(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 521 715**

(51) Int. Cl.:

***G10L 19/008*** (2013.01)

*G10L 19/20* (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **28.04.2010 E 10716830 (4)**

(97) Fecha y número de publicación de la concesión europea: **10.09.2014 EP 2425427**

(54) Título: **Aparato para suministrar uno o más parámetros ajustados para un suministro de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente, decodificador de señal de audio, transcodificador de señal de audio, procedimiento y programa de computación que utiliza información paramétrica relacionada con el objeto**

(30) Prioridad:

**28.04.2009 US 173456 P**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**13.11.2014**

(73) Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (50.0%)**
**Hansastrasse 27c**
**80686 München, DE y**
**DOLBY INTERNATIONAL AB (50.0%)**

(72) Inventor/es:

**HERRE, JÜRGEN;**
**HOELZER, ANDREAS;**
**TERENTIEV, LEONID;**
**KASTNER, THORSTEN;**
**FALCH, CORNELIA;**
**PURNHAGEN, HEIKO;**
**ENGDEGARD, JONAS y**
**RIDDERBUSCH, FALCO**

(74) Agente/Representante:

**PONTI SALES, Adelaida**

**ES 2 521 715 T3**

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

## DESCRIPCIÓN

Aparato para suministrar uno o más parámetros ajustados para un suministro de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente, decodificador de señal de audio, transcodificador de señal de audio, procedimiento y programa de computación que utiliza información paramétrica relacionada con el objeto

Campo Técnico

**[0001]** Las realizaciones de acuerdo con la invención se relacionan con un aparato para suministrar uno o más parámetros ajustados para la producción de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente e información paramétrica relacionada con el objeto.

**[0002]** Otra realización de acuerdo con la invención se relaciona con un decodificador de señales de audio.

**[0003]** Otra realización de acuerdo con la invención se relaciona con un transcodificador de señales de audio.

**[0004]** Algunas realizaciones más de acuerdo con la invención se relacionan con un procedimiento para producir uno o más parámetros ajustados.

**[0005]** Algunas realizaciones más se relacionan con un procedimiento para producir, como representación de la señal de mezcla ascendente, una pluralidad de canales de audio de mezcla ascendente sobre la base de una representación de señal de mezcla descendente, información paramétrica relacionada con el objeto y una información de representación pretendida.

**[0006]** Otra de las realizaciones se relaciona con un procedimiento para producir, como representación de la señal de mezcla ascendente, una representación de señal de mezcla descendente y una información paramétrica relacionada con los canales sobre la base de una representación de señal de mezcla descendente, información paramétrica relacionada con el objeto y una información de representación pretendida.

**[0007]** Algunas realizaciones más se relacionan con programas de computación correspondientes.

**[0008]** Algunas realizaciones más de acuerdo con la invención se relacionan con procedimientos, aparato y programas de computación para el procesamiento de señales de audio que evitan la distorsión.

Antecedentes de la Invención

**[0009]** En la técnica del procesamiento de audio, transmisión de audio y almacenamiento de audio, existe un deseo creciente de manejar el contenido de múltiples canales a fin de mejorar la impresión auditiva. El uso del contenido de audio multicanal acarrea considerables mejoras para el usuario. Por ejemplo, se puede obtener una impresión auditiva tridimensional, lo que trae aparejada una mayor satisfacción del usuario en aplicaciones de entretenimiento. Sin embargo, los contenidos de audio multicanal también son útiles en entornos profesionales, por ejemplo en aplicaciones de conferencias telefónicas, ya que se puede mejorar la inteligibilidad del parlante utilizando una reproducción de audio en múltiples canales.

**[0010]** Sin embargo, también es conveniente contar con una buena relación entre calidad de audio y requisitos de flujo de bits a fin de evitar una carga de recursos excesiva causada por aplicaciones multicanal.

**[0011]** Últimamente, se han propuesto técnicas paramétricas para la transmisión con eficiencia de flujo de bits y/o almacenamiento de escenas de audio que contienen múltiples objetos de audio, por ejemplo, la Binaural Cue Coding (Codificación Binaural de Monitoreo de Señales) (Tipo I) (véase, por ejemplo la referencia [BCC]), Joint Source Coding (Codificación Conjunta de Canales Fuente) (véase, por ejemplo, la referencia [JSC]), y Codificación de Objetos de Audio Espacial MPEG (MPEG Spatial Objeto de audio Coding) (SAOC) (véase, por ejemplo, las referencias [SAOC1], [SAOC2]).

**[0012]** Estas técnicas apuntan a reconstruir perceptualmente la escena de audio de salida en lugar de hacerlo por la paridad de una forma de onda.

**[0013]** La Fig. 8 ilustra una reseña de sistema de ese tipo de sistema (en este caso: MPEG SAOC). El sistema MPEG SAOC 800 expuesto 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 $x_1$ a $x_N$, que pueden estar representadas, por ejemplo, en forma de señales en el dominio del tiempo o en forma de señales en el dominio de tiempo–frecuencia (por ejemplo, en forma de una serie de coeficientes de transformada del tipo transformada de Fourier, o en forma de señales de subbandas QMF). El codificador SAOC 810 también recibe, por lo general, coeficientes de mezcla

descendente $d_1$ a $d_N$, que están asociados a las señales de objeto $x_1$ a $x_N$. Pueden existir series separadas de coeficientes de mezcla descendente para cada canal de la señal de mezcla descendente. El codificador SAOC 810 está configurado típicamente para obtener un canal de la señal de mezcla descendente mediante la combinación de 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 la separación (o tratamiento separado) de las señales de objeto del lado del decodificador SAOC 820, el codificador SAOC 810 suministra tanto una o más señales de mezcla descendente (designadas 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 al procesamiento específico de objetos del lado del decodificador.

**[0014]** El decodificador SAOC 820 está configurado para recibir tanto dichas una o más señales de mezcla descendente 812 como información complementaria 814. Además, el decodificador SAOC 820 está configurado típicamente para recibir una información de interacción con el usuario y/o una información de control del usuario 822, que describe una configuración pretendida de representación. Por ejemplo, la información de interacción con el usuario/ control del usuario 822 puede describir una configuración de los parlantes y la colocación espacial conveniente de los objetos que producen las señales de objeto $x_1$ a $x_N$.

**[0015]** El decodificador SAOC 820 está configurado para producir, por ejemplo, una pluralidad de señales de canales decodificadas de mezcla descendente $\hat{y}_1$ a $\hat{y}_M$. Las señales de canales de mezcla ascendente pueden estar asociadas, por ejemplo a parlantes individuales de una disposición de representación 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 $x_1$ a $x_N$ 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 $x_1$ a $x_N$, 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 $\hat{y}_1$ a $\hat{y}_M$. 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 representación (que también se denominan coeficientes de representación), que determinan la contribución de las señales de objetos reconstruidas 822 individuales a las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$.

**[0016]** 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 $\hat{y}_1$ a $\hat{y}_M$. 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 820.

**[0017]** 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 924 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/representación, lo que da lugar a una separación de la funcionalidad de decodificación de objetos de la funcionalidad de mezcla/representación, aunque trae aparejada una complejidad informática relativamente elevada.

**[0018]** 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/representación, donde los parámetros para dicho proceso conjunto de mezcla ascendente dependen tanto de la información complementaria relacionada con objetos y la información sobre la representación. 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.

**[0019]** 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.

**[0020]** 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 (Surround) 980, en lugar de un decodificador SAOC.

**[0021]** 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 representación. El transcodificador de información complementaria también está configurado para proporcionar una información complementaria sobre MPEG Envolvente (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, tomando en cuenta la información sobre la representación y, opcionalmente, la información sobre el contenido de dichas una o más señales de mezcla descendente.

**[0022]** 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, por lo que la representación de señal de mezcla descendente 988 de salida del transcodificador de SAOC a MPEG Envolvente 980 es 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 representación.

**[0023]** 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 representación 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.

**[0024]** 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 el objeto. 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.

**[0025]** 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 $x_1$ a $x_N$ como parte del procesamiento del codificador SAOC. Para una mezcla descendente mono, los coeficientes están indicados por $d_1$ a $d_N$. 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 $\hat{y}_1$ a $\hat{y}_M$) utilizando una matriz de representación. Para una salida mono, los coeficientes de la matriz de representación están expresados por $r_1$ a $r_N$

- Efectivamente, 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.

**[0026]** 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) 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 representación 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 representación 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:

**[0027]** Por cada objeto de sonido transmitido, se puede ajustar su nivel relativo y (en el caso de la representación no mono) la posición espacial de la representación. 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 = –30deg).

**[0028]** Sin embargo, se ha encontrado que la elección de parámetros para el lado del decodificador para la provisión de la representación de señal de mezcla ascendente (por ej., las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$) trae aparejadas degradaciones audibles.

**[0029]** En vista de esta situación, el objetivo de la presente invención consiste en crear un concepto que da lugar a la reducción o incluso a la eliminación de la distorsión audible al suministrar una representación de la señal de mezcla ascendente (por ejemplo, en forma de señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_M$).

Síntesis de la invención

**[0030]** Este problema se soluciona mediante un aparato para suministrar uno o más parámetros ajustados para la producción de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente e información paramétrica relacionada con el objeto de acuerdo con la reivindicación 1, un decodificador de señales de audio de acuerdo con la reivindicación 24, un transcodificador de señales de audio de acuerdo con la reivindicación 25, procedimientos de acuerdo con las reivindicaciones 26, 27 y 28, un codificador de señales de audio de acuerdo con la reivindicación 29, un procedimiento de acuerdo con la reivindicación 31, un flujo de bits de audio de acuerdo con la reivindicación 32 y un programa de computación de acuerdo con la reivindicación 34.

**[0031]** Una realización de acuerdo con la invención crea un aparato para suministrar uno o más parámetros ajustados para la producción de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente e información paramétrica relacionada con el objeto. El aparato comprende un ajustador de parámetros (por ejemplo, un regulador de coeficientes de representación) configurado para recibir uno o más parámetros de entrada (por ejemplo, un coeficiente de representación o una descripción de una matriz de representación conveniente) y producir, sobre la base de éste, uno o más parámetros ajustados. El ajustador de parámetros está configurado para producir uno o más parámetros ajustados dependiendo de uno o más parámetros de entrada y la información paramétrica relacionada con el objeto (por ejemplo, dependiendo de uno o más coeficientes de mezcla descendente y/o uno o más valores de diferencia de nivel de los objetos y/o uno o más valores de inter correlación con los objetos) de tal modo que se reduce la distorsión de la representación de señal de

mezcla ascendente, que sería causada por el uso de parámetros no óptimos, por lo menos en el caso de los parámetros de entrada que se desvían de los parámetros óptimos en un grado mayor que una desviación predeterminada.

**[0032]** Esta realización de acuerdo con la invención se basa en la idea de que las distorsiones de señales de audio que son provocadas por parámetros de entrada incorrectamente elegidos se pueden reducir produciendo parámetros ajustados para la provisión de la representación de señal de mezcla ascendente y que la provisión de los parámetros ajustados se pueden realizar con buena precisión tomando en cuenta la información paramétrica relacionada con el objeto. Se ha descubierto que el uso de la información paramétrica relacionada con el objeto permite obtener una medida estimada de distorsiones audibles, que serían provocadas por el uso de los parámetros incorrectos, lo que a su vez permite la producción de parámetros ajustados que son adecuados para mantener las distorsiones audibles dentro de un rango predeterminado o que son adecuados para reducir las distorsiones audibles al compararlos con los parámetros de entrada. La información relacionada con el objeto describe, por ejemplo, las características de los objetos de audio y/o da información acerca del procesamiento de los objetos del lado del codificador.

**[0033]** En consecuencia, se pueden reducir, o incluso evitar, las distorsiones desagradables y, con frecuencia, molestas de las señales de audio que serían causadas por el uso de parámetros inapropiados (por ejemplo, coeficientes de representación inapropiados), mediante la provisión de uno o más parámetros ajustados, donde la consideración de la información paramétrica relacionada con el objeto para el ajuste de los parámetros contribuye a garantizar una reducción y/o limitación eficaz de las distorsiones de las señales de audio al dar lugar a una estimación comparativamente confiable de las distorsiones audibles.

**[0034]** Se ha descubierto que la elección de los parámetros de representación inapropiados trae consigo una degradación significativa (y con frecuencia audible) de una representación de la señal de mezcla ascendente, que se obtiene utilizando dichos parámetros de representación incorrectamente elegidos. Además, se ha descubierto que se pueden ajustar los parámetros de representación dependiendo de la información paramétrica relacionada con el objeto, puesto que la información paramétrica relacionada con el objeto da lugar a una estimación de las distorsiones, que se introducirían merced a una elección de los parámetros de representación (que se pueden definir por los parámetros de entrada).

**[0035]** En una realización preferida, el ajustador de parámetros está configurado para obtener uno o más valores límite de los parámetros de representación que dependen de la información paramétrica relacionada con el objeto y una información de mezcla descendente que describe una contribución de las señales de objetos de audio a la representación de la señal de mezcla descendente, de manera que una distorsión métrica esté dentro de un rango predeterminado de valores de parámetros de representación que obedecen a límites definidos por los valores límite de los parámetros de representación. En este caso, el ajustador de parámetros está configurado para obtener los parámetros de representación reales dependiendo de los parámetros de representación convenientes y dicho uno o más valores límite de los parámetros de representación, de tal manera que los parámetros de representación reales obedecen los límites definidos por los valores límite de los parámetros de representación. La computación de los valores límite de los parámetros de representación constituye un mecanismo de computación sencillo y confiable para garantizar que las distorsiones audibles estén dentro de un rango permisible de acuerdo con una métrica de distorsión.

**[0036]** En una realización preferida, el ajustador de parámetros está configurado para obtener dichos uno o más valores límite de los parámetros de representación de tal manera que una contribución relativa de una señal de objeto en una superposición renderizada de una pluralidad de señales de objeto, renderizadas utilizando un parámetro de representación que obedece a dichos uno o más valores límite de los parámetros de representación, difiera de una contribución relativa de la señal del objeto en una señal de mezcla descendente en no más de una diferencia predeterminada. Se ha descubierto que las distorsiones son, por lo general, suficientemente pequeñas si la contribución de una señal de objeto en una superposición renderizada de las señales de objeto es similar a la contribución de la señal de objeto en una señal de mezcla descendente, en tanto que una diferencia importante de dichas contribuciones por lo general trae aparejadas distorsiones audibles. Esto se debe al hecho de que un fuerte cambio del nivel (relativo) de una señal de objeto en comparación con el nivel (relativo) de la señal de objeto en la representación de la señal de mezcla descendente con frecuencia trae consigo perturbaciones, puesto que a menudo no es posible separar las señales de objeto de diferentes objetos de audio de manera ideal. En consecuencia, se ha descubierto que se obtienen resultados favorables ajustando los parámetros de representación de tal manera que la contribución relativa de las señales de objeto sólo se modifique moderadamente por la elección de los parámetros de representación.

**[0037]** En otra realización, el ajustador de parámetros está configurado para obtener dichos uno o más valores límite de los parámetros de representación de tal manera que una medida de la distorsión que describe una coherencia entre una señal de mezcla descendente descripta por la representación de la señal de mezcla descendente y una señal renderizada, renderizada utilizando dichos uno o más parámetros de representación obedeciendo a dichos

uno o más valores límite de los parámetros de representación, esté dentro de un rango predeterminado. Se ha descubierto que la elección de los parámetros de representación convenientes, que forman los parámetros de entrada del ajustador de parámetros, se debe realizar de tal manera que se mantenga una “similitud” suficiente entre la señal de mezcla descendente descripta por la representación de la señal de mezcla descendente y la señal renderizada, puesto que, de lo contrario, el riesgo de obtener perturbaciones audibles en el proceso de mezcla ascendente es bastante elevado.

**[0038]** En otra de las realizaciones preferidas, el ajustador de parámetros está configurado para computar una combinación lineal entre un cuadrado de un parámetro de representación conveniente (que puede formar el parámetro de entrada del ajustador de parámetros) y el cuadrado de un parámetro de representación óptima (que puede, por ejemplo, ser definido como un parámetro de representación que minimiza una métrica de distorsión), para obtener el parámetro de representación real (que puede ser la salida del aparato en forma de parámetro ajustado). En este caso, el ajustador de parámetros está configurado para determinar una contribución del parámetro de representación conveniente y del parámetro de representación óptimo a la combinación lineal dependiendo de un parámetro umbral predeterminado T y la métrica de distorsión, donde la métrica de distorsión describe una distorsión que sería causada por el uso de uno o más parámetros de representación pretendidos, en lugar de los parámetros de representación óptimos , para obtener la representación de la señal de mezcla ascendente sobre la base de la representación de la señal de mezcla descendente. Este concepto permite reducir la distorsión en una medida aceptable y al mismo tiempo mantener un impacto suficiente de los parámetros de representación pretendidos. De acuerdo con este concepto, se puede encontrar un compromiso razonablemente favorable entre los parámetros de representación óptimos y los parámetros de representación pretendidos, tomando en cuenta un grado deseado de limitación de las distorsiones audibles

**[0039]** En una realización preferida, el ajustador de parámetros está configurado para suministrar uno o más parámetros ajustados dependiendo de una medida informática de degradación perceptual, por lo cual se limita la distorsión evaluada perceptualmente de la representación de la señal de mezcla ascendente causada por el uso de parámetros no óptimos y representada por la medida informática de degradación perceptual. De esta manera, se puede obtener el ajuste de los parámetros de acuerdo con la impresión auditiva, evitando así una impresión auditiva inadmisiblemente deficiente y a la vez brindando suficiente flexibilidad para ajustar los parámetros de acuerdo con los deseos de un usuario.

**[0040]** En una realización preferida, el ajustador de parámetros está configurado para recibir una información sobre propiedades de un objeto que describe las propiedades de una o más señales de objeto originales, que forman la base para una señal de mezcla descendente descripta por la representación de señal de mezcla descendente. En este caso, el ajustador de parámetros está configurado para considerar la información sobre propiedades del objeto a fin de producir los parámetros ajustados de tal manera que se reduzca la distorsión de la representación de la señal de mezcla ascendente con respecto a las propiedades de las señales de objeto incluidas en la representación de señal de mezcla ascendente, por lo menos en el caso de los parámetros de entrada que se desvían de los parámetros óptimos en más de una desviación predeterminada. Esta realización de acuerdo con la invención se basa en el hallazgo de que se pueden utilizar las propiedades de dichas una o más señales de objeto originales para evaluar si los parámetros de entrada son apropiados o deben ser ajustados, puesto que es conveniente producir la señal de mezcla ascendente de tal manera que las características de la señal de mezcla ascendente están relacionadas con las propiedades de dichas una o más señales de objeto originales, porque de lo contrario la impresión perceptual se degradaría significativamente en muchos casos.

**[0041]** En una realización preferida, el ajustador de parámetros está configurado para recibir y considerar, como información de propiedades del objeto, una información de tonalidad de la señal del objeto, a fin de producir dicho uno o más parámetros ajustados. Se ha encontrado que la tonalidad de las señales de objeto es una cantidad que tiene un impacto significativo sobre la impresión perceptual, y que la elección de los parámetros que cambian significativamente la impresión de tonalidad debe evitarse a fin de obtener una buena impresión auditiva.

**[0042]** En una realización preferida, el ajustador de parámetros está configurado para estimar una tonalidad de una señal de mezcla ascendente renderizada de manera ideal dependiendo de la información de tonalidad de la señal de objeto recibida y una información de potencia del objeto recibida. En ese caso, el ajustador de parámetros está configurado para suministrar dicho uno o más parámetros ajustados a fin de reducir la diferencia entre la tonalidad estimada y la tonalidad de una señal de mezcla ascendente obtenida utilizando dichos uno o más parámetros ajustados e comparación una diferencia entre la tonalidad estimada y una tonalidad de una señal de mezcla ascendente obtenida con el uso de parámetros de entrada, o para mantener una diferencia entre la tonalidad estimada y una tonalidad de una señal de mezcla ascendente obtenida utilizando dichos uno o más parámetros ajustados dentro de un rango predeterminado. Utilizando este concepto, se puede obtener una media de degradación de una impresión auditiva con alta eficiencia informática, lo que da lugar a un ajuste apropiado de los parámetros de representación.

**[0043]** En una realización preferida, el ajustador de parámetros está configurado para ejecutar un ajuste de variante en tiempo y frecuencia de los parámetros de entrada. En consecuencia, el ajuste de los parámetros de entrada para obtener parámetros ajustados, se puede realizar solo durante aquellos intervalos de tiempo o regiones de frecuencia en los cuales el ajuste en realidad trae aparejado una mejora de la impresión auditiva o evita una degradación significativa de la impresión auditiva.

**[0044]** En otra más de la de las realizaciones preferidas, el ajustador de parámetros está configurado también para considerar la representación de señal de mezcla descendente para suministrar dichos uno o más parámetros ajustados. Tomando en cuenta la representación de señal de mezcla descendente, se puede obtener un cálculo aún más preciso de la distorsión posible de la distorsión auditiva.

**[0045]** En una realización preferida, el ajustador de parámetros está configurado para obtener una medida de la distorsión total, que es una combinación de las medidas de distorsión que describen una pluralidad de tipos de perturbaciones. En este caso, el ajustador de parámetros está configurado para obtener la medida de distorsión total de tal manera que la medida de distorsión total sea una medida de las distorsiones que serían causadas por el uso de uno o más de los parámetros de representación de entrada en lugar de los parámetros de representación óptimos para obtener la representación de la señal de mezcla ascendente sobre la base de la representación de la señal de mezcla descendente. Al combinar una pluralidad de medidas de distorsión que describen una pluralidad de tipos de perturbaciones, se genera un mecanismo bien controlado para ajustar la impresión auditiva.

**[0046]** Otra realización de acuerdo con la invención da origen a un decodificador de señales de audio para suministrar, como representación de una señal de mezcla ascendente, una pluralidad de canales de audio con mezcla ascendente sobre la base de una representación de una señal de mezcla descendente, una información paramétrica relacionada con el objeto y una información de representación deseada. El decodificador de señales de audio comprende un mezclador ascendente configurado para obtener los canales de audio de mezcla ascendente sobre la base de la representación de la señal de mezcla descendente y dependiendo de la información paramétrica relacionada con el objeto y de una información de representación real que describe una asignación de una pluralidad de señales de objetos de audio descritos por la información paramétrica relacionada con el objeto a las señales de audio con mezcla ascendente. El decodificador de señales de audio también comprende un aparato para producir uno o más parámetros ajustados, como se describiera anteriormente. El aparato para producir uno o más parámetros ajustados está configurado para recibir la información de representación deseada en forma de uno o más parámetros de entrada y para presentar dichos uno o más parámetros ajustados como información de representación real. El aparato para producir dichos uno o más parámetros ajustados está también configurado para producir uno o más parámetros ajustados de manera tal que se reduzcan la distorsiones de los canales de audio con mezcla ascendente causadas por el uso de los parámetros de representación reales, que se desvían de los parámetros de representación óptimos , por lo menos para los parámetros de representación deseados que se desvían de los parámetros de representación óptimas en más de una desviación predeterminada.

**[0047]** El uso del aparato para producir uno o más parámetros ajustados en un decodificador de señales de audio permite evitar la generación fuertes distorsiones audibles, que pueden ser causadas por la ejecución de la decodificación de audio con información de representación deseada elegida de manera no apropiada.

**[0048]** Una realización de acuerdo con la invención presenta un transcodificador de señales de audio para producir, como representación de señal de mezcla ascendente, una información paramétrica relacionada con los canales, sobre la base de la representación de una señal de mezcla descendente, una información paramétrica relacionada con los objetos y una información de representación deseada. El transcodificador de señales de audio comprende un transcodificador de información complementaria configurado para obtener la información paramétrica relacionada con los canales sobre la base de la representación de señal de mezcla descendente y dependiendo de la información paramétrica relacionada con los objetos y una información de representación real que describe una asignación de una pluralidad de señales de objeto de objetos de audio descritos por la información paramétrica relacionada con los objetos a los canales de audio con mezcla ascendente. El decodificador de señales de audio comprende asimismo un aparato para producir uno o más parámetros ajustados, como se describiera anteriormente. El aparato para producir uno o más parámetros ajustados está configurado para recibir la información de representación deseada en forma de uno o más parámetros de entrada y para suministrar dichos uno o más parámetros ajustados en forma de información de representación real. Además, el aparato para producir dichos uno o más parámetros ajustados está configurado para producir dichos uno o más parámetros ajustados de manera tal que se reduzcan las distorsiones de los canales de audio con mezcla ascendente representadas por la información paramétrica relacionada con los canales (en combinación con la información de señales con mezcla descendente), que son causadas por el uso de los parámetros de representación reales que se desvían de los parámetros de representación óptimas, por lo menos en el caso de los parámetros de representación deseados que se desvían de los parámetros de representación óptimas por más de una desviación predeterminada. Se ha descubierto que el concepto de producir parámetros ajustados también es muy adecuado para usan en combinación un transcodificador de señales de audio.

**[0049]** Otra realización de acuerdo con la invención se refiere a un procedimiento para suministrar uno o más parámetros ajustados según la reivindicación 25. Este procedimiento se basa en las mismas ideas claves que el aparato antes descrito.

**[0050]** Otras realizaciones de acuerdo con la invención crean un programa de computación para implementar los procedimientos antes descritos.

Breve Descripción de las Figuras

**[0051]** A continuación se describen realizaciones de acuerdo con la invención tomando como referencia las figuras adjuntas, en las cuales:

la Figura 1 ilustra un diagrama esquemático de bloque de un aparato para suministrar uno o más parámetros ajustados para la producción de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente y una información paramétrica relacionada con el objeto;

la Figura 2 ilustra un diagrama esquemático de bloque de un sistema MPEG SAOC, de acuerdo con una realización de la invención;

la Figura 3 ilustra una diagrama esquemático de bloque de un sistema MPEG SAOC, de acuerdo con otra realización de la invención;

la Figura 4 ilustra una representación esquemática de la contribución de señales de objeto a la señal de mezcla descendente y a una señal mezclada;

la Figura 5a ilustra una diagrama esquemático de bloque de un transcodificador de SAOC a MPEG de mezcla descendente mono de acuerdo con una realización de la invención;

la Figura 5b ilustra un diagrama esquemático de bloque de un transcodificador SAOC a MPEG basado en mezcla descendente estéreo, de acuerdo con una realización de la invención;

la Figura 6 ilustra una representación esquemática de bloque de de un codificador de señales de audio de acuerdo con una realización de la invención;

la Figura 7 ilustra una representación esquemática de un flujo de bits de audio de acuerdo con una realización de la invención;

la Figura 8 ilustra un diagrama esquemático de bloque de un sistema MPEG SAOC de referencia;

la Figura 9a ilustra un diagrama esquemático de bloque de un sistema SAOC de referencia que utiliza un decodificador y mezclador separados;

la Figura 9b ilustrar un diagrama esquemático de bloque de un sistema SAOC de referencia que utiliza un decodificador y mezclador integrado; y

la Figura 9c ilustra una diagrama esquemático de bloque de un sistema SAOC de referencia que utiliza un transcodificador de SAOC a MPEG.

Descripción Detallada de las Realizaciones

**1.** Aparato para suministrar uno o más parámetros ajustados, de acuerdo con la Figura 1.

**[0052]** A continuación, se describe un aparato 100 para suministrar uno o más parámetros ajustados para la producción de una representación de señal de mezcla ascendente sobre la base de una representación de señal de mezcla descendente y una información paramétrica relacionada con los objetos tomando como referencia la Figura 1. La Figura 1 ilustra un diagrama esquemático de bloque de ese aparato 100, que está configurado para recibir uno o más parámetros de entrada 110. Los parámetros de entrada 110 pueden ser, por ejemplo, parámetros de representación convenientes. El aparato 100 también está configurado para suministrar, sobre la base de estos, uno o más parámetros ajustados 120. Los parámetros ajustados pueden ser, por ejemplo, parámetros de representación ajustados. El aparato 100 está configurado además para recibir una información paramétrica relacionada con los objetos 130. La información paramétrica relacionada con los objetos 130 puede ser, por ejemplo, una información de diferencia de niveles entre objetos y/o una información de correlación entre objetos que describe una pluralidad de objetos. El aparato100 comprende un ajustador de parámetros 140, que está configurado para recibir uno o más parámetros de entrada 110, y para suministrar sobre la base de estos, uno o más parámetros ajustados 120. El

ajustador de parámetros 140 está configurado para producir dichos uno o más parámetros ajustados 120 dependiendo de dichos uno o más parámetros de entrada 110 y la información paramétrica relacionada con los objetos 130, por lo que se reduciría la distorsión de una señal de mezcla ascendente que sería causada por el uso de parámetros no óptimos (por ejemplo, dichos uno o más parámetros de entrada 110) en un aparato para suministrar una representación de señal de mezcla ascendente sobre la base una representación de señal de mezcla descendente y la información paramétrica relacionada con los objetos 130, por lo menos en el caso de los parámetros de entrada 110 que se desvían de los parámetros óptimos en más de una desviación predeterminada.

**[0053]** En consecuencia, el aparato 100 recibe dichos uno o más parámetros de entrada 110 y produce, sobre la base de éstos, dichos uno o más parámetros ajustados 120. Al suministrar dichos uno o más parámetros ajustados 120, el aparato 100 determina, de manera explícita o implícita, si el uso inalterado de uno o más parámetros de entrada 110 causaría distorsiones inadmisiblemente elevadas si se utilizara el uno o más parámetros de entrada 110 para controlar una provisión de una señal de mezcla ascendente sobre la base de una representación de una señal de mezcla descendente y una información paramétrica relacionada con los objetos 130. De esta manera, los parámetros ajustados 120 son por lo general más adecuados para ajustar dicho aparato para la provisión de la representación de señal de mezcla ascendente que dichos uno o más parámetros de entrada 110, por lo menos si dichos uno o más parámetros de entrada 110 se eligen de manera no ventajosa.

**[0054]** En consecuencia, el aparato 100 mejora por lo general la impresión perceptual de una representación de señal de mezcla ascendente, que es provista por un productor de representación de señal de mezcla ascendente dependiendo del uno o más parámetros ajustados 120. El uso de la información paramétrica relacionada con los objetos para ajustar los uno o más parámetros de entrada, para derivar los uno o más parámetros ajustados, han mostrado traer consigo buenos resultados, puesto que la calidad de la representación de señal de mezcla ascendente es típicamente satisfactoria si dichos uno o más parámetros ajustados 120 corresponden a la información paramétrica relacionada con los objetos 130, en tanto que los parámetros que violan la relación conveniente con la información paramétrica relacionada con los objetos 130 por lo general dan lugar a distorsiones audibles. La información paramétrica relacionada con los objetos puede, por ejemplo, parámetros de mezcla descendente, que describen una contribución de señales de objeto (de una pluralidad de objetos de audio) a dichas una o más señales de mezcla descendente. La información paramétrica relacionada con los objetos puede comprender también, por otro lado o además, parámetros de diferencias de niveles de los objetos y/o parámetros de correlación entre objetos, que describen las características de las señales de objetos. Se ha encontrado que tanto los parámetros que describen un procesamiento del lado del codificador de las señales de objeto, como las parámetros que describen características de los objetos de audio Ens., pueden considerar información útil para usar por el ajustador de parámetros 120.Sin embargo, el aparato100 puede utilizar de lo contrario o además información paramétrica relacionada con los objetos 130

**[0055]** Sin embargo, se debe tener en cuenta que el ajustador de parámetros 140 puede utilizar información adicional par suministrar dichos uno o más parámetros ajustados 120 sobre la base de dichos uno o más parámetros de entrada 110. Por ejemplo, el ajustador de parámetros 140 puede evaluar opcionalmente coeficientes de mezcla descendente, una o más señales de mezcla descendente o cualquier información adicional para mejorar aún la provisión de uno o más parámetros ajustados 120.

2. Sistema de acuerdo con la Figura 2

**[0056]** A continuación se describe el sistema MPEG SAOC 200 de la Figura 2 en forma detallada.

**[0057]** Para ofrecer una buena comprensión del sistema MPEG SAOC 200, se presenta una reseña general de las especificaciones deseadas del sistema y las consideraciones de diseño. Seguidamente, se presenta una reseña estructural del sistema. Más aún, se describe una pluralidad de métricas de distorsión SAOC, y la aplicación de estas métricas de distorsión SAOC para la limitación de las distorsiones. Además, se describen otras extensiones del sistema 200.

2.1 Consideraciones de Diseño del Sistema

**[0058]** Como se comentara anteriormente, las técnicas paramétricas para la transmisión eficiente en cuanto a la velocidad de transmisión de bits escenas de almacenamiento de audio que contienen múltiples objetos de audio son por lo general eficientes, tanto en términos de velocidad de transmisión de bits como en complejidad informática. Otras ventajas para el usuario de dicho sistema del lado de la recepción, incluyen la libertad de elegir una configuración de representación y su elección (mono, estéreo, envolvente, reproducción por auriculares virtual izada, y así sucesivamente) y la característica de interactividad del usuario: la matriz de representación, y de esa manera la escena de salida, pueda puede ser ajustada y cambiada interactivamente de acuerdo con la voluntad, preferencias personales y otros criterios. Por ejemplo, es posible ubicar interlocutores de un grupo juntos en un área espacial para maximizar la discriminación de otros interlocutores restantes. Esta interactividad se obtiene incluyendo una interfaz del usuario del codificador.

**[0059]** Por cada objeto de sonido transmitido, se puede ajustar su nivel relativo y (en el caso de la representación no mono) la posición espacial de representación. Esto puede tener lugar en tiempo real al cambiar el usuario la posición de los dispositivos de deslizamiento de interfaz gráfica del usuario (GUI) asociados (por ejemplo: nivel de objeto = +5dB, posición de objeto = -30 grados). Sin embargo, se ha descubierto que debido al enfoque paramétrico basado en separación de mezcla descendente/mezcla, la calidad subjetiva de la salida de audio renderizada depende de la configuración de los parámetros de representación. Se ha encontrado que los cambios del nivel relativo de los objetos afectan la calidad de audio final más que los cambios de posición de la representación espacial ("repaneo"). Se ha encontrado también que las configuraciones extremas para los parámetros relativos (por ejemplo, +20 dB) pueden llevar incluso a una calidad de salida inadmisible. Si bien simplemente esto es el resultado de violación de algunos de los supuestos preceptúales que subyacen en este esquema, de todas maneras es inadmisible que un producto comercial produzca sonido deficiente y perturbaciones dependiendo de la configuración de la interfaz del usuario. En consecuencia, las realizaciones de acuerdo con la invención, como, por ejemplo, el sistema 200, abordan este problema de evitar estas degradaciones inadmisibles independientemente de la configuración de la interfaz del usuario (configuración de la interfaz del usuario que se deben considerar "parámetros de entrada").

**[0060]** A continuación, se describen algunos detalles con respecto a los enfoques para evitar las distorsiones SAOC. El enfoque para limitar las distorsiones SAOC presentadas en la presente se basa en los siguientes conceptos:

- Aparecen distorsiones SAOC importantes por elecciones inapropiadas de coeficientes de representación (que se pueden considerar parámetros de entrada). Esta elección es realizada habitualmente por el usuario de manera interactiva (por ejemplo, por medio de una interfaz gráfica del usuario en tiempo real (GUI) para aplicaciones interactivas). Por lo tanto, se introduce un paso adicional de procesamiento que modifica los coeficientes de representación que fueron provistos por el usuario (por ejemplo, los limita sobre la base de ciertos cálculos) y utiliza estos coeficientes modificados para el motor de representación SAOC. Por ejemplo, los coeficientes de representación que fueron provistos por el usuario se pueden considerar como parámetros de entrada, y los coeficientes modificados para el motor de representación SAOC se pueden considerar parámetros modificados.

- Para controlar la degradación excesiva de la salida de audio SAPC producida, es conveniente desarrollar una medida de computación de degradación perceptual (también designada medida de distorsión DM). Se ha hallado que esta medida de la distorsión debe cumplir ciertos criterios:

  - La medida de la distorsión debe ser fácilmente computable a partir de parámetros internos del motor de decodificación SAPC. Por ejemplo, es conveniente que no sea necesaria la computación extra de bancos de filtro para obtener la medida de distorsión.

  - El valor de medida de la distorsión se debe correlacionar con la calidad del sonido percibido subjetivamente (degradación perceptual), es decir, debe estar en línea con los principios básicos de la psicoacústica. Para este fin, la computación de la mediad de la distorsión puede realizarse preferentemente de manera selectiva de la frecuencia, y por lo general se la conoce como la codificación y el procesamiento de audio perceptual.

**[0061]** Se ha descubierto que se pueden definir y calcular una cantidad de medidas de distorsión SAOC.Sin embargo, se ha encontrado que las medidas de distorsión SAOC deben considerar preferentemente ciertos factores básicos a fin de llegar a una evaluación correcta de una calidad SAOC renderizada y por consiguiente (aunque no necesariamente) tienen ciertos puntos en común.

- Se consideran los coeficientes de mezcla descendente. Estos determinan las fracciones de mezcla relativas de cada objeto de audio dentro de dichas una o más señales de mezcla descendente. Como información de fondo, se debe notar que se ha encontrado que la distorsión SAOC depende de la relación entre la mezcla descendente y los coeficientes de representación. Si la contribución relativa del objeto definida por los coeficientes de representación es sustancialmente diferente de la contribución relativa del objeto dentro de la mezcla descendente, luego el motor de decodificación SAOC (que utiliza los parámetros modificados) tiene que ejecutar un considerable ajuste de la señal de mezcla descendente para convertirla en la salida renderizada. Se ha encontrado que esto da lugar a la distorsión SAOC.

- Se toman en cuenta los coeficientes de representación. Estos determinan la potencia de salida relativa de cada objeto de audio a cada una de las unas o más señales de salida renderizadas. Como información de fondo cabe señala que se ha encontrado que la distorsión SAOC que se produce también depende de la relación mutua de las potencias de los objetos. Si un objeto en cierto punto de tiempo tiene una potencia mucho mayor que otros objetos (y si el coeficiente de mezcla descendente de este objeto no es demasiado pequeño) luego este objeto domina la mezcla descendente y se reproduce muy favorablemente en la señal de salida renderizada. Por el

contrario, los objetos débiles solo se representan muy débilmente en la mezcla descendente y por consiguiente no pueden ser llevados a niveles de salida elevados sin distorsiones significativas.

- Toman en cuenta la potencia de objeto/nivel (relativo) de cada objeto con respecto al otro. Esta información se describe, por ejemplo, en términos de diferencias de niveles de objetos SAOC (OLDs). Como información de fondo, se debe tener en cuenta que se ha encontrado que la distorsión SAOC que tiene lugar depende además de las propiedades de las señales de objetos individuales. Por ejemplo, la amplificación de un objeto de una naturaleza tonal en una salida renderizada a niveles superiores (en tanto que los demás objetos pueden ser más de del tipo ruido) da lugar a una distorsión percibida considerable.

- Además de esto, se puede considerar otra información acerca de las propiedades de las señales de objeto originales. Estas pueden ser transmitidas por codificador SAOC como parte de la información complementaria SAOC. Por ejemplo, la información acerca de la tonalidad o el ruido de cada elemento de objeto se puede transmitir como parte de la información complementaria SAOC y utilizar con el fin de limitar las distorsiones.

2.2 Reseña General del Sistema

**[0062]** Basándose en las consideraciones antes expuestas, se presenta ahora una reseña general del sistema MPEG SAOC 200 para brindar una comprensión cabal de la presente invención. Se debe notar que el sistema SAOC 200 de acuerdo con la Figura 2 es versión extendida del sistema MPEG SAOC 800 de acuerdo con la Figura 8, por lo que también se aplica la explicación anterior. Más aún, se debe notar que el sistema MPEG SAOC 200 puede ser modificado de acuerdo con las alternativas de implementación 900, 930, 960 expuestas en las Figuras 9a, 9b y 9c, donde el codificador de objetos corresponde al codificador SAOC, donde la información de interacción con el usuario/información de control de usuario 822 corresponde a la información de control de representación/coeficiente de representación.

**[0063]** Por añadidura, el decodificador SAOC del sistema MPEG SAOC 100 puede ser reemplazado por la disposición separada del decodificador de objetos y mezclador /renderizador 920, por la disposición integrada de decodificador de objetos y mezclador/renderizador 930 o el transcodificador de SAOC a MPEG Envolvente 980.

**[0064]** Tomando ahora como referencia la Figura 2, se puede apreciar que el sistema MPEG SAOC 200 comprende un codificador SAOC 210, que está configurado para recibir una pluralidad de señales de objeto $x_1$ a $x_N$, asociados con una pluralidad de objetos numerados de 1 a N. El codificador SAOC 210 está también configurado para recibir (o de otro modo obtener) coeficientes de mezcla descendente $d_1$ a $d_N$. Por ejemplo, el codificador SAOC 210 puede obtener una serie de coeficientes de mezcla descendente $d_1$ a $d_N$ por cada canal de la señal de mezcla descendente 212 provista por el codificador SAOC 210. El codificador SAOC 210 puede estar configurado, por ejemplo, para obtener una combinación ponderada de las señales de objeto $x_1$ a $x_N$ a fin de obtener una señal de mezcla descendente, donde cada una de las señales de objeto $x_1$ a $x_N$ es ponderada con su coeficiente de mezcla descendente asociado $d_1$ a $d_N$. El codificador SAOC 210 está configurado asimismo para obtener información de relación entre objetos, que describe una relación entre diferentes señales de objetos. Por ejemplo, la información de relación entre objetos puede comprender información de diferencia del nivel de objeto, por ejemplo, en forma de parámetros OLD e información de correlación entre objetos, por ejemplo en forma de parámetros IOC. En consecuencia el codificador SAOC 200 está configurado entonces para suministrar una o más señales de mezcla descendente 212, cada una de las cuales comprende una combinación ponderada de una o más señales de objetos, ponderados de acuerdo con una serie de parámetros de mezcla descendente asociados con las respectiva señal de mezcla descendente (o un canal de la señal de mezcla descendente de múltiples canales 212). El codificador SAOC 210 también está configurado para brindar información complementaria 214, donde la información complementaria 214 comprende la información de relación entre objetos (por ejemplo, en forma de parámetros de diferencia de nivel entre objetos y parámetros de correlación entre objetos). La información complementaria 214 también comprende una información de parámetros de mezcla descendente, por ejemplo, en forma de parámetros de ganancia de mezcla descendente y parámetros de diferencias de nivel de los canales de mezcla descendente. La información complementaria 214 puede comprender además una información complementaria opcional sobre propiedades de los objetos, que puede representar las propiedades individuales de los objetos. A continuación se describen los detalles con respecto a la información complementaria opcional sobre propiedades de los objetos.

**[0065]** El sistema MPEG SAOC 200 comprende además un decodificador SAOC 220, que puede comprender la funcionalidad del decodificador SAOC 820. En consecuencia, el decodificador SAOC 220 recibe dichas una o más señales de mezcla descendente 212 e información complementaria 214 como así también los coeficientes de representación modificados ("o ajustados", o "reales") 222 y suministra, sobre la base de estos, una o más señales con mezcla ascendente $\hat{y}_1$ a $\hat{y}_N$.

**[0066]** El sistema MPEG SAOC 200 comprende además un aparato 240 para suministrar uno o más parámetros modificados (o "ajustados" o "reales"), es decir los coeficientes de representación modificados 222,dependiendo de uno o más parámetros de entrada, es decir los parámetros de entrada que describen una información de control de

representación o coeficientes de representación 242. El aparato 240 está configurado para recibir asimismo por lo menos una pare de la información complementaria 214. Por ejemplo, El aparato 240 está configurado para recibir los parámetros 214a que describen potencias de los objetos (por ejemplo, potencias de las señales de objeto $x_1$ a $x_N$). Por ejemplo, los parámetros 214a pueden comprender los parámetros de diferencia de nivel de los objetos (también designados OLD). El aparato 240 también recibe preferentemente parámetros 214b de la información complementaria 214 que describen los coeficientes de mezcla descendente. por ejemplo, los parámetros 214b describen los coeficientes de mezcla descendente $d_1$ a $d_N$. Opcionalmente, el aparato 240 puede recibir además otros parámetros adicionales 2214c, que constituyen una información complementaria sobre propiedades de los objetos individuales.

**[0067]** El aparato 240 está configurado en general para proporcionar los coeficientes de representación modificados 222 sobre la base de los coeficientes de representación de entrada 242 (que pueden ser recibidos, por ejemplo, desde una interfaz del usuario, o pueden ser computados, por ejemplo, dependiendo de la entrada del usuario o ser provistos como información predeterminada), por lo que se reduce la distorsión de la representación de la señal de mezcla ascendente que sería causada por el uso de parámetros de representación no óptimos por el decodificador SAOC 220. En otras palabras, los coeficientes de representación modificados 222 son una versión modificada de los coeficientes de representación de entrada 242, donde los cambios se realizan dependiendo de los parámetros 214a, 214b, de modo tal que se reducen o limiten todas las distorsiones audibles en las señales de canales de mezcla ascendente $\hat{y}_1$ a $\hat{y}_N$ (que forman la representación de señal de mezcla ascendente).

**[0068]** El aparato 240 para suministrar dichos uno o más parámetros ajustados 242 puede comprender por ejemplo, un ajustador de coeficientes de representación 250, que recibe los coeficientes de representación de entrada 242 y suministra, sobre la base de estos, los coeficientes de representación modificados 222. Para este fin, el ajustador de coeficientes de representación 250 puede recibir una medida de la distorsión 252, que describe las distorsiones que serían causadas por el uso de los coeficientes de representación de entrada 242, la medida de la distorsión 252 puede ser provista, por el ejemplo, por el calculador de distorsión 260 dependiendo de los parámetros 214a, 214b y los coeficientes de representación 242.

**[0069]** Sin embargo, las funcionalidades del ajustador de coeficientes de representación 250 y del calculador de distorsión 260 también pueden estar integradas en una unidad funcional única, de tal manera que los coeficientes de representación modificados 222 sean provistos sin computación explícita de una medida de distorsión 252. Por el contrario, se pueden aplicar mecanismos implícitos para reducir o limitar la medida de la distorsión.

**[0070]** Con respecto a la funcionalidad del sistema MPEG SAOC 200, se debe notar que la representación de señal de mezcla ascendente, que es producida como salida en forma de canales de mezcla de señales $\hat{y}_1$ a $\hat{y}_N$, se genera con buena calidad perceptual puesto que las distorsiones audibles que serían causadas por una elección incorrecta de la información de interacción con el usuario/información de control del usuario 822 en el sistema de referencia 800, se evitan mediante la modificación o ajuste de los coeficientes de representación. La modificación o ajuste es ejecutada por el aparato 240 de tal manera que se eviten las degradaciones severas de la impresión perceptual, o de tal manera que por lo menos se reduzcan las degradaciones de la impresión perceptual en comparación con un caso en el cual el decodificador SAOC 220 utilice directamente (sin modificación o ajuste) los coeficientes de representación de entrada 242.

**[0071]** A continuación, se resume brevemente la funcionalidad del concepto de la invención. Dada una medida de la distorsión (DM), se puede evitar la distorsión excesiva en la salida de audio calculando el valor medida de la distorsión correspondiente a las señales dadas, y modificando el algoritmo de decodificación SAOC (limitando los coeficientes de representación 212 utilizados en realidad) De tal manera que el valor medida de la distorsión no exceda un determinado umbral. En la Figura 2 se ilustra un sistema 200 de acuerdo con este concepto y que ha sido explicado con cierto detalle anteriormente.

**[0072]** Con referencia al sistema 200m se pueden realizar las siguientes observaciones.

- Los coeficientes de representación deseados 242 son ingresados por el usuario u otra interfaz.

- Antes de ser aplicados en el motor de decodificación de SAOC 220, los coeficientes de representación 242 son modificados por un ajustador de coeficientes de representación 250, que hace uso de una o más medidas de la distorsión calculadas 252, que son provistas por un calculador de distorsión 260.

- El calculador de distorsión 260 evalúa la información (por ejemplo, parámetros 214a, 214b) de la información complementaria 214 (por ejemplo la potencia relativa de los objetos/OLD, coeficientes de mezcla descendente, y – opcionalmente - información sobre propiedades de señales de objeto). Además, se basa en la entrada de coeficientes de representación 242.

**[0073]** En una realización preferida, el aparato 240 está configurado para modificar los coeficientes de representación sobre la base de una medida de la distorsión, De preferencia, los coeficientes de representación son ajustados en forma selectiva de la frecuencia, usando por ejemplo, ponderación selectiva de la frecuencia.

**[0074]** La modificación de los coeficientes de representación se puede basar en este marco (por ejemplo, en un marco actual), o los coeficientes de representación pueden ser ajustados en el tiempo no solo en forma de marco por lo marco, sino también procesado/controlado en el tiempo (por ejemplo, suavizados en el tiempo) donde se pueden aplicar posiblemente diferentes constantes de ataque/decaimiento, o en el caso de un compresor/limitador de rangos dinámicos.

**[0075]** En algunas realizaciones, la medida de la distorsión puede ser selectiva de todas las frecuencias.

**[0076]** En algunas realizaciones, la medida de la distorsión puede considerar una o más de las siguientes características.

- Potencia/energía/nivel de cada Objeto;
- Coeficientes de mezcla descendente;
- Coeficientes de representación; y/o
- Información complementaria adicional sobre propiedades de los objetos, en caso de aplicarse.

**[0077]** En algunas realizaciones, la medida de la distorsión se puede calcular por cada objeto y combinar para arribar a una distorsión total.

**[0078]** En algunas realizaciones, se puede evaluar una Información complementaria adicional sobre propiedades de los objetos 214c. La información complementaria adicional sobre propiedades de los objetos 214c puede ser extraída en un codificador SAOC mejorado, por ejemplo, en el codificador SAOC 210. La información complementaria adicional sobre propiedades de los objetos puede estar incluida, por ejemplo, en un flujo de bits SAOC potenciado, lo que se describe con referencia a la Figura 7. Además, la información complementaria adicional sobre propiedades de los objetos puede ser utilizada para limitar las distorsiones por un codificador SAOC mejorado.

**[0079]** En un caso especial, se puede utilizar el ruido/tonalidad como una propiedad del objeto descrito por la información complementaria adicional sobre propiedades de los objetos. En este caso, el ruido/tonalidad puede ser transmitido con una resolución de frecuencia mucho más gruesa que otros parámetros del objeto (por ejemplo, OLD) para guardar en la información complementaria. En un caso extremo, se puede transmitir la información complementaria sobre propiedades de objetos respecto a ruido/tonalidad con solo una información por objeto (por ejemplo, con características de banda ancha).

2.3 Métrica de Distorsión SAOC

**[0080]** A continuación, se describe una pluralidad de medidas de distorsión diferentes, que se pueden obtener, por ejemplo, utilizando el calculador de distorsión 260. Los detalles con respecto a la aplicación de estas medidas de distorsión para la limitación de los coeficientes de representación se describen más adelante en la sección 2.4.

**[0081]** En otras palabras, Esta sección esboza varias medidas de distorsión. Estas se pueden utilizar individualmente o se las puede combinar para formar una métrica compuesta, más compleja de distorsión, por ejemplo, mediante la adición ponderada de valores de métrica de distorsión individuales. Cabe señalar aquí que los términos “medida de distorsión” y “métrica de distorsión” cantidades similares y en la mayoría de los casos no es necesario distinguirlas.

**[0082]** A continuación, se describe una pluralidad de métricas de distorsión que pueden ser evaluadas por el calculador de distorsión 260 y que pueden ser utilizadas por el ajustador de coeficientes de representación 250 a fin de obtenerlos coeficientes de representación modificados 222 sobre la base de los coeficientes de representación de entrada 242.

2.3.1 Medida de Distorsión # 1

**[0083]** A continuación, se describe una primera medida de distorsión (también designada medida de distorsión # 1.

**[0084]** Para mayor simplicidad conceptual, se considera un sistema N-1-1 SAOC (por ejemplo, una señal de mezcla descendente mono (212) y un canal de mezcla ascendente único (señal)). Se realiza la mezcla descendente

de objetos de audio de entrada en una señal mono y se los renderiza en una salida mono. Como se presenta en la Figura 8 los coeficientes de mezcla descendente son designados $d_1 ... d_N$ y los coeficientes de representación están indicados por $r_1 ... r_N$...En las siguientes fórmulas, se han omitido los índices de tiempo para simplificar. Del mismo modo, se han omitido los índices de frecuencia, señalando que las ecuaciones se relacionan con las señales de subbanda. En algunas de las ecuaciones siguientes, las letras minúsculas indican coeficientes o señales, y las letras mayúsculas indican las potencias correspondientes, que se pueden ver del contexto de las ecuaciones. Cabe señalar que las señales en ocasiones están representadas por correspondientes coeficientes en el dominio de tiempo-frecuencia, en lugar en el dominio del tiempo.

**[0085]** Supongamos que el objeto #m (índice de objeto auditivo m) es un objeto de interés, por ejemplo el objeto más dominante que se incrementa en su nivel relativo y, de esa manera, limita la calidad general del sonido. Luego la señal de salida deseada (señal de canal de mezcla ascendente) está dada por

$$\hat{y}_{1;} = [x_m \cdot r_m] + [\sum_{i=1;\ i \neq m}^{N} x_i \cdot r_i] \qquad (1)$$

**[0086]** En este caso, el primer término es la contribución deseada del objeto de interés a la señal de salida, en tanto que el segundo término indica las contribuciones de todos los demás objetos (“interferencia”).

**[0087]** En realidad, sin embargo, debido al proceso de mezcla descendente, la señal de salida está dada por

$$y_{1;} = t \cdot \sum_{i=1}^{N} x_i \cdot d_i = [x_m \cdot t \cdot d_m] + [\sum_{i=1;\ i \neq m}^{N} x_i \cdot t \cdot d_i] \qquad (2)$$

es decir, la señal de mezcla descendente es escalada seguidamente por un coeficiente de transcodificación, *t*, correspondiente a la matriz “m2” de un decodificador MPEG Envolvente. Una vez más, ésta se puede dividir en un primer término (contribución real de la señal de objeto a la señal de salida) y un segundo término (“interferencia” real por otras señales de objeto). En este caso, el sistema SAOC (por ejemplo, el decodificador SAOC 220, y, opcionalmente, también el aparato 240) determina dinámicamente el coeficiente de transcodificación, *t*, de tal manera que la potencia de la señal de salida renderizada en realidad se equipare a la potencia de la señal ideal:

$$\hat{Y}_1 = Y_1 \quad \Rightarrow \quad t^2 = \frac{\sum_{i=1}^{N} r_i^2 \cdot X_i}{\sum_{i=1}^{N} d_i^2 \cdot X_i} \qquad (3)$$

**[0088]** Una medida de la distorsión (DM) se puede definir calculando la relación entre la contribución de potencia real del objeto #m y su contribución de potencia real:

$$dm_1(m) = \frac{P_{ideal}}{P_{real}} = \frac{r_m^2}{d_m^2 \cdot t^2} = \frac{r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i}{d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i} \qquad (4)$$

**[0089]** En este caso, $\sum_{i=1}^{N} r_i^2 \cdot X_i$ denota la potencia de la señal renderizada en última instancia, y $\sum_{i=1}^{N} d_i^2 \cdot X_i$ es la potencia de la señal de mezcla descendente. Nótese que, en la implementación real, los valores $X_i$ pueden ser reemplazados directamente por los valores correspondientes de *Diferencia de Nivel de los Objetos* (*OLD*$_i$) que se transmiten como parte de la información complementaria SAOC 214.

**[0090]** Para una mejor interpretación de la $dm_1$, se puede reformular su definición de la siguiente manera:

$$dm_1(m) = \frac{r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i}{d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i} = \frac{\dfrac{r_m^2 \cdot X_m}{\sum_{i=1}^{N} r_i^2 \cdot X_i}}{\dfrac{d_m^2 \cdot X_m}{\sum_{i=1}^{N} d_i^2 \cdot X_i}} \qquad (4a)$$

**[0091]** En efecto, esto significa que la métrica de distorsión es la relación de la contribución de potencia relativa de los objetos en la señal renderizada de manera ideal (salida) versus la señal de mezcla descendente (entrada). Esto coincide con el hallazgo de que el esquema SAOC da mejor resultado cuando no tiene que alterar las potencias relativas de los objetos en factores elevados.

**[0092]** El aumento de los valores de $dm_1$ indica la reducción de la calidad del sonido con respecto al objeto de sonido #m. Se ha descubierto que el valor de $dm_1$ se mantiene constante si todos los coeficientes de representación son escalados por un factor común, o si todos los coeficientes de mezcla descendente se escalan de igual manera. También se ha descubierto que aumentando el coeficiente de representación correspondiente al objeto #m (incrementando su nivel relativo) se da lugar a una mayor distorsión. Los valores de $dm_1$ se pueden interpretar de la siguiente manera:

- Un valor de 1 indica una calidad ideal con respecto al objeto #m;
- Los valores $dm_1$ en aumento 1 indican el descenso de la calidad;
- Los valores de $dm_1$ inferiores a 1 no aumentan más la calidad con respecto al objeto #m.

**[0093]** En consecuencia, la medida general de la calidad del sonido (es decir la calidad de todos los objetos) se puede calcular de la siguiente manera:

$$DM_1 = \frac{\sum_{m=1}^{N} w(m) \cdot \max\left[dm_1(m), 1\right]}{\sum_{m=1}^{N} w(m)} \qquad (5)$$

**[0094]** En esta ecuación, *w(m)* indica un factor de ponderación del objeto #m que se relaciona con la significancia y sensibilidad del objeto específico dentro de la escena de audio. Por ejemplo, se podría elegir entonces w(m) dependiendo de la potencia / volumen del objeto $w(m) = (r_m^2 X_m)^\alpha$ donde $\alpha$ puede ser elegido típicamente como 0,25 para emular, a grandes rasgos, el crecimiento del volumen psicoacústico correspondiente a este objeto. Más aun, *w(m)* podría tomar en cuenta fenómenos de tonalidad y enmascaramiento. Por otro lado, se puede ajustar w(m) a 1, lo que facilita la computación de $DM_1$.

2.3.2 Medida de Distorsión #2

**[0095]** Se puede construir una medida de la distorsión alternativa a partir de la ecuación (4) para formar una medida perceptual al estilo de una Relación Ruido a Máscara (NMR), es decir, calcular la relación entre ruido/interferencia y el umbral de enmascaramiento:

$$dm_2(m) = \frac{P_{Ruido}}{Máscara} = \frac{P_{real} - P_{-real}}{msr \cdot P_{total}} = \frac{\left(r_m^2 - d_m^2 \cdot t^2\right) \cdot X_m}{msr \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i} = \frac{\left(r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i - d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i\right) \cdot X_m}{msr \cdot (\sum_{i=1}^{N} r_i^2 \cdot X_i) \cdot (\sum_{i=1}^{N} d_i^2 \cdot X_i)}$$

(6)

**[0096]** En esta ecuación, *msr* es la Relación Máscara a Señal de la señal de audio total que depende de su tonalidad. El aumento de los valores de $dm_2$ indica una mayor distorsión con respecto al objeto de sonido #m. Una vez más, el valor de $dm_2$ se mantiene constante si todos los coeficientes de representación son escalados por un factor común, o si todos los coeficientes de mezcla descendente se escalan del mismo modo. El rango de valores de $dm_2$ se puede interpretar de la siguiente manera:

- Un valor de 0 indica la calidad ideal con respecto al objeto #m;
- El aumento de los valores $dm_2$ por encima de 1 indica degradaciones progresivas;
- Los valores de $dm_2$ inferiores a 1 indican calidad no distinguible con respecto al objeto #m.

**[0097]** En consecuencia, se puede calcular una medida general de la calidad de escena sonora (es decir la calidad correspondiente a todos los objetos) se puede computar de la siguiente manera:

$$DM_2 = \frac{\sum_{m=1}^{N} w(m) \cdot \max\left[dm_2(m), 1\right]}{\sum_{m=1}^{N} w(m)} \qquad (7)$$

**[0098]** Una vez más, *w(m)* indica un factor de ponderación del objeto #m que se relaciona con la significancia / nivel / volumen del objeto específico dentro de la escena de audio, que típicamente se selecciona en términos de *w(m)* = $(r_m^2 X_m)^\alpha$ con $\alpha = 0.25$.

**[0099]** La medida de la distorsión en la ecuación (6) computa la distorsión en términos de diferencia de las potencias (esto corresponde a una medición de "NMR con la diferencia espectral"). Por otro lado, se puede computar la distorsión sobre la base de una forma de onda que lleva a la siguiente medida que incluye un término de producto mezclado adicional:

$$dm_2'(m) = \frac{P_{Ruido}}{Máscara} = \frac{E(\left|y_{m;ideal} - \hat{y}_{m;real}\right|^2)}{msr \cdot P_{total}} =$$

$$\frac{\left| r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i + d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i - 2 \cdot d_m r_m \cdot \sqrt{(\sum_{i=1}^{N} r_i^2 \cdot X_i) \cdot (\sum_{i=1}^{N} d_i^2 \cdot X_i)} \right| \cdot X_m}{msr \cdot (\sum_{i=1}^{N} r_i^2 \cdot X_i) \cdot (\sum_{i=1}^{N} d_i^2 \cdot X_i)}$$

(8)

2.3.3 Medida de Distorsión #3

**[0100]** Se presenta una tercera medida de la distorsión que describe la coherencia entre la señal de mezcla descendente y la señal renderizada. La mayor coherencia da lugar a una mejor calidad subjetiva del sonido. Además, se puede tomar en cuenta la correlación de los objetos de audio si los datos de IOC están presentes en el decodificador SAOC.

**[0101]** De los parámetros de SAOC (por ej., los parámetros 214a, que pueden comprender los parámetros de diferencia de nivel de los objetos y los parámetros de correlación entre objetos) se puede determinar un modelo de la covarianza de los objetos

$$\mathbf{E} = \sqrt{\mathbf{OLD}^{\mathrm{T}} \cdot \mathbf{OLD}} \cdot \mathbf{IOC}$$

**[0102]** Para calcular la medida de la distorsión se integra una Matriz **M** que contiene los coeficientes de representación y mezcla descendente (**M** se puede interpretar como matriz de representación correspondiente a un sistema N–1–2 SAOC)

$$\mathbf{M} = \begin{pmatrix} r_1 & r_2 & \cdots & r_N \\ d_1 & d_2 & \cdots & d_N \end{pmatrix}$$

**[0103]** La covarianza entre la señal de mezcla descendente y renderizada **C** es entonces

$$\mathbf{C} = \mathbf{M} \cdot \mathbf{E} \cdot \mathbf{M}^* = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

**[0104]** Una medida de la distorsión $DM_3$ se define como

$$\mathrm{DM}_3 = 1 - \min\left( \frac{|c_{12}|}{\sqrt{c_{11} \cdot c_{22}}},\ 1 \right)$$

**[0105]** Los valores de $DM_3$ se pueden interpretar de la siguiente manera:

- Los valores están en el rango [0 .. 1] e indican la coherencia entre la señal de mezcla descendente y renderizada.
- Un valor de 0 indica la calidad ideal.
- El aumento de los valores $DM_3$ indica un descenso de la calidad.

2.3.4 Medida de Distorsión #4

2.3.4.1 Reseña General

**[0106]** Este enfoque propone el uso, como medida de la distorsión, de la relación ponderada promediada entre la energía de representación objetivo (UPMIX) y la energía de mezcla descendente óptima (calculada a partir de una mezcla descendente dada DMX).

**[0107]** Para obtener detalles, se hace referencia asimismo a la Fig. 4, que ilustra una representación gráfica de la mezcla descendente (DMX), la energía de mezcla descendente óptima (DMX_opt) y la energía de representación objetivo (UPMIX).

2.3.4.2 Nomenclatura

**[0108]**

$ch = \{1, 2, ..., N_{ch}\}$ índice correspondiente a los canales de mezcla ascendente

$dx = \{1, 2\}$ índice correspondiente a los canales de mezcla descendente

$ob = \{1, 2, ..., N_{ob}\}$ índice correspondiente a objetos de audio

$pb = \{1, 2, ..., N_{pb}\}$ índice correspondiente a bandas de parámetros

$r_{ch,ob,pb} = r(ch, ob, pb)$ matriz de representación correspondiente al canal ch, objeto de audio ob y banda de parámetros pb

$d_{dx,ob,pb} = d(dx, ob, pb)$ matriz de mezcla descendente correspondiente al canal de mezcla descendente dx, objeto de audio ob y banda de parámetros pb

$w_{ob,pb} = w(ob, pb)$ factor de ponderación que representa la significancia / nivel / volumen del objeto de audio ob correspondiente a la banda de parámetros pb

$NRG_{pb} = NRG(pb)$ energía absoluta del objeto del objeto de audio con la energía más elevada correspondiente a la banda de frecuencias pb

$OLD_{ob,pb} = OLD(ob, pb)$ diferencia de nivel de los objetos, que describe las diferencias de intensidad entre un objeto de audio ob y el objeto con la energía más elevada de la correspondiente banda de frecuencias pb

$IOC_{ob_i,ob_j,pb} = IOC(ob_i, ob_j, pb)$ correlación entre objetos, que describe la correlación entre dos canales de objetos de audio.

2.3.4.3 Algoritmo

**[0109]** A continuación se describen brevemente los pasos de un algoritmo para obtener la medida de la distorsión #4:

- Cálculo de las energías relativas de mezcla ascendente y mezcla descendente:

$$\hat{r}^2_{ch,ob,pb} = OLD_{ob,pb} \cdot r^2_{ch,ob,pb}, \qquad \hat{d}^2_{dx,ob,pb} = OLD_{ob,pb} \cdot d^2_{dx,ob}.$$

Normalización de las energías, de tal manera que $\sum_{ob=1}^{N_{ob}} \tilde{r}^2_{ch,ob,pb} = 1$ y $\sum_{ob=1}^{N_{ob}} \tilde{d}^2_{dm,ob,pb} = 1$:

$$\tilde{r}^2_{ch,ob,pb} = \frac{\hat{r}^2_{ch,ob,pb}}{\sum_{ob=1}^{N_{ob}} \hat{r}^2_{ch,ob,pb}}, \qquad \tilde{d}^2_{dm,ob,pb} = \frac{\hat{d}^2_{dm,ob,pb}}{\sum_{ob=1}^{N_{ob}} \hat{d}^2_{dm,ob,pb}}.$$

- Construcción de la mezcla descendente óptima $d^{2(opt)}_{ch,ob,pb}$ por cada canal y banda de mezcla ascendente:

$$d^{2(opt)}_{ch,ob,pb} = \alpha_{ch,ob,pb} \cdot \tilde{d}^2_{1,ob,pb} + \beta_{ch,ob,pb} \cdot \tilde{d}^2_{2,ob,pb}.$$

**[0110]** Las constantes de multiplicación $\alpha_{ch,ob,pb}$, $\beta_{ch,ob,pb}$ se calculan resolviendo el sistema sobredefinido de ecuaciones lineales para satisfacer la siguiente condición: $\left\| d^{2(opt)}_{ch,ob,pb} - \tilde{r}^2_{ch,ob,pb} \right\| \xrightarrow[\alpha,\beta]{} 0$.

- Cálculo de la medida de la distorsión:

$$DM_4 = \sum_{ob=1}^{N_{ob}} \sum_{ch=1}^{N_{ch}} \left| 1 - \frac{\tilde{r}^2_{ch,ob,pb}}{d^{2(opt)}_{ch,ob,pb}} \right| w_{ob,pb} \hat{r}^2_{ch,ob,pb}.$$

2.3.4.4 Control de Distorsión

**[0111]** Se logra controlar la distorsión limitando uno o más coeficientes de representación que dependen de la medida de la distorsión DM4.

**[0112]** Se apreciará que (i) la medida sólo es pertinente al caso de la mezcla descendente estéreo, y (ii) se la puede reducir a DM1 en el caso de #dx=1 y #ch=1.

2.3.4.5 Propiedades

**[0113]** A continuación se resumen brevemente las propiedades del concepto de cálculo de la medida de la distorsión número 4. El concepto

- presume la transcodificación ideal
- puede tratar la mezcla descendente estéreo y
- permite la generalización a una representación multicanal.

2.3.5 Medida de distorsión #5

**[0114]** Se sugiere un cálculo alternativo del coeficiente de transcodificación $t$. Se lo puede interpretar en términos de extensión de $t$ y lleva a la matriz de transcodificación $T$ que se caracteriza por la incorporación de la coherencia entre objetos (IOC) y al mismo tiempo se hace extensiva a la métrica actual DM#1 y DM#2 a mezcla descendente estéreo y mezcla ascendente multicanal. La implementación actual del coeficiente de transcodificación $t$ considera la paridad de la potencia de la señal de salida renderizada en realidad y la potencia de la señal renderizada ideal, es decir

$$t^2 = \frac{\sum_{i=1}^{N} r_i^2 X_i}{\sum_{i=1}^{N} d_i^2 X_i}.$$

**[0115]** La incorporación de la matriz de covarianza ***E*** da una formulación modificada correspondiente a $t$, es decir la matriz de transcodificación $T$, que toma en cuenta, también, la coherencia entre objetos. Los elementos de $E$ se computan a partir de los parámetros SAOC 214, a saber

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

**[0116]** La matriz de transcodificación representa la conversión de la señal de mezcla descendente a la de salida renderizada, de tal manera que $TDx \approx Rx$. Esto se obtiene mediante la minimización del error cuadrático medio, para dar

$$T = RED^* \left( DED^* \right)^{-1}.$$

Donde $H = RED^*$ o $h_{ij} = \sum_{l=1}^{N} \sum_{m=1}^{N} r_{il} d_{jm} e_{lm}$

y $V = DED^*$ o $v_{ij} = \sum_{l=1}^{N} \sum_{m=1}^{N} d_{il} d_{jm} e_{lm}$

la medida de la distorsión al estilo de $dm_1$ aunque en este caso para cada combinación de mezcla descendente/representación $(n,k)$ del objeto $m$ está dada por

$$dm_5^{''}(m,n,k) = \frac{r_{m,k}^2 v_{n,n}}{d_{m,n}^2 h_{k,n}}\,.$$

**[0117]** Considerando $dm_1(m)$ por separado para los canales de mezcla descendente izquierdo y derecho lleva a

$$dm_L(m,k) = \frac{r_{m,k}^2 v_{1,1}}{d_{m,1}^2 h_{k,1}} \text{ y } dm_R(m,k) = \frac{r_{m,k}^2 v_{2,2}}{d_{m,2}^2 h_{k,2}}\,.$$

**[0118]** Se puede suponer que el mejor de los dos trayectos de mezcla descendente/mezcla ascendente es relevante para la calidad de la salida renderizada, por lo que la medida corresponde al valor mínimo, es decir

$$dm_5^{'}(m,k) = \min\left[ dm_L, dm_R \right].$$

**[0119]** Una medida general de todos los canales de salida, designada por el índice k, se puede computar de la siguiente manera

$$dm_5(m) \frac{\sum_{k=1}^{N_{Ch}} dm_5^{'}(m,k)\, r_{m,k}^2 X_m}{\sum_{k=1}^{N_{Ch}} r_{m,k}^2 e_{k,k}}\,.$$

**[0120]** La medida general de todos los objetos se puede obtener según

$$DM_5 = \frac{\sum_{m=1}^{N} w(m) \max\left[dm_5(m), 1\right]}{\sum_{m=1}^{N} w(m)}$$ donde $w(m) = \left[r_m^2 X_m\right]^{\alpha}$ como antes.

**[0121]** Es posible una extensión similar de $t$ a $T$ en el caso de $dm_2$ y $dm_2^{'}$.

2.3.6. Medida de distorsión #6

**[0122]** A continuación se describe una sexta medida de la distorsión.

**[0123]** Supongamos que $e_i(t)$ es el cuadrado de la envolvente de Hilbert de la señal de objeto #i y $P_i$ la potencia de la señal de objeto #i (ambas típicamente dentro de una subbanda), luego se puede obtener una medida N de la semejanza tonalidad/ruido de un cálculo normalizado de la varianza de la envolvente de Hilbert como

$$N_i = \frac{\mathrm{var}\{e_i\}}{P_i^2}$$

Por otro lado, también se puede utilizar la potencia / varianza de señal con diferencia de envolvente de Hilbert en lugar de la varianza de la envolvente de Hilbert en sí. De todo caso, la medida describe la fuerza de la fluctuación de la envolvente en el tiempo.

**[0124]** Esta medida de la semejanza tonalidad/ruido, N, se puede determinar tanto con respecto a la mezcla de señales renderizadas idealmente como a la mezcla de sonido renderizada SAOC real y se puede computar una medida de la distorsión a partir de la diferencia entre ambas, por ej.:

$$DM_6 = \left|N_{ideal} - N_{real}\right|^{\beta}$$

donde β es un parámetro (por ej. β =2).

2.3.7. Cálculo de las energías de las imágenes de las señales de origen correspondientes a la escena de referencia y la escena renderizada por SAOC

**[0125]** Para calcular las energías de objeto de la imagen de origen en la escena de referencia y la renderizada por SAOC utilizadas para las medidas de la distorsión, hay que tomar en cuenta la matriz de transcodificación $T$ correspondiente a la escena renderizada por SAOC como se hace en “Medida de distorsión 5”, aunque también la correlación de las señales de origen tanto de la escena de referencia como de la escena renderizada.

**[0126]** *Comentario*: La notación de las señales en letras mayúsculas refleja en este caso la notación de matriz de las señales, no las energías de las señales como en los capítulos anteriores

**[0127]** En el caso de un origen arbitrario $x_m$ las partes de señal de $x_m$ de todas las fuentes $x_i$ se pueden calcular de la siguiente manera:

Se dividen todas las señales de origen $x_i$ en una parte de señal $x_{i\|m}$ que se correlaciona con el objeto de interés $x_m$ y una parte $x_{i\perp m}$ que no se correlaciona con $x_m$. Esto se puede lograr mediante la proyección

subespacial de $x_m$ sobre todas las señales $x_i$, es decir $x_i = x_{i\|m} + x_{i\perp m}$. la parte correlacionada está dada por

$$x_{i\|m} = \frac{x_m{}^T x_i}{x_m{}^T x_m} x_m = \frac{IOC_{i,m}}{\|x_m\|^2} x_m = g_{i,m} x_m .$$

2.3.7.1 Cálculo de $P_{ideal,x_m}$ a partir de la imagen de origen $y_{x_m}$ en la escena de referencia $y$:

**[0128]** Donde $Y = \mathrm{R}X$ y $X = X_{\perp m} + X_{\|m}$, la imagen $y_{x_m}$ de la fuente $x_m$ correspondiente a todos los canales renderizados se puede calcular mediante $\mathrm{Y}_{x_m} = \mathrm{R}X_{\|m}$ donde

$$X_{\|m} = \begin{pmatrix} x^T{}_{1\|m} \\ x^T{}_{2\|m} \\ \vdots \\ x^T{}_{N\|m} \end{pmatrix} = \begin{pmatrix} g_{1,m} x_m{}^T \\ g_{2,m} x_m{}^T \\ \vdots \\ g_{N,m} x_m{}^T \end{pmatrix}$$

**[0129]** $\mathrm{Y}_{x_m}$ se puede calcular según

$$\mathrm{Y}_{x_m} = RX_{\|m} = \begin{pmatrix} r_{ch_1,x_1} & r_{ch_1,x_2} & \vdots & r_{ch_1,x_N} \\ r_{ch_2,x_1} & r_{ch_2,x_2} & \vdots & r_{ch_2,x_N} \\ \cdots & \cdots & \ddots & r_{N_{ch-1},x_N} \\ r_{N_{ch},x_1} & r_{N_{ch},x_2} & r_{N_{ch},x_{n-1}} & r_{N_{ch},x_N} \end{pmatrix} \begin{pmatrix} g_{1,m} x_m{}^T \\ g_{2,m} x_m{}^T \\ \vdots \\ g_{N,m} x_m{}^T \end{pmatrix}$$

**[0130]** Por lo tanto, la energía $P_{ideal,x_m}$ de la imagen de origen $\mathrm{Y}_{x_m}$ en la escena de referencia es:

$$P_{ideal,x_m} = \begin{pmatrix} \left\| r_{ch_1,x_1} g_{1,m} + r_{ch_1,x_2} g_{2,m} + \cdots + r_{ch_1,x_N} g_{N,m} \right\|^2 \|x_m\|^2 \\ \cdots \\ \left\| r_{N_{ch},x_1} g_{1,m} + r_{N_{ch},x_2} g_{2,m} + \cdots + r_{N_{ch},x_N} g_{N,m} \right\|^2 \|x_m\|^2 \end{pmatrix} .$$

2.3.7.2 Cálculo de $P_{real,x_m}$ de la imagen de origen $\hat{y}_{x_m}$ en la escena renderizada por SAOC $\hat{y}$:

**[0131]** Esto se puede realizar de manera igual a la correspondiente a $P_{ideal,x_m}$. donde $T$ es la matriz de transcodificación y $D$ la matriz de mezcla descendente, $\hat{y}_{x_m}$ para todos los canales de la escena renderizada es:

$$\hat{Y}_{x_m} = T^{0.5} D X_{||m}.$$

**[0132]** Utilizando $D = \begin{pmatrix} d_{11} & \cdots & d_{1N} \\ d_{21} & \cdots & d_{2N} \end{pmatrix}$ y $T = \begin{pmatrix} t_{11} & t_{12} \\ \vdots & \vdots \\ t_{N_{ch}1} & t_{N_{ch}2} \end{pmatrix}$

$$\hat{Y}_{x_m} = \begin{pmatrix} \sqrt{t_{11}}d_{11} + \sqrt{t_{12}}d_{21} & \sqrt{t_{11}}d_{12} + \sqrt{t_{12}}d_{22} & \cdots & \sqrt{t_{11}}d_{1N} + t_{12}d_{2N} \\ \sqrt{t_{21}}d_{11} + \sqrt{t_{22}}d_{21} & \sqrt{t_{21}}d_{12} + \sqrt{t_{22}}d_{22} & \cdots & \sqrt{t_{21}}d_{1N} + t_{22}d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t_{N_{ch}1}}d_{11} + \sqrt{t_{N_{ch}2}}d_{21} & \sqrt{t_{N_{ch}1}}d_{12} + \sqrt{t_{N_{ch}2}}d_{22} & \cdots & \sqrt{t_{N_{ch}1}}d_{1N} + \sqrt{t_{N_{ch}2}}d_{2N} \end{pmatrix} \begin{pmatrix} g_{1,m}x_m^T \\ g_{2,m}x_m^T \\ \vdots \\ g_{N,m}x_m^T \end{pmatrix}$$

**[0133]** Por lo tanto, la energía $P_{actual,x_m}$ de la imagen de origen $\hat{Y}_{x_m}$ de la escena de referencia ha de ser:

$$P_{real,x_m} = \begin{pmatrix} \left\| g_{1,m}\left(\sqrt{t_{11}}d_{11} + \sqrt{t_{12}}d_{21}\right) + g_{2,m}\left(\sqrt{t_{11}}d_{12} + \sqrt{t_{12}}d_{22}\right) + \cdots g_{N,m}\left(\sqrt{t_{11}}d_{1N} + \sqrt{t_{12}}d_{2N}\right) \right\|^2 \|x_m\|^2 \\ \cdots \\ \left\| g_{1,m}\left(\sqrt{t_{N_{ch}1}}d_{11} + \sqrt{t_{N_{ch}2}}d_{21}\right) + g_{2,m}\left(\sqrt{t_{N_{ch}1}}d_{12} + \sqrt{t_{N_{ch}2}}d_{22}\right) + \cdots g_{N,m}\left(\sqrt{t_{N_{ch}1}}d_{1N} + \sqrt{t_{N_{ch}2}}d_{2N}\right) \right\|^2 \|x_m\|^2 \end{pmatrix}$$

2.3.7.3. Cálculo de la medida de la distorsión

**[0134]** La medida de la distorsión al estilo de $dm_1$ se puede calcular en lo que respecta a cada objeto $m$ y canal de representación de salida $k$ de la siguiente manera:

$$dm_7'(m,k) = \frac{P_{ideal}}{P_{real}} = \frac{\left\| r_{k1}IOC_{1m} + \cdots + r_{kN}IOC_{Nm} \right\|^2}{\left\| \left(\sqrt{t_{k1}}d_{11} + \sqrt{t_{k2}}d_{21}\right) IOC_{1m} + \cdots + \left(\sqrt{t_{k1}}d_{1N} + \sqrt{t_{k2}}d_{2N}\right) IOC_{Nm} \right\|^2}$$

.

$$dm_7(m)\frac{\sum_{k=1}^{N_{Ch}} dm_7^{'}(m,k)\, r_{m,k}^2 \left\| x_m \right\|^2}{\sum_{k=1}^{N_{Ch}} r_{m,k}^2 e_{k,k}}.$$

$$DM_7 = \frac{\sum_{m=1}^{N} w(m) \max\left[ dm_7(m), 1 \right]}{\sum_{m=1}^{N} w(m)}$$

donde $w(m) = \left[ r_m^2 X_m \right]^{\alpha}$ como antes.

2.3.8 Propiedades de Señales de Objeto

**[0135]** A continuación se describe un ejemplo de propiedades de las señales de objeto que pueden se utilizadas, por ejemplo, por el aparato 250 o el dispositivo reductor de perturbaciones 320 para obtener una medida de la distorsión.

**[0136]** En el procesamiento SAOC, se realiza la mezcla descendente de varias señales de objetos de audio para obtener una señal de mezcla descendente que luego es utilizada para generar la salida renderizada final. Si se mezcla una señal de objeto tonal con una segunda señal de objeto más del tipo ruido de igual potencia de señal, el resultado tiene a ser más de ruido. Lo mismo ocurre si la segunda señal de objeto tiene una potencia más elevada. Sólo que, si la segunda señal de objeto tiene una potencia sustancialmente menor a la de la primera, el resultado tiende a ser tonal. Del mismo modo, la semejanza de tonalidad /ruido de la señal de salida SAOC renderizada está determinada principalmente por la semejanza de tonalidad / ruido de la señal de mezcla descendente independientemente de los coeficientes de representación aplicados. Para obtener una calidad de salida subobjetiva favorable, también la semejanza de tonalidad/ ruido de la señal renderizada en realidad debe acercarse a la semejanza de tonalidad /ruido de la señal renderizada idealmente. Para utilizar este concepto en la medida de la distorsión, es necesario transmitir la información acerca de la semejanza de tonalidad/ ruido de cada objeto como parte del flujo de bits. A continuación se puede estimar la semejanza de tonalidad/ruido N de la salida renderizada idealmente en el decodificador SAOC en función de la semejanza de tonalidad/ruido de cada objeto $N_i$ y su potencia de objeto $P_i$ , es decir

$$N = f(N_1, P_1, N_2, P_2, N_3, P_3, \ldots)$$

y compararla con la semejanza de tonalidad/ruido de la señal de salida renderizada en realidad para computar una medida de la distorsión. Por ejemplo, se puede utilizar la siguiente función f():

$$N = \frac{\sum_i N_i \cdot P_i^{\alpha}}{(\sum_i P_i)^{\alpha}}$$

que combina los valores de semejanza de tonalidad/ruido del objeto y las potencias de los objetos en una sola salida estimando el valor de semejanza de tonalidad/ruido de las mezcla de las señales. Se puede elegir el parámetro α para optimizar la precisión del procedimiento de cálculo para una medida de semejanza de tonalidad/ruido (por ej. α=2). Se describe una métrica de distorsión adecuada basada en la semejanza de tonalidad/ ruido en la Sección 2.3.6 como medida de distorsión #6.

2.4 Esquema de limitación de Distorsiones

Reseña general de los esquemas de limitación de distorsiones

**[0137]** A continuación se presenta una breve reseña de una pluralidad de esquemas de limitación de las distorsiones. Como se señalara anteriormente, el ajustador de coeficientes de representación 250 recibe los coeficientes de representación de entrada 242 y produce, sobre la base de éstos, un coeficiente de representación 222 modificado para utilizar por el decodificador SAOC 220.

**[0138]** Se pueden distinguir diferentes conceptos para la provisión de coeficientes de representación modificados, donde los conceptos también se pueden combinar en algunas realizaciones. De acuerdo con el primer concepto, se obtiene uno o más valores límite de los parámetros de representación en un primer paso dependiendo de uno o más parámetros de información complementaria 214 (es decir, que depende de la información paramétrica relacionada con el objeto 214). Seguidamente, se obtienen los coeficientes de representación reales "(modificados o ajustados)" 222 que dependen del parámetro de representación deseado 242 y de dichos uno o más valores límite de los parámetros de representación, de manera que los parámetros de representación reales respeten los límites definidos por los valores límite de los parámetros de representación. En consecuencia, dichos parámetros de representación que exceden los valores límite de los parámetros de representación, son ajustados (modificados) para adaptarse a los valores límite de los parámetros de representación. Este primer concepto es fácil de implementar, aunque en ocasiones trae aparejado una satisfacción levemente degradada del usuario, puesto que la elección del usuario de los parámetros de representación deseados 242 queda fuera de consideración si los parámetros de representación deseados 242 definidos por el usuario exceden los valores límite de los parámetros de representación.

**[0139]** De acuerdo con el segundo concepto, el ajustador de parámetros computa una combinación entre el cuadrado de un parámetro de representación deseado y el cuadrado de un parámetro de representación óptimo, para obtener el parámetro de representación real. En este caso, el ajustador de parámetros está configurado para determinar una contribución del parámetro de representación deseado y del parámetro de representación óptimo a la combinación de lineal dependiendo de un parámetro umbral predeterminado y una métrica de distorsión (descripta anteriormente).

**[0140]** Además, se puede distinguir si la medida de la distorsión (métrica de distorsión) se computa utilizando las propiedades de relaciones entre objetos y/o las propiedades de los objetos individuales. En algunas realizaciones, sólo se evalúan las propiedades de la relación entre objetos, dejando a las propiedades de los objetos individuales (que se relacionan sólo con un objeto) fuera de consideración. En algunas realizaciones adicionales, solo se tienen en cuenta las propiedades de los objetos individuales, dejando las propiedades de relación entre objetos fuera de consideración. Sin embargo, en algunas realizaciones, se evalúa una combinación de propiedades de relación entre objetos y propiedades de los objetos individuales.

**[0141]** Basándose en las consideraciones que anteceden, y también basándose en la descripción anterior de las diferentes medidas de distorsión, se pasa a definir un número de esquemas para limitar la distorsión, esbozados en las siguientes subsecciones. Estos esquemas para limitar la distorsión pueden ser aplicados por el ajustador de coeficientes de representación 250 para obtener los coeficientes de representación modificados dependiendo de los coeficientes de representación de entrada 242.

2.4.2 Esquema de Limitación de Distorsiones #1

**[0142]** En la subsección 2.3.1 se define una sencilla medida de distorsión calculando la relación entre la contribución de potencia ideal del objeto #m y su contribución de potencia real (ecuación 4):

$$dm_1(m) = \frac{P_{ideal}}{P_{real}} = \frac{r_m^2}{d_m^2 \cdot t^2} = \frac{r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i}{d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i} \qquad (4)$$

**[0143]** En esta ecuación, las únicas variables que están bajo el control del renderizador SAOC son los coeficientes de representación que se utilizan en el proceso de transcodificación. Por lo tanto, si la métrica de distorsión obtenida no excede un determinado valor umbral, *T*, esto impone una condición al correspondiente coeficiente de matriz de representación:

$$dm_1(m) = \frac{r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i}{d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i} \leq T \Leftrightarrow r_m^2 \leq \hat{r}_m^2 = T \cdot \frac{d_m^2 \cdot \sum_{i=1,\, i \neq m}^{N} r_i^2 \cdot X_i}{\left| \sum_{i=1}^{N} d_i^2 \cdot X_i - T \cdot d_m^2 \cdot X_m \right|} \quad \text{(6.1.a)}$$

**[0144]** Para encontrar una solución para todos los $\hat{r}_m^2$ se puede establecer una serie de ecuaciones lineales $\mathbf{Ax} = \mathbf{b}$ donde

$$\mathbf{x} = \begin{bmatrix} \hat{r}_1^2 \\ \hat{r}_2^2 \\ \vdots \\ \hat{r}_N^2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \sum_{i=1}^{N} r_i^2 \end{bmatrix} \text{ y } \mathbf{A} = \begin{bmatrix} -c_1 & d_1^2 X_2 & \cdots & d_1^2 X_N \\ d_2^2 X_1 & -c_2 & \cdots & d_2^2 X_N \\ \vdots & \vdots & \ddots & \vdots \\ d_N^2 X_1 & d_N^2 X_2 & \cdots & -c_N \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

donde $c_m = \frac{1}{T}\left( \sum_{i=1}^{N} d_i^2 \cdot X_i - T \cdot d_m^2 \cdot X_m \right)$.

**[0145]** Las primeras N filas de **A** derivan directamente de la ecuación (6.1.a). Además, se agrega una restricción de manera que la energía de los nuevos coeficientes (limitados) de representación sea igual a la energía de los coeficientes especificados por el usuario. Se obtiene entonces una solución correspondiente a $\hat{r}_m^2$ (que se considera como valores límite de los parámetros de representación) de la siguiente manera:

$$\mathbf{x} = \left( \mathbf{A}^{\mathbf{T}} \mathbf{A} \right)^{-1} \mathbf{A}^{\mathbf{T}} \mathbf{b}$$

**[0146]** A partir de esto, se puede ver un primer esquema de limitación de las distorsiones, a saber: En lugar de usar los coeficientes de la matriz de representación 242 tal como son ingresados al decodificador SAOC desde la interfaz del usuario, el coeficiente de representación utilizado efectivamente $r_m'$, 222 para el objeto #m es modificado / limitado (por ejemplo, por el ajustador de coeficientes de 240 cuadro por cuadro antes de utilizarlo para el proceso de decodificación SAOC:

$$r_m'^2 = \min\left( r_m^2 \ , \ \hat{r}_m^2 \right)$$

**[0147]** Nótese que el proceso de limitación depende de las energías de los objetos individuales en cada cuadro específico. El enfoque es sencillo y tiene los siguientes inconvenientes leves:

- No toma en cuenta el volumen del objeto ni el enmascaramiento perceptual y
- Sólo captura los efectos de incremento de ganancia de un objeto específico, pero no captura los efectos atenuando las ganancias de objeto. Esto podría solucionarse estableciendo asimismo un límite más bajo para el valor dm.

2.4.3 Esquema de limitación #2

2.4.3.1 Reseña general del Esquema de limitación

**[0148]** Esta sección describe una función de limitación que toma en cuenta los siguientes aspectos:

- la medida de la distorsión está restringida por un umbral limitante,

- la derivación de la matriz de representación limitada se basa en la función de limitación y en su distancia de la matriz de representación inicial.

**[0149]** Esta función de limitación (o esquema de limitación) puede ser ejecutado, por ejemplo, por el ajustador de coeficientes de representación 250 en combinación con el calculador de distorsiones 260.

**[0150]** La medida de la distorsión es una función de la matriz de representación, por lo que

- una matriz de representación inicial (descripta, por ejemplo, por los coeficientes de representación de entrada 242) da una medida de la distorsión inicial,

- la medida de distorsión óptima produce una matriz de representación óptima, pero la distancia de esta matriz de representación óptima a la matriz de representación inicial puede no ser óptima,

- la medida de la distorsión es inversamente proporcional lineal a la distancia de una matriz de representación con la matriz de representación inicial,

- para un determinado umbral, se deriva la matriz de representación limitada (descripta, por ejemplo, por los coeficientes de representación ajustados o modificados 222) mediante interpolación (por ejemplo, interpolación lineal) entre el punto de trabajo inicial y el óptimo.

**[0151]** Además, se puede suponer que la potencia de la señal renderizada en cada punto de trabajo es aproximadamente constante, por lo que

$$\sum_{i=1}^{N_{ob}} r_i^2 X_i \approx \sum_{i=1}^{N_{ob}} r_{lim,i}^2 X_i \approx \sum_{i=1}^{N_{ob}} r_{opt,i}^2 X_i \ .$$

**[0152]** Se puede utilizar el esquema de limitación #2 en combinación con diferentes medidas de distorsión, como se describe a continuación.

2.4.3.2 <u>Limitación de la medida de distorsión #1</u>

**[0153]** Por cada banda de parámetros, la medida de la distorsión $dm_1(m)$ correspondiente a un objeto de interés $m$ se define como

$$dm_1(m) = \frac{r_m^2 \sum_{i=1}^{N_{ob}} d_i^2 X_i}{d_m^2 \sum_{i=1}^{N_{ob}} r_i^2 X_i} \ .$$

**[0154]** La matriz de representación óptima se produce cuando se ajusta $dm_1(m)$ en su valor óptimo, es decir $dm_{1,opt}(m) = 1$

$$r_{opt,m}^2 = d_m^2 \frac{\sum_{i=1}^{N_{ob}} r_i^2 X_i}{\sum_{i=1}^{N_{ob}} d_i^2 X_i} \ .$$

**[0155]** En consecuencia, se pueden obtener los valores óptimos de la matriz de representación $r_{opt,m}^2$ utilizando una serie de ecuaciones, donde $r_i^2$ es reemplazado por $r_{opt,i}^2$.

**[0156]** Dentro del umbral predefinido $T$ correspondiente a $dm_1(m)$ la matriz de representación limitada está dada por

$$r_{lim,m}^2 = \frac{T-1}{dm_1(m)}\left(r_m^2 - r_{opt,m}^2\right) + r_{opt,m}^2 .$$

2.4.3.3 Limitación de la medida de distorsión #2a

**[0157]** La medida de distorsión $dm_{2a}(m)$, que en ocasiones se designa también, brevemente, "$dm_2(m)$", se define como

$$dm_{2a}(m) = \frac{\left(r_m^2 \sum_{i=1}^{N_{ob}} d_i^2 X_i - d_m^2 \sum_{i=1}^{N_{ob}} r_i^2 X_i\right) X_m}{msr \sum_{i=1}^{N_{ob}} r_i^2 X_i \sum_{i=1}^{N_{ob}} d_i^2 X_i} = \frac{\frac{r_m^2 X_m}{\sum_{i=1}^{N_{ob}} r_i^2 X_i} - \frac{d_m^2 X_m}{\sum_{i=1}^{N_{ob}} d_i^2 X_i}}{msr}$$

en el caso del objeto $m$ y cada banda de parámetros. Con respecto a ciertas bandas de parámetros $pb$ la relación máscara a señal $msr(pb)$ es una función de la potencia de la señal renderizada

$$msr(pb) = \left[\sum_{i=1}^{N_{ob}} r_i^2 X_i M_k\right]_{k=\max(pb)} = \left[\sum_{i=1}^{N_{ob}} r_i^2 X_i\right]_{k=\max(pb)} \left[M_k\right]_{k=\max(pb)} .$$

**[0158]** El valor óptimo de la medida de la distorsión es cero, es decir $dm_{2a,opt}(m) = 0$. Esto corresponde a un proceso de transcodificación prefecto que no introduce error alguno. Por ende, la matriz de representación óptima da

$$r_{opt,m}^2 = d_m^2 \frac{\sum_{i=1}^{N_{ob}} r_i^2 X_i}{\sum_{i=1}^{N_{ob}} d_i^2 X_i} .$$

**[0159]** Donde $dm_{2a}(m)=T$ la matriz de la matriz de representación limitada, que puede ser descripta por los coeficientes de representación modificados 222, se convierte en

$$r_{lim,m}^2=\frac{T-1}{dm_{2a}(m)}\left(r_m^2-r_{opt,m}^2\right)+r_{opt,m}^2 .$$

2.4.3.4 Limitación de la medida de distorsión #2b

**[0160]** La medida de la distorsión $dm_{2b}(m)$, la que en ocasiones también se denomina en forma breve $dm_{2'}(m)$, también puede ser utilizada por el aparato 240 para obtener la matriz de representación limitada, que puede ser descripta por los coeficientes de representación modificados 222, que dependen de los coeficientes de representación de entrada 242.

2.4.3.5 Limitación de la medida de distorsión #4

**[0161]** La medida de la distorsión $dm_4(m)$ se define de la siguiente manera:

$$dm_4(m)=\left|1-\frac{r_m^2\sum_{i=1}^{N_{ob}}d_i^2X_i}{d_m^2\sum_{i=1}^{N_{ob}}r_i^2X_i}\right|$$

en el caso del objeto $m$ y cada banda de parámetros y su valor óptimo es $dm_{4,opt}(m)=0$. En consecuencia, las matrices de representación óptima y limitada dan como resultado

$$r_{opt,m}^2=d_m^2\frac{\sum_{i=1}^{N_{ob}}r_i^2X_i}{\sum_{i=1}^{N_{ob}}d_i^2X_i}$$

y

$$r_{lim,m}^2=\frac{T-1}{dm_4(m)}\left(r_m^2-r_{opt,m}^2\right)+r_{opt,m}^2 .$$

**[0162]** En consecuencia, el aparato 240 puede suministrar los coeficientes de representación modificados 222 dependiendo de los coeficientes de representación de entrada 242 y dependiendo asimismo de la medida de la distorsión 252, que puede ser igual a la cuarta medida de distorsión $dm_4(m)$.

2.4.4 Esquema de limitación #3

**[0163]** Con respecto a la fórmula (6.1.a) se puede calcular el coeficiente de representación limitado correspondiente al objeto *m* para la medida de la distorsión #3 de la siguiente manera. Con las abreviaturas

$$c_1 = \sum_{i=1}^{N}\sum_{j=1}^{N} d_i\, d_j e_{ij} \;\;,\;\; c_2 = \sum_{i=1,\, i\neq m}^{N} r_i\, e_{im} \;\;,\; c_3 = \sum_{i=1,\, i\neq m}^{N}\;\sum_{j=1,\, j\neq m}^{N} r_i\, r_j\, e_{ij} \;\;,\; c_4 = \sum_{i=1}^{N} d_i\, e_{mi}$$

y $c_5 = \sum_{i=1,\, i\neq m}^{N}\;\sum_{j=1}^{N} r_i\, d_j e_{ij}$

se establece una ecuación cuadrática

$$\hat{r}_m^2\left(\left(1-T\right)^2\cdot c_1 e_{mm} - c_4^2\right) + \hat{r}_m \cdot 2\cdot\left(\left(1-T\right)^2\cdot c_1 c_2 - c_4 c_5\right) + \left(1-T\right)^2\cdot c_1 c_3 - c_5^2 \;\overset{!}{=}\; a\cdot\hat{r}_m^2 + b\cdot\hat{r}_m + c \;=\; 0$$

cuya solución (positiva) es

$$\hat{r}_m = \frac{-b+\sqrt{b^2-4ac}}{2a} \qquad (6.2.a)$$

**[0164]** En consecuencia, el aparato 240 puede comprender valores límite de los parámetros de representación $\hat{r}_m$, y puede limitar los coeficientes de representación ajustados (o modificados) 222 de acuerdo con dichos valores límite de los parámetros de representación.

2.4.5 Otras mejoras opcionales

**[0165]** El concepto antes descrito para limitar los coeficientes de representación 222, que es ejecutado, en forma individual o en combinación, por el aparato 240, puede ser mejorado aun. Por ejemplo, se puede ejecutar una generalización para la representación de M canales. Para este fin, se puede utilizar la suma de los cuadrados/potencia de los coeficientes de representación en lugar de un solo coeficiente de representación.

**[0166]** Además, se puede ejecutar una generalización a una mezcla descendente estéreo. Para este fin, se puede utilizar una suma de los cuadrados/potencias de los coeficientes de mezcla descendente en lugar de un solo coeficiente de mezcla descendente.

**[0167]** En algunas realizaciones, se puede combinar la mezcla de distorsión a través de la frecuencia para obtener una sola que se utiliza para controlar la degradación. Por otro lado, puede ser mejor (y más sencillo) en algunos casos, realizar el control de la distorsión independientemente por cada banda de frecuencia.

**[0168]** Se pueden aplicar diferentes conceptos para efectuar, en realidad, el control de la distorsión. Por ejemplo, se puede limitar dichos uno o más coeficientes de representación. Por otro lado, o además, se puede limitar un coeficiente de matriz m2 (por ejemplo de una decodificación de MPEG Envolvente). Por otro lado, o además, se puede limitar una ganancia relativa del objeto.

3. Realización de acuerdo con la Fig. 3

**[0169]** A continuación se describe otra realización de un decodificador SAOC tomando como referencia la Fig. 3. Para facilitar la comprensión, en primer lugar se presenta una breve explicación de las consideraciones subyacentes. La salida de un sistema de “codificación espacial de objetos de audio” (SAOC) (como el que se ha normalizado como ISO/IEC 23003–2) puede exhibir perturbaciones que dependen de las propiedades del objeto de audio y la relación entre la matriz de representación y la matriz de mezcla descendente. Para explicar este problema, se considera aquí el caso en que las matrices de mezcla descendente y representación tienen la misma dimensión

sin desmedro de la generalidad. Se aplican consideraciones correspondientes si el número de canales de la escena de mezcla descendente y renderizada son diferentes.

**[0170]** Se ha descubierto que, en general, el riesgo de perturbaciones se incrementa cuando la matriz de representación se torna significativamente diferente de la matriz de mezcla descendente. Se pueden distinguir diferentes tipos de perturbaciones:

**1.** Imperfecciones de la representación, es decir que la matriz de representación "efectiva" difiere de la matriz de representación deseada que es ingresada como entrada al decodificador SAOC (la atenuación lograda en efecto o la ganancia de un objeto es diferente de la especificada en la matriz de representación) Este es típicamente el efecto de la superposición de objetos en ciertas bandas de parámetros.

**2.** Cambios desfavorables y, posiblemente, temporalmente variantes del timbre de un objeto. Esta perturbación es especialmente severa cuando la "fuga" mencionada en 1. sólo tiene lugar localmente en caso de una sola banda de parámetros.

**3.** Perturbaciones, como señales de objeto moduladas, tonos musicales o ruido modulado, causadas por el procesamiento de señales de tiempo y frecuencia variantes en el decodificador SAOC.

**[0171]** Se ha descubierto que es ventajoso minimizar todo tipo de perturbaciones.

**[0172]** Una estrategia generalizada para abordar este problema y minimizar las perturbaciones consiste en emplear un post–procesamiento variante en tiempo–frecuencia de la matriz de representación deseada antes de enviarla al decodificador SAOC. Esta estrategia está expuesta en la Fig. 3.

**[0173]** La Fig. 3 ilustra un diagrama esquemático de bloques de una disposición de decodificador SAOC 300. El decodificador SAOC 300 puede ser designado, en pocas palabras, decodificador de señales de audio. El decodificador de señales de audio 300 comprende un núcleo de decodificador SAOC 310, que está configurado para recibir una representación de señal de mezcla descendente 312 y un flujo de bits SAOC 314 y para producir, sobre la base de estos, una descripción 316 de una escena renderizada, por ejemplo, en forma de representación de una pluralidad de canales de audio de mezcla ascendente.

**[0174]** El decodificador de señales de audio 300 comprende asimismo un dispositivo reductor de perturbaciones 320, que se puede presentar, por ejemplo, en forma de aparato para suministrar uno o más parámetros ajustados dependiendo de uno o más parámetros de entrada. El dispositivo reductor de perturbaciones 320 está configurado para recibir información 322 acerca de una matriz de representación deseada. La información 322 puede tomar la forma, por ejemplo, de una pluralidad de parámetros de representación deseados, que pueden constituir los parámetros de entrada del dispositivo reductor de perturbaciones. El dispositivo reductor de perturbaciones 320 está configurado además para recibir la representación de la señal de mezcla descendente 312 y el flujo de bits SAOC 314, donde el flujo de bits SAOC 314 puede acarrear información paramétrica relacionada con el objeto. El dispositivo reductor de perturbaciones 320 está configurado además para producir una matriz de representación modificada 324 (por ejemplo, en forma de una pluralidad de parámetros de representación ajustados) dependiendo de la información 322 acerca de la matriz de representación deseada.

**[0175]** En consecuencia, el núcleo del decodificador SAOC 310 puede estar configurado para transmitir la representación 316 de la escena renderizada dependiendo de la representación de la señal de mezcla descendente 312, el flujo de bits SAOC 314 y la matriz de representación modificada 324.

**[0176]** A continuación se presentan algunos detalles con respecto a la funcionalidad del decodificador de señales de audio. Se ha descubierto que, para evaluar el riesgo de perturbaciones debidas a las capacidades de separación potencialmente limitadas del sistema SAOC con respecto a una matriz de representación deseada, es conveniente tomar en cuenta tanto la señal de mezcla descendente (descripta por la representación de la señal de mezcla descendente 312) y el flujo de bits SAOC 314. Contando con esta información, es posible intentar mitigar estas perturbaciones, por ejemplo, mediante la modificación de la matriz de representación. Esto es ejecutado por el dispositivo reductor de perturbaciones 320. Las estrategias de mitigación avanzadas toman en cuenta tanto las limitaciones (superposición) de la selectividad en tiempo y frecuencia del sistema SAOC como los efectos perceptuales, es decir, que deberían tratar de lograr que la señal renderizada sea lo más similar posible a la señal de salida deseada y a la vez tener la menor cantidad posible de perturbaciones audibles.

**[0177]** Un enfoque preferido para la reducción de perturbaciones, que es utilizado en el decodificador de señales de audio 300 expuesto en la Fig. 3, se basa en una medida de distorsión global que es una combinación ponderada de las medidas de distorsión que evalúan los diferentes tipos de perturbaciones antes enumeradas. Estas ponderaciones determinan una relación adecuada entre los diferentes tipos de perturbaciones antes enumerados.

Cabe señalar que las ponderaciones para estos tipos de perturbaciones pueden depender de la aplicación en la cual se utiliza el sistema SAOC.

**[0178]** En otras palabras, el dispositivo reductor de perturbaciones 320 puede estar configurado para obtener medidas de distorsión correspondientes a una pluralidad de tipos de perturbaciones. Por ejemplo, el dispositivo reductor de perturbaciones 320 puede aplicar algunas de las medidas de la distorsión $dm_1$ a $dm_6$ descriptas anteriormente. Por otro lado, o además, el dispositivo reductor de perturbaciones 320 puede utilizar otras medidas de distorsión que describen otros tipos de perturbaciones, como se describe dentro de esta sección. Asimismo, el dispositivo reductor de perturbaciones puede estar configurado para obtener la matriz de representación modificada 324 sobre la base de la matriz de representación deseada 322 empleando uno o más de los esquemas de limitación de distorsiones que ya ha sido descrito en párrafos anteriores (por ejemplo, en las secciones 2.4.2, 2.4.3 y 2.4.4), o esquemas de limitación de perturbaciones comparables.

4. Transcodificadores de señales de audio de acuerdo con las Figs. 5a y 5b

4.1 Transcodificador de señales de audio de acuerdo con la Fig. 5a

**[0179]** Se debe notar que los conceptos antes descritos se pueden aplicar tanto a un decodificador de señales de audio como un transcodificador de señales de audio. Tomando como referencia las Figs. 2 y 3, se ha descrito el concepto en combinación con decodificadores de audio. A continuación se describe en forma sucinta el concepto de la presente invención en combinación con transcodificadores de señales de audio.

**[0180]** Con respecto a este tema, se debe tener en cuenta que ya se han descrito las similitudes entre los decodificadores de señales de audio y los transcodificadores de señales de audio con referencia a las Figs. 9a, 9b y 9c, razón por la cual las explicaciones presentadas con respecto a las Figs. 9a, 9b y 9c son aplicables al concepto de la invención.

**[0181]** La Fig. 5a ilustra un diagrama esquemático de bloques de un transcodificador de señales de audio 500 en combinación con un decodificador MPEG Envolvente 510. Como se puede apreciar, el transcodificador de señales de audio 500, que puede consistir en un transcodificador de SAOC a MPEG Envolvente, está configurado para recibir un flujo de bits SAOC 520 y para producir, sobre la base de éste, un flujo de bits MPEG Envolvente 522 sin afectar (ni modificar) la representación de una señal de mezcla descendente 524. El transcodificador de señales de audio 500 comprende un analizador SAOC 530, que está configurado para recibir el flujo de bits SAOC 520 y para extraer los parámetros SAOC deseados del flujo de bits SAOC 530. El transcodificador de señales de audio 500 comprende además un motor de representación de escenas 540, que está configurado para recibir los parámetros SAOC provistos por el analizador SAOC 530 e información de la matriz de representación 542, que puede ser considerada información de (matriz) representación real, y que puede estar representada, por ejemplo, en forma de una pluralidad de parámetros de representación ajustados (o modificados). El motor de representación de escenas 540 está configurado para transmitir el flujo de bits de MPEG Envolvente 522 que depende de dichos parámetros SAOC y la matriz de representación 542. Para este fin, el motor de representación de escenas 540 está configurado para computar los parámetros del flujo de bits MPEG Envolvente 522, que son parámetros relativos a los canales (también designada información paramétrica). Por consiguiente, el motor de representación de escenas 540 está configurado para transformar (o “transcodificar”) los parámetros del flujo de bits SAOC 520, lo que constituye información paramétrica relacionada con el objeto, en los parámetros del flujo de bits MPEG Envolvente, lo que constituye una información paramétrica relacionada con los canales, que depende de la matriz de representación 542.

**[0182]** El transcodificador de señales de audio 500 comprende además la generación de una matriz de representación 550, que está configurado para recibir una información acerca de una matriz de representación deseada, por ejemplo, en forma de información 552 acerca de una configuración de reproducción y una información 554 acerca de posiciones de objetos. Por otro lado, la generación de matrices de representación 550 puede recibir información acerca de los parámetros de representación (por ej., ingresos a la matriz de representación). El generador de matrices de representación también está configurado para recibir el flujo de bits SAOC 520 (o, por lo menos, una subserie de información paramétrica relacionada con el objeto representada por el flujo de bits SAOC 520). El generador de matrices de representación 550 también está configurado para producir la matriz de representación real (ajustada o modificada) 542 sobre la base de la información recibida. En tal medida, el generador de matrices de representación 550 puede asumir la funcionalidad del aparato 100 o del aparato 240.

**[0183]** El decodificador MPEG Envolvente 510 está configurado típicamente para obtener una pluralidad de señales de canales de mezcla ascendente sobre la base de la señal de mezcla descendente información 524 y el flujo de bits MPEG Envolvente 522 provisto por el motor de representación de escenas 540.

**[0184]** Para resumir, el transcodificador de señales de audio 500 está configurado para producir el flujo de bits MPEG Envolvente 522 de tal manera que el flujo de bits MPEG Envolvente 522 da lugar a la producción de una

representación de señal de mezcla ascendente sobre la base de la representación de la señal de mezcla descendente 524, donde la representación de señal de mezcla ascendente provista en realidad por el decodificador MPEG Envolvente 510. El generador de matrices de representación 550 ajusta la matriz de representación 542 utilizada por el motor de representación de escenas 540 de manera que la representación de señal de mezcla ascendente generada por el decodificador MPEG Envolvente 510 no comprenda una distorsión audible inadmisible.

4.2 Transcodificador de señales de audio ce acuerdo con la Fig. 5b

**[0185]** La Fig. 5b ilustra otra disposición de un transcodificador de señales de audio 560 y un decodificador MPEG Envolvente 510. Cabe señalar que la disposición de la Fig. 5b es muy similar a la disposición de la Fig. 5a, razón por la cual se designan los medios y señales idénticos con idénticos números de referencia. El transcodificador de señales de audio 560 difiere del transcodificador de señales de audio 500 en que el transcodificador de señales de audio 560 comprende un transcodificador de mezcla descendente 570, que está configurado para recibir la representación de mezcla descendente de entrada 524 y para producir una representación de mezcla descendente modificada 574, que es alimentada al decodificador MPEG Envolvente 510. La modificación de la representación de la señal de mezcla descendente se realiza para obtener más flexibilidad en la definición del resultado de audio buscado. Esto se debe a que el flujo de bits de MPEG Envolvente 522 no puede representar ciertos mapeos de la señal de entrada del decodificador MPEG Envolvente 510 sobre las señales de canales de mezcla ascendente producidas como salida por el decodificador MPEG Envolvente 510. En consecuencia, la modificación de la representación de la señal de mezcla descendente utilizando el transcodificador de mezcla descendente 570 puede traer aparejada un aumento de la flexibilidad.

**[0186]** Una vez más, el generador de matrices de representación 550 puede asumir la funcionalidad del aparato 100 o el aparato 240, garantizando así que las distorsiones audibles en la representación de la señal de mezcla ascendente provista por el decodificador MPEG Envolvente 510 se mantengan suficientemente bajas.

5. Codificador de Señales de Audio de acuerdo con la Fig. 6

**[0187]** A continuación se describe un codificador de señales de audio 600 tomando como referencia la Fig. 6, que ilustra un diagrama esquemático de bloques de ese tipo de codificador de señales de audio. El codificador de señales de audio 600 está configurado para recibir una pluralidad de señales de objeto 612a, 612N (también designadas $x_1$ a $x_N$) y para producir, sobre la base de éstas, una representación de señal de mezcla descendente 614 e información paramétrica relacionada con el objeto 616. El codificador de señales de audio 600 comprende un dispositivo de mezcla descendente 620 configurado para producir una o más señales de mezcla descendente (que constituyen la representación de la señal de mezcla descendente 614) dependiendo de los coeficientes de mezcla descendente $d_1$ a $d_N$ asociados a las señales de objeto, de tal manera que dichas una o más señales de mezcla descendente comprendan una superposición de una pluralidad de señales de objeto. El codificador de señales de audio 600 comprende además un proveedor de información complementaria 630, que está configurado para producir información complementaria de relación entre objetos que describe las diferencias de nivel y características de correlación de dos o más señales de objeto 612a a 612N. El proveedor de información complementaria 630 también está configurado para suministrar una información complementaria sobre objetos individuales que describe una o más propiedades individuales de las señales de objeto individuales. El codificador de señales de audio 600 produce entonces la información paramétrica relacionada con el objeto 616 por lo que la información paramétrica relacionada con el objeto comprende tanto la información complementaria de relación entre objetos como información complementaria de objetos individuales.

**[0188]** Se ha descubierto que dicha información paramétrica relacionada con el objeto, que describe tanto una relación entre señales de objeto como las características individuales de señales de objetos individuales permite la provisión de una señal de audio multicanal en un decodificador de señales de audio, como se describiera anteriormente. La información complementaria de relación entre objetos puede ser aprovechada por el decodificador de señales de audio para recibir la información paramétrica relacionada con el objeto 616 a fin de extraer, por lo menos aproximadamente, señales de objetos individuales de la representación de la señal de mezcla descendente. La información complementaria sobre objetos individuales, que también está incluida en la información paramétrica relacionada con el objeto 614, puede ser utilizada por el decodificador de señales de audio para verificar si el proceso de mezcla ascendente trae aparejadas distorsiones de señal demasiado fuertes, por lo que es necesario ajustar los parámetros de mezcla ascendente (por ejemplo, los parámetros de representación).

**[0189]** De preferencia, el proveedor de información complementaria 630 está configurado para suministrar la información complementaria sobre objetos individuales, por lo que la información complementaria sobre objetos individuales describe una tonalidad de las señales de objetos individuales. Se ha descubierto que se puede emplear una información de tonalidad como criterio seguro para evaluar si el proceso de mezcla ascendente trae consigo, o no, distorsiones significativas.

**[0190]** Se debe notar asimismo que el codificador de señales de audio 600 puede ser complementado por medio de cualquiera de las características y funcionalidades descriptas en este documento con respecto a los codificadores de señales de audio, y que la representación de la señal de mezcla descendente 614 y la información paramétrica relacionada con el objeto 616 pueden ser provistas por el codificador de señales de audio 600 de tal manera que comprendan las características descriptas con respecto al decodificador de señales de audio de la invención.

6. Flujo de Bits de Audio de acuerdo con la Fig. 7

**[0191]** Una realización de acuerdo con la invención da origen a un flujo de bits de audio 700, una representación esquemática del cual está expuesta en la Fig. 7. El flujo de bits representa una pluralidad de señales de objeto en forma codificada.

**[0192]** El flujo de bits 700 comprende una representación de señal de mezcla descendente710 que representa una o más señales de mezcla descendente, donde por lo menos una de las señales de mezcla descendente comprende una superposición de una pluralidad señales de objeto. El flujo de bits 700 comprende además una información complementaria de relación entre objetos 720 que describe diferencias de nivel y características de correlación de señales de objeto. El flujo de bits comprende asimismo una información complementaria sobre objetos individuales 730 que describe una o más propiedades individuales de las señales de objetos individuales (que constituyen la base de la representación de la señal de mezcla descendente 710).

**[0193]** La información complementaria de relación entre objetos y la información de objetos individuales pueden ser consideradas, en su conjunto, como información complementaria paramétrica relacionada con el objeto.

**[0194]** En una realización preferida, la información complementaria sobre objetos individuales describe tonalidades de las señales de objetos individuales.

**[0195]** Naturalmente, como el flujo de bits 700 es provisto por lo general por un codificador de señales de audio de acuerdo con lo descrito en este documento, y evaluado por un decodificador de señales de audio, que se describe en el presente. El flujo de bits puede comprender las características descriptas con respecto al codificador de señales de audio y el decodificador de señales de audio. En consecuencia, el flujo de bits 700 puede ser muy adecuado para la producción de una señal de audio multicanal utilizando un decodificador de señales de audio, como se describe aquí.

7. Conclusión

**[0196]** Las realizaciones de acuerdo con la invención ofrecen soluciones para reducir o evitar el problema de distorsiones antes explicado, que se originan por el hecho de que no se pueden reconstruir las señales de objetos únicos originales perfectamente de las pocas señales de mezcla descendente transmitidas. Hay otras soluciones sencillas que se aplican para solucionar este problema:

- Una estrategia sencilla consistiría en limitar el rango de ganancia relativa de los objetos a, por ej. +/– 12dB. Aunque es cierto que la configuración de grandes ganancias de objeto pueden llevar a degradaciones audibles (por ejemplo: incrementar la ganancia un objeto en 20dB dejando, al mismo tiempo, los demás niveles de objeto a 0dB): esto no es necesario, sin embargo. Por ejemplo, si se incrementa la ganancia de todos los niveles relativos de los objetos en el mismo factor se produce una salida del sistema sin alteraciones.

- Un enfoque más elaborado consistiría en considerar las diferencias de los niveles relativos de los objetos. Para la representación de dos objetos de audio, la diferencia de los niveles relativos de ambos objetos ofrece un gancho para las posibles degradaciones de la salida renderizada. Sin embargo, no está claro cómo se generaliza esta idea a más de dos objetos de audio renderizados.

**[0197]** En vista de esta situación, las realizaciones de acuerdo con la presente invención ofrecen medios para abordar este problema y, de esa manera, prevenir una experiencia desfavorable del usuario. Algunas realizaciones de acuerdo con la presente invención pueden dar lugar a soluciones aun más elaboradas que las descriptas en la sección anterior.

**[0198]** En consecuencia, se puede obtener una buena impresión auditiva utilizando la presente invención, aun en el caso en que un usuario provee parámetros de representación inapropiados.

**[0199]** En términos generales, las realizaciones de acuerdo con la invención se relacionan con un aparato, un procedimiento o un programa de computación para codificar una señal de audio o para decodificar una señal de audio codificada, o con una señal de audio codificada (por ejemplo, en forma de flujo de bits de audio) de acuerdo con lo descrito anteriormente.

8. Alternativas de Implementación

**[0200]** Si bien se han descrito ciertos aspectos en el contexto de un aparato, es claro que estos aspectos representan asimismo una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a un paso del procedimiento o una característica de un paso del procedimiento. De manera análoga, los aspectos descritos en el contexto de un paso del procedimiento también representan una descripción de un bloque o elemento o característica correspondiente de un aparato respectivo. Algunos o todos los pasos del procedimiento pueden ser ejecutados (o utilizados) por un aparato de hardware, como por ejemplo un microprocesador, una computadora programable o un circuito electrónico. En algunas realizaciones, uno o más de los pasos más importantes del procedimiento pueden ser ejecutados por ese aparato.

**[0201]** La señal de audio codificada o el flujo de bits de audio de la invención puede ser almacenada en un medio de almacenamiento digital o pueden ser transmitidos en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión alámbrico tal como la internet.

**[0202]** Dependiendo de ciertos requisitos 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.

**[0203]** 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 descritos.

**[0204]** 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.

**[0205]** Otras realizaciones comprenden el programa de computación para ejecutar uno de los procedimientos aquí descritos, almacenados en un portador legible por una máquina.

**[0206]** 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 descritos al ejecutarse el programa de computación en una computadora.

**[0207]** 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 descritos en la presente.

**[0208]** 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 descritos. El flujo de datos o la secuencia de señales puede estar configurada, por ejemplo, para ser transferido a través de una conexión de comunicación de datos, por ejemplo a través de la Internet.

**[0209]** 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 descritos.

**[0210]** Otra realización comprende una computadora en la cual se ha instalado el programa de computación para ejecutar uno de los procedimientos aquí descritos.

**[0211]** 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 descritos. En algunas realizaciones, una matriz de puertas programable en el campo puede cooperar con un microprocesador para ejecutar uno de los procedimientos aquí descritos. En general, los procedimientos se ejecutan preferentemente mediante cualquier aparato de hardware.

**[0212]** 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 descritos 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.

Referencias

**[0213]**

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

[JSC] C. Faller, “Parametric Joint–Coding of Audio Sources”, 120a Convención de AES, París, 2006, Preprint 6752

[SAOC1] J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC To SAOC – Recent Developments in Parametric Coding of Spatial Audio", 22a Conferencia Regional en el RU de AES, Cambridge, RU, April 2007

[SAOC2] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers y W. Oomen: " Spatial Objeto de audio Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124a Convención de AES, Amsterdam 2008, Preimpresión 7377

## REIVINDICACIONES

**1.** Aparato (100; 240; 320; 550) para suministrar uno o varios parámetros ajustados (120; 222; 324; *rm', r*1*im,m*) para un suministro de una representación de señal de mezcla ascendente ( $\hat{y}_1$ a $\hat{y}_N$; 316; 522, 524; 522, 574), que es una representación de señal audio de mezcla ascendente, sobre la base de una representación de señal de mezcla ascendente (212; 312; 524), en el cual varias señales de objeto de audio se mezclan hacia abajo para obtener la señal de mezcla descendente, y una información paramétrica relativa al objeto (214; 314; 520), comprendiendo el aparato:

un ajustador de parámetros (140; 240) configurado para recibir uno o varios parámetros de entrada (110; 242; 322; 552, 554; *ri*) y para suministrar, sobre la base de estos últimos, uno o varios parámetros ajustados (120; 222; 324; *r*1*im,m*),
en el cual el ajustador de parámetros está configurado para suministrar los uno o varios parámetros ajustados en función de los uno o varios parámetros de entrada y de la información paramétrica relativa al objeto (130; 214a, 214b, 214c; 314; 520), de manera que una deformación de la representación de señal de mezcla ascendente provocada por la utilización de parámetros no óptimos se reduzca al menos para los parámetros de entrada que se separan de los parámetros óptimos en más de una separación predeterminada;
en el cual el aparato está configurado para recibir, como parámetros de entrada (110; 242; 322; 552, 554; *ri*), los parámetros de representación deseados que describen un escalonamiento de intensidad deseada de una pluralidad de señales de objeto de audio (*x*1 a *xN*) en uno o varios canales de audio descritos por la representación de señal de mezcla ascendente ( $\hat{y}_1$ a $\hat{y}_N$; 316; 522, 524; 522, 574); y
en el cual el ajustador de parámetros está configurado para suministrar, como parámetros ajustados (120, 222; 324; *rm*', *r*1*im,m*), uno o varios parámetros de representación reales en función de los uno o varios parámetros de representación deseados y de la información paramétrica relativa al objeto.

**2.** Aparato según la reivindicación 1, en el cual el ajustador de parámetros está configurado para obtener uno o varios valores límite de parámetros de representación en función de la información paramétrica relativa al objeto (130; 214a, 214b, 214c; 314; 520) y de una información de mezcla descendente (214b; *di*), que describen una contribución de las señales de objeto de audio (*x*1 a *xN*) de la representación de señal de mezcla descendente, de manera que una métrica de distorsión (dm1(m),dm2(m),dm5(m),dm6(m), DM1, DM2, DM3, DM4, DM5,DM6) se sitúa en un intervalo predeterminado para la representación de valores de los parámetros que satisfacen los límites definidos por los valores límite de parámetros de representación, y en el cual el ajustador de parámetros está configurado para obtener los parámetros de representación reales (*r*1*im,m*) en función de los parámetros de representación deseados (*r*1) y de los uno o varios valores límite de parámetro de representación, de manera que los parámetros de representación reales satisfagan los límites definidos por los valores límite de parámetro de representación.

**3.** Aparato según la reivindicación 2, en el cual el ajustador de parámetros está configurado para obtener los uno o varios valores límite de parámetro de representación ( $\hat{r}_m^2$ ) de manera que una contribución relativa de una señal de objeto (*x*1 a *xN*) en una superposición representada de una pluralidad de señales de objetos, representadas con ayuda de uno o varios parámetros de representación (*rm'*, *r*1*im,m*) que obedecen a los uno o varios valores límite de parámetro de representación, difiere de una contribución relativa de la señal de objeto (*x*1 a *xN)* en una señal de mezcla descendente (212; 312; 524) de no más de una diferencia predeterminada.

**4.** Aparato según la reivindicación 3, en el cual el ajustador de parámetros está configurado para determinar los uno o varios valores de parámetro de representación ( $\hat{r}_m^2$ ) de manera que la ecuación

$$dm_1(m) = \frac{r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i}{d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i} \leq T$$

se satisfaga para uno o varios objetos de audio designados por un índice de objeto m, donde *rm* designa un parámetro de representación que describe una contribución de una señal de objeto de un objeto de audio que presenta el índice de objeto m en un canal determinado (*y*1 a *yN*) de la señal de mezcla ascendente, donde *dm* designa un parámetro de mezcla descendente que describe una contribución de la señal de objeto (*x*1 a *xN*) del objeto que presenta el índice m en una señal de mezcla descendente, y donde *Xi* designa una medida de energía del objeto de audio que presenta el índice de objeto m, medida de energía que se determina mediante la información paramétrica relativa al objeto.

**5.** Aparato según la reivindicación 2, en el cual el ajustador de parámetros está configurado para obtener los uno o varios valores límite de parámetro de representación ($\hat{r}_m^2$) de manera que una medida de distorsión (DM3) que describe una coherencia entre una señal de mezcla descendente esté descrita por la representación de señal de mezcla descendente y que una señal representación, representada con ayuda de uno o varios parámetros de representación (*rm*) que satisfacen uno o varios valores límite de parámetro de representación ($\hat{r}_m^2$) se sitúa en un intervalo predeterminado.

**6.** Aparato según la reivindicación 5, en el cual el ajustador de parámetros está configurado para obtener uno o varios valores límite de parámetro de representación $\hat{r}_m^2$ de manera que la medida de distorsión

$$\mathrm{DM}_3 = 1 - \min\left( \frac{|c_{12}|}{\sqrt{c_{11} \cdot c_{22}}},\ 1 \right)$$

adopte un valor predeterminado, donde **C** se define como siendo

$$\mathbf{C} = \mathbf{M} \cdot \mathbf{E} \cdot \mathbf{M}^* = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

Donde

$$\mathbf{M} = \begin{pmatrix} r_1 & r_2 & \cdots & r_N \\ d_1 & d_2 & \cdots & d_N \end{pmatrix}$$

es una matriz que comprende una primera fila de parámetros de representación *r*1 a *rn* y una segunda fila de parámetros de mezcla descendente *d*1 a *dn* que describe una contribución de las señales de objeto de audio a la representación de señal de mezcla descendente; donde **E** es una matriz de covarianza de objeto que se obtiene con ayuda de los parámetros (OLD, IOC) de la información paramétrica relativa al objeto, y donde "*" designa un operador conjugado complejo.

**7.** Aparato según la reivindicación 1, en el cual el ajustador de parámetros está configurado para calcular una combinación lineal entre un cuadrado de un parámetro de representación deseado (*rm*) y un cuadrado de un parámetro de representación óptimo (*ropt,m*), para obtener el parámetro de representación real (*r*1*im,m*), en el cual el ajustador de parámetros está configurado para determinar una contribución del parámetro de representación deseado (*rm*) y del parámetro de representación óptimo (*ropt,m*) a la combinación lineal en función de un parámetro de umbral predeterminado T y de una métrica de distorsión (*dm*1, *dm*2, *dm*3, *dm*4, *dm*5, *dm*6), en el cual la métrica de distorsión describe una distorsión que sería provocada por la utilización de los uno o varios parámetros de representación deseados (*rm*) antes que parámetros de representación óptimos (*ropt,m*), para obtener la representación de señal de mezcla ascendente sobre la base de la representación de señal de mezcla descendente.

**8.** Aparato según la reivindicación 7, en el cual el ajustador de parámetros está configurado para evaluar la ecuación

$$r_{lim,m}^2 = \frac{T-1}{dm_x(m)}\left(r_m^2 - r_{opt,m}^2\right) + r_{opt,m}^2$$

para obtener el parámetro de representación real *r*1*im,m* que describe una contribución de una señal de objeto de un objeto que presenta el índice de objeto m en un canal determinado de la señal de mezcla ascendente,
donde T designa un parámetro de umbral de distorsión predeterminado,
donde *dmx* (m) designa una métrica de distorsión asociada al parámetro de representación deseado *rm* que describe una contribución deseada de una señal de objeto de un objeto de audio que presenta el índice de objeto m en un canal determinado de la señal de mezcla ascendente;
donde *ropt,m* designa un parámetro de representación óptimo que describe una contribución óptima de una señal de objeto del objeto de audio que presenta un índice de objeto m en el canal determinado de la señal de mezcla ascendente.

**9.** Aparato según la reivindicación 7 o la reivindicación 8, en el cual el ajustador de parámetros está configurado para obtener la métrica de distorsión de manera que la métrica de distorsión dependa de una relación entre una contribución relativa de una señal de objeto determinada en una superposición representada de una pluralidad de señales de objeto, representada según los parámetros de representación deseados, y una contribución relativa de la señal de objeto determinada en una señal de mezcla descendente que comprende la señal de objeto determinada.

**10.** Aparato según la reivindicación 7, 8 o la 9, en el cual el ajustador de parámetros está configurado para obtener la métrica de distorsión (dm1) de manera que la métrica de distorsión dependa de una relación entre una contribución relativa de una señal de objeto determinada (*x*1 a *xN*) en una superposición representada de una pluralidad de

señales de objeto, representada en función de los parámetros de representación deseados (*rm*), y una contribución relativa de la señal de objeto del objeto determinada (*x*1 a *xN*) en una señal de mezcla descendente que comprende la señal de objeto determinada (*x*1 a *xN*).

**11.** Aparato según cualquiera de las reivindicaciones 7 a 10, en el cual el ajustador de parámetros está configurado para calcular la métrica de distorsión *dmx* (m) según

$$dm_x(m) = dm_1(m) = \frac{r_m^2 \sum_{i=1}^{N_{ob}} d_i^2 X_i}{d_m^2 \sum_{i=1}^{N_{ob}} r_i^2 X_i}$$

donde *rm y ri* designan los parámetros de representación deseados asociados a los objetos de audio que presentan los índices de objeto respectivamente m y *i*;
donde *dm y di* designan parámetros de mezcla descendente que describen una contribución de señales de objeto de los objetos de audio que presentan los índices de objeto respectivamente m y i a una señal de mezcla descendente de la representación de señal de mezcla descendente;
donde *Nob* designa un número de objetos de audio considerados;
donde *Xi* designa las medidas de energía asociadas a las señales de objeto de los objetos de audio que presentan los índices de objeto i.

**12.** Aparato según la reivindicación 7, 8 o la 9, en el cual el ajustador de parámetros está configurado para obtener la métrica de distorsión (*dm2*) de manera que la métrica de distorsión dependa de una diferencia entre una contribución relativa de una señal de objeto determinada (*x*1 a *xN*) en una superposición representada de una pluralidad de señales de objeto, representada en función de los parámetros de representación deseados (*rm*), y de una contribución relativa de la señal de objeto determinada (*x*1 a *xN*) en una señal de mezcla descendente que comprende la señal de objeto determinada (*x*1 a *xN*).

**13.** Aparato según cualquiera de las reivindicaciones 7 a 12, en el cual el ajustador de parámetros está configurado para calcular la métrica de distorsión (*dm2*) de manera que la métrica de distorsión dependa de una relación máscara-señal (msr), de manera que la métrica de distorsión (*dm*2) disminuye, indicando una distorsión más reducida, si la relación máscara-señal aumenta.

**14.** Aparato según cualquiera de las reivindicaciones 7 a 11, en el cual el ajustador de parámetros está configurado para calcular la métrica de distorsión según

$$dm_x(m) = dm_2(m) = \frac{\left( r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i - d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i \right) \cdot X_m}{msr \cdot (\sum_{i=1}^{N} r_i^2 \cdot X_i) \cdot (\sum_{i=1}^{N} d_i^2 \cdot X_i)}$$

O

$$dm_x = dm_2'(m) = \frac{P_{ruido}}{Máscara} = \frac{E\{|y_{m;ideal} - \hat{y}_{m;real}|^2\}}{msr \cdot P_{total}} =$$

$$\frac{\left| r_m^2 \cdot \sum_{i=1}^{N} d_i^2 \cdot X_i + d_m^2 \cdot \sum_{i=1}^{N} r_i^2 \cdot X_i - 2 \cdot d_m r_m \cdot \sqrt{(\sum_{i=1}^{N} r_i^2 \cdot X_i) \cdot (\sum_{i=1}^{N} d_i^2 \cdot X_i)} \right| \cdot X_m}{msr \cdot (\sum_{i=1}^{N} r_i^2 \cdot X_i) \cdot (\sum_{i=1}^{N} d_i^2 \cdot X_i)}$$

donde *rm y ri* designan los parámetros de representación deseados asociados a unos objetos de audio que tienen índices de objeto m y i, respectivamente;
donde *dm y di* designan los parámetros de mezcla descendente que describen una contribución de señales de objeto de los objetos de audio que presentan los índices de objeto respectivamente m y i a una señal de mezcla descendente de la representación de señal de mezcla descendente;
donde N designa un número de objetos de audio considerados; donde *Xi y Xm* designan las medidas de energía asociadas a las señales de objeto de los objetos de audio que presentan los índices de objetos respectivamente i y m; y
donde msr define una relación máscara-señal.

**15.** Aparato según cualquiera de las reivindicaciones 1 a 14, en el cual el ajustador de parámetros está configurado para suministrar los uno o varios parámetros ajustados en función de una medida de cálculo de la degradación perceptual, de manera que una distorsión evaluada de manera perceptual de la representación de señal de mezcla ascendente provocada por la utilización de parámetros no óptimos y representada por la medida de cálculo de la degradación perceptual sea limitada.

**16.** Aparato según cualquiera de las reivindicaciones 1 a 15, en el cual el ajustador de parámetros está configurado para recibir una información de propiedades de objeto individuales que describen las propiedades individuales de una o varias señales de objeto originales que constituyen la base de una señal de mezcla descendente descrita por la representación de señal de mezcla descendente; y
en el cual el ajustador de parámetros está configurado para considerar las informaciones de propiedades de objeto individuales, y para suministrar los parámetros ajustados de manera que una distorsión de la representación de señal de mezcla ascendente con respecto a una representación de señal de mezcla ascendente representada idealmente se reduzca a menos para los parámetros de entrada que se separan de los parámetros óptimos en más de una separación predeterminada.

**17.** Aparato según la reivindicación 16, en el cual el ajustador de parámetros está configurado para recibir y considerar, a título de información de propiedades de objeto individuales, una información de tonalidad de señal de objeto, para suministrar los uno o varios parámetros ajustados.

**18.** Aparato según la reivindicación 17, en el cual el ajustador de parámetros está configurado para estimar una tonalidad (N) de una señal de mezcla ascendente representada idealmente en función de la información de tonalidad de señal de objeto recibida y de la información de energía de objeto recibida (OLD, P); y
en el cual el ajustador de parámetros está configurado para suministrar los uno o varios parámetros ajustados para reducir una diferencia entre la tonalidad estimada y la tonalidad de una señal de mezcla ascendente obtenida con ayuda de los uno o varios parámetros ajustados, en comparación con una diferencia entre la tonalidad estimada y una tonalidad de una señal de mezcla ascendente obtenida con ayuda de los uno o varios parámetros de entrada, o para mantener una diferencia entre la tonalidad estimada y una tonalidad de una señal de mezcla ascendente obtenida con ayuda de los uno o varios parámetros ajustados en un intervalo predeterminado.

**19.** Aparato según cualquiera de las reivindicaciones 1 a 18, en el cual el ajustador de parámetros está configurado para realizar un ajuste variable en el tiempo y en frecuencia de los parámetros de entrada.

**20.** Aparato según cualquiera de las reivindicaciones 1 a 19, en el cual el ajustador de parámetros está configurado para considerar también la representación de señal de mezcla descendente para suministrar los uno o varios parámetros ajustados.

**21.** Aparato según cualquiera de las reivindicaciones 1 a 20, en el cual el ajustador de parámetros está configurado para obtener una medida de deformación global que es una combinación ponderada de las medidas de distorsión que describen una pluralidad de tipos de artefactos;
en el cual el ajustador de parámetros está configurado para obtener la medida de deformación global de manera que la medida de deformación global sea una medida de deformaciones que estarían provocadas por la utilización de uno o varios de los parámetros de representación de entrada antes que parámetros de representación óptimos, para obtener la representación de señal de mezcla ascendente sobre la base de la representación de señal de mezcla descendente.

**22.** Aparato según la reivindicación 21, en el cual el ajustador de parámetros está configurado para combinar al menos dos de las medidas de distorsión siguientes, para obtener la medida de deformación global:

- una medida que describe un cambio parásito de timbre de un objeto de audio;
- una medida que describe una modulación parásito de una señal de objeto asociada a un objeto de audio;
- una medida que describe la presencia de un sonido musical parásito;
- una medida que describe la presencia de un ruido modulado parásito.

**23.** Descodificador de señal audio (220, 240; 300) para suministrar, como representación de señal de mezcla ascendente, una pluralidad de canales de audio de mezcla ascendente ($\hat{y}_1$ a $\hat{y}_N$; 316) sobre la base de una representación de señal de mezcla descendente (212; 312), de una información paramétrica relativa al objeto (214; 314) y de una información de representación deseada (242; 322), comprendiendo el descodificador de señal audio:

un mezclador ascendente (220; 310) configurado para obtener los canales de audio mezclados hacia arriba ($\hat{y}_1$ a $\hat{y}_N$; 316) sobre la base de la representación de señal de mezcla descendente (212; 312) y en función de la información paramétrica relativa al objeto (214; 314) y una información de representación real (222; 324) que

describe una atribución de una pluralidad de señales de objeto de los objetos de audio descritos por la información paramétrica relativa al objeto en los canales de audio de mezcla ascendente; y
un aparato (100; 240; 320) para suministrar uno o varios parámetros ajustados, según cualquiera de las reivindicaciones 1 a 22,
en el cual el aparato para suministrar uno o varios parámetros ajustados está configurado para recibir la información de representación deseada (242; 322) como los uno o varios parámetros de entrada (110) y para suministrar los uno o varios parámetros ajustados (222, 324) como la información de representación real; y en el cual el aparato para suministrar los uno o varios parámetros ajustados está configurado para suministrar los uno o varios parámetros ajustados de manera que las distorsiones de los canales de audio de mezcla ascendente ($\hat{y}_1$ a $\hat{y}_N$, 316) provocadas por la utilización de los parámetros de representación real (*rm*', *r*1*im,m*), que se separan de los parámetros de representación óptimos (*ropt,m*) sean reducidas al menos para los parámetros de representación deseados (*ri*) que se separan de los parámetros de representación óptimos (*ropt,m*) en más de una separación predeterminada.

**24.** Transcodificador de señal audio (500; 560) para suministrar, como representación de señal de mezcla ascendente (522), una información paramétrica relativa al canal sobre la base de una representación de señal de mezcla descendente (524), de una información paramétrica relativa al objeto (520) y de una información de representación deseada (552, 554), comprendiendo el transcodificador de señal audio:

un transcodificador de información lateral (540) configurado para obtener la información paramétrica relativa al canal (522) sobre la base de la representación de señal de mezcla descendente (524) y en función de la información paramétrica relativa al objeto (520) y de una información de representación real (542) que describe una atribución de una pluralidad de señales de objeto de los objetos de audio descritos por la información paramétrica relativa al objeto (522) en los canales de audio de mezcla ascendente descritos por la información paramétrica relativa al canal; y
un aparato (100; 550) para suministrar uno o varios parámetros ajustados (542) según cualquiera de las reivindicaciones 1 a 22, en el cual el aparato para suministrar uno o varios parámetros ajustados está configurado para recibir la información de representación deseada (552, 554) como los uno o varios parámetros de entrada (110) y para suministrar los uno o varios parámetros ajustados (120) como información de representación real (542); y
en el cual el aparato para suministrar los uno o varios parámetros ajustados está configurado para suministrar los uno o varios parámetros ajustados (120) de manera que las distorsiones de los canales de audio de mezcla ascendente provocadas por la utilización de los parámetros de representación reales (542), que se separan de los parámetros de representación óptimos, sean reducidas al menos para los parámetros de representación deseados (552, 554) que se separan de los parámetros de representación óptimos en más de una separación predeterminada.

**25.** Procedimiento para suministrar uno o varios parámetros ajustados para un suministro de una representación de señal de mezcla ascendente, que es una representación de señal audio de mezcla ascendente, sobre la base de una representación de señal de mezcla descendente, en el cual varias señales de objeto de audio se mezclan hacia abajo para obtener la señal de mezcla descendente, y de una información paramétrica relativa al objeto, comprendiendo el procedimiento:

recibir uno o varios parámetros de entrada y suministrar, sobre la base de estos últimos, uno o varios parámetros ajustados,
en el cual los uno o varios parámetros ajustados se suministran en función de los uno o varios parámetros de entrada y de la información paramétrica relativa al objeto, de manera que una deformación de la representación de señal de mezcla ascendente provocada por la utilización de parámetros no óptimos se reduzca al menos para los parámetros de entrada que se separan de los parámetros óptimos en más de una separación predeterminada;
en el cual los parámetros de representación deseados que describen un escalonamiento de intensidad deseado de una pluralidad de señales de objeto de audio en uno o varios canales de audio descritos por la representación de señal mezcla ascendente son recibidos como parámetros de entrada; y
en el cual uno o varios parámetros de representación reales se suministran, como parámetros ajustados, en función de los uno o varios parámetros de representación deseados y de la información paramétrica relativa al objeto.

**26.** Procedimiento para suministrar, como representación de señal de mezcla ascendente, una pluralidad de canales de audio de mezcla ascendente sobre la base de una representación de señal de mezcla descendente, de una información paramétrica relativa al objeto y de una información de representación deseada, comprendiendo el procedimiento:

suministrar uno o varios parámetros ajustados, según el procedimiento de la reivindicación 25, en el cual la información de representación deseada es recibida como los uno o varios parámetros de entrada y en el cual los uno o varios parámetros ajustados se suministran como información de representación real, y en el cual los uno o varios parámetros ajustados se suministran de manera que las distorsiones de los canales de audio de mezcla ascendente provocadas por la utilización de los parámetros de representación reales, que se separan de los parámetros de

representación óptimo, sean reducidas al menos para los parámetros de representación deseado que se separan de los parámetros de representación óptimos en más de una separación predeterminada; y
obtener los canales de audio de mezcla ascendente sobre la base de la representación de señal de mezcla descendente y en función de la información paramétrica relativa al objeto y de la información de representación real que describe una atribución de una pluralidad de señales de objeto de los objetos de audio descritos por la información paramétrica relativa al objeto a los canales de audio de mezcla ascendente.

**27.** Procedimiento para suministrar, como representación de señal de mezcla ascendente, una información paramétrica relativa al canal sobre la base de una representación de señal de mezcla descendente, una información paramétrica relativa al objeto y una información de representación deseada, comprendiendo el procedimiento:

suministrar uno o varios parámetros ajustados según el procedimiento de la reivindicación 25, donde la información de representación deseada es recibida como los uno o varios parámetros de entrada y donde los uno o varios parámetros ajustados se suministran como información de representación real, y donde los uno o varios parámetros ajustados se suministran de manera que las distorsiones de los canales de audio de mezcla ascendente provocadas por la utilización de los parámetros de representación reales, que se separan de los parámetros de representación óptimo, sean reducidas al menos para los parámetros de representación deseados que se separan de los parámetros de representación óptimos en más de una separación predeterminada; y
obtener la información paramétrica relativa al canal, que describe los canales de audio de mezcla ascendente, sobre la base de la representación de señal de mezcla descendente y en función de la información paramétrica relativa al objeto y de la información de representación real que describe una atribución de una pluralidad de señales de objeto de los objetos de audio descritos por la información paramétrica relativa al objeto a los canales de audio de mezcla ascendente, canales de audio de mezcla ascendente que se describen por la información paramétrica relativa al canal.

**28.** Programa de ordenador adaptado para realizar, cuando se ejecuta en un ordenador, uno de los procesos según cualquiera de las reivindicaciones 25, 26 o 27.

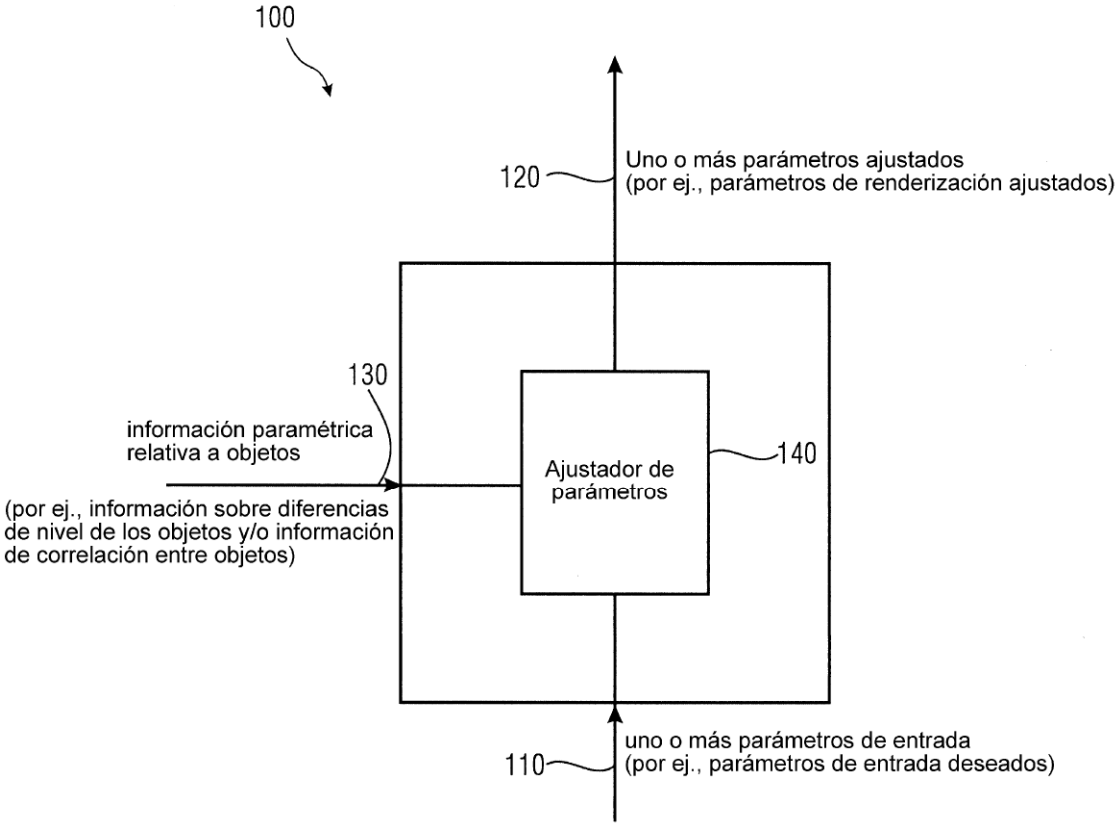

100
120
Uno o más parámetros ajustados
(por ej., parámetros de renderización ajustados)
130
información paramétrica
relativa a objetos
(por ej., información sobre diferencias
de nivel de los objetos y/o información
de correlación entre objetos)
Ajustador de
parámetros
140
110
uno o más parámetros de entrada
(por ej., parámetros de entrada deseados)

**FIG. 1**

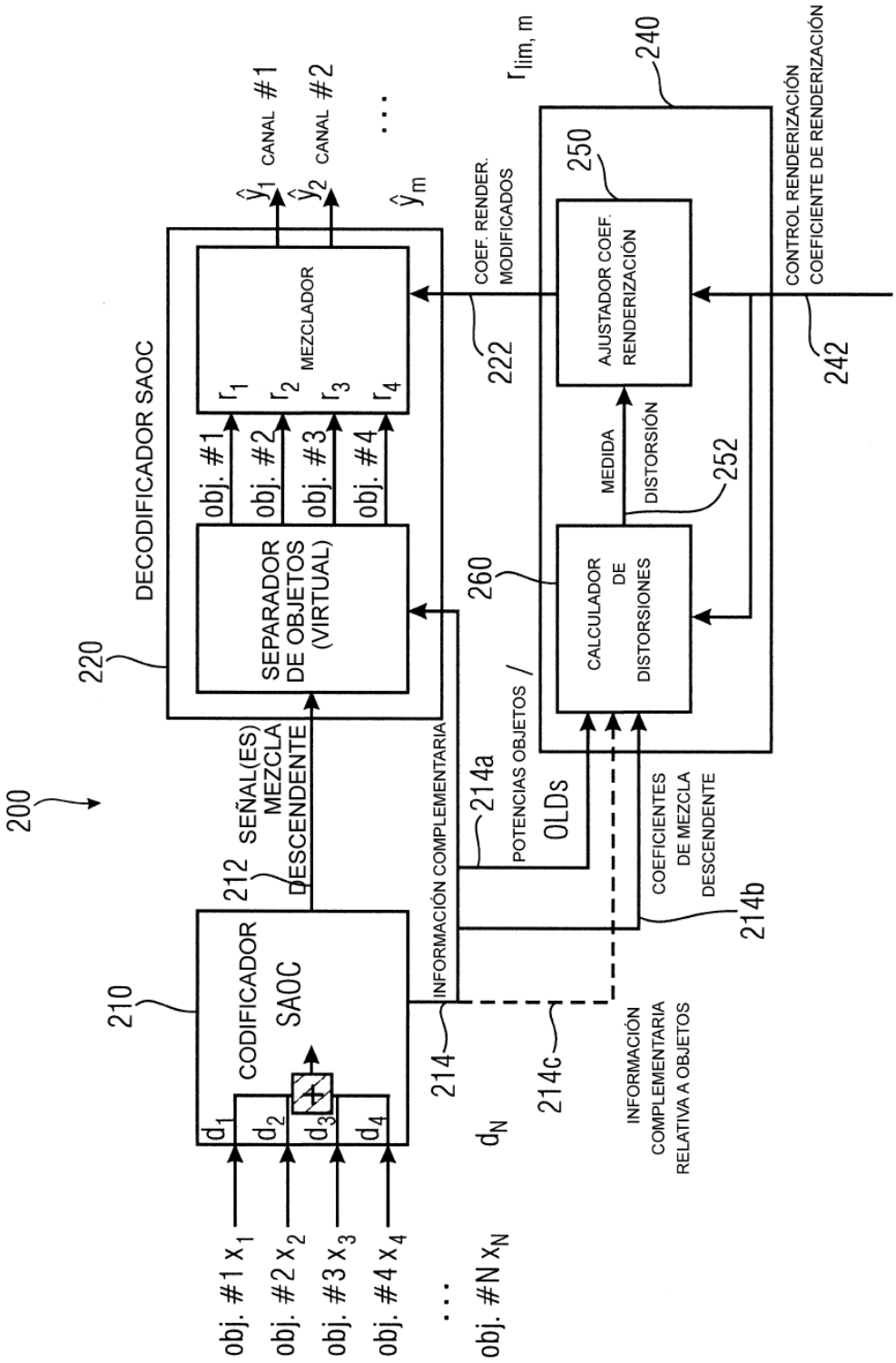

200
210
CODIFICADOR SAOC
212
SEÑAL(ES) MEZCLA DESCENDENTE
220
DECODIFICADOR SAOC
SEPARADOR DE OBJETOS (VIRTUAL)
obj. #1
obj. #2
obj. #3
obj. #4
MEZCLADOR
CANAL #1
CANAL #2
obj. #1 x1
obj. #2 x2
obj. #3 x3
obj. #4 x4
obj. #N xN
d1
d2
d3
d4
dN
214
INFORMACIÓN COMPLEMENTARIA
214a
POTENCIAS OBJETOS / OLDs
214b
COEFICIENTES DE MEZCLA DESCENDENTE
214c
INFORMACIÓN COMPLEMENTARIA RELATIVA A OBJETOS
222
COEF. RENDER. MODIFICADOS
240
250
AJUSTADOR COEF. RENDERIZACIÓN
252
MEDIDA DISTORSIÓN
260
CALCULADOR DE DISTORSIONES
242
CONTROL RENDERIZACIÓN COEFICIENTE DE RENDERIZACIÓN

FIG. 2

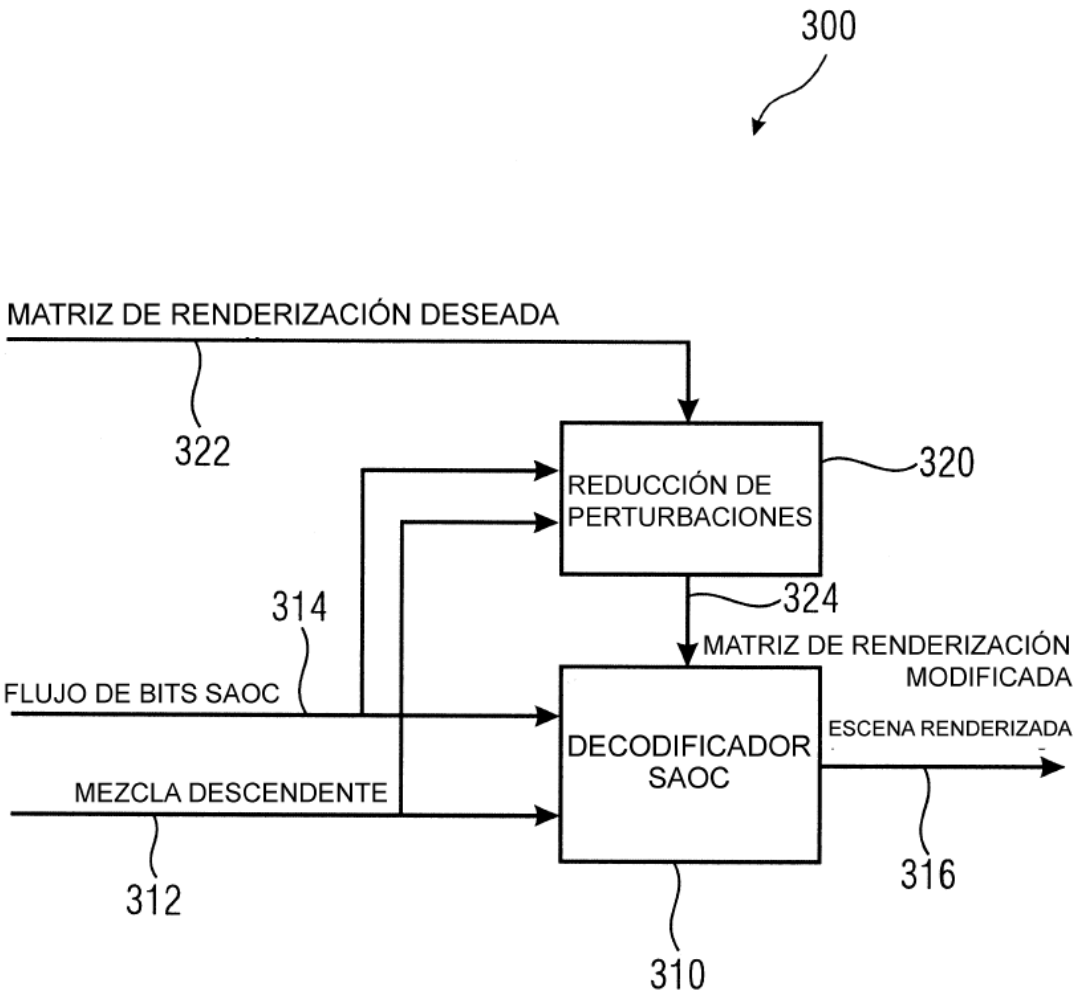

300
MATRIZ DE RENDERIZACIÓN DESEADA
322
REDUCCIÓN DE PERTURBACIONES
320
314
324
MATRIZ DE RENDERIZACIÓN MODIFICADA
FLUJO DE BITS SAOC
DECODIFICADOR SAOC
ESCENA RENDERIZADA
MEZCLA DESCENDENTE
312
316
310

FIG. 3

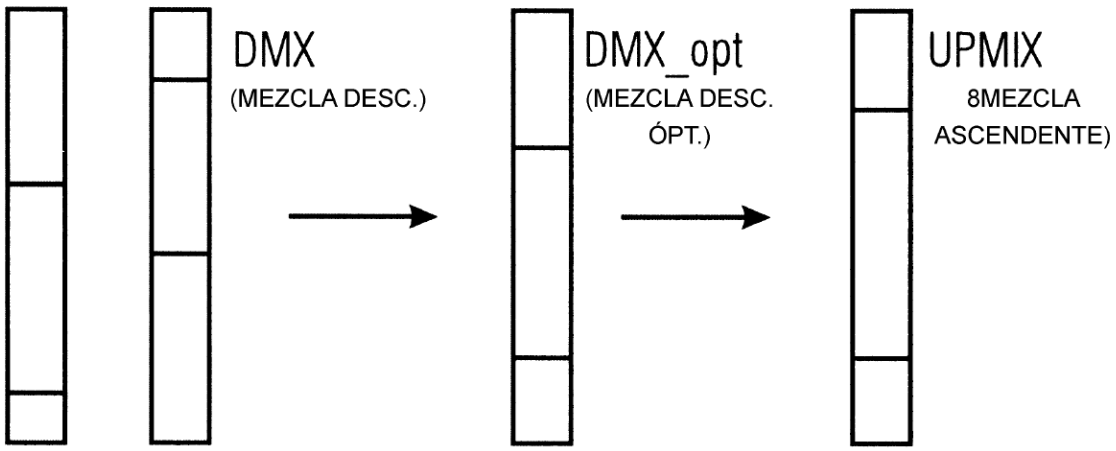
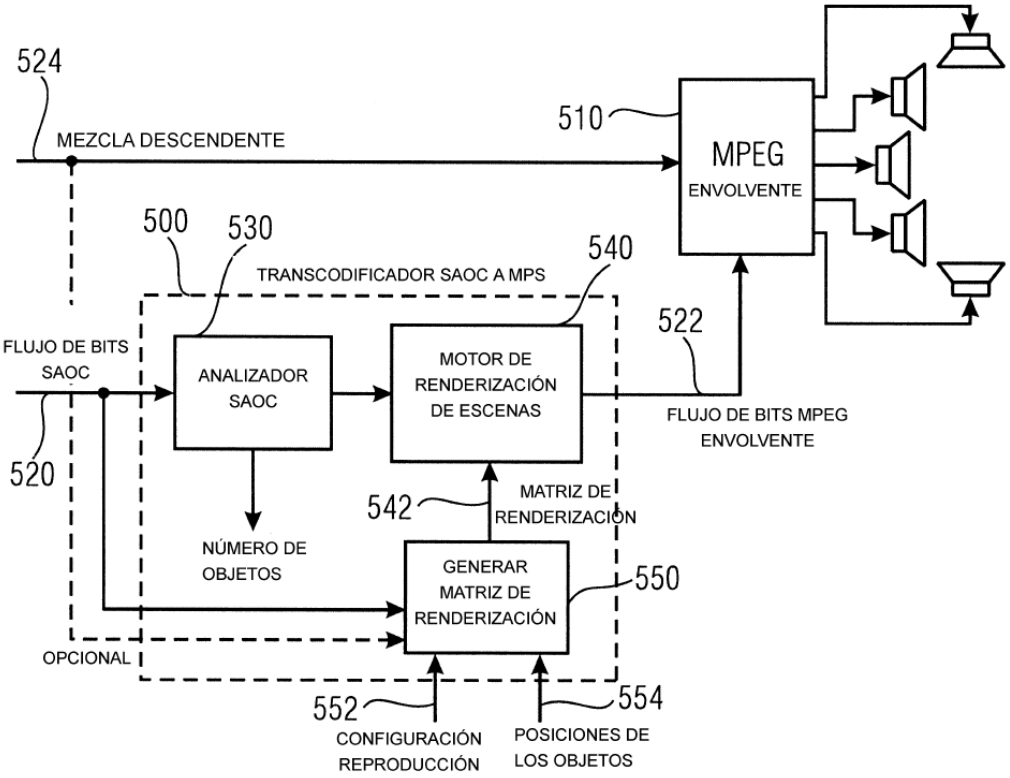
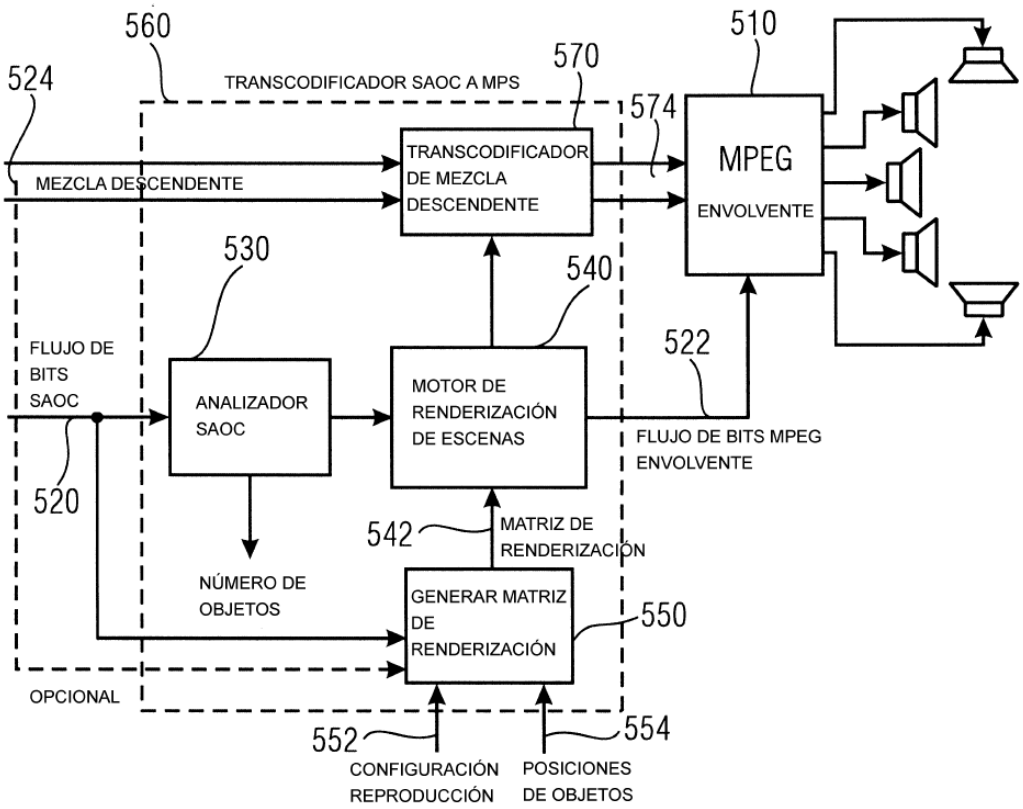

DMX
(MEZCLA DESC.)
DMX_opt
(MEZCLA DESC.
ÓPT.)
UPMIX
8MEZCLA
ASCENDENTE)

FIG. 4




(A) TRANSCODIFICADOR MONO BASADO EN MEZCLA DESCENDENTE

**FIG. 5 A**




(B) TRANSCODIFICADOR ESTEREO BASADO EN MEZCLA DESCENDENTE

**FIG. 5 B**

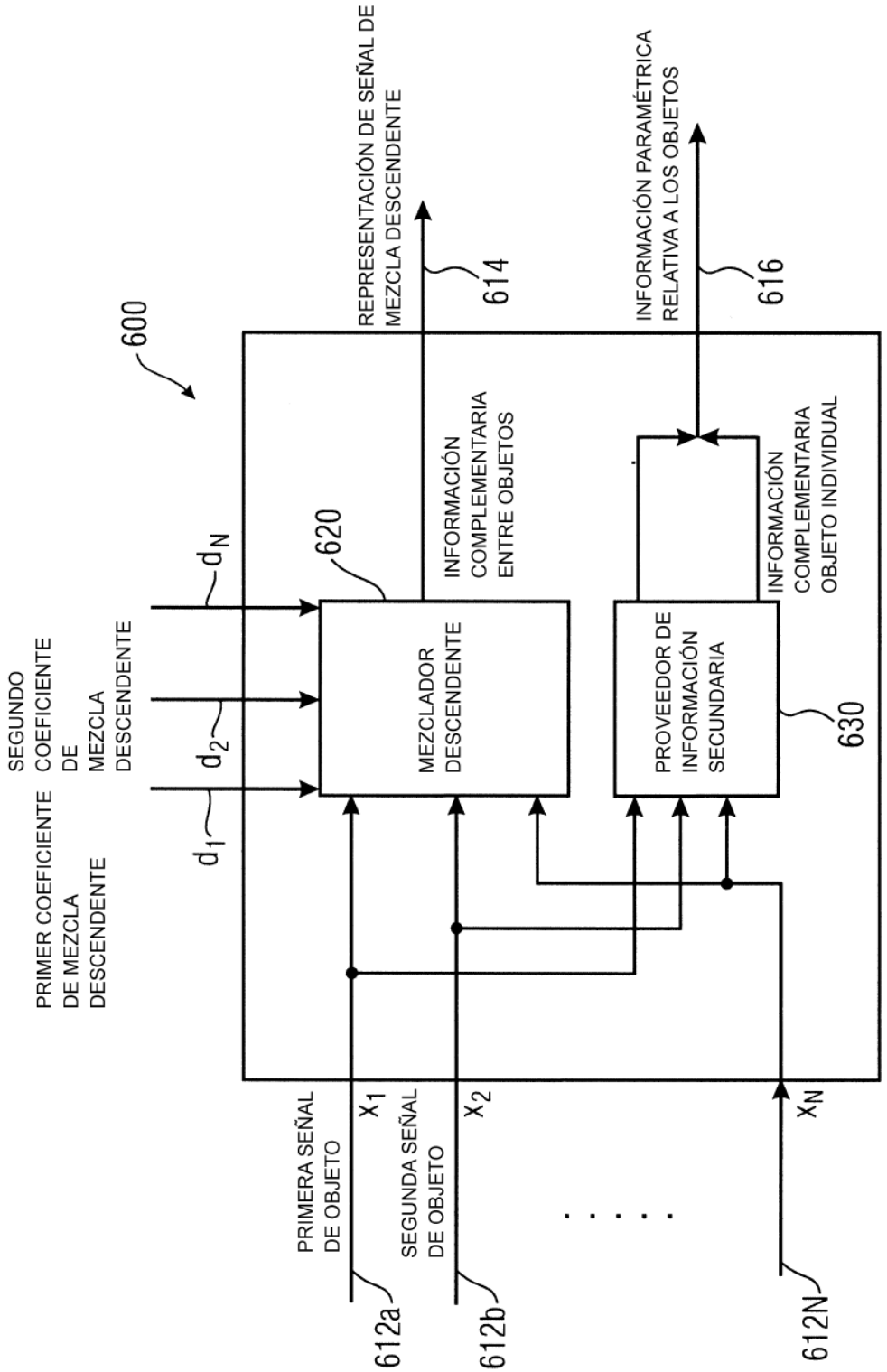

600
PRIMER COEFICIENTE DE MEZCLA DESCENDENTE
SEGUNDO COEFICIENTE DE MEZCLA DESCENDENTE
$d_1$
$d_2$
$d_N$
620
MEZCLADOR DESCENDENTE
REPRESENTACIÓN DE SEÑAL DE MEZCLA DESCENDENTE
614
INFORMACIÓN COMPLEMENTARIA ENTRE OBJETOS
PROVEEDOR DE INFORMACIÓN SECUNDARIA
630
INFORMACIÓN PARAMÉTRICA RELATIVA A LOS OBJETOS
616
INFORMACIÓN COMPLEMENTARIA OBJETO INDIVIDUAL
PRIMERA SEÑAL DE OBJETO
$x_1$
612a
SEGUNDA SEÑAL DE OBJETO
$x_2$
612b
$x_N$
612N

FIG. 6

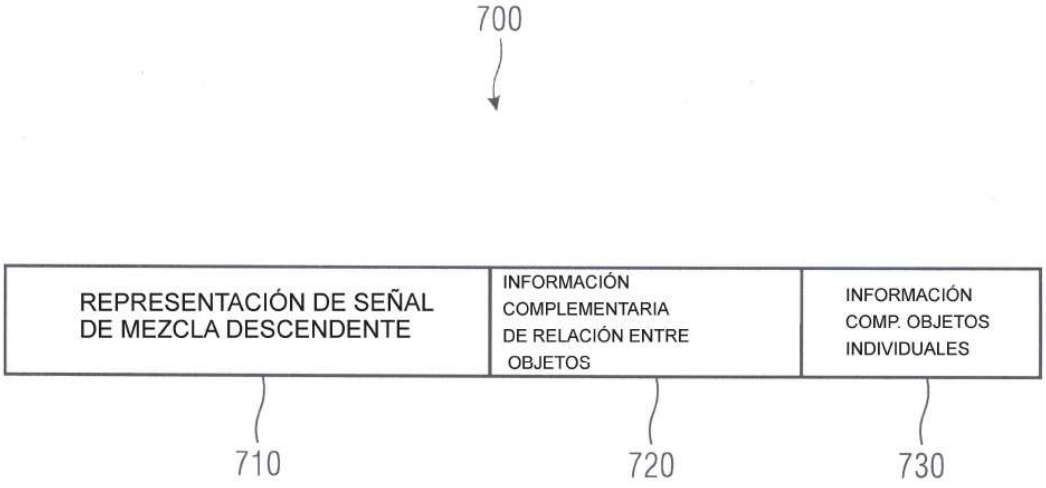
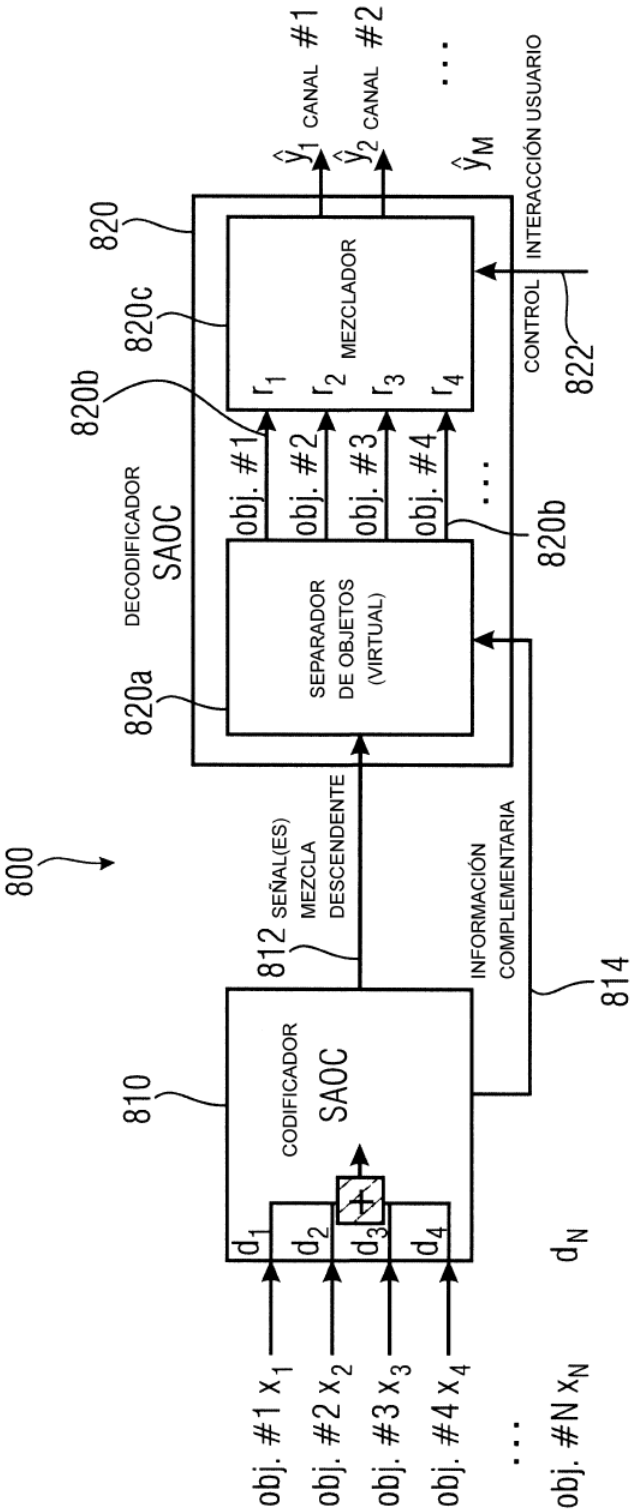
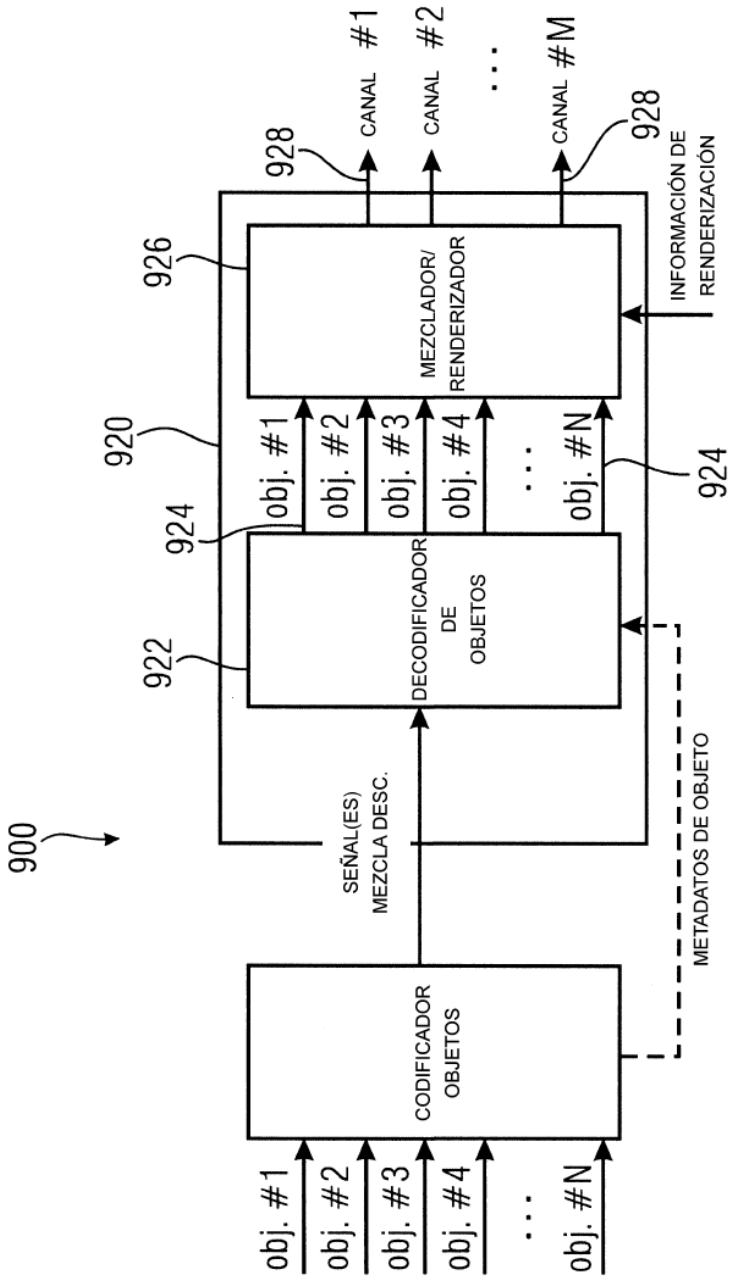

700
REPRESENTACIÓN DE SEÑAL DE MEZCLA DESCENDENTE
INFORMACIÓN COMPLEMENTARIA DE RELACIÓN ENTRE OBJETOS
INFORMACIÓN COMP. OBJETOS INDIVIDUALES
710
720
730

**FIG. 7**



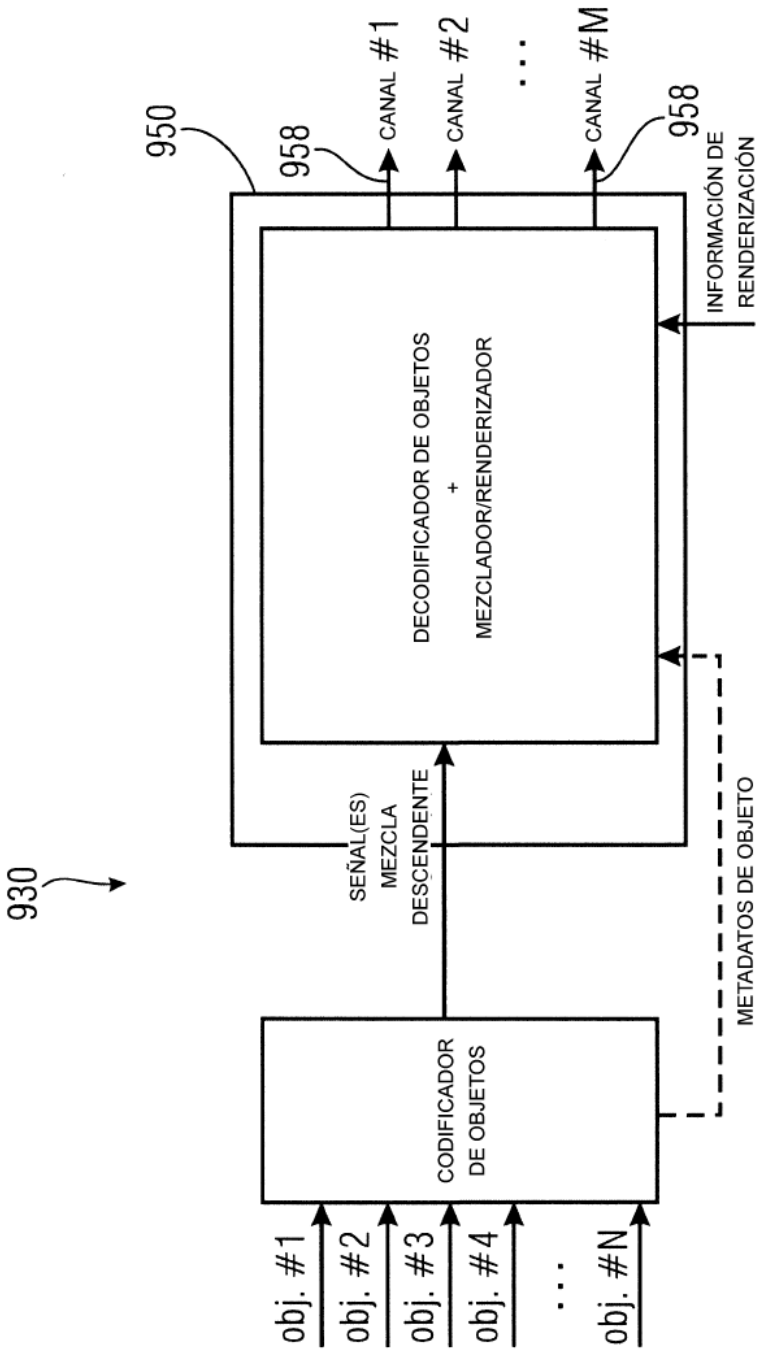

RESEÑA GENERAL SISTEMA MPEG SAOC

**FIG. 8**




RESEÑA GENERAL SISTEMA MPEG SAOC

**FIG. 9 A**




DECODIFICADOR Y MEZCLADOR INTEGRADOS

**FIG. 9 B**

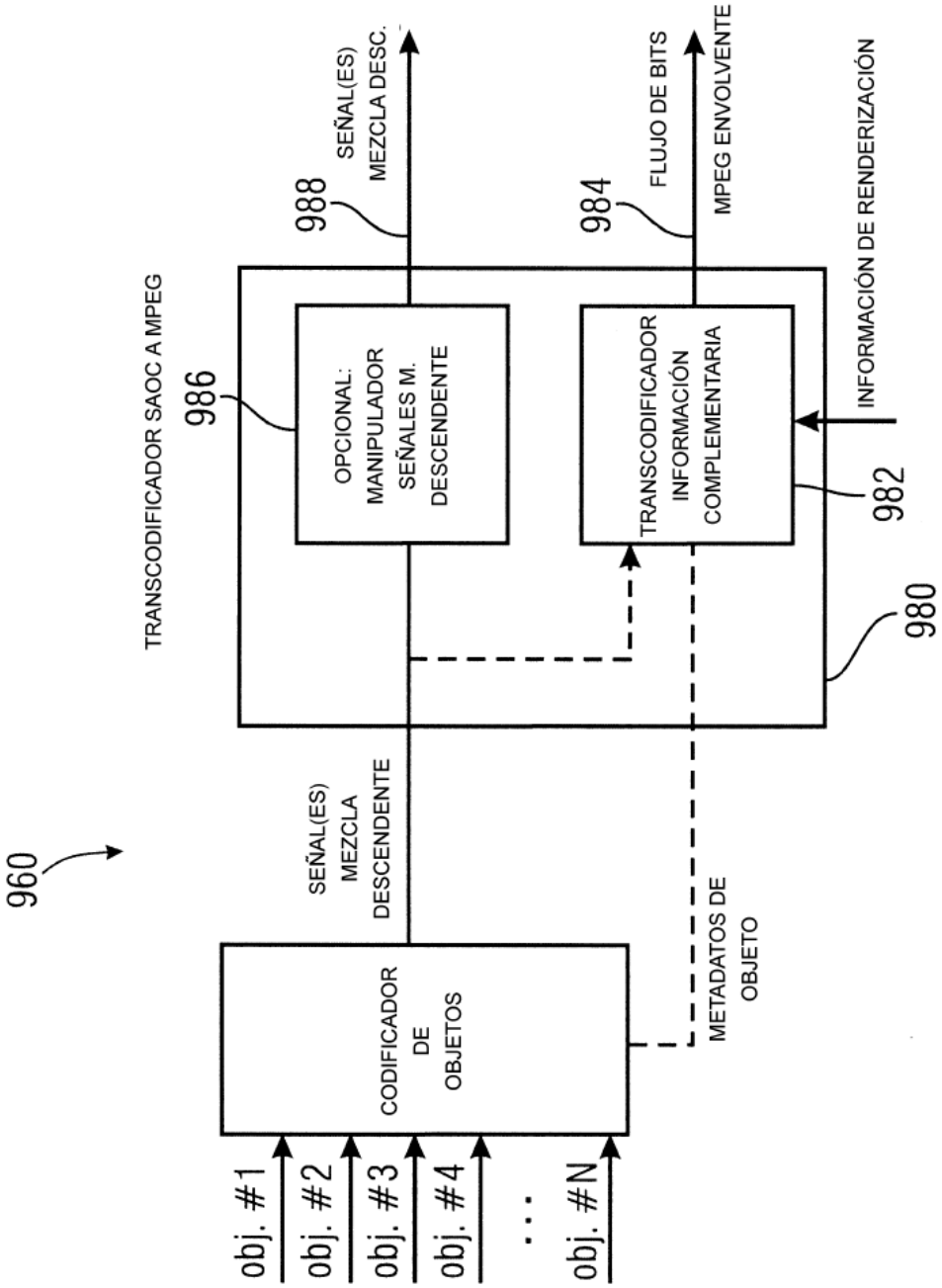

960
TRANSCODIFICADOR SAOC A MPEG
986
OPCIONAL:
MANIPULADOR
SEÑALES M.
DESCENDENTE
988
SEÑAL(ES)
MEZCLA DESC.
984
FLUJO DE BITS
MPEG ENVOLVENTE
TRANSCODIFICADOR
INFORMACIÓN
COMPLEMENTARIA
INFORMACIÓN DE RENDERIZACIÓN
982
980
SEÑAL(ES)
MEZCLA
DESCENDENTE
METADATOS DE
OBJETO
CODIFICADOR
DE
OBJETOS
obj. #1
obj. #2
obj. #3
obj. #4
...
obj. #N

FIG. 9 C