

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 768 431**

51 Int. Cl.:

G10L 19/008 (2013.01)

H04S 3/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.07.2014 PCT/EP2014/065290**

87 Fecha y número de publicación internacional: **29.01.2015 WO15010999**

96 Fecha de presentación y número de la solicitud europea: **16.07.2014 E 14742188 (7)**

97 Fecha y número de publicación de la concesión europea: **13.11.2019 EP 3025333**

54 Título: **Aparato y método para realizar una mezcla descendente SAOC de contenido de audio 3D**

30 Prioridad:

22.07.2013 EP 13177371

22.07.2013 EP 13177357

22.07.2013 EP 13177378

18.10.2013 EP 13189281

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

22.06.2020

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 768 431 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 y método para realizar una mezcla descendente SAOC de contenido de audio 3D

5 La presente invención se refiere a codificación/ decodificación de audio, en particular, a codificación de audio espacial y codificación de un objeto de audio espacial, y, más particularmente, a un aparato y método para realizar una mezcla descendente SAOC de contenido de audio 3D y a un aparato y método para decodificar en forma eficiente la mezcla descendente SAOC de contenido de audio 3D.

10 Las herramientas de codificación de audio espacial se conocen muy bien en la técnica y están, por ejemplo, estandarizadas en el estándar envolvente MPEG. La codificación de audio espacial comienza con canales de entrada originales tales como cinco o siete canales que se identifican por su colocación en una configuración de reproducción, es decir, un canal izquierdo, un canal central, un canal derecho, un canal envolvente izquierdo, un canal envolvente derecho y un canal de potencia de baja frecuencia. Un codificador de audio espacial normalmente
 15 deriva uno o más canales de mezcla descendente de los canales originales y, en forma adicional, deriva datos paramétricos relacionados con señales espaciales tales como diferencias de nivel entre canales, diferencias de fase entre canales, diferencias de tiempo entre canales, etc. El uno o más canales de mezcla descendente se transmiten junto con la información lateral paramétrica lo que indica las señales espaciales a un decodificador de audio espacial que decodifica el canal de mezcla descendente y los datos paramétricos asociados con el fin de obtener finalmente
 20 canales de salida que son una versión aproximada de los canales de entrada originales. La colocación de los canales en la configuración de salida es normalmente fija y es, por ejemplo, un formato 5.1, un formato 7.1, etc.

Dichos formatos de audio basados en canales se utilizan ampliamente para almacenar o transmitir contenido de audio multicanal donde cada canal se relaciona con un altavoz específico en una posición dada. Una reproducción
 25 fiel de esta clase de formatos requiere una configuración de altavoz donde los altavoces se colocan en las mismas posiciones que los altavoces que se usan durante la producción de señales de audio. Mientras el aumento de la cantidad de altavoces mejora la reproducción de escenas de audio 3D realmente multisensorial, se torna cada vez más difícil cumplir con este requerimiento – en especial en un entorno doméstico como una sala de estar.

30 La necesidad de tener una configuración específica de un altavoz puede superarse por medio de un método basado en el objeto donde las señales del altavoz se procesan específicamente para la configuración de reproducción.

Por ejemplo, las herramientas de codificación de un objeto de audio espacial se conocen muy bien en la técnica y se estandarizan en el estándar SAOC de MPEG (SAOC = Codificación de un objeto de audio espacial). En contraste
 35 una codificación de audio espacial que se inicia en canales originales, la codificación de un objeto de audio espacial comienza con objetos de audio que no se dedican en forma automática para una determinada configuración de la reproducción de procesamiento. En su lugar, la colocación de los objetos de audio en la escena de reproducción es flexible y puede ser determinada por el usuario al ingresar cierta información de procesamiento en una codificación de un objeto de decodificador de audio espacial. Alternativamente o en forma adicional, la información de
 40 procesamiento, es decir, la información en cuya posición en la configuración de reproducción un determinado objeto de audio deberá colocarse normalmente con el paso del tiempo puede transmitirse como información lateral adicional o metadatos. Con el objetivo de obtener una determinada compresión de datos, una cantidad de objetos de audio se codifican por medio de un codificador de SAOC que calcula, a partir de los objetos de entrada, uno o más canales de transporte por la realización de mezcla descendente de los objetos de acuerdo con cierta información del
 45 proceso de mezcla descendente. Además, el codificador de SAOC calcula información lateral paramétrica que representa señales entre objetos tales como diferencias en el nivel de objetos (OLD), valores de coherencia de objetos, etc. Los datos paramétricos entre objetos se calculan para mosaicos de frecuencia/tiempo paramétrico, es decir, para un determinado cuadro de la señal de audio que comprende, por ejemplo, 1024 o 2048 muestras, 28, 20, 14 o 10, etc., se consideran las bandas de procesamiento de modo tal que, al final, existen datos paramétricos para
 50 cada cuadro y cada banda de procesamiento. Como ejemplo, cuando una pieza de audio tiene 20 cuadros y cuando cada cuadro se subdivide en 28 bandas de procesamiento, entonces la cantidad de mosaicos de tiempo/frecuencia es 560.

En un enfoque basado en objetos, el campo de sonido se describe por medio de objetos de audio individuales. Esto
 55 requiere metadatos de objetos que describen entre otros la posición variante con el tiempo de cada fuente de sonido en espacio 3D.

Un primer concepto de codificación de metadatos en la técnica anterior es el formato de intercambio de descripción de sonido espacial (SpatDIF), un formato de descripción de la escena de audio que aún se encuentra en etapa de
 60 desarrollo [M1]. Está diseñado como un formato de intercambio para escenas de sonido basadas en objetos y no proporciona ningún método de compresión para trayectorias de objetos. SpatDIF usa el formato de Control de Sonido Abierto basado en texto (OSC) para estructurar los metadatos de objetos [M2]. Una representación simple basada en texto, sin embargo, no es una opción para la transmisión comprimida de trayectorias de objetos.

Otro concepto de metadatos en la técnica previo es el Formato de Descripción de la Escena de Audio (ASDF) [M3], una solución basada en texto que tiene la misma desventaja. Los datos están estructurados por una extensión del Lenguaje de Interacción Multimedia Sincronizada (SMIL) que es un subconjunto del Lenguaje de Marcación Extensible (XML) [M4], [M5].

5 Un concepto de metadatos adicional en la técnica anterior es el formato binario de audio para escenas (AudioBIFS), un formato binario que es parte de la especificación de MPEG-4 [M6], [M7]. Está estrechamente relacionado con el Lenguaje de Modelado de Realidad Virtual basado en XML (VRML) que se desarrolló para la descripción de escenas audiovisuales 3D y aplicaciones de realidad virtual interactiva [M8]. La especificación AudioBIFS compleja usa
10 gráficos de escena para especificar vías de movimientos de objetos. Una desventaja principal de AudioBIFS es que no está diseñado para la operación en tiempo real donde una demora limitada del sistema y acceso aleatorio al torrente de datos son un requerimiento. Además, la codificación de las posiciones del objeto no explota el funcionamiento de localización limitada de los oyentes humanos. Para una posición de oyente fijo dentro de la
15 escena audiovisual, los datos de objetos pueden cuantificarse con una cantidad mucho menor de bits [M9]. Por lo tanto, la codificación de los metadatos de objetos que se aplica en AudioBIFS no es eficiente con respecto a la compresión de datos.

El documento US 2010/174548 A1 da a conocer un aparato y un método para la codificación y decodificación de una
20 señal de audio de múltiples objetos. El aparato incluye un mezclador descendente para el mezclado descendente de las señales de audio en una señal de audio mezclada de manera descendente y extracción de información complementaria incluyendo información de cabecera e información de señales espaciales para cada una de las
25 señales de audio, un codificador para la codificación de la señal de audio mezclada de manera descendente, y un codificador de información complementaria para generar la información complementaria como una corriente de bits. La información de cabecera incluye información de identificación para cada una de las señales de audio e información de canal para las señales de audio.

El objetivo de la presente invención es proporcionar conceptos mejorados para realizar una mezcla descendente del
30 contenido de audio. El objetivo de la presente invención se resuelve por un aparato según la reivindicación 1, por un aparato según la reivindicación 9, por un sistema según la reivindicación 11, por un método según la reivindicación 12, por un método según la reivindicación 13 y por un programa informático según la reivindicación 14.

Todas las siguientes ocurrencias de la palabra "realización (es)", si se refiere a combinaciones de características
35 diferentes de las definidas por las reivindicaciones independientes, se refieren a ejemplos que se presentaron originalmente pero que no representan realizaciones de la invención actualmente reivindicada; estos ejemplos todavía se muestran solo para fines ilustrativos.

Según las realizaciones, se realiza el transporte eficiente y se proporcionan medios de cómo decodificar la mezcla
descendente para contenido de audio 3D.

40 Se proporciona un aparato para generar uno o más canales de salida de audio. El aparato comprende un procesador de parámetros para calcular la información de mezclado del canal de salida y un procesador de mezcla descendente para generar el uno o más canales de salida de audio. El procesador de mezcla descendente se configura para recibir una señal de transporte de audio que comprende uno o más canales de transporte de audio, en donde dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la
45 cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio. La señal de transporte de audio depende de una primera regla de mezclado y en una segunda regla de mezclado. La primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados. Más aún, la segunda regla de mezclado indica cómo
50 mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio. El procesador de parámetros se configura para recibir información sobre la segunda regla de mezclado, en donde la información sobre la segunda regla de mezclado indica cómo mezclar la pluralidad de señales previamente mezcladas de manera tal que se obtienen el uno o más canales de transporte de audio. Más aún, el procesador de parámetros se configura para calcular la información de mezclado del canal de salida dependiendo de una cantidad de objetos de audio lo que indica la cantidad de las dos o más señales de
55 objeto de audio, dependiendo de una cantidad de canales previamente mezclados lo que indica la cantidad de la pluralidad de canales previamente mezclados, y dependiendo de la información sobre la segunda regla de mezclado. El procesador de mezcla descendente se configura para generar el uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información de mezclado del canal de salida.

60 Además, se proporciona un aparato para generar una señal de transporte de audio que comprende uno o más canales de transporte de audio. El aparato comprende un mezclador de objetos para generar la señal de transporte de audio que comprende el uno o más canales de transporte de audio a partir de dos o más señales de objeto de audio, de manera tal que las dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o

más señales de objeto de audio, y una interfaz de salida para emitir la señal de transporte de audio. El mezclador de objetos se configura para generar el uno o más canales de transporte de audio de la señal de transporte de audio dependiendo de una primera regla de mezclado y dependiendo de una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio. La primera regla de mezclado depende de una cantidad de objetos de audio, lo que indica la cantidad de las dos o más señales de objeto de audio, y depende de una cantidad de canales previamente mezclados, lo que indica la cantidad de la pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado depende de la cantidad de canales previamente mezclados. La interfaz de salida se configura para emitir información sobre la segunda regla de mezclado.

Además, se proporciona un sistema. El sistema comprende un aparato para generar una señal de transporte de audio como se describió con anterioridad y un aparato para generar uno o más canales de salida de audio como se describió con anterioridad. El aparato para generar uno o más canales de salida de audio se configura para recibir la señal de transporte de audio e información sobre la segunda regla de mezclado desde el aparato para generar una señal de transporte de audio. Además, el aparato para generar uno o más canales de salida de audio está configurado para generar uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información sobre la segunda regla de mezcla.

Además, se proporciona un método para generar uno o más canales de salida de audio. El método comprende:

- Recibir una señal de transporte de audio que comprende uno o más canales de transporte de audio, en donde dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio, en donde la señal de transporte de audio depende de una primera regla de mezclado y en una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio.

- Recibir información sobre la segunda regla de mezclado, en donde la información sobre la segunda regla de mezclado indica cómo mezclar la pluralidad de señales previamente mezcladas de manera tal que se obtienen el uno o más canales de transporte de audio.

- Calcular la información de mezclado del canal de salida dependiendo de una cantidad de objetos de audio lo que indica la cantidad de las dos o más señales de objeto de audio, dependiendo de una cantidad de canales previamente mezclados lo que indica la cantidad de la pluralidad de canales previamente mezclados, y dependiendo de la información sobre la segunda regla de mezclado. Y:

- Generar uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información de mezclado del canal de salida.

Además, se proporciona un método para generar una señal de transporte de audio que comprende uno o más canales de transporte de audio. El método comprende:

- Generar la señal de transporte de audio que comprende el uno o más canales de transporte de audio a partir de dos o más señales de objeto de audio.

- Emitir la señal de transporte de audio. y:

- Emitir información sobre la segunda regla de mezclado.

Generar la señal de transporte de audio que comprende el uno o más canales de transporte de audio a partir de dos o más señales de objeto de audio se conduce de manera tal que las dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio. Generar el uno o más canales de transporte de audio de la señal de transporte de audio se conduce dependiendo de una primera regla de mezclado y dependiendo de una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio. La primera regla de mezclado depende de una cantidad de objetos de audio, lo que indica la cantidad de las dos o más señales de objeto de audio, y depende de una cantidad de canales previamente mezclados, lo que indica la cantidad de la pluralidad de canales

previamente mezclados. La segunda regla de mezclado depende de la cantidad de canales previamente mezclados.

Más aún, se proporciona un programa informático para implementar el método descrito con anterioridad cuando se ejecuta en un ordenador o procesador de señal.

5 A continuación, se describen las realizaciones de la presente invención en mayor detalle con referencia a las figuras, en las cuales:

10 La figura 1 ilustra un aparato para generar uno o más canales de salida de audio de acuerdo con una realización,

La figura 2 ilustra un aparato para generar una señal de transporte de audio que comprende uno o más canales de transporte de audio de acuerdo con una realización,

15 La figura 3 ilustra un sistema de acuerdo con una realización,

La figura 4 ilustra una primera realización de un codificador de audio 3D,

La figura 5 ilustra una primera realización de un decodificador de audio 3D,

20 La figura 6 ilustra una segunda realización de un codificador de audio 3D,

La figura 7 ilustra una segunda realización de un decodificador de audio 3D,

25 La figura 8 ilustra una tercera realización de un codificador de audio 3D,

La figura 9 ilustra una tercera realización de un decodificador de audio 3D,

30 La figura 10 ilustra la posición de un objeto de audio en un espacio tridimensional a partir de un origen expresado por azimut, elevación y radio, y

La figura 11 ilustra posiciones de objetos de audio y una configuración de un altavoz asumida por el generador del canal de audio.

35 Antes de describir las realizaciones preferidas de la presente invención en detalle, se describe el nuevo Sistema de Códec de Audio 3D.

En la técnica anterior, no existe tecnología flexible que combina codificación de canales por un lado y codificación de objetos por otro lado de modo tal que se obtienen calidades de audio aceptables a bajas transferencias de bits.

40 Esta limitación se supera por medio del nuevo Sistema de Códec de Audio 3D.

Antes de describir las realizaciones preferidas en detalle, se describe el nuevo Sistema de Códec de Audio 3D.

45 La figura 4 ilustra un codificador de audio 3D de acuerdo con una realización de la presente invención. El codificador de audio 3D se configura para codificar datos 101 de entrada de audio para obtener datos 501 de salida de audio. El codificador de audio 3D comprende una interfaz de salida para recibir una pluralidad de canales de audio indicados por CH y una pluralidad de objetos de audio indicados por OBJ. Además, como se ilustra en la figura 4, la interfaz 1100 de entrada en forma adicional recibe metadatos relacionados con uno o más de la pluralidad de objetos de audio OBJ. Además, el codificador de audio 3D comprende un mezclador 200 para mezclar la pluralidad de objetos y la pluralidad de canales para obtener una pluralidad de canales mezclados previamente, en donde cada canal mezclado previamente comprende datos de audio de un canal y datos de audio de por lo menos un objeto.

50 Además, el codificador de audio 3D comprende un codificador 300 central para codificar en forma central datos de entrada del codificador central, un compresor 400 de metadatos para comprimir los metadatos relacionados con el uno o más de la pluralidad de objetos de audio.

60 Además, el codificador de audio 3D puede comprender un controlador 600 de modos para controlar el mezclador, el codificador central y/o una interfaz 500 de salida en uno de muchos modos de operación, en donde en el primer modo, el codificador central se configura para codificar la pluralidad de audio canales y la pluralidad de objetos de audio recibidos por la interfaz 1100 de entrada sin ninguna interacción por el mezclador, es decir, sin ningún mezclado hecho por el mezclador 200. En un segundo modo, sin embargo, en el cual el mezclador 200 era activo, el codificador central codifica la pluralidad de canales mixtos, es decir, la salida generada por el bloque 200. En este último caso, se prefiere no codificar ya ningún dato de objetos. En su lugar, los metadatos que indican las posiciones de los objetos de audio ya se usan por el mezclador 200 para procesar los objetos sobre los canales como se indica

por los metadatos. En otras palabras, el mezclador 200 usa los metadatos relacionados con la pluralidad de objetos de audio para preprocesar los objetos de audio y luego los objetos de audio preprocesados se mezclan con los canales para obtener canales mixtos en la salida del mezclador. En esta realización, cualquier objeto puede no necesariamente transmitirse y esto también aplica para metadatos comprimidos como salida para el bloque 400. Sin embargo, si no todos los objetos entran en la interfaz 1100 se mezclan pero únicamente una cierta cantidad de objetos se mezcla, entonces no sólo los objetos no mezclados anteriormente y los metadatos asociados no obstante se transmiten al codificador 300 central o el compresor 400 de metadatos, respectivamente.

La figura 6 ilustra una realización adicional de un codificador de audio 3D que, en forma adicional, comprende un codificador 800 de SAOC. El codificador 800 de SAOC se configura para generar uno o más canales de transporte y datos paramétricos a partir de los datos de entrada del codificador de objetos de audio espacial. Como se ilustra en la figura 6, los datos de entrada del codificador de objeto de audio espacial son objetos que no han sido procesados por el preprocesador/ mezclador. Alternativamente, siempre que el preprocesador/ mezclador se ha traspasado como en el modo uno donde una codificación de objetos/ canal individual está activa, todas las entradas de objetos en la interfaz 1100 de entrada se codifican por medio del codificador 800 de SAOC.

Además, como se ilustra en la figura 6, el codificador 300 central se implementa con preferencia como un codificador USAC, es decir, como un codificador como se define y se estandariza en el estándar MPEG-USAC (USAC = Unified Speech and Audio Coding). La emisión del codificador de audio 3D completo ilustrado en la figura 6 es un caudal de datos MPEG 4, caudal de datos MPEG H o caudal de datos de audio 3D, que tienen estructuras similares a un contenedor para tipos de datos individuales. Además, los metadatos se indican como datos "OAM" y el compresor 400 de metadatos en la figura 4 corresponde al codificador 400 OAM para obtener datos OAM comprimidos que se ingresan en el codificador USAC 300 que, como puede observarse en la figura 6, en forma adicional comprende la interfaz de salida para obtener el caudal de datos de reproducción de MP4 que no solamente tiene datos de objeto/ canal codificados sino que también tiene los datos OAM comprimidos.

La figura 8 ilustra una realización adicional del codificador de audio 3D, donde en contraste con la figura 6, el codificador de SAOC puede configurarse en forma indistinta para codificar, con el algoritmo de codificación SAOC, los canales provistos en el preprocesador/mezclador 200 que no está activo en este modo o, alternativamente, para codificar por SAOC los canales previamente procesados más objetos. De este modo, en la figura 8, el codificador 800 de SAOC puede operar en tres clases diferentes de datos de entrada, es decir, canales sin ningún objeto previamente procesado, canales y objetos previamente procesados o únicamente objetos. Además, se prefiere proporcionar un decodificador 420 OAM adicional en la figura 8 de modo tal que el codificador 800 de SAOC use, para su procesamiento, los mismos datos que se encuentran del lado del decodificador, es decir, datos obtenidos por una compresión de pérdida en lugar de los datos OAM originales.

La figura 8 el decodificador de audio 3D puede operar en varios modos individuales.

Adicionalmente del primero y segundo modos como se discute en el contexto de la figura 4, el codificador de audio 3D de la figura 8 puede operar en forma adicional en un tercer modo en el cual el codificador central genera el uno o más canales de transporte a partir de los objetos individuales cuando el preprocesador/mezclador 200 no estaba activo. Alternativamente o en forma adicional, en este tercer modo el codificador 800 de SAOC puede generar uno o más canales de transporte alternativos o adicionales de los canales originales, es decir, nuevamente cuando el preprocesador/mezclador 200 correspondiente al mezclador 200 de la figura 4 no estaba activo.

Finalmente, el codificador 800 de SAOC puede codificar, cuando el codificador de audio 3D se configura en el cuarto modo, los canales más objetos previamente procesados según se generan por el preprocesador/ mezclador. De este modo, en el cuarto modo las aplicaciones con menor transferencia de bits proporcionarán buena calidad debido al hecho de que los canales y objetos se han transformado completamente en canales de transporte SAOC individuales y la información lateral asociada según se indica en las figuras. 3 y 5 como "SAOC-SI" y, en forma adicional, cualquier metadato no comprimido no tiene que transmitirse en este cuarto modo.

La figura 5 ilustra un decodificador de audio 3D de acuerdo con una realización de la presente invención. El codificador de audio 3D recibe, como entrada, los datos de audio codificados, es decir, los datos 501 de la figura 4.

El codificador de audio 3D comprende un descompresor 1400 de metadatos, un decodificador 1300 central, un procesador 1200 de objetos, un controlador 1600 de modos y un postprocesador 1700.

Específicamente, el codificador de audio 3D se configura para decodificar datos de audio codificados y la interfaz de entrada se configura para recibir los datos de audio codificados, los datos de audio codificados que comprende una pluralidad de canales codificados y la pluralidad de objetos codificados y metadatos comprimidos relacionados con la pluralidad de objetos en un determinado modo.

Además, el decodificador 1300 central se configura para decodificar la pluralidad de canales codificados y la

pluralidad de objetos codificados y, en forma adicional, el descompresor de metadatos se configura para descomprimir los metadatos comprimidos.

5 Además, el procesador 1200 de objetos se configura para procesar la pluralidad de objetos decodificados según se generan por el decodificador 1300 central con el uso de metadatos comprimidos para obtener una cantidad predeterminada de canales de salida que comprende datos de objeto y los canales decodificados. Estos canales de salida como se indica en 1205 luego se ingresan en un postprocesador 1700. El postprocesador 1700 se configura para convertir la cantidad de canales 1205 de salida en un determinado formato de entrada que puede ser un formato de reproducción binaural o un formato de reproducción por altavoz tales como un formato de reproducción 10 5.1, 7.1, etc.

Preferiblemente, el codificador de audio 3D comprende un controlador 1600 de modos el cual se configura para analizar los datos codificados para detectar una indicación de modo. Por lo tanto, el controlador 1600 de modo se conecta a la interfaz 1100 de entrada en la figura 5. Sin embargo, alternativamente, el controlador de modo no necesariamente tiene que estar allí. En su lugar, el decodificador de audio flexible puede preconfigurarse por cualquier otra clase de datos de control tales como una entrada del usuario o cualquier otro control. El codificador de audio 3D en la figura 5 y, preferiblemente controlado por el controlador 1600 de modo, se configura o bien para 15 traspasar el procesador de objetos y para alimentar la pluralidad de canales decodificados en el postprocesador 1700. Este es el funcionamiento en el modo 2, es decir, en el cual se reciben únicamente los canales previamente procesados, es decir, cuando se ha aplicado el modo 2 en el codificador de audio 3D de la figura 4. Alternativamente, cuando se ha aplicado el modo 1 en el codificador de audio 3D, es decir, cuando el codificador de audio 3D ha realizado codificación individual de canales/objetos, luego el procesador 1200 de objetos no se atraviesa, pero la pluralidad de canales decodificados y la pluralidad de objetos decodificados se colocan en el procesador 1200 de objetos junto con metadatos descomprimidos generados por el descompresor 1400 de 20 metadatos.

Preferiblemente, la indicación de si debe aplicarse el modo 1 o el modo 2 se incluye los datos de audio codificados y luego el controlador 1600 de modo analiza los datos codificados para detectar una indicación de modo. Se usa el modo 1 cuando la indicación del modo indica que los datos de audio codificados comprenden canales codificados y objetos codificados y el modo 2 se aplica cuando la indicación del modo indica que los datos de audio codificados no 30 contienen ningún objeto de audio, es decir, únicamente contienen canales preprocesados obtenidos por el modo 2 del codificador de audio 3D de la figura 4.

La figura 7 ilustra una realización preferida comparada con la del codificador de audio 3D de la figura 5 y la realización de la figura 7 corresponde al codificador de audio 3D de la figura 6. Adicionalmente de la implementación del codificador de audio 3D de la figura 5, el codificador de audio 3D en la figura 7 comprende un decodificador 1800 de SAOC. Además, el procesador 1200 de objetos de la figura 5 se implementa como un procesador separado de objetos 1210 y el mezclador 1220 mientras que, dependiendo del modo, la funcionalidad del procesador 1210 de 35 objetos también pueda ser implementada por el decodificador 1800 de SAOC.

Además, el postprocesador 1700 puede implementarse como un procesador 1710 binaural o un conversor 1720 de formato. Alternativamente, una emisión 1205 directa de datos de la figura 5 también puede implementarse como se ilustra en 1730. Por lo tanto, se prefiere realizar el procesamiento en el decodificador sobre la cantidad más elevada de canales tales como 22.2 o 32 con el objetivo de tener flexibilidad y luego postprocesar si se requiere un formato 45 más pequeño. Sin embargo, cuando queda claro desde el principio que solo se requiere un formato diferente con un número menor de canales como un formato 5.1 se requiere, entonces se prefiere, como se indica por la figura 9 por el método 1727 simplificado, que un determinado control sobre el decodificador de SAOC y/o el decodificador USAC puede aplicarse con el fin de evitar operaciones innecesarias de mezcla ascendente y las posteriores operaciones de mezcla descendente.

50 En una realización preferida de la presente invención, el procesador 1200 de objetos comprende el decodificador 1800 de SAOC y el decodificador de SAOC se configura para decodificar uno o más canales de transporte emitidos por el decodificador central y datos paramétricos asociados y con el uso de metadatos descomprimidos para obtener la pluralidad de objetos de audio procesados. Hasta este punto, la salida de OAM se conecta a la caja 1800.

55 Además, el procesador 1200 de objetos se configura para procesar objetos decodificados emitidos por el decodificador central que no se codifican en los canales de transporte de SAOC pero que se codifican individualmente en normalmente elementos en canales individuales como se indica por el procesador 1210 de objetos. Además, el decodificador comprende una interfaz de salida que corresponde a la salida 1730 para emitir 60 una salida del mezclador a los altavoces.

En una realización adicional, el procesador 1200 de objetos comprende a codificación 1800 de un objeto de decodificador de audio espacial para decodificar uno o más canales de transporte e información lateral paramétrica asociada que representa señales de audio codificadas o canales de audio codificados, en donde la codificación de

5 un objeto de decodificador de audio espacial se configura para transcodificar la información paramétrica asociada y los metadatos descomprimidos en información lateral paramétrica transcodificada susceptible de usarse para procesar directamente el formato de salida, como se define por ejemplo en una versión anterior de SAOC. El postprocesador 1700 se configura para calcular audio canales del formato de salida con el uso de los canales de transporte codificados y la información lateral paramétrica transcodificada. El procesamiento realizado por el post procesador puede ser similar al procesamiento Envoltente MPEG o puede ser cualquier otro procesamiento tales como procesamiento BCC y demás.

10 En una realización adicional, el procesador 1200 de objetos comprende una codificación 1800 de un objeto de decodificador de audio espacial configurado para mezclar/unificar (mezcla directa) y procesar señales de canales para el formato de salida con el uso de los canales de transporte decodificados (por el decodificador central) y la información lateral paramétrica

15 Además, y muy importante, el procesador 1200 de objetos de la figura 5 en forma adicional comprende el mezclador 1220 que recibe, como entrada, datos generados por el decodificador USAC 1300 directamente cuando existen objetos previamente procesados mezclados con canales, es decir, cuando el mezclador 200 de la figura 4 estaba activo. Adicionalmente El bloque VBAP con preferencia recibe los datos OAM decodificados para derivar las matrices de renderización (procesamiento)., el mezclador 1220 recibe datos del procesador de objetos que realiza procesamiento de objetos sin decodificación SAOC. Además, el mezclador recibe datos de salida del decodificador SAOC, es decir, objetos renderizados por SAOC.

20 El mezclador 1220 se conecta a la interfaz 1730 de salida, el procesador 1710 binaural y el conversor 1720 de formato. El procesador 1710 binaural se configura para renderizar los canales de salida en dos canales binaurales con el uso de funciones de transferencia relacionadas con el cabezal o respuestas a los impulsos de sala binaural (BRIR). El conversor 1720 de formato se configura para convertir los canales de salida en un formato de salida que tiene una cantidad menor de canales que los canales 1205 de salida del mezclador y el conversor 1720 de formato requiere información sobre la disposición de salida tales como altavoces 5.1 y demás.

25 El decodificador de audio 3D de la figura 9 es diferente del decodificador de audio 3D de la figura 7 en el hecho de que decodificador de SAOC no sólo puede generar objetos renderizados sino también canales renderizados y este es el caso en el que el codificador de audio 3D de la figura 8 se ha usado y la conexión 900 entre los canales/objetos previamente procesados y la interfaz de entrada del codificador 800 de SAOC está activa.

30 Además, una etapa 1810 de panoramización de amplitud de base del vector (VPAP) se configura que recibe, del decodificador de SAOC, la información sobre la disposición de salida y que emite una matriz de renderización (procesamiento) al decodificador de SAOC de modo tal que el decodificador de SAOC puede, al final, proporcionar canales renderizados sin ninguna operación adicional del mezclador en el formato del canal alto de 1205, es decir, 32 altavoces.

35 El bloque VBAP con preferencia recibe los datos OAM decodificados para derivar las matrices de renderización. Más en general, preferiblemente requiere información geométrica no sólo de la disposición de salida sino también de las posiciones donde las señales de entrada deben renderizarse en la disposición de salida. Estos datos de entrada geométricos pueden ser datos OAM para información de la posición del canal u objetos para canales que se han transmitido con el uso de SAOC.

40 Sin embargo, si sólo una interfaz de salida específica se requiere luego el estado VBAP 1810 ya puede proporcionar la matriz de renderización requerida para la salida, por ejemplo 5.1. El decodificador 1800 de SAOC entonces realiza una renderización directa de los canales de transporte de SAOC, los datos paramétricos asociados y metadatos descomprimidos, una renderización directa en el formato de salida requerido sin ninguna interacción del mezclador 1220. Sin embargo, cuando se aplica una determinada mezcla entre modos, es decir, donde varios canales se codifican con SAOC pero no todos los canales se codifican con SAOC o donde varios objetos se codifican con SAOC pero no todos los objetos se codifican con SAOC o cuando sólo una cierta cantidad de objetos previamente procesados con canales están decodificados por SAOC y los canales restantes no se procesan con SAOC luego el mezclador unificará los datos de las porciones de entrada individuales, es decir, directamente del decodificador 1300 central, del procesador 1210 de objetos y del decodificador 1800 de SAOC.

45 En el audio 3D, un ángulo de azimut, un ángulo de elevación y un radio se usa para definir la posición de un objeto de audio. Más aún, puede transmitirse una ganancia para un objeto de audio.

50 El ángulo de azimut, ángulo de elevación y radio define en forma no ambigua la posición de un objeto de audio en un espacio 3D de un origen. Esto se ilustra con referencia a la figura 10.

55 La figura 10 ilustra la posición 410 de un objeto de audio en un espacio tridimensional (3D) de un origen 400 expresado por azimut, elevación y radio.

5 El ángulo de azimut específica, por ejemplo, un ángulo en el plano xy (el plano definido por el eje x y el eje y). El ángulo de elevación define, por ejemplo, un ángulo en el plano xz (el plano definido por el eje x y el eje z). Al especificar el ángulo de azimut y el ángulo de elevación, la línea 415 recta a través del origen 400 y la posición 410 del objeto de audio puede definirse. Mediante la especificación en forma adicional del radio, la posición exacta 410 del objeto de audio puede definirse.

10 En una realización, el ángulo de azimut se define para el rango: $-180^\circ < \text{azimut} \leq 180^\circ$, el ángulo de elevación se define para el rango: $-90^\circ < \text{elevación} \leq 90^\circ$ y el radio puede, por ejemplo, definirse en metros [m] (mayor que o igual a 0 m). La esfera descrita por el azimut, elevación y ángulo puede dividirse en dos hemisferios: hemisferio izquierdo ($0^\circ < \text{azimut} \leq 180^\circ$) y hemisferio derecho ($-180^\circ < \text{azimut} \leq 0^\circ$), o hemisferio superior ($0^\circ < \text{elevación} \leq 90^\circ$) y hemisferio inferior ($-90^\circ < \text{elevación} \leq 0^\circ$)

15 En otra realización, donde puede suponerse, por ejemplo, que todos los valores x de las posiciones del objeto de audio en el sistema de coordenadas xyz son mayores que o iguales a cero, el ángulo de azimut puede definirse para el rango: $-90^\circ \leq \text{azimut} \leq 90^\circ$, el ángulo de elevación puede definirse para el rango: $-90^\circ < \text{elevación} \leq 90^\circ$, y el radio puede, por ejemplo, definirse en metros [m].

20 El procesador 120 de mezcla descendente puede, por ejemplo, configurarse para generar uno o más canales de audio dependiendo de uno o más señales de objeto de audio dependiendo de los valores de información de metadatos reconstruidos, en donde los valores de información de metadatos reconstruidos pueden, por ejemplo, indicar la posición de los objetos de audio.

25 En una realización los valores de información de metadatos pueden, por ejemplo, indicar, el ángulo de azimut definido para el rango: $-180^\circ < \text{azimut} \leq 180^\circ$, el ángulo de elevación definido para el rango: $-90^\circ < \text{elevación} \leq 90^\circ$ y el radio puede, por ejemplo, definirse en metros [m] (mayor que o igual a 0m).

30 La figura 11 ilustra posiciones de objetos de audio y una configuración de un altavoz asumida por el generador del canal de audio. Se ilustra el origen 500 del sistema de coordenadas xyz. Más aún, se ilustra la posición 510 de un primer objeto de audio y la posición 520 de un segundo objeto de audio. Además, la figura 11 ilustra un escenario, donde el generador 120 de canales de audio genera cuatro canales de audio para cuatro altavoces. El generador 120 de canales de audio supone que los cuatro altavoces 511, 512, 513 y 514 se ubican en las posiciones que se muestran en la figura 11.

35 En la figura 11, el primer objeto de audio se ubica en una posición 510 cercanas a las supuestas posiciones de altavoces 511 y 512, y se ubica alejado de los altavoces 513 y 514. Por lo tanto, el generador 120 de canales de audio puede generar los cuatro canales de audio de manera tal que el primer objeto de audio 510 se reproduce a través de los altavoces 511 y 512 pero no por los altavoces 513 y 514.

40 En otras realizaciones, el generador 120 de canales de audio puede generar los cuatro canales de audio de manera tal que el primer objeto de audio 510 se reproduce con un alto nivel por los altavoces 511 y 512 y con un bajo nivel por los altavoces 513 y 514.

45 Más aún, el segundo objeto de audio se ubica en una posición 520 cercana a las supuestas posiciones de altavoces 513 y 514, y se ubica alejado de los altavoces 511 y 512. Por lo tanto, el generador 120 de canales de audio puede generar los cuatro canales de audio de manera tal que el segundo objeto de audio 520 se reproduce a través de los altavoces 513 y 514 pero no por los altavoces 511 y 512.

50 En otras realizaciones, el procesador 120 de mezcla descendente puede generar los cuatro canales de audio de manera tal que el segundo objeto de audio 520 se reproduce con un alto nivel por los altavoces 513 y 514 y con un bajo nivel por los altavoces 511 y 512.

55 En realizaciones alternativas, únicamente dos valores de información de metadatos se usan para especificar la posición de un objeto de audio. Por ejemplo, únicamente el azimut y el radio puede especificarse, por ejemplo, cuando se supone que todos los objetos de audio se ubican dentro de un plano único.

60 En otras realizaciones adicionales, para cada objeto de audio, sólo un valor de información de metadatos individual de una señal de metadatos se codifica y se transmite como información de la posición. Por ejemplo, únicamente un ángulo de azimut puede especificarse como información de la posición para un objeto de audio (por ejemplo puede suponerse que todos los objetos de audio se ubican en el mismo plano que tiene la misma distancia desde un punto central, y así se supone que tienen el mismo radio). La información del azimut puede, por ejemplo, ser suficiente para determinar que un objeto de audio se ubica cerca de un altavoz izquierdo y lejos de un altavoz derecho. En dicha situación, el generador 120 de canales de audio puede, por ejemplo, generar el uno o más audio canales de manera tal que el objeto de audio se reproduce por el altavoz izquierdo, pero no por el altavoz derecho.

Por ejemplo, puede emplearse la panoramización de amplitud de base de vector para determinar el peso de un objeto de señal de audio dentro de cada uno de los canales de salida de audio (véase, por ejemplo [VBAP]). Con respecto a VBAP, se supone que un objeto de señal de audio se asigna a una fuente virtual, y se supone en forma adicional que un canal de salida de audio es un canal de un altavoz.

En las realizaciones, un valor de información de metadatos adicionales por ejemplo de una señal adicional de metadatos puede especificar un volumen, por ejemplo un aumento (por ejemplo, expresado en decibeles [dB]) para cada objeto de audio.

Por ejemplo, en la figura 11, un primer valor de aumento puede especificarse por un valor de información de metadatos adicionales para el primer objeto de audio ubicado en la posición 510 que es mayor que un segundo valor de aumento especificada por otra información de metadatos adicionales para el segundo objeto de audio ubicado en la posición 520. En dicha situación, los altavoces 511 y 512 pueden reproducir el primer objeto de audio con un nivel que es superior al nivel con el cual los altavoces 513 y 514 reproducen el segundo objeto de audio.

De acuerdo con la técnica de SAOC, un codificador de SAOC recibe una pluralidad de señales de objeto de audio X y produce un mezcla descendente de ellas mediante el empleo de una matriz de mezcla descendente D para obtener una señal de transporte de audio. Y que comprende uno o más canales de transporte de audio. La fórmula

$$Y = DX$$

puede emplearse. El codificador de SAOC transmite la señal de transporte de audio Y e información sobre la matriz de mezcla descendente D (por ejemplo coeficientes de la matriz de mezcla descendente D) al decodificador de SAOC. Más aún, el codificador de SAOC transmite información sobre una matriz de covarianza E (por ejemplo coeficientes de la matriz de covarianza E) al decodificador de SAOC.

Del lado del decodificador, las señales de objeto de audio X podrían reconstruirse para obtener objetos de audio reconstruidos \hat{X} mediante el empleo de la fórmula

$$\hat{X} = GY$$

en donde G es una matriz de estimación de fuentes paramétricas con $G = E D H (D E D H)^{-1}$.

Entonces, uno o más canales de salida de audio Z podría generarse por aplicación de una matriz de renderización R sobre los objetos de audio reconstruidos \hat{X} de acuerdo con la fórmula:

$$Z = R \hat{X} .$$

Generar el uno o más canales de salida de audio Z a partir de la señal de transporte de audio puede también conducirse, sin embargo en un paso simple mediante el empleo de matriz U de acuerdo con la fórmula:

$$Z = UY , \quad \text{con } U = RG .$$

Cada hilera de la matriz de renderización R se asocia con uno de los canales de salida de audio que se generará. Cada coeficiente dentro de una de las hileras de la matriz de renderización R determina el peso de una de las señales reconstruidas de objeto de audio dentro del canal de salida de audio, con el cual dicha hilera de la matriz de renderización R se relaciona.

Por ejemplo, la matriz de renderización R puede depender de la información sobre la posición para cada una de las señales de objeto de audio transmitidas al decodificador de SAOC dentro de la información de metadatos. Por ejemplo, un objeto de señal de audio que tiene una posición que se ubica cerca de una posición de altavoz supuesta o real puede, por ejemplo tener un peso más alto dentro del canal de salida de audio de dicho altavoz que el peso de un objeto de señal de audio, la posición de que se ubica lejos de dicho altavoz (véase la figura 5). Por ejemplo, puede emplearse la panoramización de amplitud de base de vector para determinar el peso de un objeto de señal de audio dentro de cada uno de los canales de salida de audio (véase, por ejemplo [VBAP]). Con respecto a VBAP, se supone que un objeto de señal de audio se asigna a una fuente virtual, y se supone en forma adicional que un canal de salida de audio es un canal de un altavoz.

En las figuras 6 y 8, se presenta un codificador de SAOC. El codificador 800 de SAOC se usa para codificar paramétricamente una cantidad de objetos/canales de entrada por la realización de mezcla descendente a ellos a

una cantidad menor de canales de transporte y extracción de la información auxiliar necesaria que se sumerge en el caudal de transferencia de bits de Audio 3D.

5 La realización de mezcla descendente a una cantidad menor de canales de transporte se realiza con el uso de coeficientes de mezcla descendente para cada señal de entrada y canal de mezcla descendente (por ejemplo mediante el empleo de una matriz de mezcla descendente).

10 El estado de la técnica anterior en el procesamiento de señales de objeto de audio es el sistema SAOC MPEG. Una propiedad principal de dicho sistema es que las señales de mezcla descendente intermedias (o los canales de transporte de SAOC de acuerdo con la figura 6 y 8) pueden escucharse con dispositivos de legado incapaces de decodificar la información SAOC. Esto impone restricciones sobre los coeficientes de mezcla descendente que se usarán, los que usualmente son provistos por el creador de contenidos.

15 El Sistema de Códec de Audio 3D tiene la finalidad de usar tecnología SAOC para aumentar la eficiencia para codificar una cantidad grande de objetos o canales. La realización de mezcla descendente de una cantidad grande de objetos a una cantidad pequeña de canales de transporte ahorra la transferencia de bits.

20 La figura 2 ilustra un aparato para generar una señal de transporte de audio que comprende uno o más canales de transporte de audio de acuerdo con una realización.

25 El aparato comprende un mezclador 210 de objetos para generar la señal de transporte de audio que comprende el uno o más canales de transporte de audio a partir de dos o más señales de objeto de audio, de manera tal que las dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio.

Más aún, el aparato comprende una interfaz 220 de salida para emitir la señal de transporte de audio.

30 El mezclador 210 de objetos se configura para generar el uno o más canales de transporte de audio de la señal de transporte de audio dependiendo de una primera regla de mezclado y dependiendo de una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio. La primera regla de mezclado depende de una cantidad de objetos de audio, lo que indica la cantidad de las dos o más señales de objeto de audio, y depende de una cantidad de canales previamente mezclados, lo que indica la cantidad de la pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado depende de la cantidad de canales previamente mezclados. La interfaz 220 de salida se configura para emitir información sobre la segunda regla de mezclado.

40 La figura 1 ilustra un aparato para generar uno o más canales de salida de audio de acuerdo con una realización.

El aparato comprende un procesador 110 de parámetros para calcular la información de mezclado del canal de salida y un procesador 120 de mezcla descendente para generar el uno o más canales de salida de audio.

45 El procesador 120 de mezcla descendente se configura para recibir una señal de transporte de audio que comprende uno o más canales de transporte de audio, en donde dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio. La señal de transporte de audio depende de una primera regla de mezclado y en una segunda regla de mezclado. La primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados. Más aún, la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio.

50 El procesador 110 de parámetros se configura para recibir información sobre la segunda regla de mezclado, en donde la información sobre la segunda regla de mezclado indica cómo mezclar la pluralidad de señales previamente mezcladas de manera tal que se obtienen el uno o más canales de transporte de audio. El procesador 110 de parámetros se configura para calcular la información de mezclado del canal de salida dependiendo de una cantidad de objetos de audio lo que indica la cantidad de las dos o más señales de objeto de audio, dependiendo de una cantidad de canales previamente mezclados lo que indica la cantidad de la pluralidad de canales previamente mezclados, y dependiendo de la información sobre la segunda regla de mezclado.

60 El procesador 120 de mezcla descendente se configura para generar el uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información de mezclado del canal de salida.

De acuerdo con una realización, el aparato puede, por ejemplo, configurarse para recibir por lo menos una cantidad de objetos de audio y la cantidad de canales previamente mezclados.

5 En otra realización, el procesador 110 de parámetros puede, por ejemplo, configurarse para determinar, dependiendo de la cantidad de los objetos de audio y dependiendo de la cantidad de canales previamente mezclados, la información sobre la primera regla de mezclado, de manera tal que la información sobre la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener la pluralidad de canales previamente mezclados. En dicha realización, el procesador 110 de parámetros puede, por ejemplo, configurarse para calcular la información de mezclado del canal de salida, dependiendo de la información sobre la primera regla de mezclado y dependiendo de la información sobre la segunda regla de mezclado.

15 De acuerdo con una realización, el procesador 110 de parámetros puede, por ejemplo, configurarse para determinar, dependiendo de la cantidad de los objetos de audio y dependiendo de la cantidad de canales previamente mezclados, una pluralidad de coeficientes de una primera matriz P como la información sobre la primera regla de mezclado, en donde la primera matriz P indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio. En dicha realización, el procesador 110 de parámetros, puede, por ejemplo, configurarse para recibir una pluralidad de coeficientes de una segunda matriz P como la información sobre la segunda regla de mezclado, en donde la segunda matriz Q indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio. El procesador 110 de parámetros de dicha realización puede, por ejemplo configurarse para calcular la información de mezclado del canal de salida dependiendo de la primera matriz P y dependiendo de la segunda matriz Q.

25 Las realizaciones se basan en el hallazgo de que cuando se realiza la mezcla descendente de las dos o más señales de objeto de audio X para obtener una señal de transporte de audio Y del lado del codificador mediante el empleo de matriz de mezcla descendente D de acuerdo con la fórmula

$$Y = DX,$$

30 luego la matriz de mezcla descendente D puede dividirse en dos matrices más pequeñas P y Q de acuerdo con la fórmula

$$D = QP.$$

35 Aquí, la primera matriz P realiza el mezclado a partir de las señales de objeto de audio X a la pluralidad de canales previamente mezclados X_{pre} de acuerdo con la fórmula:

$$X_{pre} = PX.$$

40 La segunda matriz Q realiza el mezclado a partir de la pluralidad de canales de premezcla X_{pre} al uno o más canales de transporte de audio de la señal de transporte de audio Y de acuerdo con la fórmula:

$$Y = Q X_{pre}.$$

45 De acuerdo con las realizaciones, la información sobre la segunda regla de mezclado, por ejemplo sobre los coeficientes de la segunda matriz de mezclado Q, se transmite al decodificador.

50 Los coeficientes de la primera matriz de mezclado P no tienen que transmitirse al decodificador. En su lugar, el decodificador recibe información sobre la cantidad de señales de objeto de audio e información sobre la cantidad de canales previamente mezclados. A partir de esta información, el decodificador es capaz de reconstruir la primera matriz de mezclado P. Por ejemplo, el codificador y decodificador determinan la matriz de mezclado P del mismo modo, cuando se mezcla una primera cantidad de $N_{objetos}$ señales de objeto de audio a una segunda cantidad de N_{pre} canales previamente mezclados.

55 La figura 3 ilustra un sistema de acuerdo con una realización. El sistema comprende un aparato 310 para generar una señal de transporte de audio como se describió con anterioridad con referencia a la figura 2 y un aparato 320 para generar uno o más canales de salida de audio como se describió con anterioridad con referencia a la figura 1.

60 El aparato 320 para generar uno o más canales de salida de audio se configura para recibir la señal de transporte de audio e información sobre la segunda regla de mezclado desde el aparato 310 para generar una señal de transporte de audio. Más aún, el aparato 320 para generar uno o más canales de salida de audio se configura para generar el uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información sobre la segunda regla de mezclado.

Por ejemplo, el procesador 110 de parámetros puede, por ejemplo, configurarse para recibir la información de metadatos que comprende la información sobre la posición para cada una de las dos o más señales de objeto de audio, y determina la información sobre la primera regla de mezcla descendente dependiendo de la información de la posición de cada una de las dos o más señales de objeto de audio, por ejemplo mediante el empleo de panoramización de amplitud de base vertical. Por ejemplo, el codificador también puede tener acceso a la información de la posición de cada una de las dos o más señales de objeto de audio y también puede emplear panoramización de amplitud de base de vector para determinar los pesos de las señales de objeto de audio en los canales previamente mezclados, y por esto determina los coeficientes de la primera matriz P del mismo modo como lo realiza más tarde el decodificador (por ejemplo tanto en codificador como el decodificador pueden suponer el mismo posicionamiento de los altavoces supuestos asignados a los N_{pre} canales previamente mezclados).

Al recibir los coeficientes de la segunda matriz Q y al determinar la primera matriz P, el decodificador puede determinar la matriz de mezcla descendente D de acuerdo con $D = QP$.

En una realización, el procesador 110 de parámetros puede, por ejemplo, configurarse para recibir información de covarianza, por ejemplo coeficientes de una matriz de covarianza E (por ejemplo desde el aparato para generar la señal de transporte de audio), lo que indica una diferencia en el nivel de objetos para cada una de las dos o más señales de objeto de audio, y, posiblemente, lo que indica uno o más correlaciones entre objetos entre una de las señales de objeto de audio y otra de las señales de objeto de audio.

En dicha realización, el procesador 110 de parámetros puede configurarse para calcular la información de mezclado del canal de salida dependiendo de la cantidad de los objetos de audio, dependiendo de la cantidad de canales previamente mezclados, dependiendo de la información sobre la segunda regla de mezclado, y dependiendo de la información de covarianza.

Por ejemplo, con el uso de la matriz de covarianza E, las señales de objeto de audio X podrían reconstruirse para obtener objetos de audio reconstruidos \hat{X} mediante el empleo de la fórmula

$$\hat{X} = GY$$

en donde G es una matriz de estimación de fuentes paramétricas con $G = E D^H (D E D^H)^{-1}$.

Entonces, uno o más canales de salida de audio Z podría generarse por aplicación de una matriz de renderización R sobre los objetos de audio reconstruidos \hat{X} de acuerdo con la fórmula:

$$Z = R \hat{X} .$$

Generar el uno o más canales de salida de audio Z a partir de la señal de transporte de audio puede, sin embargo, también conducirse en un paso simple mediante el empleo de matriz U de acuerdo con la fórmula:

$$Z = UY , \quad \text{con } S = UG .$$

Dicha matriz S es un ejemplo para una información de mezclado del canal de salida determinada por el procesador 110 de parámetros.

Por ejemplo, como ya se explicó con anterioridad, cada hilera de la matriz de renderización R puede asociarse con uno de los canales de salida de audio que se generará. Cada coeficiente dentro de una de las hileras de la matriz de renderización R determina el peso de una de las señales reconstruidas de objeto de audio dentro del canal de salida de audio, con el cual dicha hilera de la matriz de renderización R se relaciona.

De acuerdo con una realización, en donde el procesador 110 de parámetros puede, por ejemplo, configurarse para recibir la información de metadatos que comprende la información sobre la posición para cada una de las dos o más señales de objeto de audio, puede, por ejemplo configurarse para determinar la información de procesamiento, por ejemplo los coeficientes de la matriz de renderización R dependiendo de la información de la posición de cada una de las dos o más señales de objeto de audio, y puede, por ejemplo, configurarse para calcular la información de mezclado del canal de salida (por ejemplo la matriz anterior S) dependiendo de la cantidad de los objetos de audio, dependiendo de la cantidad de canales previamente mezclados, dependiendo de la información sobre la segunda regla de mezclado, y dependiendo de la información de procesamiento (por ejemplo matriz de renderización R).

De este modo, la matriz de renderización R puede, por ejemplo, depender de la información sobre la posición para cada una de las señales de objeto de audio transmitidas al decodificador de SAOC dentro de la información de

metadatos. Por ejemplo, un objeto de señal de audio que tiene una posición que se ubica cerca de una posición de altavoz supuesta o real puede, por ejemplo tener un peso más alto dentro del canal de salida de audio de dicho altavoz que el peso de un objeto de señal de audio, la posición de que se ubica lejos de dicho altavoz (véase la figura 5). Por ejemplo, puede emplearse la panoramización de amplitud de base de vector para determinar el peso de un objeto de señal de audio dentro de cada uno de los canales de salida de audio (véase, por ejemplo [VBAP]). Con respecto a VBAP, se supone que un objeto de señal de audio se asigna a una fuente virtual, y se supone en forma adicional que un canal de salida de audio es un canal de un altavoz. El correspondiente coeficiente de la matriz de renderización R (el coeficiente que se asigna al considerado canal de salida de audio y la considerada señal del objeto de audio) entonces puede configurarse al valor dependiendo de dicho peso. Por ejemplo, el peso mismo puede ser el valor de dicho correspondiente coeficiente dentro de la matriz de renderización R.

A continuación, las realizaciones que realizan mezcla descendente espacial para señales basadas en objetos se explican en detalle.

Se hace referencia a las siguientes notas y definiciones:

- $N_{Objetos}$ cantidad de señales de entrada de objeto de audio
- $N_{Ca}N_{ales}$ cantidad de canales de entrada
- N cantidad de señales de entrada;
- N puede ser igual con $N_{Objetos}$, $N_{Ca}N_{ales}$ o $N_{Objetos} + N_{Ca}N_{ales}$.
- N_{DmxCh} cantidad de canales con mezcla descendente (procesados)
- N_{pre} cantidad de canales de premezcla
- $N_{Muestras}$ cantidad de muestras de datos procesados
- D matriz de mezcla descendente, tamaño $N_{DmxCh} \times N$
- X señal de entrada de audio que comprende las dos o más señales de entrada de audio, tamaño $N \times N_{Muestras}$
- Y señal de audio de mezcla descendente (la señal de transporte de audio), tamaño $N_{DmxCh} \times N_{Muestras}$, definido como $Y = DX$
- DMG datos de aumento de mezcla descendente para cada señal de entrada, canal de mezcla descendente, y parámetro fijado

D_{DMG} es la matriz tridimensional que retiene los datos de DMG mapeados, no cuantificados para cada señal de entrada, canal de mezcla descendente, y parámetro fijado

Sin pérdida de generalidad, con el fin de mejorar la legibilidad de ecuaciones, para todas las variables introducidas los índices que denotan dependencia en tiempo y frecuencia se omiten.

Si no se especifica limitación respecto de las señales de entrada (canales u objetos), los coeficientes de mezcla descendente se computan del mismo modo para señales de canal de entrada y señales de objeto de entrada. Se usa la indicación para la cantidad de señales de entrada N.

Algunas realizaciones pueden, por ejemplo diseñarse para realizar un mezcla descendente de las señales de objeto en un modo diferente de las señales del canal, guiada por la información espacial disponible en los metadatos de objetos.

El mezcla descendente puede separarse en dos pasos:

- En un primer paso, los objetos se renderizan previamente a la disposición de salida con la cantidad más alta de altavoces N_{pre} (por ejemplo $N_{pre} = 22$ dado por la configuración 22.2). Por ejemplo puede emplearse la primera matriz P.

- En una segunda etapa, las señales N_{pre} previamente renderizadas obtenidas se procesan por mezcla descendente de la cantidad de canales disponibles de transporte (N_{DmxCh}) (por ejemplo de acuerdo con un

algoritmo de distribución de mezcla descendente ortogonal). Por ejemplo puede emplearse la segunda matriz Q.

Sin embargo, en algunas realizaciones, la mezcla descendente se realiza en un paso simple, por ejemplo mediante el empleo de matriz D que se define de acuerdo con la fórmula: $D = QP$, y al aplicar $Y = DX$ con $D = QP$.

5 Entre otras, una ventaja adicional de los conceptos propuestos es, por ejemplo que las señales de objeto de entrada que se supone que se renderizarán en la misma posición espacial, en la escena de audio, se procesan por mezcla descendente juntas en los mismos canales de transporte. Como consecuencia en el lado del decodificador una mejor separación de las señales previamente renderizadas se obtiene, evitando la separación de objetos de audio que se mezclarán nuevamente juntos en la escena de reproducción final.

De acuerdo con particular las realizaciones preferidas, la mezcla descendente puede describirse como una multiplicación de matriz por:

15
$$X_{pre} = PX \quad \text{y} \quad Y = QX_{pre} .$$

donde P de tamaño $(N_{pre} \times N_{Objetos})$ y Q de tamaño $(N_{DmxCh} \times N_{pre})$ se computan como se explica a continuación.

20 Los coeficientes de mezclado en P se construyen a partir de los metadatos de las señales de objeto (ángulos de radio, aumento, azimut y ángulo de elevación) con el uso de un algoritmo de panoramización (por ej. panoramización de amplitud de base de vector). El algoritmo de panoramización debe ser el mismo que el utilizado en el lado del decodificador para la construcción de los canales de salida.

25 Los coeficientes de mezclado en Q se dan en el lado del codificador para N_{pre} señales de entrada y N_{DmxCh} canales disponibles de transporte.

Con el fin de reducir la complejidad computacional, la mezcla descendente de dos etapas puede simplificarse a una al computar los aumentos de la mezcla descendente final como:

30
$$D = QP .$$

Luego las señales de la mezcla descendente se dan por:

35
$$Y = DX .$$

Los coeficientes de mezclado en P no se transmiten dentro del flujo de bits. En su lugar, se reconstruyen en el lado del decodificador con el uso del mismo algoritmo de panoramización. Por lo tanto la tasa de transferencia se reduce al enviar únicamente los coeficientes de mezclado en Q. En particular, ya que los coeficientes de mezclado en P usualmente son una variante de tiempo, y como P no se transmite, puede lograrse una reducción elevada en la transferencia de bits.

A continuación, se considera la sintaxis del flujo de bits de acuerdo con una realización.

45 Para señalar el método de mezcla descendente usado y la cantidad de canales N_{pre} para prerenderizar los objetos en la primera etapa, la síntesis del flujo de bits SAOC de MPEG se extiende con 4 bits:

Método bsSaocDmx	Modo	Significado
0	Modo directo	La matriz de mezcla descendente se construye directamente a partir de los DMG descuantificados (aumentos de mezcla descendente).
1, ..., 15	Modo premezcla	La matriz de mezcla descendente se construye como un producto de la matriz obtenida de los DMG descuantificados y una matriz de premezcla obtenida de la información espacial de los objetos de audio de entrada.

bsNumPremixedChannels

bsSaocDmxMethod	bsNumPremixedChannels
0	0

1	22
2	11
3	10
4	8
5	7
6	5
7	2
8,..., 14	reservado
15	valor de escape

En el contexto de MPEG SAOC, esto puede lograrse por la siguiente codificación:

bsSaocDmxMethod : Indica cómo se construye la matriz de mezcla descendente

5 Sintaxis de SAOC3DSpecificConfig() - Señalización

```

bsSaocDmxMethod;                                4    uimsbf
si (bsSaocDmxMethod == 15) {
    bsNumPremixedChannels;                        5    uimsbf
}
    
```

10 Sintaxis de Saoc3DFrame(): el modo que los DMGs se leen para diferentes modos

```

si (bsNumSaocDmxObjects==0) {
    para (i=0; i< bsNumSaocDmxChannels; i++) {
        idxDMG[i] = EcDataSaoc(DMG, 0, NumInputSignals);
    }
} o bien {
    dmglDx = 0;
    para (i=0; i<bsNumSaocDmxChannels; i++) {
        idxDMG[i] = EcDataSaoc(DMG, 0, bsNumSaocChannels);
    }
    dmglDx = bsNumSaocDmxChannels;
    si (bsSaocDmxMethod == 0) {
        para (i=dmglDx; i<dmglDx + bsNumSaocDmxObjects; i++) {
            idxDMG[i] = EcDataSaoc(DMG, 0, bsNumSaocObjects);
        }
    } o bien {
        para (i= dmglDx; i<dmglDx + bsNumSaocDmxObjects; i++) {
            idxDMG[i] = EcDataSaoc(DMG, 0, bsNumPremixedChannels);
        }
    }
}
    
```

15 **bsNumSaocDmxChannels** Define la cantidad de canales para mezcla descendente para contenido basado en los canales. Si no se encuentran canales presentes en la mezcla descendente, bsNumSaocDmxChannels se fija en cero.

bsNumSaocChannels Define la cantidad de canales de entrada para los cuales los parámetros 3D de SAOC se transmiten. Si bsNumSaocChannels = 0 no hay canales presentes en la mezcla descendente.

20 **BsNumSaocDmxObjects** Define la cantidad de canales para mezcla descendente para contenido basado en objetos. Si no hay objetos presentes en la mezcla descendente bsNumSaocDmxObjects se fija en cero.

25 **bsNumPremixedChannels** Define la cantidad de canales de premezcla para los objetos de entrada de audio. Si bsSaocDmxMethod es igual a 15 luego la cantidad real de canales previamente mezclados se señala directamente por el valor de bsNumPremixedChannels. En todos los otros casos bsNumPremixedChannels se fija de

acuerdo con la tabla previa.

De acuerdo con una realización, la matriz de mezcla descendente \mathbf{D} aplicada a las señales de entrada de audio \mathbf{S} determina la señal de mezcla descendente como

$$\mathbf{X} = \mathbf{D}\mathbf{S}$$

La matriz de mezcla descendente \mathbf{D} de tamaño $N_{\text{dmx}} \times N$ se obtiene como:

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

La matriz \mathbf{D}_{dmx} y matriz $\mathbf{D}_{\text{premix}}$ tienen tamaños diferentes dependiendo del modo de procesamiento.

La matriz \mathbf{D}_{dmx} se obtiene de los parámetros de DMG como:

$$d_{i,j} = \begin{cases} 0 \\ 10^{0.05 DMG_{i,j}} \end{cases},$$

si no hay datos de DMG para el par (i,j) presentes en el flujo de bits, de otro modo.

Aquí, los parámetros de mezcla descendente descuantificados se obtienen como:

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

En el caso del modo directo, no se usa mezclado previo. La matriz $\mathbf{D}_{\text{premix}}$ tiene un tamaño $N \times N$ y se da por: $\mathbf{D}_{\text{premix}} = \mathbf{I}$. La matriz \mathbf{D}_{dmx} tiene un tamaño $N_{\text{dmx}} \times N$ y se obtiene de los parámetros de DMG.

En el caso del modo premezcla la matriz $\mathbf{D}_{\text{premix}}$ tiene un tamaño $(N_{\text{ch}} + N_{\text{premix}}) \times N$ y se da por:

$$\mathbf{D}_{\text{premix}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix},$$

donde la matriz de premezcla \mathbf{A} de tamaño $N_{\text{premix}} \times N_{\text{obj}}$ se recibe como entrada al decodificador de 3D SAOC, del procesador de objetos.

La matriz \mathbf{D}_{dmx} tiene un tamaño $N_{\text{dmx}} \times (N_{\text{ch}} + N_{\text{premix}})$ y se obtiene de los parámetros de DMG.

Aunque algunos aspectos se han descrito en el contexto de un aparato, resulta claro que estos aspectos representan, además, una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa del método o una característica de una etapa del método. De forma análoga, los aspectos descritos en el contexto de una etapa del método representan, además, una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.

La señal descompuesta de la invención puede almacenarse en un medio de almacenamiento digital o puede transmitirse en un medio de transmisión tales como un medio de transmisión inalámbrico o un medio de transmisión alámbrico tales como la Internet.

Dependiendo de ciertos requerimientos de implementación, las realizaciones la invención pueden implementarse en hardware o en software. La implementación puede realizarse con el uso de un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, un ROM, un PROM, una memoria EPROM, EEPROM o FLASH, que tiene

almacenadas en su interior señales de control legibles en forma electrónica, que cooperan (o son capaces de cooperar) con un sistema de ordenadores programables de manera tal que se realiza el método respectivo.

5 Algunas realizaciones de acuerdo con la invención comprenden un portador de datos no transitorio que tiene señales de control legibles por medios electrónicos, que son capaces de cooperar con un sistema de ordenadores programables, de manera tal que se realiza uno de los métodos descritos en este documento.

10 En general, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, el código de programa es operativo para realizar uno de los métodos cuando el producto del programa informático se ejecuta en un ordenador. El código de programa puede almacenarse por ejemplo en un portador legible en la máquina.

15 Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en este documento, almacenados en un portador legible en la máquina.

En otras palabras, una realización del método de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los métodos descritos en este documento, cuando el programa informático se ejecuta en un ordenador.

20 Una realización adicional de los métodos de la invención es, por lo tanto, un portador de datos (o un medio de almacenamiento digital, o un medio que se lee por ordenador) que comprende, grabado allí, el programa informático para realizar uno de los métodos descritos en este documento.

25 Una realización adicional del método de la invención es, por lo tanto, un caudal de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en este documento. El caudal de datos o la secuencia de señales puede por ejemplo configurarse para transferirse a través de una conexión de comunicación de datos, por ejemplo a través de la Internet.

30 Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado para o adaptado para realizar uno de los métodos descritos en este documento.

Una realización adicional comprende un ordenador que tiene instalado el programa informático para realizar uno de los métodos descritos en este documento.

35 En algunas realizaciones, un dispositivo lógico programable (por ejemplo un ordenamiento de acceso programable en campo) puede usarse para realizar todas o algunas de las funcionalidades de los métodos descritos en este documento. En algunas realizaciones, un ordenamiento de acceso programable en campo puede cooperar con un microprocesador con el fin de realizar uno de los métodos descritos en este documento. En general, los métodos con preferencia se realizan por cualquier aparato de hardware.

40 Las realizaciones descritas con anterioridad son simplemente ilustrativas para los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en este documento serán obvias para otras personas con experiencia en la técnica. Es la intención, por lo tanto, limitarse únicamente limitarse por el alcance de las reivindicaciones de la patente pendiente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de este documento.

Referencias

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

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

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

60 [VBAP] Ville Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning"; J. Audio Eng. Soc., Level 45, Issue 6, pp. 456-466, June 1997.

[M1] Peters, N., Lossius, T. and Schacher J. C., "SpatDIF: Principles, Specification, and Examples", 9th Sound and Music Computing Conference, Copenhagen, Denmark, Jul. 2012.

- [M2] Wright, M., Freed, A., "Open Sound Control: A New Protocol for Communicating with Sound Synthesizers", International Computer Music Conference, Thessaloniki, Greece, 1997.
- 5 [M3] Matthias Geier, Jens Ahrens, and Sascha Spors. (2010), "Object-based audio reproduction and the audio scene description format", *Org. Sound*, Vol. 15, No. 3, pp. 219-227, December 2010.
- [M4] W3C, "Synchronized Multimedia Integration Language (SMIL 3.0)", Dec. 2008.
- 10 [M5] W3C, "Extensible Markup Language (XML) 1.0 (Fifth Edition)", Nov. 2008.
- [M6] MPEG, "ISO/IEC International Standard 14496-3 - Coding of audio-visual objects, Part 3 Audio", 2009.
- 15 [M7] Schmidt, J.; Schroeder, E. F. (2004), "New and Advanced Features for Audio Presentation in the MPEG-4 Standard", 116th AES Convention, Berlin, Germany, May 2004.
- [M8] Web3D, "International Standard ISO/IEC 14772-1:1997 - The Virtual Reality Modeling Language (VRML), Part 1: Functional specification and UTF-8 encoding", 1997.
- 20 [M9] Sporer, T. (2012), "Codierung räumlicher Audiosignale mit leichtgewichtigen Audio-Objekten", Proc. Annual Meeting of the German Audiological Society (DGA), Erlangen, Germany, Mar. 2012.

REIVINDICACIONES

1. Aparato para generar uno o más canales de salida de audio, en donde el aparato comprende:

5 un procesador (110) de parámetros para calcular la información de mezclado del canal de salida, y

10 un procesador (120) de mezcla descendente para generar el uno o más canales de salida de audio, en donde el procesador (120) de mezcla descendente se configura para recibir una señal de transporte de audio que comprende uno o más canales de transporte de audio, en donde dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio,

15 en donde la señal de transporte de audio depende de una primera regla de mezclado y en una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio,

20 en donde el procesador (110) de parámetros se configura para recibir información sobre la segunda regla de mezclado, en donde la información sobre la segunda regla de mezclado indica cómo mezclar la pluralidad de señales previamente mezcladas de manera tal que se obtienen el uno o más canales de transporte de audio,

25 en donde el procesador (110) de parámetros se configura para calcular la información de mezclado del canal de salida dependiendo de una cantidad de objetos de audio lo que indica la cantidad de las dos o más señales de objeto de audio, dependiendo de una cantidad de canales previamente mezclados lo que indica la cantidad de la pluralidad de canales previamente mezclados, y dependiendo de la información sobre la segunda regla de mezclado, y

30 en donde el procesador (120) de mezcla descendente se configura para generar el uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información de mezclado del canal de salida.

35 2. Aparato según la reivindicación 1, en donde el aparato se configura para recibir por lo menos una cantidad de objetos de audio y la cantidad de canales previamente mezclados.

3. Aparato según la reivindicación 1 o 2,

40 en donde el procesador (110) de parámetros se configura para determinar, dependiendo de la cantidad de los objetos de audio y dependiendo de la cantidad de canales previamente mezclados, la información sobre la primera regla de mezclado, de manera tal que la información sobre la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener la pluralidad de canales previamente mezclados, y

45 en donde el procesador (110) de parámetros se configura para calcular la información de mezclado del canal de salida, dependiendo de la información sobre la primera regla de mezclado y dependiendo de la información sobre la segunda regla de mezclado.

50 4. Aparato según la reivindicación 3,

55 en donde el procesador (110) de parámetros se configura para determinar, dependiendo de la cantidad de los objetos de audio y dependiendo de la cantidad de canales previamente mezclados, una pluralidad de coeficientes de una primera matriz (P) como la información sobre la primera regla de mezclado, en donde la primera matriz (P) indica cómo mezclar las dos o más señales de objeto de audio para obtener la pluralidad de canales previamente mezclados,

60 en donde el procesador (110) de parámetros se configura para recibir una pluralidad de coeficientes de una segunda matriz (Q) como la información sobre la segunda regla de mezclado, en donde la segunda matriz (Q) indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio, y

en donde el procesador (110) de parámetros se configura para calcular la información de mezclado del canal de salida dependiendo de la primera matriz (P) y dependiendo de la segunda matriz (Q).

5. Aparato según una de las reivindicaciones anteriores,
- 5 en donde el procesador (110) de parámetros se configura para recibir la información de metadatos que comprende la información sobre la posición para cada una de las dos o más señales de objeto de audio,
- en donde el procesador (110) de parámetros se configura para determinar la información sobre la primera regla de mezcla dependiendo de la información de la posición de cada una de las dos o más señales de objeto de audio.
- 10 6. Aparato según la reivindicación 5,
- en donde el procesador (110) de parámetros se configura para determinar la información de procesamiento dependiendo de la información de la posición de cada una de las dos o más señales de objeto de audio, y
- 15 en donde el procesador (110) de parámetros se configura para calcular la información de mezclado del canal de salida dependiendo de la cantidad de los objetos de audio, dependiendo de la cantidad de canales previamente mezclados, dependiendo de la información sobre la segunda regla de mezclado, y dependiendo de la información de procesamiento.
- 20 7. Aparato según las reivindicaciones anteriores,
- en donde el procesador (110) de parámetros se configura para recibir información de covarianza lo que indica una diferencia en el nivel de objetos para cada una de las dos o más señales de objeto de audio, y
- 25 en donde el procesador (110) de parámetros se configura para calcular la información de mezclado del canal de salida dependiendo de la cantidad de los objetos de audio, dependiendo de la cantidad de canales previamente mezclados, dependiendo de la información sobre la segunda regla de mezclado, y dependiendo de la información de covarianza.
- 30 8. Aparato según la reivindicación 7,
- en donde la información de covarianza indica además por lo menos una correlación entre objetos entre una de las dos o más señales de objeto de audio y otra de las dos o más señales de objeto de audio, y
- 35 en donde el procesador (110) de parámetros se configura para calcular la información de mezclado del canal de salida dependiendo de la cantidad de los objetos de audio, dependiendo de la cantidad de canales previamente mezclados, dependiendo de la información sobre la segunda regla de mezclado, dependiendo de la diferencia del nivel de objeto de cada una de las dos o más señales de objeto de audio y dependiendo de la por lo menos una correlación entre objetos entre una de las dos o más señales de objeto de audio y otra de las dos o más señales de objeto de audio.
- 40
9. Aparato para generar una señal de transporte de audio que comprende uno o más canales de transporte de audio, en donde el aparato comprende:
- 45 un mezclador (210) de objetos para generar la señal de transporte de audio que comprende el uno o más canales de transporte de audio a partir de dos o más señales de objeto de audio, de manera tal que las dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio, y
- 50 una interfaz (220) de salida para emitir la señal de transporte de audio, en donde el aparato se configura para transmitir la señal de transporte de audio a un decodificador.
- 55 en donde el mezclador (210) de objetos se configura para generar el uno o más canales de transporte de audio de la señal de transporte de audio dependiendo de una primera regla de mezclado y dependiendo de una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio,
- 60 en donde la primera regla de mezclado depende de una cantidad de objetos de audio, lo que indica la cantidad de las dos o más señales de objeto de audio, y depende de una cantidad de canales previamente mezclados, lo que indica la cantidad de la pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado depende de la cantidad de canales previamente mezclados, y

- 5 en donde el mezclador (210) de objetos se configura para generar el uno o más canales de transporte de audio de la señal de transporte de audio dependiendo de una primera matriz (P), en donde la primera matriz (P) indica cómo mezclar las dos o más señales de objeto de audio para obtener la pluralidad de canales previamente mezclados, y dependiendo de una segunda matriz (Q), en donde la segunda matriz (Q) indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio,
- 10 en donde los coeficientes de la primera matriz (P) indica información sobre la primera regla de mezclado, y en donde los coeficientes de la segunda matriz (Q) indica información sobre la segunda regla de mezclado,
- 15 en donde el aparato se configura para transmitir los coeficientes de la segunda matriz de mezclado (Q) al decodificador, y en donde el aparato se configura para no transmitir los coeficientes de la primera matriz de mezclado (P) al decodificador.
10. Aparato según la reivindicación 9,
- 20 en donde el mezclador (210) de objetos se configura para recibir la información sobre la posición para cada una de las dos o más señales de objeto de audio, y
- en donde el mezclador (210) de objetos se configura para determinar la primera regla de mezclado dependiendo de la información de la posición de cada una de las dos o más señales de objeto de audio.
11. Sistema, que comprende:
- 25 un aparato (310) según la reivindicación 9 o 10 para generar una señal de transporte de audio, y
- 30 un aparato (320) según una de las reivindicaciones 1 a 8 para generar uno o más canales de salida de audio,
- en donde el aparato (320) según una de las reivindicaciones 1 a 8 se configura para recibir la señal de transporte de audio e información sobre la segunda regla de mezclado desde el aparato (310) según la reivindicación 9 o 10, y
- 35 en donde el aparato (320) según una de las reivindicaciones 1 a 8 se configura para generar el uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información sobre la segunda regla de mezclado.
12. Método para generar uno o más canales de salida de audio, en donde el método comprende:
- 40 recibir una señal de transporte de audio que comprende uno o más canales de transporte de audio, en donde dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, y en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio, en donde la señal de transporte de audio depende de una primera regla de mezclado y en una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio,
- 50 recibir información sobre la segunda regla de mezclado, en donde la información sobre la segunda regla de mezclado indica cómo mezclar la pluralidad de señales previamente mezcladas de manera tal que se obtienen el uno o más canales de transporte de audio,
- 55 calcular la información de mezclado del canal de salida dependiendo de una cantidad de objetos de audio lo que indica la cantidad de las dos o más señales de objeto de audio, dependiendo de una cantidad de canales previamente mezclados lo que indica la cantidad de la pluralidad de canales previamente mezclados, y dependiendo de la información sobre la segunda regla de mezclado, y
- 60 generar uno o más canales de salida de audio a partir de la señal de transporte de audio dependiendo de la información de mezclado del canal de salida.
13. Método para generar una señal de transporte de audio que comprende uno o más canales de transporte de audio, en donde el método comprende:

generar la señal de transporte de audio que comprende el uno o más canales de transporte de audio a partir de dos o más señales de objeto de audio,

5 emitir la señal de transporte de audio, y transmitir la señal de transporte de audio a un decodificador, y

transmitir los coeficientes de una segunda matriz de mezclado (Q) al decodificador, y no transmitir los coeficientes de una primera matriz de mezclado (P) al decodificador,

10 en donde generar la señal de transporte de audio que comprende el uno o más canales de transporte de audio a partir de dos o más señales de objeto de audio se conduce de manera tal que las dos o más señales de objeto de audio se mezclan dentro de la señal de transporte de audio, en donde la cantidad del uno o más canales de transporte de audio es menor que la cantidad de las dos o más señales de objeto de audio, y

15 en donde generar el uno o más canales de transporte de audio de la señal de transporte de audio se conduce dependiendo de una primera regla de mezclado y dependiendo de una segunda regla de mezclado, en donde la primera regla de mezclado indica cómo mezclar las dos o más señales de objeto de audio para obtener una pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado indica cómo mezclar la pluralidad de canales previamente mezclados para obtener el uno o más canales de transporte de audio de la señal de transporte de audio, en donde la primera regla de mezclado depende de una cantidad de objetos de audio, lo que indica la cantidad de las dos o más señales de objeto de audio, y depende de una cantidad de canales previamente mezclados, lo que indica la cantidad de la pluralidad de canales previamente mezclados, y en donde la segunda regla de mezclado depende de la cantidad de canales previamente mezclados,

20 en donde generar el uno o más canales de transporte de audio de la señal de transporte de audio dependiendo de la primera matriz (P), en donde la primera matriz (P) indica cómo mezclar las dos o más señales de objeto de audio para obtener la pluralidad de canales previamente mezclados, y dependiendo de la segunda matriz (Q), en donde la segunda matriz (Q) indica cómo mezclar la pluralidad de canales mezclados previamente para obtener el uno o más canales de transporte de audio de la señal de transporte de audio,

25 en donde los coeficientes de la primera matriz (P) indica información sobre la primera regla de mezclado, y en donde los coeficientes de la segunda matriz (Q) indica información sobre la segunda regla de mezclado,

30 35 14. Programa informático para implementar el método de la reivindicación 12 o 13 cuando se ejecuta en un ordenador o procesador de señal.

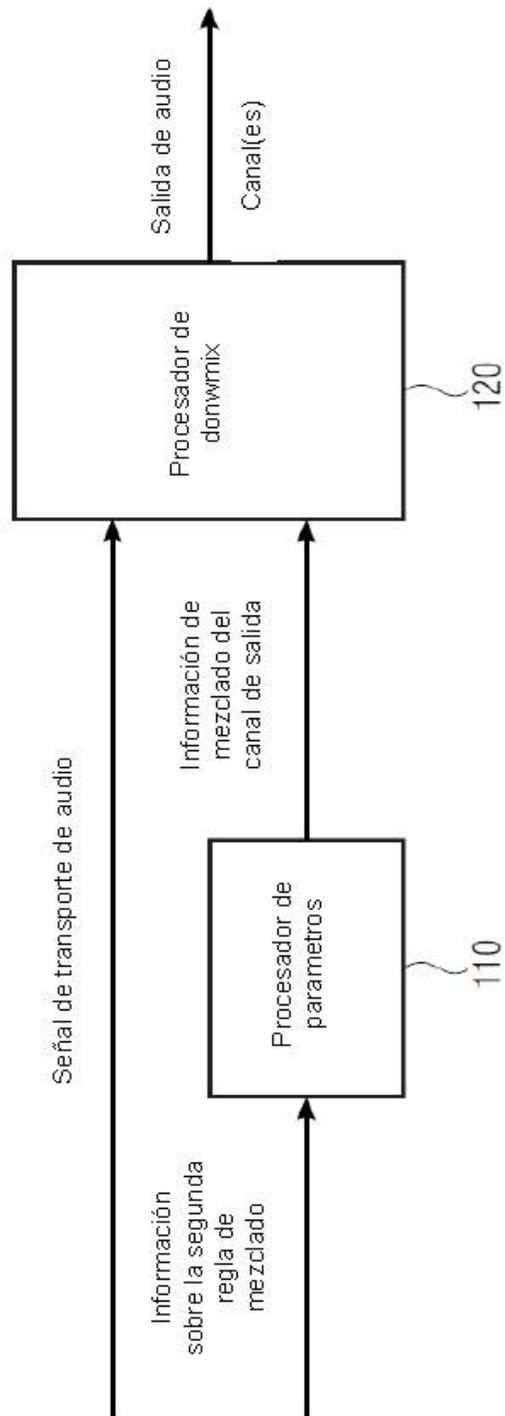


FIG 1

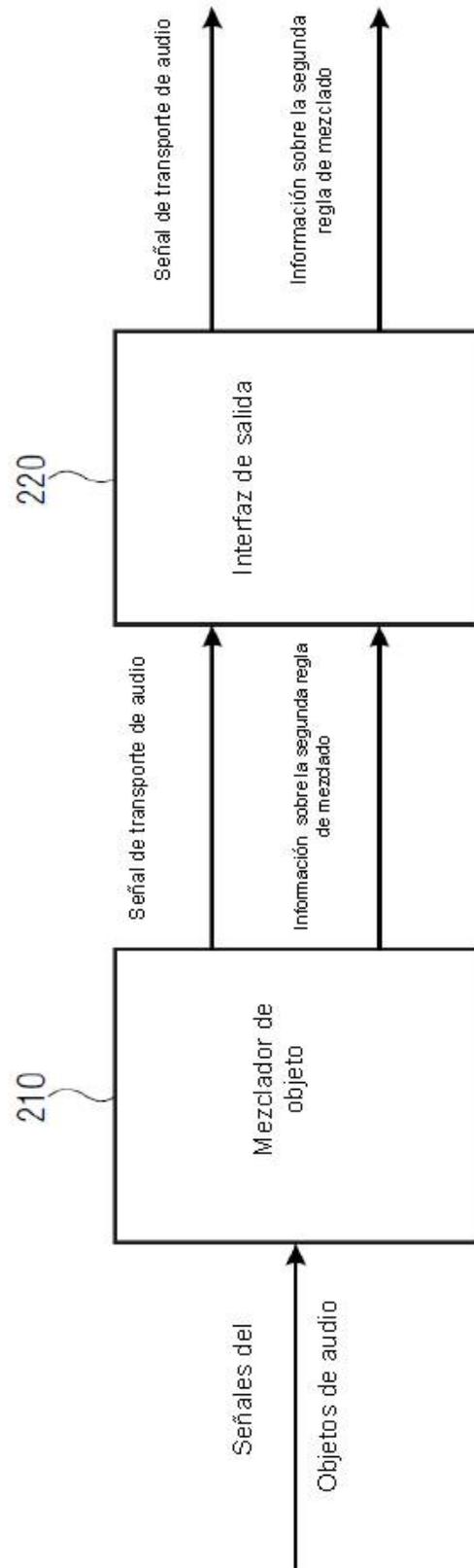


FIG 2

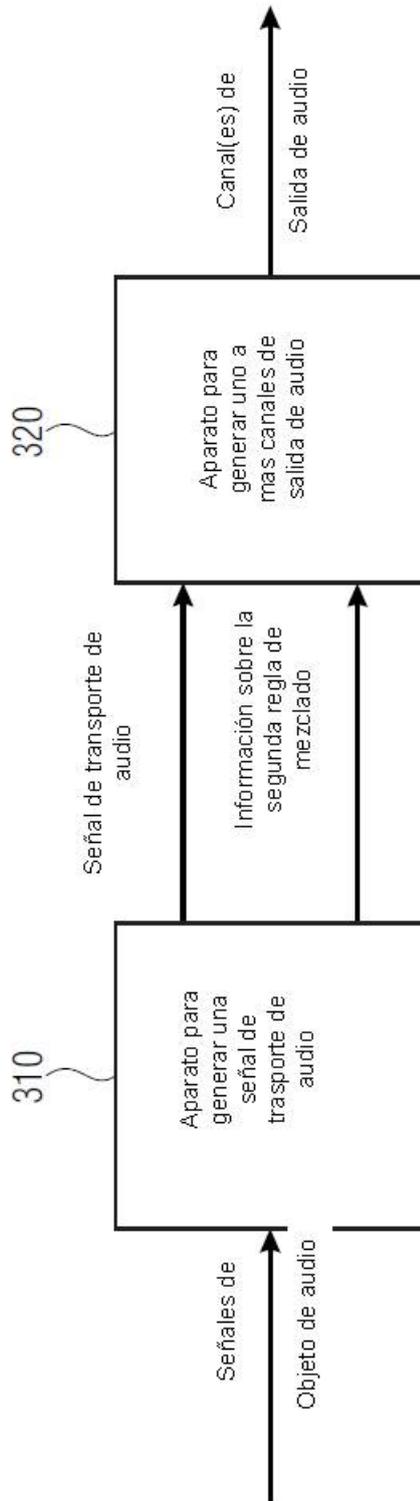


FIG 3

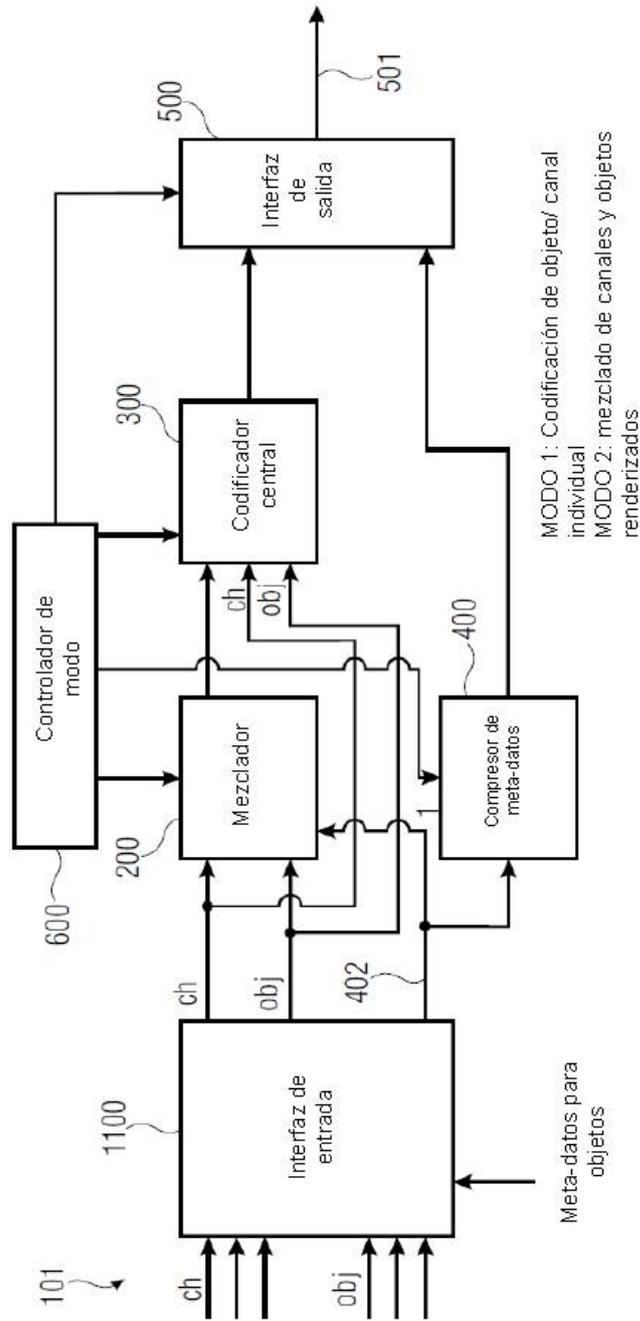


FIG 4
(CODIFICADOR)

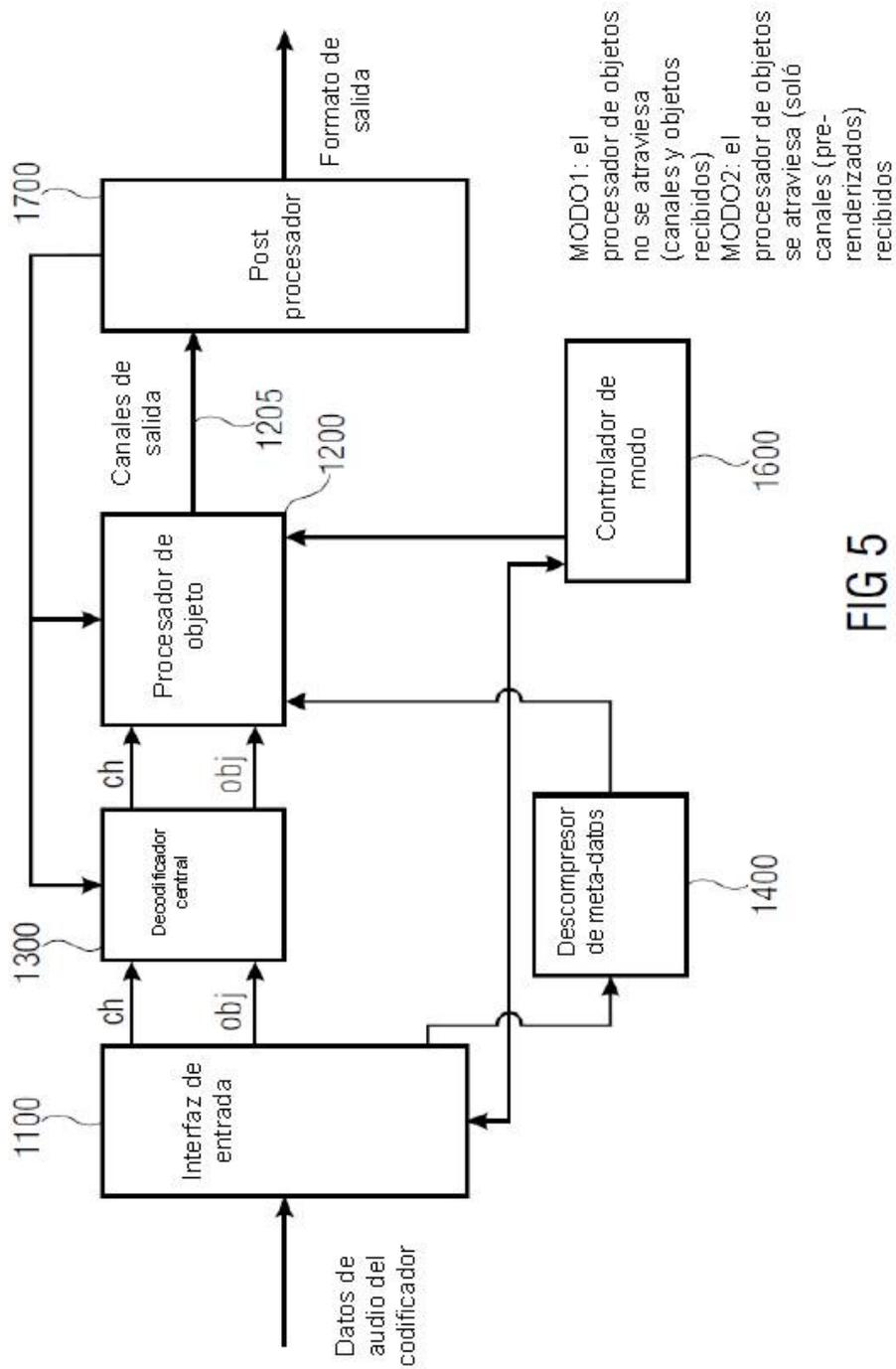


FIG 5
(DECODIFICADOR)

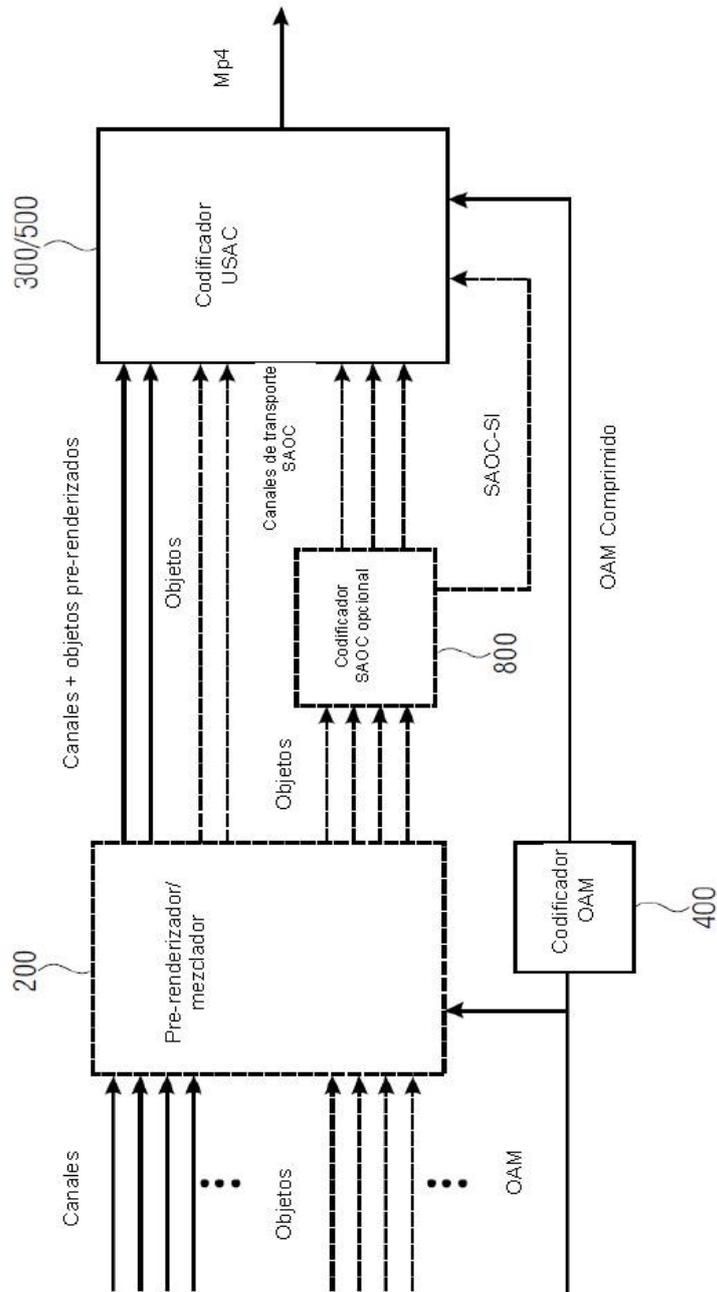


FIG 6
(codificador)

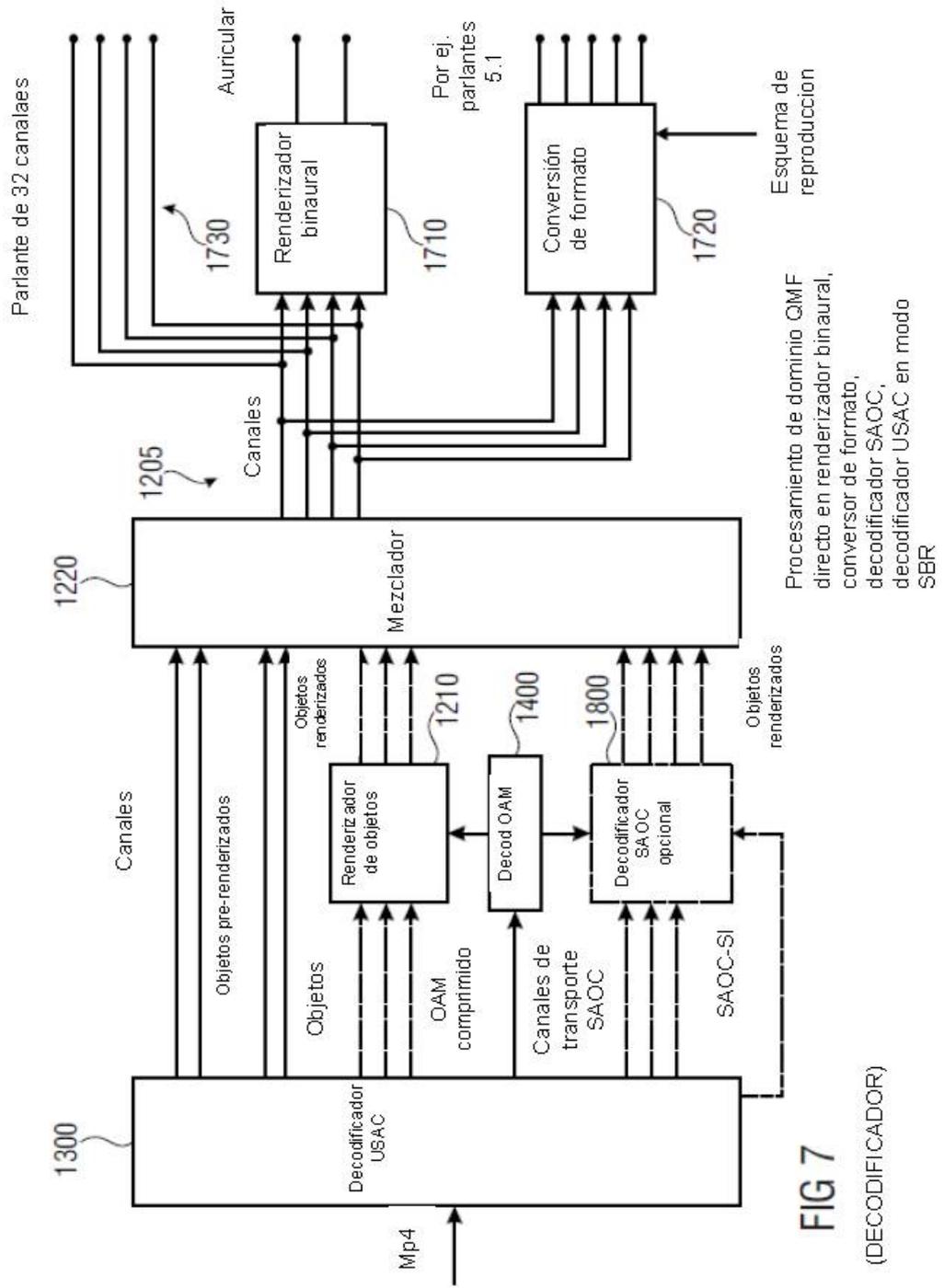


FIG 7
(DECODIFICADOR)

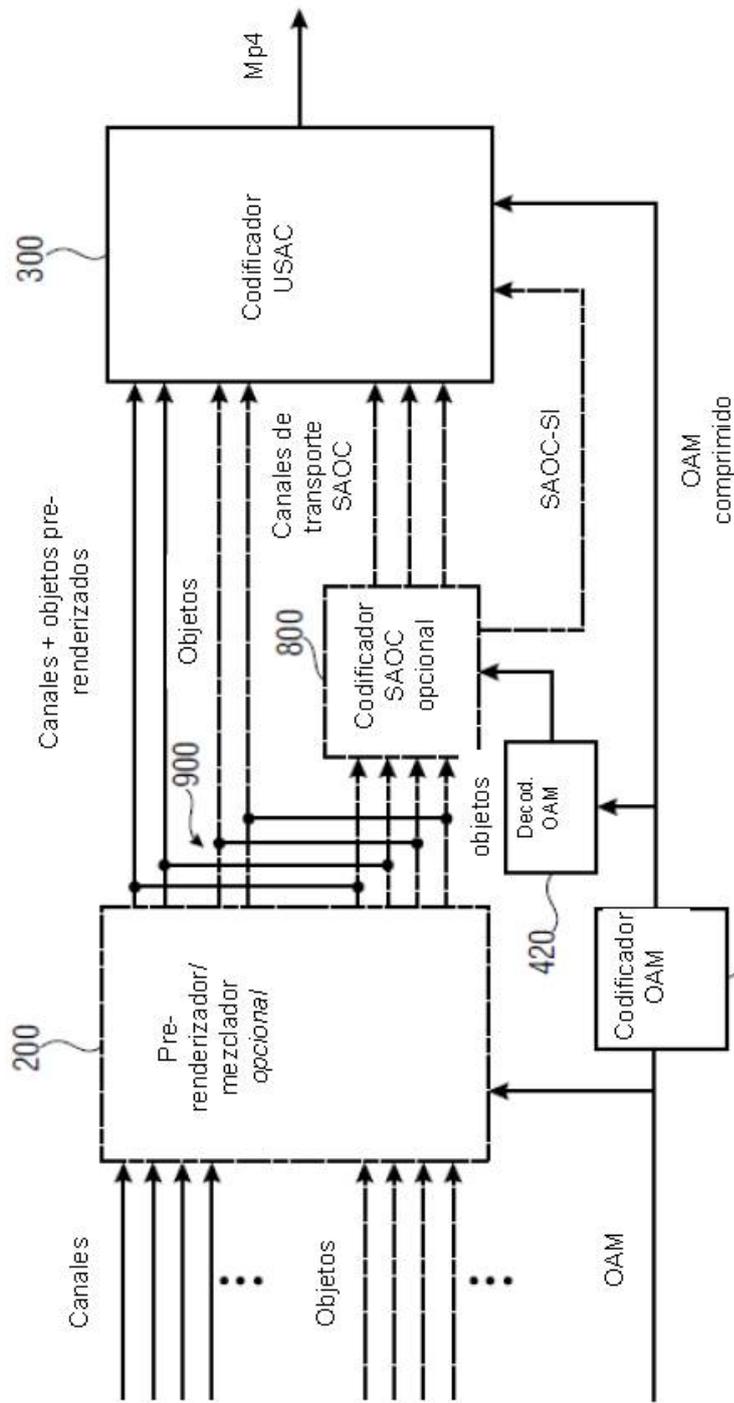


FIG 8
(CODIFICADOR)

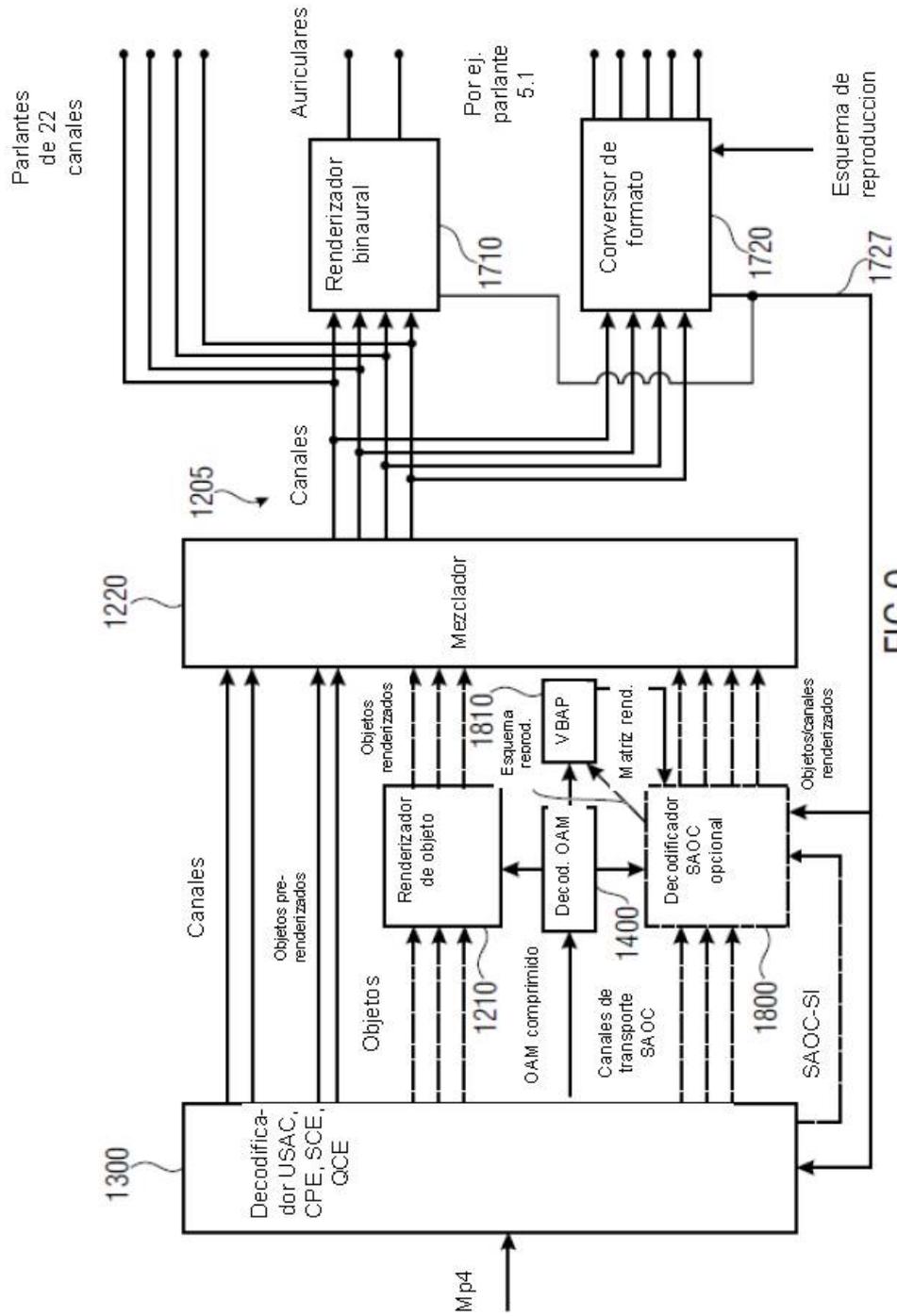


FIG 9
(DECODIFICADOR)

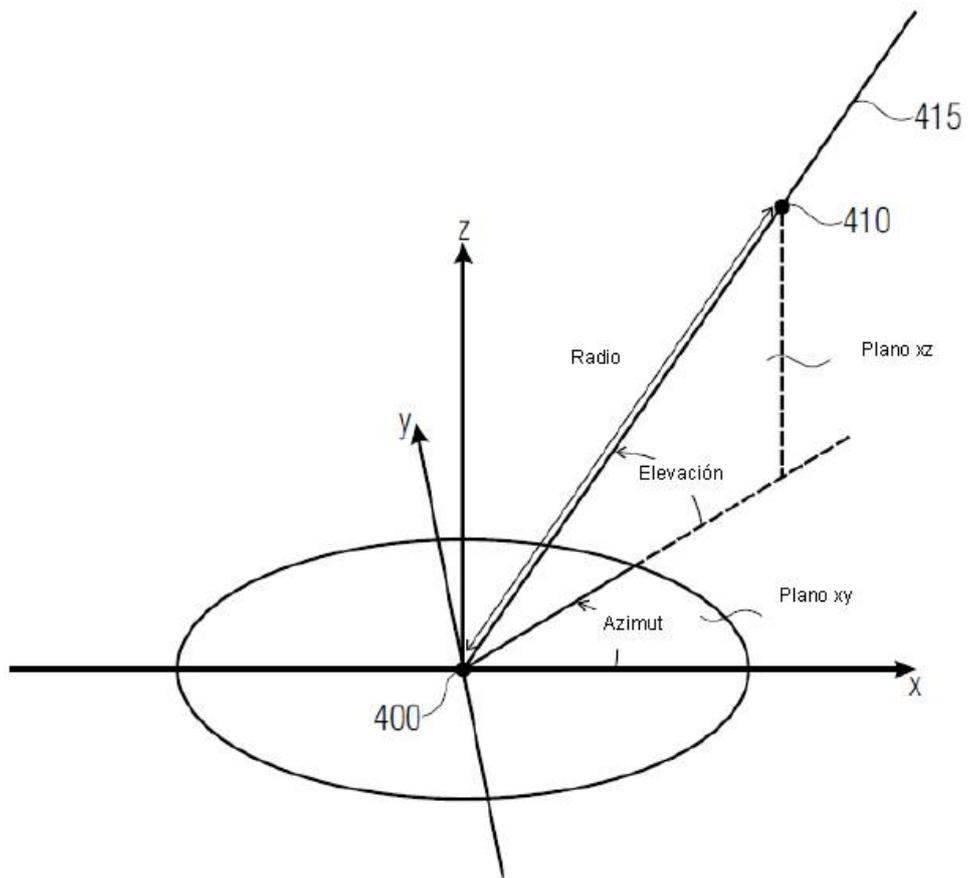


FIG 10

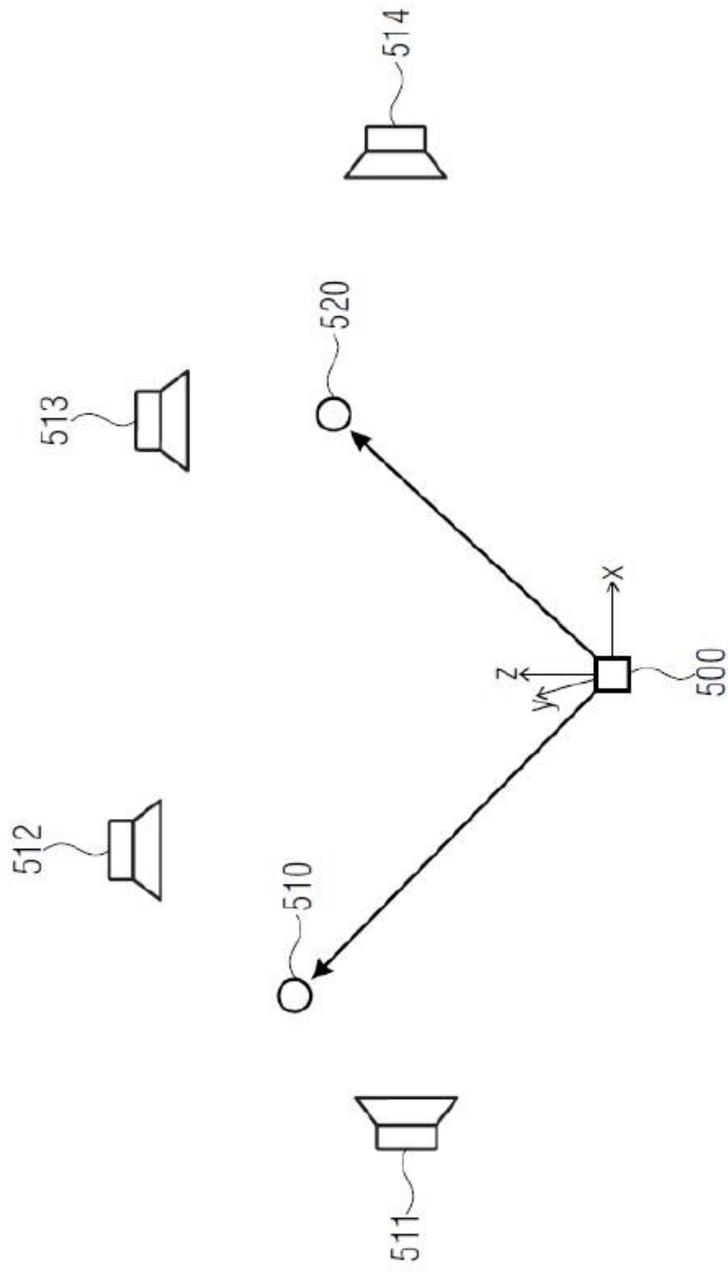


FIG 11