



# OFICINA ESPAÑOLA DE PATENTES Y MARCAS

**ESPAÑA** 



11) Número de publicación: 2 524 428

(51) Int. CI.:

G10L 19/008 (2013.01) G10L 19/20 (2013.01) H04S 7/00 (2006.01) G10H 1/36 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA

**T3** 

- (96) Fecha de presentación y número de la solicitud europea: 23.06.2010 E 12183562 (3) (97) Fecha y número de publicación de la concesión europea: 27.08.2014 EP 2535892
- (54) Título: Decodificador de señales de audio, procedimiento para decodificar una señal de audio y programa de computación que utiliza etapas en cascada de procesamiento de objetos de audio
- ③ Prioridad:

24.06.2009 US 220042 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente: 09.12.2014

(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 **TERENTIV, LEONID** 

(74) Agente/Representante:

PONTI SALES, Adelaida

### **DESCRIPCIÓN**

Decodificador de señales de audio, procedimiento para decodificar una señal de audio y programa de computación que utiliza etapas en cascada de procesamiento de objetos de audio

#### Campo Técnico

5

10

15

20

25

30

35

40

45

50

55

[0001] Las realizaciones de acuerdo con la invención se relacionan con un decodificador de señales de audio para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto.

**[0002]** Otras realizaciones de acuerdo con la invención se relacionan con un procedimiento para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto.

[0003] Otras realizaciones de acuerdo con la invención se relacionan con un programa de computación.

[0004] Algunas realizaciones de acuerdo con la invención se relacionan con un sistema SAOC de Karaoke/Solos mejorado.

#### Antecedentes de la Invención

[0005] En las disposiciones de audio modernas, es conveniente transferir y almacenar información de audio de manera eficiente con respecto a la velocidad de transmisión de bits. Además, con frecuencia es conveniente reproducir un contenido de audio empleando una pluralidad de dos o incluso más parlantes, que están espacialmente distribuidos en un recinto. En esos casos, es conveniente aprovechar las capacidades de dicha disposición de múltiples parlantes para permitir que un usuario identifique espacialmente diferentes contenidos de audio o diferentes elementos de un único contenido de audio. Esto se puede lograr distribuyendo individualmente los diferentes contenidos de audio en los diferentes parlantes.

[0006] Dicho de otro modo, en la técnica del procesamiento de audio, la transmisión de audio y el almacenamiento de audio, existe un creciente interés por manejar contenidos de canales múltiples para mejorar la impresión auditiva. El uso de contenido de audio de múltiples canales trae aparejadas significativas mejoras para el usuario. Por ejemplo, se puede obtener una impresión auditiva tridimensional, lo que trae consigo una mayor satisfacción del usuario en aplicaciones de entretenimiento. Sin embargo, los contenidos de audio en múltiples canales también son de utilidad en entornos profesionales, por ejemplo en aplicaciones de conferencias telefónicas, puesto que se puede mejorar la inteligibilidad de los parlantes utilizando una reproducción de audio en canales múltiples.

[0007] Sin embargo, también es conveniente contar con un buen equilibrio entre buena calidad de audio y requerimientos de velocidad de transferencia de bits, a fin de evitar una excesiva carga de recursos causada por aplicaciones de canales múltiples.

[0008] Últimamente se han propuesto técnicas paramétricas para la transmisión eficiente de bits y/o para el almacenamiento de escenas de audio que contienen múltiples objetos de audio, por ejemplo, Binaural Cue Coding (Type I) (Ver, por ejemplo la referencia [BCC]), Joint Source Coding (ver, por ejemplo, la referencia [JSC J), y MPEG Spatial Audio Object Coding (SAOC) (ver, por ejemplo, las referencias [SAOCI], [SAOC2]).

[0009] Estas técnicas apuntan a reconstruir perceptualmente la escena de salida de audio ventajosa en lugar de hacerlo por coincidencia de formas de onda.

[0010] La Fig. 8 ilustra una reseña general de la disposición correspondiente a ese sistema (en este caso: MPEG SAOC). El sistema MPEG SAOC 800 ilustrado en la Fig. 8 comprende un codificador SAOC 810 y un decodificador SAOC 820. El codificador SAOC 810 recibe una pluralidad de señales de objeto X<sub>1</sub> a X<sub>N</sub>, que pueden estar representadas, por ejemplo, en forma de señales en el dominio del tiempo o en forma señales en el dominio del tiempo y la frecuencia (por ejemplo, en forma de una serie de coeficientes de transformación del tipo de transformadas de Fourier, o en forma de señales de subbanda QMF). El codificador SAOC 810 también recibe, por lo general, coeficientes de mezcla descendente d<sub>1</sub> a d<sub>N</sub>, que están asociados con las señales de objeto x<sub>1</sub> a x<sub>N</sub>. Se puede disponer de series independientes de de coeficientes de mezcla descendente por cada canal de de la señal de mezcla descendente. El codificador SAOC 810 está configurado típicamente para obtener un canal de la señal de mezcla descendente

combinando las señales de objeto  $X_1$  a  $X_N$  de acuerdo con los coeficientes de mezcla descendente asociados  $d_1$  a  $d_N$ . Por lo general, hay menos canales de mezcla descendente que señales de objeto  $X_1$  a  $X_N$ . Para dar lugar (por lo menos aproximadamente) a una separación (o tratamiento separado) de las señales de objeto del lado del decodificador SAOC 820, el codificador SAOC 8 10 suministra tanto la una o más señales de mezcla descendente (designados canales de mezcla descendente) 812 como una información complementaria 814. La información complementaria 814 describe características de las señales de objeto  $X_1$  a  $X_N$ , para dar lugar a un procesamiento específico de objetos del lado del decodificador.

5

10

15

20

25

30

45

50

55

[0011] El decodificador SAOC 820 está configurado para recibir tanto dicha una o más señales de mezcla descendente 812 como la información complementaria 814. Además, el decodificador SAOC 820 está configurado por lo general para recibir una información sobre interacción con el usuario y/o una información de interacción con el usuario y/o una información de control del usuario 822, que describe una configuración de renderización. Por ejemplo, la información de interacción con el usuario/información de control del usuario 822 puede describir una configuración de parlantes y la colocación espacial pretendida de los objetos provistos por las señales de objeto X<sub>1</sub> a X<sub>N</sub>, El decodificador SAOC 820 está configurado para producir, por ejemplo, una pluralidad de señales decodificadas de canales de mezcla ascendente

[0012] El decodificador SAOC 820 está configurado para producir, por ejemplo, una pluralidad de señales de canales decodificadas de mezcla descendente ŷ<sub>1</sub> a ŷ<sub>M</sub>. Las señales de canales de mezcla ascendente pueden estar asociadas, por ejemplo a parlantes individuales de una disposición de renderización de múltiples parlantes. El decodificador SAOC 820 puede comprender, por ejemplo, un separador de objetos 820a, que está configurado para reconstruir, por lo menos aproximadamente, las señales de objeto x1 a xN sobre la base de dichas una o más señales de mezcla descendente 812 e información complementaria 814, a fin de obtener así las señales de objetos reconstruidas 820b. Sin embargo, las señales de objetos reconstruidas 820b se pueden desviar hasta cierto punto de las señales de objetos originales x<sub>1</sub> a x<sub>N</sub>, por ejemplo, porque la información complementaria 814 no basta totalmente para una reconstrucción perfecta debido a restricciones de velocidad de transmisión de bits. El decodificador SAOC 820 puede comprender además un mezclador 820c, que puede estar configurado para recibir las señales de objetos reconstruidas 820b y la información de interacción con el usuario/información de control de usuarios 822, y para suministrar, sobre la base de éstas, las señales de canales de mezcla ascendente ŷ<sub>1</sub> a ŷ<sub>M</sub>. El mezclador 820 puede estar configurado para usar la información de interacción con el usuario /información de control de usuarios 822 para determinar la contribución de las señales de objetos reconstruidas 820b individuales a las señales de canales de mezcla ascendente  $\hat{y}_1$  a  $\hat{y}_M$ . La información de interacción con el usuario/información de control de usuarios 822 puede comprender, por ejemplo, parámetros de renderización (que también se denominan coeficientes de renderización), que determinan la contribución de las señales de objetos reconstruidas 822 individuales a las señales de canales de mezcla ascendente ŷ<sub>1</sub> a ŷ<sub>M</sub>.

[0013] Sin embargo, cabe señalar que, en muchas realizaciones, la separación de los objetos, que está indicada por el separador de objetos 820a de la Fig. 8, y la mezcla, que está indicada por el mezclador 820c de la Fig. 8, se llevan a cabo en un solo paso. Para este fin, se pueden computar los parámetros generales que describen un mapeo directo de dichas una o más señales de mezcla descendente 812 sobre las señales de canales de mezcla ascendente ŷ<sub>1</sub> a ŷ<sub>M</sub>. Estos parámetros se pueden computar sobre la base de la información complementaria y la información de interacción con el usuario/ información de control de usuarios 822.

[0014] Tomando, ahora, como referencia las Figs. 9a, 9b y 9c, se describen diferentes aparatos para obtener una representación de la señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente e información complementaria relacionada con objetos. La Fig. 9a ilustra un diagrama esquemático de bloques de un sistema MPEG SAOC 900 que comprende un decodificador SAOC 920. El decodificador SAOC 920 comprende, como bloques funcionales separados, un decodificador de objetos 922 y un mezclador/renderizador 926. El decodificador de objetos 922 produce una pluralidad de señales de objetos reconstruidas 924 que depende de la representación de la señal de mezcla descendente (por ejemplo, en forma de una o más señales de mezcla descendente representadas en el dominio del tiempo o en el dominio del tiempo—frecuencia) e información complementaria relacionada con objetos (por ejemplo, en forma de metadatos de objeto). El mezclador/renderizador 926 recibe las señales de objetos reconstruidas 924 asociadas a una pluralidad de objetos N y produce, sobre la base de éstas, una o más señales de canales de mezcla ascendente 928. En el decodificador SAOC 920, la extracción de las señales de objeto 924 se realiza en forma independiente de la mezcla/renderización, lo que da lugar a una separación de la funcionalidad de decodificación de objetos de la funcionalidad de mezcla/renderización, aunque trae aparejada una complejidad informática relativamente elevada.

[0015] Haciendo referencia, ahora, a la Fig. 9b, se describe brevemente otro sistema MPEG SAOC 930, que comprende un decodificador SAOC 950. El decodificador SAOC 950 produce una pluralidad de señales de canales de mezcla ascendente 958 que depende de una representación de señal de mezcla descendente (por ejemplo, en forma de una o

más señales de mezcla descendente) e información complementaria relacionada con un objeto (por ejemplo, en forma de metadatos de objeto). El decodificador SAOC 950 comprende una combinación de decodificador de objeto y mezclador/renderizador, que está configurado para obtener las señales de canales de mezcla ascendente 958 en un proceso de mezclado conjunto sin separación de la decodificación de objetos y la mezcla/renderización, donde los parámetros para dicho proceso conjunto de mezcla ascendente dependen tanto de la información complementaria relacionada con objetos y la información sobre la renderización. El proceso conjunto de mezcla ascendente depende asimismo de la información sobre mezcla descendente, que se considera parte de la información complementaria relacionada con objetos.

10 **[0016]** Para resumir lo anterior, la provisión de las señales de canales de mezcla ascendente 928, 958 se puede realizar en un proceso de un solo paso o un proceso de dos pasos.

5

15

20

25

30

35

40

[0017] Haciendo referencia, ahora, a la Fig. 9c, se describe un sistema MPEG SAOC 960. El sistema SAOC 960 comprende un transcodificador de SAOC a MPEG Envolvente 980, en lugar de un decodificador SAOC.

[0018] El transcodificador de SAOC a MPEG Envolvente comprende un transcodificador de información complementaria 982, que está configurado para recibir la información complementaria relacionada con objetos (por ejemplo, en forma de metadatos de objeto) u, opcionalmente, información de dicha una o más señales de mezcla descendente y la información sobre la renderización. El transcodificador de información complementaria también está configurado para proporcionar una información complementaria sobre MPEG Envolvente 984 (por ejemplo, en forma de flujo de bits de MPEG Envolvente) sobre la base de ciertos datos recibidos. En consecuencia, el transcodificador de información complementaria 982 está configurado para transformar una información complementaria relacionada con objetos (paramétrica), que es emitida por un codificador de objetos, en una información complementaria (paramétrica) relacionada con los canales 984, tomando en cuenta la información sobre la renderización y, opcionalmente, la información sobre el contenido de dichas una o más señales de mezcla descendente.

[0019] Opcionalmente, el transcodificador de SAOC a MPEG Envolvente 980 puede estar configurado para manipular dichas una o más señales de mezcla descendente, descriptas, por ejemplo, por la representación de la señal de mezcla descendente, para obtener una representación de señal de mezcla descendente manipulada 988. Sin embargo, se puede omitir el manipulador de señal de mezcla descendente 986, de manera que la representación de señal de mezcla descendente 988 de salida del transcodificador de SAOC a MPEG Envolvente 980 sea idéntica a la representación de señal de mezcla descendente de entrada del transcodificador de SAOC a MPEG Envolvente. El manipulador de señales de mezcla descendente 986 se puede utilizar, por ejemplo, en caso de que la información complementaria de MPEG Envolvente relacionada con los canales 984 no permita la producción de una impresión auditiva adecuada sobre la base de la representación de señal de mezcla descendente de entrada del transcodificador de SAOC a MPEG Envolvente 980, lo que puede ocurrir en algunas constelaciones de renderización.

[0020] En consecuencia, el transcodificador de SAOC a MPEG Envolvente 980 da origen a la representación de la señal de mezcla descendente 988 y el flujo de bits de MPEG Envolvente 984 razón por la cual se puede generar una pluralidad de señales de canales de mezcla ascendente, que representan los objetos de audio de acuerdo con la información sobre la renderización ingresada en el transcodificador de SAOC a MPEG Envolvente 980 utilizando un decodificador de MPEG Envolvente que recibe el flujo de bits de MPEG Envolvente 984 y la representación de la señal de mezcla descendente 988.

[0021] Para resumir lo anterior, se pueden emplear diferentes conceptos para decodificar señales de audio codificadas por SAOC. En algunos casos, se utiliza un decodificador SAOC, que produce señales de canales de mezcla ascendente (por ejemplo, las señales de canales de mezcla ascendente 928, 958) que dependen de la representación de la señal de mezcla descendente y la información complementaria paramétrica relacionada con los objetos. Se pueden ver ejemplos de este concepto en las Figs. 9a y 9b. Por otro lado, se puede transcodificar la información de audio codificada por SAOC para obtener una representación de señal de mezcla descendente (por ejemplo, una representación de señal de mezcla descendente 988) y una información complementaria relacionada con los canales (por ejemplo, el flujo de bits de MPEG Envolvente relacionado con los canales 984), que puede ser utilizada por un decodificador de MPEG Envolvente para producir las pretendidas señales de canales de mezcla ascendente.

[0022] En el sistema MPEG SAOC 800, una reseña general de la cual se presenta en la Fig. 8, se lleva a cabo el procesamiento general en forma selectiva de la frecuencia y se puede describir de la siguiente manera dentro de cada banda de frecuencias:

- Se realiza la mezcla descendente de N señales de audio de entrada de objeto x<sub>1</sub> a x<sub>N</sub> como parte del procesamiento del codificador SAOC. Para una mezcla descendente mono, los coeficientes están indicados por d<sub>1</sub> a d<sub>N</sub>. Además, el codificador SAOC 810 extrae información complementaria 814 que describe las características de los objetos de audio de entrada. En el caso de la MPEG SAOC, las relaciones de las potencias de objeto entre sí son la forma más básica de dicha información complementaria.
- La señal (o señales) de mezcla descendente 812 y la información complementaria 814 son transmitidas y/o almacenadas. Para este fin, se puede comprimir la señal de audio de mezcla descendente empleando codificadores de audio perceptuales muy conocidos tales como MPEG-1 Capa II o III (también conocido como ".mp3"), Codificación de Audio Avanzada MPEG (AAC), o cualquier otro codificador de audio
- Del lado del receptor, el decodificador SAOC 820 intenta conceptualmente restablecer la señal de objeto original ("separación de objetos") utilizando la información complementaria transmitida 814 (y, naturalmente, la una o más señales de mezcla descendente 812). Estas señales de objeto aproximadas (que también se denominan señales de objetos reconstruidas 820b) se mezclan a continuación en una escena objetivo representada por M canales de salida de audio (que pueden estar representados, por ejemplo, por las señales de canales de mezcla ascendente ŷ<sub>1</sub> a ŷ<sub>M</sub>) utilizando una matriz de renderización. En el caso de una salida mono, los coeficientes de la matriz de renderización están expresados por r<sub>1</sub> a r<sub>N</sub>.
- 20 En efecto, raramente se ejecuta la separación de las señales de objeto (o incluso nunca se las ejecuta), puesto que tanto el paso de separación (indicado por el separador de objetos 820a) como el paso de mezcla (indicado por el mezclador 820c) se combinan para obtener un solo paso de transcodificación, que con frecuencia da lugar a una enorme reducción de la complejidad informática.
- [0023] Se ha encontrado que ese tipo de esquema es tremendamente eficiente, tanto en términos de velocidad de transmisión de bits (sólo es necesario transmitir unos pocos canales de mezcla descendente más cierta información complementaria en lugar de N señales discretas de audio de objeto o un sistema discreto) y complejidad informática (la complejidad de procesamiento se relaciona principalmente con el número de canales de salida en lugar del número de objetos de audio). Otras ventajas para el usuario del lado de la recepción incluyen la libertad de elegir una configuración de renderización de su elección (mono, estéreo, envolvente, reproducción virtualizada con auriculares y demás) y la característica de interactividad con el usuario: se puede ajustar la matriz de renderización y, por consiguiente, la escena de salida, puede ser ajustada y cambiada interactivamente por el usuario según su voluntad, preferencias personales u otros criterios. Por ejemplo, es posible localizar los interlocutores de un grupo juntos en un área espacial para maximizar la discriminación de las demás personas que conversan. Esta interactividad se obtiene produciendo una interfaz de usuario del decodificador.
  - [0024] Por cada objeto de sonido transmitido, se puede ajustar su nivel relativo y (en el caso de la renderización no mono) la posición espacial de la renderización. Esto puede ocurrir en tiempo real al cambiar el usuario la posición de los dispositivos deslizantes de interfaz gráfica asociada (GUI) (por ejemplo: nivel de objeto = +5dB, posición de objeto = -30grados).
  - [0025] Sin embargo, se ha encontrado que es difícil manejar objetos de audio de diferentes tipos de objetos de audio en ese tipo de sistema. En particular, se ha encontrado que es difícil procesar objetos de audio de diferentes tipos de objetos de audio, por ejemplo, objetos de audio con los cuales se asocia diferente información complementaria, si el número total de objetos de audio a procesar no ha sido predeterminado.
  - [0026] En vista de esta situación, el objetivo de la presente invención consiste en crear un concepto que da lugar a la decodificación informáticamente eficiente y flexible de una señal de audio que comprende una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto, donde la información paramétrica relacionada con los objetos describe objetos de audio de dos o más tipos de objetos de audio diferentes.

#### Síntesis de la Invención

5

10

15

40

45

50

55

[0027] Este objetivo se logra mediante decodificadores de señales de audio para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto, procedimientos para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto y un programa de computación, de acuerdo con lo definido por las reivindicaciones independientes.

[0028] Una realización de acuerdo con la invención tal como se establece en las reivindicaciones independientes 1 a 3 crean decodificadores de señales de audio para producir una representación de señal de mezcla ascendente que depende de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto. El decodificador de señales de audio comprende un separador de objetos configurado para descomponer representación de señal de mezcla descendente, para producir una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio que depende de la representación de la señal de mezcla descendente y usar por lo menos parte de la información paramétrica relacionada con los objetos. Los decodificadores de señales de audio comprenden además un procesador de señales de audio configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información 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 ascendente.

[0029] Es una idea clave de la presente invención que se puede obtener un procesamiento eficiente de diferentes tipos de objetos de audio en una estructura de etapas en cascada, que permite la separación de los diferentes tipos de objetos de audio utilizando por lo menos parte de la información paramétrica relacionada con los objetos en un primer paso de procesamiento ejecutado por el separador de objetos, y que da lugar a un procesamiento espacial adicional en un segundo paso de procesamiento ejecutado dependiendo de por lo menos parte de la información paramétrica relacionada con los objetos por el procesador de señales de audio. Se ha descubierto que la extracción de una segunda información de audio, que comprende objetos de audio del segundo tipo de objetos de audio, de una representación de señal de mezcla descendente se puede ejecutar con moderada complejidad incluso si hay un gran número de objetos del segundo tipo de objetos de audio. Además, se ha encontrado que se puede ejecutar un procesamiento espacial de los objetos de audio del segundo tipo de audio de manera eficiente, una vez separada la segunda información de audio de la primera información de audio.

[0030] Por añadidura, se ha encontrado que el algoritmo de procesamiento ejecutado por el separador de objetos para separar la primera información de audio y la segunda información de audio puede ser ejecutado con relativamente poca complejidad si se posterga el procesamiento individual por objetos de los objetos de audio del segundo tipo de objetos de audio al procesador de señales de audio y no se ejecuta al mismo tiempo que la separación de la primera información de audio y la segunda información de audio.

[0031] Por ejemplo, el decodificador de señales de audio puede estar configurado para producir la representación de señal de mezcla ascendente dependiendo de la representación de señal de mezcla descendente, la información paramétrica relacionada con los objetos y una información residual asociada con una subserie de objetos de audio representada por la representación de señal de mezcla descendente. En este caso, el separador de objetos puede estar configurado para descomponer la representación de señal de mezcla descendente para producir la primera información de audio que describe la primera serie de uno o más objetos de audio (por ejemplo, los objetos en primer plano FGO) del primer tipo de objetos de audio con el cual está asociada la información residual y la segunda información de audio que describe la segunda serie de uno o más objetos de audio (por ejemplo, los objetos de fondo BGO) del segundo tipo de objetos de audio con el cual no hay ninguna información residual asociada que depende de la representación de la señal de mezcla descendente para usar por lo menos parte de la información paramétrica relacionada con los objetos y la información residual.

[0032] Esta implementación se basa en el hallazgo de que se puede obtener una separación particularmente precisa entre la primera información de audio que describe la primera serie de objetos de audio del primer tipo de objetos de audio y la segunda información de audio que describe la segunda serie de objetos de audio del segundo tipo de objetos de audio utilizando una información residual además de la información paramétrica relacionada con los objetos. Se ha encontrado que el mero uso de la información paramétrica relacionada con los objetos daría lugar, en muchos casos, a distorsiones, que se pueden reducir significativamente, o incluso eliminar por completo, por el uso de la información residual. La información residual describe, por ejemplo, una distorsión residual, que se estima que ha de quedar si se aísla un objeto de audio del primer tipo de objetos de audio meramente utilizando la información paramétrica relacionada con los objetos. La información residual es calculada, por lo general, por un codificador de señales de audio. Aplicando la información residual, se puede mejorar la separación entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio.

[0033] Esto permite obtener la primera información de audio y la segunda información de audio con una separación particularmente buena entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo

tipo de objetos de audio, lo que, a su vez, permite obtener un procesamiento altamente espacial de los objetos de audio del segundo tipo de objetos de audio al procesar la segunda información de audio en el procesador de señales de audio.

[0034] En una implementación, el separador de objetos puede estar configurado, por lo tanto, para proporcionar la primera información de audio de tal manera que se realcen los objetos de audio del primer tipo de objetos de audio con respecto a los objetos de audio del segundo tipo de objetos de audio de la primera información de audio. El separador de objetos también puede estar configurado para transmitir la segunda información de audio de tal manera que se realcen los objetos de audio del segundo tipo de objetos de audio con respecto a los objetos de audio del primer tipo de objetos de audio de la segunda información de audio.

5

10

15

20

25

30

35

40

45

50

55

[0035] En una implementación, el decodificador de señales de audio puede estar configurado para ejecutar un procesamiento en dos etapas, de tal manera que el procesamiento de la segunda información de audio del procesador de señales de audio se ejecute después de la separación entre la primera información de audio que describe la primera serie de uno o más objetos de audio del primer tipo de objetos de audio y la segunda información de audio que describe la segunda serie de uno o más objetos de audio del segundo tipo de objetos de audio.

[0036] En una implementación, el procesador de señales de audio puede estar configurado para procesar la segunda información de audio dependiendo de la información paramétrica relacionada con los objetos asociada a los objetos de audio del segundo tipo de objetos de audio e independiente de la información paramétrica relacionada con los objetos asociada a los objetos de audio del primer tipo de objetos de audio. En consecuencia, se puede obtener un procesamiento separado de los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio.

[0037] En una implementación, el separador de objetos puede estar configurado para obtener la primera información de audio y la segunda información de audio utilizando una combinación lineal de uno o más canales de mezcla descendente y uno o más canales residuales. En este caso, el separador de objetos puede estar configurado para obtener parámetros de combinación para ejecutar la combinación lineal dependiendo de los parámetros de mezcla descendente asociados con los objetos de audio del primer tipo de objetos de audio y dependiendo de los coeficientes de predicción de canales de los objetos de audio del primer tipo de objetos de audio. La computación de los coeficientes de predicción de canales de los objetos de audio del primer tipo de objetos de audio puede tomar en cuenta, por ejemplo, los objetos de audio del segundo tipo de objetos de audio como único objeto de audio común. En consecuencia, se puede ejecutar un proceso de separación con una complejidad informática suficientemente baja que puede ser, por ejemplo, casi independiente del número de objetos de audio del segundo tipo de objetos de audio.

[0038] En una implementación, el separador de objetos puede estar configurado para aplicar una matriz de renderización a la primera información de audio para mapear las señales de objeto de la primera información de audio sobre los canales de audio de la representación de señal de audio de mezcla ascendente. Esto se puede hacer porque el separador de objetos puede tener la capacidad de extraer señales de audio separadas que individualmente representan los objetos de audio del primer tipo de objetos de audio. En consecuencia, es posible mapear las señales de objeto de la primera información de audio directamente sobre los canales de audio de la representación de señal de audio de mezcla ascendente.

[0039] En una implementación, el procesador de audio puede estar configurado para ejecutar un procesamiento en estéreo de la segunda información de audio que depende de una información de renderización, una información de covarianza relacionada con los objetos y una información de mezcla descendente, para obtener los canales de audio de la representación de señal de audio de mezcla ascendente.

[0040] En consecuencia, el procesamiento en estéreo de los objetos de audio del segundo tipo de objetos de audio puede ser independiente de la separación entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio. De esa manera, no resulta afectada (ni se degrada) la separación eficiente entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio por el procesamiento en estéreo, lo que por lo general da lugar a una distribución de los objetos de audio en toda una pluralidad de canales de audio sin producir un alto grado de separación de objetos, que se puede obtener en el separador de objetos, por ejemplo, empleando la información residual.

[0041] En una implementación, el procesador de audio puede estar configurado para ejecutar un postprocesamiento de la segunda información de audio dependiendo de una información de renderización, una información de covarianza relacionada con los objetos y una información de mezcla descendente. Esta forma de postprocesamiento da lugar a una ubicación espacial de los objetos de audio del segundo tipo de objetos de audio dentro de una escena de audio. No

obstante, debido al concepto de procesamiento en cascada, la complejidad informática del procesador de audio se puede mantener suficientemente baja, puesto que no es necesario que el procesador de audio tome en cuenta la información paramétrica relacionada con los objetos asociada a los objetos de audio del primer tipo de objetos de audio.

5 **[0042]** Además, el procesador de audio puede ejecutar diferentes tipos de procesamiento, como por ejemplo un procesamiento mono a binaural, un procesamiento mono a estéreo, un procesamiento estéreo a binaural o un procesamiento estéreo a estéreo.

10

15

20

25

30

35

40

45

50

55

[0043] En una implementación, el separador de objetos puede estar configurado para tratar objetos de audio del segundo tipo de objetos de audio, a los cuales no está asociada ninguna información residual, en forma de objeto de audio único. Asimismo, el procesador de señales de audio está configurado para tomar en cuenta parámetros de renderización de objetos específicos para ajustar las contribuciones de los objetos del segundo tipo de objetos de audio a la representación de señal de mezcla ascendente. Por consiguiente, el separador de objetos considera los objetos de audio del segundo tipo de objetos de audio como único objeto de audio, lo que reduce significativamente la complejidad del separador de objetos y también permite contar con una información residual distintiva, que es independiente de los parámetros de renderización asociados con los objetos de audio del segundo tipo de objetos de audio.

[0044] En una implementación, el separador de objetos puede estar configurado para obtener un valor de diferencia de nivel común de los objetos correspondiente a una pluralidad de objetos de audio del segundo tipo de objetos de audio. El separador de objetos está configurado para usar el valor de diferencia de nivel común de los objetos para la computación de los coeficientes de predicción de canales. Además, el separador de objetos puede estar configurado para usar los coeficientes de predicción de canales para obtener uno o dos canales de audio que representan la segunda información de audio. Para obtener un valor común de diferencia de nivel de los objetos, los objetos de audio del segundo tipo de objetos de audio pueden ser tratados de manera eficiente como único objeto de audio por el separador de objetos.

[0045] En una implementación, el separador de objetos puede estar configurado para obtener un valor común de diferencia de nivel de los objetos correspondiente a una pluralidad de objetos de audio del segundo tipo de objetos de audio y el separador de objetos puede estar configurado para utilizar el valor común de diferencia de nivel de los objetos correspondiente a la computación de las anotaciones de una matriz de mapeo en modo de energía. El separador de objetos puede estar configurado para usar la matriz de mapeo en modo de energía para obtener dicho uno o más canales de audio que representan la segunda información de audio. Una vez más, el valor común de diferencia de nivel de los objetos permite un tratamiento común informáticamente eficiente de los objetos de audio del segundo tipo de objetos de audio por el separador de objetos.

[0046] En una implementación, el separador de objetos puede estar configurado para obtener selectivamente un valor común de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio del segundo tipo de objetos de audio del segundo tipo de objetos de audio y para ajustar el valor de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio a cero si se halla que hay más o menos de dos objetos de audio del segundo tipo de objetos de audio. El separador de objetos puede estar configurado para usar el valor común de correlación entre objetos asociado a los objetos de audio del segundo tipo de objetos de audio para obtener el uno o más canales de audio que representan la segunda información de audio. Empleando esta estrategia, se aprovecha el valor de correlación entre objetos si se lo puede obtener con gran eficiencia de computación, es decir, si hay dos objetos de audio del segundo tipo de objetos de audio. De lo contrario, sería una gran exigencia de computación la obtención de valores de correlación entre objetos. En consecuencia, se ha encontrado que se obtendría un buen equilibrio en términos de impresión auditiva y complejidad informática la fijación del valor de correlación entre objetos de audio del segundo tipo de objetos de audio a cero si hay más o menos de dos objetos de audio del segundo tipo de objetos.

[0047] En una implementación, el procesador de señales de audio puede estar configurado para renderizar la segunda información de audio dependiendo de (por lo menos parte de) la información paramétrica relacionada con los objetos, para obtener una representación renderizada de los objetos de audio del segundo tipo de objetos de audio en forma de versión procesada de la segunda información de audio. En este caso, la renderización se puede realizar independientemente de los objetos de audio del primer tipo de objetos de audio.

[0048] En una implementación, el procesador de señales de audio puede estar configurado para proporcionar la segunda información de audio de tal manera que la segunda información de audio describa más de dos objetos de audio del segundo tipo de objetos de audio. Las implementaciones permiten un ajuste flexible del número de objetos de audio del segundo tipo de objetos de audio, lo que se ve considerablemente facilitado por la estructura de etapas en cascada del procesamiento.

[0049] En una implementación, el separador de objetos está configurado para obtener, como segunda información de audio, una representación de señal de audio de un canal único o una representación de señal de audio de dos canales que representan más de dos objetos de audio del segundo tipo de objetos de audio. La extracción de uno o dos canales de señales de audio puede ser ejecutada por el separador de objetos con baja complejidad informática. En particular, la complejidad del separador de objetos se puede mantener significativamente más baja en comparación con un caso en el cual el separador de objetos necesitaría tratar más de dos objetos de audio del segundo tipo de objetos de audio. Sin embargo, se ha encontrado que es una representación informáticamente eficiente de los objetos de audio del segundo tipo de objetos de audio el uso de uno o dos canales de una señal de audio.

10

5

[0050] En una implementación, el procesador de señales de audio puede estar configurado para recibir la segunda información de audio y para procesar la segunda información de audio dependiendo de (por lo menos parte de) la información paramétrica relacionada con los objetos, tomando en cuenta información paramétrica relacionada con los objetos asociada a más de dos objetos de audio del segundo tipo de objetos de audio. En consecuencia, el procesador de audio ejecuta un procesamiento de objetos individuales, en tanto que el separador de objetos no lleva a cabo un procesamiento de objetos individuales respecto de los objetos de audio del segundo tipo de objetos de audio.

15

20

[0051] En una implementación, el decodificador de audio puede estar configurado para extraer una información de número total de objetos y una información de número de objetos en primer plano a partir de una información de configuración relacionada con la información paramétrica relacionada con los objetos. El decodificador de audio también puede estar configurado para determinar un número de objetos de audio del segundo tipo de objetos de audio formando una diferencia entre la información del número total de objetos y la información del número de objetos en primer plano. En consecuencia, se obtiene una señalización eficiente del número de objetos de audio del segundo tipo de objetos de audio. Más aun, este concepto ofrece un alto grao de flexibilidad con respecto al número de objetos de audio del segundo tipo de objetos de audio.

25

30

[0052] En una implementación, el separador de objetos puede estar configurado para usar información paramétrica relacionada con los objetos asociada a  $N_{\text{EAO}}$  objetos de audio del primer tipo de objetos de audio para obtener, como primera información de audio,  $N_{\text{EAO}}$  señales de audio que representan (preferentemente en forma individual), los  $N_{\text{EAO}}$  objetos de audio del primer tipo de objetos de audio, y para obtener, como segunda información de audio, una o dos señales de audio que representan los  $N_{\text{NEAO}}$  objetos de audio del segundo tipo de objetos de audio, tratar los  $N_{\text{NEAO}}$  objetos de audio del segundo tipo de audio monocanal o de dos canales. El procesador de señales de audio está configurado para renderizar individualmente los  $N_{\text{NEAO}}$  objetos de audio representados por la una o dos señales de audio de la segunda información de audio usando la información paramétrica relacionada con los objetos asociada a los  $N_{\text{NEAO}}$  objetos de audio del segundo tipo de objetos de audio. En consecuencia, la separación de objetos de audio entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio es independiente del procesamiento posterior de los objetos de audio del segundo tipo de objetos de audio.

35

**[0053]** Unas realizaciones de acuerdo con la invención crean un procedimiento tal como se establece en las reivindicaciones independientes 4 a 6, para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con un objeto.

45

40

[0054] Otra realización de acuerdo con la invención genera un programa de computación para poner en práctica dicho procedimiento, tal como se establece en la reivindicación 7.

Breve Descripción de las Figuras.

50

[0055] A continuación se describen las realizaciones de acuerdo con la invención con referencia a las Figs. adjuntas, en las cuales:

La Fig. 1 ilustra un diagrama esquemático de bloques de un decodificador de señales de audio, de acuerdo con una realización de la invención;

55

La Fig. 2 ilustra un diagrama esquemático de bloques de otro decodificador de señales de audio de acuerdo con una realización de la invención;

Las Figs. 3a y 3b ilustran diagramas esquemáticos de bloques de un procesador residual que se puede utilizar como separador de objetos en una realización de la invención;

- Las Figs. 4a a 4e ilustran diagramas esquemáticos de bloques de procesadores de señales de audio que se pueden utilizar en un decodificador de señales de audio de acuerdo con una realización de la invención:
- 5 La Fig. 4f ilustra un diagrama de bloques de un modo de procesamiento del transcodificador SAOC;
  - La Fig. 4g ilustra un diagrama de bloques de un modo de procesamiento del decodificador SAOC;
- La Fig. 5a ilustra un diagrama esquemático de bloques de un decodificador de señales de audio de acuerdo con una realización de la invención:
  - La Fig. 5b ilustra un diagrama esquemático de bloques de otro decodificador de señales de audio de acuerdo con una realización de la invención;
- 15 La Fig. 6a ilustra una Tabla que representa una descripción del diseño de una prueba de audición;
  - La Fig. 6b ilustra una Tabla que representa los sistemas en estudio;
  - La Fig. 6c ilustra una Tabla que representa los elementos de la prueba de audición y las matrices de renderización;
  - La Fig. 6d ilustra una representación gráfica de puntuaciones MUSHRA correspondientes a la prueba de audición de la renderización del tipo Karaoke/Solo;
- La Fig. 6e ilustra una representación gráfica de puntuaciones MUSHRA correspondientes a una prueba de audición de renderización clásica:
  - La Fig. 7 ilustra un gráfico de flujo de un procedimiento para producir una representación de señal de mezcla ascendente de acuerdo con una realización de la invención;
- 30 La Fig. 8 ilustra un diagrama esquemático de bloques de un sistema MPEG SAOC de referencia;
  - La Fig. 9a ilustra un diagrama esquemático de bloques de un sistema SAOC de referencia que utiliza decodificador y mezclador separados;
- La Fig. 9b ilustra un diagrama esquemático de bloques de un sistema SAOC de referencia que utiliza un decodificador y mezclador integrados y
  - La Fig. 9c ilustra un diagrama esquemático de bloques de un sistema SAOC de referencia que utiliza un transcodificador SAOC a MPEG.
  - Descripción Detallada de las Realizaciones

20

40

50

55

- 1. Decodificador de señales de audio de acuerdo con la Fig. 1
- 45 **[0056]** La Fig. 1 ilustra un diagrama esquemático de bloques de un decodificador de señales de audio 100 de acuerdo con una realización de la invención.
  - [0057] El decodificador de señales de audio 100 está configurado para recibir una información paramétrica relacionada con los objetos 110 y una representación de señal de mezcla descendente 112. El decodificador de señales de audio 100 está configurado para transmitir una representación de señal de mezcla ascendente 120 dependiendo de la representación de la señal de mezcla descendente y la información paramétrica relacionada con los objetos 110. El decodificador de señales de audio 100 comprende un separador de objetos 130, que está configurado para descomponer la representación de señal de mezcla descendente 112 para ofrecer una primera información de audio 132 que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio 134 que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio que depende de la representación de señal de mezcla descendente 112 y utilizar por lo menos parte de la información paramétrica relacionada con los objetos 110. El decodificador de señales de audio 100 comprende asimismo un procesador de señales de audio 140, que está configurado para recibir la segunda información de audio 134 y para procesar la segunda información de audio que depende de por lo menos parte de la información paramétrica relacionada

con los objetos 112, para obtener una versión procesada 142 de la segunda información de audio 134. El decodificador de señales de audio 100 comprende asimismo un combinador de señales 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 ascendente 120.

[0058] El decodificador de señales de audio 100 implementa un procesamiento en etapas en cascada de la representación de señal de mezcla descendente, que representa objetos de audio del primer tipo de objetos de audio y objetos de audio del segundo tipo de objetos de audio de manera combinada.

[0059] En una primera etapa de procesamiento, que es ejecutado por el separador de objetos 130, la segunda información de audio que describe una segunda serie de objetos de audio del segundo tipo de objetos de audio se separa de la primera información de audio 132 que describe una primera serie de objetos de audio de un primer tipo de objetos de audio empleando la información paramétrica relacionada con los objetos 110. Sin embargo, la segunda información de audio 134 es por lo general una información de audio (por ejemplo, una señal de audio de un solo canal o una señal de audio de dos canales) que describe los objetos de audio del segundo tipo de objetos de audio de manera combinada.

**[0060]** En la segunda etapa de procesamiento, el procesador de señales de audio 140 procesa la segunda información de audio 134 que depende de la información paramétrica relacionada con los objetos. En consecuencia, el procesador de señales de audio 140 tiene la capacidad de ejecutar un procesamiento de objetos individuales o de renderizar los objetos de audio del segundo tipo de objetos de audio, lo que se describe en la segunda información de audio 134, y que por lo general no es ejecutado por el separador de objetos 130.

[0061] Por consiguiente, si bien es preferible que el separador de objetos 130 no procese los objetos de audio del segundo tipo de obietos de audio de manera individual. los obietos de audio del segundo tipo de obietos de audio son procesados, por cierto, de manera individual por objeto (por ejemplo, renderizados de manera individual por objeto) en la segunda etapa de procesamiento, que es ejecutada por el procesador de señales de audio 140. Por consiguiente, la separación entre los objetos de audio del primer tipo de objetos de audio y los objetos de audio del segundo tipo de objetos de audio, que es ejecutada por el separador de objetos 130, es independiente del procesamiento individual por objetos de los objetos de audio del segundo tipo de objetos de audio, que es ejecutada posteriormente por el procesador de señales de audio 140. En consecuencia, el procesamiento ejecutado por el separador de objetos 130 es sustancialmente independiente de un número de objetos de audio del segundo tipo de objetos de audio. Además, el formato (por ejemplo, señal de audio de un solo canal o señal de audio de dos canales) de la segunda información de audio 134 depende, por lo general, del número de objetos de audio del segundo tipo de objetos de audio. Por consiguiente, se puede variar el número de objetos de audio del segundo tipo de objetos de audio sin necesidad de modificar la estructura del separador de objetos 130. Dicho de otro modo, los objetos de audio del segundo tipo de obietos de audio son tratados como único obieto de audio (por ejemplo, de un canal o de dos canales) con respecto al cual el separador de objetos 140 obtiene una información paramétrica común relacionada con los objetos (por ejemplo, un valor común de diferencia de nivel de los objetos asociado con uno o dos canales de audio).

[0062] En consecuencia, el decodificador de señales de audio 100 de acuerdo con la Fig. 1 es apto para manejar un número variable de objetos de audio del segundo tipo de objetos de audio sin una modificación estructural del separador de objetos 130. Además, se pueden aplicar diferentes algoritmos de procesamiento de objetos mediante el separador de objetos 130 y el procesador de señales de audio 140. En consecuencia, por ejemplo, es posible realizar una separación de objetos de audio mediante el uso por el separador de objetos 130 de información residual, lo que da lugar a una separación particularmente eficaz de diferentes objetos de audio, haciendo uso de la información residual, que constituye una información complementaria para mejorar la calidad de una separación de objetos. Por el contrario, el procesador de señales de audio 140 puede ejecutar un procesamiento individual por objeto sin emplear una información residual. Por ejemplo, el procesador de señales de audio 140 puede estar configurado para ejecutar un procesamiento de señales de audio del tipo convencional de codificación espacial de objetos de audio (SAOC) para renderizar los diferentes objetos de audio.

2. Decodificador de señales de audio de acuerdo con la Fig. 2

5

20

25

30

35

40

45

50

55

[0063] A continuación se describe un decodificador de señales de audio 200 de acuerdo con una realización de la invención. En la Fig. 2 se ilustra un diagrama esquemático de bloques de este decodificador de señales de audio 200.

[0064] El decodificador de audio 200 está configurado para recibir una señal de mezcla descendente 210, un denominado flujo de bits SAOC 212, información de la matriz de renderización 2 14 y, opcionalmente, parámetros de

transferencia de cabeza (HRTF) 216. El decodificador de señales de audio 200 también está configurado para emitir una señal de mezcla descendente de salida/MPS 220 y (opcionalmente) un flujo de bits MPS 222.

2.1. Señales de entrada y señales de salida del decodificador de señales de audio 200

5

15

20

35

55

[0065] A continuación, se describen diversos detalles con respecto a las señales de entrada y las señales de salida del decodificador de audio 200.

[0066] La señal de mezcla descendente 200 puede ser, por ejemplo, una señal de audio de un solo canal o una señal de audio de dos canales. La señal de mezcla descendente 210 puede derivar, por ejemplo, de una representación codificada de la señal de mezcla descendente.

[0067] El flujo de bits por codificación espacial de objetos de audio (Flujo de bits SAOC) 212 puede comprender, por ejemplo, información paramétrica relacionada con los objetos. Por ejemplo, el flujo de bits SAOC 212 puede comprender información de diferencia de nivel de los objetos, por ejemplo, en forma de parámetros de diferencia de nivel de los objetos OLD, una información de correlación entre objetos, por ejemplo, en forma de parámetros de correlación entre objetos IOC.

[0068] Por otra parte, el flujo de bits SAOC 212 puede comprender una información de mezcla descendente que describe de que manera se han producido las señales de mezcla descendente sobre la base de una pluralidad de señales de objeto de audio utilizando un proceso de mezcla descendente. Por ejemplo, el flujo de bits SAOC puede comprender un parámetro de ganancia de mezcla descendente DMG y (opcionalmente) parámetros de diferencia de nivel de canales de mezcla descendente DCLD.

[0069] La información de matrices de renderización 214 puede describir, por ejemplo, de qué manera se los diferentes objetos de audio han de ser renderizados por el decodificador de audio. Por ejemplo, la información de matrices de renderización 214 puede describir una asignación de un objeto de audio a uno o más canales de la señal de mezcla descendente de salida/MPS 220.

30 **[0070]** La información opcional de parámetros de función de transferencia de cabeza (HRTF) 216 puede describir además una función de transferencia para derivar una señal binaural para auriculares.

[0071] La señal de mezcla descendente de salida/MPEG Envolvente (también abreviada como "señal de mezcla descendente de salida/MPS") 220 representa uno o más canales de audio, por ejemplo, en forma de una representación de una señal de audio en el dominio de tiempo o una representación de una señal de audio en el dominio de la frecuencia. Sola o en combinación con el flujo de bits opcional MPEG-Envolvente (flujo de bits MPS) 222, que comprende parámetros de MPEG-Envolvente que describen un mapeo de la señal de mezcla descendente de salida/MPS 220 sobre una pluralidad de canales de audio, se forma una representación de señal de mezcla ascendente.

40 2.2. Estructura y funcionalidad del decodificador de señales de audio 200

[0072] A continuación se describe en forma más detallada la estructura del decodificador de señales de audio 200, que puede cumplir la función de un transcodificador SAOC o la función de un decodificador SAOC.

[0073] El decodificador de señales de audio 200 comprende un procesador de mezcla descendente 230, que está configurado para recibir la señal de mezcla descendente 210 y para producir, sobre la base de ésta, la señal de mezcla descendente de salida/MPS 220. El procesador de mezcla descendente 230 también está configurado para recibir por lo menos parte de la información de flujo de bits SAOC 212 y por lo menos parte de la información de la matriz de renderización 214. Además, el procesador de mezcla descendente 230 puede recibir asimismo una información sobre parámetros SAOC procesados 240 procedente de un procesador un procesador de parámetros 250.

[0074] El procesador de parámetros 250 está configurado para recibir la información sobre flujo de bits SAOC 212, información sobre matrices de renderización 214 y, opcionalmente, la información de parámetros de función de transferencia de cabeza 260, y para producir, sobre la base de ésta, el flujo de bits MPEG Envolvente 222 que acarrea los parámetros de MPEG envolvente (en caso de que sean necesarios los parámetros de MPEG envolvente, como ocurre, por ejemplo, en el modo de operación de transcodificación). Además, el procesador de parámetros 250 suministra la información SAOC procesada (en caso de ser necesaria esta información de SAOC procesada).

[0075] A continuación se describe en forma más detallada la estructura y funcionalidad del procesador de mezcla descendente 230.

[0076] El procesador de mezcla descendente 230 comprende un procesador residual 260, que está configurado para recibir la señal de mezcla descendente 210 y para producir, sobre la base de ésta, una primera señal de objetos de audio 262 que describe los denominados objetos de audio realzados (EAOs), que se pueden considerar objetos de audio de un primer tipo de objetos de audio. La primera señal de objetos de audio puede comprender uno o más canales de audio y se la puede considerar como primera información de audio. El procesador residual 260 también está configurado para producir una segunda señal de objetos de audio 264, que describe objetos de audio de un segundo tipo de objetos de audio y puede ser considerada segunda información de audio. La segunda señal de objetos de audio 264 puede comprender uno o más canales y por lo general puede comprender uno o dos canales de audio que describen una pluralidad de objetos de audio. Generalmente, la segunda señal de objetos de audio puede describir aun más de dos objetos de audio del segundo tipo de objetos de audio.

5

10

25

30

35

40

55

15 **[0077]** El procesador de mezcla descendente 230 comprende asimismo un preprocesador de mezcla descendente SAOC 270, que está configurado para recibir la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, una versión procesada 272 de la segunda señal de objetos de audio 264, que se puede considerar una versión procesada de la segunda información de audio.

[0078] El procesador de mezcla descendente 230 comprende asimismo un combinador de señales de audio 280, que está configurado para recibir la primera señal de objetos de audio 262 y la versión procesada 272 de la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, la señal de mezcla descendente de salida/MPS 220, que puede ser considerada, sola o junto con el correspondiente flujo de bits MPEG–Envolvente (opcional) 222, una representación de señal de mezcla ascendente.

[0079] A continuación se describe en forma más detallada la funcionalidad de las unidades individuales del procesador de mezcla descendente 230.

[0080] El procesador residual 260 está configurado para aportar, por separado, la primera señal de objetos de audio 262 y la segunda señal de objetos de audio 264. Para este fin, el procesador residual 260 puede estar configurado para aplicar por lo menos parte de la información de flujo de bits SAOC 212. Por ejemplo, el procesador residual 260 puede estar configurado para evaluar una información paramétrica relacionada con los objetos asociada a los objetos de audio del primer tipo de objetos de audio, es decir, los denominados "objetos de audio realzados" EAO. Además, el procesador residual 260 puede estar configurado para obtener una información general que describe comúnmente los objetos de audio del segundo tipo de objetos de audio, por ejemplo, los denominados "objetos de audio no realzados". El procesador residual 260 puede estar configurado asimismo para evaluar una información residual, que está incluida en la información de flujo de bits SAOC 212, para la separación entre objetos de audio realzados (objetos de audio del primer tipo de objetos de audio) y objetos de audio no realzados (objetos de audio del segundo tipo de objetos de audio). La información residual puede codificar, por ejemplo, una señal residual en el dominio del tiempo, que se aplica para obtener una separación particularmente limpia entre los objetos de audio realzados y los objetos de audio no realzados. Además, el procesador residual 260 puede evaluar, opcionalmente, por lo menos una parte de la información de matrices de renderización 214, por ejemplo, a fin de determinar la distribución de los objetos de audio realzados por los canales de audio de la primera señal de objetos de audio 262.

[0081] El preprocesador de mezcla descendente SAOC 270 comprende un redistribuidor de canales 274, que está configurado para recibir el uno o más canales de audio de la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, uno o más canales de audio (típicamente dos) de la segunda señal de objetos de audio procesada 272. Además, el preprocesador de mezcla descendente SAOC 270 comprende un productor de señales descorrelacionadas 276, que está configurado para recibir el uno o más canales de audio de la segunda señal de objetos de audio 264 y para producir, sobre la base de ésta, uno o más señales descorrelacionadas 278a, 278b, que se agregan a las señales provistas por el re—distribuidor de canales 274 a fin de obtener la versión procesada 272 de la segunda señal de objetos de audio 264.

[0082] Más adelante se explican más detalles con respecto al procesador SAOC de mezcla descendente.

[0083] El combinador de señales de audio 280 combina la primera señal de objetos de audio 262 con la versión procesada 272 de la segunda señal de objetos de audio. Para este fin, se puede realizar una combinación por canales. De manera acorde se obtiene la señal de mezcla descendente de salida/MPS 220.

[0084] El procesador de parámetros 250 está configurado para obtener los parámetros (opcionales) de MPEG-Envolvente, que conforman el flujo de bits MPEG-Envolvente 222 de la representación de señal de mezcla ascendente, sobre la base del flujo de bits SAOC, tomando en cuenta la información de matrices de renderización 214 y, opcionalmente, la información de parámetros HRTF 21 6. Dicho de otro modo, el procesador de parámetros SAOC 252 está configurado para traducir la información de parámetros relacionados con los objetos, que es la descripta por la información de flujo de bits SAOC 212, en una información paramétrica relacionada con los canales, descripta por el flujo de bits MPEG Envolvente 222.

5

10

15

20

25

30

50

55

[0085] A continuación se presenta una breve reseña general de la estructura de la arquitectura del transcodificador/decodificador SAOC expuesta en la Fig. 2. La codificación espacial de objetos de audio (SAOC) es una técnica paramétrica de codificación de múltiples objetos. Está destinada a transmitir un número de objetos de audio en una señal de audio (por ejemplo la señal de audio de mezcla descendente 210) que comprende M canales. Junto con esta señal de mezcla descendente inversa compatible, se transmiten parámetros de objetos (por ejemplo, utilizando la información de flujo de bits SAOC 212) que dan lugar a la recreación y manipulación de las señales de objeto originales. Un codificador SAOC (que no se ilustra aquí) produce una mezcla descendente de las señales de objeto en su entrada y extrae estos parámetros de objeto. El número de objetos que se puede manejar no está limitado en principio. Los parámetros de objeto son cuantificados y codificados eficientemente para integrar el flujo de bits SAOC 212. La señal de mezcla descendente 210 puede ser comprimida y transmitida sin necesidad de actualizar los codificadores e infraestructuras existentes. Los parámetros de objetos, o información complementaria SAOC, se transmiten por un canal secundario de baja velocidad de transmisión de bits, por ejemplo, la porción de datos auxiliares del flujo de bits de mezcla descendente.

[0086] Del lado del decodificador, los objetos de entrada son reconstruidos y renderizados a un número de canales de reproducción. La información de renderización que contiene el nivel de reproducción y la posición de paneo por cada objeto es provista por el usuario o se la puede extraer del flujo de bits SAOC (por ejemplo, como información preestablecida). La información de renderización puede variar temporalmente. Las configuraciones de salida pueden variar de mono a multicanal (por ejemplo, 5.1) y son independientemente tanto del número de objetos de entrada como del número de canales de mezcla descendente. La renderización binaural de objetos es posible, incluyendo el azimut y la elevación de las posiciones virtuales de los objetos. Una interfaz de efectos opcionales da lugar a la manipulación avanzada de señales de objeto, aparte de la modificación de nivel y paneo.

[0087] Los objetos en sí pueden ser señales mono, señales estereofónicas, como así también señales multicanales (por ejemplo 5.1 canales). Las configuraciones típicas de mezcla descendente son mono y estéreo.

35 [0088] A continuación, se pasa a explicar la estructura básica del transcodificador/decodificador SAOC, que se ilustra en la Fig. 2. El módulo de transcodificador/decodificador SAOC aquí descripto puede actuar como decodificador autoportante o como transcodificador de un flujo de bits SAOC a MPEG-envolvente, dependiendo de la configuración pretendida de los canales de salida. 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 40 modo de decodificador y la salida del módulo SAOC es una salida de modulación por pulsos codificados (salida PCM). En el primer caso, no es necesario un decodificador MPEG envolvente. Por el contrario, la representación de señal de mezcla ascendente puede comprender sólo la señal de salida 220, y la vez se puede omitir la provisión del flujo de bits MPEG envolvente 222. En un segundo caso, la configuración de la señal de salida es una configuración multicanal con más de dos canales de salida. El módulo SAOC puede ser operativo en un modo de transcodificador. La salida del 45 módulo SAOC puede comprender tanto una señal de mezcla descendente 220 como un flujo de bits MPEG envolvente 222 en este caso, como se ilustra en la Fig. 2. En consecuencia, se necesita un decodificador MPEG envolvente para obtener una representación final de la señal de audio para la salida de los parlantes.

[0089] La Fig. 2 ilustra la estructura básica de la arquitectura del transcodificador/decodificador SAOC. El procesador residual 216 extrae el objeto de audio realzado de la señal de mezcla descendente de entrada 210 usando la información residual contenida en el flujo de bits SAOC 212. El preprocesador de mezcla descendente 270 procesa los objetos de audio normales (que son, por ejemplo, objetos de audio no realzados, es decir, objetos de audio con respecto a los cuales no se transmite información en el flujo de bits SAOC 212). Los objetos de audio realzados (representados por la primera señal de objetos de audio 262) y los objetos de audio normales procesados (representados, por ejemplo, por la versión procesada 272 de la segunda señal de objetos de audio 264) se combinan en la señal de salida 220 correspondiente al modo de decodificador SAOC o a la señal de mezcla descendente MPEG envolvente 220 correspondiente al modo de transcodificador SAOC. A continuación se presentan descripciones detalladas de los bloques de procesamiento.

### 3. Arquitectura y funcionalidad del Procesador Residual y del Procesador de Modo de Energía

[0090] A continuación, se describen detalles referentes a un procesador residual que puede asumir, por ejemplo, la funcionalidad del separador de objetos 130 del decodificador de señales de audio 100 o del procesador residual 260 del decodificador de señales de audio 200. Para este fin, las Figs. 3a y 3b ilustran diagramas esquemáticos de boques de ese tipo de procesador residual 300, que puede tomar el lugar del separador de objetos 130 o del procesador residual 260. La Fig. 3a presenta menos detalles que la Fig. 3b. Sin embargo, la siguiente descripción se aplica al procesador residual 300 de acuerdo con la Fig. 3a y también al procesador residual 380 de acuerdo con la Fig. 3b.

- [0091] El procesador residual 300 está configurado para recibir una señal de mezcla descendente SAOC 3 10, que puede ser equivalente a la representación de señal de mezcla descendente 112 de la Fig. 1 o la representación de señal de mezcla descendente 210 de la Fig. 2. El procesador residual 300 está configurado para proporcionar, sobre la base de ésta, una primera información de audio 320 que describe uno o más objetos de audio realzados, que pueden ser equivalentes, por ejemplo, a la primera información de audio 132 o a la primera señal de objetos de audio 262. Además, el procesador residual 300 puede suministrar una segunda información de audio 322 que describe uno o más objetos de audio adicionales (por ejemplo, objetos de audio no realzados, con respecto a los cuales no se dispone de información residual), donde la segunda información de audio 322 puede ser equivalente a la segunda información de audio 134 o a la segunda señal de objetos de audio 264.
- [0092] El procesador residual 300 comprende una unidad 1-a-N/2-a-N (unidad OTN/TTN) 330, que recibe la señal de mezcla descendente SAOC 310 y que también recibe datos SAOC y residuales 332. La unidad 1-a-N/2-a-N 330 también emite una señal de objetos de audio realzados 334, que describe los objetos de audio realzados (EAO) contenidos en la señal de mezcla descendente SAOC 310. Asimismo, la unidad 1-a-N/2-a-N 330 proporciona la segunda información de audio 322. El procesador residual 300 comprende asimismo una unidad de renderización 340, que recibe la señal de objetos de audio realzados 334 e información de matrices de renderización 342 y produce, sobre la base de ésta, la primera información de audio 320.
  - [0093] A continuación, se describe en forma más detallada el procesamiento de objetos de audio realzados (procesamiento de EAO), que es ejecutado por el procesador residual 300.

# 3.1. Introducción a la Operación del procesador residual 300

5

30

35

45

50

55

[0094] Con respecto a la funcionalidad del procesador residual 300, cabe señalar que la tecnología SAOC permite la manipulación individual de un número de objetos de audio en términos de su amplificación/atenuación de nivel sin desmedro significativo de la calidad del sonido obtenido como resultado, sólo de manera limitada. Una configuración especial de la aplicación del tipo "karaoke " requiere una supresión total (o casi total) de los objetos específicos, por lo general la primera voz, manteniendo indemne la calidad perceptual del sonido de fondo.

[0095] Un caso típico de aplicación contiene hasta cuatro señales de objetos de audio realzados (EAO), que pueden representar, por ejemplo, dos objetos estéreo independientes (por ejemplo, dos objetos estéreo independientes preparados para ser retirados del lado del decodificador).

[0096] Se debe tener en cuenta que los (uno o más) objetos de audio de calidad realzada (o, más precisamente, las contribuciones a la señal de audio asociadas a los objetos de audio realzados) están incluidos en la señal de mezcla descendente SAOC 310. Por lo general, las contribuciones de las señales de audio asociadas a los (uno o más) objetos de audio realzados son mezcladas por el procesamiento de mezcla descendente ejecutado por el codificador de señales de audio, con las contribuciones de señales de audio de otros objetos de audio, que no son objetos de audio realzados. Además, también se debe notar que las contribuciones de señales de audio de una pluralidad de objetos de audio realzados también se superponen o mezclan, por lo general, en el procesamiento de mezcla descendente ejecutado por el codificador de señales de audio.

## 3.2 Arquitectura SOAC que sirve como soporte para los Objetos de audio realzados

[0097] A continuación, se describen detalles con respecto al procesador residual 300. El procesamiento de objetos de audio realzados incorpora las unidades 1–a–N o 2–a–N, dependiendo del modo de mezcla descendente SAOC. La unidad de procesamiento 1–a–N está dedicada a la señal de mezcla descendente mono y la unidad de procesamiento 2–a–N está dedicada a una señal de mezcla descendente estéreo 310. Estas dos unidades representan una modificación generalizada y realzada de la caja 2–a–2 (caja TTT) que fuera dada a conocer por ISO/IEC 23003–1:2007. En el codificador, se combinan señales regulares y EAO en la mezcla descendente. Las unidades de procesamiento

OTN<sup>-1</sup>–/TTN<sup>-1</sup> (que son unidades de procesamiento uno a N inversas o unidades de procesamiento 2 a N inversas) son utilizadas para producir y codificar las correspondientes señales residuales.

[0098] Las señales EAO y regulares son recuperadas de la mezcla descendente 310 por las unidades OTN/TTN 330 empleando la información complementaria SAOC y las señales residuales incorporadas. Los EAOs recuperados (descriptos por las señales de objetos de audio realzados 334) son alimentados a la unidad de renderización 340 que representa (o crea el producto de la matriz de renderización correspondiente (descripta por la información sobre matrices de renderización 342) y la salida obtenida como resultado de la unidad OTN/TTN. Los objetos de audio normales (descriptos por la segunda información de audio 322) son transmitidos al preprocesador de mezcla descendente SAOC, por ejemplo, el preprocesador de mezcla descendente SAOC 270, para continuar su procesamiento. Las Figs. 3a y 3b ilustran la estructura general del procesador residual, es decir, la arquitectura del procesador residual.

[0099] Las señales de salida del procesador residual 320, 322 se computan de la siguiente manera:

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

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

15

5

10

[0100] donde  $X_{obj}$  representa la señal de mezcla descendente de los objetos de audio normales (es decir, los no-EAOs) y  $X_{EAO}$  es la señal de salida EAO renderizada para el modo de decodificador SAOC o la correspondiente señal de mezcla descendente de EAO para el modo de transcodificador SAOC.

20

[0101] El procesador residual puede operar en el modo de predicción (utilizando la información residual) o en el modo de energía (sin información residual). La señal de entrada ampliada X<sub>res</sub> se define de manera consiguiente:

$$\mathbf{X}_{res} = \begin{cases} \left(\frac{\mathbf{X}}{res}\right), & \text{para el modo de predicción} \\ \mathbf{X}, & \text{para el modo de energía} \end{cases}$$

25

**[0102]** En este caso, **X** puede representar, por ejemplo, el uno o más canales de la representación de señal de mezcla descendente 310, que pueden ser transportados en el flujo de bits que representa el contenido de audio multicanal. **res** puede designar una más señales residuales, que pueden ser descriptas por el flujo de bits que representan el contenido de audio de múltiples canales.

30

[0103] El procesamiento OTN/TTN está representado por la matriz M y el procesador EAO por la matriz A<sub>EAO</sub>.

[0104] La matriz de procesamiento OTN/TTN M se define de acuerdo con el modo de operación de EAO (es decir, el de predicción o energía) de la siguiente manera:

predicción o energía) de la siguiente manera: 35

$$\mathbf{M} = egin{cases} \mathbf{M}_{ ext{predicción}} \;,\;\; ext{en el caso del modo de predicción} \ \mathbf{M}_{ ext{energía}} \;,\;\;\; ext{en el caso del modo de energía} \end{cases}$$

[0105] La matriz de procesamiento OTN/TTN M está representada por

$$\mathbf{M} = \left(\frac{\mathbf{M}_{OBJ}}{\mathbf{M}_{EAO}}\right),$$

40

donde la matriz N se relaciona con los objetos de audio normales (es decir, los no EAOs) y  $M_{EAO}$  con los objetos de audio realzados (EAOs).

[0106] En algunas realizaciones, uno o más objetos de fondo multicanales (MBO) pueden ser tratados de igual manera por el procesador residual 300.

- [0107] Un Objeto de Fondo Multicanal (MBO) es una mezcla descendente mono o estéreo MPS que forma parte de la mezcla descendente SAOC. A diferencia del uso de objetos SAOC individuales para cada canal en una señal multicanal, se puede utilizar un MBO que habilite a SAOC para manejar con más eficiencia un objeto multicanal. En el caso del MBO, la sobrecarga de SAOC se reduce a medida que los parámetros SAOC de los MBOs sólo se relacionan con los canales de mezcla descendente en lugar de hacerlo con todos los canales de mezcla ascendente.
  - 3.3 Otras Definiciones

20

35

40

- 3.3.1 Dimensionalidad de las Señales y Parámetros
- 15 **[0108]** A continuación, se describe brevemente la dimensionalidad de las señales y parámetros a fin de arrojar más claridad sobre la frecuencia con que se realizan los diferentes cálculos.
  - [0109] Las señales de audio se definen por cada franja horaria n y cada subbanda híbrida (que puede ser una subbanda de frecuencia) k. Los correspondientes parámetros SAOC se definen por cada franja horaria de parámetros 1 y la banda de procesamiento m. Un mapeo subsiguiente entre el dominio híbrido y de los parámetros está especificado por la tabla A. 31 ISO/IEC 23003–1:2007. Por ende, todos los cálculos se realizan con respecto a ciertos índices de banda horaria y las correspondientes dimensionalidades están implícitas por cada variable introducida.
- [0110] Sin embargo, a continuación se omiten, en ocasiones, los índices de bandas de tiempo y frecuencia para mantener la notación concisa.
  - 3.3.2 Cálculo de la matriz A<sub>EAO</sub>,
- [0111] La matriz de pre-renderización de EAO A<sub>EAO</sub> se define de acuerdo con el número de canales de salida (es decir, mono, estéreo o binaural) de la siguiente manera:

$$\mathbf{A}_{EAO} = egin{cases} \mathbf{A}_1^{EAO}, & ext{en el caso de mono} \ \mathbf{A}_2^{EAO}, & ext{para otros casos} \end{cases}$$
 .

[0112] Las matrices  $A_1^{EAO}$  del tamaño 1 x  $N_{EAO}$  y  $A_2^{EAO}$  del tamaño 2x  $N_{EAO}$  se definen de la siguiente manera:

$$\mathbf{A}_{1}^{\textit{EAO}} = \mathbf{D}_{16}^{\textit{EAO}} \mathbf{M}_{\textit{ren}}^{\textit{EAO}} \;, \qquad \quad \mathbf{D}_{16}^{\textit{EAO}} = \begin{pmatrix} w_{1}^{\textit{EAO}} & w_{2}^{\textit{EAO}} & w_{3}^{\textit{EAO}} & w_{1}^{\textit{EAO}} & w_{2}^{\textit{EAO}} \end{pmatrix},$$

$$\mathbf{A}_{2}^{EAO} = \mathbf{D}_{26}^{EAO} \mathbf{M}_{ren}^{EAO} \,, \qquad \mathbf{D}_{26}^{EAO} = \begin{pmatrix} w_{1}^{EAO} & 0 & \frac{w_{3}^{EAO}}{\sqrt{2}} & \frac{w_{2}^{EAO}}{\sqrt{2}} & w_{1}^{EAO} & 0 \\ 0 & w_{2}^{EAO} & \frac{w_{3}^{EAO}}{\sqrt{2}} & \frac{w_{2}^{EAO}}{\sqrt{2}} & 0 & w_{2}^{EAO} \end{pmatrix},$$

donde la submatriz de renderización  $M_{res}^{EAO}$  corresponde a la renderización EAO (y describe un mapeo conveniente de los objetos de audio realzados sobre los canales de la representación de señal de mezcla ascendente).

- [0113] Los valores W<sub>1</sub><sup>EAO</sup> se computan dependiendo de la información de renderización asociada a los objetos de audio realzados usando los correspondientes elementos de EAO y utilizando las ecuaciones de la sección 4.2.2.1.
- [0114] En el caso de la renderización binaural la matriz  $A_2^{EAO}$  está definida por las ecuaciones presentadas en la sección 4.1.2, para la cual la matriz binaural objetivo correspondiente contiene sólo elementos relacionados con los EAO.
  - 3.4 Cálculo de los Elementos OTN/TTN en el Modo Residual

[0115] A continuación, se describe cómo la señal de mezcla descendente SAOC 310, que por lo general comprende uno o dos canales de audio, se mapea sobre la señal de objetos de audio realzados 334, que típicamente comprende uno o más canales de objetos de audio realzados, y la segunda información de audio 322, que por lo general comprende uno o dos canales de objetos de audio normales.

5

[0116] La funcionalidad de la unidad 1-a-N o la unidad 2-a-N 330 se puede implementar, por ejemplo, utilizando una multiplicación de vectores de matriz, de tal manera que se obtiene un vector que describe tanto los canales de la señal de objetos de audio realzados 334 y los canales de la segunda información de audio 322 multiplicando un vector que describe los canales de la señal de mezcla descendente SAOC 3 y (opcionalmente) una o más señales residuales con una matriz M<sub>Predicción</sub> o M<sub>Energía</sub>. En consecuencia, la determinación de la matriz M<sub>Predicción</sub> o M<sub>Energía</sub> es un paso importante en la derivación de la primera información de audio 320 y la segunda información de audio 322 de la mezcla descendente SAOC 310.

15

10

[0117] Para resumir, el proceso de mezcla descendente OTN/TTN se presenta en una matriz en el modo M<sub>Predicción</sub> en el caso del modo de predicción o M<sub>Energía</sub> en el caso del modo de energía.

20

[0118] El procedimiento de codificación/decodificación basado en la energía está destinado a la codificación de preservación no de forma de onda de la señal de mezcla descendente. Por consiguiente, la matriz de mezcla ascendente OTN/TTN para el correspondiente modo de energía no se basa en formas de onda específicas, pero sólo describen la distribución relativa de energía de los objetos de audio de entrada, como se describe más adelante en forma más detallada.

(

3.4.1 Modo de predicción

25

[0119] En el caso del modo de predicción, la matriz M<sub>predicción</sub> se define utilizando la información de mezcla descendente contenida en la matriz D<sup>-1</sup> y los datos y los datos de CPC de la matriz C:

30

[0120] Con respecto a los varios modos SAOC, la matriz de mezcla descendente ampliada y la matriz CPC C exhiben las siguientes dimensiones y estructuras:

[0121] 3.4.1.1 Los modos de mezcla descendente estéreo (TTN): En el caso de los modos de mezcla descendente estéreo (TTN) (por ejemplo, en el caso de una mezcla descendente estéreo sobre la base de dos canales de objetos de

35

audio normales y  $N_{\text{EAO}}$  canales de objetos de audio realzados), la matriz de mezcla descendente (ampliada)  $\mathbf{p}$  y la matriz de CPC  $\mathbf{c}$  se puede obtener de la siguiente manera:

$$\tilde{\mathbf{D}} = \begin{pmatrix} 1 & 0 & m_0 & \dots & m_{N_{EAO}-1} \\ 0 & 1 & n_0 & \dots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & -1 & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & 0 & \dots & -1 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{pmatrix}.$$

[0122] En el caso de una mezcla descendente estéreo, cada EAO j tiene dos CPCs, c j.o y c j.o, para dar la matriz C.

[0123] Las señales de salida del procesador residual se computan de la siguiente manera:

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \begin{bmatrix} l_0 \\ -\frac{r_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{bmatrix},$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Predicción}} \begin{pmatrix} l_{o} \\ \underline{r_{o}} \\ res_{o} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}.$$

5

10

[0124] En consecuencia, se obtienen dos señales, y<sub>L</sub> e y<sub>R</sub> (que están representadas por X<sub>obi</sub>), que representan uno o dos, o incluso más de dos objetos de audio normales (también denominados objetos de audio no ampliados)). Asimismo, las señales, N<sub>EAO</sub> (representadas por X<sub>EAO</sub>) que representan los N<sub>EAO</sub> objetos de audio realzados. Estas señales se obtienen sobre la base de dos señales de mezcla descendente SAOC señal de mezcla descendentes I<sub>0</sub>r<sub>0</sub> y N<sub>EAO</sub> o de señales residuales res<sub>0</sub> a resN<sub>EAO-1</sub>, que son codificadas en la Información complementaria SAOC, por ejemplo, como parte de la información paramétrica relacionada con los objetos.

15

[0125] Se debe tener en cuenta que las señales y<sub>L</sub> e y<sub>R</sub> pueden ser equivalentes a la señal 322, y que las señales y<sub>0</sub> e yN<sub>EAO-1</sub>, EAO (que están representadas por X<sub>EAO</sub>) pueden ser equivalentes a las señales 320.

20

[0126] La matriz de A<sup>EAO</sup> es una matriz de renderización. Los asientos de la matriz pueden describir, por ejemplo, un mapeo de objetos de audio realzados con los canales de las señales de objetos de audio realzados 334 (X<sub>EAO</sub>).

[0127] En consecuencia, una elección apropiada de la matriz de A<sup>EAO</sup> puede dar lugar a una integración opcional de la funcionalidad de la unidad de renderización 340, de tal manera que la multiplicación del vector que describe los canales (I<sub>0</sub>,r<sub>o</sub>) de la señal de mezcla descendente SAOC 310 y una o más señales residuales (res<sub>0</sub> y resN<sub>EAO-1</sub>) con la matriz A<sup>EAO</sup>M<sup>Predicción</sup>

~puede dar origen a una representación X<sub>EAO</sub> de la primera información de audio 320.

30

25

3.4.1.2 Modos de mezcla descendente mono (OTN):

101281 A continuación, se describe la derivación de las señales de objetos de audio realzados 320 (o. de lo contrario, de las señales de objetos de audio realzados 334) y de la señal de objetos de audio normales 322 con respecto al caso en que la señal de mezcla descendente SAOC 310 comprende sólo un canal de señal.

[0129] En el caso de los modos de mezcla descendente mono (OTN) (por ej., una mezcla descendente mono basada en un canal de objetos de audio normales y canales de objetos de audio realzados NEAO), se puede obtener la matriz de

mezcla descendente (ampliada)  $\tilde{\mathbf{D}}$  y la matriz de CPC  $\mathbf{C}$  de la siguiente manera:

$$\tilde{\mathbf{D}} = \begin{pmatrix} \frac{1}{m_0} & \frac{1}{m_0} & \frac{1}{m_{N_{EAO}-1}} \\ \frac{1}{m_0} & -1 & \frac{1}{m_0} \\ \vdots & 0 & \ddots & \vdots \\ \frac{1}{m_{N_{EAO}-1}} & 0 & \dots & -1 \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} -\frac{1}{c_{0,0}} & \frac{1}{1} & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ c_{N_{EAO}-1,0} & 0 & \dots & 1 \end{pmatrix}.$$

**[0130]** Con una mezcla descendente mono mezcla descendente, un EAO *j* es previsto sólo por un coeficiente *c j* para dar la matriz C. Todos los elementos de matriz *c j* se obtienen, por ejemplo, de los parámetros SAOC (por ejemplo, de los datos SAOC 322 de acuerdo con las relaciones provistas más adelante en la sección 3.4.1.4

[0131] Las señales de salida del procesador residual se computan de la siguiente manera:

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

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

5

10

15

**[0132]** La señal de salida X<sub>OBJ</sub> comprende, por ejemplo, un canal que describe los objetos de audio normales (objetos de audio no realzados). La señal de salida X<sub>EAO</sub> comprende, por ejemplo, uno, dos o incluso más canales que describen los objetos de audio realzados (preferentemente canales N<sub>EAO</sub> que describen los objetos de audio realzados). Una vez más, dichas señales son equivalentes a las señales 320, 322.

3.4.1.3 Cálculo de la matriz de mezcla descendente ampliada inversa

[0133] La matriz  $\tilde{\mathbf{D}}^{-1}$  es la inversa de la matriz de mezcla descendente ampliada  $\tilde{\mathbf{D}}$  y C implica los CPCs.

[0134] La matriz  $\tilde{\mathbf{D}}^{-1}$  es la inversa de la matriz de mezcla descendente ampliada  $\mathbf{D}$  y se la puede calcular de la siguiente manera:

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

[0135] Los elementos  $\tilde{d}_{i,j}$  (por ejemplo, de la  $\tilde{\mathbf{D}}^{-1}$  inversa de la matriz de mezcla descendente ampliada  $\tilde{\mathbf{D}}$  de un tamaño de 6 x 6, se derivan utilizando los siguientes valores:

**[0136]** Los coeficientes m, y *n*, de la matriz de mezcla descendente ampliada descendente por cada EAO *j* correspondiente a los canales de mezcla descendente derecho e izquierdo según

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

5

10

[0137] Los elementos d<sub>ij</sub> de la matriz de mezcla descendente be se obtienen utilizando la información de ganancia de mezcla descendente DMG y la información (opcional) de diferencias de nivel de los canales de mezcla descendente DCLD, que está incluida en la información de SAOC 332, que está representada, por ejemplo, por la información paramétrica relacionada con los objetos 110 o por la información de flujo de bits SAOC 212.

**[0138]** En el caso de la mezcla descendente estéreo la matriz de mezcla descendente **D** del tamaño 2 x N con los elementos  $d_{i,j}$  (i = 0,1; j = 0,..., N-1) se obtiene de los parámetros DMG y DCLD, a saber

$$d_{\text{0},j} = 10^{0.05\,\text{DMG}_{\text{J}}} \sqrt{\frac{10^{0.1D\,\text{CLD}_{\text{J}}}}{1+10^{0.1D\,\text{CLD}_{\text{J}}}}}} \,, \qquad \qquad d_{\text{1},j} = 10^{0.05\,\text{DMG}_{\text{J}}} \sqrt{\frac{1}{1+10^{0.1D\,\text{CLD}_{\text{J}}}}} \,.$$

**[0139]** En el caso de la mezcla descendente mono mezcla descendente la matriz de mezcla descendente **D** del tamaño 1 x N con los elementos  $d_{i,j}$  (j = 0,1; j = 0,..., N-1) se obtiene de los parámetros DMG de la siguiente manera:

$$d_{0,j} = 10^{0,05 DMG j}$$

[0140] En este caso, se obtienen los parámetros de mezcla descendente DMGj y DCLDj descuantificados, por ejemplo, de la información parametrica complementaria 110 o del flujo de bits SAOC 2 12.

[0141] La función EAO(j) determina el mapeo entre los índices de canales de objeto de audio de entrada y las señales de EAO:

$$EAO_{j} = N - 1 - j$$
  $j = 0,...,N_{EAO} - 1$ 

3.4.1.4 Cálculo de la matriz C

15

20

35

40

[0142] La matriz C incluye los CPCs y se deriva de los parámetros SAOC transmitidos (es decir, los OLDs, IOCs, DMGs y DCLDs) de la siguiente manera:

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

[0143] Dicho de otro modo, se obtienen los CPCs de restricción de acuerdo con las ecuaciones anteriores, que se pueden considerar algoritmos de constricción. Sin embargo, los CPCs de restricción también se pueden derivar de los valores  $\widetilde{c}_{j,0}$ ,  $\widetilde{c}_{j,1}$  utilizando una estrategia de limitación diferente (algoritmo de restricción) o pueden ser fijados en valores iguales a  $\widetilde{c}_{j,0}$ ,  $\widetilde{c}_{j,1}$ .

[0144] Se ha de notar que las anotaciones de la matriz c j,1 (y las cantidades intermedias sobre la base de las cuales se computan las anotaciones de la matriz c j,1), sólo son necesarias por lo general si la señal de mezcla descendente es una señal de mezcla descendente estéreo.

[0145] Los CPCs son restringidos por las siguientes funciones limitantes:

$$\gamma_{j,1} = \frac{m_{j}OLD_{L} + n_{j}e_{L,R} - \sum_{i=0}^{N_{EAO}-1} m_{i}e_{i,j}}{2\left(OLD_{L} + \sum_{i=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} m_{i}m_{k}e_{i,k}\right)}, \qquad \gamma_{j,2} = \frac{n_{j}OLD_{R} + m_{j}e_{L,R} - \sum_{i=0}^{N_{EAO}-1} n_{i}e_{i,j}}{2\left(OLD_{R} + \sum_{i=0}^{N_{EAO}-1} \sum_{k=0}^{N_{EAO}-1} n_{i}n_{k}e_{i,k}\right)},$$

donde el factor  $\lambda$  se determina de la siguiente manera:

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

**[0146]** En el caso de un canal EAO específico  $j = 0...N_{EAO} - 1$  los CPCs sin restricción son calculados por

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

45 [0147] Las cantidades de energía  $P_{lo}$ ,  $P_{ro}$ ,  $P_{LoRo}$ ,  $P_{LoCo}$ ,  $P_{LoRo,j}$ ,  $P_{LoCo,j}$ , se computan como

$$\begin{split} P_{Lo} &= OLD_L + \sum_{j=0}^{N_{EMO}-1} \sum_{k=0}^{N_{EMO}-1} m_j m_k e_{j,k} \;, \\ P_{Ro} &= OLD_R + \sum_{j=0}^{N_{EMO}-1} \sum_{k=0}^{N_{EMO}-1} n_j n_k e_{j,k} \;, \\ \\ P_{LoRo} &= e_{L,R} + \sum_{j=0}^{N_{EMO}-1} \sum_{k=0}^{N_{EMO}-1} m_j n_k e_{j,k} \;, \\ \\ P_{LoCo,j} &= m_j OLD_L + n_j e_{L,R} - m_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EMO}-1} m_i e_{i,j} \;, \\ \\ P_{RoCo,j} &= n_j OLD_R + m_j e_{L,R} - n_j OLD_j - \sum_{\substack{i=0 \\ i \neq j}}^{N_{EMO}-1} n_i e_{i,j} \;. \end{split}$$

5

**[0148]** La matriz de covarianza  $e_{l,j}$ , se define de la siguiente manera: La matriz de covarianza E del tamaño  $N \times N$  con los elementos  $e_{l,j}$ , representa una aproximación de la matriz de covarianza de señales original E  $\approx$  **SS**\* y se la obtiene de los parámetros OLD y LOC de la siguiente manera

10

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

**[0149]** En este caso se obtienen los parámetros de objetos descuantificados OLD<sub>i</sub>, LOC<sub>i,j</sub>, por ejemplo de la información paramétrica complementaria 110 o del flujo de bits SAOC 212.

15

[0150] Además, se puede obtener  $e_{L,R}$  de la siguiente manera

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

20

[0151] Los parámetros *OLD<sub>L</sub>*, *OLD<sub>R</sub>* y *IOC<sub>L,R</sub>* corresponden a los objetos (de audio) normales y se los puede derivar utilizando la información de mezcla descendente:

$$\begin{split} OLD_{L} &= \sum_{i=0}^{N-N_{EAO}-1} d_{0,i}^{2}OLD_{i} \;, \\ OLD_{R} &= \sum_{i=0}^{N-N_{EAO}-1} d_{1,i}^{2}OLD_{i} \;, \\ IOC_{L,R} &= \begin{cases} IOC_{0,1}, & N-N_{EAO} = 2, \\ 0, & \text{de lo contrario}. \end{cases} \end{split}$$

30

25

[0152] Como se puede ver, se computan dos valores diferentes de niveles de objeto OLD<sub>L</sub> y OLD<sub>R</sub> correspondientes a los objetos de audio normales en el caso de una señal de mezcla descendente estéreo (que implica preferentemente una señal de objetos de audio normales de dos canales). Por el contrario, sólo se computa un valor de nivel de objeto diferente OLD<sub>L</sub> para los objetos de audio normales en el caso de una señal de mezcla descendente (mono) monocanal (que preferentemente conlleva una señal de objetos de audio normales de de un canal).

[0153] Como se puede apreciar, se obtiene el primer (en el caso de una señal de mezcla descendente de dos canales o el único (en el caso de una señal de mezcla descendente de un solo canal) valor común de diferencia de niveles de los objetos OLD<sub>L</sub> sumando las contribuciones de los objetos de audio normales que tienen un índice (o índices) de objetos de audio i correspondiente al canal izquierdo (o el único canal) de la señal de mezcla descendente SAOC 310.

35

**[0154]** El segundo valor común de diferencia de niveles de los objetos  $OLD_R$  (que se utiliza en el caso de una señal de mezcla descendente de dos canales) se obtiene sumando las contribuciones de los objetos de audio normales que tienen un índice (o índices) de objetos de audio i correspondiente al canal derecho de la señal de mezcla descendente SAOC 310.

5

[0155] La contribución  $OLD_L$  de los objetos de audio normales (que tienen los índices de objetos de audio i =  $N-N_{EAO}-1$ ) sobre la señal del canal izquierdo (o la señal del único canal) de la señal de mezcla descendente SAOC 710 se computa, por ejemplo, tomando en cuenta la ganancia de mezcla descendente ganancia de mezcla descendente  $d_{0,i}$ , que describe la ganancia de mezcla descendente aplicada al objeto de audio normal que tiene el índice de objeto de audio i al obtener la señal del canal izquierdo de la señal de mezcla descendente SAOC 310, y también el nivel de objeto del objeto de audio normal que tiene el objeto de audio i, que está representado por el valor  $OLD_i$ .

15

10

[0156] De manera similar, el valor común de diferencia de nivel de los objetos OLD<sub>R</sub> se obtiene usando los coeficientes de mezcla descendente d<sub>i,i</sub>, que describen la ganancia de mezcla descendente que se aplica al objeto de audio normal que tiene el índice de objeto de audio i al formar la señal del canal izquierdo de la señal de mezcla descendente SAOC 310, y la información de nivel OLD<sub>i</sub> asociada al objeto de audio normal que tiene el índice de objeto de audio i.

20

**[0157]** Como se puede ver, las ecuaciones para el cálculo de las cantidades  $P_{lo}$ ,  $P_{ro}$ ,  $P_{LoRo}$ ,  $P_{LoRo}$ ,  $P_{LoRo, j}$  y  $P_{LoCo, j}$ , no distinguen entre los objetos de audio normales individuales, sino meramente hacen uso de los valores comunes de diferencia de nivel de los objetos OLD<sub>L</sub>, OLD<sub>R</sub>, considerando así los objetos de audio normales (que tienen los índices de objeto de audio i) como un solo objeto de audio.

25

[0158] Además, se establece el valor de correlación entre objetos IOC<sub>L,R</sub> que está asociado a los objetos de audio normales, de 0 a menos que haya dos objetos de audio normales.

۷:

[0159] La matriz de covarianza  $e_{i,j}$  (y  $e_{L,R}$ ) se define de la siguiente manera:

**[0160]** La matriz de covarianza  $\mathbf{E}$  del tamaño  $\mathbf{N}$  x  $\mathbf{N}$  con los elementos  $\mathbf{e}_{i,j}$  representan una aproximación de la matriz de covarianza de señal original  $\mathbf{E} \approx \mathbf{S}\mathbf{S}^*$  es obtenida de los parámetros OLD y IOC según

30

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

[0161] Por ejemplo

35

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

donde OLD<sub>L</sub> y OLD<sub>R</sub> y IOC<sub>L,R</sub> se computan de la manera antes descripta.

40

[0162] En este caso, los parámetros de objeto descuantificados se obtienen como

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

45

donde  $\mathbf{D}_{\text{OLD}}$  y  $\mathbf{D}_{\text{IOC}}$  son matrices que comprenden parámetros de diferencias de nivel de los objetos y parámetros de correlación entre objetos.

<del>4</del>0

3.4.2. Modo de Energía

50

**[0163]** A continuación se describe otro concepto que se puede utilizar para separar las señales de objeto de audio ampliadas 320 y las señales de objetos de audio normales (objeto de audio no ampliado) 322 y que se puede utilizar en combinación con una codificación de audio sin conservación de la forma de onda de los canales SAOC de mezcla descendente 3 10.

55

[0164] Dicho de otro modo, el procedimiento de codificación/decodificación basado en la energía está destinado a la codificación de audio sin conservación de la forma de onda de la señal de mezcla descendente. Por consiguiente, la matriz de mezcla ascendente OTN/TTN para el correspondiente modo de energía no se basa en formas de onda específicas, sino que sólo describe la distribución relativa de energía de los objetos de audio de entrada.

**[0165]** Además, el concepto aquí explicado, que se designa concepto de "modo de energía", se puede utilizar sin transmitir una información de señal residual. Una vez más, los objetos de audio normales (objetos de audio no realzados) son tratados en forma de objetos de audio de uno o dos canales que tienen uno o dos valores comunes de diferencia de nivel de los objetos, OLD<sub>L</sub>, OLD<sub>R</sub>.

[0166] En el caso del modo de energía, la matriz M<sub>energía</sub> se define utilizando la información de mezcla descendente y los OLDs, como se describe a continuación.

10 3.4.2.1. Modo de energía correspondiente a los Modos de Mezcla Descendente Estéreo (TTN)

[0167] En el caso de un estéreo (por ejemplo, una mezcla descendente estéreo basado en dos canales de objetos de audio normales y  $N_{EAO}$  canales de objetos de audio realzados), las matrices  $\mathbf{M}_{OBJ}^{Energia}$  y  $\mathbf{M}_{EAO}^{Energia}$  se obtienen de los OLDs correspondientes de acuerdo con

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

$$\mathbf{M}_{EAO}^{\text{Energia}} = \begin{pmatrix} \sqrt{\frac{m_0^2 O L D_0}{O L D_L} + \sum_{i=0}^{N_{EAO}-1} m_i^2 O L D_i} & \sqrt{\frac{n_0^2 O L D_0}{O L D_R} + \sum_{i=0}^{N_{EAO}-1} n_i^2 O L D_i} \\ \vdots & \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^2 O L D_{N_{EAO}-1}}{O L D_L} + \sum_{i=0}^{N_{EAO}-1} m_i^2 O L D_i} & \sqrt{\frac{n_{N_{EAO}-1}^2 O L D_{N_{EAO}-1}}{O L D_R + \sum_{i=0}^{N_{EAO}-1} n_i^2 O L D_i}} \end{pmatrix}$$

15

5

10

[0168] El procesador residual señales de salida se computa de la siguiente manera:

$$\begin{split} \mathbf{X}_{OBJ} &= \mathbf{M}_{OBJ}^{\text{Energia}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}, \\ \mathbf{X}_{EAO} &= \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energia}} \begin{pmatrix} l_0 \\ r_0 \end{pmatrix}. \end{split}$$

20

**[0169]** Lasa señales  $y_L$  e  $y_R$ , que están representadas por la señal  $X_{OBJ}$ , describen los objetos de audio normales (y pueden ser equivalentes a la señal 322), y las señales  $y_{0,EAO}$  a  $y_{NEAO-1,EAO}$ , que están descriptas por la señal  $X_{EAO}$ , describen los objetos de audio realzados (y pueden ser equivalentes a la señal 334 o a la señal 320).

**[0170]** Si se pretende una señal de mezcla ascendente mono para el caso de una señal de mezcla descendente estéreo, se puede efectuar un procesamiento de 2a 1, por ejemplo lo puede realizar el preprocesador 270 basándose en la señal bicanal X<sub>OBJ</sub>.

5 3.4.2.2. Modo de energía correspondiente a Modos de Mezcla Descendente Mono (OTN)

[0171] En el caso mono (por ejemplo, una mezcla descendente mono sobre la base de un canal de objetos de audio  $\mathbf{M}_{OBJ}^{\text{Energia}}$  y  $\mathbf{M}_{EAO}^{\text{Energia}}$  se obtienen a partir de los correspondientes OLDs de acuerdo con

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

$$\mathbf{M}_{EAO}^{Energia} = \begin{pmatrix} \frac{m_{0}^{2}OLD_{0}}{OLD_{L} + \sum_{i=0}^{N_{EAO}-1} m_{i}^{2}OLD_{i}} \\ \vdots \\ \frac{m_{N_{EAO}-1}^{2}OLD_{N_{EAO}-1}}{OLD_{L} + \sum_{i=0}^{N_{EAO}-1} m_{i}^{2}OLD_{i}} \end{pmatrix}$$

[01721] El procesador residual señales de salida se computa de la siguiente manera

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{Energia} \left( d_{\theta} \right),$$

$$\mathbf{X}_{\mathit{EAO}} = \mathbf{A}^{\mathit{EAO}} \mathbf{M}_{\mathit{EAO}}^{\mathsf{Energia}} \left(d_{o}\right).$$

15

10

[0173] Se puede obtener un canal único de objetos de audio normales 322 (representado por  $X_{OBJ}$ ) y  $N_{EAO}$  canales de objetos de audio realzados 320 (representados por  $X_{EAO}$ ) aplicando las matrices  $\mathbf{M}_{OBJ}^{Energia}$  y  $\mathbf{M}_{EAO}^{Energia}$  a una representación de una señal de mezcla descendente SAOC de canal único 310 (representado aquí por  $d_0$ ).

20

**[0174]** Si se desea una señal de mezcla ascendente de dos canales (estéreo) para el caso de una señal de mezcla descendente de un canal (mono), se puede ejecutar un procesamiento de 1 a 2, por ejemplo mediante el preprocesador 270, sobre la base de una señal monocanal  $X_{OBJ}$ .

25 <u>4. Arquitectura y operación del preprocesador de mezcla descendente SAOC</u>

**[0175]** A continuación se describe el funcionamiento del preprocesador de mezcla descendente SAOC 270 tanto con respecto a algunos modos de decodificación como en el caso de los modos de operación de transcodificación.

30 <u>4.1 Operación en los Modos de Decodificación</u>

# 4.1.1 Introducción

[0176] A continuación se describe un procedimiento para obtener una señal de salida utilizando parámetros SAOC e información de paneo (o información de renderización) asociada con cada objeto de audio. El decodificador SAOC 495 está ilustrado en la Fig. 4g y consiste en el procesador de parámetros SAOC 496 y el procesador de mezcla descendente 497.

[0177] Se debe notar que se puede utilizar el decodificador SAOC 494 para procesar los objetos de audio normales, y por lo tanto recibir, como señal de mezcla descendente 497a, la segunda señal de objetos de audio 264 o la señal de objetos de audio normales 322 o la segunda información de audio 134. En consecuencia, el procesador de mezcla descendente 497 puede producir, como sus señales de salida 497b, la versión procesada 272 de la segunda señal de objetos de audio 264 o la versión procesada 142 de la segunda información de audio 134. En consecuencia, el procesador de mezcla descendente 497 puede asumir el rol del preprocesador de mezcla descendente SAOC 270, o el rol del procesador de señales de audio 140.

15 **[0178]** El procesador de parámetros SAOC 496 puede asumir la función del procesador de parámetros SAOC 252 y, en consecuencia, proporciona la información de mezcla descendente 496a.

#### 4.1.2 Procesador de mezcla descendente

5

10

25

30

35

40

45

50

[0179] A continuación se describe en forma más detallada el procesador de mezcla descendente, que es parte del procesador de señales de audio 140, y que ha sido designado "preprocesador de mezcla descendente SAOC" 270 en la realización de la Fig. 2, y que ha sido designada 497 en el decodificador SAOC 495.

[0180] En el caso del modo de decodificador del sistema SAOC, la señal de salida 142, 272, 497b del procesador de mezcla descendente (representado en el dominio híbrido QMF) es alimentada al correspondiente banco de filtros de síntesis (no se ilustra en las Figs. 1 y 2) de acuerdo con lo descripto en ISO/IEC 23003–1: 2007 para dar la señal PCM de salida final. Sin embargo, la señal de salida 142, 272, 497b del procesador de mezcla descendente se combina por lo general con una o más señales de audio 132, 262 que representan los objetos de audio realzados. Esta combinación se puede realizar antes del banco de filtro de síntesis correspondiente (de tal manera que se transmita una señal combinada que combina la salida del procesador de mezcla descendente y dichas una o más señales que representan los objetos de audio realzados como entrada al banco de filtros de síntesis). Por otro lado, se puede combinar la salida del procesador de mezcla descendente con una o más señales de audio que representan los objetos de audio realzados sólo después del procesamiento en el banco de filtros de. En consecuencia, la representación de señal de mezcla ascendente 120, 220 puede ser una representación en el dominio QMF o una representación en el dominio PCM (o cualquier otra representación apropiada). El procesamiento de mezcla descendente incorpora, por ejemplo, el procesamiento mono, el procesamiento en estéreo y, si se necesita, el posterior procesamiento binaural.

[0181] La señal de salida  $\hat{\mathbf{X}}$  del procesador de mezcla descendente 270, 497 (también designada con los números 142, 272, 497b) se computa a partir de la señal de mezcla descendente mono X (también indicada con 134, 264, 497a) y la señal de mezcla descendente mono descorrelacionada  $\mathbf{X}_d$  según

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{X} + \mathbf{P}_2\mathbf{X}_d.$$

[0182] La señal de mezcla descendente mono descorrelacionada  $X_d$  se computa de la siguiente manera

$$X_d$$
 = Func. descor ( $\mathbf{X}$ ).

[0183] Las señales descorrelacionadas  $X_d$  se generan a partir de descorrelacionador descripto en ISO/IEC 23003–1:2007, subcláusula 6.6.2. Siguiendo este esquema, se debe emplear la configuración bsDecorrConfig == 0 con un índice de descorrelación, X = 8, de acuerdo con la Tabla A.26 a la Tabla A.29 de ISO/IEC 23003–1:2007. Por ende, la decorrFunc() denota el proceso de descorrelación:

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

[0184] En el caso de la salida binaural se aplican los parámetros de mezcla ascendente G y  $P_2$  derivados de los datos de SAOC, la información de renderización  $M_{\text{ren}}^{l,m}$  y los parámetros HRTF a la señal de mezcla descendente X (y  $X_d$ ) para

dar origen a la salida binaural  $\hat{\mathbf{X}}$ , ver la Fig. 2, número de referencia 270, donde se ilustra la estructura básica del procesador de mezcla descendente.

[0185] La matriz de renderización binaural objetivo  $A^{1,m}$  del tamaño 2xN consiste en los elementos  $a^{l,m}_{x,y}$ . Cada

elemento  $a_{x,y}^{l,m}$ . deriva de los parámetros HRTF y la matriz de renderización  $\mathbf{M}_{\text{ren}}^{l,m}$  con los elementos  $\mathbf{p}_{y,i}^{l,m}$ , por ejemplo, por el procesador de parámetros SAOC. La matriz de renderización binaural objetivo  $\mathbf{A}^{l,m}$  representa la relación entre todos los objetos de audio de entrada y la salida binaural pretendida.

$$a_{y,1}^{l,m} = \sum_{i=0}^{M_{\rm HRTF}-1} m_{y,i}^{l,m} H_{i,L}^{m} \exp\left(j\frac{\phi_{i}^{m}}{2}\right), \qquad a_{y,2}^{l,m} = \sum_{i=0}^{M_{\rm IRTF}-1} m_{y,i}^{l,m} H_{i,R}^{m} \exp\left(-j\frac{\phi_{i}^{m}}{2}\right).$$

**[0186]** Los parámetros HRTF están dados por  $H_{i,L}^m$ ,  $H_{i,R}^m$ ,  $H_{i,R}^m$  por cada banda de procesamiento m. Las posiciones espaciales para las cuales se dispone de parámetros HRTF se caracterizan por el índice i. Estos parámetros han sido descriptos en ISO/IEC 23003–1:2007.

## 4.1.2.1 Reseña general

5

10

15

20

35

[0187] A continuación, se presenta una reseña general con respecto al procesamiento de mezcla descendente con referencia a las Figs. 4a y 4b, que ilustran una representación de bloques del procesamiento de mezcla descendente, que puede ser ejecutado por el procesador de señales de audio 140 o por la combinación de procesador de parámetros SAOC 252 y el preprocesador de mezcla descendente SAOC 270, o por la combinación del procesador de parámetros SAOC 496 y el procesador de mezcla descendente 497.

[0188] Haciendo referencia, ahora, a la Fig. 4a, el procesamiento de mezcla descendente recibe una matriz de renderización M, una información sobre diferencias de nivel de los objetos OLD, una información sobre correlación entre objetos IOC, una información de ganancia de mezcla descendente DMG y (opcionalmente) una información sobre diferencias de nivel de canales de mezcla descendente DCLD. El procesamiento de mezcla descendente 400 de acuerdo con la Fig. 4a obtiene una matriz de renderización A sobre la base de la matriz de renderización M, por ejemplo, usando un ajustador de parámetros y un mapeo de M con A. Además, las anotaciones de una matriz de covarianza E se obtienen dependiendo de la información sobre diferencias de nivel de los objetos OLD y la información sobre correlación entre objetos IOC, por ejemplo, de acuerdo conde acuerdo con lo descripto anteriormente. De manera similar, las anotaciones de una matriz de mezcla descendente D se obtienen dependiendo de la información sobre ganancia de mezcla descendente DMG y la información sobre diferencias de nivel de los canales de mezcla descendente DCLD.

**[0189]** Las anotaciones f de una matriz de covarianza conveniente F se obtienen dependiendo de la matriz de renderización A y la matriz de covarianza E. Asimismo, se obtiene un valor escalar v que depende de la matriz de covarianza E y la matriz de mezcla descendente D (o dependiendo de las anotaciones de las mismas).

40 [0190] Los valores de ganancia P<sub>L</sub> y P<sub>R</sub> correspondientes a dos canales se obtienen dependiendo de las anotaciones de la matriz de covarianza pretendida F y el valor escalar v. Además, se obtiene un valor de diferencias de fases entre canales φ<sub>C</sub> dependiendo de las anotaciones f de la matriz de covarianza pretendida F. También se obtiene un ángulo de rotación α que depende de las anotaciones f de la matriz de covarianza pretendida F, tomando en cuenta, por ejemplo, una constante c. Por añadidura, se obtiene un segundo ángulo de rotación β, dependiendo de las ganancias de canales P<sub>L</sub> y P<sub>R</sub> c y del primer ángulo de rotación α. Las anotaciones de una matriz G se obtienen, por ejemplo, dependiendo de los dos valores de ganancia de los canales P<sub>L</sub>, P<sub>R</sub> y dependiendo asimismo de la diferencia de fases entre canales φ<sub>C</sub> y, opcionalmente, de los ángulos de rotación α, β. De manera similar, las anotaciones de una matriz P<sub>2</sub> se determinan dependiendo de algunos o todos dichos valores P<sub>L</sub>, P<sub>R</sub>, φ<sub>C</sub>, α, β.

50 **[0191]** A continuación se describe la manera en que se puede obtener la matriz **G** y/o **P2** (o las anotaciones de las mismas), que se puede aplicar con los diferentes modos de procesamiento.

# 4.1.2.2 Modo de Procesamiento Mono a Binaural "x-1-b"

5

10

15

25

[0192] A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por las señales de mezcla descendente monocanales 134, 264, 322, 497a y en las cuales se busca la renderización binaural.

[0193] Los parámetros de mezcla ascendente G<sup>1,m</sup> y P<sub>2</sub><sup>1,m</sup> se computan de la siguiente manera

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

$$\mathbf{P}_{2}^{l,m} = \begin{pmatrix} P_{L}^{l,m} \exp\left(j\frac{\phi_{C}^{l,m}}{2}\right) \sin\left(\beta^{l,m} + \alpha^{l,m}\right) \\ P_{R}^{l,m} \exp\left(-j\frac{\phi_{C}^{l,m}}{2}\right) \sin\left(\beta^{l,m} - \alpha^{l,m}\right) \end{pmatrix}.$$

[0194] Las ganancias  $P_L^{l,m}$  y  $P_R^{l,m}$  correspondientes a los canales de salida derecho e izquierdo son

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

[0195] La matriz de covarianza pretendida F $^{1,m}$  del tamaño 2x2 con los elementos  $f_{i,i}^{l,m}$  se expresan como

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

20 **[0196]** El escalar v<sup>1,m</sup> se computa de la siguiente manera

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

[0197] La diferencia de fases entre canales  $\phi_C^{l,m}$  se expresa como

$$\phi_C^{l,m} = \begin{cases} \arg\left(f_{1,2}^{l,m}\right), & 0 \le m \le 11, \\ 0, & \text{de otro modo} \end{cases} \quad \rho_C^{l,m} \ge 0.6,$$

[0198] La coherencia entre canales  $ho_{c}^{i,m}$  se computa según

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

[0199] Los ángulos de rotación  $\alpha^{1,m}$  y  $\beta^{1,m}$  se expresan como

$$\alpha^{l,m} = \begin{cases} \frac{1}{2}\arccos\left(\rho_C^{l,m}\cos\left(\arg\left(f_{1,2}^{l,m}\right)\right)\right), & 0 \le m \le 11, \quad \rho_C^{l,m} < 0.6, \\ & \frac{1}{2}\arccos\left(\rho_C^{l,m}\right), & \text{de otro modo.} \end{cases}$$

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

4.1.2.3 Modo de Procesamiento Mono a Estéreo "x-1-2"

5

10

15

20

25

30

[0200] A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por la señal monocanal 134, 264, 222, y en los cuales se desea una renderización estéreo.

[0201] En el caso de la salida estéreo se puede aplicar el modo de procesamiento "x-1-b" sin utilizar la información HRTF. Esto se puede efectuar derivando todos los elementos  $a_{x,y}^{l,m}$  de la matriz de renderización **A**, para dar:

$$a_{1,y}^{l,m} = m_{Lf,y}^{l,m}, \qquad a_{2,y}^{l,m} = m_{Rf,y}^{l,m}.$$

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

[0202] A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por un canal de señal 134, 264, 322, 497a y en el cual se busca una renderización de dos canales de los objetos de audio normales.

[0203] En el caso de la salida mono, se puede aplicar el modo de procesamiento "x-1-2" con las siguientes anotaciones:

 $a_{1,\nu}^{l,m}=m_{C,\nu}^{l,m}, \quad a_{2,\nu}^{l,m}=0$ 

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

[0204] A continuación, se describe un modo de procesamiento en el cual los objetos de audio normales están representados por una señal de canales 134, 264, 322, 497a, y en el cual se pretende una renderización binaural de los objetos de audio normales.

[0205] Los parámetros de mezcla ascendente  $\mathbf{G}^{1,m}$  y  $\mathbf{P}_{2}^{1,m}$  se computan de la siguiente manera:

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

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

$$P_{L}^{l,m,x}, P_{R}^{l,m,x}$$
 y  $P_{L}^{l,m}, P_{R}^{l,m}$ 

5 [0206] Las ganancias correspondientes

a los canales de salida izquierdo y derecho son

$$\begin{split} P_L^{l,m,x} &= \sqrt{\max\left(\frac{f_{1,1}^{l,m,x}}{v^{l,m,x}}, \varepsilon^2\right)}, \quad P_R^{l,m,x} &= \sqrt{\max\left(\frac{f_{2,2}^{l,m,x}}{v^{l,m,x}}, \varepsilon^2\right)} \\ P_L^{l,m} &= \sqrt{\max\left(\frac{c_{1,1}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}, \qquad P_R^{l,m} &= \sqrt{\max\left(\frac{c_{2,2}^{l,m}}{v^{l,m}}, \varepsilon^2\right)}. \end{split}$$

[0207] La matriz de covarianza pretendida F  $^{\text{1},m,x}$  del tamaño 2x2 con los elementos

10 se expresan de la siguiente manera

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

15

 $C_{u,v}^{l,m}$  [0208] La matriz de covarianza  $C^{l,m}$  del tamaño 2 x 2 con los elementos de la señal binaural "seca" se estima de la siguiente manera

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

20

donde

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

[0219] Los correspondientes escalares  $v^{1,m,x}$  y  $v^{1,m}$  se computan según

5

10

15

20

25

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

[0210] La matriz de mezcla descendente  $D^{l,x}$  del tamaño 1 x N con los elementos  $d_1^{l,m}$  se puede encontrar como

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

[0211] La matriz de mezcla descendente estéreo  $\mathbf{D}^1$  del tamaño 2 x N con los elementos  $\mathbf{d}^1_{x,j}$  se pueden encontrar como  $\mathbf{d}^1_{x,j} = \mathbf{d}^{l,x}_{l}$ .

[0212] La matriz  $E^{l,m,x}$  con los elementos  $e^{l,m,x}_l$  se pueden obtener según

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

[0213] Las diferencias de fase entre canales  $\phi_C^{l,m}$  se expresan de la siguiente manera:

$$\phi^{l,m,x} = \begin{cases} \arg\left(f_{1,2}^{l,m,x}\right), & 0 \le m \le 11, \\ 0, & \text{de otro modo.} \end{cases} \qquad \rho_C^{l,m} > 0.6,$$

[0214] Los ICCs  $\rho_C^{l,m}$  y  $\rho_T^{l,m}$  se computan como

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

[0215] Los ángulos de rotación  $\alpha^{l,m}$  y  $\beta^{l,m}$  se expresan como

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

### 4.1.2.6 Modo de procesamiento Estéreo a estéreo "x-2-2"

[0216] A continuación se describe un modo de procesamiento en el cual los objetos de audio normales son descriptos por una señal de dos canales (estéreo) 134, 264, 322, 497a y en el cual se pretende una renderización de 2 canales (estéreo).

[0217] En el caso de la salida estéreo, se aplica directamente el preprocesamiento estéreo, que se describe más adelante en la Sección 4.2.2.3.

# 10 <u>4.1.2.7 Modo de procesamiento Estéreo a mono "x-2-1"</u>

[0218] A continuación se describe un modo de procesamiento en el cual los objetos de audio normales están representados por una señal de dos canales (estéreo) 134, 264, 322, 497a y en el cual se pretende una renderización de un canal (mono).

[0219] En el caso de la salida mono, se aplica el preprocesamiento estéreo con una sola anotación activa de la matriz de renderización, como se describe más adelante en la Sección 4.2.2.3.

#### 4.1.2.8 Conclusión

20 <u>4.1.2.8 Conciu</u>

5

15

25

40

45

50

55

[0220] Tomando como referencia, una vez más, las Figs. 4a y 4b, se describe un procesamiento que se puede aplicar a una señal de 1 canal o una señal de dos canales 134, 264, 322,497a que representa los objetos de audio normales con posterioridad a la separación entre los objetos de audio ampliados y los objetos de audio normales. Las Figs. 4a y 4b ilustran el procesamiento, donde el procesamiento de las Figs. 4a y 4b difiere en que se introduce un ajuste opcional de los parámetros en diferentes etapas del procesamiento.

# 4.2. Operación en los modos de transcodificación

# 4.2.1 Introducción

30 <u>4.2.1 introducció</u>

[0221] A continuación se explica un procedimiento para combinar parámetros SAOC e información de paneo (o información de renderización) asociada a cada objeto de audio (o, preferentemente, a cada objeto de audio normal) en un flujo de bits MPEG envolvente compatible standard (flujo de bits MPS).

[0222] El transcodificador SAOC 490 está ilustrado en la Fig. 4f y consiste en un procesador de parámetros SAOC 491 y un procesador de mezcla descendente 492 aplicados a una mezcla descendente estéreo.

[0223] El transcodificador SAOC 490 puede asumir, por ejemplo, la funcionalidad del procesador de señales de audio 140. Por otro lado, el transcodificador SAOC 490 puede asumir la funcionalidad del preprocesador de mezcla descendente SAOC 270 tomado en combinación con el procesador de parámetros SAOC 252.

[0224] Por ejemplo, el procesador de parámetros SAOC 491 puede recibir un flujo de bits SAOC 491a, que es equivalente a la información paramétrica relacionada con los objetos 110 o al flujo de bits SAOC 212. Además, el procesador de parámetros SAOC 491 puede recibir una información de matriz de renderización 491 que puede estar incluida en la información paramétrica relacionada con los objetos 110, o que puede ser equivalente a la información de matriz de renderización 214. El procesador de parámetros SAOC 491 puede producir asimismo la información de procesamiento de mezcla descendente 491c al procesador de mezcla descendente 492, que puede ser equivalente a la información 240. Más aun, el procesador de parámetros SAOC 491 puede producir un flujo de bits MPEG envolvente (o un flujo de bits de parámetros MPEG envolvente) 491d, que comprende una información paramétrica de envolvente que es compatible con la norma sobre MPEG envolvente. El flujo de bits MPEG envolvente 491d puede ser parte, por ejemplo, de la versión procesada 142 de la segunda información de audio, o puede ser parte o tomar el lugar, por ejemplo, del flujo de bits MPS 222.

[0225] El procesador de mezcla descendente 492 está configurado para recibir una señal de mezcla descendente 492a, que es preferentemente una señal de mezcla descendente de un canal o una señal de mezcla descendente de dos canales, y que preferentemente es equivalente a la segunda información de audio 134, o a la segunda señal de objetos de audio 264, 322. El procesador de mezcla descendente 492 puede producir asimismo una señal de mezcla descendente MPEG envolvente 492b, que es equivalente (o forma parte de) la versión procesada 142 de la segunda

información de audio 134, o es equivalente (o es parte de) la versión procesada 272 de la segunda señal de objetos de audio 264.

[0226] Sin embargo, hay diferentes maneras de combinar la señal de mezcla descendente MPEG envolvente 492b con la señal de objetos de audio realzados 132, 262. La combinación se puede realizar en el dominio MPEG envolvente.

[0227] Por otro lado, no obstante, la representación MPEG envolvente, que comprende el flujo de bits de parámetros MPEG envolvente 491d y la señal de mezcla descendente MPEG envolvente 492b, de los objetos de audio normales se puede convertir de vuelta a una representación en el domino del tiempo multicanal o una representación multicanal en el dominio de la frecuencia (que individualmente representan diferentes canales de audio) por un decodificador MPEG envolvente y seguidamente se la puede combinar con las señales de objetos de audio realzados.

[0228] Se ha de notar que los modos de transcodificación comprenden tanto uno o más modos de procesamiento de mezcla descendente mono como uno o más modos de procesamiento de mezcla descendente estéreo. Sin embargo, a continuación sólo se describe el modo de procesamiento de mezcla descendente estéreo, puesto que el procesamiento de las señales de objetos de audio normales es más elaborado en el modo de procesamiento de mezcla descendente estéreo.

4.2.2 Procesamiento de mezcla descendente; en el modo de procesamiento de mezcla descendente estéreo ("x-2-5")

4.2.2.1 Introducción

5

10

15

20

25

30

35

40

45

50

55

[0229] En la siguiente sección se presenta una descripción del modo de transcodificación SAOC correspondiente al caso de la mezcla descendente estéreo.

[0230] Los parámetros de objetos (diferencia de nivel de los objetos OLD, correlación entre objetos IOC, la ganancia de mezcla descendente DMG y la diferencias de nivel de canales de mezcla descendente DCMD) del flujo de bits SAOC se transcodifican en parámetros espaciales (preferentemente relacionados con los canales) (diferencias de nivel de los canales CLD, correlación entre canales ICC, coeficiente de predicción de canales CPC) correspondientes al flujo de bits MPEG envolvente de acuerdo con la información de renderización. La mezcla descendente se modifica de acuerdo con los parámetros de objetos y una matriz de renderización.

[0231] Haciendo referencia, ahora, a las Figs. 4c, 4d y 4e, se presenta una reseña general del procesamiento, y en particular de la modificación de la mezcla descendente.

[0232] La Fig. 4c ilustra una representación en bloques de un procesamiento que se ejecuta para modificar la señal de mezcla descendente, por ejemplo la señal de mezcla descendente 134, 264, 322,492a que describe uno o, preferentemente, más objetos de audio normales. Como se puede apreciar en las Figs. 4c, 4d y 4e, el procesamiento recibe una matriz de renderización M<sub>ren</sub>, una información de ganancia de mezcla descendente DMG, una información de diferencias de nivel de los canales de mezcla descendente DCLD, una información de diferencias de nivel de los objetos OLD, y una información de correlación entre objetos IOC. La matriz de renderización puede ser modificada opcionalmente mediante un ajuste de parámetros, como se ilustra en la Fig. 4c. Las anotaciones de una matriz de mezcla descendente D se obtienen dependiendo de la información de ganancia de mezcla descendente DMG y la información de diferencias de nivel de los canales de mezcla descendente DCLD. Las anotaciones de una matriz de coherencia E se obtienen dependiendo de la información de diferencias de nivel de los objetos OLD y la información de correlación entre objetos IOC. Además, se puede obtener una matriz J dependiendo de la matriz de mezcla descendente D y de la matriz de coherencia E, o supeditada a las anotaciones de las mismas. Seguidamente se puede obtener una matriz C<sub>3</sub> dependiendo de la matriz de renderización M<sub>ren</sub>, la matriz de mezcla descendente D, la matriz de coherencia E y la matriz J. Se puede obtener una matriz G dependiendo de una matriz D<sub>TTT</sub>, que puede ser una matriz que consta de anotaciones predeterminadas, y que también depende de la matriz C3. La matriz G puede ser modificada, opcionalmente, para obtener una matriz modificada  $G_{mod}$ . Se puede utilizar la matriz G o la versión modificada  $G_{mod}$  de la misma para derivar 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 (donde la segunda información de audio 134, 264 se designa X y donde la

versión procesada 142,272 está indicada con X.

[0233] A continuación se describe la renderización de la energía de los objetos, que es ejecutada a fin de obtener los parámetros MPEG envolvente. Además, el preprocesamiento estéreo, que es ejecutado a fin de obtener la versión procesada 142, 272,492b de la segunda información de audio 134,264,492a que representa los objetos de audio normales.

### 4.2.2.2 Renderización de energías de los objetos

[0234] El transcodficador determina los parámetros para el decodificador MPS de acuerdo con la renderización objetivo según lo descripto por la matriz de renderización M<sub>ren</sub>. La covarianza objetivo de seis canales está indicada con F y dada por

$$F = YY^* = M_{ren}S(M_{ren}S)^* = M_{ren}(SS^*)M_{ren}^+ = M_{ren}EM_{ren}^+$$

[0235] El proceso de transcodificación se puede dividir conceptualmente en dos partes. En una parte, se ejecuta una renderización de tres canales al canal izquierdo, derecho y central. En esta etapa, se obtienen los parámetros correspondientes a la modificación de la mezcla descendente, como así también los parámetros de predicción correspondientes a la caja TTT para el decodificador MPS. En la otra parte se determinan los parámetros CLD e ICC para la renderización entre los canales anterior y envolvente (parámetros OTT, anterior izquierdo –envolvente izquierdo, anterior derecho –envolvente derecho).

## 4.2.2.2.1 Renderización al canal izquierdo, derecho y central

[0236] En esta etapa se determinan los parámetros espaciales que controlan la renderización a los canales izquierdo y derecho, que consisten en las señales anterior y envolvente. Estos parámetros describen la matriz de predicción de la caja TTT para la decodificación MPS C<sub>TTT</sub> (parámetros CPC para el decodificador MPS) y la matriz del convertidor de mezcla descendente G.

[0237] C<sub>TTT</sub> es la matriz de predicción para obtener la renderización objetivo a partir de la mezcla descendente modificada.

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

$$C_{TTT}\hat{X} = C_{TTT}GX \approx A_3S$$
.

[0238]  $A_3$  es una matriz de renderización reducida del tamaño 3 x N, que describe la renderización al canal izquierdo, derecho y central, respectivamente. Se la obtiene en términos de  $A_3 = D_{36}M_{ren}$ , donde la matriz de mezcla descendente parcial 6 a 3  $D_{36}$  se define según

$$\mathbf{D}_{36} = \begin{pmatrix} w_1 & 0 & 0 & 0 & w_1 & 0 \\ 0 & w_2 & 0 & 0 & 0 & w_2 \\ 0 & 0 & w_3 & w_3 & 0 & 0 \end{pmatrix}.$$

[0239] Los pesos parciales de mezcla descendente  $\mathbf{w}_p$  p=1,2,3 se ajustan de tal manera que la energía  $\det^{w_p(y_{2p-1}+y_{2p})}$  es igual a la suma de las energías  $\|y_{2p-1}\|^2 + \|y_{2p}\|$  hasta un factor límite.

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

donde  $f_{i,j}$  indica los elementos de  ${f F}$ 

40

donde

5

25

**[0240]** Para la estimación de la matriz de predicción pretendida  $C_{TTT}$  y la matriz de procesamiento de mezcla descendente G definimos una matriz de predicción  $C_3$  del tamaño  $3 \times 2$ , que conduce a la renderización objetivo.

 $C_3X \approx A_3S$ 

[0241] Dicha matriz se deriva teniendo en cuenta las ecuaciones normales

5

10

15

20

30

35

40

$$C_3(DED^*) \approx A_3ED^*$$
.

[0242] La solución a las ecuaciones normales da la mejor coincidencia posible de la forma de onda para la salida objetivo dado el modelo de covarianza de objetos. Ahora se obtiene G y  $C_{TTT}$  resolviendo el sistema de ecuaciones

$$C_{TTT}G = C_3$$

**[0243]** Para evitar problemas numéricos al calcular el término  $\mathbf{J} = (\mathbf{DED}^*)^{-1}$ , se modifica  $\mathbf{J}$ . En primer lugar se calculan los valores propios (término en otro idioma)  $\lambda_{1,2}$  de  $\mathbf{J}$ , resolviendo  $\det(\mathbf{J} - \lambda_{1,2}\mathbf{I}) = 0$ .

**[0244]** Los valores propios se clasifican en orden descendente  $(\lambda_1 \ge \lambda_2)$  y el vector propio correspondiente al valor propio más alto se calcula de acuerdo con la ecuación anterior. Con seguridad ha de yacer en el plano x positivo (el primer elemento tiene que ser positivo). El segundo vector propio se obtiene a partir del primero por una rotación de -90 grados:

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

25 [0245] Se computa una matriz de ponderación a partir de la matriz de mezcla descendente **D** y la matriz de predicción **C**<sub>3</sub>, **W** = (**D** diag(**C**<sub>3</sub>)).

**[0246]** Dado que  $C_{TTT}$  es una función de los parámetros de predicción  $c_1$  y  $c_2$  (definidos en ISO/IEC 23003–1:2007),  $C_{TTT}G=C_3$  se reformula de la siguiente manera, para encontrar el punto estacionario o los puntos de la función,

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

donde  $\mathbf{r} = \mathbf{D}_{TTT} \mathbf{C}_3$ ) W  $(\mathbf{D}_{TTT} \mathbf{C}_3)^*$  y  $\mathbf{b} = \mathbf{GWC}_3 \mathbf{v}$ , donde

$$\mathbf{D}_{\mathsf{TTT}} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad \mathbf{v} \quad \mathbf{v} = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}.$$

**[0247]** Si r no ofrece una solución única (det(r) <  $10^{-3}$ ), se elige el punto que yace más cerca del punto que da lugar a un pasaje TTT. Como primer paso, se elige la fila i de r  $\gamma=[\gamma_{i,1} \ \gamma_{i,2}]$  donde los elementos contienen la mayor cantidad de energía, por consiguiente,  $\gamma_{i,1}^2 + \gamma_{i,2}^2 \ge \gamma_{j,1}^2 + \gamma_{j,2}^2$ , j=1,2.

[0248] Luego se determina una solución de tal manera que

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

[0249] Si la solución obtenida para  $\tilde{\mathcal{C}}_{i}$   $\tilde{\mathcal{C}}_{7}$  está fuera del rango de coeficientes de predicción permitidos que se define como  $-2 \leq \tilde{\mathcal{C}}_{f} \leq 3$  (según definición en ISO/IEC 23003–1:2007), se calculan de acuerdo con lo siguiente.

[0250] En primer lugar se define la serie de puntos,  $x_p$ , de acuerdo con:

$$\mathbf{X}_{p} \in \begin{bmatrix} \left(\min\left(3, \max\left(-2, -\frac{-2\gamma_{1,2} - b_{1}}{\gamma_{1,1} + \varepsilon}\right)\right)\right), \left(\min\left(3, \max\left(-2, -\frac{3\gamma_{1,2} - b_{1}}{\gamma_{1,1} + \varepsilon}\right)\right)\right) \\ -2 \\ \left(\min\left(3, \max\left(-2, -\frac{-2\gamma_{2,1} - b_{2}}{\gamma_{2,2} + \varepsilon}\right)\right)\right), \left(\min\left(3, \max\left(-2, -\frac{3\gamma_{2,1} - b_{2}}{\gamma_{2,2} + \varepsilon}\right)\right)\right) \end{bmatrix}, \left(\min\left(3, \max\left(-2, -\frac{3\gamma_{2,1} - b_{2}}{\gamma_{2,2} + \varepsilon}\right)\right)\right) \end{bmatrix}$$

y la función de distancia,

5

10

15

25

Functist 
$$(\mathbf{x}_p) = \mathbf{x}_p^{\dagger} \mathbf{\Gamma} \mathbf{x}_{pl} - 2\mathbf{b} \mathbf{x}_p$$
.

[0251] Luego se definen los parámetros de predicción de acuerdo con

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

[0252] Se restringen los parámetros de predicción de acuerdo con:

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

20 donde  $\lambda$ ,  $\gamma_1$  y  $\gamma_2$  se definen de la siguiente manera

$$\begin{split} \gamma_1 &= \frac{2f_{1,1} + 2f_{5,5} - f_{3,3} + f_{1,3} + f_{5,3}}{2f_{1,1} + 2f_{5,5} + 2f_{3,3} + 4f_{1,3} + 4f_{5,3}}, \\ \gamma_2 &= \frac{2f_{2,2} + 2f_{6,6} - f_{3,3} + f_{2,3} + f_{6,3}}{2f_{2,2} + 2f_{6,6} + 2f_{3,3} + 4f_{2,3} + 4f_{6,3}}, \end{split}$$

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

[0253] En el caso del decodificador MPS, se obtienen los CPCs y los correspondiente ICC<sub>TTT</sub> de la siguiente manera

$$D_{CPC_{-1}} = c_1 (I,m), D_{CPC_{-2}} = c_2 (I,m) y D_{ICC,TTT=1}$$

4.2.2.2.2 Renderización entre canales anteriores y envolventes

5 **[0254]** Los parámetros que determinan la renderización entre canales anteriores y envolventes se puede estimar directamente de la matriz de covarianza objetivo **F** 

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

10 donde (a,b) = (1,2) y (3,4).

[0255] Los parámetros MPS se producen en la forma

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

por cada caja OTT h.

4.2.2.3 Procesamiento estéreo

[0256] A continuación, se describe el procesamiento estéreo de las señales de objetos de audio normales 134 a 64, 322. El procesamiento en estéreo se utiliza para derivar un proceso a la representación general 142, 272 sobre la base de una representación de dos canales de los objetos de audio normales.

25 [0257] La mezcla descendente estéreo X, que está representada por las señales de objetos de audio normales 134, 264,

492a se procesa para obtener la señal de mezcla descendente modificada  $\ddot{X}$ , que está representada por las señales de objeto de audio normales 142, 272:

$$\hat{\mathbf{X}}_{=GX}$$

donde

15

20

30

$$G = D_{TTT}C_3 = D_{TTT} M_{ren}ED^*J.$$

35 **[0258]** La salida estéreo final del transcodificador SAOC  $\hat{\mathbf{X}}$  se produce mezclando  $\mathbf{X}$  con un componente de señal descorrelacionada de acuerdo con:

$$\hat{\mathbf{X}} = \mathbf{G}_{\mathsf{Mod}} \mathbf{X} + \mathbf{P}_2 \mathbf{X}_{\mathsf{d}}.$$

donde la señal descorrelacionada  $X_d$  se calcula de acuerdo con lo descripto anteriormente, y las matrices de mezcla  $G_{Mod}$  y  $P_2$  de acuerdo con lo siguiente:

$$\hat{\mathbf{X}} = \mathbf{G}_{\mathsf{Mod}} \mathbf{X} + \mathbf{P}_2 \mathbf{X}_{\mathsf{d}} ,$$

[0259] En primer lugar, se define la matriz de error de renderización de mezcla ascendente de la siguiente manera:

$$R = A_{dif}EA_{dif},$$

donde

$$A_{dif} = D_{TTT}A_3 - GD_1$$

y, además, se define la matriz de covarianza de la señal predicha  $\hat{\mathbf{R}}$  de la siguiente manera:

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{r}_{1,1} & \hat{r}_{1,2} \\ \hat{r}_{2,1} & \hat{r}_{2,2} \end{pmatrix} = \mathbf{GDED}^* \mathbf{G}^*.$$

[0260] A continuación se puede calcular el vector de ganancia  $g_{vec}$  de la siguiente manera:

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

y la matriz de mezclado se expresa como

5

10

15

25

30

$$\mathbf{G}_{\mathsf{Mod}} = \begin{cases} diag(\mathbf{g}_{\mathsf{vec}})\mathbf{G}, & r_{1,2} > 0, \\ \mathbf{G}, & \text{de otro} \\ modo & \end{cases}$$

[0261] De manera similar, la matriz de mezclado P2 se expresa en los siguientes términos:

$$\mathbf{P}_{2} = \begin{cases} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, r_{1,2} > 0, \\ \mathbf{v}_{R} diag(\mathbf{W}_{d}), \text{ de otro modo.} \end{cases}$$

20 [0262] Para derivar  $v_R$  y  $W_d$ , se debe resolver la ecuación característica de R:

 $det(R - \lambda_{1,2}I)v_{R1,R2} = 0$ , lo que da los valores propios  $\lambda_1$  y  $\lambda_2$ 

[0263] Los vectores propios correspondientes V<sub>R</sub>, y V<sub>R2</sub> de **R** se pueden calcular resolviendo el sistema de ecuaciones:

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

[0264] Los valores propios se clasifican en orden descendente  $(\lambda_1 \ge \lambda_2)$  y el vector propio correspondiente al valor propio más elevado se calcula de acuerdo con la ecuación anterior. Con seguridad ha de yacer 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:

$$\mathbf{R} = \left(\mathbf{v}_{\mathsf{R}1}\mathbf{v}_{\mathsf{R}2}\right) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \left(\mathbf{v}_{\mathsf{R}1}\mathbf{v}_{\mathsf{R}2}\right)^*.$$

35 [0265] La incorporación de  $P_1 = 1$  1)G,  $R_d$  se puede calcular de acuerdo con:

$$\mathbf{R}_{d} = \begin{pmatrix} \mathbf{r}_{d11} & \mathbf{r}_{d12} \\ \mathbf{r}_{d21} & \mathbf{r}_{d22} \end{pmatrix} = diag(\mathbf{P}_{1}(\mathbf{DED}^{*})\mathbf{P}_{1}^{*}),$$

lo que da

5

10

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

y por último la matriz de mezcla,

 $\mathbf{P}_2 = \begin{pmatrix} \mathbf{v}_{\mathsf{R}1} & \mathbf{v}_{\mathsf{R}2} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{\mathsf{d}1} & 0 \\ 0 & \mathbf{w}_{\mathsf{d}2} \end{pmatrix}$ 

4.2.2.4 Modo Dual

[0266] El transcodificador SAOC puede dar lugar al cálculo de las matrices de mezcla P<sub>1</sub>, P<sub>2</sub> y la matriz de predicción C<sub>3</sub> de acuerdo con un esquema alternativo para el rango de frecuencias superiores. Este esquema alternativo es especialmente ventajoso para las señales de mezcla descendente, donde el rango de frecuencias superiores es codificado por un algoritmo de codificación sin conservación de la forma de onda, por ejemplo SBR en AAC de Alta Eficiencia.

[0267] En el caso de las bandas de parámetros superiores, definidas por **bsTttBandsLow**  $\leq$  pb < numBands ,  $P_1$ ,  $P_2$  y  $C_3$  se deben calcular de acuerdo con el esquema alternativo descripto a continuación:

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

25 [0268] Se define la mezcla descendente de energía y los vectores objetivos de energía, respectivamente:

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

y la matriz de ayuda

30

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

[0269] Luego se calcula el vector de ganancia

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

que, por último, da la nueva matriz de predicción

$$\mathbf{C}_{3} = \begin{pmatrix} g_{1}t_{1,1} & g_{1}t_{1,2} \\ g_{2}t_{2,1} & g_{2}t_{2,2} \\ g_{3}t_{3,1} & g_{3}t_{3,2} \end{pmatrix}.$$

- [0270] 5. Modo combinado de decodificación/transcodificación EKS SAOC, codificador da la Fig. 10 y sistemas de acuerdo con las Figs. 5a, 5b
- [0271] A continuación, se presenta una breve descripción del esquema de procesamiento combinado EKS SAOC. Se propone un esquema de procesamiento combinado "EKS SAOC", donde el procesamiento EKS se integra a la cadena de decodificación/transcodificación SAOC normal mediante un esquema de etapas en cascada.
  - 5.1. Codificador de señales de audio de acuerdo con la Fig. 5
- [0272] En un primer paso, se identifican los objetos especiales para el procesamiento EKS (procesamiento Karaoke/solo realzado) como objetos en primer plano (FGO) y su número N<sub>FGO</sub> (también designado N<sub>EAO</sub>) se determina mediante una variable de flujo de bits "bsNumGroupsFGO". Dicha variable de flujo de bits puede estar incluida, por ejemplo, en un flujo de bits SAOC, de acuerdo con lo descripto anteriormente.
- 25 **[0273]** Para la generación del flujo de bits (en un codificador de señales de audio), los parámetros de todos los objetos de entrada N<sub>obj</sub> se reordenan de tal manera que los objetos en primer plano FGO comprendan los últimos parámetros N<sub>FGO</sub> (o, por otro lado, N<sub>EAO</sub>) en cada caso, por ejemplo, OLD₁ para [N<sub>obj</sub> − N<sub>FGO</sub> ≤ i ≤ N<sub>obj</sub> − 1].
- [0274] Del resto de los objetos que son, por ejemplo, objetos BGO de fondo u objetos de audio no realzados, se genera una señal de mezcla descendente a "estilo SAOC normal " que, al mismo tiempo, sirve como objeto BGO de fondo. A continuación, se ejecuta la mezcla descendente del objeto de fondo y los objetos en primer plano al "estilo de procesamiento EKS" y se extrae la información residual de cada objeto de fondo. De esta manera, no es necesario introducir ninguna etapa adicional. Por consiguiente, no se requiere ningún cambio de sintaxis del flujo de bits.
- [0275] Dicho de otro modo, del lado del codificador, los objetos de audio no realzados se distinguen de los objetos de audio realzados. Se produce una señal de audio de mezcla descendente de objetos de audio de un canal o de dos canales (objetos de audio no realzados), en la cual puede haber uno, dos o incluso más objetos de audio normales

(objetos de audio no realzados). A continuación se combina la señal de mezcla descendente de objetos de audio de un canal o de dos canales con una o más señales de objeto de audio realzados (que pueden ser, por ejemplo, señales de un canal o señales de dos canales) para obtener una señal de mezcla descendente común (que puede ser, por ejemplo, una señal de mezcla descendente de dos canales) combinando las señales de audio de los objetos de audio realzados y la señal de mezcla descendente de objetos de audio normales.

[0276] A continuación se describe en forma sucinta la estructura básica de dicho codificador de etapas en cascada tomando como referencia la Fig. 10, que ilustra una representación esquemática de bloques de un codificador SAOC 1000, de acuerdo con una realización de la invención. El codificador SAOC 1000 comprende un primer dispositivo de mezcla descendente SAOC 1010, que por lo general es un dispositivo de mezcla descendente que no produce información residual. El dispositivo de mezcla descendente SAOC 1010 está configurado para recibir una pluralidad de señales de objetos de audio N<sub>BGO</sub> 1012 de objetos de audio normales (no realzados). Además, el dispositivo de mezcla descendente SAOC 1010 está configurado para proporcionar una señal de mezcla descendente de objetos de audio normales 1014 sobre la base de los objetos de audio normales 1012, de tal manera que la señal de mezcla descendente de objetos de audio normales 1014 combine las señales de objetos de audio normales 1012 de acuerdo con los parámetros de mezcla descendente. El dispositivo de mezcla descendente SAOC 1010 también produce una información SAOC de objetos de audio normales 1016, que describe las señales de objetos de audio normales y la mezcla descendente. Por ejemplo, la información SAOC de objetos de audio normales 1016 puede comprender una información de ganancia de mezcla descendente DMG y una información de diferencias de nivel de mezcla descendente de los canales DCLD que describe la mezcla descendente ejecutada por el dispositivo de mezcla descendente SAOC 1010. Además, la información SAOC de objetos de audio normales 1016puede comprender una información de diferencias de nivel de los objetos y una información de correlación entre objetos que describe una relación entre los objetos de audio normales descripta por la señal de objetos de audio normales 1012.

- [0277] El codificador 1000 comprende asimismo un segundo dispositivo de mezcla descendente SAOC 1020, que por lo general está configurado para producir una información residual. El segundo dispositivo de mezcla descendente SAOC 1020 está configurado preferentemente para recibir una o más señales de objetos de audio realzados 1022 y también para recibirla señal de mezcla descendente de objetos de audio normales 1014.
- [0278] El segundo dispositivo de mezcla descendente SAOC 1020 está configurado asimismo para producir una señal de mezcla descendente SAOC común 1024 sobre la base de las señales de objetos de audio 1022 y la señal de mezcla descendente de objetos de audio normales 1014. Al producir señal de mezcla descendente SAOC común, el segundo dispositivo de mezcla descendente SAOC 1020 trata típicamente la señal de mezcla descendente de objetos de audio normales 1014 como señal única de de objetos de un canal o de dos canales.

[0279] El segundo dispositivo de mezcla descendente SAOC 1020 está configurado además para producir una información SAOC sobre objetos de audio realzados que describe, por ejemplo, diferencias de nivel de mezcla descendente de los valores de los canales DCLD asociadas a los objetos de audio realzados, valores OLD de diferencia de nivel de los objetos asociados a los objetos de audio realzados y valores de correlación entre objetos IOC asociados a los objetos de audio realzados. Además, el segundo SAOC 1020 está configurado preferentemente para proporcionar información residual asociada a cada uno de los objetos de audio realzados, por lo que la información residual asociada a los objetos de audio realzados describe la diferencia entre una señal de objetos de audio realzados individual original y una señal de objetos de audio realzados individual estimada que se puede extraer de la señal de mezcla descendente usando la información de mezcla descendente DMG, DCLD y la información de objetos OLD, IOC.

[0280] El codificador de audio 1000 es muy adecuado para cooperar con el decodificador de audio aquí descripto.

5.2. Decodificador de señales de audio de acuerdo con la Fig. 5a

5

10

15

20

40

45

55

[0281] A continuación, se describe la estructura básica de un decodificador combinado EKS SAOC 500, un diagrama esquemático del cual está expuesto en la Fig. 5a.

[0282] El decodificador de audio 500 de acuerdo con la Fig. 5a está configurado para recibir una señal de mezcla descendente 510, información de flujo de bits SAOC 512 y una información sobre matrices de renderización 514. El decodificador de audio 500 comprende un procesamiento realzado Karaoke/Solo y una renderización de objetos en primer plano 520, que está configurado para producir una primera señal de objetos de audio 562, que describe objetos renderizados en primer plano, y una segunda señal de objetos de audio 564, que describe los objetos de fondo. Los objetos en primer plano pueden ser, por ejemplo, los denominados "objetos de audio realzados" y los objetos de fondo pueden ser, por ejemplo, los denominados "objetos de audio normales" u "objetos de audio no realzados". El

decodificador de audio 500 comprende asimismo decodificación SAOC normal 570, que está configurado para recibir la segunda señal de objetos de audio 562 y para producir, sobre la base de ésta, una versión procesada 572 de la segunda señal de objetos de audio 564. El decodificador de audio 500 comprende asimismo un combinador 580, que está configurado para combinar la primera señal de objetos de audio 562 y la versión procesada 572 de la segunda señal de objetos de audio 564, para obtener una señal de salida 520.

[0283] A continuación, se describe la funcionalidad del decodificador de audio 500 en forma algo más detallada. Del lado de la decodificación/transcodificación SAOC, el proceso de mezcla ascendente da lugar a un esquema de etapas en cascada que comprende, en primer lugar, un procesamiento realzado Karaoke—Solo (procesamiento EKS) para descomponer la señal de mezcla descendente para obtener el objeto de fondo (BGO) y los objetos en primer plano (FGOs). Las diferencias de nivel requeridas de los objetos (OLDs) y las correlaciones entre objetos (IOCs) correspondientes a los objetos de fondo se derivan de la información sobre objetos y sobre mezcla descendente (que es, en ambos casos, información paramétrica relacionada con los objetos, y que, en ambos casos, está típicamente incluida en el flujo de bits SAOC):

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

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

$$IOC_{LR} = \begin{cases} IOC_{0,1}, & N - N_{FGO} = 2, \\ 0, & \text{de otro modo.} \end{cases}$$

[0284] Además, esta etapa (que por lo general es ejecutada por el procesamiento EKS y la renderización de objetos en primer plano 520) incluye el mapeo de los objetos en primer plano contra los canales de salida finales 25 (por lo que, por ejemplo, la primera señal de objetos de audio 562 es una señal multicanal en la cual se mapean los objetos en primer plano contra uno o más canales cada uno). El objeto de fondo (que por lo general comprende una pluralidad de los denominados "objetos de audio normales") se renderiza a los correspondientes canales de salida mediante un proceso de decodificación SAOC normal (o, de lo contrario, en algunos casos por un proceso de transcodificación SAOC). Este proceso puede ser ejecutado, por ejemplo, por la decodificación SAOC normal 570. La etapa de mezcla final (por ejemplo, el combinador 580) produce una combinación ventajosa de los objetos renderizados en primer plano y las señales de objetos de fondo a la salida.

**[0285]** Este sistema combinado EKS SAOC representa una combinación de todas las propiedades ventajosas del sistema SAOC normal y s modo EKS. Esta estrategia permite obtener la correspondiente eficiencia utilizando el sistema propuesto con el mismo flujo de bits para entornos de reproducción tanto clásicos (renderización moderada) como similares a Karaoke/Solo (renderización extrema).

5.3. Estructura generalizada de acuerdo con la Fig. 5b

5

10

15

20

25

30

35

40

[0286] A continuación, se describe una estructura generalizada de un sistema combinado EKS SAOC 590 haciendo referencia a la Fig. 5b, que ilustra un diagrama esquemático de bloques de ese sistema combinado EKS SAOC. El sistema combinado EKS SAOC 590 de la Fig. 5b puede ser considerado asimismo como decodificador de audio.

[0287] El sistema combinado EKS SAOC 590 está configurado para recibir una señal de mezcla descendente 510a, una información de flujo de bits SAOC 512a e información sobre matrices de renderización 514a. Además, el sistema combinado EKS SAOC 590 está configurado para producir una señal de salida 520a sobre la base de ésta.

[0288] El sistema combinado EKS SAOC 590 comprende una etapa de procesamiento I del tipo SAOC 520a, que recibe la señal de mezcla descendente 510a, la información de flujo de bits SAOC 512a (o por lo menos parte de la misma) y la información de matrices de renderización 514a (o por lo menos parte de la misma). En particular, la etapa de procesamiento I del tipo SAOC 520a recibe valores de diferencias de nivel de los objetos de la primera etapa (OLD,). La etapa de procesamiento I del tipo SAOC 520a suministra una o más señales 562a que describe una primera serie de objetos (por ejemplo, objetos de audio de un primer tipo de objetos de audio). La etapa de procesamiento I del tipo SAOC 520a proporciona asimismo una o más señales 564a que describen una segunda serie de objetos.

[0289] El sistema combinado EKS SAOC comprende asimismo una etapa de procesamiento II del tipo SAOC 570a, que está configurada para recibir dichas una o más señales 564a que describe la segunda serie de objetos y para producir, sobre la base de ésta, una o más señales 572a que describen una tercera serie de objetos usando las diferencias de nivel de los objetos de la segunda etapa, que están incluidos en la información de flujo de bits SAOC 512a, y también por lo menos parte de la información de matrices de renderización 514. El sistema combinado EKS SAOC comprende asimismo un combinador 580a, que puede ser, por ejemplo, un sumador, para producir las señales de salida 520a combinando dichas una o más señales 562a que describen la primera serie de objetos y dichas una o más señales 570a que describen la tercera serie de objetos (donde la tercera serie de objetos puede ser una versión procesada de la segunda serie de objetos).

[0290] Para resumir lo antedicho, la Fig. 5b ilustra una forma generalizada de la estructura básica descripta con referencia a la Fig. 5a anterior de otra realización de la invención.

- 6. Evaluación Perceptual del Esquema de Procesamiento Combinado EKS SAOC
- 6.1 Metodología de Ensayo, Diseño y Elementos

5

10

15

25

30

35

40

45

50

55

[0291] Estas pruebas de audición subjetivas se llevaron a cabo en un recinto de audición acústicamente aislado que está diseñado para permitir la audición de alta calidad. La reproducción se realizó utilizando auriculares (STAX SR Lambda Pro con Convertidor Lake—People DIA y Monitor STAX SRM). El procedimiento de ensayo siguió los procedimientos standard empleados en las pruebas de verificación de audio espacial, basadas en el procedimiento de "estímulo múltiple con referencia y anclas ocultos" (MUSHRA) para la evaluación subjetiva de la calidad de audio intermedia (ver la referencia [7]).

[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, los oyentes fueron instruidos para comparar todas las condiciones de ensayo contra la referencia. Las condiciones de ensayo fueron distribuidas al azar por cada elemento de ensayo y por cada oyente. Las respuestas subjetivas fueron registradas por un programa MUSHRA de computación en una escala en el rango de 0 a 100.

[0293] Se permitió un desplazamiento instantáneo entre los elementos en estudio. La prueba MUSHRA se llevó a cabo para evaluar la eficiencia perceptual de los modos SAOC considerados y el sistema propuesto descripto en la tabla de la Fig. 6a, que presenta una descripción del diseño de ensayo de audición. Las correspondientes señales de mezcla descendente fueron codificadas utilizando un codificador de núcleo AAC con una velocidad de transmisión de bits de 128 kbps. Para evaluar la calidad perceptual del sistema combinado EKS SAOC, se lo compara contra el sistema SAOC RM normal (sistema de modelos de referencia SAOC) y el modo EKS actual (modo realzado de Karaoke–Solo) correspondientes a dos entornos de ensayo de renderización diferentes descriptos en la tabla de la Fig. 6b, que describe los sistemas en estudio.

[0294] Se aplicó la codificación residual con una velocidad de transmisión de bits de 20 kbps para el modo EKS actual y un sistema combinado EKS SAOC propuesto. Se debe notar que en el caso del modo EKS actual es necesario generar un objeto de fondo estéreo (BGO) con anterioridad al procedimiento actual de codificación/decodificación, ya que este modo tiene limitaciones con respecto al número y tipo de objetos de entrada.

[0295] El material de ensayo de audición y los correspondientes parámetros de mezcla descendente y renderización utilizados en las pruebas realizadas fueron seleccionados de la serie de elementos de audio propuestos (CfP) descriptos en el documento [2]. Los correspondientes datos sobre configuraciones de aplicaciones de renderización para "Karaoke" y "Clásico" se pueden encontrar en la tabla de la Fig. 6c, que describe los elementos de la prueba de audición y las matrices de renderización.

6.2 Resultados de la Prueba de Audición

[0296] Se puede encontrar una breve reseña general en términos de diagramas que demuestran los resultados obtenidos de la prueba de audición en las Figs. 6d y 6e, donde la Fig. 6d ilustra las puntuaciones MUSHRA promediadas correspondientes a la prueba de audición de renderización del tipo Karaoke/Solo y la Fig. 6e ilustra las puntuaciones MUSHRA promediadas correspondientes a la prueba de audición de renderización clásica. Los trazados ilustran la calificación MUSHRA promedio por elemento asignada por todos los oyentes y el valor medio estadístico de todos los elementos evaluados junto con los intervalos de confianza asociados de 95%.

[0297] Se pueden plantear las siguientes conclusiones basándose en los resultados de las pruebas de audición realizadas:

- La Fig. 6d representa la comparación del modo actual EKS con el sistema combinado EKS SAOC para aplicaciones del tipo Karaoke. Con respecto a todos los elementos analizados, no se puede observar diferencia significativa alguna (en el sentido estadístico) en el rendimiento entre estos dos sistemas. De esta observación se puede concluir que el sistema combinado EKS SAOC puede explotar eficientemente la información residual que alcanza la eficiencia del modo EKS. También se puede notar que la eficiencia del sistema SAOC normal (sin residual) es inferior a los otros dos sistemas.
- La Fig. 6e representa la comparación del SAOC normal actual con el sistema combinado EKS SAOC para entornos de renderización clásica. En el caso de todos los elementos analizados, la eficiencia de estos dos sistemas es estadísticamente igual. Esto demuestra la funcionalidad correcta del sistema combinado EKS SAOC para una situación de renderización clásica.

[0298] Por lo tanto, se puede concluir que el sistema unificado propuesto que combina el modo EKS con el SAOC normal mantiene las ventajas de la calidad auditiva subjetiva respecto de los tipos de renderización correspondientes.

[0299] Tomando en cuenta el hecho de que el sistema combinado EKS SAOC propuesto ya no tiene restricciones con respecto a los objetos BGO, sino que tiene una capacidad de renderización totalmente flexible del modo SAOC normal y puede utilizar el mismo flujo de bits para todos los tipos de renderización, parece ser ventajoso incorporarlo a la norma MPEG SAOC.

7. Procedimiento de acuerdo con la Fig. 7

5

15

20

30

35

40

45

50

55

[0300] A continuación se describe un procedimiento para producir una representación de señal de mezcla ascendente dependiendo de una representación de señal de mezcla descendente e información paramétrica relacionada con los objetos con referencia a la Fig. 7, que ilustra un gráfico de flujo de dicho procedimiento.

[0301] El procedimiento 700 comprende una etapa 710 de descomposición de una representación de señal de mezcla descendente, para dar origen a una primera información de audio que describe una primera serie de uno o más objetos de audio de un primer tipo de objetos de audio y una segunda información de audio que describe una segunda serie de uno o más objetos de audio de un segundo tipo de objetos de audio que depende de la representación de señal de mezcla descendente y por lo menos parte de la información paramétrica relacionada con los objetos. El procedimiento 700 comprende asimismo una etapa 720 de procesamiento de la segunda información de audio que depende de la información paramétrica relacionada con los objetos, para obtener una versión procesada de la segunda información de audio.

[0302] El procedimiento 700 comprende asimismo una etapa 730 de combinación de la primera información de audio con la versión procesada de la segunda información de audio, para obtener la representación de señal de mezcla ascendente.

**[0303]** El procedimiento 700 de acuerdo con la Fig. 7 puede ser complementado con cualquiera de las características y funcionalidades aquí descriptas con respecto al aparato de la invención. Además, el procedimiento 700 trae aparejadas las ventajas mencionadas con respecto al aparato de la invención.

8. Alternativas de Implementación

[0304] Si bien se han descripto algunos aspectos en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa del procedimiento o a una característica de una etapa del procedimiento. De manera análoga, los aspectos descriptos en

el contexto de una etapa del procedimiento también representan una descripción de un correspondiente bloque o elemento o característica de un aparato correspondiente. Algunas o todas las etapas del procedimiento pueden ser ejecutadas por (o utilizando) un aparato de hardware, como por ejemplo un microprocesador, una computadora programable o un circuito electrónico. En algunas realizaciones, una o más de las etapas más importantes del procedimiento pueden ser ejecutadas por ese tipo de aparato.

5

10

15

20

25

40

45

55

[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 inalámbrica o un medio de transmisión cableado tal como la Internet.

[0306] Dependiendo de ciertos requerimientos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación se puede ejecutar utilizando un medio de almacenamiento digital, por ejemplo un disco blando, un DVD, un Blue–Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, con señales de control legibles electrónicamente almacenados en el mismo, que cooperan (o pueden cooperar) con un sistema de computación programable a fin de ejecutar el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por computación.

[0307] Algunas realizaciones de acuerdo con la invención comprenden un portador de datos que cuenta con señales de control legibles electrónicamente, con capacidad para cooperar con un sistema de computación programable para la ejecución de los procedimientos aquí descriptos.

[0308] En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto programa de computación con un código de programa, donde el código de programa es operativo para realizar uno de los procedimientos al ejecutarse el programa de computación en una computadora. El código de programa puede ser almacenado, por ejemplo, en un portador legible por una máquina.

[0309] Otras realizaciones comprenden el programa de computación para ejecutar uno de los procedimientos aquí descriptos, almacenados en un portador legible por una máquina.

[0310] En otras palabras, una realización del procedimiento de la invención consiste, por lo tanto, en un programa de computación que consta de un código de programa para realizar uno de los procedimientos aquí descriptos al ejecutarse el programa de computación en una computadora.

[0311] Otra realización del procedimiento de la invención consiste, por lo tanto, en un portador de datos (o un medio de almacenamiento digital, o un medio legible por computadora) que comprende, grabado en el mismo, el programa de computación para ejecutar uno de los procedimientos descriptos en la presente. El portador de datos, el medio de almacenamiento digital o el medio registrado son por lo general tangibles y/no no de transmisión.

[0312] Otra realización del procedimiento de la invención consiste, por lo tanto, en un flujo de datos o una secuencia de señales que representan el programa de computación para ejecutar uno de los procedimientos aquí descriptos. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferido a través de una conexión de comunicación de datos, por ejemplo a través de la Internet.

[0313] Otra realización comprende un medio de procesamiento, por ejemplo una computadora o un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los procedimientos aquí descriptos.

[0314] Otra realización comprende una computadora en la cual se ha instalado el programa de computación para ejecutar uno de los procedimientos aquí descriptos.

[0315] En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en el campo) para ejecutar algunas o todas las funcionalidades de los procedimientos aquí descriptos. En algunas realizaciones, una matriz de puertas programable en el campo puede cooperar con un microprocesador para ejecutar uno de los procedimientos aquí descriptos. En general, los procedimientos se ejecutan preferentemente mediante cualquier aparato de hardware.

[0316] Las realizaciones precedentemente descriptas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles descriptos en este documento han de ser evidentes para otras personas con capacitación en la técnica. Por lo tanto, sólo se pretende

limitarse al alcance de las reivindicaciones de patente siguientes y no a los detalles específicos presentados a título de descripción y explicaciones de las presentes realizaciones.

## 9. Conclusiones

5

10

20

25

[0317] A continuación se resumen brevemente algunos aspectos y ventajas del sistema combinado EKS SAOC de acuerdo con la presente invención. En el caso de los entornos de reproducción de Karaoke y Solo, el modo de procesamiento SAOC EKS soporta tanto la reproducción de los objetos de fondo/objetos en primer plano) de estos grupos de objetos.

[0318] Asimismo, el primer modo se considera el objetivo principal del procesamiento EKS, el último aporta mayor flexibilidad.

[0319] Se ha encontrado que una generalización de la funcionalidad EKS, en consecuencia, conlleva el esfuerzo de combinar el EKS con el modo de procesamiento SAOC normal para obtener un sistema unificado. Los potenciales de dicho sistema unificado son:

- Una única estructura clara de decodificación/transcodificación SAOC;
- Un flujo de bits tanto para el modo EKS como para el SAOC normal;
  - no hay limitación con respecto al número de objetos de entrada que comprenden el objeto de fondo (BGO), por lo que no hay necesidad de generar el objeto de fondo antes de la etapa de codificación SAOC y
  - Soporte de una codificación residual correspondiente a objetos en primer plano que da una calidad perceptual mejorada en situaciones demandantes de reproducción para Karaoke/Solo.

[0320] Estas ventajas se hacen posibles merced al sistema unificado aquí descripto.

## Referencias

## [0321]

- [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.
- [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 of MPEG SAOC RM0", 83rd MPEG Meeting, Antalya, Turkey, January 2008.
- [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 and 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.
- [8] ISO/IEC 23003-1:2007, Information technology MPEG audio technologies Part 1: MPEG Surround.

5

## **REIVINDICACIONES**

1. Descodificador de señal de audio (100; 200; 500; 590) para suministrar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo (112; 210; 510; 510a) y de una información paramétrica relativa al 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 suministrar una primera información de audio (132; 262; 562; 562a) que describe un primer conjunto de uno o varios 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 varios objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y con ayuda de al menos una parte de la información paramétrica relativa al objeto;

un procesador de señal de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para tratar la segunda información de audio en función de la información paramétrica relativa al objeto, para obtener una versión tratada (142; 272; 572; 572a) de la segunda información de audio; y

un combinador de señal de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión tratada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;

en el cual el separador de objeto está configurado para obtener la primera información de audio y la segunda información de audio según

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\mathsf{Predicción}} \begin{pmatrix} l_0 \\ -\frac{r_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$
 $\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\mathsf{Predicción}} \begin{pmatrix} l_0 \\ -\frac{r_0}{res_0} \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$ 

25 donde

5

10

20

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

donde

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

30 donde X<sub>OBJ</sub> representa los canales de la segunda información de audio;

donde X<sub>EAO</sub> representa las señales de objeto de la primera información de audio;

donde D<sup>-1</sup> representa una matriz que es la inversa de una matriz de mezcla hacia abajo extendida:

donde C describe una matriz que representa una pluralidad de coeficientes de predicción de canal, c j,0, c j,1;

donde  $l_0$  y  $r_0$  representan los canales de la representación de señal de mezcla hacia abajo; donde res0 a res $N_{\it EAO}$  -1 representan los canales residuales; y

5

15

donde  $\mathbf{A}^{\text{EAO}}$  es una matriz de pre-representación de EAO cuyas entradas describen un cartografiado de objetos de audio mejorados en los canales de una señal de objeto de audio mejorado  $\mathbf{X}_{\text{EAO}}$ ;

en el cual el separador de objetos está configurado para obtener la matriz de mezcla hacia abajo inversa **D**<sup>-1</sup> como la inversa de una matriz de mezcla hacia abajo extendida D que se define como

$$\tilde{\mathbf{D}} = \begin{pmatrix} 1 & 0 & | m_0 & \dots & m_{N_{EAO}-1} \\ 0 & 1 & | n_0 & \dots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & | -1 & \dots & 0 \\ \vdots & \vdots & | 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & | 0 & \dots & -1 \end{pmatrix}$$

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

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{pmatrix}$$

donde m0 a m<sub>NEAO</sub> -1 son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio;

donde no a n<sub>NEAO</sub> <sup>-1</sup> son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio;

en el cual el separador de objeto está configurado para calcular los coeficientes de predicción c j.o y c j.1 como

to está configurado para calcular los coeficientes de predicción 
$$ilde{c}_{j,0} = rac{P_{LoCo,j}P_{Ro} - P_{RoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2} \ ilde{c}_{j,1} = rac{P_{RoCo,j}P_{Lo} - P_{LoCo,j}P_{LoRo}}{P_{Lo}P_{Ro} - P_{LoRo}^2};$$

y en el cual el separador de objeto está configurado para derivar los coeficientes de predicción restringidos c  $_{j,0}$  y c  $_{j,1}$  a partir de los coeficientes de predicción c  $_{j,0}$  y c  $_{j,1}$  com ayuda de un algoritmo de restricción, o para utilizar los coeficientes de predicción c  $_{i,0}$  y c  $_{i,1}$  como coeficientes de predicción c  $_{i,0}$  y c  $_{i,1}$ ;

en el cual las cantidades de energía PLo, PRo, PLoRo, PLoCoj y PRoCoj se definen como

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

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

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

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

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

donde los parámetros OLDL, OLDR, y IOCL,R corresponden a unos objetos de audio del segundo tipo de objeto de audio y se definen según

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

$$OLD_L = \sum_{i=0}^{N-N_{EAO^{-1}}} d_{0,i}^2 OLD_i,$$
 
$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N-N_{EAO} = 2, \\ 0, & autrement. \end{cases}$$

donde d $_{0,i}$ y d $_{1,i}$  son unos valores de mezcla hacia abajo asociados a los objetos de audio del segundo tipo de objeto de audio:

donde OLD<sub>i</sub> son unos valores de diferencia de nivel de objeto asociados a los objetos de audio del segundo tipo de objeto de audio;

donde N es el número total de objetos de audio;

15

20

25

35

donde Nese el número de objetos de audio,

donde N<sub>EAO</sub> es el número de objetos de audio del prime

donde IOC<sub>0,1</sub> es un valor de correlación entre objetos a *En los demás casos.* audio del segundo tipo de objeto de audio;

donde  $e_{ij}$  y  $e_{L,R}$  son unos valores de covariancia derivados de los parámetros de diferencia de nivel de objeto y de los parámetros de correlación entre objetos; y

donde  $e_{ij}$  está asociada a un par de objetos de audio del primer tipo de objeto de audio y  $e_{L,R}$  está asociada a un par de objetos de audio del segundo tipo de objeto de audio.

2. Descodificador de señal de audio (100; 200; 500; 590) para suministrar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo (112; 210; 510; 510a) y de una información paramétrica relativa al 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 suministrar una primera información de audio (132; 262; 562; 562a) que describe un primer conjunto de uno o varios 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 varios objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y con ayuda de al menos una parte de la información paramétrica relativa al objeto;

un procesador de señal de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y

para tratar la segunda información de audio en función de la información paramétrica relativa al objeto, para obtener una versión tratada (142; 272; 572; 572a) de la segunda información de audio; y

un combinador de señal de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión tratada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;

en el cual el separador de objeto está configurado para obtener la primera información de audio y la segunda información de audio según

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

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{\text{Energia}} \begin{pmatrix} l_o \\ r_o \end{pmatrix}$$

donde XOBJ representa los canales de la segunda información de audio;

donde X<sub>EAO</sub> representa las señales de objeto de la primera información de audio;

donde

5

10

25

30

$$\mathbf{M}_{OBJ}^{\text{Energia}} = \begin{pmatrix} \frac{OLD_{L}}{OLD_{L} + \sum_{i=0}^{N_{EAO}-1} m_{i}^{2}OLD_{i}} & 0 \\ 0 & \sqrt{\frac{OLD_{R}}{OLD_{R} + \sum_{i=0}^{N_{EAO}-1} n_{i}^{2}OLD_{i}}} \\ 0 & \sqrt{\frac{OLD_{R}}{OLD_{R} + \sum_{i=0}^{N_{EAO}-1} n_{i}^{2}OLD_{i}}} \\ \mathbf{M}_{EAO}^{\text{Energia}} = & \vdots & \vdots \\ \sqrt{\frac{m_{N_{EAO}-1}^{2}OLD_{N_{EAO}-1}}{OLD_{L} + \sum_{i=0}^{N_{EAO}-1} OLD_{N_{EAO}-1}}} & \sqrt{\frac{n_{N_{EAO}-1}^{2}OLD_{N_{EAO}-1}}{N_{N_{EAO}-1} OLD_{N_{EAO}-1}}} \\ \sqrt{\frac{n_{N_{EAO}-1}^{2}OLD_{N_{EAO}-1}}{OLD_{L} + \sum_{i=0}^{N_{EAO}-1} n_{i}^{2}OLD_{i}}} & \sqrt{\frac{n_{N_{EAO}-1}^{2}OLD_{N_{EAO}-1}}{OLD_{R} + \sum_{i=0}^{N_{EAO}-1} n_{i}^{2}OLD_{i}}} \\ \sqrt{\frac{n_{N_{EAO}-1}^{2}OLD_{N_{EAO}-1}}{OLD_{L} + \sum_{i=0}^{N_{EAO}-1} n_{i}^{2}OLD_{i}}} & \sqrt{\frac{n_{N_{EAO}-1}^{2}OLD_{N_{EAO}-1}}{OLD_{R} + \sum_{i=0}^{N_{EAO}-1} n_{i}^{2}OLD_{i}}} \\ \end{pmatrix}$$

donde  $m_0$  a  $mN_{EAO}^{-1}$  son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio;

donde no a  $nN_{EAO}^{-1}$  son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio;

donde OLD<sub>i</sub> son unos valores de diferencia de nivel de objeto asociados a los objetos de audio del primer tipo de objeto de audio;

donde  $OLD_L$  y  $OLD_R$  son unos valores de diferencia de nivel de objeto comunes asociados a los objetos de audio del segundo tipo de objeto de audio; y

- donde **A**<sup>EAO</sup> es una matriz de pre-representación EAO cuyas entradas describen un cartografiado de objetos de audio mejorados en los canales de una señal de objeto de audio mejorado **X**<sub>EAO</sub>.
  - 3. Descodificador de señal de audio (100; 200; 500; 590) para suministrar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo (112; 210; 510; 510a) y de una información paramétrica relativa al 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 suministrar una primera información de audio (132; 262; 562; 562a) que describe un primer conjunto de uno o varios 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 varios objetos de audio de un segundo tipo de objeto de audio en

función de la representación de señal de mezcla hacia abajo y con ayuda de al menos una parte de la información paramétrica relativa al objeto;

un procesador de señal de audio configurado para recibir la segunda información de audio (134; 264; 564; 564a) y para tratar la segunda información de audio en función de la información paramétrica relativa al objeto, para obtener una versión tratada (142; 272; 572; 572a) de la segunda información de audio; y

un combinador de señal de audio (150; 280; 580; 580a) configurado para combinar la primera información de audio con la versión tratada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;

en el cual el separador de objeto está configurado para obtener la primera información de audio y la segunda información de audio según

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energia}} \left( d_0 \right)$$

$$\mathbf{X}_{\scriptscriptstyle EAO} = \mathbf{A}^{\scriptscriptstyle EAO} \mathbf{M}_{\scriptscriptstyle EAO}^{\scriptscriptstyle (Energia)} \left(d_{\scriptscriptstyle 0}
ight)$$

donde X<sub>OBJ</sub> representa un canal de la segunda información de audio;

donde X<sub>EAO</sub> representa las señales de objeto de la primera información de audio; donde

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

$$\mathbf{M}_{EAO}^{\text{Energia}} = \begin{pmatrix} \frac{m_0^2 O L D_0}{O L D_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 O L D_i} \\ \vdots \\ \frac{m_{N_{EAO}-1}^2 O L D_{N_{EAO}-1}}{O L D_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 O L D_i} \end{pmatrix}$$

20

10

donde  $m_0$  a  $m_{NEAO}$  <sup>-1</sup> son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio;

donde OLD<sub>i</sub> son unos valores de diferencia de nivel de objeto asociados a los objetos de audio del primer tipo de objeto de audio;

donde  $OLD_L$  es un valor de diferencia de nivel de objeto común asociado a los objetos de audio del segundo tipo de objeto de audio; y

donde **A**<sup>EAO</sup> es una matriz de pre-representación EAO cuyas entradas describen un cartografiado de objetos de audio mejorados en los canales de una señal de objeto de audio mejorado XEAO;

donde las matrices  $\mathbf{M}$  Energía  $_{OBJ}$  y  $\mathbf{M}$  Energía  $_{EAO}$  se aplican a una representación  $\mathbf{d}_0$  de una señal de mezcla hacia abajo SAOC única.

4. Procedimiento para suministrar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y de una información paramétrica relativa al objeto, comprendiendo el procedimiento:

descomponer la representación de señal de mezcla hacia abajo, para suministrar una primera información de audio que describe un primer conjunto de uno o varios 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 varios objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y con ayuda de al menos una parte de la información paramétrica relativa al objeto; y

tratar la segunda información de audio en función de la información paramétrica relativa al objeto, para obtener una versión tratada de la segunda información de audio; y

combinar la primera información de audio con la versión tratada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;

en el cual la primera información de audio y la segunda información de audio se obtienen según

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Predicción}} \begin{bmatrix} l_0 \\ r_0 \\ -res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{bmatrix}$$

$$\mathbf{X}_{EAO} = \mathbf{A}^{EAO} \mathbf{M}_{EAO}^{ ext{Predicción}} egin{pmatrix} l_0 \\ \underline{r_0} \\ res_0 \\ \vdots \\ res_{N_{EAO}-1} \end{pmatrix}$$

donde

20

30

5

10

15

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

donde

donde X<sub>OBJ</sub> representa los canales de la segunda información de audio; 25

donde X<sub>EAO</sub> representa las señales de objeto de la primera información de audio;

donde D<sup>-1</sup> representa una matriz que es la inversa de una matriz de mezcla hacia abajo extendida;

donde C describe una matriz que representa una pluralidad de coeficientes de predicción de canal, c i.o, c i.t,

donde l<sub>0</sub> y r<sub>0</sub> representan los canales de la representación de la señal de mezcla hacia abajo:

donde **res**<sub>0</sub> a **res**<sub>NEAO</sub> -1 representan los canales residuales; y donde **A**<sup>EAO</sup> es una matriz de pre-representación EAO cuyas entradas describen un cartografiado de objetos de audio

mejorados en los canales de una señal de objeto de audio  $\mathbf{X}_{\mathsf{EAO}}$  mejorado; donde la matriz de mezcla hacia abajo inversa  $\mathbf{D}^{-1}$  se obtiene como la inversa de una matriz de mezcla hacia abajo extendida D que se define como

$$\tilde{\mathbf{D}} = \begin{pmatrix} 1 & 0 & | m_0 & \dots & m_{N_{EAO}-1} \\ 0 & 1 & | n_0 & \dots & n_{N_{EAO}-1} \\ \hline m_0 & n_0 & | -1 & \dots & 0 \\ \vdots & \vdots & | 0 & \ddots & \vdots \\ m_{N_{EAO}-1} & n_{N_{EAO}-1} & | 0 & \dots & -1 \end{pmatrix}$$

donde la matriz C se obtiene como

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \hline c_{0,0} & c_{0,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N_{EAO}-1,0} & c_{N_{EAO}-1,1} & 0 & \cdots & 1 \end{pmatrix}$$

donde  $m_0$  a  $m_{NEAO}^{-1}$  son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio:

donde no a  $n_{NEAO}^{-1}$  son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio:

donde los coeficientes de predicción c  $_{\it i,0}$  y c  $_{\it i,1}$  se calculan como

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

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

10

5

donde los coeficientes de predicción restringidos c  $_{j,0}$  y c  $_{j,1}$  se derivan de los coeficientes de predicción c  $_{j,0}$  y c  $_{j,1}$  con ayuda de un algoritmo de restricción, o bien donde los coeficientes de predicción c  $_{j,0}$  y c  $_{j,1}$  se utilizan como coeficientes de predicción c  $_{j,0}$  y c  $_{j,1}$ ;

donde las cantidades de energía P<sub>Lo</sub>,P<sub>Ro</sub>,P<sub>LoRo</sub>,P<sub>LoCoj</sub> y P<sub>RoCoj</sub> se definen como

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

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

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

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

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

donde los parámetros OLD<sub>L</sub>,OLD<sub>R</sub>, y IOC<sub>L,R</sub> corresponden a los objetos de audio del segundo tipo de objeto de audio y se definen según

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

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

$$IOC_{L,R} = \begin{cases} IOC_{0,1}, & N - N_{EAO} = 2, \\ 0, & \text{de otro modo.} \end{cases}$$

donde  $d_{0,i}$  y  $d_{1,i}$  son unos valores de mezcla hacia abajo asociados a los objetos de audio del segundo tipo de objeto de audio:

donde OLD<sub>i</sub> son unos valores de diferencia de nivel de objeto asociados a los objetos de audio del segundo tipo de objeto de audio;

donde N es el número total de objetos de audio;

donde N<sub>EAO</sub> es el número de objetos de audio del primer tipo de objeto de audio;

donde IOC<sub>0,1</sub> es un valor de correlación entre objetos asociado a un par de objetos de audio del segundo tipo de objeto de audio;

donde e<sub>ij</sub> y e<sub>L,R</sub> son unos valores de covariancia derivados de parámetros de diferencia de nivel de objeto y de los parámetros de correlación entre objetos; y

donde  $e_{ij}$  está asociado a un par de objetos de audio del primer tipo de objeto de audio y  $e_{L,R}$  está asociado a un par de objetos de audio del segundo tipo de objeto de audio.

5. Procedimiento para suministrar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y de una información paramétrica relativa al objeto, comprendiendo el procedimiento:

descomponer la representación de señal de mezcla hacia abajo, para suministrar una primera información de audio que describe un primer conjunto de uno o varios 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 varios objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y con ayuda de al menos una parte de la información paramétrica relativa al objeto; y

tratar la segunda información de audio en función de la información paramétrica relativa al objeto, para obtener una versión tratada de la segunda información de audio; y

combinar la primera información de audio con la versión tratada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;

30 en el cual la primera información de audio y la segunda información de audio se obtienen según

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

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

donde X<sub>OBJ</sub> representa los canales de la segunda información de audio;

donde X<sub>EAO</sub> representa las señales de objeto de la primera información de audio;

donde

5

20

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

donde m<sub>0</sub> a m<sub>NEAO</sub> -1 son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio:

donde al no a n<sub>NEAO</sub> -1 son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio;

donde OLDi son unos valores de diferencia de nivel de objeto asociados a los objetos de audio del primer tipo de objeto de audio:

donde OLD<sub>L</sub> y OLD<sub>R</sub> son unos valores de diferencia de nivel de objeto comunes asociados a los objetos de audio del segundo tipo de objeto de audio; y donde  $\mathbf{A}^{\text{EAO}}$  es una matriz de pre-representación EAO cuyas entradas describen un cartografiado de objetos de audio

10 mejorados en los canales de una señal de objeto de audio mejorado X<sub>EAO</sub>.

6. Procedimiento para suministrar una representación de señal de mezcla hacia arriba en función de una representación de señal de mezcla hacia abajo y de una información paramétrica relativa al objeto, comprendiendo el procedimiento:

descomponer la representación de señal de mezcla hacia abajo, para suministrar una primera información de audio que describe un primer conjunto de uno o varios 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 varios objetos de audio de un segundo tipo de objeto de audio en función de la representación de señal de mezcla hacia abajo y con ayuda de al menos una parte de la información paramétrica relativa al objeto; y

tratar la segunda información de audio en función de la información paramétrica relativa al objeto, para obtener una versión tratada de la segunda información de audio; y

combinar la primera información de audio con la versión tratada de la segunda información de audio, para obtener la representación de señal de mezcla hacia arriba;

en el cual la primera información de audio y la segunda información de audio se obtienen según 25

$$\mathbf{X}_{OBJ} = \mathbf{M}_{OBJ}^{\text{Energia}} \left( d_0 \right)$$

$$\mathbf{X}_{\scriptscriptstyle EAO} = \mathbf{A}^{\scriptscriptstyle EAO} \mathbf{M}_{\scriptscriptstyle EAO}^{\scriptscriptstyle (Energia)} \left(d_{\scriptscriptstyle 0}
ight)$$

donde X<sub>OBJ</sub> representa un canal de la segunda información de audio;

donde X<sub>EAO</sub> representan señales de objeto de la primera información de audio;

donde

30

5

15

20

$$\begin{split} \mathbf{M}_{OBJ}^{\text{Energia}} &= \left( \sqrt{\frac{OLD_L}{OLD_L} + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} OLD_{N_{EAO}-1}}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} OLD_{N_{EAO}-1}}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_L + \sum_{i=0}^{N_{EAO}-1} m_i^2 OLD_i}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_0}} \right) \\ \mathbf{M}_{EAO}^{\text{Energia}} &= \left( \sqrt{\frac{m_0^2 OLD_0}{OLD_0} \right) \\ \mathbf{M}_{EAO}^{\text$$

donde  $m_0$  a  $m_{NEAO}$  <sup>-1</sup> son unos valores de mezcla hacia abajo asociados a los objetos de audio del primer tipo de objeto de audio;

- donde OLD<sub>i</sub> son unos valores de diferencia de nivel de objeto asociados a los objetos de audio del primer tipo de objeto de audio:
  - donde  $OLD_L$  es un valor de diferencia de nivel de objeto común asociado a los objetos de audio del segundo tipo de objeto de audio; y
- 10 donde **A**<sup>EAO</sup> es una matriz de pre-representación EAO cuyas entradas describen un cartografiado de objetos de audio mejorados en los canales de una señal de objeto de audio mejorado **X**<sup>EAO</sup>;
- donde las matrices  $\mathbf{M}$  <sup>Energía</sup><sub>OBJ</sub> y  $\mathbf{M}$  <sup>Energía</sup><sub>EAO</sub> se aplican a una representación d<sub>0</sub> de una señal de mezcla hacia abajo SAOC única.
  - **7.** Programa de ordenador para realizar el procedimiento según cualquiera de las reivindicaciones 4 a 6 cuando el programa de ordenador se ejecuta en un ordenador.

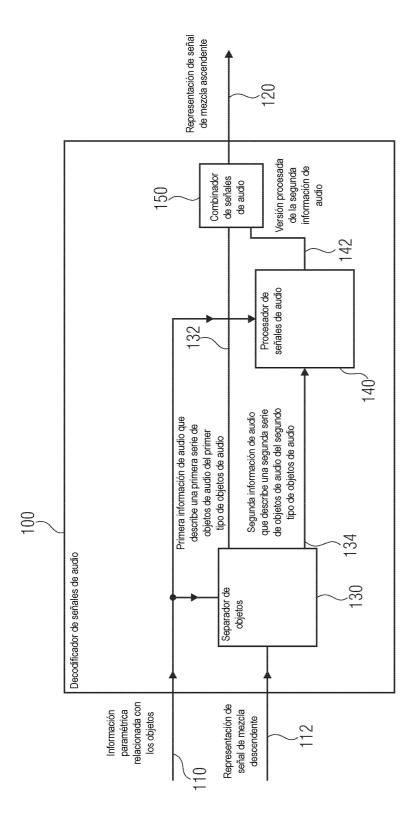
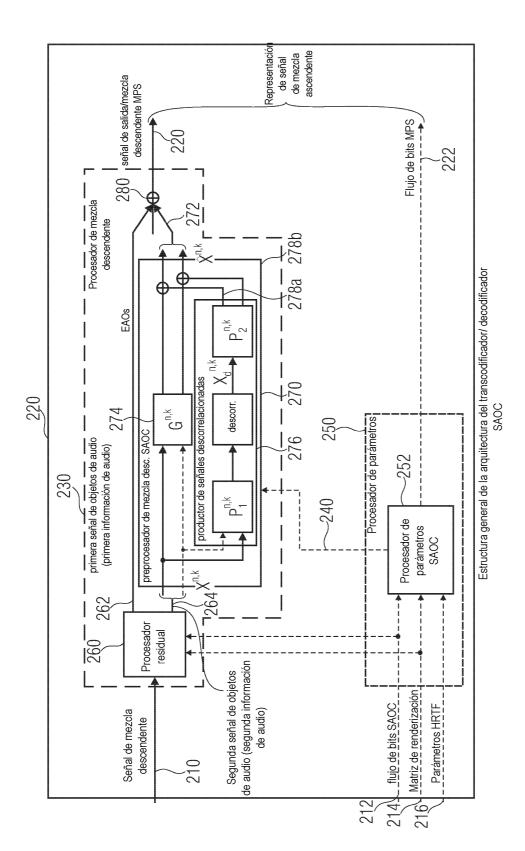


FIG. 1



**FIG. 2** 

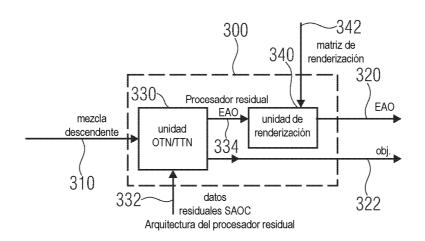
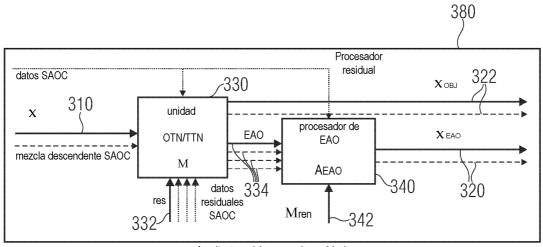


FIG. 3 A



Arquitectura del procesador residual

FIG. 3 B

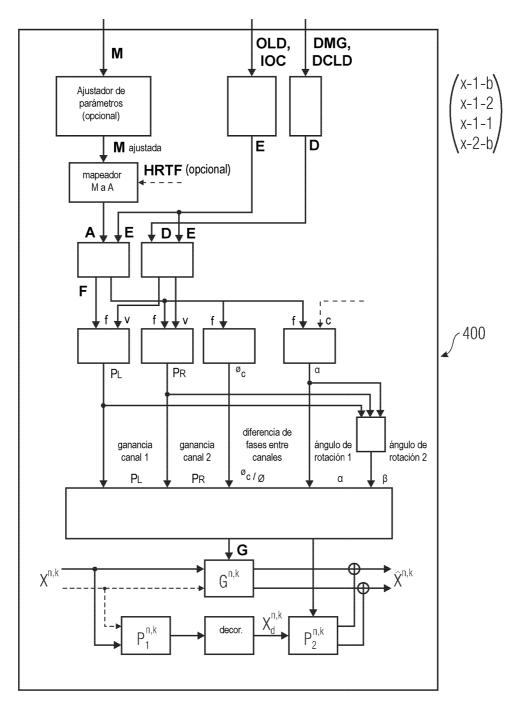


FIG. 4 A

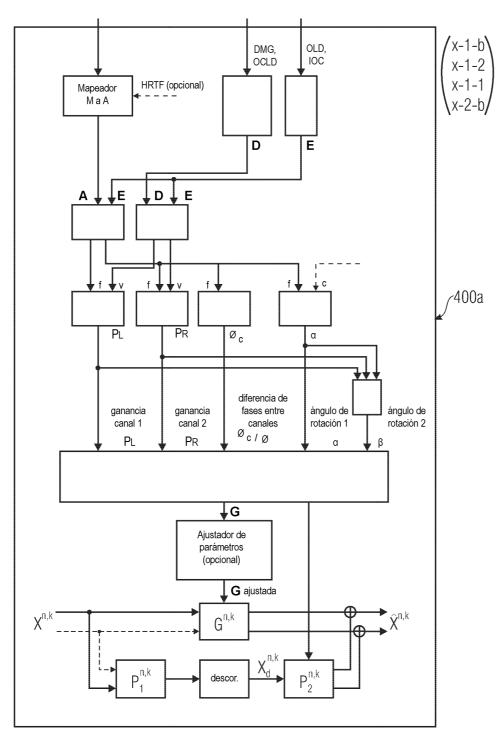


FIG. 4B

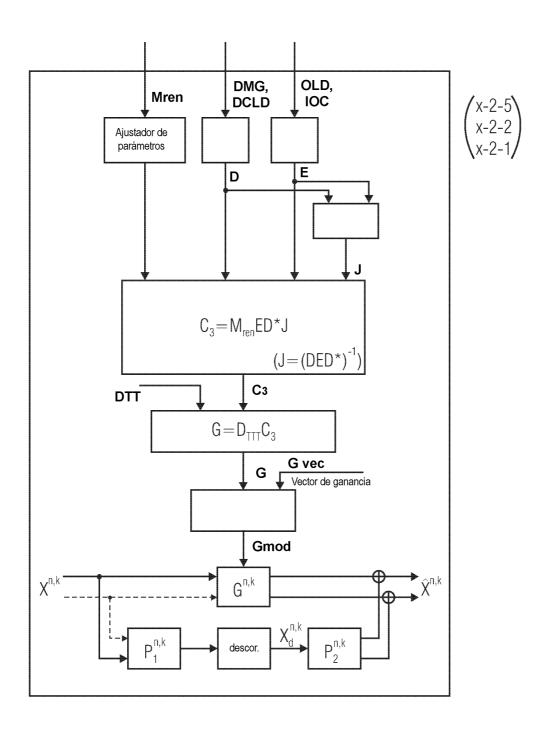


FIG. 4 C

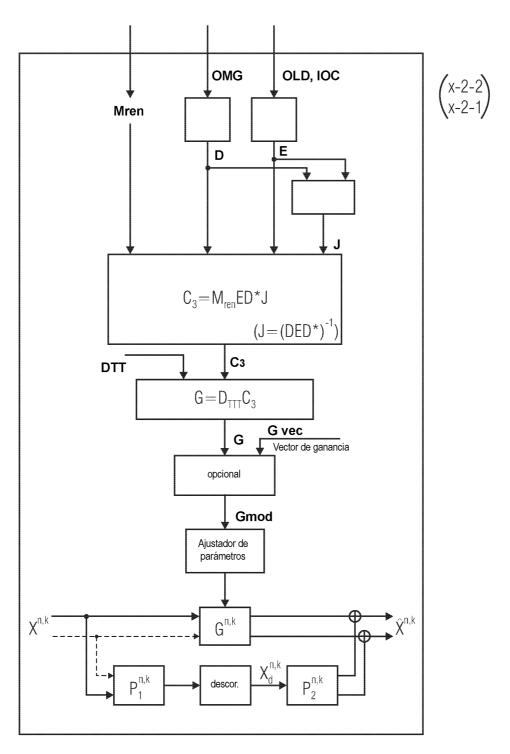


FIG. 4 D

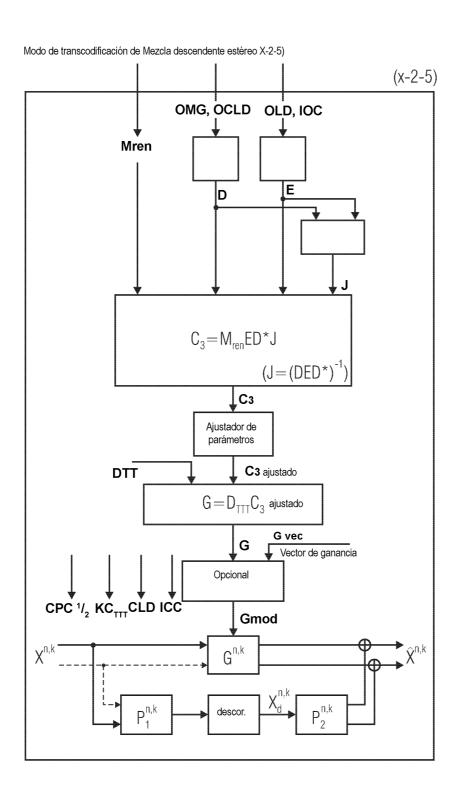


FIG. 4 E

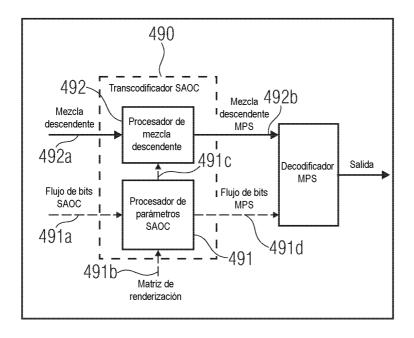


FIG. 4 F

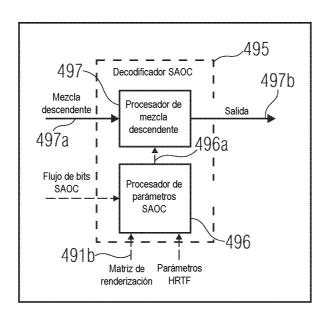


FIG. 4 G

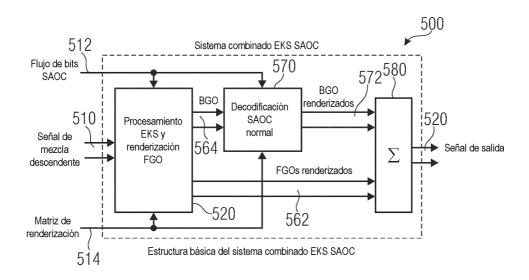


FIG. 5 A

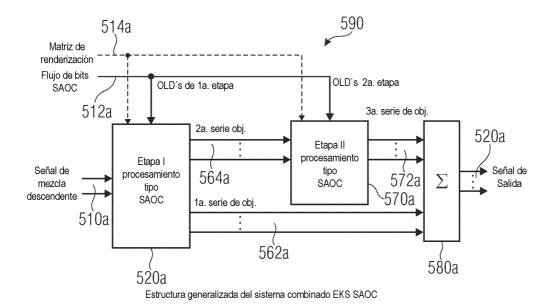


FIG. 5 B

Nombre codificador	Descripción
Referencia oculta	escena de audio c/renderiz. ideal (referencia oculta)
Anclaje inferior	anclaje inferior @ 3.5 kHz
SAOC normal	Modo de decodificación SAOC normal
EKS actual	Modo actual SAOC EKS
Sistema combinado	Sistema combinado (EKS + SAOC) (sistema propuesto)

FIG. 6 A

Sistema	Flujo de bits	Renderización Karaoke	Renderización Clásica
Sistema Combinado SAOC normal	SAOC A + datos res. A datos SAOC	X (X)	X X
EKS actual	SAOC B + datos res. B	X	-

FIG. 6 B

Elemento	Tipo	Matriz de renderización	Matriz de mezcla descendente
K- po p0 1	Karaoke	[0.6 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.038   [0.0 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.6 0.0 1.0 0.0 0.0 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.038]   0.6 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.038]	[0.0 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.6 0.6 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.6]
	Clásico	[0.6 0.4 0.0 1.0 0.0 0.6 0.0 1.0 0.6 0.0 0.0 0.0 0.4 0.0 1.0 0.0 0.6 1.2 0.6]	
R	Karaoke	[1.0 0.0 0.707 0.707 0.000 0.038 0.000 0.707 0.000 0.707 0.038]	707.0 0.00 0.707 0.707 0.000
ock	Clásico	[1.0 0.0 0.6 0.0 0.707 0.000 0.4 0.0 0.000 0.707 0.6]	0.0 1.0 0.707 0.000 0.707 0.707]

FIG. 6 C

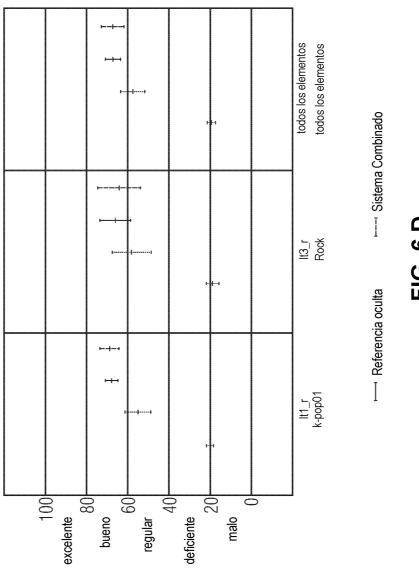
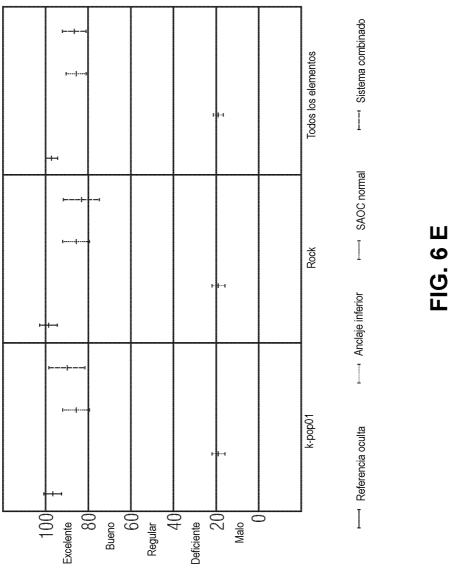


FIG. 6 D



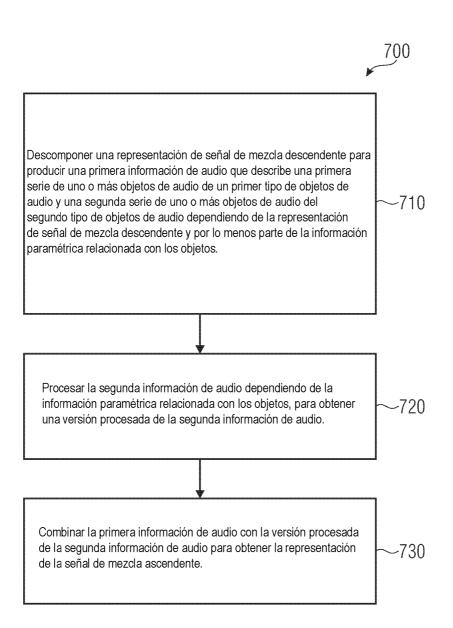


FIG. 7

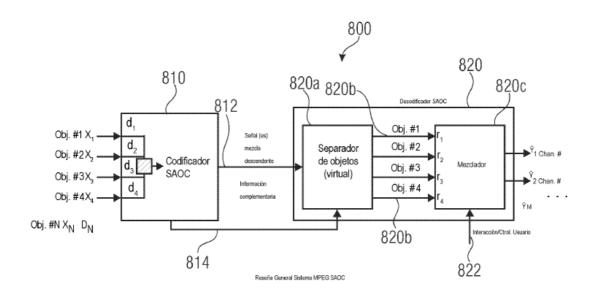


FIG. 8

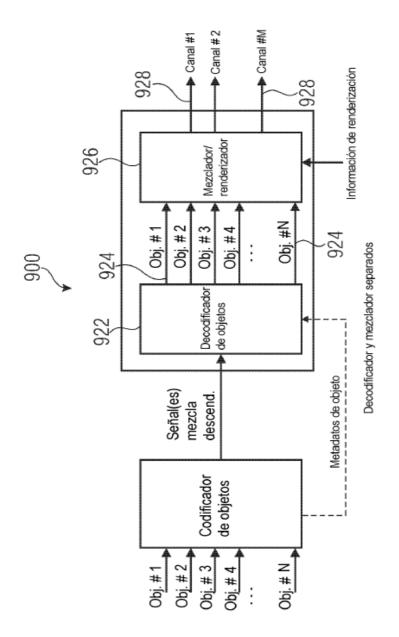


FIG. 9 A

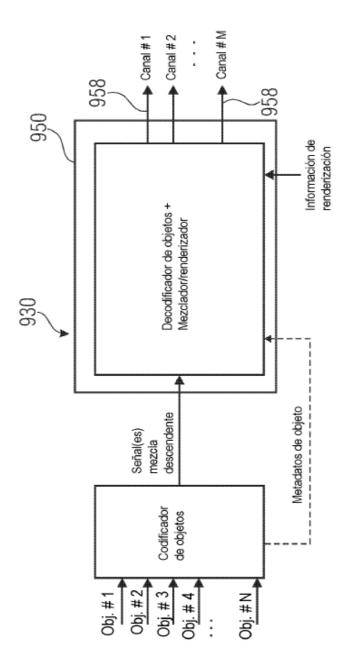


FIG. 9 B

Mezclador y Decodificador integrados

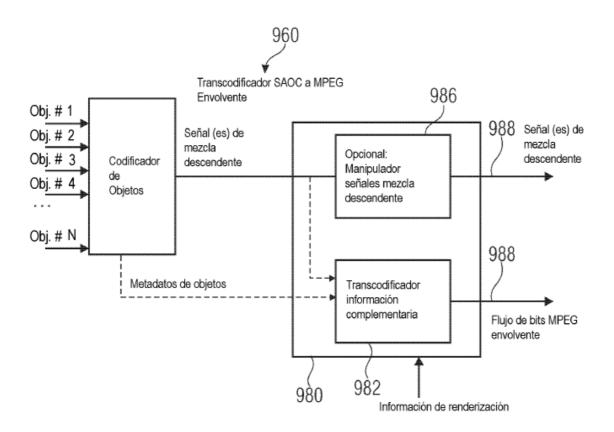
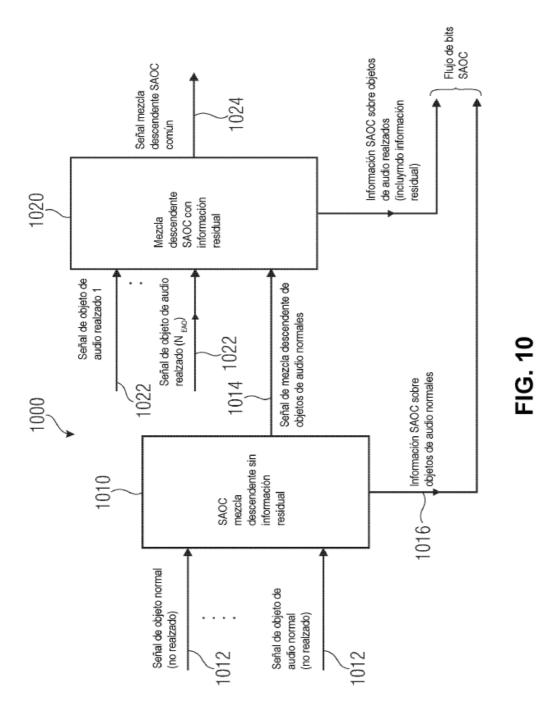


FIG. 9 C



80