de mezcla hacia abajo estéreo.

4.2.2 Modo de Procesamiento de procesamiento de mezcla hacia abajo en la mezcla hacia abajo estéreo ("x-2-5")

4.2.2.1 Introducción

10 **[0230]** En la siguiente sección, se dará una descripción del modo de transcodificación SAOC para el caso de mezcla estéreo.

15 **[0231]** Los parámetros de objeto (diferencia de niveles de objeto OLD, correlación entre objetos IOC, ganancia de mezcla hacia abajo DMG y diferencias de niveles de canales de mezcla hacia abajo DCMD) de la corriente de bits SAOC se transcodifican en parámetros espaciales (preferentemente relativos a canal) (diferencia de niveles de canales CLD, correlación entre canales ICC, coeficiente de predicción de canal CPC) para la corriente de bits MPEG Surround según la información de representación. La mezcla hacia abajo se modifica según parámetros de objeto y una matriz de representación.

[0232] Haciendo referencia a continuación a las figuras 4c, 4d y 4e, se dará una visión general del procesamiento, y en particular de la modificación de mezcla hacia abajo.

20 **[0233]** La figura 4c muestra una representación de bloques de un procesamiento que se realiza modificando la señal de mezcla hacia abajo, por ejemplo la señal de mezcla hacia abajo 134, 264, 322,492a que describe el único, preferentemente, o más objetos de audio regulares. Como se puede ver a partir de las figuras 4c, 4d y 4e, el procesamiento recibe una matriz de representación M_{ren} , una información de ganancia de mezcla hacia abajo DMG, una información de diferencia de niveles de canales de mezcla hacia abajo DCLD, una información de diferencia de niveles de objetos OLD, y una información de correlación entre objetos IOC. La matriz de representación puede de manera opcional modificarse mediante un ajuste de parámetros, tal como se muestra en la figura 4c. Las entradas de una matriz de mezcla hacia abajo D se obtienen en función de la información de ganancia de mezcla hacia abajo DMG y la información de diferencia de niveles de canales de mezcla hacia abajo DCLD. Las entradas de una matriz de coherencia E se obtienen en función de la información de diferencia de niveles de objetos OLD y la información de correlación entre objetos IOC. Además, se puede obtener una matriz J obtenida en función de la matriz de mezcla hacia abajo D y la matriz de coherencia E , o en función de las entradas de esta. Subsecuentemente, puede obtenerse una matriz C_3 en función de la matriz de representación M_{ren} , la matriz de mezcla hacia abajo D , la matriz de coherencia E y la matriz J . Puede obtenerse una matriz G en función de una matriz D_{TTT} , que puede ser una matriz que tiene entradas predeterminadas, y también en función de la matriz C_3 . La matriz G puede, de manera opcional, ser modificada, para obtener una matriz modificada G_{mod} . La matriz G o la versión modificada G_{mod} de esta puede ser utilizada para obtener la versión procesada 142, 272,492b de la segunda información de audio 134, 264 a partir de la segunda información de audio 134, 264,492a (en la que la segunda información de audio 134, 264 se designa por X , y en la que la versión procesada 142, 272 de esta se designa por X).

40 **[0234]** A continuación, se discutirá la energía de representación de objeto, que se realiza con la finalidad de obtener los parámetros MPEG Surround,. También se describirá el reprocesamiento estéreo, que se realiza con la finalidad de obtener la versión procesada 142, 272,492b de la segunda información de audio 134, 264,492a que representa los objetos de audio regulares.

4.2.2.2 Energías de representación de objetos

45 **[0235]** El transcodificador determina los parámetros para el decodificador MPS de acuerdo con la representación de destino tal como se describe por la matriz de representación M_{ren} . La covarianza objetivo de seis canales se denota con F y viene dada por

$$F = YY^* = M_{ren} S(M_{ren} S)^* = M_{ren} (SS^*) M_{ren}^* = M_{ren} E M_{ren}^* .$$

50 **[0236]** El proceso de transcodificación puede dividirse conceptualmente en dos partes. En una parte se realiza una representación de tres canales a unos canales izquierdo, derecho y central. En esta etapa se obtienen los parámetros para la modificación de mezcla hacia abajo así como los parámetros de predicción para la caja TTT para el de codificador MPS. En la otra parte se determinan los parámetros CLD y ICC para la representación entre los

canales frontal y Surround (parámetros OTT, Surround izquierdo frontal izquierdo, Surround derecho frontal derecho).

4.2.2.2.1 Representación en los canales izquierdo, derecho y central

5 [0237] En esta etapa se determinan los parámetros espaciales de control de la representación a canales izquierdo y derecho, que consiste en señales frontal y Surround. Estos parámetros describen la caja de matriz de predicción TTT para la descodificación MPS C_{TTT} (parámetros CPC para el descodificador MPS) y la matriz de conversión de mezcla hacia abajo G .

[0238] C_{TTT} es la matriz de predicción para obtener la representación objetivo a partir de la mezcla hacia abajo modificada $X=GX$:

$$C_{TTT} \hat{X} = C_{TTT} GX \approx A_3 S.$$

10

[0239] A_3 es una matriz de representación reducida de tamaño $3 \times N$, que describe la representación a los canales izquierdo, derecho y central respectivamente. Se obtiene como $A_3 = D_{36} M_{ren}$ con la matriz parcial 6 a 3 de mezcla hacia abajo D_{36} definida por

$$D_{36} = \begin{pmatrix} w_1 & 0 & 0 & 0 & w_1 & 0 \\ 0 & w_2 & 0 & 0 & 0 & w_2 \\ 0 & 0 & w_3 & w_3 & 0 & 0 \end{pmatrix}.$$

15 [0240] Las ponderaciones de mezcla hacia abajo parciales w_p , $p = 1,2,3$ se ajustan de modo que la energía de $w_p(y_{2p-1} + y_{2p})$ es igual a la suma de las energías $y_{2p-1}^2 + y_{2p}^2$ hasta un factor límite.

$$w_1 = \frac{f_{1,1} + f_{5,5}}{f_{1,1} + f_{5,5} + 2f_{1,5}}, \quad w_2 = \frac{f_{2,2} + f_{6,6}}{f_{2,2} + f_{6,6} + 2f_{2,6}}, \quad w_3 = 0.5,$$

Donde $f_{i,j}$ indican los elementos de F .

20 [0241] Para la estimación de la matriz de predicción deseada C_{TTT} y la matriz de preprocesamiento de mezcla hacia abajo G se define una predicción matriz C_3 de tamaño 3×3 , que conduce a la representación de destino

$$C_3 X \approx A_3 S.$$

[0242] Esta matriz se deriva teniendo en cuenta las ecuaciones normales

$$C_3 (DED^*) \approx A_3 ED^*.$$

25 [0243] La solución de las ecuaciones normales produce la mejor coincidencia posible de forma de onda de la salida de destino dado el modelo de covarianza de objeto. G y C_{TTT} se obtiene ahora resolviendo el sistema de ecuaciones

$$C_{TTT} G = C_3.$$

[0244] Para evitar problemas numéricos en el cálculo del término $J = (\mathbf{DED}^*)^{-1}$, se modifica J . Primero se calculan los valores propios $\lambda_{1,2}$ de J , resolviendo $\det(\mathbf{J} - \lambda_{1,2}\mathbf{I}) = 0$.

[0245] Los valores propios son ordenados en orden descendente ($\lambda_1 \geq \lambda_2$) y el vector propio correspondiente al valor propio más grande se calcula de acuerdo con la ecuación anterior. Se asegura que se está en el plano x positivo (el primer elemento tiene que ser positivo). El segundo vector propio se obtiene del primero mediante una rotación de -90 grados:

$$\mathbf{J} = (\mathbf{v}_1 \mathbf{v}_2) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} (\mathbf{v}_1 \mathbf{v}_2)^T.$$

[0246] Una matriz de ponderación se calcula a partir de la matriz de mezcla hacia abajo \mathbf{D} y la matriz de predicción \mathbf{C}_3 , $\mathbf{W} = (\mathbf{D} \text{diag}(\mathbf{C}_3))$.

[0247] Puesto que \mathbf{CTTT} es una función de los parámetros de predicción MPS c_1 y c_2 (tal como se define en ISO/IEC 23003-1:2007), $\mathbf{CTTTG} = \mathbf{C}_3$ se reescribe de la siguiente manera, para encontrar el o los puntos estacionarios de la función,

$$\mathbf{\Gamma} \begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \end{pmatrix} = \mathbf{b},$$

con $\mathbf{\Gamma} = (\mathbf{DTTT} \mathbf{C}_3) \mathbf{W} (\mathbf{DTTT} \mathbf{C}_3)^*$ y $\mathbf{b} = \mathbf{GWC}_3 \mathbf{v}$, donde $\mathbf{v} = (1 \ 1 \ -1)$.

[0248] Si φ no proporciona una solución única ($\det(\mathbf{\Gamma}) < 10^{-3}$), se elige el punto que se encuentra más cercano al punto lo que da una TTT que pasa a través de. Como primera etapa, se escoge la fila i de $\mathbf{\Gamma}$ $\gamma =$

$[\gamma_{i,1} \ \gamma_{i,2}]$ donde los elementos contienen más energía, y por lo tanto $\gamma_{i,1}^2 + \gamma_{i,2}^2 \geq \gamma_{j,1}^2 + \gamma_{j,2}^2, j=1,2$.

[0249] A continuación, una solución se determina de tal manera que

$$\begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 3\mathbf{y} \quad \text{with} \quad \mathbf{y} = \frac{b_{i,3}}{\left(\sum_{j=1,2} (\gamma_{i,j})^2 \right) + \epsilon} \mathbf{y}^T.$$

[0250] Si la solución obtenida para c_1 y c_2 se encuentra fuera del rango permitido para coeficientes de predicción que se definen como $-2 \leq c_j \leq 3$ (tal como se definen en la norma ISO/IEC 23003-1:2007), c_j se calcularán según se indica más abajo.

[0251] En primer lugar se define el conjunto de puntos, \mathbf{x}_p como:

$$\mathbf{x}_p \in \left[\begin{array}{l} \left(\min \left(3, \max \left(-2, -\frac{-2\gamma_{1,2} - b_1}{\gamma_{1,1} + \varepsilon} \right) \right) \right), \left(\min \left(3, \max \left(-2, -\frac{3\gamma_{1,2} - b_1}{\gamma_{1,1} + \varepsilon} \right) \right) \right) \\ -2 \qquad \qquad \qquad 3 \\ -2 \qquad \qquad \qquad 3 \\ \left(\min \left(3, \max \left(-2, -\frac{-2\gamma_{2,1} - b_2}{\gamma_{2,2} + \varepsilon} \right) \right) \right), \left(\min \left(3, \max \left(-2, -\frac{3\gamma_{2,1} - b_2}{\gamma_{2,2} + \varepsilon} \right) \right) \right) \end{array} \right],$$

y la función de distancia,

$$distFunc(\mathbf{x}_p) = \mathbf{x}_p^T \Gamma \mathbf{x}_p - 2\mathbf{b} \mathbf{x}_p.$$

5 **[0252]** A continuación, los parámetros de predicción se definen según:

$$\begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \end{pmatrix} = \arg \min_{\mathbf{x} \in \mathbf{x}_p} (distFunc(\mathbf{x})).$$

[0253] Los parámetros de predicción están limitados según:

$$c_1 = (1 - \lambda) \tilde{c}_1 + \lambda \gamma_1, \quad c_2 = (1 - \lambda) \tilde{c}_2 + \lambda \gamma_2,$$

10 donde λ , γ_1 y γ_2 se definen como

$$\gamma_1 = \frac{2f_{1,1} + 2f_{5,5} - f_{3,3} + f_{1,3} + f_{5,3}}{2f_{1,1} + 2f_{5,5} + 2f_{3,3} + 4f_{1,3} + 4f_{5,3}},$$

$$\gamma_2 = \frac{2f_{2,2} + 2f_{6,6} - f_{3,3} + f_{2,3} + f_{6,3}}{2f_{2,2} + 2f_{6,6} + 2f_{3,3} + 4f_{2,3} + 4f_{6,3}},$$

$$\lambda = \left(\frac{(f_{1,2} + f_{1,6} + f_{5,2} + f_{5,6} + f_{1,3} + f_{5,3} + f_{2,3} + f_{6,3} + f_{3,3})^2}{(f_{1,1} + f_{5,5} + f_{3,3} + 2f_{1,3} + 2f_{5,3})(f_{2,2} + f_{6,6} + f_{3,3} + 2f_{2,3} + 2f_{6,3})} \right)^8.$$

[0254] Para el decodificador MPS, se proporcionan los CPCs y los ICC_{TTT} correspondientes tal como se expone a continuación

$$\mathbf{D}_{\text{CPC}_1} = c_1(l, m), \mathbf{D}_{\text{CPC}_2} = c_2(l, m) \text{ and } \mathbf{D}_{\text{ICC}_{\text{TTT}}} = 1.$$

4.2.2.2 representación entre los canales frontales y Surround

[0255] Los parámetros que determinan la representación entre los canales frontales y Surround se puede estimar directamente a partir de la matriz de covarianza de destino \mathbf{F}

$$CLD_{a,b} = 10 \log_{10} \left(\frac{\max(f_{a,a}, \varepsilon^2)}{\max(f_{b,b}, \varepsilon^2)} \right), \quad ICC_{a,b} = \frac{\max(f_{a,b}, \varepsilon^2)}{\sqrt{\max(f_{a,a}, \varepsilon^2) \max(f_{b,b}, \varepsilon^2)}},$$

5

con $(a,b) = (1,2)$ y $(3,4)$.

[0256] Los parámetros MPS se proporcionan en la forma

$$CLD_h^{l,m} = \mathbf{D}_{\text{CLD}}(h, l, m) \text{ and } ICC_h^{l,m} = \mathbf{D}_{\text{ICC}}(h, l, m),$$

para cada caja h OTT.

10 4.2.2.3 Procesado estéreo

[0257] A continuación, se describirá un procesado estéreo de la señal de objeto de audio regular 134 a 64, 322. El procesado estéreo se utiliza para derivar un proceso de representación general 142, 272 a partir de una representación de los objetos de audio regulares de dos canales.

15 [0258] La mezcla hacia abajo estéreo \mathbf{X} , que está representada por las señales de objeto de audio regulares 134, 264, 492a se procesa como la señal de mezcla hacia abajo modificada $\hat{\mathbf{X}}$, que está representada por las señales de objeto de audio regulares procesadas 142, 272:

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{X},$$

donde

$$\mathbf{G} = \mathbf{D}_{\text{TTT}}\mathbf{C}_3 = \mathbf{D}_{\text{TTT}}\mathbf{M}_{\text{ren}}\mathbf{E}\mathbf{D}'\mathbf{J}.$$

20 [0259] La salida estéreo final a partir del transcodificador SAOC \mathbf{X} se produce mediante la mezcla de \mathbf{X} con un componente de señal decorrelacionada de acuerdo con:

$$\hat{\mathbf{X}} = \mathbf{G}_{\text{Mod}}\mathbf{X} + \mathbf{P}_2\mathbf{X}_d,$$

donde la señal decorrelacionada \mathbf{X}_d se calcula tal como se ha descrito más arriba, y las matrices de mezcla \mathbf{G}_{Mod} y \mathbf{P}_2 de la siguiente manera.

[0260] En primer lugar, se define la matriz de error de mezcla hacia arriba de representación como

$$\mathbf{R} = \mathbf{A}_{\text{diff}} \mathbf{E} \mathbf{A}_{\text{diff}}'$$

donde

$$\mathbf{A}_{\text{diff}} = \mathbf{D}_{\text{TTT}} \mathbf{A}_3 - \mathbf{G} \mathbf{D},$$

5 y además se define la matriz de covarianza de la señal predicha \mathbf{R} como

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{r}_{1,1} & \hat{r}_{1,2} \\ \hat{r}_{2,1} & \hat{r}_{2,2} \end{pmatrix} = \mathbf{G} \mathbf{D} \mathbf{E} \mathbf{D}' \mathbf{G}' .$$

[0261] El vector de ganancia \mathbf{g}_{vec} puede ser calculado subsecuentemente como:

$$\mathbf{g}_{\text{vec}} = \left(\min \left(\sqrt{\max \left(\frac{\hat{r}_{1,1} + r_{1,1} + \varepsilon^2}{r_{1,1} + \varepsilon^2}, 0 \right)}, 1.5 \right), \min \left(\sqrt{\max \left(\frac{\hat{r}_{2,2} + r_{2,2} + \varepsilon^2}{r_{2,2} + \varepsilon^2}, 0 \right)}, 1.5 \right) \right),$$

y la matriz de mezcla \mathbf{G}_{Mod} se da como:

$$\mathbf{G}_{\text{Mod}} = \begin{cases} \text{diag}(\mathbf{g}_{\text{vec}}) \mathbf{G}, & r_{1,2} > 0, \\ \mathbf{G}, & \text{En los casos restantes} \end{cases}$$

10

[0262] De manera similar, la matriz de mezcla \mathbf{P}_2 se da como:

$$\mathbf{P}_2 = \begin{cases} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & r_{1,2} > 0, \\ \mathbf{v}_R \text{diag}(\mathbf{W}_d) & \text{En los casos restantes} \end{cases}$$

[0263] Para derivar \mathbf{v}_R y \mathbf{W}_d , hay que resolver la ecuación característica \mathbf{R} :

$$\det(\mathbf{R} - \lambda_{1,2} \mathbf{I}) = 0, \text{ Dados los valores propios } \lambda_1 \text{ y } \lambda_2$$

15 [0264] Los vectores propios correspondientes \mathbf{v}_{R1} y \mathbf{v}_{R2} de \mathbf{R} pueden calcularse resolviendo el sistema de ecuaciones:

$$(\mathbf{R} - \lambda_{1,2} \mathbf{I}) \mathbf{v}_{R1,R2} = 0.$$

5 [0265] Los valores propios se ordenan en orden descendente ($\lambda_1 \geq \lambda_2$) y el vector propio correspondiente al valor propio más grande se calcula según la ecuación anterior. Se garantiza que está en el plano x positivo (el primer elemento tiene que ser positivo). El segundo vector propio se obtiene a partir del primero mediante una rotación de -90 grados:

$$\mathbf{R} = (\mathbf{v}_{R1} \ \mathbf{v}_{R2}) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} (\mathbf{v}_{R1} \ \mathbf{v}_{R2})^*.$$

[0266] Incorporando $\mathbf{P}_1 = (1 \ 1)\mathbf{G}$, \mathbf{R}_d se puede calcular según:

$$\mathbf{R}_d = \begin{pmatrix} \mathbf{r}_{d11} & \mathbf{r}_{d12} \\ \mathbf{r}_{d21} & \mathbf{r}_{d22} \end{pmatrix} = \text{diag}(\mathbf{P}_1 (\mathbf{D}\mathbf{E}\mathbf{D}^*) \mathbf{P}_1^*),$$

lo que da

$$\mathbf{w}_{d1} = \min\left(\sqrt{\frac{\lambda_1}{r_{d1} + \varepsilon}}, 2\right), \quad \mathbf{w}_{d2} = \min\left(\sqrt{\frac{\lambda_2}{r_{d2} + \varepsilon}}, 2\right),$$

10

y, finalmente, la matriz de mezcla,

$$\mathbf{P}_2 = (\mathbf{v}_{R1} \ \mathbf{v}_{R2}) \begin{pmatrix} \mathbf{w}_{d1} & 0 \\ 0 & \mathbf{w}_{d2} \end{pmatrix}.$$

4.2.2.4 Modo Dual

15 [0267] El transcodificador SAOC puede permitir calcular las matrices de mezcla \mathbf{P}_1 , \mathbf{P}_2 y la matriz de predicción \mathbf{C}_3 según un esquema alternativo para el rango de frecuencia superior. Este esquema alternativo es particularmente útil para las señales de mezcla hacia abajo en las que el rango de frecuencia superior está codificado por un algoritmo de codificación que no preserva la forma de onda por ejemplo SBR en AAC de alta eficiencia.

20 [0268] Para las bandas de parámetros superiores, definidas por $\text{bsTttBandsLow} \leq pb < \text{numBands}$, \mathbf{P}_1 , \mathbf{P}_2 y \mathbf{C}_3 se calcularán de acuerdo con el esquema alternativo que se describe a continuación:

$$\begin{cases} \mathbf{P}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \\ \mathbf{P}_2 = \mathbf{G}. \end{cases}$$

[0269] Definir los vectores de mezcla de energía hacia abajo y energía objetivo, respectivamente:

$$\begin{cases} \mathbf{e}_{\text{dmx}} = \begin{pmatrix} e_{\text{dmx1}} \\ e_{\text{dmx2}} \end{pmatrix} = \text{diag}(\mathbf{DED}^*) + \varepsilon \mathbf{I}, \\ \mathbf{e}_{\text{tar}} = \begin{pmatrix} e_{\text{tar1}} \\ e_{\text{tar2}} \\ e_{\text{tar3}} \end{pmatrix} = \text{diag}(\mathbf{A}_3 \mathbf{E} \mathbf{A}_3^*), \end{cases}$$

y la matriz de ayuda

$$\mathbf{T} = \begin{pmatrix} t_{1,1} & t_{1,2} \\ t_{2,1} & t_{2,2} \\ t_{3,1} & t_{3,2} \end{pmatrix} = \mathbf{A}_3 \mathbf{D}^* + \varepsilon \mathbf{I}.$$

[0270] A continuación, se calcula el vector de ganancia

$$\mathbf{g} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} \frac{e_{\text{tar1}}}{\sqrt{t_{1,1}^2 e_{\text{dmx1}} + t_{1,2}^2 e_{\text{dmx2}}}} \\ \frac{e_{\text{tar2}}}{\sqrt{t_{2,1}^2 e_{\text{dmx1}} + t_{2,2}^2 e_{\text{dmx2}}}} \\ \frac{e_{\text{tar3}}}{\sqrt{t_{3,1}^2 e_{\text{dmx1}} + t_{3,2}^2 e_{\text{dmx2}}}} \end{pmatrix},$$

5

que finalmente da la nueva matriz de predicción

$$\mathbf{C}_3 = \begin{pmatrix} g_1 t_{1,1} & g_1 t_{1,2} \\ g_2 t_{2,1} & g_2 t_{2,2} \\ g_3 t_{3,1} & g_3 t_{3,2} \end{pmatrix}.$$

- 10 5. Modo de decodificación / transcodificación combinado EKS SAOC, codificador de acuerdo con la figura 10 y sistemas según las figuras 5a, 5b

[0271] A continuación se dará una breve descripción del esquema de procesamiento EKS SAOC combinado. Se propone un esquema preferido de procesamiento "EKS SAOC combinado", donde el procesamiento EKS se integra en la cadena de decodificación / transcodificación SAOC regular mediante un esquema en cascada.

- 15 5.1. Codificador de señal de audio según la figura 5

[0272] En una primera etapa, los objetos dedicados al procesamiento EKS (procesamiento Karaoke/solo) se identifican como objetos de primer plano (FGO) y su número NFGO (también designado como NEAO) se determina

por mediante una variable de corriente de bits "bsNumGroupsFGO". Esta variable de corriente de bits puede, por ejemplo, incluirse en una corriente de bits SAOC, tal como se ha descrito más arriba.

5 **[0273]** Para la generación de la corriente de bits (en un codificador de señal de audio), los parámetros de todos los objetos de entrada N_{obj} se reordenan de modo que los objetos de primer plano FGO comprenden los últimos NFGO (o como alternativa, NEAO) parámetros en cada case, por ejemplo, OLD_i para $[N_{obj} - NFGO \leq i \leq N_{obj} - 1]$.

10 **[0274]** De los restantes objetos que son, por ejemplo, objetos de segundo plano FGO u objetos de audio no mejorados, se genera una señal de mezcla hacia abajo al " estilo SAOC regular " que al mismo tiempo sirve como objeto de Fondo BGO. A continuación, el objeto de fondo y los objetos de primer plano se mezclan hacia abajo con el "estilo de Procesamiento EKS " y la información residual se extrae de cada objeto de segundo plano. De esta manera, no se necesitan etapas de procesamiento extras. Por lo tanto, no se requiere cambio alguno en la sintaxis de la corriente de bits.

15 **[0275]** Dicho de otro modo, en el lado del codificador, los objetos de audio no mejorados se distinguen de los objetos de audio mejorados. Se proporciona una señal de mezcla hacia abajo de objetos de audio regulares de uno o dos canales que representan los objetos de audio regulares (objetos de audio no mejorados), en la que puede haber uno, dos o incluso más objetos de audio regulares (objetos de audio no mejorados). La señal de mezcla hacia abajo de objetos de audio regular de uno o dos canales se combina entonces con uno o más objetos de señales de audio mejorados (que pueden, por ejemplo, ser señales de un canal o señales de dos canales), para obtener una señal común de mezcla hacia abajo (que puede, por ejemplo, ser una señal de mezcla hacia abajo de un canal o una señal de mezcla hacia abajo de dos canales) combinando las señales de audio de los objetos de audio mejorados y la señal de objeto de audio de mezcla hacia abajo regular.

20 **[0276]** A continuación, la estructura básica de este codificador en cascada se describirá brevemente haciendo referencia a la figura 10, que muestra una representación esquemática en bloque de un codificador SAOC 1000, según una realización de la invención. El codificador SAOC 1000 comprende un primer mezclador hacia abajo SAOC 1010, que es típicamente un mezclador hacia abajo SAOC que no proporciona una información residual. El mezclador hacia abajo SAOC 1010 está configurado para recibir una pluralidad de N_{BGO} señales de objeto de audio 1012 a partir de objetos de audio (no mejorados). También, el mezclador hacia abajo SAOC 1010 está configurado para proporcionar una señal de mezcla hacia abajo de objeto de audio regular 1014 a partir de los objetos de audio regulares 1012, de modo que la señal de mezcla hacia abajo de objeto de audio regular 1014 combina las señales de objeto de audio regulares 1012 de acuerdo con parámetros de mezcla hacia abajo. El mezclador hacia abajo SAOC 1010 también proporciona una información de objeto de audio regular SAOC 1016, que describe las señales de objeto de audio regulares y la mezcla hacia abajo. Por ejemplo, la información de objeto de audio regular SAOC 1016 puede comprender una información de ganancia de mezcla hacia abajo DMG y una información de diferencia de niveles de canales de mezcla hacia abajo DCLD que describe la mezcla hacia abajo realizada por el mezclador hacia abajo SAOC 1010. Además, la información de objeto de audio regular SAOC 1016 puede comprender una información de diferencia de niveles de objetos y una información de correlación entre objetos que describe una relación entre los objetos de audio regulares descrita por la señal de objeto de audio regular 1012.

35 **[0277]** El codificador 1000 también comprende un segundo mezclador hacia abajo SAOC 1020, que está típicamente configurado para proporcionar una información residual. El segundo mezclador hacia abajo SAOC 1020 está preferentemente configurado para recibir una o más señales de objeto de audio mejoradas 1022 y también para recibir la señal de mezcla hacia abajo de objeto de audio regular 1014.

40 **[0278]** El segundo mezclador hacia abajo SAOC 1020 también está configurado para proporcionar una señal de mezcla hacia abajo SAOC común 1024 a partir de las señales de objeto de audio mejoradas 1022 y la señal de mezcla hacia abajo de objeto de audio regular 1014. Al proporcionar la señal de mezcla hacia abajo SAOC común, el segundo mezclador hacia abajo SAOC 1020 típicamente procesa la señal de mezcla hacia abajo de objeto de audio regular 1014 como una señal de objeto de un canal o de dos canales.

45 **[0279]** El segundo mezclador hacia abajo SAOC 1020 también está configurado para proporcionar una información de objeto de audio SAOC mejorado que describe, por ejemplo, valores de diferencia de niveles de canales de mezcla hacia abajo DCLD asociados con los objetos de audio mejorados, valores de diferencia de nivel de objeto OLD asociados con los objetos de audio mejorados y valores de correlación entre objetos IOC asociados con los objetos de audio mejorados. Además, el segundo SAOC 1020 está preferentemente configurado para proporcionar información residual asociada con cada uno de los objetos de audio mejorados, de modo que la información residual asociada con los objetos de audio mejorados describe la diferencia entre un señal de objeto de audio mejorado individual original y un señal de objeto de audio mejorado individual esperado que se puede extraer de la señal de mezcla hacia abajo empleando la información de mezcla hacia abajo DMG, DCLD y la información de objeto OLD, IOC.

[0280] El codificador de audio 1000 es apropiado para cooperar con el descodificador de audio aquí descrito.

5.2. Descodificador de señal de audio según la figura 5a

[0281] A continuación, se describirá la estructura básica de un descodificador EKS SAOC combinado 500, cuyo diagrama de bloques esquemático se muestra en la figura 5a.

[0282] El descodificador de audio 500 según la figura 5a está configurado para recibir una señal de mezcla hacia abajo 510, una información de corriente de bits SAOC 512 y una información de matriz de representación 514. El descodificador de audio 500 comprende un procesamiento de Karaoke/Solo mejorado y una representación de objeto de primer plano 520, que está configurado para proporcionar una primera señal de objeto de audio 562, que describe objetos de primer plano representados, y una segunda señal de objeto de audio 564, que describe los objetos de fondo. Los objetos de primer plano pueden, por ejemplo, llamarse "objetos de audio mejorados" y los objetos de fondo pueden, por ejemplo, llamarse "objetos de audio regulares" u "objetos de audio no mejorados". El descodificador de audio 500 también comprende una descodificación SAOC regular 570, que está configurada para recibir la segunda señal de objeto de audio 562 y para proporcionar, a partir de esta, una versión procesada 572 de la segunda señal de objeto de audio 564. El descodificador de audio 500 también comprende un combinador 580, que está configurado para combinar la primera señal de objeto de audio 562 y la versión procesada 572 de la segunda señal de objeto de audio 564, para obtener una señal de salida 520.

[0283] A continuación, se discutirá con más detalle la funcionalidad del descodificador de audio 500. En el lado de descodificación/ transcodificación, el proceso de mezcla hacia arriba da como resultado un esquema en cascada que comprende primero un procesamiento Karaoke- Solo mejorado (Procesamiento EKS) para descomponer la señal de mezcla hacia abajo en el objeto de fondo (BGO) y los objetos de primer plano (FGOs). Las diferencias de nivel de objetos requeridas (OLDs) y correlaciones entre-objetos (IOCs) para el objeto de fondo se derivan de la información de objeto y de mezcla hacia abajo (que son ambas una información paramétrica relativa a un objeto, y que están ambas típicamente incluidas en la corriente de bits SAOC):

$$OLD_L = \sum_{i=0}^{N-N_{FGO}-1} d_{0,i}^2 OLD_i$$

$$OLD_R = \sum_{i=0}^{N-N_{FGO}-1} d_{1,i}^2 OLD_i ,$$

$$IOC_{LR} = \begin{cases} IOC_{0,1}, & N - N_{FGO} = 2, \\ 0, & \text{En los casos restantes} \end{cases}$$

[0284] Además, esta etapa (que normalmente se ejecuta por el procesamiento EKS y la representación de objetos de primer plano 520) incluye la cartografía de los objetos de primer plano con los canales de salida finales (de tal manera que, por ejemplo, la primera señal de objeto de audio 562 es una señal multicanal en la que a los objetos de primer plano se les asignan uno o más canales). El objeto de fondo (que comprende típicamente una pluralidad de los llamados "objetos de audio regulares") se representa en los canales de salida correspondientes por un proceso de decodificación SAOC regular (o, de forma alternativa, en algunos casos por un proceso de transcodificación SAOC). Este proceso puede, por ejemplo, ser realizado por la descodificación SAOC regular 570. La etapa final de mezclado (por ejemplo, el combinador 580) proporciona una combinación deseada de objetos de primer plano representados y señales de objeto de fondo en la salida.

[0285] Este sistema EKS SAOC combinado representa una combinación de todas las propiedades beneficiosas del sistema SAOC regular y su modo EKS. Este enfoque permite conseguir las prestaciones correspondientes empleando el sistema propuesto con la misma corriente de datos tanto para el escenario de reproducción clásico (representación moderada) como el similar a Karaoke/Solo (representación extrema).

5.3. Estructura generalizada según la figura 5b

[0286] A continuación, se describirá una estructura generalizada de un sistema EKS SAOC combinado 590 haciendo referencia a la figura 5b, que muestra un diagrama de bloques esquemático de este sistema EKS SAOC

combinado generalizado. El sistema EKS SAOC combinado 590 de la figura 5b puede también ser considerado como un descodificador de audio.

5 **[0287]** El sistema EKS SAOC combinado 590 está configurado para recibir una señal de mezcla hacia abajo 510a, una información de corriente de bits SAOC 512a y la información de matriz de representación 514a. También, el sistema EKS SAOC combinado 590 está configurado para proporcionar una señal de salida 520a a partir de esta.

10 **[0288]** El sistema EKS SAOC combinado 590 comprende una etapa I de procesamiento de tipo SAOC 520a, que recibe la señal de mezcla hacia abajo 510a, la información de corriente de bits SAOC 512a (o al menos una parte de esta) y la información de matriz de representación 514a (o al menos una parte de esta). En particular, la etapa I de procesamiento de tipo SAOC 520a recibe valores de diferencia de nivel de objeto de primera etapa (OLDs). La etapa I de procesamiento de tipo SAOC 520a proporciona una o más señales 562a que describen un primer conjunto de objetos (por ejemplo, objetos de audio de un primer tipo de objeto de audio). La etapa I de procesamiento de tipo SAOC 520a también proporciona una o más señales 564a que describen un segundo conjunto de objetos.

15 **[0289]** El sistema EKS SAOC combinado también comprende una etapa II de procesamiento de tipo SAOC 570a, que está configurada para recibir la o las más señales 564a que describen el segundo conjunto de objetos y para proporcionar, a partir de esta, una o más señales 572a que describen un tercer conjunto de objetos empleando diferencias de niveles de objeto de segunda etapa, que están incluidos en la información de corriente de bits SAOC 512a, y también al menos una parte de la información de matriz de representación 514. El sistema EKS SAOC combinado también comprende un combinador 580a, que puede, por ejemplo, ser un sumador, para proporcionar las señales de salida 520a mediante la combinación de la o las más señales 562a que describen el primer conjunto de objetos y la o las más señales 570a que describen el tercer conjunto de objetos (en el que el tercer conjunto de objetos puede ser una versión procesada del segundo conjunto de objetos).

20 **[0290]** Para resumir lo anterior, la figura 5b muestra una forma generalizada de la estructura básica descrita con referencia a la figura 5a de más arriba en otra realización de la invención.

6. Evaluación Perceptual del esquema de procesamiento EKS SAOC combinado

25 6.1 Prueba de Metodología, Diseño y Artículos

[0291] Estas pruebas de escucha subjetivas se llevaron a cabo en una sala de escucha acústicamente aislada que está diseñada para permitir la escucha de alta calidad. La reproducción se realiza con auriculares (STAX SR Lambda Pro con convertidor Lake-People D/A y Monitor STAX SRM). El procedimiento de ensayo siguió los procedimientos estándar utilizados en las pruebas de verificación de audio espacial, a partir del procedimiento "estímulo múltiple con referencia y anclajes ocultos" (MUSHRA) para la evaluación subjetiva de la calidad de audio intermedia (véase la referencia [7]).

35 **[0292]** Un total de ocho oyentes participaron en la prueba realizada. Todos los sujetos pueden ser considerados oyentes experimentados. De acuerdo con la metodología MUSHRA, se instruyó a los oyentes para comparar todas las condiciones de prueba frente a la de referencia. Las condiciones de ensayo fueron aleatorizadas de forma automática para cada elemento de prueba y para cada oyente. Las respuestas subjetivas fueron registradas por un programa MUSHRA basado en ordenador en una escala que va de 0 a 100. Se permitió una conmutación instantánea entre los elementos bajo prueba. La prueba de MUSHRA se ha llevado a cabo con el fin de evaluar el rendimiento perceptual de los modos SAOC considerados y el sistema propuesto se describe en la tabla de la figura 6a, que proporciona una descripción del diseño de pruebas de escucha.

40 **[0293]** Las señales de mezcla hacia abajo correspondientes se codificaron usando un codificador de núcleo AAC con una velocidad de bits de 128 kbps. Con el fin de evaluar la calidad percibida del sistema EKS SAOC combinado propuesto, se compara con el sistema SAOC RM regular (sistema de modelo de referencia SAOC) y modo EKS actual (modo Karaoke-Solo mejorado) para dos escenarios de prueba de representación diferentes descritos en la tabla de la figura 6b, que describe los sistemas bajo prueba.

45 **[0294]** Se aplicó la codificación residual con una velocidad de bits de 20 kbps para el modo EKS actual y un sistema EKS SAOC combinado propuesto. Cabe señalar que para el modo EKS actual es necesario generar un objeto de fondo estéreo (BGO) antes del procedimiento de codificación / decodificación real, ya que este modo tiene limitaciones sobre el número y tipo de objetos de entrada.

50 **[0295]** El material de prueba de escucha y los parámetros de mezcla hacia abajo y de representación correspondientes usados en las pruebas realizadas se seleccionaron de un concurso de propuestas de artículos de audio (CfP Concurso de propuestas) de elementos de audio descritos en el documento [2]. Los datos correspondientes a los escenarios de aplicación de representación "Karaoke" y "Classic" se pueden encontrar en la tabla de la figura 6c, que describe elementos de prueba de audición y matrices de representación.

6.2 Resultados de pruebas de escucha

[0296] Una visión general corta en términos de los diagramas que demuestran los resultados de las pruebas de escucha obtenidos se puede encontrar en las figuras 6d y 6e, donde la figura 6d muestra valores MUSHRA promediados para la prueba de audición de representación de tipo Karaoke/Solo, y la figura 6e muestra valores MUSHRA promediados para la prueba de audición de representación clásica. Los gráficos muestran la valoración MUSHRA promediada por artículo entre todos los oyentes y el valor medio estadístico sobre todos los elementos evaluados junto con los intervalos de confianza del 95% asociados.

[0297] Las siguientes conclusiones se pueden extraer de los resultados de las pruebas de sonido realizadas:

- La figura 6d representa la comparación para el modo EKS actual con el sistema EKS SAOC combinado para aplicaciones de tipo Karaoke. Para todos los artículos ensayados no se puede observar ninguna diferencia significativa (en el sentido estadístico) en el rendimiento entre estos dos sistemas. De esta observación se puede concluir que el sistema de EKS SAOC combinado es capaz de explotar de manera eficiente la información residual alcanzando el rendimiento del modo EKS. También se puede observar que el rendimiento del sistema SAOC regular (sin residual) está por debajo de los otros dos sistemas.

- La figura 6e representa la comparación del sistema actual SAOC regular con el sistema EKS SAOC combinado para escenarios de representación clásicos. Para todos los elementos probados el rendimiento de estos dos sistemas es estadísticamente el mismo. Esto demuestra el funcionamiento correcto del sistema EKS SAOC combinado para un escenario de representación clásico.

[0298] Por lo tanto, se puede concluir que el sistema unificado propuesto que combina el modo de EKS con el SAOC regular conserva las ventajas en la calidad de audio subjetiva para los tipos correspondientes de representación.

[0299] Teniendo en cuenta el hecho de que el sistema propuesto EKS SAOC combinado ya no tiene restricciones sobre el objeto BGO, pero tiene capacidad de representación totalmente flexible del modo SAOC regular y puede utilizar la misma corriente de datos para todos los tipos de representación, parece ser ventajoso incorporarlo en el estándar MPEG SAOC.

7. Procedimiento Según la figura 7

[0300] A continuación, se describirá un procedimiento para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto con referencia a la figura 7, que muestra un diagrama de flujo de un procedimiento como este.

[0301] El procedimiento 700 comprende una etapa 710 de descomponer una representación de señal de mezcla hacia abajo, para proporcionar una primera información de audio que describe un primer conjunto de uno o más objetos de audio de un primer tipo de objeto de audio y una segunda información de audio que describe un segundo conjunto de uno o más objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y al menos una parte de la información paramétrica relativa a un objeto. El procedimiento 700 también comprende una etapa 720 de procesar la segunda información de audio en función de la información paramétrica relativa a un objeto, para obtener una versión procesada de la segunda información de audio.

[0302] El procedimiento 700 también comprende una etapa 730 de combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba.

[0303] El procedimiento 700 según la figura 7 puede suplementarse con cualquiera de las características y funcionalidades que se discuten aquí con respecto al aparato de la invención. También, el procedimiento 700 aporta las ventajas discutidas con respecto al aparato de la invención.

8. Alternativas de Implementación

[0304] Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque correspondiente o un elemento o característica de un aparato correspondiente. Algunas o todas las etapas del procedimiento pueden ser ejecutadas por (o mediante) 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 dicho aparato.

[0305] La señal de audio codificada de la invención puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable, tal como la Internet.

5 **[0306]** Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación puede realizarse utilizando un medio de almacenamiento digital, por ejemplo, un disquete, un DVD, un Blue-Ray, un CD, una memoria ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control legibles electrónicamente almacenadas, que coopera (o es capaz de cooperar) con un sistema informático programable de tal manera que se lleva a cabo el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

10 **[0307]** Algunas realizaciones de acuerdo con la invención comprenden un portador de datos que tiene señales de control legibles electrónicamente, que es capaz de cooperar con un sistema informático programable, de tal manera que se lleva a cabo uno de los procedimientos descritos en este documento.

[0308] En general, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, siendo el código de programa operativo para llevar a cabo uno de los procedimientos, cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede, por ejemplo, ser almacenado en un soporte legible por máquina.

15 **[0309]** Otras formas de realización comprenden el programa de ordenador para llevar a cabo uno de los procedimientos descritos en este documento, almacenado en un soporte legible por máquina.

[0310] 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 realizar uno de los procedimientos descritos en este documento, cuando el programa informático se ejecuta en un ordenador.

20 **[0311]** Una realización adicional de los procedimientos de la invención 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 de ordenador para llevar a cabo uno de los procedimientos descritos en este documento. El portador de datos, el medio de almacenamiento digital o el soporte de registro son típicamente tangibles y / o no transmisores.

25 **[0312]** Una realización adicional del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa de ordenador para llevar a cabo uno de los procedimientos descritos en este documento. El flujo de datos o la secuencia de señales pueden por ejemplo estar configurados para ser transferidos a través de una conexión de comunicación de datos, por ejemplo a través de Internet.

30 **[0313]** Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los procedimientos descritos en este documento.

[0314] Una realización adicional comprende un ordenador que tiene instalado el programa de ordenador para realizar uno de los procedimientos descritos en este documento.

35 **[0315]** En algunas formas de realización, un dispositivo lógico programable (por ejemplo, una matriz de puertas programable) puede ser utilizado para llevar a cabo algunas o todas las funcionalidades de los procedimientos descritos en este documento. En algunas formas de realización, una FPGA puede cooperar con un microprocesador con el fin de realizar uno de los procedimientos descritos en este documento. En general, los procedimientos se llevan a cabo preferentemente mediante cualquier aparato de hardware.

[0316] Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención.

40 Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en este documento serán evidentes para otros expertos en la materia. Es la intención, por lo tanto, estar limitado sólo por el alcance de las reivindicaciones de patente adjuntas y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en este documento.

9. Conclusiones

45 **[0317]** A continuación, algunos aspectos y ventajas del sistema combinado EKS SAOC de acuerdo con la presente invención se resumirán brevemente. Para los escenarios de Karaoke y reproducción en solitario, el modo de procesamiento SAOC EKS soporta tanto la reproducción de los objetos de primer y segundo plano exclusivamente y una mezcla arbitraria (definida por la matriz de representación) de estos grupos de objetos.

50 **[0318]** Además, el primer modo se considera que es el objetivo principal del procesamiento EKS, este último proporciona una flexibilidad adicional.

[0319] Se ha encontrado que una generalización de la funcionalidad EKS implica por consiguiente el esfuerzo de combinar EKS con el modo de procesamiento SAOC regular para obtener un sistema unificado. Las posibilidades de un sistema unificado como este son:

- Una estructura única clara de decodificación / transcodificación SAOC;
 - Un flujo de bits tanto para EKS como para el modo SAOC regular;
 - No hay limitación en el número de objetos de entrada que comprende el objeto de segundo plano (BGO), de tal manera que no hay necesidad de generar el objeto de segundo plano antes de la etapa de codificación SAOC; y
- 5 - Apoyo a una codificación residual de los objetos en primer plano que produce una mayor calidad percibida en situaciones de reproducción de Karaoke / Solo exigentes.

[0320] Estas ventajas se pueden obtener por el sistema unificado descrito en este documento.

Referencias

[0321]

- 10 [1] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N8853, " Call for Proposals on Spatial Audio Object Coding", 79th MPEG Meeting, Marrakech, January 2007.
- [2] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N9099, " Final Spatial Audio Object Coding Evaluation Procedures and Criterion ", 80th MPEG Meeting, San Jose, April 2007.
- 15 [3] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N9250, " Report on Spatial Audio Object Coding RM0 Selection ", 81st MPEG Meeting, Lausanne, July 2007.
- [4] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M15123, "Information and Verification Results for CE on Karaoke/ Solo system improving the performance de MPEG SAOC RM0", 83rd MPEG Meeting, Antalya, Turkey, January 2008.
- 20 [5] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N10659, " Study on ISO/IEC 23003-2:200x Spatial Audio Object Coding (SAOC)", 88th MPEG Meeting, Maui, USA, April 2009.
- [6] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M10660, "Status y Workplan on SAOC Core Experiments", 88th MPEG Meeting, Maui, USA, April 2009.
- [7] EBU Technical recommendation: " MUSHRA-EBU Method for Subjective Listening Tests of Intermediate Audio Quality ", Doc. B/AIM022, October 1999.
- 25 [8] ISO/IEC 23003-1:2007, Information technology - MPEG audio technologies - Part 1: MPEG Surround.

REIVINDICACIONES

- 5 **1.** Descodificador de señal de audio (100; 200; 500; 590) para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo (112; 210; 510; 510) y una información paramétrica relativa a un objeto (110; 212; 512; 512a), comprendiendo el descodificador de señal de audio:
- un separador de objetos (130; 260; 520; 520a) configurado para descomponer la representación de señal de mezcla hacia abajo, para proporcionar una primera información de audio (132; 262; 562; 562a) que describe un primer conjunto de uno o más objetos de audio de un primer tipo de objeto de audio, y una segunda información de audio (134; 264; 564; 564a) que describe un segundo conjunto de uno o más objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y empleando al menos una parte de la información paramétrica relativa a un objeto, en el que la segunda información de audio es una información de audio que describe los objetos de audio del segundo tipo de objeto de audio de una manera combinada;
- 10 un procesador de señal de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para procesar la segunda información de audio en función de la información paramétrica relativa a un objeto, para obtener una versión procesada (142; 272; 572; 572a) de la segunda información de audio; y
- 15 un combinador de señal de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;
- 20 en el que el descodificador de señal de audio está configurado para proporcionar la representación de señal de mezcla hacia arriba en función de una información residual asociada con un subconjunto de objetos de audio representados por la representación de señal de mezcla hacia abajo, en el que el separador de objetos está configurado para descomponer la representación de señal de mezcla hacia abajo para proporcionar la primera información de audio que describe un primer conjunto de uno o más objetos de audio de un primer tipo de objeto de audio al cual la información residual está asociada, y describiendo la segunda información de audio un segundo conjunto de uno o más objetos de audio de un segundo tipo de objeto de audio, que no tiene información residual asociada, en función de la representación de señal de mezcla hacia abajo y empleando la información residual; y
- 25 en el que el procesador de señal de audio está configurado para procesar la segunda información de audio, para realizar un procesamiento individual por objeto de los objetos de audio del segundo tipo de objeto de audio, empleando información paramétrica relativa a un objeto asociado con más de dos objetos de audio del segundo tipo de objeto de audio; y
- 30 en el que la información residual describe una distorsión residual, que se prevé que permanezca si un objeto de audio del primer tipo de objeto de audio se aísla simplemente empleando la información paramétrica relativa a un objeto.
- 35 **2.** El descodificador de señal de audio (100; 200; 500; 590) según la reivindicación 1, en el que el separador de objetos está configurado para proporcionar la primera información de audio empleando la información residual de modo que uno o más objetos de audio del primer tipo de objeto de audio se enfatizan por encima de objetos de audio del segundo tipo de objeto de audio en la primera información de audio, y en el que el separador de objetos está configurado para proporcionar la segunda información de audio empleando la información residual de modo que objetos de audio del segundo tipo de objeto de audio se enfatizan por encima de objetos de audio del primer tipo de objeto de audio en el segundo información de audio.
- 40 **3.** El descodificador de señal de audio (100; 200; 500; 570) según una de las reivindicaciones 1 a 2, en el que el procesador de señal de audio está configurado para procesar la segunda información de audio (134; 264; 564; 564a) en función de la información paramétrica relativa a un objeto (110; 212; 512; 512a) asociada con los objetos de audio del segundo tipo de objeto de audio e independiente de la información paramétrica relativa a un objeto (110; 212; 512; 512a) asociada con los objetos de audio del primer tipo de objeto de audio.
- 45 **4.** El descodificador de señal de audio (100; 200; 500; 590) según una de las reivindicaciones 1 a 3, en el que el separador de objetos está configurado para obtener la primera información de audio (132; 262; 562; 562a, X_{EAO}) y la segunda información de audio (134; 264; 564; 564a, X_{OBJ}) empleando una combinación lineal de uno o más canales de señal de mezcla hacia abajo de la representación de señal de mezcla hacia abajo y uno o más canales residuales, en el que el separador de objetos está configurado para obtener parámetros de combinación para realizar la combinación lineal en función de parámetros de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio ($m_{0,0}$... m_{NEAO-1} ; $n_{0,0}$... n_{NEAO-1}) y en función de coeficientes de predicción de canal ($c_{j,0}$, $c_{j,1}$) de los objetos de audio del primer tipo de objeto de audio.
- 50 **5.** El descodificador de señal de audio (100; 200; 500; 590) según una de las reivindicaciones 1 a 4, en el que el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio según
- 55

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Prediction} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Prediction} \begin{pmatrix} l_0 \\ r_0 \\ \hline res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

donde

$$\mathbf{M}_{Prediction} = \tilde{\mathbf{D}}^{-1} \mathbf{C} ,$$

donde

$$\mathbf{M}^{Prediction} = \begin{pmatrix} \mathbf{M}_{OBJ}^{Prediction} \\ \hline \mathbf{M}_{EAO}^{Prediction} \end{pmatrix}$$

- 5 en el que \mathbf{X}_{OBJ} representan canales de la segunda información de audio;
- en el que \mathbf{X}_{EAO} representan señales de objeto de la primera información de audio;
- en el que \mathbf{D}^{-1} representa una matriz que es una inversa de una matriz de mezcla hacia abajo extendida;
- en el que \mathbf{C} describe una matriz que representa una pluralidad de coeficientes de predicción de canal, $c_{j,0}$, $c_{j,1}$;
- 10 en el que l_0 y r_0 representan canales de la representación de señal de mezcla hacia abajo;
- en el que res_0 a $res_{N_{EAO}-1}$ representan canales residuales; y
- en el que \mathbf{A}^{EAO} es una matriz de pre-representación EAO, cuyas entradas describen un cartografiado de objetos de audio mejorados en los canales de un señal de objeto de audio mejorado \mathbf{X}_{EAO} ;
- 15 en el que el separador de objetos está configurado para obtener la matriz de mezcla hacia abajo inversa \mathbf{D}^{-1} como inversa de una matriz de mezcla hacia abajo extendida \mathbf{D} que se define como

$$\tilde{\mathbf{D}} = \left(\begin{array}{cc|ccc} 1 & 0 & m_0 & \dots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \dots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \dots & -1 \end{array} \right)$$

en el que el separador de objetos está configurado para obtener la matriz C como

$$\mathbf{C} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \dots & 1 \end{array} \right)$$

5 en el que m_0 a $m_{N_{EAO}-1}$ son valores de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio;

en el que n_0 a $n_{N_{EAO}-1}$ son valores de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio;

en el que el separador de objetos está configurado para calcular los coeficientes de predicción $c_{j,0}$ y $c_{j,1}$ como

$$\tilde{c}_{j,0} = \frac{P_{LoCo,j} P_{Ro} - P_{RoCo,j} P_{LoRo}}{P_{Lo} P_{Ro} - P_{LoRo}^2}$$

$$\tilde{c}_{j,1} = \frac{P_{RoCo,j} P_{Lo} - P_{LoCo,j} P_{LoRo}}{P_{Lo} P_{Ro} - P_{LoRo}^2};$$

10

y en el que el separador de objetos está configurado para derivar coeficientes de predicción restringidos $c_{j,0}$ y $c_{j,1}$ de los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ empleando un algoritmo de restricción, o para utilizar los coeficientes de predicción $\tilde{c}_{j,0}$ y $\tilde{c}_{j,1}$ como los coeficientes de predicción $c_{j,0}$ y $c_{j,1}$;

en el que las cantidades de energía P_{Lo} , P_{Ro} , P_{LoRo} , P_{LoCoj} y P_{RoCoj} se definen como

$$P_{Lo} = OLD_L + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j m_k e_{j,k}$$

$$P_{Ro} = OLD_R + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_j n_k e_{j,k}$$

$$P_{LoRo} = e_{L,R} + \sum_{j=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_j n_k e_{j,k}$$

$$P_{LoCo,j} = m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} m_i e_{i,j}$$

$$P_{RoCo,j} = n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EAO}-1} n_i e_{i,j}$$

en el que los parámetros OLD_L , OLD_R y $IOC_{L,R}$ corresponden a objetos de audio del segundo tipo de objeto de audio y se definen según

$$OLD_L = \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^2 OLD_i,$$

$$OLD_R = \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^2 OLD_i,$$

5

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & \text{En los demás casos} \end{cases}$$

en el que $d_{0,i}$ y $d_{1,i}$ son valores de mezcla hacia abajo asociados con los objetos de audio del segundo tipo de objeto de audio;

10 en el que OLD_i son valores de diferencia de nivel de objeto asociados con los objetos de audio del segundo tipo de objeto de audio;

en el que N es un número total de objetos de audio;

en el que N_{EAO} es un número de objetos de audio del primer tipo de objeto de audio;

en el que $IOC_{0,1}$ es un valor de correlación entre objetos asociado con un par de objetos de audio del segundo tipo de objeto de audio;

en el que e_{ij} y $e_{L,R}$ son valores de covarianza derivados de parámetros de diferencia de nivel de objeto y parámetros de correlación entre objetos; y

en el que e_{ij} están asociados con un par de objetos de audio del primer tipo de objeto de audio y $e_{L,R}$ está asociado con un par de objetos de audio del segundo tipo de objeto de audio.

- 5 **6.** El decodificador de señal de audio (100; 200; 500; 590) según una de las reivindicaciones 1 a 4, en el que el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio según

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Prediction} \begin{pmatrix} \frac{d_0}{\text{---}} \\ res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Prediction} \begin{pmatrix} \frac{d_0}{\text{---}} \\ res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

donde

$$\mathbf{M}_{Prediction} = \tilde{\mathbf{D}}^{-1} \mathbf{C}$$

- 10 en el que \mathbf{X}_{OBJ} representa un canal de la segunda información de audio;
 en el que \mathbf{X}_{EAO} representa señales de objeto de la primera información de audio;
 en el que \mathbf{D}^{-1} representa una matriz que es una inversa de una matriz de mezcla hacia abajo extendida;
 en el que \mathbf{C} describe una matriz que representa una pluralidad de coeficientes de predicción de canal, $c_{j,0}$, $c_{j,1}$;
 15 en el que d_0 representa un canal de la representación de señal de mezcla hacia abajo; y
 en el que res_0 a $res_{N_{EAO}-1}$ representan canales residuales; y
 en el que \mathbf{A}^{EAO} es una matriz de pre-representación EAO.
 20 **7.** El decodificador de señal de audio según la reivindicación 6, en el que el separador de objetos está configurado para obtener la matriz de mezcla hacia abajo inversa \mathbf{D}^{-1} es una inversa de una matriz de mezcla hacia abajo extendida \mathbf{D} que se define como

$$\tilde{\mathbf{D}} = \left(\begin{array}{c|ccc} 1 & m_0 & \dots & m_{N_{EAO}-1} \\ \hline m_0 & -1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & 0 & \dots & -1 \end{array} \right)$$

en el que el separador de objetos está configurado para obtener la matriz \mathbf{C} como

$$\mathbf{C} = \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline c_0 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1} & 0 & \dots & 1 \end{array} \right);$$

en la que m_0 a $m_{N_{EAO}-1}$ son valores de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio.

- 5 8. El decodificador de señal de audio (100; 200; 500; 590) según una de las reivindicaciones 1 a 4, en el que el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio según

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energy} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Energy} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}$$

en el que \mathbf{X}_{OBJ} representa canales de la segunda información de audio;

en el que \mathbf{X}_{EAO} representa señales de objeto de la primera información de audio;

- 10 donde

$$\mathbf{M}_{OBJ}^{Energy} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & 0 \\ 0 & \sqrt{\frac{OLD_R}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_0^2 OLD_0}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \\ \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} & \sqrt{\frac{n_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 OLD_i}} \end{pmatrix}$$

en el que m_0 a $m_{N_{EAO}-1}$ son valores de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio;

5 en el que n_0 a $n_{N_{EAO}-1}$ son valores de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio;

en el que OLD_i son valores de diferencia de nivel de objeto asociados con los objetos de audio del primer tipo de objeto de audio;

en el que OLD_L y OLD_R son valores de diferencia de nivel de objeto comunes asociados con los objetos de audio del segundo tipo de objeto de audio; y

10 en el que \mathbf{A}_{EAO} es una matriz de pre-representación EAO.

9. El decodificador de señal de audio según una de las reivindicaciones 1 a 3, en el que el separador de objetos está configurado para obtener la primera información de audio y la segunda información de audio según

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energy} d_0$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{Energy} d_0$$

en el que \mathbf{X}_{OBJ} representa un canal de la segunda información de audio;

15 en el que \mathbf{X}_{EAO} representan señales de objeto de la primera información de audio;

donde

$$\mathbf{M}_{OBJ}^{Energy} = \begin{pmatrix} \sqrt{\frac{OLD_L}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

$$\mathbf{M}_{EAO}^{Energy} = \begin{pmatrix} \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \\ \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 OLD_{N_{EAO}-1}}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \end{pmatrix}$$

en el que m_0 a $m_{N_{EAO}-1}$ son valores de mezcla hacia abajo asociados con los objetos de audio del primer tipo de objeto de audio;

5 en el que OLD_i son valores de diferencia de nivel de objeto asociados con los objetos de audio del primer tipo de objeto de audio;

en el que OLD_L es un valor de diferencia de niveles de objeto común asociado con los objetos de audio del segundo tipo de objeto de audio; y

en el que \mathbf{A}_{EAO} es una matriz de pre-representación EAO;

en el que las matrices

10 $\mathbf{M}_{OBJ}^{Energy}$ y $\mathbf{M}_{EAO}^{Energy}$

se aplican a una representación do de una única señal de mezcla hacia abajo SAOC.

15 **10.** El descodificador de señal de audio (100; 200; 500; 590) según una de las reivindicaciones 1 a 9, en el que el separador de objetos está configurado para aplicar una matriz de representación a la primera información de audio (132; 262; 562; 562a) para cartografiar señales de objeto de la primera información de audio sobre canales de audio de la representación de señal de audio de mezcla hacia arriba (120; 220, 222; 562; 562a).

20 **11.** El descodificador de señal de audio (100; 200; 500; 590) según una de las reivindicaciones 1 a 10, en el que el procesador de señal de audio (140; 270; 570; 570a) está configurado para realizar un reprocesamiento estéreo de la segunda información de audio (134; 264; 564; 564a) en función de una información de representación (M_{ren}), una información de covarianza relativa a objeto (E), una información de mezcla hacia abajo (D), para obtener canales de audio de la versión procesada de la segunda información de audio;

25 **12.** El descodificador de señal de audio (100; 200; 500; 590) según la reivindicación 11, en el que el procesador de señal de audio (140; 270; 570; 570a) está configurado para realizar el procesado estéreo para cartografiar una contribución de objeto de audio estimada ($ED \cdot JX$) de la segunda información de audio (134; 264; 564; 564a) sobre una pluralidad de canales de la representación de señal de audio de mezcla hacia arriba en función de una información de representación y una información de covarianza.

30 **13.** El descodificador de señal de audio según la reivindicación 11 o la reivindicación 12, en el que el procesador de señal de audio está configurado para añadir una contribución de señal de audio descorrelacionada ($P2Xd$), obtenida a partir de uno o más canales de audio de la segunda información de audio, a la segunda información de audio, o una información derivados de la segunda información de audio, en función de una información de error de mezcla hacia arriba (R) y uno o más valores de escalado de intensidad de señal descorrelacionados ($wd1$, $wd2$).

35 **14.** El descodificador de señal de audio según una de las reivindicaciones 1 a 10, en el que el procesador de señal de audio (140; 270; 570; 570a) está configurado para realizar un postprocesado de la segunda información de audio (134; 264; 564; 564a) en función de una información de representación (A), una información de covarianza relativa a objeto (E) y una información de mezcla hacia abajo (D).

15. El descodificador de señal de audio según la reivindicación 14, en el que el procesador de señal de audio está configurado para realizar un procesado mono a binaural de la segunda información de audio, para cartografiar un único canal de la segunda información de audio sobre dos canales de la representación de señal de mezcla hacia arriba, teniendo en cuenta una función de transferencia relativa a la cabeza.

16. El descodificador de señal de audio según la reivindicación 14, en el que el procesador de señal de audio está configurado para realizar un procesado mono a estéreo de la segunda información de audio, para cartografiar un único canal de la segunda información de audio sobre dos canales de la representación de señal de mezcla hacia arriba.
- 5 17. El descodificador de señal de audio según la reivindicación 14, en el que el procesador de señal de audio está configurado para realizar un procesado estéreo a binaural de la segunda información de audio, para cartografiar dos canales de la segunda información de audio sobre dos canales de la representación de señal de mezcla hacia arriba, teniendo en cuenta una función de transferencia relativa a la cabeza.
- 10 18. El descodificador de señal de audio según la reivindicación 14, en el que el procesador de señal de audio está configurado para realizar un procesado estéreo a estéreo de la segunda información de audio, para cartografiar dos canales de la segunda información de audio sobre dos canales de la representación de señal de mezcla hacia arriba.
- 15 19. El descodificador de señal de audio según una de las reivindicaciones 1 a 18, en el que el separador de objetos está configurado para procesar objetos de audio del segundo tipo de objeto de audio, que no tiene información residual asociada, como un objeto de audio único, y en el que el procesador de señal de audio (140; 270; 570; 570a) está configurado para considerar parámetros de representación específicos del objeto asociados con los objetos de audio del segundo tipo de objeto de audio para ajustar contribuciones de los objetos de audio del segundo tipo de objeto de audio a la representación de señal de mezcla hacia arriba.
- 20 20. El descodificador de señal de audio según una de las reivindicaciones 1 a 19, en el que el separador de objetos está configurado para obtener uno o dos valores de diferencia de nivel de objeto comunes (OLDL, OLDR) para una pluralidad de objetos de audio del segundo tipo de objeto de audio; y
- en el que el separador de objetos está configurado para utilizar el valor de diferencia de niveles de objetos común para un cálculo de coeficientes de predicción de canal (CPC); y
- 25 en el que el separador de objetos está configurado para utilizar los coeficientes de predicción de canal para obtener uno o dos canales de audio que representan la segunda información de audio.
- 30 21. El descodificador de señal de audio según una de las reivindicaciones 1 a 20, en el que el separador de objetos está configurado para obtener uno o dos valores de diferencia de nivel de objeto comunes (OLDL, OLDR) para una pluralidad de objetos de audio del segundo tipo de objeto de audio, y en el que el separador de objetos está configurado para utilizar el valor de diferencia de niveles de objetos común para un cálculo de entradas de una matriz (**M**); y
- en el que el separador de objetos está configurado para utilizar la matriz (**M**) para obtener uno o más canales de audio que representan la segunda información de audio.
- 35 22. El descodificador de señal de audio según una de las reivindicaciones 1 a 21, en el que el separador de objetos está configurado para obtener selectivamente un valor de correlación entre objetos común (IOCL,R) asociado con el objeto de audio del segundo tipo de objeto de audio en función de la información paramétrica relativa a un objeto si se comprueba que hay dos objetos de audio del segundo tipo de objeto de audio, y para ajustar el valor de correlación entre objetos asociado con los objetos de audio del segundo tipo de objeto de audio a cero si se comprueba que hay más o menos que dos objetos de audio del segundo tipo de objeto de audio; y
- 40 en el que el separador de objetos está configurado para utilizar el valor de correlación entre objetos común para un cálculo de entradas de una matriz (**M**); y
- en el que el separador de objetos está configurado para utilizar el valor de correlación entre objetos común asociada con los objetos de audio del segundo tipo de objeto de audio para obtener el o los más canales de audio que representan la segunda información de audio.
- 45 23. El descodificador de señal de audio según una de las reivindicaciones 1 a 22, en el que el procesador de señal de audio está configurado para representar la segunda información de audio en función de la información paramétrica relativa a un objeto, para obtener una representación parametrizada de los objetos de audio del segundo tipo de objeto de audio como la versión procesada de la segunda información de audio.
- 50 24. El descodificador de señal de audio según una de las reivindicaciones 1 a 23, en el que el separador de objetos está configurado para proporcionar la segunda información de audio de modo que la segunda información de audio describe más de dos objetos de audio del segundo tipo de objeto de audio.
25. El descodificador de señal de audio según la reivindicación 24, en el que el separador de objetos está configurado para obtener, como la segunda información de audio, una representación de señal de audio de un canal o una representación de señal de audio de dos canales que representa más de dos objetos de audio del segundo tipo de objeto de audio.

26. El descodificador de señal de audio según una de las reivindicaciones 1 a 25, en el que el procesador de señal de audio está configurado para recibir la segunda información de audio y para procesar la segunda información de audio en función de la información paramétrica relativa a objeto, teniendo en cuenta información paramétrica relativa a un objeto asociado con más de dos objetos de audio del segundo tipo de objeto de audio.
- 5 27. El descodificador de señal de audio según una de las reivindicaciones 1 a 26, en el que el descodificador de señal de audio está configurado para extraer una información de número total de objetos (bsNumObjects) y una información de número de objetos de primer plano (bsNum- GroupsFGO) a partir de una información de configuración (SAOCSpecificConfig) de la información paramétrica relativa a un objeto, y para determinar el número de objetos de audio del segundo tipo de objeto de audio por formar una diferencia entre la información de número total de objetos y la información de número de objetos de primer plano.
- 10 28. El descodificador de señal de audio según una de las reivindicaciones 1 a 27, en el que el separador de objetos está configurado para utilizar información paramétrica relativa a un objeto asociado con N-NEAO objetos de audio del primer tipo de objeto de audio para obtener, como la primera información de audio, N-NEAO señales de audio (X_{EAO}) que representan los N-NEAO objetos de audio del primer tipo de objeto de audio y para obtener, como la segunda información de audio, una o dos señales de audio (X_{OBJ}) que representan los N-NEAO objetos de audio del segundo tipo de objeto de audio, procesar los N-NEAO objetos de audio del segundo tipo de objeto de audio como un objeto de audio de un único canal o de dos canales; y
- 15 en el que el procesador de señal de audio está configurado para representar individualmente los N-NEAO objetos de audio representados por la una o las dos señales de audio de la segunda información de audio empleando la información paramétrica relativa a un objeto asociado con los N-NEAO objetos de audio del segundo tipo de objeto de audio.
- 20 29. Procedimiento para proporcionar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y una información paramétrica relativa a un objeto, comprendiendo el procedimiento:
- 25 descomponer la representación de señal de mezcla hacia abajo, para proporcionar una primera información de audio que describe un primer conjunto de uno o más objetos de audio de un primer tipo de objeto de audio, y una segunda información de audio que describe un segundo conjunto de uno o más objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y empleando al menos una parte de la información paramétrica relativa a un objeto, en el que la segunda información de audio es una información de
- 30 audio que describe los objetos de audio del segundo tipo de objeto de audio de una manera combinada; y
- procesar la segunda información de audio en función de la información paramétrica relativa a un objeto, para obtener una versión procesada de la segunda información de audio; y
- combinar la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;
- 35 en el que la representación de señal de mezcla hacia arriba se proporciona en función de una información residual asociada con un subconjunto de objetos de audio representados por la representación de señal de mezcla hacia abajo, en el que se descompone la representación de señal de mezcla hacia abajo, para proporcionar la primera información de audio que describe un primer conjunto de uno o más objetos de audio de un primer tipo de objeto de audio al cual la información residual está asociada, y describiendo la segunda información de audio un segundo conjunto de uno o más objetos de audio de un segundo tipo de objeto de audio, que no tiene información residual asociada, en función de la representación de señal de mezcla hacia abajo y empleando la información residual; en el que se realiza un procesamiento individual por objeto de los objetos de audio del segundo tipo de objeto de audio, empleando información paramétrica relativa a un objeto asociado con más de dos objetos de audio del segundo tipo de objeto de audio; y en el que la información residual describe una distorsión residual, que se prevé que
- 40 permanezca si un objeto de audio del primer tipo de objeto de audio se aísla simplemente empleando la información paramétrica relativa a un objeto.
- 45 30. Programa de ordenador para realizar el procedimiento según la reivindicación 29 cuando el programa de ordenador se ejecuta en un ordenador.

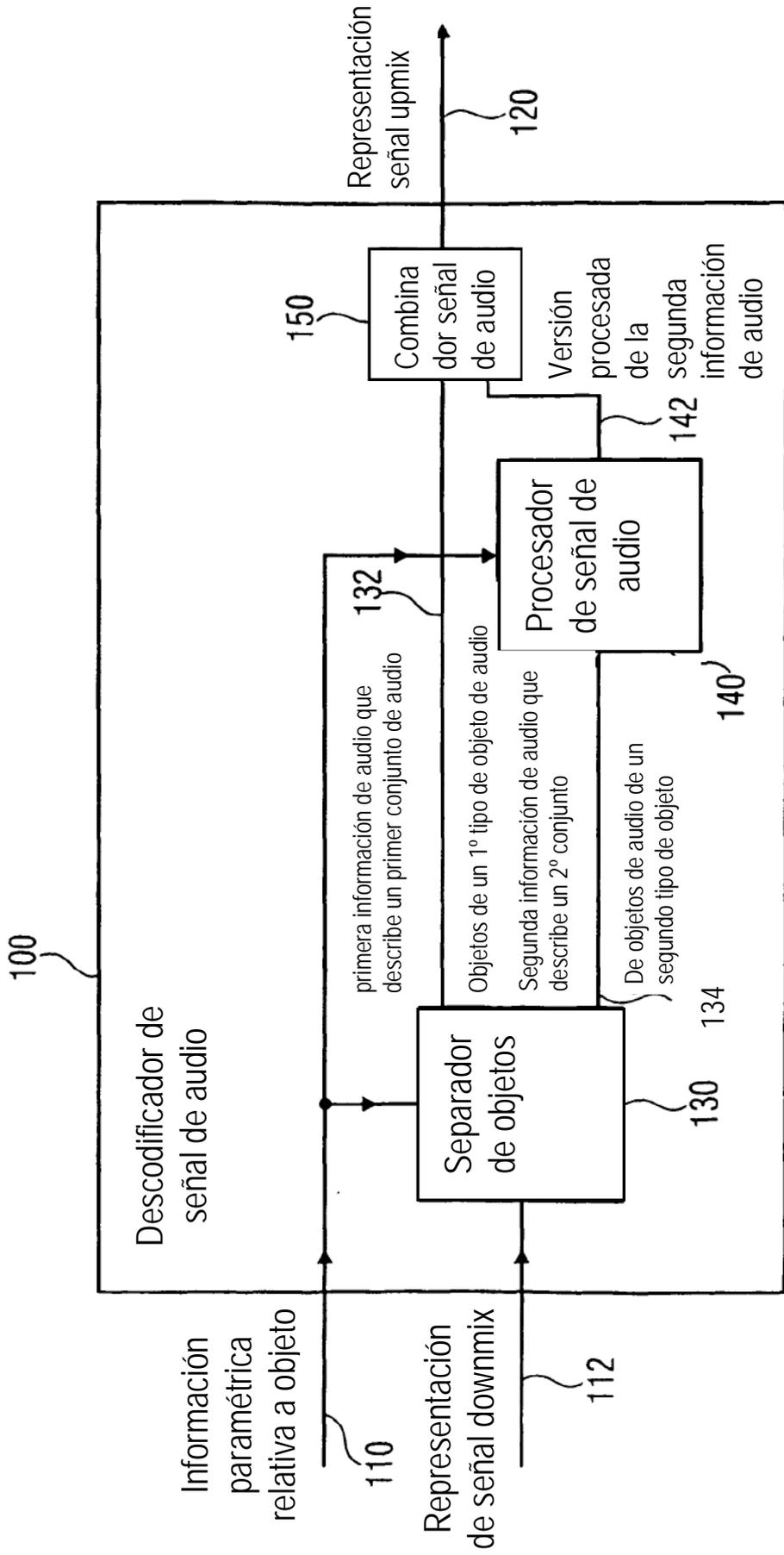
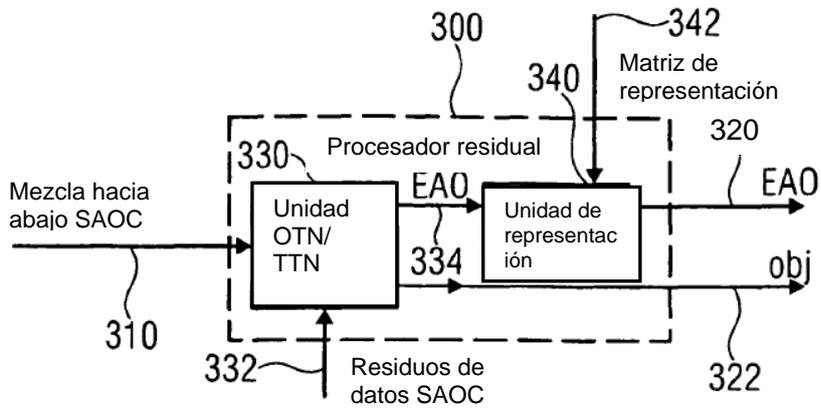
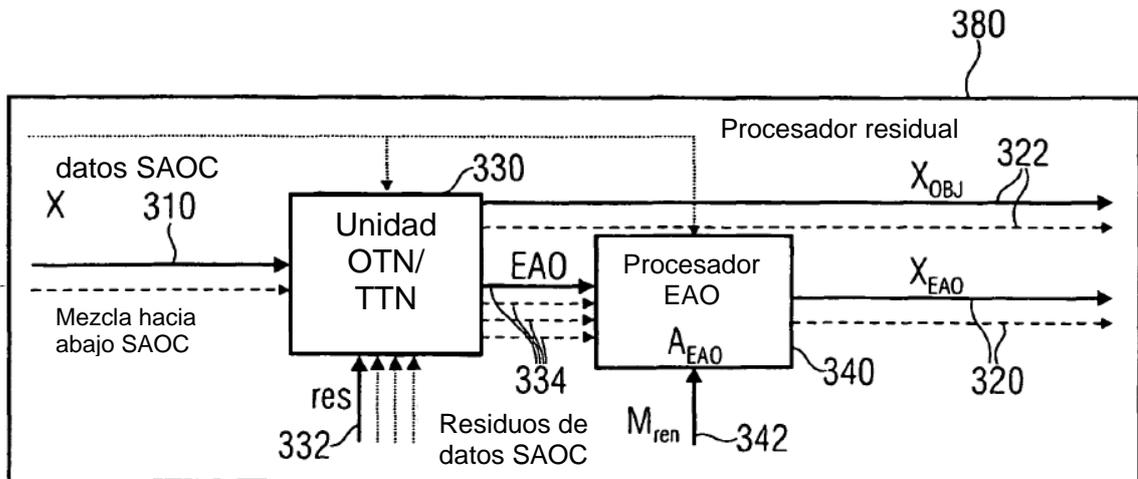


FIG 1



ARQUITECTURA DEL PROCESADOR RESIDUAL

FIG 3A



ARQUITECTURA DEL PROCESADOR RESIDUAL

FIG 3B

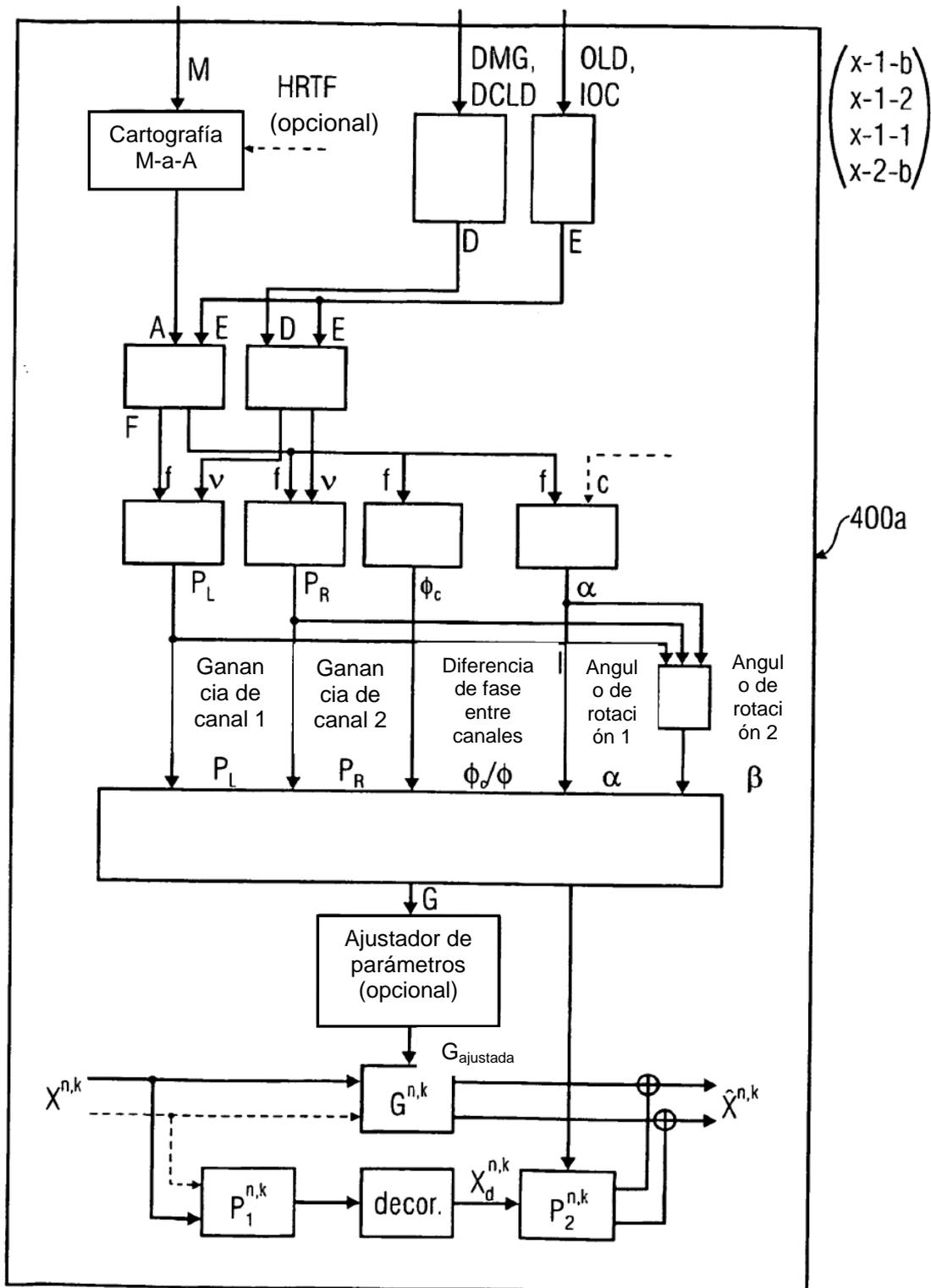


FIG 4B

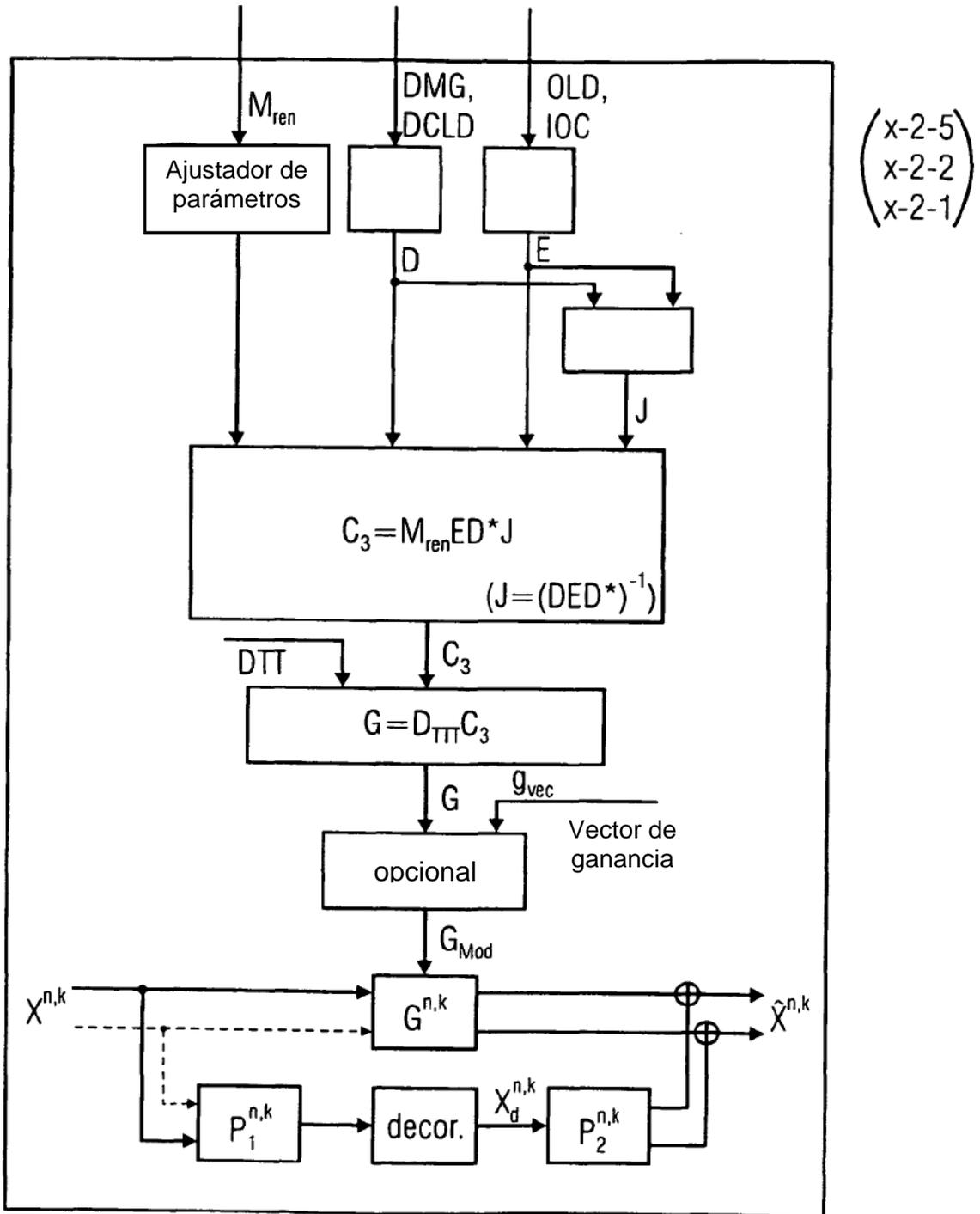


FIG 4C

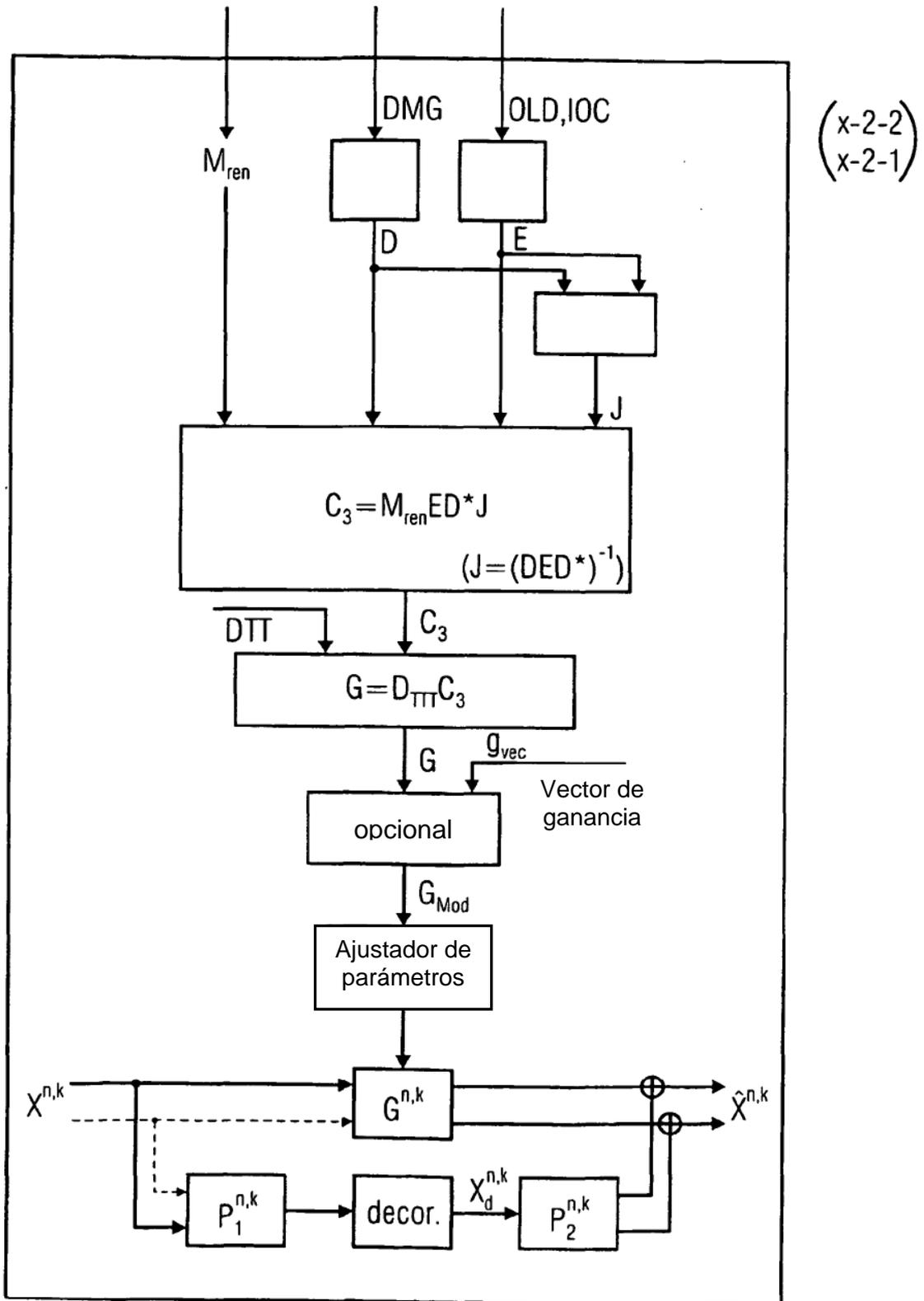


FIG 4D

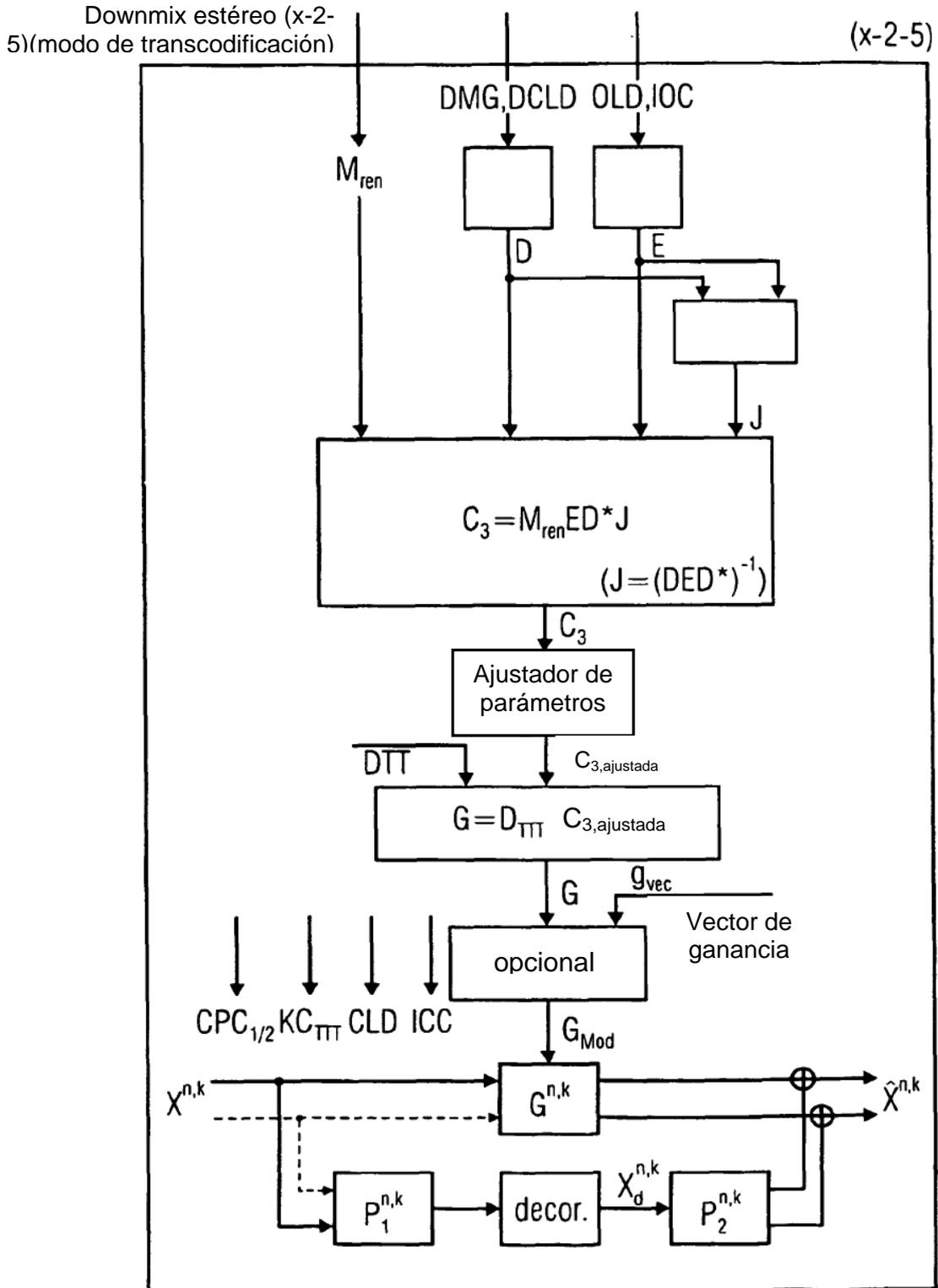


FIG. 4E

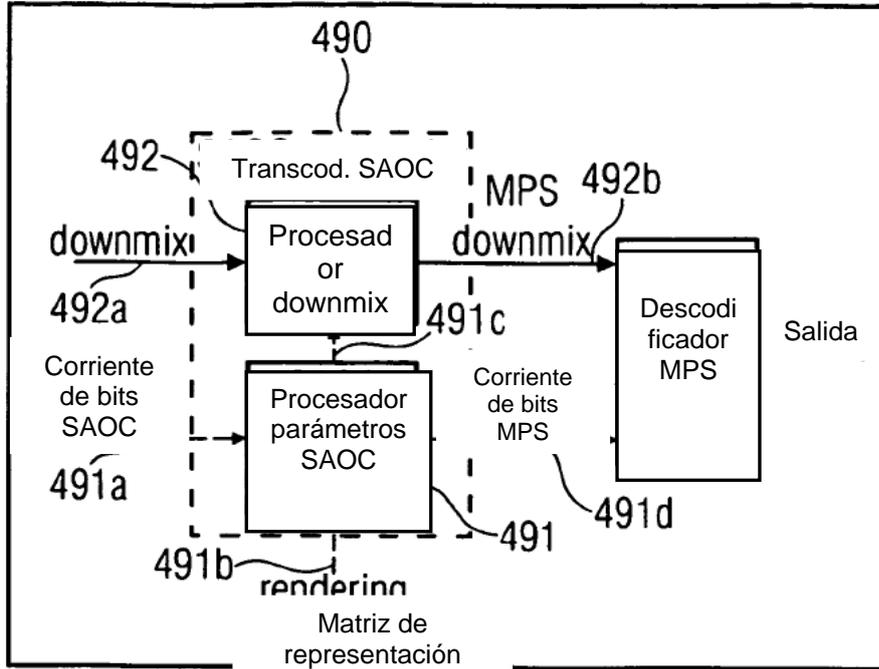


FIG 4F

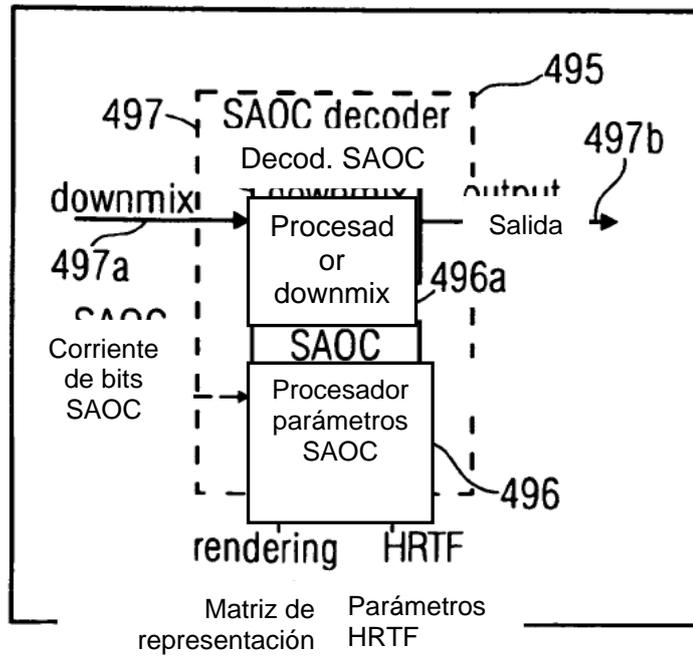
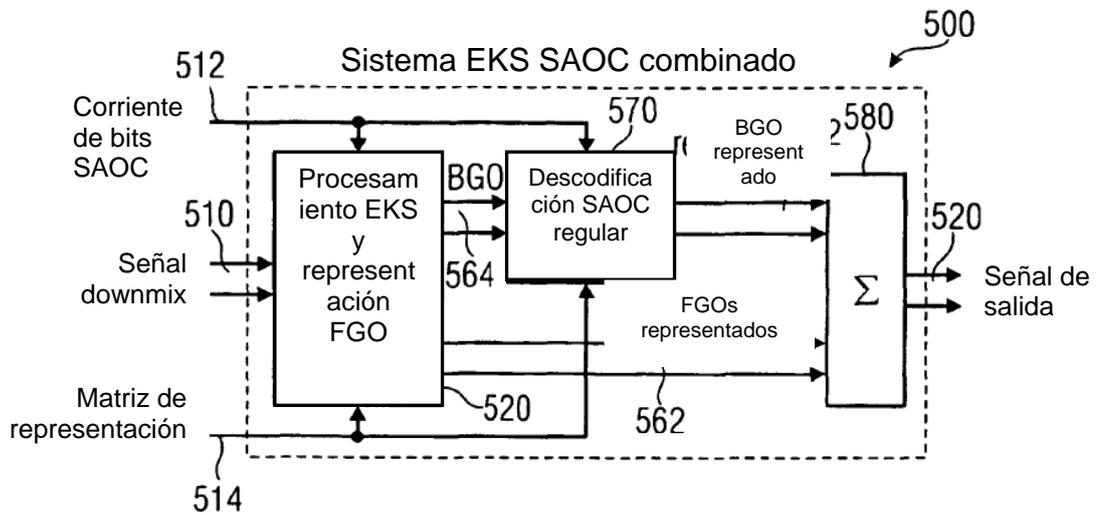
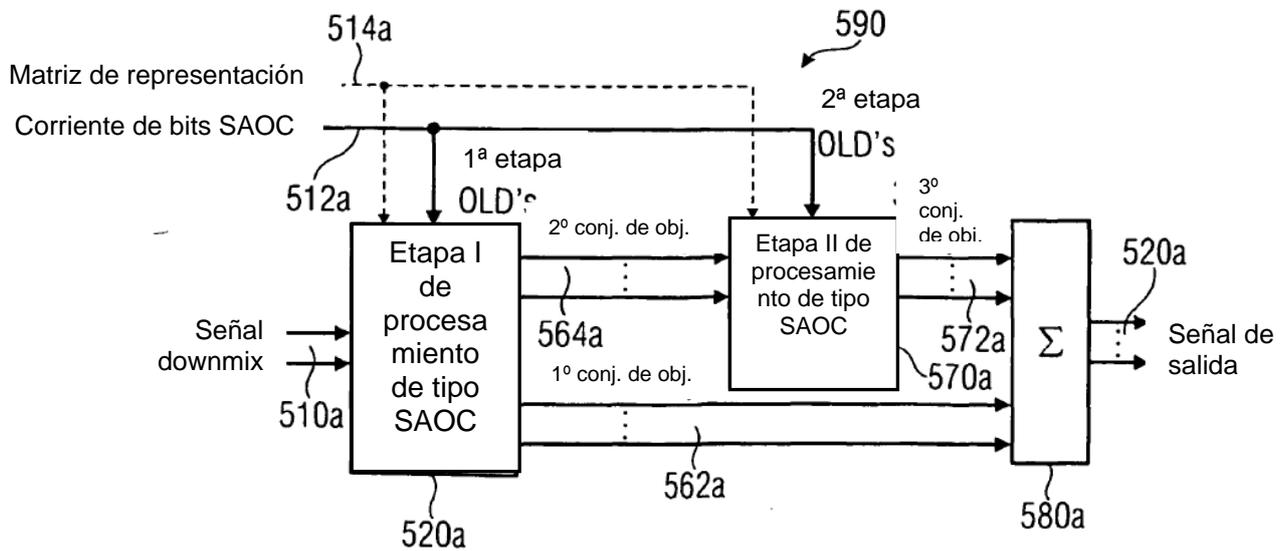


FIG 4G



ESTRUCTURA BÁSICA DEL SISTEMA EKS SAOC COMBINADO

FIG 5A



ESTRUCTURA GENERALIZADA DEL SISTEMA EKS SAOC COMBINADO

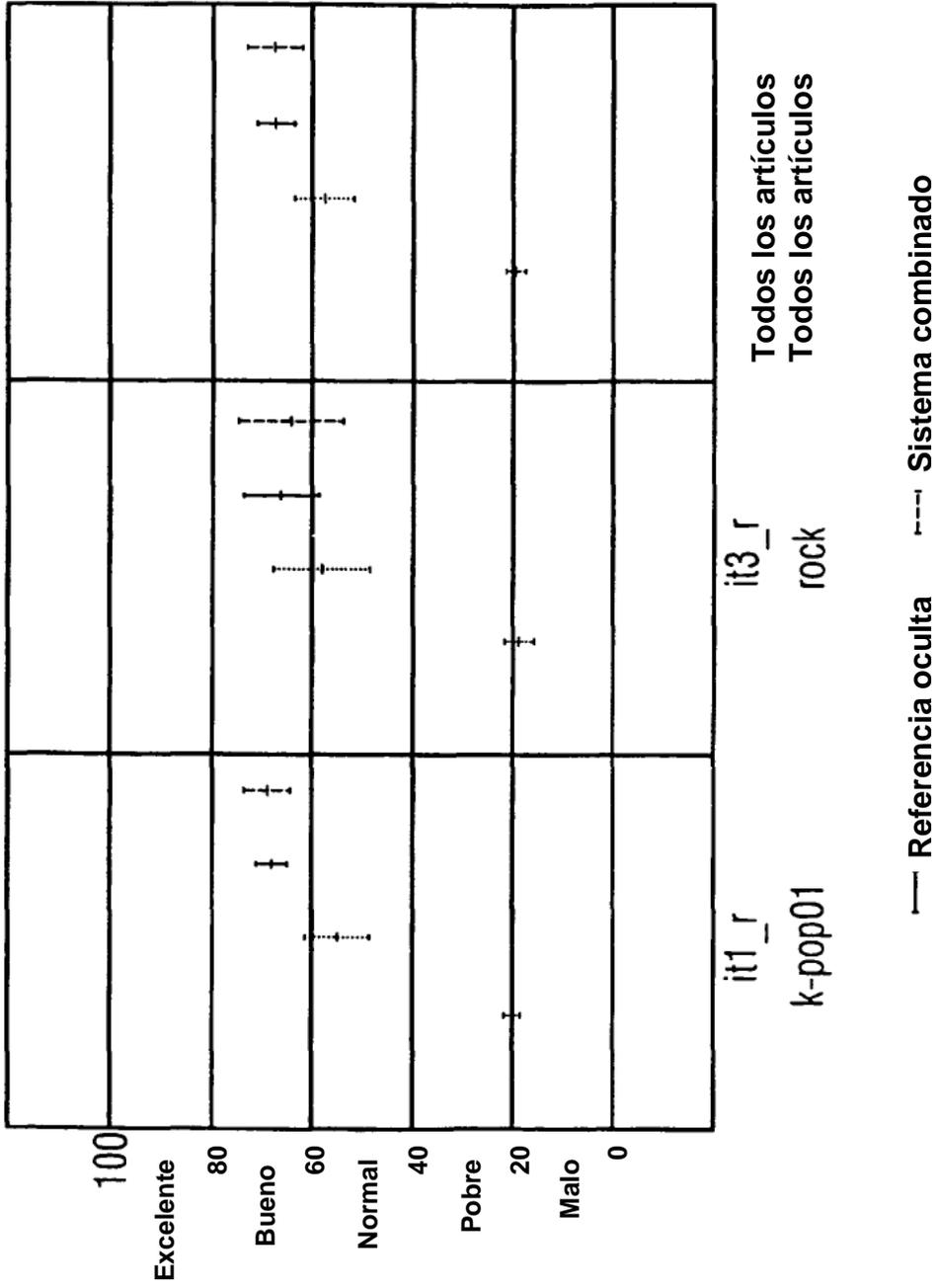
FIG 5B

Nombre de codificador	Descripción
Referencia oculta Anclaje inferior SAOC regular EKS actual Sistema combinado	Escena de audio idealmente representada (ref. oculta) Anclaje inferior @ 3.5kHz Modo decodificación SAOC regular Modo SAOC EKS actual Sistema combinado (EKS+SAOC) (Sistema propuesto)

FIG 6A

Sistema	Corriente de bits	Repres. Karaoke	Repres. Clásica
Sistema combinado SAOC regular EKS actual	Datos SAOC A+res. A Datos A SAOC Datos SAOC B+res. B	X (X) X	X X -

FIG 6B



Todos los artículos
 Todos los artículos

it1_r
 k-pop01

FIG 6D

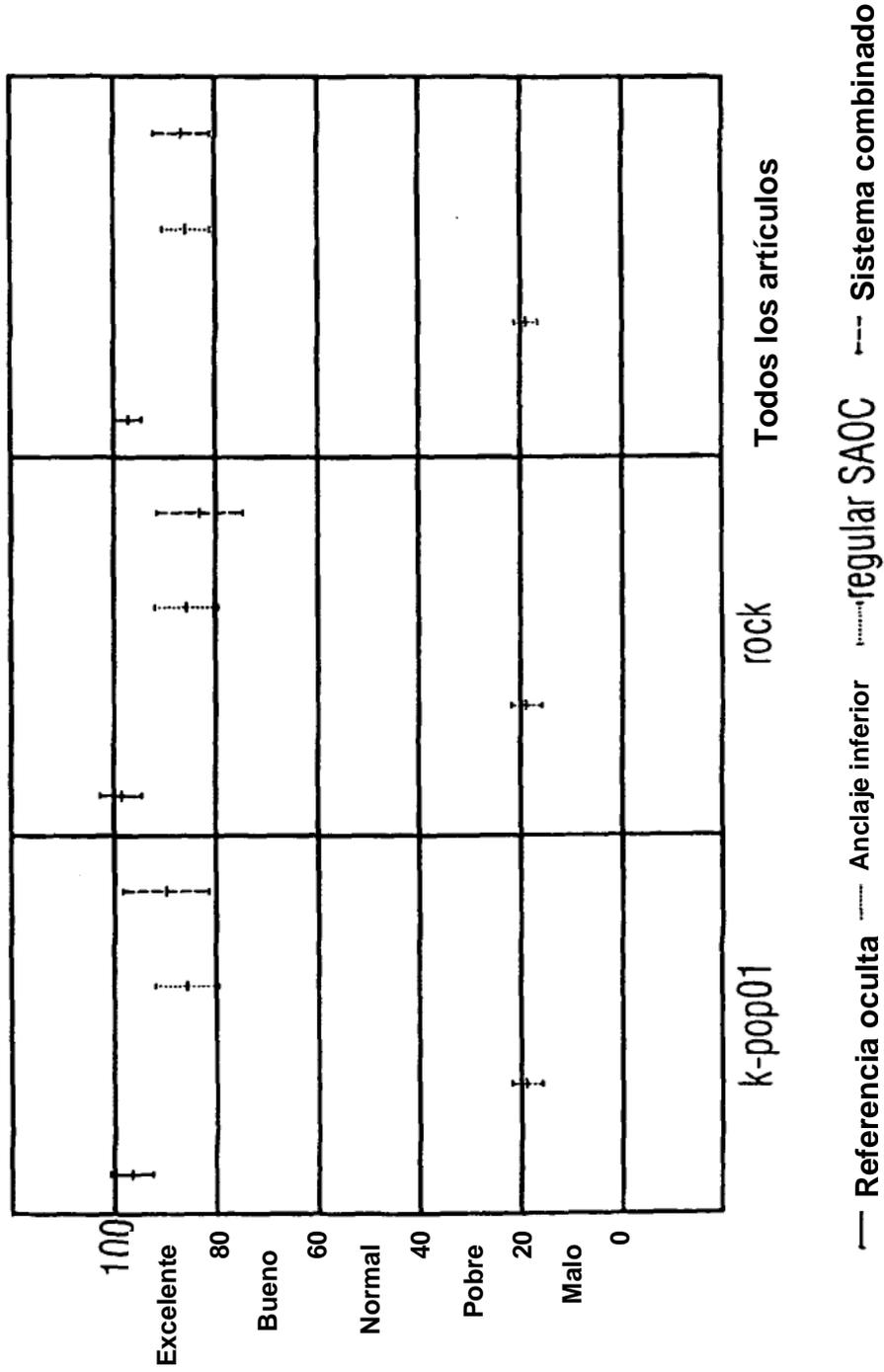


FIG 6E

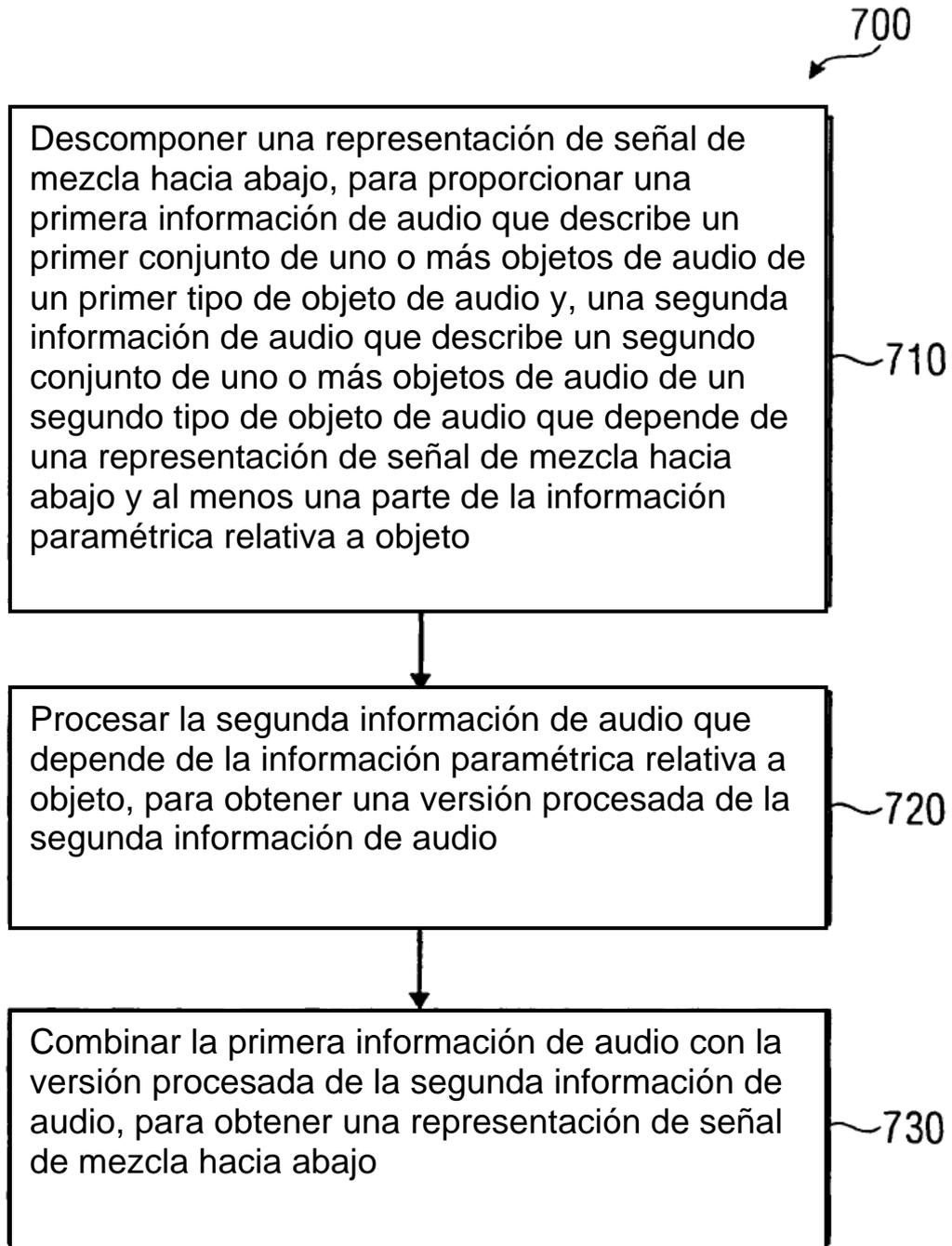
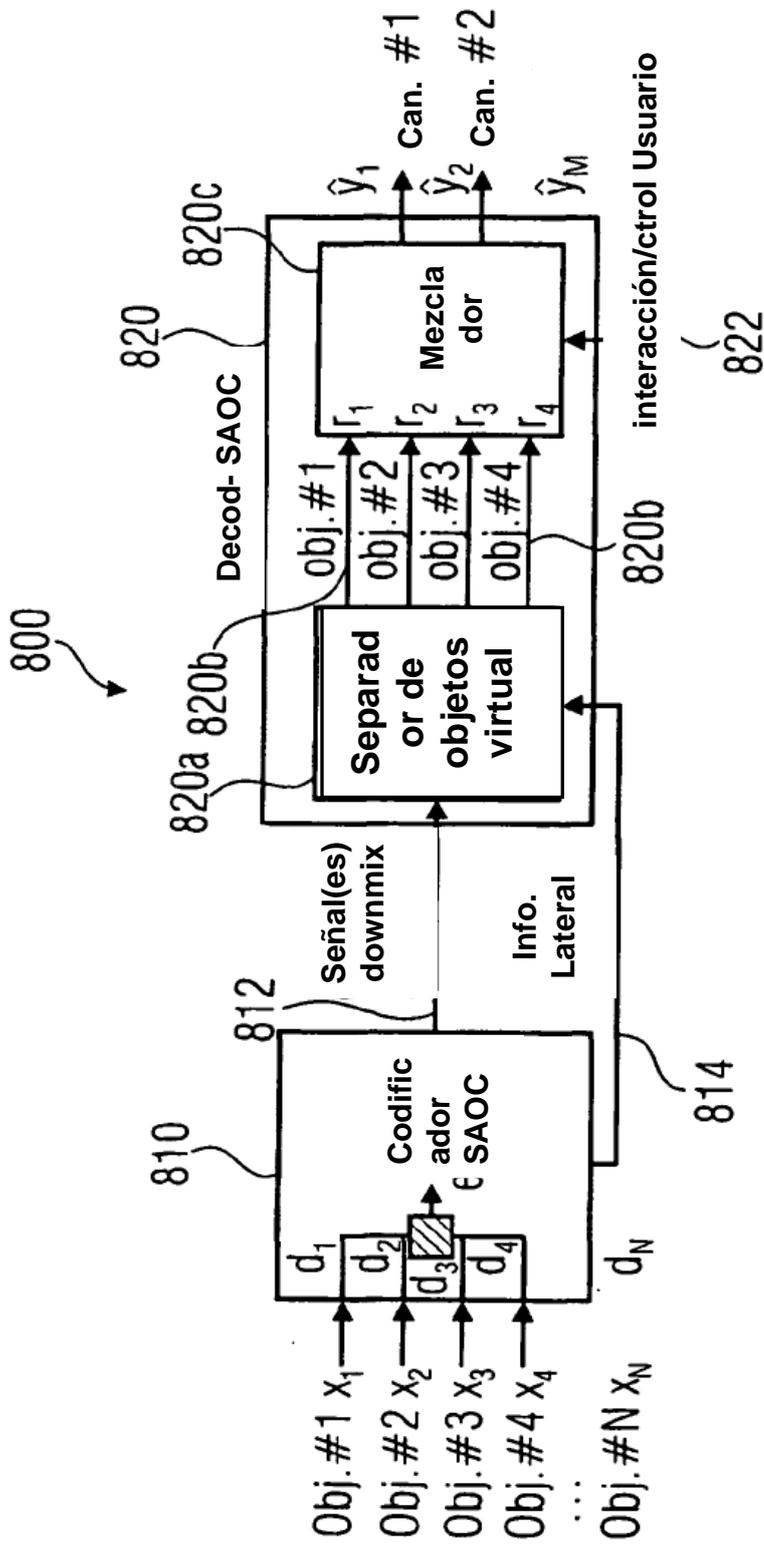
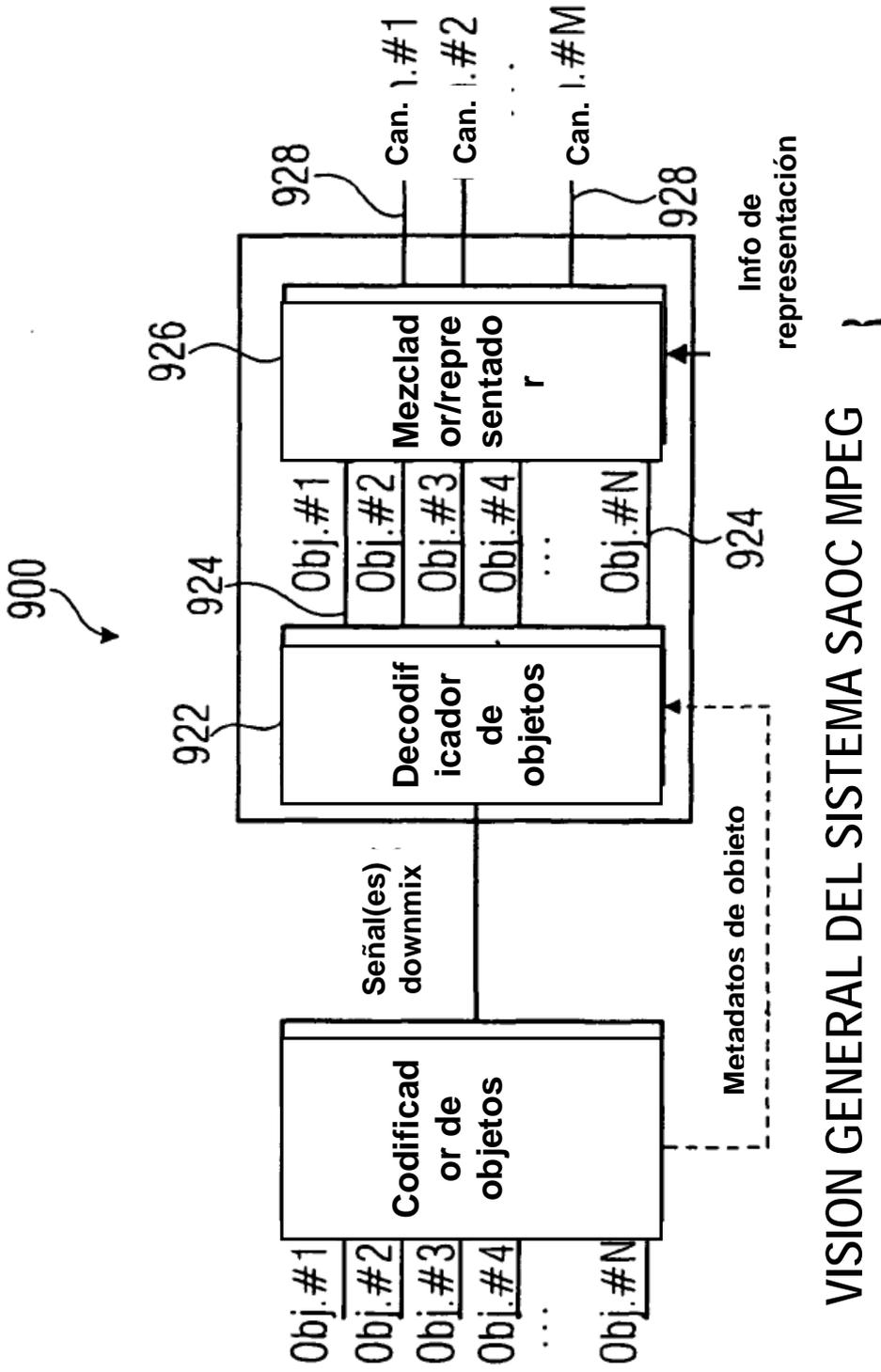


FIG 7



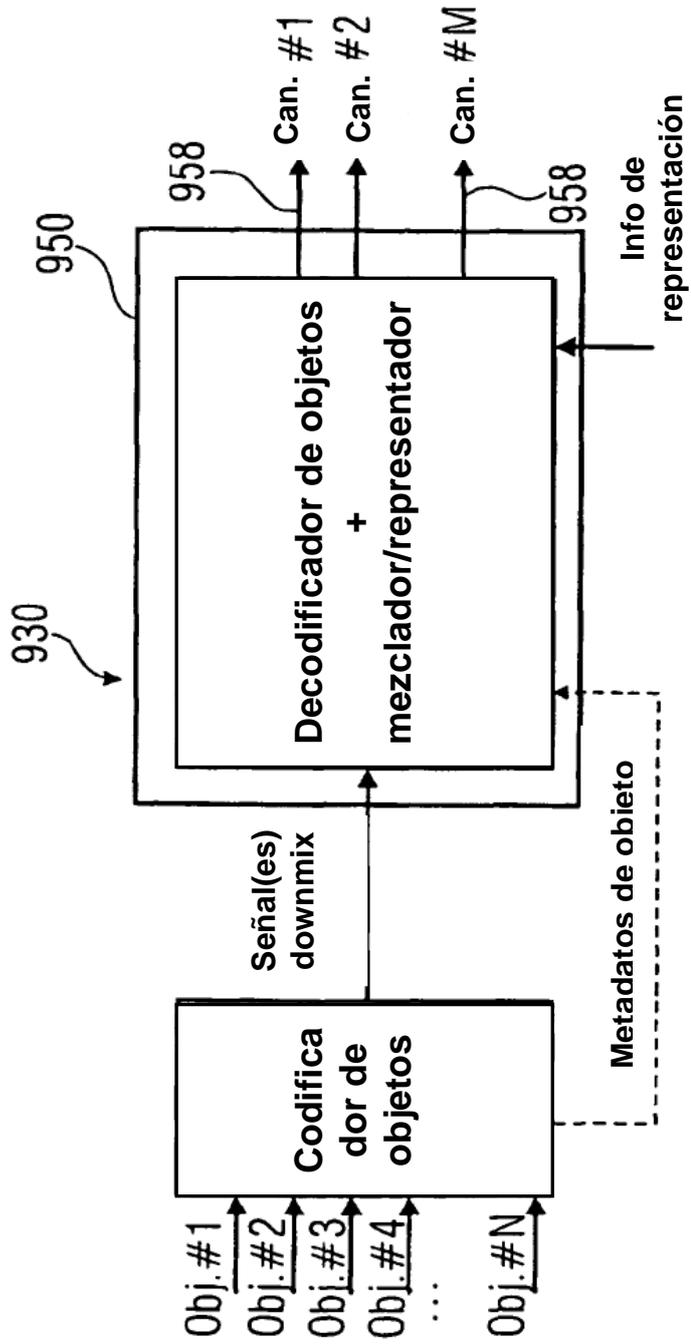
VISION GENERAL DEL SISTEMA SAOC MPEG

FIG 8



VISION GENERAL DEL SISTEMA SAOC MPEG

FIG 9A



DECODIFICADOR Y MEZCLADOR INTEGRADOS

FIG 9B

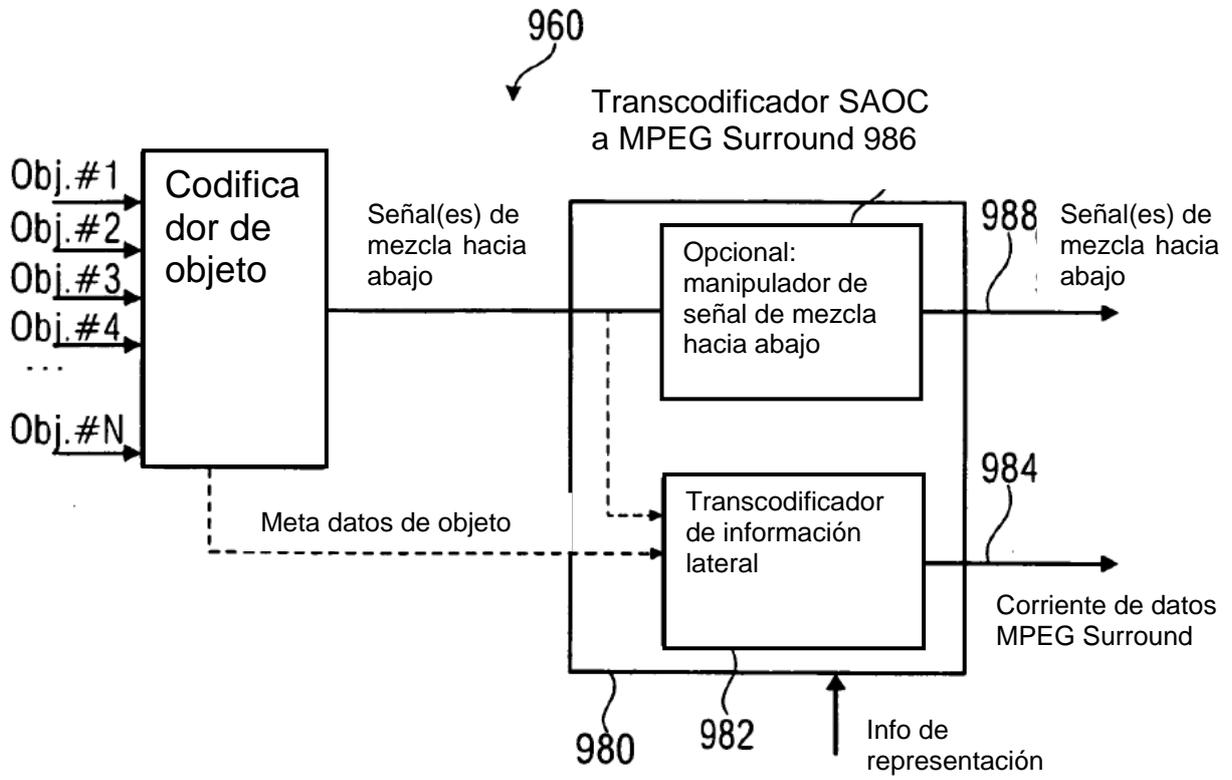


FIG 9C

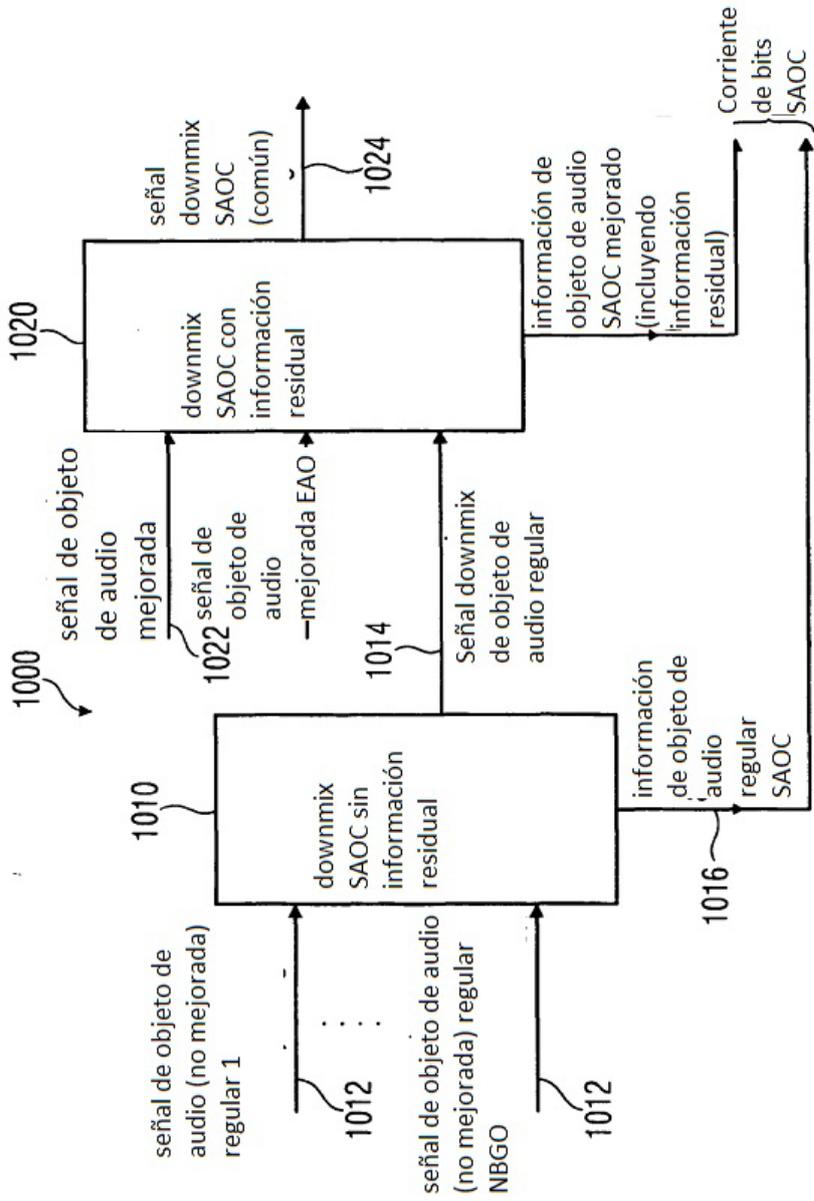


FIG 10