

(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



(11) Número de publicación: **3 013 669**

(51) Int. Cl.:

G10L 19/012 (2013.01)

G10L 19/008 (2013.01)

H04J 3/17 (2006.01)

G10L 19/16 (2013.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Fecha de presentación y número de la solicitud internacional: **31.05.2021 PCT/EP2021/064576**

(87) Fecha y número de publicación internacional: **03.02.2022 WO20222876**

(96) Fecha de presentación y número de la solicitud europea: **31.05.2021 E 21729320 (8)**

(97) Fecha y número de publicación de la concesión europea: **15.01.2025 EP 4189674**

(54) Título: **Aparato, método y programa informático para codificar una escena de audio**

(30) Prioridad:

30.07.2020 EP 20188707

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

14.04.2025

(73) Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.00%)
Hansastraße 27c
80686 München, DE**

(72) Inventor/es:

**FUCHS, GUILLAUME;
TAMARAPU, ARCHIT;
EICHENSEER, ANDREA;
KORSE, SRIKANTH;
DÖHLA, STEFAN y
MULTRUS, MARKUS**

(74) Agente/Representante:

RUÍZ GALLEGOS, Natalia

ES 3 013 669 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, método y programa informático para codificar una escena de audio

- 5 Este documento se refiere, entre otros, a un aparato para generar una escena de audio codificada, y a un aparato para decodificar y/o procesar una escena de audio codificada. El documento también se refiere a los métodos relacionados y unidades de almacenamiento no transitorio, que almacenan instrucciones que cuando se ejecutan por un procesador, hacen que el procesador ejecute un método relacionado.
- 10 Este documento discute métodos sobre el modo de transmisión discontinua (DTX) y la generación de ruido de confort (CNG) para escenas de audio, para las cuales la imagen espacial se codificó paramétricamente por el paradigma de codificación de audio direccional (DirAC) o transmitida en un formato de audio espacial asistido por metadatos (MASA).
- 15 Las realizaciones se refieren con la transmisión discontinua de un audio espacial codificado paramétricamente, tal como un modo de DTX para DirAC y MASA.
- 20 Las realizaciones de la presente invención se refieren a la transmisión y representación eficientes del habla conversacional, por ejemplo, capturada con micrófonos de campo sonoro. La señal de audio así capturada, en general, es llamada un audio tridimensional (3D), puesto que los eventos de sonido pueden localizarse en el espacio tridimensional, lo que refuerza la inmersividad e incrementa tanto la inteligibilidad como la experiencia del usuario.
- 25 La transmisión de una escena de audio, por ejemplo, en tres dimensiones, requiere manipular múltiples canales que usualmente generan una gran cantidad de datos para transmitir. Por ejemplo, la técnica de codificación de audio direccional (DirAC) [1], puede utilizarse para reducir la gran velocidad de los datos originales. La DirAC se considera un enfoque eficiente para analizar la escena de audio y representarla paramétricamente. Tiene una motivación perceptiva, y representa el campo sonoro con la ayuda de una dirección de la llegada (DOA) y la difusión medida por la banda de frecuencia. Se basa en el supuesto de que en un instante del tiempo y para una banda crítica, la resolución espacial del sistema auditivo está limitada para decodificar una indicación para la dirección y otra para la coherencia interaural. El sonido espacial se reproduce a continuación en el dominio de frecuencia, mediante el fundido cruzado de dos corrientes: una corriente difusa no direccional y una corriente no difusa direccional.
- 30 Además, en una conversación típica, cada hablante está en silencio durante aproximadamente sesenta por ciento del tiempo. Al distinguir los cuadros de la señal de audio que contienen el habla ("cuadros activos") de los cuadros que contienen únicamente el ruido de fondo o silencio ("cuadros inactivos"), los codificadores del habla pueden ahorrar una velocidad de los datos significativa. Los cuadros inactivos se perciben típicamente como que portan poca o ninguna información, y los codificadores del habla se configuran usualmente para reducir su tasa de bits para tales cuadros, o incluso no transmitir información. En tal caso, los codificadores se ejecutan en un llamado modo de transmisión discontinua (DTX), que es una manera eficiente para reducir de forma drástica la velocidad de transmisión de un códec de comunicación en la ausencia de entrada de voz. En este modo, la mayoría de los cuadros que se ha determinado que consisten de ruido de fondo únicamente, se omiten de la transmisión, y se reemplazan por alguna generación de ruido de confort (CNG) en el decodificador. Para estos cuadros, una representación paramétrica de muy baja velocidad de la señal, se transmite por los cuadros del descriptor de inserción de silencio (SID) enviados regularmente, pero no en cada cuadro. Esto permite que la CNG en el decodificador produzca un ruido artificial que se asemeja al ruido de fondo real.
- 35 40 45 50
- 40 Las realizaciones de la presente invención se refieren a un sistema de DTX y especialmente, con un SID y CNG para las escenas de audio en 3D, capturadas, por ejemplo, mediante un micrófono para el campo sonoro, y que pueden codificarse paramétricamente por un esquema de codificación basado en el paradigma de DirAC y similar. La presente invención permite una reducción drástica de la demanda de la tasa de bits para transmitir el habla conversacional inmersiva.

Técnica anterior

- 55 El documento US 9,514,757 B2 se refiere a la codificación y decodificación DTX estéreo. El procedimiento cambia entre un codificador estéreo y un codificador estéreo DTX. Cuando se utiliza el codificador DTX, los datos se pueden guardar al no transmitir los datos para todos los cuadros.
- 60 El documento EP 3 511 934 A1 describe cómo implementar la codificación discontinua de una señal de mezcla descendente y un parámetro estéreo de un audio multicanal.
- 60 El documento WO 2020/002448 A1 propone codificar un parámetro de ganancia lateral promediado para un segmento inactivo y una mezcla descendente más parámetros de ganancia lateral para un segmento activo en el contexto de la generación de un parámetro de ruido de confort.

Técnica anterior adicional:

- [1] V. Pulkki, M-V. Laitinen, J. Vilkamo, J. Ahonen, T. Lokki, y T. Pihlajamäki, "Directional audio coding - perception-based reproduction of spatial sound", International Workshop on the Principles and Application on Spatial Hearing, 5 noviembre de 2009, Zao; Miyagi, Japón.
- [2] 3GPP TS 26.194; Voice Activity Detector (VAD); - 3GPP technical specification Retrieved on 2009-06-17.
- [3] 3GPP TS 26.449, "Codec for Enhanced Voice Services (EVS); Comfort Noise Generation (CNG) Aspects".
10 [4] 3GPP TS 26.450, "Codec for Enhanced Voice Services (EVS); Discontinuous Transmission (DTX)"
- [5] A. Lombard, S. Wilde, E. Ravelli, S. Döhla, G. Fuchs y M. Dietz, "Frequency-domain Comfort Noise Generation for Discontinuous Transmission in EVS," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, 2015, pp. 5893-5897, doi: 10.1109/ICASSP.2015.7179102.
15
- [6] V. Pulkki, "Virtual source positioning using vector base amplitude panning", J. Audio Eng. Soc., 45(6):456-466, junio de 1997.
- 20 [7] J. Ahonen y V. Pulkki, "Diffuseness estimation using temporal variation of intensity vectors", en Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA, Mohonk Mountain House, New Paltz, 2009.
- [8] T. Hirvonen, J. Ahonen, and V. Pulkki, "Perceptual compression methods for metadata in Directional Audio Coding applied to audiovisual teleconference", AES 126th Convention 2009, mayo 7-10, Múnich, Alemania.
25
- [9] Vilkamo, Juha & Bäckström, Tom & Kuntz, Achim. (2013). Optimized Covariance Domain Framework for Time--Frequency Processing of Spatial Audio. Journal of the Audio Engineering Society. 61.
- [10] M. Laitinen and V. Pulkki, "Converting 5.1 audio recordings to B-format for directional audio coding reproduction," 30 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Praga, 2011, pp. 61-64, doi: 10.1109/ICASSP.2011.5946328.

Sumario

- 35 La invención se define en las reivindicaciones independientes. Se definen las realizaciones útiles en las reivindicaciones dependientes.
En particular, la invención se refiere a un aparato para generar una escena de audio codificada de una señal de audio, que tiene un primer cuadro y un segundo cuadro, que comprende:
40 un generador de parámetros de campo sonoro para determinar una primera representación de parámetros de campo sonoro para el primer cuadro de la señal de audio en el primer cuadro, y una segunda representación de parámetros de campo sonoro para el segundo cuadro de la señal de audio en el segundo cuadro; y
45 un detector de actividad para analizar la señal de audio para determinar, dependiendo de la señal de audio, que el primer cuadro es un cuadro activo, y que el segundo cuadro es un cuadro inactivo,
en donde el generador de parámetros de campo sonoro está configurado para determinar, del segundo cuadro (de la señal de audio), una pluralidad de fuentes de sonido individuales y determinar, para cada fuente de sonido, una descripción paramétrica para el segundo cuadro,
50 en donde el generador de campo sonoro está configurado para descomponer el segundo cuadro en una pluralidad de contenedores de frecuencia, cada contenedor de frecuencia representa una fuente de sonido individual, y para determinar, para cada contenedor de frecuencia, al menos un parámetro espacial inactivo como la segunda representación de parámetros de campo sonoro para el segundo cuadro, el al menos un parámetro espacial inactivo que comprende un parámetro de dirección, un parámetro de dirección de llegada, un parámetro de difusividad o un parámetro de relación de energía,
55 el aparato que comprende además:
60 un codificador de señal de audio para generar una señal de audio codificada, la señal de audio codificada que proporciona una señal de audio codificada para el primer cuadro es el cuadro activo y la descripción paramétrica para el segundo cuadro es el cuadro inactivo; y

un formador de señales codificadas para componer la escena de audio codificada juntando la primera representación de parámetros de campo sonoro para el primer cuadro, la segunda representación de parámetros de campo sonoro para el segundo cuadro, la señal de audio codificada para el primer cuadro, y la descripción paramétrica para el segundo cuadro.

5

Figuras

La figura 1 (que se divide entre la figura 1a y la figura 1b) muestra un ejemplo de acuerdo con la técnica anterior que puede utilizarse para el análisis y la síntesis de acuerdo con los ejemplos.

10

La figura 2 muestra un ejemplo de un decodificador y un codificador, de acuerdo con los ejemplos.

La figura 3 muestra un ejemplo de un codificador, de acuerdo con la invención reivindicada.

15

Las figuras 4 y 5 muestran ejemplos de los componentes.

La figura 5 muestra un ejemplo de un componente, de acuerdo con un ejemplo.

Las figuras 6-11 muestran ejemplos de decodificadores útiles para comprender la invención.

20

Realizaciones

En primer lugar, se proporciona alguna discusión de los paradigmas conocidos (DTX, DirAC, MASA, etc.), con la descripción de las técnicas, algunas de las cuales pueden, al menos en algunos casos, implementarse en los ejemplos de la invención.

DTX

Los generadores de ruido de confort se utilizan usualmente en la transmisión discontinua (DTX) del habla. En tal modo, el habla se clasifica primero en cuadros activos e inactivos mediante un detector de actividad de voz (VAD). Un ejemplo de un VAD puede encontrarse en [2]. Basándose en el resultado del VAD, únicamente los cuadros con el habla activa se codifican y transmiten a la tasa de bits nominal. Durante las pausas largas, en donde sólo está presente el ruido de fondo, la tasa de bits se disminuye o pone en cero, y el ruido de fondo se codifica de manera episódica y paramétrica. La tasa de bits promedio se reduce entonces de manera significativa. El ruido se genera durante los cuadros inactivos en el lado del decodificador por un generador de ruido de confort (CNG). Por ejemplo, los codificadores del habla AMR-WB [2] y 3GPP EVS [3, 4], tienen ambos la posibilidad de ejecutarse en el modo DTX. Un ejemplo de una CNG eficiente se proporciona en [5].

Las realizaciones de la presente invención extienden este principio en una forma que aplica el mismo principio al habla conversacional inmersiva con localización espacial de los eventos de sonido.

DirAC

DirAC es una reproducción motivada perceptualmente del sonido espacial. Se supone que en un instante del tiempo y para una banda crítica, la resolución espacial del sistema auditivo se limita a decodificar una indicación para la dirección y otra para la coherencia interaural.

Basándose en estas suposiciones, la DirAC representa el sonido espacial en una banda de frecuencia fundiendo de manera cruzada dos corrientes: una corriente difusa no direccional y una corriente no difusa direccional. El procesamiento de DirAC se realiza en dos fases: el análisis y la síntesis como se muestra en la figura 1 (las figuras 1a que muestra una síntesis, la figura 1b que muestra un análisis).

En la etapa de análisis DirAC, un micrófono coincidente de primer orden en formato B se considera como la entrada, y la difusión y la dirección de la llegada del sonido se analiza en el dominio de frecuencia.

55

En la etapa de síntesis DirAC, el sonido se divide en dos corrientes, la corriente no difusa y la corriente difusa. La corriente no difusa se reproduce como fuentes puntuales utilizando la panoramización de amplitud, que puede hacerse utilizando la panoramización de amplitud base del vector (VBAP) [6]. La corriente difusa es en general, responsable de la sensación de envolvimiento, y se produce transmitiéndola a las señales decorrelacionadas mutuamente de los altavoces.

Los parámetros DirAC, también llamados metadatos parciales o metadatos DirAC en lo siguiente, consisten de tuplas de difusión y dirección. La dirección puede representarse en coordenadas esféricas mediante dos ángulos, el azimut y la elevación, mientras que la difusión puede ser un factor escalar de entre 0 y 1.

Algunos trabajos se han hecho para reducir el tamaño de los metadatos, para permitir que el paradigma DirAC se utilice para la codificación del audio espacial, y en escenarios de teleconferencia [8].

- 5 Según los conocimientos de los inventores, ningún sistema de DTX se ha construido o propuesto alrededor de un codificador de audio espacial paramétrico e incluso menos, basado en un paradigma DirAC. Esto es el objeto de las realizaciones de la presente invención.

MASA

- 10 El audio espacial asistido por metadatos (MASA), es un formato de audio espacial derivado del principio DirAC, que puede calcularse directamente de las señales sin tratar del micrófono y transmitirse a un códec de audio sin la necesidad de pasar a través de un formato intermedio como Ambisonics. Un conjunto de parámetros, que puede consistir de un parámetro de dirección, por ejemplo, en las bandas de frecuencia (por ejemplo, indicando la proporción 15 de energía del sonido que es direccional), también puede utilizarse como los metadatos espaciales para un códec o renderizador de audio. Estos parámetros pueden estimarse de las señales de audio capturadas por un micrófono matriz; por ejemplo, una señal mono o estéreo puede generarse de las señales del micrófono matriz para transmitirse con los metadatos espaciales. La señal mono o estéreo podría codificarse, por ejemplo, con un codificador principal como 3GPP EVS o un derivado del mismo. Un decodificador puede decodificar las señales de audio en y procesar el 20 sonido en las bandas de frecuencia (utilizando los metadatos espaciales transmitidos), para obtener la salida espacial, que podría ser una salida binaural, una señal con múltiples canales del altavoz o una señal con múltiples canales en un formato Ambisonics.

Motivación

- 25 La comunicación con habla inmersiva es un nuevo dominio de investigación y existen muy pocos sistemas, además, ningún sistema de DTX se diseñó para tal aplicación.
- Sin embargo, puede ser sencillo combinar las soluciones existentes. Uno puede por ejemplo, aplicar de manera independiente el DTX en cada señal con múltiples canales individuales. Este enfoque simplista enfrenta varios 30 problemas. Para esto, uno necesita transmitir de manera discreta cada canal individual, que es incompatible con las restricciones de comunicación de baja tasa de bits, y por lo tanto, difícilmente compatible con el DTX, que se diseña para casos de comunicación con una baja tasa de bits. Además, se requiere a continuación, sincronizar las decisiones 35 del VAD a través de los canales, para evitar rarezas y efectos no enmascarados, y también para aprovechar completamente la reducción de la tasa de bits del sistema de DTX. En realidad, para interrumpir la transmisión y beneficiarse de esto, uno necesita asegurarse que las decisiones de actividad de la voz se sincronicen a través de todos los canales.

- Otro problema surge en el lado del receptor, cuando genera el ruido de fondo que falta durante los cuadros inactivos 40 por el/los generador(es) de ruido de confort. Para las comunicaciones inmersivas, especialmente cuando se aplica directamente el DTX a los canales individuales, se requiere un generador por canal. Si estos generadores, que típicamente muestran un ruido aleatorio, se utilizan de manera independiente, la coherencia entre los canales será cero o cercana a cero, y puede desviarse perceptualmente del paisaje sonoro original. Por otra parte, si sólo se utiliza 45 un generador y el ruido de confort resultante se copia a todos los canales de salida, la coherencia será muy alta, y la inmersividad se reducirá drásticamente.

- Estos problemas pueden solucionarse parcialmente aplicando el DTX no directamente a los canales de entrada o 50 salida del sistema, sino en su lugar, después de un esquema de codificación del audio espacial paramétrico, como DirAC, en los canales de transporte resultantes, que son usualmente una versión mezclada de manera descendente o reducida de la señal con múltiples canales original. En este caso, es necesario definir cómo se parametrizan los cuadros inactivos y cómo luego los espacializa el sistema DTX. Esto no es trivial, y es el objeto de las realizaciones de la presente invención. La imagen espacial debe ser consistente entre los cuadros activos e inactivos, y debe ser tan perceptualmente fiel como sea posible al ruido de fondo original.

- 55 La figura 3 muestra un codificador 300 de acuerdo con un ejemplo de la invención reivindicada. El codificador 300 puede generar una escena de audio codificada 304 de una señal de audio 302.

- La señal de audio 304 o la escena de audio 304 (y también otras señales de audio descritas a continuación), pueden dividirse en cuadros (por ejemplo, puede ser una secuencia de cuadros). Los cuadros pueden estar asociados a 60 segmentos temporales, que pueden definirse posteriormente unos con otros (en algunos ejemplos, un aspecto precedente puede superponerse con un cuadro subsecuente). Para cada cuadro, los valores en el dominio del tiempo (TD) o el dominio de frecuencia (FD), pueden escribirse en la secuencia de bits 304. En el TD, los valores pueden proporcionarse para cada muestra (cada cuadro tiene, por ejemplo, una secuencia discreta de muestras). En el FD, los valores pueden proporcionarse para cada contenedor de frecuencia. Como se explicará posteriormente, cada

cuadro puede clasificarse (por ejemplo, mediante un detector de actividad), ya sea como un cuadro activo 306 (por ejemplo, un cuadro no vacío) o un cuadro inactivo 308 (por ejemplo, cuadros vacíos, o cuadros de silencio, o cuadros con únicamente ruido). Diferentes parámetros (por ejemplo, parámetros espaciales activos 316 o parámetros espaciales inactivos 318), también pueden proporcionarse en asociación con el cuadro activo 306 y el cuadro inactivo 308 (en el caso en que no haya datos, el número de referencia 319 muestra que no se proporcionan datos).

La señal de audio 302 puede, por ejemplo, ser una señal de audio multicanal (por ejemplo, con dos canales o más). La señal de audio 302 puede ser, por ejemplo, una señal de audio estéreo. La señal de audio 302 puede ser, por ejemplo, una señal Ambisonics, por ejemplo, en el formato A o el formato B. La señal de audio 302 puede tener, por ejemplo, un formato MASA (audio espacial asistido por metadatos). La señal de audio 302 puede tener un formato de entrada que es un formato Ambisonics de primer orden, un formato Ambisonics de orden superior, un formato multicanal asociado con una instalación del altavoz dada, tal como 5.1 o 7.1 o 7.1 + 4, o uno o más canales de audio que representan uno o varios objetos de audio diferentes, localizados en un espacio, como se indica por la información incluida en los metadatos asociados, o un formato de entrada que es una representación del audio espacial asociada con los metadatos. La señal de audio 302 puede comprender una señal del micrófono como captada por micrófonos reales o micrófonos virtuales. La señal de audio 302 puede comprender una señal del micrófono creada de manera sintética (por ejemplo, estando en un formato Ambisonics de primer orden, o un formato Ambisonics de orden superior).

La escena de audio 304 puede comprender al menos uno o una combinación de:

una primera representación de parámetros de campo sonoro (por ejemplo, parámetro espacial activo) 316 para el primer cuadro 306;

una segunda representación de parámetros de campo sonoro (por ejemplo, parámetro espacial inactivo) 318 para el segundo cuadro 308;

una señal de audio codificada 346 para el primer cuadro 306; y

una descripción paramétrica 348 para el segundo cuadro 308 (en algunos ejemplos, el parámetro espacial inactivo 318 puede incluirse en la descripción paramétrica 348, pero la descripción paramétrica 348 también puede incluir otros parámetros, que no son parámetros espaciales).

Los cuadros activos 306 (primeros cuadros) pueden ser aquellos cuadros que contienen habla (o, en algunos ejemplos, también otros sonidos de audio diferentes de ruido puro). Puede entenderse que los cuadros inactivos 308 (segundos cuadros) son aquellos cuadros que no comprenden habla (o, en algunos ejemplos, también otros sonidos de audio diferentes de ruido puro), y pueden entenderse como que contienen únicamente ruido.

Un analizador de escena de audio (generador de parámetros de campo sonoro) 310 puede proporcionarse, por ejemplo, para generar una versión del canal de transporte 324 (subdividido entre 326 y 328) de la señal de audio 302. Aquí, podemos referirnos al/los canal(es) de transporte 326 de cada primer cuadro 306 y/o los canales de transporte 328 de cada segundo cuadro 308 (puede entenderse que el/los canal(es) de transporte 328 proporcionan una descripción paramétrica del silencio o el ruido, por ejemplo). El/los canal(es) de transporte 324 (326, 328) puede(n) ser una versión mezclada de manera descendente del formato de entrada 302. En términos generales, cada uno de los canales de transporte 326, 328 puede ser, por ejemplo, un canal único si la señal de audio de entrada 302 es un canal estéreo. Si la señal de audio de entrada 302 tiene más de dos canales, la versión mezclada de manera descendente 324 de la señal de audio de entrada 302, puede tener menos canales que la señal de audio de entrada 302, pero todavía más de un canal en algunos ejemplos (por ejemplo, si la señal de audio de entrada 302 tiene cuatro canales, la versión mezclada de manera descendente 324 puede tener uno, dos o tres canales).

El analizador de señal de audio 310 puede, de manera adicional o alterna, proporcionar los parámetros de campo sonoro (parámetros espaciales), indicados con 314. En particular, los parámetros de campo sonoro 314 pueden incluir los parámetros espaciales activos (primeros parámetros espaciales o primera representación del parámetro espacial) 316 asociados con el primer cuadro 306, y los parámetros espaciales inactivos (segundos parámetros espaciales o segunda representación del parámetro espacial) 318 asociados con el segundo cuadro 308. Cada parámetro espacial activo 314 (316, 318) puede comprender (por ejemplo, ser), un parámetro que indica una característica espacial de la señal de audio (302), por ejemplo, con respecto a una posición del oyente. En algunos otros ejemplos, el parámetro espacial activo 314 (316, 318) puede comprender (por ejemplo, ser), al menos parcialmente, un parámetro que indica una característica de la señal de audio 302, con respecto a la posición de los altavoces. En algunos ejemplos, el parámetro espacial activo 314 (316, 318) puede ser, o comprender al menos parcialmente, las características de la señal de audio tomadas de la fuente de la señal.

Por ejemplo, los parámetros espaciales 314 (316, 318) pueden incluir, parámetros de difusión: por ejemplo, uno o más parámetro(s) de difusión que indican una relación de difusión a la señal con respecto al sonido en el primer cuadro 306 y/o en el segundo cuadro 308, o uno o más parámetro(s) de relación de energía, que indican una relación de

energía de un sonido directo y un sonido difuso en el primer cuadro 306, y/o en el segundo cuadro 308, o un parámetro de la coherencia intercanal/entorno en el primer cuadro 306 y/o en el segundo cuadro 308, o una(s) relación(es) de potencia coherente a difusa en el primer cuadro 306 y/o en el segundo cuadro 308, o una(s) relación(es) de la señal a la difusión en el primer cuadro 306 y/o en el segundo cuadro 308.

- 5 En los ejemplos, el/los parámetro(s) espaciales activos (primera representación de parámetros de campo sonoro) 316 y/o el/los parámetro(s) espacial(es) inactivo(s) 318 (segunda representación de parámetros de campo sonoro), pueden obtenerse de la señal de entrada 302 en su versión del canal completo, o un subconjunto del mismo, como el componente de primer orden de una señal de entrada Ambisonics de orden superior.
 - 10 El aparato 300 puede incluir un detector de actividad 320. El detector de actividad 320 puede analizar la señal de audio de entrada (ya sea en su versión de entrada 302 o en su versión mezclada de manera descendente 324), para determinar, dependiendo de la señal de audio (302 o 324), si un cuadro es un cuadro activo 306 o un cuadro inactivo 308, realizando así una clasificación del cuadro. Como puede observarse de la figura 3, puede suponerse que el detector de actividad 320 controla (por ejemplo, a través del control 321), un primer desviador 322 y un segundo desviador 322a. El primer desviador 322 puede seleccionar entre el parámetro espacial activo 316 (primera representación de parámetros de campo sonoro) y los parámetros espaciales inactivos 318 (segunda representación de parámetros de campo sonoro). Por lo tanto, el detector de actividad 320 puede decidir si los parámetros espaciales activos 316 o los parámetros espaciales inactivos 318 se van a emitir (por ejemplo, señalizar en la secuencia de bits 304). El mismo control 321 puede controlar el segundo desviador 322a, que puede seleccionar entre emitir el primer cuadro 326 (306) en el canal de transporte 324, o el segundo cuadro 328 (308) (por ejemplo, la descripción paramétrica) en el canal de transporte 326. Las actividades del primer y segundo desviadores 322 y 322a, están coordinadas unas con otras: cuando los parámetros espaciales activos 316 se emiten, entonces los canales de transporte 326 del primer cuadro 306 también se emiten, y cuando los parámetros espaciales inactivos 318 se emiten, entonces los canales de transporte 328 del primer cuadro 306, los canales de transporte se emiten. Esto es debido a que los parámetros espaciales activos 316 (primera representación de parámetros de campo sonoro), describen las características espaciales del primer cuadro 306, mientras que los parámetros espaciales inactivos 318 (segunda representación de parámetros de campo sonoro), describen las características espaciales del segundo cuadro 308.
 - 15
 - 20
 - 25
 - 30
 - 35
 - 40
 - 45
 - 50
 - 55
 - 60
- El detector de actividad 320 puede, por lo tanto, decidir básicamente cuál entre el primer cuadro 306 (326, 346), y sus parámetros relacionados (316), y el segundo cuadro 308 (328, 348), y sus parámetros relacionados (318), se van a emitir. El detector de actividad 320 también puede controlar la codificación de alguna señalización en la secuencia de bits, que señalan si el cuadro es uno activo o uno inactivo (otras técnicas pueden utilizarse).
- El detector de actividad 320 puede realizar el procesamiento en cada cuadro 306/308 de la señal de audio de entrada 302 (por ejemplo, midiendo la energía en el cuadro, por ejemplo, en todos, o al menos una pluralidad de los contenedores de frecuencia de los cuadros particulares de la señal de audio), y puede clasificar el cuadro particular como que es un primer cuadro 306 o un segundo cuadro 308. En términos generales, el detector de actividad 320 puede decidir un solo resultado de la clasificación para un solo cuadro completo, sin distinguir entre diferentes contenedores de frecuencia y diferentes muestras del mismo cuadro. Por ejemplo, un resultado de la clasificación podría ser "habla" (que equivaldría al primer cuadro 306, 326, 346, descrito espacialmente por los parámetros espaciales activos 316) o "silencio" (que equivaldría al segundo cuadro 308, 328, 348, descrito espacialmente por los parámetros espaciales inactivos 318). Por lo tanto, de acuerdo con la clasificación ejercida por el detector de actividad 320, los desviadores 322 y 322a pueden ejecutar su conmutación, y su resultado es en principio válido para todos los contenedores de frecuencia (y las muestras) del cuadro clasificado.
- El aparato 300 puede incluir un codificador de señal de audio 330. El codificador de señal de audio 330 puede generar una señal de audio codificada 344. El codificador de señal de audio 330 puede, en particular, proporcionar una señal de audio codificada 346 para el primer cuadro (306, 326), por ejemplo, generada por un canal de transporte codificador 340, que puede ser parte del codificador de señal de audio 330. La señal de audio codificada 344 puede ser o incluir una descripción paramétrica 348 del silencio (por ejemplo, la descripción paramétrica de ruido), y puede generarse, por un descriptor SI del canal de transporte 350, que puede ser parte del codificador de señal de audio 330. El segundo cuadro 348 generado puede corresponder a al menos un segundo cuadro 308 de la señal de entrada de audio original 302, y a al menos un segundo cuadro 328 de la señal de mezcla descendente 324, y puede describirse espacialmente por los parámetros espaciales inactivos 318 (segunda representación de parámetros de campo sonoro). En particular, la señal de audio codificada 344 (ya sea 346 o 348), también puede estar en el canal de transporte (y puede, por lo tanto, ser una señal de mezcla descendente 324). La señal de audio codificada 344 (ya sea 346 o 348) puede comprimirse, para reducir su tamaño.
- El aparato 300 puede incluir un formador de señal codificada 370. El formador de señal codificada 370 puede escribir una versión codificada de al menos la escena de audio codificada 304. El formador de señal codificada 370, puede operar juntando la primera representación de parámetros de campo sonoro (activo) 316 para el primer cuadro 306, la segunda representación de parámetros de campo sonoro (inactivo) 318 para el segundo cuadro 308, la señal de audio codificada 346 para el primer cuadro 306, y la descripción paramétrica 348 para el segundo cuadro 308. En

consecuencia, la escena de audio 304 puede ser una secuencia de bits, que puede ya sea transmitirse o almacenarse (o ambos), y utilizarse por un decodificador genérico para generar una señal de audio a ser emitida, que es una copia de la señal de entrada original 302. En la escena de audio (secuencia de bits) 304, la secuencia de "primeros cuadros"/"segundos cuadros", puede obtenerse así, para permitir una reproducción de la señal de entrada 306.

- 5 La figura 2 muestra un ejemplo de un codificador 300 y un decodificador 200 (el decodificador 200 es útil para entender la invención). El codificador 300 puede ser el mismo de (o una variación de) aquél de la figura 3 en algunos ejemplos (en algunos otros ejemplos, pueden ser realizaciones diferentes). El codificador 300 tiene una entrada de la señal de audio 302 (que puede, por ejemplo, estar en un formato B), y puede tener un primer cuadro 306 (que puede ser, por ejemplo, un cuadro activo), y un segundo cuadro 308 (que puede ser, por ejemplo, un cuadro inactivo). La señal de audio 302 puede proporcionarse como la señal 324 (por ejemplo, como la señal de audio codificada 326 para el primer cuadro y la señal de audio codificada 328, o la representación paramétrica, para el segundo cuadro), al codificador de señal de audio 330, después de una selección interna en el selector 320 (que puede incluir el audio asociado con los desviadores 322 y 322a). Particularmente, el bloque 320 también puede tener las capacidades para formar la mezcla descendente de la señal de entrada 302 (306, 308) en los canales de transporte 324 (326, 328). Básicamente, puede entenderse que el bloque 320 (bloque de formación del haz/selección de la señal) incluye las funcionalidades del detector de actividad 320 de la figura 3, pero algunas otras funcionalidades (tales como la generación de los parámetros espaciales 316 y 318), que en la figura 3, se realizan por el bloque 310, pueden realizarse por el "bloque de análisis DirAC" 310 de la figura 2. Por lo tanto, la señal del canal 324 (326, 328) puede ser una versión mezclada de manera descendente de la señal original 302. En algunos casos, sin embargo, podría también ser posible que no se realice un mezclado descendente en la señal 302, y una señal 324 es simplemente una selección entre el primer y segundo cuadros. El codificador de señal de audio 330 puede incluir, al menos uno de los bloques 340 y 350, como se explicó anteriormente. El codificador de señal de audio 330 puede emitir la señal de audio codificada 344, ya sea para el primer cuadro 346 o para el segundo cuadro 348. La figura 2 no muestra el formador de la señal codificada 370, que puede no obstante, estar presente.

Como se muestra, el bloque 310 puede incluir, un bloque del análisis DirAC (o más en general, el generador de parámetros de campo sonoro 310). El bloque 310 (generador de parámetros de campo sonoro), puede incluir un análisis de banco de filtros 390. El análisis de banco de filtros 390 puede subdividir cada cuadro de la señal de entrada 302 en una pluralidad de contenedores de frecuencia, que puede ser la salida 391 del análisis de banco de filtros 390. Un bloque de estimación de difusión 392a puede proporcionar los parámetros de difusión 314a (que pueden ser un parámetro de difusión del/los parámetro(s) espaciales activos 316 para un cuadro activo 306 o un parámetro de difusión de/los parámetro(s) espaciales inactivos 318 para un cuadro inactivo 308), por ejemplo, para cada contenedor de frecuencia de la pluralidad de contenedores de frecuencia 391 emitidos por el análisis de banco de filtros 390. El generador de parámetros de campo sonoro 310 puede incluir un bloque de estimación de la dirección 392b, cuya salida 314b puede ser un parámetro de dirección (que puede ser un parámetro de dirección de/los parámetro(s) espacial(es) activo(s) 316 para un cuadro activo 306 o un parámetro de dirección de/los parámetro(s) espacial(es) inactivo(s) 318 para un cuadro inactivo 308), por ejemplo, para cada contenedor de frecuencia de la pluralidad de contenedores de frecuencia 391 emitidos por el análisis de banco de filtros 390.

- 40 La figura 4 muestra un ejemplo del bloque 310 (generador de parámetros de campo sonoro). El generador de parámetros de campo sonoro 310 puede ser el mismo que el de la figura 2 y/o puede ser el mismo o implementar al menos las funcionalidades del bloque 310 de la figura 3, a pesar del hecho de que el bloque 310 de la figura 3 también es capaz de realizar un mezclado descendente de la señal de entrada 302, aunque esto no se muestra (o no se implementa) en el generador de parámetros de campo sonoro 310 de la figura 4.

- 45 El generador de parámetros de campo sonoro 310 de la figura 4, puede incluir un bloque de análisis de banco de filtros 390 (que puede ser igual al bloque de análisis de banco de filtros 390 de la figura 2). El bloque de análisis de banco de filtros 390 puede proporcionar la información del dominio de frecuencia 391 para cada cuadro y para cada contenedor (tesela de frecuencia). La información de dominio de frecuencia 391 puede proporcionarse a un bloque de análisis de difusión 392a y/o a un bloque de análisis de dirección 392b, que pueden ser aquéllos mostrados en la figura 3. El bloque de análisis de difusión 392a y/o el bloque de análisis de dirección 392b, puede proporcionar la información de difusión 314a y/o la información de dirección 314b. Éstas pueden proporcionarse para cada primer cuadro 306 (346) y para cada segundo cuadro 308 (348). Complejamente, la información proporcionada por el bloque 392a y 392b, se considera los parámetros de campo sonoro 314, que abarcan tanto los primeros parámetros de campo sonoro 316 (parámetros espaciales activos), como los segundos parámetros de campo sonoro 318 (parámetros espaciales inactivos). Los parámetros espaciales activos 316 pueden proporcionarse a un codificador de los metadatos espaciales activos 396, y los parámetros espaciales inactivos 318 pueden proporcionarse a un codificador de los metadatos espaciales inactivos 398. El resultado son la primera y segunda representaciones de parámetros de campo sonoro (316, 318, indicado complejamente con 314), que pueden codificarse en la secuencia de bits 304 (por ejemplo, a través del formador de señal de codificador 370), y almacenarse para reproducirse posteriormente por un decodificador. Si el codificador de los metadatos espaciales activos 396 o los parámetros espaciales inactivos 318 van a codificar un cuadro, esto puede controlarse mediante un control tal como el control 321 en la figura 3 (el desviador 322 no se muestra en la figura 2), por ejemplo, a través de la clasificación operada por el detector de actividad. (Deberá

notarse que los codificadores 396, 398, también pueden realizar una cuantificación, en algunos ejemplos).

La figura 5 muestra otro ejemplo del posible generador de parámetros de campo sonoro 310, que puede ser alterno a aquél de la figura 4, y que también puede implementarse en los ejemplos de las figuras 2 y 3. En este ejemplo, la señal de audio de entrada 302 ya está en formato MASA, en el cual los parámetros espaciales ya son parte de la señal de audio de entrada 302 (por ejemplo, como metadatos espaciales), por ejemplo, para cada contenedor de frecuencia de una pluralidad de contenedores de frecuencia. En consecuencia, no hay necesidad de tener un bloque de análisis de difusión y/o un bloque direccional, sino que pueden sustituirse por un lector MASA 390M. El lector MASA 390M puede leer los campos de datos específicos en la señal de audio 302, que ya contiene la información tal como el/los parámetro(s) espacial(es) activo(s) 316 y el/los parámetro(s) espacial(es) inactivo(s) 318 (de acuerdo con el hecho de que el cuadro de la señal 302 es un primer cuadro 306 o un segundo cuadro 308). Los ejemplos de los parámetros que pueden codificarse en la señal 302 (y que pueden leerse por el lector MASA 390M), pueden incluir al menos uno de una dirección, una relación de la energía, una coherencia del entorno, una coherencia de la dispersión, y así sucesivamente. Corriente abajo del lector MASA 390M, puede proporcionarse un codificador de los metadatos espaciales activos 396 (por ejemplo, como el de la figura 4) y un codificador de los metadatos espaciales inactivos 398 (por ejemplo, como el de la figura 4), para emitir la primera representación de parámetros de campo sonoro 316 y la segunda representación de parámetros de campo sonoro 318, respectivamente. Si la señal de audio de entrada 302 es una señal MASA, entonces el detector de actividad 320 puede implementarse como un elemento que lee un campo de datos determinado en la señal MASA de entrada 302, y lo clasifica como un cuadro activo 306 o un cuadro inactivo 308, basándose en el valor codificado en el campo de datos. El ejemplo de la figura 5 puede generalizarse para una señal de audio 302, que ya tiene codificada en la misma, la información espacial que puede ser codificada como un parámetro espacial activo 316 o un parámetro espacial inactivo 318.

Las realizaciones de la presente invención se aplican en un sistema de codificación del audio espacial, por ejemplo, ilustrado en la figura 2, en donde se describe un codificador y un decodificador del audio espacial basado en DirAC. Una discusión de los mismos sigue a continuación.

El codificador 300 puede analizar usualmente la escena de audio espacial en el formato B. De manera alterna, el análisis DirAC puede ajustarse para analizar diferentes formatos de audio como objetos de audio o señales con múltiples canales, o la combinación de cualesquier formatos de audio espacial.

El análisis DirAC (por ejemplo, realizado en cualquiera de las etapas 392a, 392b), puede extraer una representación paramétrica de la escena de audio de entrada 302 (señal de entrada). Una dirección de la llegada (DOA) 314b y/o una difusión 314a medida por unidad de tiempo-frecuencia, forman el/los parámetro(s) 316, 318. El análisis DirAC (por ejemplo, realizado en cualquiera de las etapas 392a, 392b), puede ser seguido por un codificador de los metadatos espaciales (por ejemplo, 396 y/o 398), que puede cuantificar y/o codificar los parámetros DirAC para obtener una representación paramétrica de baja tasa de bits (en las figuras, las representaciones paramétricas de baja tasa de bits 316, 318, se indican con los mismos números de referencia de las representaciones paramétricas corriente arriba de los codificadores de los metadatos espaciales 396 y/o 398).

Junto con los parámetros 316 y/o 318, una señal de mezcla descendente 324 (326) derivada de la(s) diferente(s) fuente(s) (por ejemplo, diferentes micrófonos) o señal(es) de entrada de audio (por ejemplo, diferentes componentes de una señal con múltiples canales) 302, puede codificarse (por ejemplo, para la transmisión y/o para el almacenamiento), mediante un codificador principal de audio convencional. En la realización preferida, un codificador de audio EVS (por ejemplo, 330, figura 2), puede preferirse para codificar la señal de mezcla descendente 324 (326, 328), pero las realizaciones de la invención no están limitadas a este codificador principal, y pueden aplicarse a cualquier codificador principal de audio. La señal de mezcla descendente 324 (326, 328) puede consistir, por ejemplo, de diferentes canales, también llamados canales de transporte: la señal 324 puede, por ejemplo, ser o comprender las señales de cuatro coeficientes que componen una señal en formato B, un par estéreo o un mezclado descendente monofónico, dependiendo de la tasa de bits seleccionada. Los parámetros espaciales codificados 328 y la secuencia de bits del audio codificado 326, pueden multiplexarse antes de transmitirse en el canal de comunicación (o almacenarse).

En el decodificador (véase a continuación), los canales de transporte 344 se decodifican por un decodificador principal, mientras que los metadatos DirAC (por ejemplo, los parámetros espaciales 316, 318), pueden decodificarse primero antes de ser transportados con los canales de transporte decodificados a la síntesis DirAC. La síntesis DirAC utiliza los metadatos decodificados para controlar la reproducción de la corriente del sonido directo y su mezcla con la corriente del sonido difuso. El campo sonoro reproducido puede reproducirse en una instalación de altavoces arbitraria o puede generarse en un formato Ambisonics (HOA/FOA) con un orden arbitrario.

Estimación del parámetro DirAC

Se explica aquí, una técnica no limitante para estimar los parámetros espaciales 316, 318 (por ejemplo, difusión 314a, dirección 314b). Se proporciona el ejemplo del formato B.

En cada banda de frecuencia (por ejemplo, obtenida del análisis de banco de filtros 390), puede estimarse la dirección de la llegada 314a del sonido, junto con la difusión 314b del sonido. A partir del análisis de tiempo-frecuencia de los componentes de formato B de entrada $w^i(n), x^i(n), y^i(n), z^i(n)$, los vectores de presión y velocidad se pueden determinar de la siguiente manera:

$$P^i(n, k) = W^i(n, k)$$

$$U^i(n, k) = X^i(n, k)\mathbf{e}_x + Y^i(n, k)\mathbf{e}_y + Z^i(n, k)\mathbf{e}_z$$

10

en donde i es el índice de la entrada 302 y, k y n índices de tiempo y frecuencia de la tesela de tiempo-frecuencia, y $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ representan los vectores unitarios cartesianos. $P(n, k)$ y $U(n, k)$ pueden necesariamente, en algunos ejemplos, calcular los parámetros DirAC (316, 318), a saber, la DOA 314a y la difusión 314a a través de, por ejemplo, el cálculo del vector de la intensidad:

15

$$I(k, n) = \frac{1}{2} \Re\{P(k, n) \cdot \overline{U(n, k)}\},$$

en donde (\cdot) denota conjugación compleja. La difusión de campo sonoro combinado se proporciona por:

$$\psi(k, n) = 1 - \frac{\|E\{I(k, n)\}\|}{cE\{E(k, n)\}}$$

20

en donde $E\{\cdot\}$ denota el operador de la promediación temporal, $E(k, n)$ es la velocidad del sonido y la energía de campo sonoro dada por:

$$E(n, k) = \frac{\rho_0}{4} \|U(n, k)\|^2 + \frac{1}{\rho_0 c^2} |P(n, k)|^2$$

25

La difusión de campo sonoro se define como la relación entre la intensidad del sonido y la densidad de la energía, que tiene valores de entre 0 y 1.

La dirección de llegada (DOA) se expresa por medio de la *dirección* del vector unitario (n, k) , definida como

$$\text{dirección}(n, k) = \frac{I(n, k)}{\|I(n, k)\|}$$

30

La dirección de llegada 314b se puede determinar mediante un análisis energético (por ejemplo, en 392b) de la señal de entrada de formato B 302 y se puede definir como la dirección opuesta del vector de intensidad. La dirección se define en las coordenadas cartesianas, pero pueden, por ejemplo, transformarse fácilmente a coordenadas esféricas definidas por un radio unitario, el ángulo del azimut y el ángulo de la elevación.

40

En el caso de la transmisión, los parámetros 314a, 314b (316, 318) necesitan transmitirse al lado del receptor (por ejemplo, lado del decodificador), mediante una secuencia de bits (por ejemplo, 304). Para una transmisión más robusta sobre una red con capacidad limitada, una secuencia de bits de baja tasa de bits es preferida, o incluso necesaria, lo que puede lograrse diseñando un esquema de codificación eficiente para los parámetros DirAC 314a, 314b (316, 318).

45

Pueden emplearse, por ejemplo, técnicas tales como agrupación de la banda de frecuencia promediando los parámetros en diferentes bandas y/o unidades de tiempo de frecuencia, predicción, cuantificación y codificación por entropía. En el decodificador, los parámetros transmitidos pueden decodificarse para cada unidad del tiempo/frecuencia (k, n) , en caso de que no ocurra un error en la red. Sin embargo, si las condiciones de la red no son suficientemente buenas para asegurar una transmisión del paquete apropiada, un paquete puede perderse durante la transmisión. Las realizaciones de la presente invención tienen el objeto de proporcionar una solución para el último caso.

Decodificador

50

La figura 6 muestra un ejemplo de aparato decodificador 200 útil para entender la invención. Puede ser un aparato para procesar una escena de audio codificada (304) que comprende, en un primer cuadro (346), una primera representación de parámetros de campo sonoro (316), y una señal de audio codificada (346), en donde un segundo cuadro (348) es un cuadro inactivo. El aparato decodificador 200 puede comprender al menos uno de:

55

un detector de actividad (2200) para detectar que el segundo cuadro (348) es el cuadro inactivo, y para proporcionar

una descripción paramétrica (328) para el segundo cuadro (308);

un sintetizador de la señal sintética (210) para sintetizar una señal de audio sintética (228) para el segundo cuadro (308), utilizando la descripción paramétrica (348) para el segundo cuadro (308);

5 un decodificador de audio (230) para decodificar la señal de audio codificada (346) para el primer cuadro (306); y

un renderizador espacial (240), para renderizar espacialmente la señal de audio (202) para el primer cuadro (306), utilizando la primera representación de parámetros de campo sonoro (316), y utilizando la señal de audio sintética (228) para el segundo cuadro (308).

10 En particular, el detector de actividad (2200) puede ejercer un comando 221', que puede determinar si el cuadro de entrada se clasifica como un cuadro activo 346 o un cuadro inactivo 348. El detector de actividad 2200 puede determinar la clasificación del cuadro de entrada, por ejemplo, de una información 221 que es ya sea señalizada, o 15 determinada de la longitud del cuadro obtenido.

20 El sintetizador de señal sintética (210) puede, por ejemplo, generar el ruido 228, por ejemplo, utilizando la información (por ejemplo, la información paramétrica) obtenida de la representación paramétrica 348. El renderizador espacial 220 puede 25 generar la señal de salida 202, de tal manera que los cuadros inactivos 228 (obtenidos de los cuadros codificados 348), se procesan a través del/los parámetro(s) espacial(es) inactivo(s) 318, para lograr que un oyente humano tenga una impresión espacial en 3D de la proveniencia del sonido.

Nótese que en la figura 6, los números 314, 316, 318, 344, 346, 348 son los mismos que los números de la figura 3, puesto que corresponden, por ser obtenidos de la secuencia de bits 304. No obstante, puede ser que estén presentes 25 algunas ligeras diferencias (por ejemplo, debido a la cuantificación).

La figura 6 también muestra un control 221' que puede controlar un desviador 224', de manera que puede 30 seleccionarse la señal 226 (emitida por el sintetizador de la señal sintética 210) o la señal de audio 228 (emitida por el decodificador de audio 230), por ejemplo, a través de la clasificación operada por el detector de actividad 220.

Particularmente, la señal 224 (ya sea 226 o 228), puede ser todavía una señal de mezcla descendente, que puede proporcionarse al renderizador espacial 220, de manera que el renderizador espacial genera la señal de salida 202 a 35 través de los parámetros espaciales activos o inactivos 314 (316, 318). En algunos ejemplos, la señal 224 (ya sea 226 o 228), puede mezclarse de forma ascendente, de modo que el número de canales de la señal 224 se incrementa con respecto a la versión codificada 344 (346, 348). En algunos ejemplos, a pesar de estar mezclada de forma ascendente, el número de canales de la señal 224 puede ser menor que el número de canales de la señal de salida 202.

Aquí a continuación, se proporcionan otros ejemplos del aparato decodificador 200. Las figuras 7-10 muestran los ejemplos del aparato decodificador 700, 800, 900, 1000, que pueden incorporar el aparato decodificador 200.

40 Aunque en las figuras 7-10 algunos elementos se muestran como internos al renderizador espacial 220, pueden no obstante, estar fuera del renderizador espacial 220 en algunos ejemplos. Por ejemplo, el sintetizador sintético 210 puede ser parcial o totalmente externo al renderizador espacial 220.

45 En esos ejemplos, un procesador de parámetros 275 (que puede ser interno o externo al renderizador espacial 220), puede incluirse. El procesador de parámetros 275 también puede considerarse como que está presente en el decodificador de la figura 6, a pesar de no ser mostrado.

El procesador de parámetros 275 de cualquiera de las figuras 7-10 puede incluir, por ejemplo, un decodificador del parámetro espacial inactivo 278, para proporcionar que los cuadros inactivos puedan ser los parámetros 318 (por ejemplo, obtenidos de la señalización en la secuencia de bits 304), y/o un bloque 279 ("recuperar los parámetros espaciales en el decodificador de los cuadros no transmitidos"), que proporciona los parámetros espaciales inactivos que no se leen en la secuencia de bits 304, pero que se obtienen (por ejemplo, recuperados, reconstruidos, extrapolados, inferidos, etc.), por ejemplo, mediante extrapolación o se generan de manera sintética.

55 Por lo tanto, la segunda representación de parámetros de campo sonoro también puede ser un parámetro generado 219, que no estaba presente en la secuencia de bits 304. Como se explicará posteriormente, los parámetros espaciales 219 recuperados (reconstruidos, extrapolados, inferidos, etc.), pueden obtenerse, por ejemplo, a través de una "estrategia de retención", a una "extrapolación de estrategia de dirección" y/o a través de un "tramo (dithering) de dirección" (véase a continuación). El procesador de parámetros 275 puede, por lo tanto, extrapolar u obtener de 60 cualquier manera los parámetros espaciales 219 de los cuadros previos. Como puede observarse en las figuras 6-9, un conmutador 275' puede seleccionar entre los parámetros espaciales inactivos 318 asignados en la secuencia de bits 304 y los parámetros espaciales recuperados 219. Como se explicó anteriormente, la codificación de los cuadros del silencio 348 (SID) (y también de los parámetros espaciales inactivos 318), se actualiza a una tasa de bits más baja que la codificación de los primeros cuadros 346: los parámetros espaciales inactivos 318 se actualizan con una

frecuencia más baja con respecto a los parámetros espaciales activos 316, y algunas estrategias se realizan por el procesador de parámetros 275 (1075), para recuperar los parámetros espaciales no señalizados 219 para los cuadros inactivos no transmitidos. En consecuencia, el conmutador 275' puede seleccionar entre los parámetros espaciales inactivos señalizados 318 y los parámetros espaciales inactivos no señalizados 219 (pero recuperados o reconstruidos de otra manera). En algunos casos, el procesador de parámetros 275' puede almacenar una o más representaciones de parámetros de campo sonoro 318 para varios cuadros que ocurren antes del segundo cuadro, o que ocurren en un tiempo posterior al segundo cuadro, para extrapolar (o interpolar) los parámetros de campo sonoro 219 para el segundo cuadro. En términos generales, el renderizador espacial 220 puede utilizar, para renderizar la señal de audio sintética 202 para el segundo cuadro 308, el uno o más parámetros de campo sonoro 318 para el segundo cuadro 219. De manera adicional o alterna, el procesador de parámetros 275 puede almacenar las representaciones de parámetros de campo sonoro 316 para los parámetros espaciales activos (mostrados en la figura 10), y sintetizar los parámetros de campo sonoro 219 para el segundo cuadro (cuadro inactivo), utilizando la primera representación de parámetros de campo sonoro 316 almacenada (cuadros activos) para generar el parámetro espacial recuperado 319. Como se muestra en la figura 10 (pero también implementable en cualquiera de las figuras 6-9), también es posible incluir también un decodificador del parámetro espacial activo 276 del cual pueden obtenerse los parámetros espaciales activos 316 de la secuencia de bits 304. Esto puede realizar un tramo con las direcciones incluidas en las al menos dos representaciones de parámetros de campo sonoro que ocurren en el tiempo antes o después del segundo cuadro (308), cuando se extrapolan o interpolan para determinar el uno o más parámetros de campo sonoro para el segundo cuadro (308).

El sintetizador de señal sintética 210 puede ser interno al renderizador espacial 220, o puede ser externo o, en algunos casos, puede tener una porción interna y una porción externa. El sintetizador sintético 210 puede operar en los canales de mezcla descendente de los canales de transporte 228 (que son menores que los canales de salida) (nótese aquí, que M es el número de los canales mezclados y N es el número de los canales de salida). El generador de señal sintética 210 (otro nombre para el sintetizador de señal sintética) puede generar, para el segundo cuadro, una pluralidad de señales de audio sintéticas del componente (en al menos uno de los canales de la señal de transporte o en al menos un componente individual del formato de audio de salida), para los componentes individuales relacionados con un formato externo del renderizador espacial, como la señal de audio sintética. En algunos casos, esto puede ser en los canales de la señal de mezcla descendente 228 y en algunos casos, puede estar en uno de los canales internos de renderización espacial.

La figura 7 muestra un ejemplo en el cual al menos K canales 228a obtenidos de la señal de audio sintética 228 (por ejemplo, en su versión 228b corriente abajo a un análisis de banco de filtros 720), pueden decorrelacionarse.

Esto se obtiene, por ejemplo, cuando el sintetizador sintético 210 genera la señal de audio sintética 228 en al menos uno de los M canales de la señal de audio sintética 228. Este procesamiento de correlación 730 puede aplicarse a la señal 228b (o al menos uno o alguno de sus componentes), corriente abajo del bloque de análisis de banco de filtros 720, de manera que al menos K canales (con $K \geq M$ y/o $K \leq N$, con N siendo el número de canales de salida), pueden obtenerse. Posteriormente, los K canales decorrelacionados 228a y/o los M canales de la señal 228b, pueden proporcionarse a un bloque 740 para generar las ganancias/matrizes del mezclado, que a través de los parámetros espaciales 218, 219 (véase anteriormente), pueden proporcionar una señal mezclada 742. La señal mezclada 742 puede someterse a un bloque de síntesis del banco de filtros 746, para obtener la señal de salida en los N canales de salida 202. Básicamente, los números de referencia 228a de la figura 7 pueden ser una señal de audio sintética del componente individual, que se decorrelaciona de la señal de audio sintética de componente individual 228b, de manera que el renderizador espacial (y el bloque 740), hace uso de una combinación del componente 228a y el componente 228b. La figura 8 muestra un ejemplo en el cual los canales 228 completos se generan en K canales.

Además, en la figura 7, el decorrelador 730 se aplica a los K canales decorrelacionados 228b corriente abajo del bloque de análisis de banco de filtros 720. Esto puede realizarse, por ejemplo, para el campo difuso. En algunos casos, M canales de la señal 228b corriente abajo del bloque de análisis de retroalimentación 720 pueden proporcionarse al bloque 744 que genera la ganancia/matrices del mezclado. Un método de covarianza puede utilizarse para reducir los problemas de los decorreladores 730, por ejemplo, escalando los canales 228b por un valor asociado con un valor complementario a la covarianza entre los diferentes canales.

La figura 8 muestra un ejemplo de un sintetizador de señal sintética 210, que está en el dominio de frecuencia. Puede utilizarse un método de covarianza para el sintetizador sintético 210 (810) de la figura 8. En particular, el sintetizador de audio sintético 210 (810) proporciona su salida 228c en K canales (con $K \geq M$), mientras que el canal de transporte 228 estaría en M canales.

La figura 9 muestra un ejemplo del decodificador 900 (realización del decodificador 200), que puede entenderse como que hace uso de una técnica híbrida del decodificador 800 de la figura 8 y el decodificador 700 de la figura 7. Como puede observarse aquí, el sintetizador de señal sintética 210 incluye una primera porción 210 (710) que genera una señal de audio sintética 228 en los M canales de la señal de mezcla descendente 228. La señal 228 puede introducirse a un bloque de análisis de banco de filtros 730, que puede proporcionar una salida 228b, en la cual bandas de filtro

plurales se distinguen unas de otras. En este momento, los canales 228b pueden decorrelacionarse para obtener la señal decorrelacionada 228a en K canales. Mientras tanto, la salida 228b del análisis de banco de filtros en M canales, se proporciona a un bloque 740 para generar las matrices de la ganancia de mezclado, que pueden proporcionar una versión mezclada de la señal mezclada 742. La señal mezclada 742 puede tener en cuenta los parámetros espaciales inactivos 318 y/o los parámetros espaciales recuperados (reconstruidos) para los cuadros inactivos 219. Nótese que la salida 228a del decorrelador 730 también puede agregarse, en un sumador 920, a una salida 228d de una segunda porción 810 del sintetizador de señal sintética 210, que proporciona una señal sintética 228d en K canales. La señal 228d puede sumarse, en el bloque de adición 920, a la señal decorrelacionada 228a para proporcionar la señal sumada 228e al bloque de mezclado 740. Por lo tanto, es posible renderizar la señal de salida final 202, utilizando una combinación de componente 228b y el componente 228e, que toma en cuenta a ambos de los componentes 228a decorrelacionados y los componentes 228d generados. Puede entenderse que los componentes 228b, 228a, 228d, 228e (están presentes) de las figuras 8 y 7, por ejemplo, son los componentes difusos y no difusos de la señal sintética 228. En particular, con referencia al decodificador 900 de la figura 9, básicamente las bandas de frecuencia baja de señal 228e pueden obtenerse del canal de transporte 710 (y se obtienen de 228a), y las bandas de frecuencia alta de señal 228e pueden generarse en el sintetizador 810 (y están en los canales 228d), su adición en el sumador 920 permite tener ambas en la señal 228e.

Particularmente, en las figuras 7-10 anteriores, no se muestra el decodificador del canal de transporte para los cuadros activos.

La figura 10 muestra un ejemplo del decodificador 1000 (realización del decodificador 200), en el cual se muestra tanto el decodificador de audio 230 (que proporciona los canales 226 decodificados) como el sintetizador de señal sintética 210 (aquí, considerado como que está dividido entre una primera porción externa 710 y una segunda porción interna 810). Se muestra un conmutador 224', que puede ser análogo a aquél de la figura 6 (por ejemplo, controlado por el control o comando 221', proporcionado por el detector de actividad 220). Básicamente, es posible seleccionar entre un modo en el cual la escena de audio decodificada 226 se proporciona al renderizador espacial 220, y otro modo en el cual se proporciona la señal de audio sintética 228. La señal mezclada 224 (226, 228) está en M canales, que en general, son menos que los N canales de salida de la señal de salida 202.

La señal 224 (226, 228) puede emitirse a un bloque del análisis de banco de filtros 720. La salida 228b del análisis de banco de filtros 720 (en una pluralidad de contenedores de frecuencia), puede introducirse a un bloque de adición de mezclado ascendente 750, que puede también introducirse por una señal 228d proporcionada por la segunda porción 810 del sintetizador de señal sintética 210. La salida 228f del bloque de adición de mezclado ascendente 750, puede introducirse al procesamiento del decorrelador 730. La salida 228a del procesamiento del decorrelador 730 puede proporcionarse, junto con la salida 228f del bloque de adición de mezclado ascendente 750, al bloque 740 para generar las ganancias y las matrices del mezclado. El bloque de adición de mezclado ascendente 750 puede, por ejemplo, incrementar el número de canales de M a K (y, en algunos casos, puede escalarlos, por ejemplo, mediante la multiplicación por coeficientes constantes), y puede sumar los K canales con los K canales 228d generados por el sintetizador de señal sintética 210 (por ejemplo, segunda porción interna 810). Con el fin de renderizar un primer cuadro (activo), el bloque de mezclado 740 puede considerar al menos uno de los parámetros espaciales activos 316 como proporcionados en la secuencia de bits 304, los parámetros espaciales recuperados (reconstruidos) 210 como extrapolados u obtenidos de otra manera (véase anteriormente).

En algunos ejemplos, la salida del bloque de análisis de banco de filtros 720 puede estar en M canales, pero pueden tomarse en consideración diferentes bandas de frecuencia. Para los primeros cuadros (y el conmutador 224' y el conmutador 222' siendo colocados como en la figura 10), la señal decodificada 226 (en al menos dos canales) puede proporcionarse al análisis de banco de filtros 720 y puede, por lo tanto, ponderarse en el bloque de adición de mezclado ascendente 750 a través de K canales de ruido 228d (canales de señal sintética), para obtener la señal 228f en K canales. Recuérdese que $K \geq M$ y puede comprender, por ejemplo, un canal difuso y un canal direccional. En particular, el canal difuso puede decorrelacionarse por el decorrelador 730 para obtener una señal decorrelacionada 228a. En consecuencia, la señal de audio decodificada 224 puede ponderarse (por ejemplo, en el bloque 750), con la señal de audio sintética 228d, que puede enmascarar la transición entre los cuadros activos e inactivos (primeros cuadros y segundos cuadros). A continuación, la segunda parte 810 del sintetizador de señal sintética 210, se utiliza no sólo para los cuadros activos, sino también para los cuadros inactivos.

La figura 11 muestra otro ejemplo del decodificador 200, que puede comprender un primer cuadro (346), una primera representación del parámetro de campo sonoro (316), y una señal de audio codificada (346), en donde un segundo cuadro (348) es un cuadro inactivo, el aparato comprende un detector de actividad (220) para detectar que el segundo cuadro (348) es el cuadro inactivo, y para proporcionar una descripción paramétrica (328) para el segundo cuadro (308); un sintetizador de señal sintética (210) para sintetizar una señal de audio sintética (228) para el segundo cuadro (308), utilizando la descripción paramétrica (348) para el segundo cuadro (308); un decodificador de audio (230) para decodificar la señal de audio codificada (346) para el primer cuadro (306); y un renderizador espacial (240) para renderizar espacialmente la señal de audio (202) para el primer cuadro (306), utilizando la primera representación de parámetros de campo sonoro (316), y utilizando la señal de audio sintética (228) para el segundo cuadro (308), o un

transcodificador para generar un formato de salida asistido por metadatos, que comprende la señal de audio (346) para el primer cuadro (306), la primera representación de parámetros de campo sonoro (316) para el primer cuadro (306), la señal de audio sintética (228) para el segundo cuadro (308), y una segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308).

- 5 Con referencia al sintetizador de señal sintética 210 en los ejemplos anteriores, como se explicó anteriormente, puede comprender (o incluso ser) un generador de ruido (por ejemplo, un generador de ruido de confort). En los ejemplos, el generador de señal sintética (210) puede comprender un generador de ruido y la primera señal de audio sintética del componente individual se genera por un primer muestreo del generador de ruido, y la segunda señal de audio sintética del componente individual se genera por un segundo muestreo del generador de ruido, en donde el segundo muestreo es diferente del primer muestreo.

10 De manera adicional o alterna, el generador de ruido comprende una tabla de ruido, y en donde la primera señal de audio sintética del componente individual se genera tomando una primera porción de la tabla de ruido, y en donde la segunda señal de audio sintética del componente individual se genera tomando una segunda porción de la tabla de ruido, en donde la segunda porción de la tabla de ruido es diferente de la primera porción de la tabla de ruido.

15 En los ejemplos, el generador de ruido comprende un generador de pseudo ruido, y en donde la primera señal de audio sintética del componente individual se genera utilizando una primera semilla para el generador de pseudo ruido, y en donde la segunda señal de audio sintética del componente individual se genera utilizando una segunda semilla para el generador de pseudo ruido.

20 En términos generales, el renderizador espacial 220, en los ejemplos de las figuras 6, 7, 9, 10 y 11, puede operar en un primer modo para el primer cuadro (306), utilizando un mezclado de una señal directa y una señal difusa generada por un decorrelador (730) de la señal directa bajo un control de la primera representación de parámetros de campo sonoro (316), y en un segundo modo para el segundo cuadro (308), utilizando un mezclado de una primera señal sintética de componente, y la segunda señal sintética de componente, en donde la primera y segunda señales sintéticas de componente se generan por el sintetizador de la señal sintética (210), mediante diferentes ejecuciones de un proceso de ruido o un proceso de pseudo ruido.

25 30 Como se explicó anteriormente, el renderizador espacial (220) puede configurarse para controlar el mezclado (740) en el segundo modo, mediante un parámetro de difusión, un parámetro de distribución de energía, o un parámetro de coherencia derivado del segundo cuadro (308) por un procesador de parámetros.

35 40 45 Los ejemplos anteriores también consideran un método para generar una escena de audio codificada de una señal de audio que tiene un primer cuadro (306) y un segundo cuadro (308), que comprende: determinar una primera representación del parámetros de campo sonoro (316) para el primer cuadro (306) de la señal de audio en el primer cuadro (306), y una segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308) de la señal de audio en el segundo cuadro (308); analizar la señal de audio para determinar, dependiendo de la señal de audio, que el primer cuadro (306) es un cuadro activo y el segundo cuadro (308) es un cuadro inactivo; generar una señal de audio codificada para el primer cuadro (306), que es el cuadro activo y generar una descripción paramétrica (348) para el segundo cuadro (308), que es el cuadro inactivo; y componer la escena de audio codificada juntando la primera representación de parámetros de campo sonoro (316) para el primer cuadro (306), la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308), la señal de audio codificada para el primer cuadro (306), y la descripción paramétrica (348) para el segundo cuadro (308).

50 55 Los ejemplos anteriores también consideran un método para procesar una escena de audio codificada, que comprende, en un primer cuadro (306), una primera representación de parámetros de campo sonoro (316) y una señal de audio codificada, en donde un segundo cuadro (308) es un cuadro inactivo, el método comprende: detectar que el segundo cuadro (308) es el cuadro inactivo, y proporcionar una descripción paramétrica (348) para el segundo cuadro (308); sintetizar una señal de audio sintética (228) para el segundo cuadro (308), utilizando la descripción paramétrica (348) para el segundo cuadro (308); decodificar la señal de audio codificada para el primer cuadro (306); y renderizar espacialmente la señal de audio para el primer cuadro (306), utilizando la primera representación de parámetros de campo sonoro (316), y utilizando la señal de audio sintética (228) para el segundo cuadro (308), o generar un formato de salida asistido por metadatos, que comprende la señal de audio para el primer cuadro (306), la primera representación del parámetro de campo sonoro (316) para el primer cuadro (306), la señal de audio sintética (228) para el segundo cuadro (308), y una segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308).

60 También se proporciona una escena de audio codificada (304) que comprende: una primera representación de parámetros de campo sonoro (316) para un primer cuadro (306); una segunda representación de parámetros de campo sonoro (318) para un segundo cuadro (308); una señal de audio codificada para el primer cuadro (306); y una descripción paramétrica (348) para el segundo cuadro (308).

En los ejemplos anteriores, puede ser que los parámetros espaciales 316 y/o 318 se transmitan para cada banda de frecuencia (subbanda).

De acuerdo con algunos ejemplos, esta descripción paramétrica del silencio 348 puede contener este parámetro parcial 318, que puede, por lo tanto, ser parte del SID 348.

El parámetro espacial 318 para los cuadros inactivos puede ser válido para cada subbanda de frecuencia (o banda o frecuencia).

10 Los parámetros espaciales 316 y/o 318 discutidos anteriormente, transmitidos o codificados, durante la fase activa 346 y en el SID 348, puede tener diferente resolución de frecuencia y de manera adicional o alterna, los parámetros espaciales 316 y/o 318 discutidos anteriormente, transmitidos o codificados, durante la fase activa 346 y en el SID 348, pueden tener diferente resolución del tiempo y de manera adicional o alterna, los parámetros espaciales 316 y/o 318 discutidos anteriormente, transmitidos o codificados, durante la fase activa 346 y en el SID 348, pueden tener 15 diferente resolución de cuantificación.

Nótese que el dispositivo de decodificación y el dispositivo de codificación pueden ser dispositivos como CELP o DCX o módulos de extensión del ancho de banda.

20 También es posible hacer uso de un esquema de codificación basado en MDCT (transformada discreta del coseno modificada).

En los ejemplos presentes del aparato decodificador 200 (en cualquiera de sus realizaciones, por ejemplo, aquéllas de las figuras 6-11), es posible sustituir el decodificador de audio 230 y el renderizador espacial 240, con un 25 transcodificador para generar un formato de salida asistido por metadatos, que comprende la señal de audio para el primer cuadro, la primera representación de parámetros de campo sonoro para el primer cuadro, la señal de audio sintética para el segundo cuadro, y una segunda representación de parámetros de campo sonoro para el segundo cuadro.

30 Discusión

Las realizaciones de la presente invención proponen una forma para extender el DTX a la codificación paramétrica de audio espacial. Se propone, por lo tanto, aplicar un DTX/CNG convencional en los canales de mezclado descendente/transporte (por ejemplo, 324, 224), y extenderlo con los parámetros espaciales (llamados posteriormente 35 el SID espacial), por ejemplo, 316, 318 y una renderización espacial en los cuadros inactivos (por ejemplo, 308, 328, 348, 228), en el lado del decodificador. Para restituir la imagen espacial de los cuadros inactivos (por ejemplo, 308, 328, 348, 228), el SID del canal de transporte 326, 226 se enmienda con algunos parámetros espaciales (SID espaciales) 319 (o 219), diseñados especialmente y relevantes para los ruidos de fondo inmersivos. Las realizaciones 40 de la presente invención (discutidas a continuación y/o anteriormente), cubren al menos dos aspectos:

40 • Extender el SID del canal de transporte para la renderización espacial. Para esto, el descriptor se enmienda con los parámetros espaciales 318, por ejemplo, derivados del paradigma DirAC o el formato MASA. Al menos uno de los parámetros 318 como la difusión 314a, y/o la dirección de la llegada 314b, y/o las coherencias intercanal/entorno, y/o las relaciones de energía, pueden transmitirse junto con el SID 328 del canal de transporte (348). En ciertos casos y bajo ciertas suposiciones, algunos de los parámetros 318 podrían descartarse. Por ejemplo, si suponemos que el ruido de fondo es completamente difuso, podemos descartar la transmisión de las direcciones 314b, que entonces son irrelevantes.

50 • Espacializar en el lado del receptor los cuadros inactivos renderizando el canal de transporte CNG en el espacio: El principio de síntesis de DirAC o una de sus derivadas puede emplearse guiado por los parámetros espaciales eventualmente transmitidos 318 dentro del descriptor Sid espacial de ruido de fondo. Existen al menos dos opciones, que pueden incluso combinarse: la generación de ruido de confort del canal de transporte puede generarse únicamente para los canales de transporte 228 (esto es el caso de la figura 7, en donde el ruido de confort 228 se genera por el sintetizador de señal sintética 710); o la CNG del canal de transporte puede también generarse para los canales de 55 transporte y también para los canales adicionales utilizados en el renderizador para el mezclado ascendente (este es el caso de la figura 9, en donde algo de ruido de confort 228 se genera por la primera porción del sintetizador de señal sintética 710, pero algún otro ruido de confort 228d se genera por la segunda porción del sintetizador de señal sintética 810). En el último caso, la segunda porción de CNG 710, por ejemplo, el muestreo de un ruido aleatorio 228d con diferente semilla, puede decorrelacionar de manera automática los canales 228d generados y reducir al mínimo el empleo de los decorreladores 730, que podrían ser las fuentes de artefactos típicos. Además, la CNG también puede emplearse (como se muestra en la figura 10), en los cuadros activos pero, en algunos ejemplos, con resistencia reducida para facilitar la transición entre las fases activa e inactiva (cuadros), y también para enmascarar los artefactos 60 eventuales del codificador de canal de transporte y el paradigma DirAC paramétrico.

La figura 3 describe visión general de las realizaciones del aparato codificador 300. En el lado del codificador, la señal puede analizarse mediante el análisis DirAC. La DirAC puede analizar las señales como el formato B o Ambisonics de primer orden (FOA). Sin embargo, también es posible extender el principio a Ambisonics de orden superior (HOA), e incluso a señales con múltiples canales asociadas con una instalación del altavoz dada como 5.1, o 7.1 o 7.1 + 4, como se propone en [10]. El formato de entrada 302 también puede ser los canales de audio individuales que representan uno, o varios objetos de audio diferentes localizados en el espacio, mediante la información incluida en los metadatos asociados. Alternativamente, el formato de entrada 302 puede ser audio espacial asociado con metadatos (MASA). En este caso, los parámetros espaciales y los canales de transporte se transmiten directamente al aparato codificador 300. El análisis de la escena de audio (por ejemplo, como se muestra en la figura 5), puede saltarse entonces, y sólo tiene que realizarse una recuantificación y remuestreo eventual del parámetro espacial, para el conjunto inactivo de los parámetros espaciales 318 o para ambos conjuntos activo e inactivo de los parámetros espaciales 316, 318.

El análisis de la escena de audio puede hacerse para ambos cuadros activos e inactivos 306, 308 y producir dos conjuntos de parámetros espaciales 316, 318. Un primer conjunto 316 en el caso del cuadro activo 308, y otro (318) en el caso del cuadro inactivo 308. Es posible no tener parámetros espaciales inactivos, pero en la realización preferida de la invención, los parámetros espaciales inactivos 318 son menos y/o son cuantificados más gruesos que los parámetros espaciales activos 316. Después de eso, pueden estar disponibles dos versiones de los parámetros espaciales (también llamados metadatos DirAC). De manera importante, las realizaciones de la presente invención pueden dirigirse principalmente a las representaciones espaciales de la escena de audio desde la perspectiva del oyente. Por lo tanto, se consideran los parámetros espaciales, como los parámetros DirAC 318, 316 que incluyen una o varias direcciones junto con un factor de difusión o relación de energía eventual. A diferencia de los parámetros intercanal, estos parámetros espaciales desde la perspectiva del oyente, tienen la gran ventaja de ser indiferentes de la captura del sonido y el sistema de reproducción. Esta parametrización no es específica para algún micrófono matriz o instalación de altavoces particular.

El detector de actividad de voz (o más en general, un detector de actividad) 320, puede aplicarse entonces en la señal de entrada 302 y/o los canales de transporte 326 producidos por el analizador de escena de audio. Los canales de transporte son menos que el número de canales de entrada; usualmente un monomezclado, un mezclado estéreo, un formato A, o una señal Ambisonics de primer orden. Basándose en la decisión del VAD, el cuadro actual bajo el proceso se define como activo (306, 326) o inactivo (308, 328). En el caso de los cuadros activos (306, 326), se realiza una codificación convencional del habla o el audio de los canales de transporte. Los datos del código resultante se combinan entonces con los parámetros espaciales activos 316. En el caso de los cuadros inactivos (308, 328), una descripción de información de silencio 328 de los canales de transporte 324, se produce de manera episódica, usualmente a intervalos regulares del cuadro durante la fase inactiva, por ejemplo, cada 8 cuadros activos (306, 326, 346). El SID de canal de transporte (328, 348) puede enmendarse entonces en el multiplexor (formador de la señal codificada) 370, con los parámetros espaciales inactivos. En el caso de que los parámetros espaciales inactivos 318 sean nulos, únicamente el SID de canal de transporte 348 se transmite entonces. El SID general puede usualmente ser una descripción con una tasa de bits muy baja, que es, por ejemplo, tan baja como 2,4 o 4,25 kbps. La tasa de bits promedio es todavía más reducida en la fase inactiva, puesto que la mayoría del tiempo no se hace una transmisión y no se envían datos.

En la realización preferida de la invención, el SID de canal de transporte 348 tiene un tamaño de 2,4 kbps, y el SID general incluyendo los parámetros espaciales, tiene un tamaño de 4,25 kbps. El cálculo de los parámetros espaciales inactivos se describe en la figura 4 para DirAC, que tiene como entrada una señal con múltiples canales como el FOA, que podría derivarse directamente de un Ambisonics de orden superior (HOA), en la figura 5 para el formato de entrada MASA. Como se describió anteriormente, los parámetros espaciales inactivos 318 pueden derivarse en paralelo con los parámetros espaciales activos 316, promediando y/o recuantificando los parámetros espaciales activos 318 ya codificados. En caso de una señal con múltiples canales como FOA como un formato de entrada 302, puede realizarse un análisis de banco de filtros de señal con múltiples canales 302, antes de calcular los parámetros espaciales, la dirección y la difusión, para cada tiempo y tesela de frecuencia. Los codificadores de metadatos 396, 398 podrían promediar los parámetros 316, 318 sobre diferentes bandas de frecuencia y/o segmentos temporales antes de aplicar un cuantificador y una codificación de parámetros cuantificados. El codificador de metadatos espaciales inactivos puede además, heredar algunos de los parámetros cuantificados derivados en el codificador de metadatos espaciales activos, para utilizarlos directamente en los parámetros espaciales inactivos o para recuantificarlos. En el caso del formato MASA (por ejemplo, figura 5), primero, los metadatos de entrada pueden leerse y proporcionarse a los codificadores de los metadatos 396, 398 a una resolución del tiempo-frecuencia y profundidad del bit dadas. Los codificadores de metadatos 396, 398 los procesarán entonces adicionalmente, convirtiendo eventualmente algunos parámetros, adaptando su resolución (es decir, disminuyendo la resolución, por ejemplo, promediándolos), y recuantificándolos antes de codificarlos mediante un esquema de codificación por entropía, por ejemplo.

En el lado del decodificador como se describe, por ejemplo, en la figura 6, la información del VAD 221 (por ejemplo, si el cuadro se clasifica como activo o inactivo), se recupera primero, ya sea detectando el tamaño del paquete transmitido (por ejemplo, cuadro), o detectando la no transmisión de un paquete. En los cuadros activos 346, el

decodificador se ejecuta en el modo activo y la carga del codificador de canal de transporte se decodifica, así como los parámetros espaciales activos. El renderizador espacial 220 (síntesis DirAC) mezcla/espacializa entonces los canales de transporte decodificados, utilizando los parámetros espaciales decodificados 316, 318, en el formato espacial de salida. En los cuadros inactivos, un ruido de confort puede generarse en los canales de transporte por la porción de CNG del canal de transporte 810 (por ejemplo, en la figura 10). La CNG se guía por el SID de canal de transporte para ajustar usualmente la energía y la forma espectral (a través de, por ejemplo, los factores escalares aplicados en el dominio de frecuencia o los coeficientes de codificación predictiva lineal aplicados a través de un filtro de síntesis de dominio de tiempo). Los ruidos de confort 228d, 228a, etc., se renderizan/espacializan entonces en el renderizador espacial (síntesis DirAC) 740, guiado esta vez por los parámetros espaciales inactivos 318. El formato espacial de salida 202 puede ser una señal binaural (2 canales), múltiples canales para una instalación del altavoz dada, o una señal con múltiples canales en un formato Ambisonic. En una realización alterna, el formato de salida puede ser audio espacial asistido por metadatos (MASA), lo que significa que los canales de transporte decodificados o los ruidos de confort de canal de transporte se emiten directamente junto con los parámetros espaciales activos o inactivos, respectivamente, para la renderización por un dispositivo externo.

15

Codificación y decodificación de los parámetros espaciales inactivos

Los parámetros espaciales inactivos 318 pueden consistir de una de múltiples direcciones en las bandas de frecuencia y las relaciones de energía asociadas en las bandas de frecuencia, que corresponden a la relación del componente de una dirección con respecto a la energía total. En el caso de una dirección, como en una realización preferida, la relación de la energía puede reemplazarse por la difusión, que es complementaria a la relación de energía, y a continuación seguir el conjunto de parámetros DirAC original. Puesto que se espera que los componentes direccionales en general sean menos relevantes que la parte difusa en los cuadros inactivos, también pueden transmitirse en menos bits, utilizando un esquema de cuantificación más grueso, como en los cuadros activos y/o promediando la dirección con el tiempo o la frecuencia, para obtener una resolución más gruesa del tiempo y/o frecuencia. En una realización preferida, la dirección puede enviarse cada 20 ms en lugar de 5 ms para los cuadros activos, pero utilizando la misma resolución de frecuencia de 5 bandas no uniformes.

En una realización preferida, la difusión 314a puede transmitirse con el mismo tiempo/frecuencia que en los cuadros activos, pero en menos bits, forzando un índice de cuantificación mínimo. Por ejemplo, si la difusión 314a se cuantifica en 4 bits en los cuadros activos, se transmite entonces únicamente en 2 bits, evitando la transmisión de los índices originales de 0 a 3. El índice decodificado se sumará entonces con una compensación de +4.

También es posible evitar completamente enviar la dirección 314b o de manera alterna evitar enviar la difusión 314a y reemplazarla en el decodificador mediante un valor por defecto o estimado, en algunos ejemplos.

Además, uno puede considerar transmitir una coherencia intercanal si los canales de entrada corresponden a los canales colocados en el dominio espacial. Las diferencias del nivel intercanal también son una alternativa para las direcciones.

40

Es más relevante enviar una coherencia del entorno que se define como la relación de la energía difusa, que es coherente en el campo sonoro. Puede aprovecharse en el renderizador espacial (síntesis DirAC), por ejemplo, redistribuyendo la energía entre las señales directa y difusa. La energía de los componentes coherentes del entorno se elimina de la energía difusa para ser redistribuida a los componentes direccionales, que a continuación se panoramizarán de manera más uniforme en el espacio.

Naturalmente, cualesquier combinaciones de los parámetros listados previamente, podrían considerarse para los parámetros espaciales inactivos. También podría considerarse para propósitos de ahorro de bits, no enviar ningún parámetro en la fase inactiva.

50

Un pseudo código a modo de ejemplo del codificador de los metadatos espaciales inactivos, se proporciona a continuación:

```
bistream = inactive_spatial_metadata_encoder (
    azimuth, /* i: azimuth values from active spatial metadata encoder */
    elevation, /* i: elevation values from active spatial metadata encoder */
    diffuseness_index, /* i/o: diffuseness indices from active spatial metadata encoder */
    metadata_sid_bits /* i bits allocated to inactive spatial metadata (spatial SID) */

)
{
    /* Signalling 2D*/
    not_in_2D = 0;
    para ( b = start_band; b < nbands; b++ )
    {

```

```

    para ( m = 0; m < nblocks; m++ )
    {
        not_in_2D += elevation[b][m];
    }
5     }
     write_next_indice( bitstream, (not_in_2D > 0 ), 1 ); /*2D flag*/

/*Count required bits */
bits_dir = 0;
10    bits_diff = 0;
para ( b = start_band; b < nbands; b++ )
{
    diffuseness_index[b] = max( diffuseness_index[b], 4 );
    bits_diff += get_bits_diffuseness(diffuseness_index[b] - 4, DIRAC_DIFFUSE_LEVELS - 4);
15    si ( not_in_2D == 0 )
    {
        bits_dir += get_bits_azimuth(diffuseness_index[b]);
    }
    else
20    {
        bits_dir += get_bits_spherical(diffuseness_index[b]);
    }
}

25 /* Reduce bit demand by increasing diffuseness index*/
bits_delta = metadata_sid_bits - 1 - bits_diff - bits_dir;
while ( ( bits_delta < 0 ) && (not_in_2D > 0 ) )
{
30    para ( b = nbands - 1; b >= start_band && ( bits_delta < 0 ); b-- )
    {
        if ( diffuseness_index[b] < ( DIRAC_DIFFUSE_LEVELS - 1 ) )
        {
            bits_delta += get_bits_spherical(diffuseness_index[b]);
            diffuseness_index[b]++;
35            bits_delta -= get_bits_spherical(diffuseness_index[b]);
        }
    }
}

40 /*Write diffuseness indices*/
para ( b = start_band; b < nbands; b++ )
{
    Write_diffuseness(bitstream, diffuseness_index[b]- 4, DIRAC_DIFFUSE_LEVELS - 4);
}
45 /* Compute and Quantize an average direction per band*/
para ( b = start_band; b < nbands; b++ )
{
    set_zero( avg_direction_vector, 3 );
50    para ( m = 0; m < nblocks; m++ )
    {
        /*compute the average direction */
        azimuth_elevation_to_direction_vector(azimuth[b][m], elevation[b][m], direction_vector );
        v_add( avg_direction_vector, direction_vector, avg_direction_vector, 3 );
55    }
    direction_vector_to_azimuth_elevation( avg_direction_vector, &avg_azimuth[b], &avg_elevation[b] );

    /* Quantize the average direction */
    if ( not_in_2D > 0 )
60    {
        Code_and_write_spherical_angles(bitsream, avg_elevation[b], avg_azimuth[b],
get_bits_spherical(diffuseness_index[b]));
    }
    else

```

```

    {
        Code_and_write_azimuth (bitsream, avg_azimuth[b],           get_bits_azimuth(diffuseness_index[b]));
    }
}
5
Para(i=0; i<delta_bits; i++)
{
    Write_next_bit ( bitstream, 0); /*fill bit with value 0*/
}
10
}

```

Un pseudo código a modo de ejemplo del decodificador de los metadatos espaciales inactivos se proporciona a continuación:

```

15 [diffuseness, azimuth, elevation] = inactive_spatial_metadata_decoder(bitstream)

    /* Read 2D signalling*/
    not_in_2D = read_next_bit(bitstream);

20    /* Decode diffuseness*/
    para ( b = start_band; b < nbands; b++ )
    {
        diffuseness_index[b] = read_diffuseness_index( bitstream, DIFFUSE_LEVELS - 4 ) + 4;
        diffuseness_avg = diffuseness_reconstructions[diffuseness_index[b]];
25        para ( m = 0; m < nblocks; m++ )
            diffuseness[b][m] = diffusenessavg;
    }

30    /* Decoder DOAs*/
    if (not_in_2D > 0)
    {
        para ( b = start_band; b < nbands; b++ )
        {
35            bits_spherical = get_bits_spherical(diffuseness_index[b]);
            spherical_index = Read_spherical_index( bitstream, bits_spherical);
            azimuth_avg = decode_azimuth(spherical_index, bits_spherical);
            elevation_avg = decode_elevation(spherical_index, bits_spherical);
            para ( m = 0; m < nblocks; m++ )
40            {
                elevation[b][m] *= 0.9f;
                elevation[b][m] += 0.1f * elevation_avg;
                azimuth[b][m] *= 0.9f;
                azimuth[b][m] += 0.1f * azimuth_avg;
45            }
        }
    }
    else
    {
50        para ( b = start_band; b < nbands; b++ )
        {
            bits_azimuth = get_bits_azimuth(diffuseness_index[b]);
            azimuth_index = Read_azimuth_index( bitstream, bits_azimuth);
            azimuth_avg = decode_azimuth(diffuseness_index, bits_azimuth);
55        para ( m = 0; m < nblocks; m++ )
            {
                elevation[b][m] *= 0.9f;
                azimuth[b][m] *= 0.9f;
                azimuth[b][m] += 0.1f * azimuth_avg;
60            }
        }
    }
}

```

Recuperación del parámetro espacial en caso de no haya transmisión en el lado del decodificador

En caso de SID durante la fase inactiva, los parámetros espaciales pueden decodificarse completa o parcialmente y a continuación utilizarse para la síntesis DirAC subsiguiente.

- 5 En el caso en que no haya transmisión de datos o si ningún parámetro espacial 318 se transmite junto con el canal de transporte 348, los parámetros espaciales 219 podrían necesitar restituirse. Esto puede lograrse generando sintéticamente los parámetros 219 que faltan (por ejemplo, las figuras 7-10), considerando los parámetros recibidos en el pasado (por ejemplo, 316 y/o 318). Una imagen espacial inestable puede percibirse como desagradable, especialmente con ruido de fondo considerado constante y que no se desarrolla rápidamente. Por otra parte, una 10 imagen espacial estrictamente constante, puede percibirse como no natural. Pueden aplicarse diferentes estrategias:

Estrategia de retención:

Generalmente es seguro considerar que la imagen espacial deba ser relativamente estable con el tiempo, lo que puede traducirse para los parámetros DirAC, es decir, la DOA y la difusión, que no cambian mucho entre los cuadros. Por esta razón, un enfoque simple pero efectivo es mantener, como los parámetros espaciales recuperados 219, los parámetros espaciales 316 y/o 318 recibidos al último. Esto es un enfoque muy robusto, al menos para la difusión, que tiene una característica a largo plazo. Sin embargo, para la dirección, deben contemplarse diferentes estrategias, como se lista a continuación.

- 20 **Extrapolación de la dirección:**

De manera adicional o alterna, puede contemplarse estimar la trayectoria de los eventos de sonido en la escena de audio y a continuación, tratar de extrapolar la trayectoria estimada. Es especialmente relevante si el evento de sonido está bien localizado en el espacio como una fuente puntual, que se refleja en el modelo DirAC mediante una difusión baja. La trayectoria estimada puede calcularse de las observaciones de las direcciones pasadas y ajustando una curva entre estos puntos, lo que puede desarrollar ya sea en interpolación o alisamiento o suavizado. También puede emplearse un análisis de regresión. La extrapolación del parámetro 219 puede realizarse entonces, evaluando la curva ajustada más allá del intervalo de los datos observados (por ejemplo, incluyendo los parámetros 316 y/o 318 previos). 30 Sin embargo, este enfoque podría resultar menos relevante para los cuadros inactivos 348, en donde el ruido de fondo es inútil y se espera que sea muy difuso.

Tramado de dirección:

- 35 Cuando el evento del sonido es más difuso, que es especialmente el caso para el ruido de fondo, las direcciones son menos relevantes y pueden considerarse como la ejecución de un proceso estocástico. El tramado puede ayudar entonces a hacer más natural y más agradable el campo sonoro renderizado, inyectando un ruido aleatorio a las direcciones previas, antes de utilizarlo para los cuadros no transmitidos. El ruido inyectado y su varianza puede ser una función de difusión. Por ejemplo, las varianzas σ_{azi} y σ_{ele} de los ruidos inyectados en el azimuth y la elevación, 40 pueden seguir una función de modelo simple de difusión Ψ , como sigue:

$$\sigma_{azi} = 65\Psi^{3.5} + \sigma_{ele}$$

$$\sigma_{ele} = 33.25\Psi + 1.25$$

45 ***Generación de ruido de confort y espacialización (lado de decodificador)***

Algunos ejemplos, proporcionados anteriormente, se discuten ahora.

- 50 En una primera realización, el generador de ruido de confort 210 (710) se hace en el decodificador principal, como se describe en la figura 7. Los ruidos de confort resultantes se inyectan en los canales de transporte y a continuación se espacializan en la síntesis DirAC con la ayuda de los parámetros espaciales inactivos 318 transmitidos, o en el caso de que no haya transmisión, utilizando los parámetros espaciales 219 deducidos como se describió previamente. La espacialización puede realizarse entonces en la forma descrita anteriormente, por ejemplo, generando dos corrientes, 55 una direccional y una no direccional, que se derivan de los canales de transporte decodificados, en el caso de los cuadros inactivos, de los ruidos de confort de canal de transporte. Las dos corrientes se mezclan de manera ascendente y se mezclan juntas entonces en el bloque 740, dependiendo de los parámetros espaciales 318.

Alternativamente, el ruido de confort o una parte del mismo, podría generarse directamente dentro de la síntesis DirAC 60 en el dominio de banco de filtros. En realidad, DirAC puede controlar la coherencia de la escena restituída con la ayuda de los canales de transporte 224, los parámetros espaciales 318, 316, 319, y algunos decorreladores (por ejemplo, 730). Los decorreladores 730 pueden reducir la coherencia de campo sonoro sintetizado. La imagen espacial se percibe entonces con más amplitud, profundidad, difusión, reverberación o externalización, en el caso de reproducción con auriculares. Sin embargo, los decorreladores con frecuencia son propensos a artefactos auditivos típicos, y es

deseable reducir su uso. Esto puede lograrse, por ejemplo, mediante el llamado método de síntesis de covarianza [5], aprovechando el componente incoherente ya existente de los canales de transporte. Sin embargo, este enfoque puede tener limitaciones, especialmente en el caso de un canal de transporte monofónico.

- 5 En el caso de ruido de confort generado por el ruido aleatorio, es ventajoso generar para cada uno de los canales de salida, o al menos un subconjunto de ellos, un ruido de confort dedicado. Más específicamente, es ventajoso aplicar la generación de ruido de confort no sólo en los canales de transporte, sino también a los canales de audio intermedios utilizados en el renderizador espacial (síntesis DirAC) 220 (y en el bloque de mezclado 740). La decorrelación de campo difuso se dará entonces directamente utilizando diferentes generadores de ruido, más que utilizando los decorreladores 730, lo que puede disminuir la cantidad de artefactos, pero también la complejidad general. En realidad, diferentes ejecuciones de un ruido aleatorio, son por definición, decorrelacionadas. Las figuras 8 y 9 ilustran dos formas para lograr esto, generando el ruido de confort completa o parcialmente dentro del renderizador espacial 220. En la figura 8, el CN se hace en el dominio de frecuencia, como se describe en [5], puede generarse directamente con el dominio de banco de filtros de renderizador espacial, evitando tanto el análisis de banco de filtros 720 como los decorreladores 730. Aquí, K, el número de canales para los cuales se genera un ruido de confort, es igual o mayor que M, el número de canales de transporte, y menor o igual que N, el número de canales de salida. En el caso más simple, K = N.

La figura 9 ilustra otra alternativa para incluir la generación de ruido de confort 810 en el renderizador.

- 20 La generación de ruido de confort se divide entre el interior (en 710) y el exterior (en 810) del renderizador espacial 220. El ruido de confort 228d dentro del renderizador 220 se agrega (en el sumador 920) a la salida eventual del decorrelador 228a. Por ejemplo, una banda baja puede generarse en el exterior en el mismo dominio que en el codificador principal, con el fin de poder actualizar fácilmente las memorias necesarias. Por otra parte, la generación de ruido de confort puede realizarse directamente en el renderizador para las frecuencias altas.

Además, la generación de ruido de confort también puede aplicarse durante los cuadros activos 346. En lugar de apagar completamente la generación de ruido de confort durante los cuadros activos 346, puede mantenerse activa reduciendo su intensidad. Sirve entonces para enmascarar la transición entre los cuadros activos e inactivos, enmascarando también los artefactos y las imperfecciones en el codificador principal y en el modelo del audio espacial paramétrico. Esto se propuso en [11] para la codificación del habla monofónica. El mismo principio puede extenderse a la codificación del habla espacial. La figura 10 ilustra una implementación. Esta vez, la generación de los ruidos de confort en el renderizador espacial 220 se enciende en la fase activa e inactiva. En la fase inactiva 348, es complementaria a la generación de ruido de confort realizada en los canales de transporte. En el renderizador, el ruido de confort se hace en K canales iguales o mayores que los M canales de transporte, con el objeto de reducir el uso de los decorreladores. La generación de ruido de confort en el renderizador espacial 220 se suma a la versión mezclada de manera ascendente 228f de los canales de transporte, lo que puede lograrse mediante una simple copia de los M canales en los K canales.

40 **Características adicionales de las figuras**

la figura 1 análisis y síntesis DirAC de [1]

- 45 la figura 2 Diagrama de bloques detallado del análisis y síntesis DirAC en el codificador de audio 3D con tasa de bits baja

la figura 3 Diagrama de bloques de decodificador

la figura 4 Diagrama de bloques del analizador de escena de audio en modo DirAC

- 50 la figura 5 Diagrama de bloques del analizador de escena de audio para el formato de entrada MASA

la figura 6 Diagrama de bloques de decodificador

- 55 la figura 7 Diagrama de bloques de renderizador espacial (síntesis DirAC) con CNG en los canales de transporte en el exterior del renderizador

la figura 8 Diagrama de bloques del renderizador espacial (síntesis DirAC) con CNG realizada directamente en el dominio de banco de filtros de renderizador para los K canales, $K \geq M$ canales de transporte.

- 60 la figura 9 Diagrama de bloques del renderizador espacial (síntesis DirAC) con CNG realizada tanto en el exterior como en el interior del renderizador espacial.

la figura 10 Diagrama de bloques del renderizador espacial (síntesis DirAC) con CNG realizada tanto en el exterior

como en el interior del renderizador espacial y también activado para los cuadros activos e inactivos.

Ventajas

5 Las realizaciones de la presente invención permiten extender el DTX a la codificación del audio espacial paramétrico de una manera eficiente. Puede restituir con una alta fidelidad perceptual el ruido de fondo, incluso para los cuadros inactivos para los cuales la transmisión puede interrumpirse para ahorrar el ancho de banda de la comunicación.

10 Para esto, el SID de los canales de transporte se extiende por los parámetros espaciales inactivos relevantes para describir la imagen espacial de ruido de fondo. El ruido de confort generado se aplica en los canales de transporte antes de espacializarse por el renderizador (síntesis DirAC). Alternativamente, para una mejora en la calidad, la CNG puede aplicarse a más canales que los canales de transporte dentro de la renderización. Permite ahorro de la complejidad y reduce la molestia de los artefactos del decorrelador.

15 **Otros aspectos**

Debe mencionarse aquí que todas las alternativas o aspectos discutidos anteriormente y todos los aspectos definidos por los aspectos independientes en los siguientes aspectos, pueden utilizarse de manera individual, es decir, sin ningún otra alternativa u objeto que la alternativa, objeto o aspecto independiente contemplados. Sin embargo, en 20 otras realizaciones, dos o más de las alternativas o los aspectos o los aspectos independientes, pueden combinarse unos con otros, y en otras realizaciones, todos los aspectos o alternativas y todos los aspectos independientes pueden combinarse unos con otros.

Una señal codificada de forma inventiva se puede almacenar en un medio de almacenamiento digital o un medio de almacenamiento no transitorio o se puede transmitir en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión alámbrico tal como Internet.

Si bien se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del método correspondiente, en el cual un bloque o dispositivo corresponde a una etapa del método o a una característica de una etapa del método. De manera análoga, los aspectos descritos en el contexto de una etapa del método también representan una descripción de un bloque o elemento correspondiente o de una característica de un aparato correspondiente.

Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede realizar utilizando un medio de almacenamiento digital, por ejemplo, un disquete, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga señales de control electrónicamente legibles almacenadas en el mismo, que cooperan (o sean capaces de cooperar) con un sistema informático programable de manera que se realiza el método respectivo.

40 Algunas realizaciones de acuerdo con la invención comprenden un transportador no transitorio de datos que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se ejecute uno de los métodos descritos en el presente documento.

En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto de programa informático con un código de programa, donde el código de programa cumple la función de ejecutar uno de los métodos al ejecutarse el programa informático en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un portador legible por una máquina.

50 Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en la presente, almacenado en un portador legible por máquina o un medio de almacenamiento no transitorio.

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

55 Una realización adicional de los métodos inventivos es, por lo tanto, un portador de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los métodos descritos en el presente documento.

60 Otra realización del método inventivo es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en la presente. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferida a través de una conexión de comunicación de datos, por ejemplo a través de Internet.

Otra realización comprende un medio de procesamiento, por ejemplo un ordenador o un dispositivo lógico programable, configurado o adaptado para realizar uno de los métodos descritos en el presente documento.

5 Otra realización comprende un ordenador que tiene instalado en él el programa informático para realizar uno de los métodos descritos en el presente documento.

En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en campo) para realizar algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programables en campo puede cooperar con un 10 microprocesador para realizar uno de los métodos descritos en el presente documento. Por lo general, los métodos son ejecutados preferentemente por cualquier aparato de hardware.

Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los detalles aquí descritos resultarán evidentes para 15 los expertos en la técnica. Es la intención, por consiguiente, limitarse sólo por el alcance de las reivindicaciones de patente que siguen y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

20 Los aspectos definidos posteriormente para el primer conjunto de realizaciones y el segundo conjunto de realizaciones, pueden combinarse, de manera que ciertas características de un conjunto de realizaciones pueden incluirse en el otro conjunto de realizaciones.

REIVINDICACIONES

1. Aparato (300) para generar una escena de audio codificada (304) de una señal de audio (302) que tiene un primer cuadro (306) y un segundo cuadro (308), que comprende:
 - 5 un generador de parámetros de campo sonoro (310) para determinar una primera representación de parámetros de campo sonoro (316) para el primer cuadro (306) de la señal de audio (302) en el primer cuadro (306), y una segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308) de la señal de audio (302) en el segundo cuadro(308); y
 - 10 un detector de actividad (320) para analizar la señal de audio (302) para determinar, dependiendo de la señal de audio (302), que el primer cuadro es un cuadro activo (306), y que el segundo cuadro es un cuadro inactivo (308),
 - 15 en donde el generador de parámetros de campo sonoro (310) está configurado para determinar, del segundo cuadro (308) de la señal de audio, una pluralidad de fuentes de sonido individuales y determinar, para cada fuente de sonido, una descripción paramétrica (328) para el segundo cuadro,
 - 20 en donde el generador de campo sonoro (310) está configurado para descomponer el segundo cuadro (308) en una pluralidad de contenedores de frecuencia, cada contenedor de frecuencia representa una fuente de sonido individual, y para determinar, para cada contenedor de frecuencia, al menos un parámetro espacial inactivo como la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308), el al menos un parámetro espacial inactivo que comprende un parámetro de dirección, un parámetro de dirección de llegada, un parámetro de difusividad o un parámetro de relación de energía,
 - 25 el aparato que comprende además:
 - un codificador de señal de audio (330) para generar una señal de audio codificada (344), la señal de audio codificada (344), que proporciona una señal de audio codificada (346) para el primer cuadro que es el cuadro activo (306) y la descripción paramétrica (348) para el segundo cuadro que es el cuadro inactivo(308); y
 - 30 un formador de señales codificadas para (370) componer la escena de audio codificada (304) juntando la primera representación de parámetros de campo sonoro (316) para el primer cuadro (306), la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308), la señal de audio codificada (346) para el primer cuadro (306), y la descripción paramétrica (348) para el segundo cuadro (308).
 2. Aparato de la reivindicación 1, en donde el generador de parámetros de campo sonoro (310) se configura para generar la segunda representación de parámetros de campo sonoro (318), de manera que la segunda representación de parámetros de campo sonoro (318) comprende un parámetro que indica una característica de la señal de audio (302), con respecto a una posición del oyente.
 3. Aparato de la reivindicación 1 o 2, en donde la primera representación de parámetros de campo sonoro (316) comprende uno o más parámetros de dirección, que indican una dirección de sonido con respecto a una posición del oyente en el primer cuadro (306), o uno o más parámetros de difusión, que indican una porción de un sonido difuso con respecto a un sonido directo en el primer cuadro (306), o uno o más parámetros de relación de energía, que indican una relación de energía de un sonido directo y un sonido difuso en el primer cuadro (306), o un parámetro de coherencia intercanal/entorno en el primer cuadro (306).
 4. Aparato de una de las reivindicaciones anteriores, en donde la señal de audio para el primer cuadro (306) y el segundo cuadro (308) comprende un formato de entrada que tiene una pluralidad de componentes que representan un campo sonoro con respecto a un oyente,
 - 50 en donde el generador de parámetros de campo sonoro (310) se configura para calcular uno o más canales de transporte (324, 326, 328) para el primer cuadro (306) y el segundo cuadro (308), utilizando un mezclado descendente de la pluralidad de componentes, y para analizar el formato de entrada para determinar la primera representación de parámetro, relacionada con el uno o más canales de transporte, o
 - 55 en donde el generador de parámetros de campo sonoro (310) se configura para calcular uno o más canales de transporte (324, 326, 328), utilizando un mezclado descendente de la pluralidad de componentes, y
 - 60 en donde el detector de actividad (320) se configura para analizar el uno o más canales de transporte (328), derivados de la señal de audio en el segundo cuadro (308).
 5. Aparato de una de las reivindicaciones 1 a 4,

- en donde la señal de audio para el primer cuadro (306) o el segundo cuadro (308) comprende un formato de entrada que tiene, para cada cuadro del primer y el segundo cuadros, uno o más canales de transporte y metadatos asociados con cada cuadro,
- 5 en donde el generador de parámetros de campo sonoro (310) se configura para leer los metadatos del primer cuadro (306) y el segundo cuadro (308) y utilizar o procesar los metadatos para el primer cuadro (306) como la primera representación de parámetros de campo sonoro (316) y para procesar los metadatos del segundo cuadro (308) para obtener la segunda representación de parámetros de campo sonoro (318), en donde el procesamiento para obtener la segunda representación de parámetros de campo sonoro (318) es tal, que una cantidad de las unidades de información requerida para la transmisión de los metadatos para el segundo cuadro (308), se reduce con respecto a una cantidad requerida antes del procesamiento.
- 10 6. Aparato de la reivindicación 5,
- 15 en donde el generador de parámetros de campo sonoro (310) se configura para procesar los metadatos para el segundo cuadro (308) para reducir un número de elementos de información en los metadatos o para remuestrear los elementos de información en los metadatos a una resolución más baja, tal como una resolución de tiempo o una resolución de frecuencia, o para recuantificar las unidades de información de los metadatos para el segundo cuadro (308) a una representación más gruesa con respecto a una situación antes de la recuantificación.
- 20 7. Aparato de una de las reivindicaciones anteriores,
- 25 en donde el codificador de señal de audio (330) se configura para determinar una descripción de información de silencio para el cuadro inactivo como la descripción paramétrica (348),
- en donde la descripción de información de silencio comprende una información relacionada con la amplitud, tal como una energía, una potencia o una intensidad sonora para el segundo cuadro (308), y una información de configuración, tal como información de configuración espectral, o información relacionada con la amplitud para el segundo cuadro (308), tal como una energía, una potencia, o una intensidad sonora, y una codificación predictiva lineal, LPC, de los parámetros para el segundo cuadro (308), o los parámetros de escala para el segundo cuadro (308) con una resolución de frecuencia asociada variable, de manera que diferentes parámetros de escala se refieren a las bandas de frecuencia con diferentes anchos.
- 30 35 8. Aparato de una de las reivindicaciones anteriores,
- en donde el codificador de señal de audio (330) se configura para codificar, para el primer cuadro (306), la señal de audio, utilizando un modo de codificación en el dominio del tiempo o en el dominio de frecuencia, la señal de audio codificada comprende muestras codificadas del dominio del tiempo, muestras codificadas del dominio espectral, muestras codificadas del dominio LPC e información secundaria obtenida de los componentes de la señal de audio u obtenida de uno o más canales de transporte derivados de los componentes de la señal de audio mediante una operación de mezclado descendente.
- 40 45 9. Aparato de una de las reivindicaciones anteriores,
- en donde la señal de audio (302) comprende un formato de entrada que es un formato Ambisonics de primer orden, un formato Ambisonics de orden superior, un formato multicanal asociado con una instalación del altavoz dada, tal como 5.1 o 7.1 o 7.1 + 4, o uno o más canales de audio que representan uno o varios objetos de audio diferentes, localizados en un espacio, como se indica por la información incluida en los metadatos asociados, o un formato de entrada que es una representación del audio espacial asociada con los metadatos,
- 50 55 en donde el generador de parámetros de campo sonoro (310) se configura para determinar la primera representación de parámetros de campo sonoro (316) y la segunda representación de campo sonoro, de manera que los parámetros representan un campo sonoro con respecto a una posición definida del oyente.
- 60 10. Aparato de una de las reivindicaciones 1 a 8,
- en donde la señal de audio (302) comprende un formato de entrada que es un formato Ambisonics de primer orden, un formato Ambisonics de orden superior, un formato multicanal asociado con una instalación del altavoz dada, tal como 5.1 o 7.1 o 7.1 + 4, o uno o más canales de audio que representan uno o varios objetos de audio diferentes, localizados en un espacio, como se indica por la información incluida en los metadatos asociados, o un formato de entrada que es una representación del audio espacial asociada con los metadatos,

en donde la señal de audio comprende una señal del micrófono como captada por un micrófono real o un micrófono virtual o una señal de micrófono creada sintéticamente, por ejemplo, que está en un formato Ambisonics de primer orden, o un formato Ambisonics de orden superior.

- 5 11. Aparato de una de las reivindicaciones anteriores,
en donde el detector de actividad (320) se configura para detectar una fase de inactividad en el segundo cuadro (308) y uno o más cuadros después del segundo cuadro (308), y
10 en donde el detector de actividad (320) se configura para determinar una fase inactiva que comprende el segundo cuadro (308) y ocho cuadros después del segundo cuadro (308), y en donde el codificador de señal de audio (330) se configura para generar una descripción paramétrica para un cuadro inactivo sólo en cada octavo cuadro, y en donde el generador de parámetros de campo sonoro (310) se configura para generar una representación de parámetros de campo sonoro para cada octavo cuadro inactivo.
15 12. Aparato de una de las reivindicaciones 1 a 10,
en donde el detector de actividad (320) se configura para detectar una fase de inactividad en el segundo cuadro (308) y uno o más cuadros después del segundo cuadro (308), y
20 en donde el generador de parámetros de campo sonoro (310) se configura para generar una representación de parámetros de campo sonoro para cada cuadro inactivo, incluso cuando el codificador de señal de audio (330) no genera una descripción paramétrica para un cuadro inactivo.
25 13. Aparato de una de las reivindicaciones 1 a 10,
en donde el generador de parámetros de campo sonoro (310) se configura para determinar un representación de parámetros con una velocidad del cuadro más alta que la del codificador de señal de audio (330) que genera la descripción paramétrica para uno o más cuadros inactivos.
30 14. Aparato de una de las reivindicaciones anteriores,
en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
35 utilizando los parámetros espaciales para una o más direcciones en las bandas de frecuencia y las relaciones de energía asociadas en las bandas de frecuencia que corresponde a una relación de un componente direccional con respecto a la energía total.
40 15. Aparato de una de las reivindicaciones 1 a 13,
en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
45 para determinar un parámetro de difusión que indica una relación de sonido difuso o el sonido directo.
16. Aparato de una de las reivindicaciones 1 a 13,
en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
50 para determinar una información de dirección utilizando un esquema de cuantificación más gruesa en comparación con una cuantificación en el primer cuadro (306).
55 17. Aparato de una de las reivindicaciones 1 a 13,
en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
60 utilizando una promediación de una dirección con el tiempo o frecuencia para obtener una resolución más gruesa del tiempo o de frecuencia
18. Aparato de una de las reivindicaciones 1 a 13,

- en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
- para determinar una representación de parámetros de campo sonoro para uno o más cuadros inactivos con la misma resolución de frecuencia que en la primera representación de parámetros de campo sonoro (316) para un cuadro activo, y con una ocurrencia temporal que es menor que la ocurrencia temporal para los cuadros activos, con respecto a una información de dirección en la representación de parámetros de campo sonoro para el cuadro inactivo.
- 10 19. Aparato de una de las reivindicaciones 1 a 13,
 en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
- 15 para determinar la segunda representación de parámetros del campo sonoro (318), que tiene un parámetro de difusión, en donde el parámetro de difusión se transmite con la misma resolución del tiempo o de frecuencia que para los cuadros activos, pero con una cuantificación más gruesa.
- 20 20. Aparato de una de las reivindicaciones 1 a 13,
 en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
- 25 para cuantificar un parámetro de difusión para la segunda representación de campo sonoro con un primer número de bits, y en donde se transmite sólo un segundo número de bits de cada índice de cuantificación, el segundo número de bits es más pequeño que el primer número de bits.
21. Aparato de una de las reivindicaciones 1 a 13,
 en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
- 30 para determinar, para la segunda representación de parámetros de campo sonoro (318), una coherencia intercanal si la señal de audio tiene canales de entrada que corresponden a los canales colocados en un dominio espacial o las diferencias del nivel intercanal si la señal de audio tiene canales de entrada que corresponden a los canales colocados en el dominio espacial.
- 35 22. Aparato de una de las reivindicaciones 1 a 13,
 en donde el generador de parámetros de campo sonoro (310) se configura para determinar la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308)
- 40 para determinar una coherencia del entorno que se define como una relación de energía difusa que es coherente en un campo sonoro representado por la señal de audio.
- 45 23. Método para generar una escena de audio codificada (304) de una señal de audio (302) que tiene un primer cuadro (306) y un segundo cuadro (308), que comprende:
 determinar una primera representación de parámetros de campo sonoro (316) para el primer cuadro (306) de la señal de audio en el primer cuadro (306), y una segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308) de la señal de audio en el segundo cuadro (308); y
- 50 para analizar la señal de audio (302) para determinar, dependiendo de la señal de audio, que el primer cuadro (306) es un cuadro activo, y que el segundo cuadro (308) es un cuadro inactivo;
- 55 en donde la determinación de la primera representación de parámetros de campo sonoro incluye determinar, del segundo cuadro (308) de la señal de audio, una pluralidad de fuentes de sonido individuales y determinar, para cada fuente de sonido, una descripción paramétrica (328) para el segundo cuadro (308),
- 60 en donde la determinación de la primera representación de parámetros de campo sonoro incluye descomponer el segundo cuadro (308) en una pluralidad de contenedores de frecuencia, cada contenedor de frecuencia representa una fuente de sonido individual de la pluralidad de fuentes de sonido individuales, y determinar, para cada contenedor de frecuencia, al menos un parámetro espacial inactivo como la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308), el al menos un

- parámetro espacial inactivo comprende un parámetro de dirección, un parámetro de dirección de llegada, un parámetro de difusión, o un parámetro de relación de energía,
- el método comprende además:
- 5 para generar una señal de audio codificada (344), la señal de audio codificada (344), que proporciona una señal de audio codificada (346) para el primer cuadro que es el cuadro activo (306) y la descripción paramétrica (348) para el segundo cuadro que es el cuadro inactivo (308); y
- 10 componer la escena de audio codificada (304) juntando la primera representación de parámetros de campo sonoro (316) para el primer cuadro (306), la segunda representación de parámetros de campo sonoro (318) para el segundo cuadro (308), la señal de audio codificada (346) para el primer cuadro (306), y la descripción paramétrica (348) para el segundo cuadro (308).
- 15 24. Programa informático para realizar, cuando se ejecuta en un ordenador o procesador, el método de la reivindicación 23.

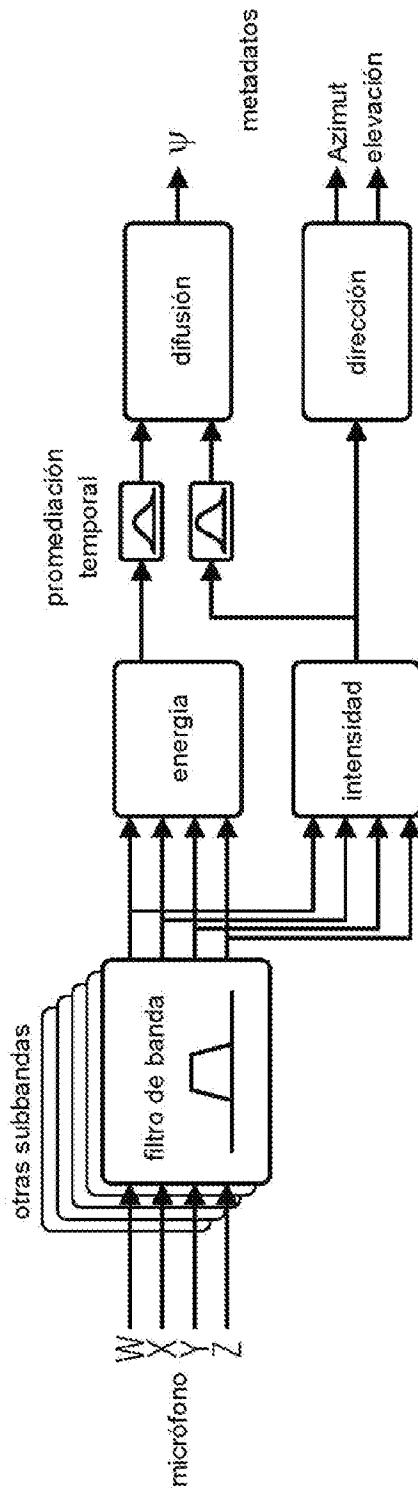


FIGURA 1a

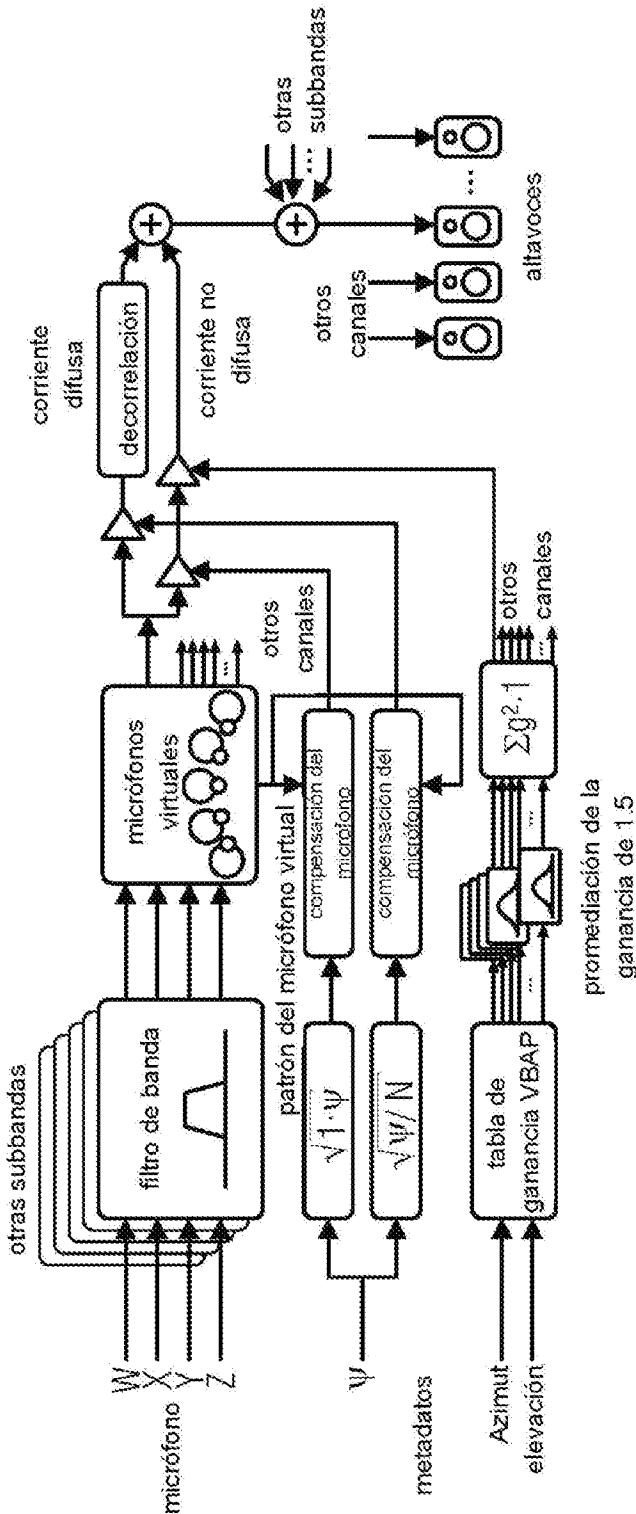


FIGURA 1b

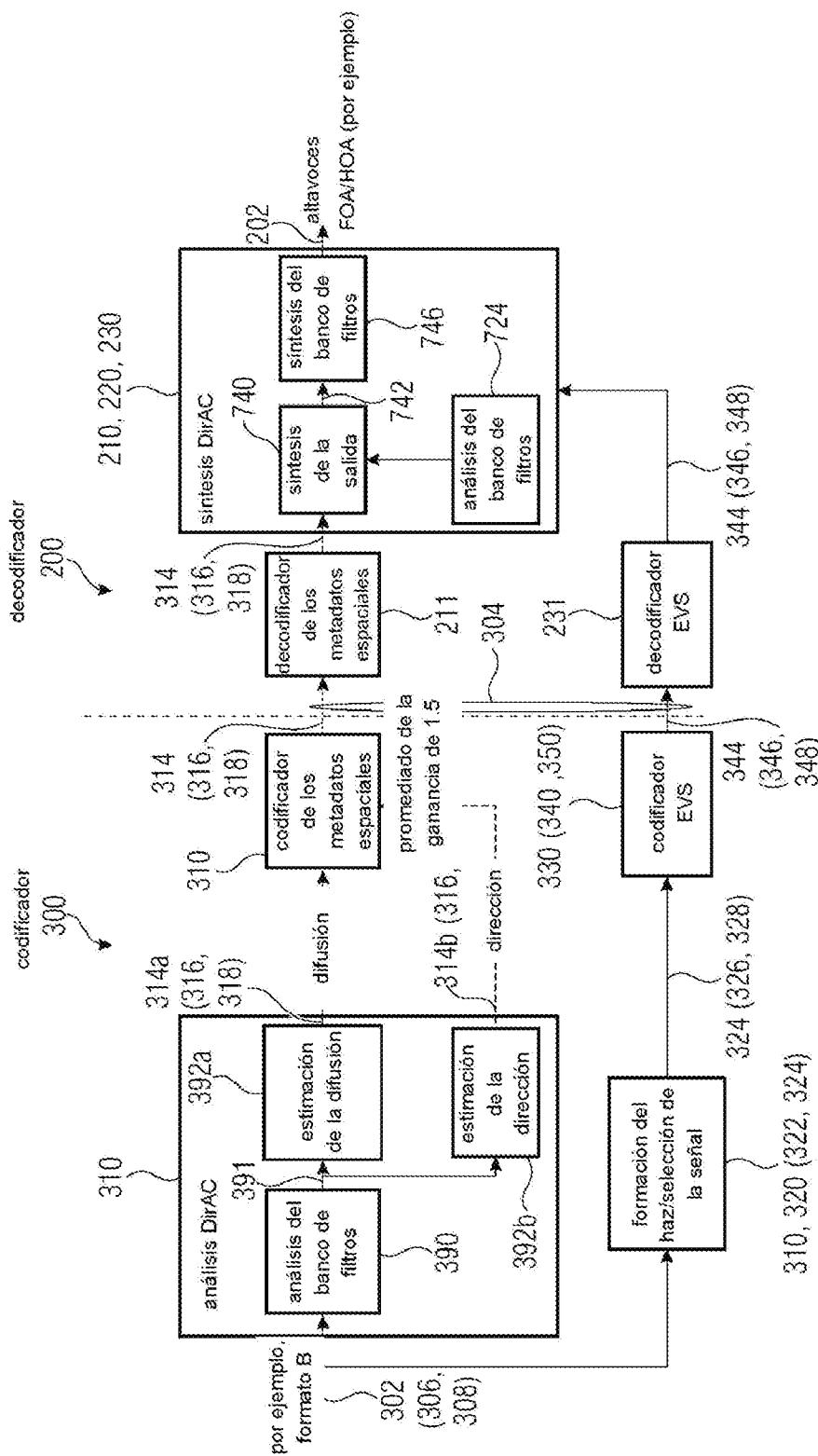


FIGURA 2

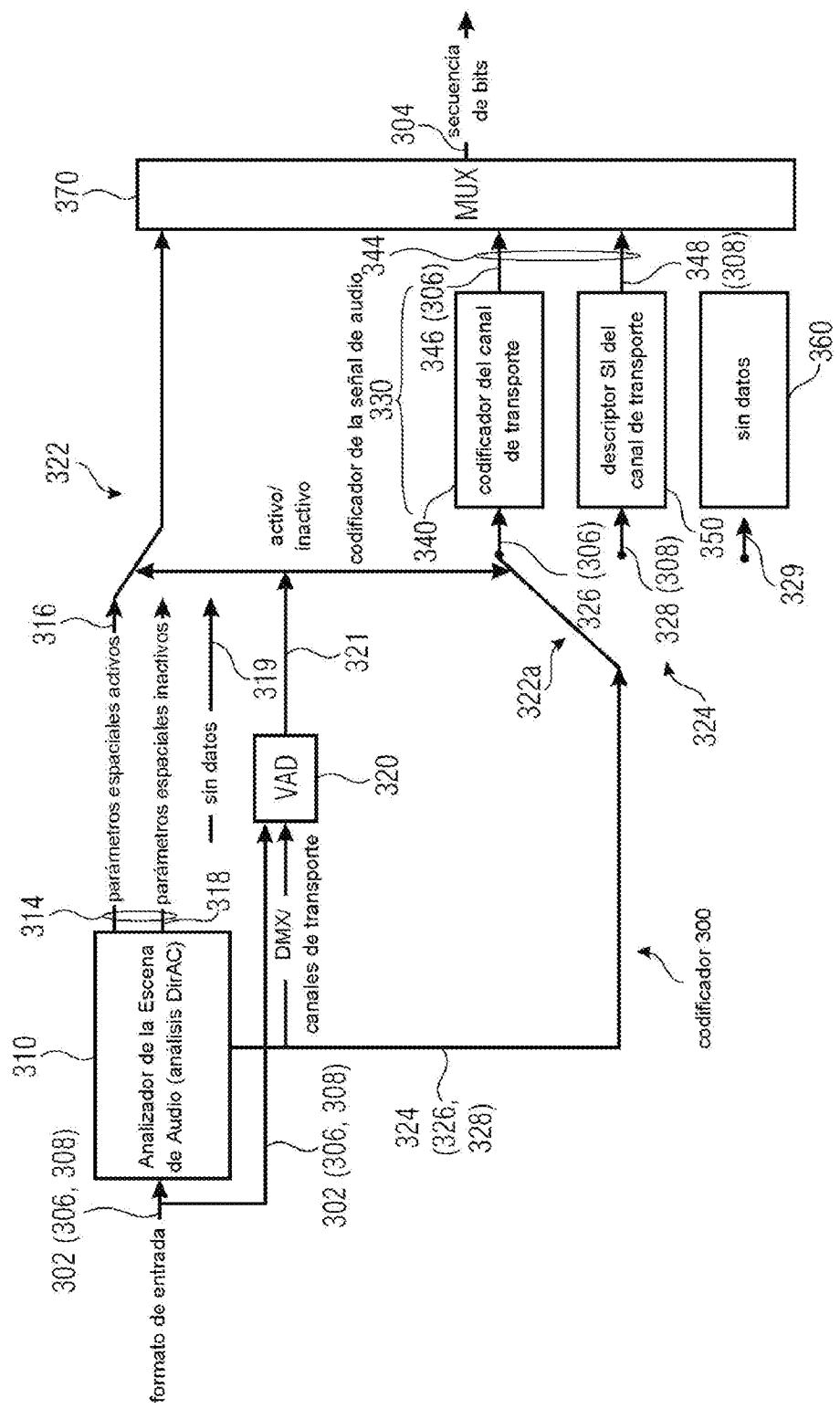


FIGURA 3

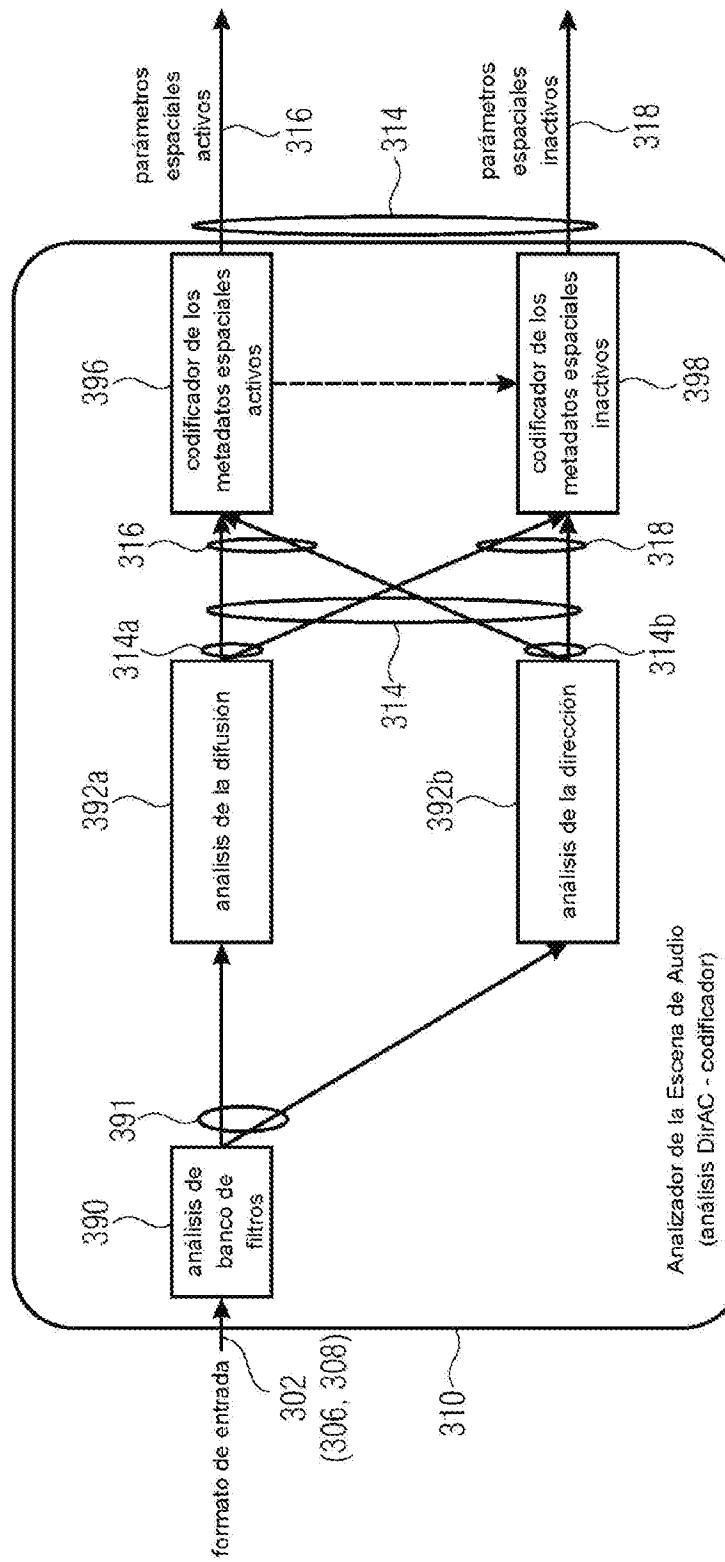


FIGURA 4

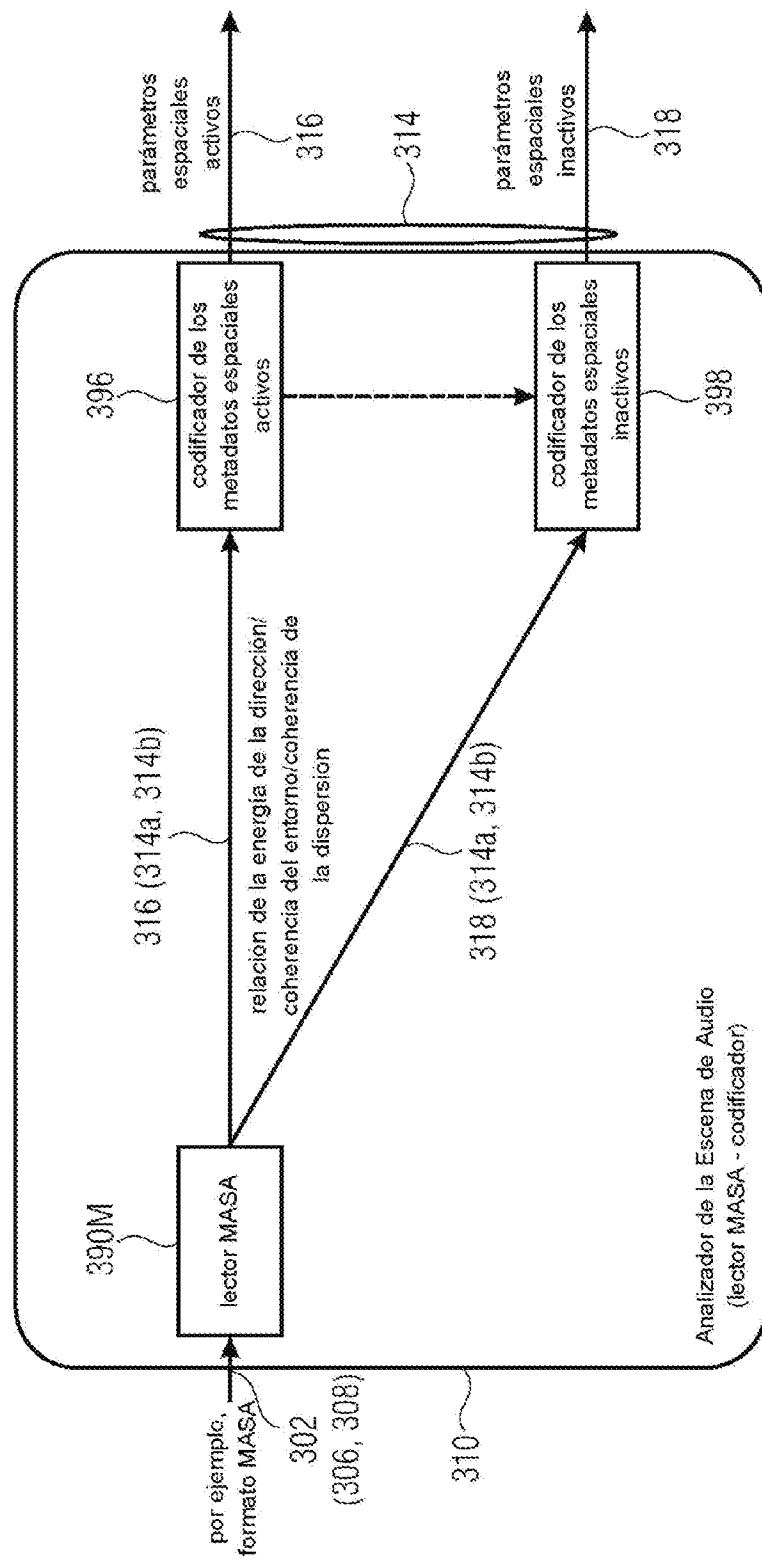


FIGURA 5

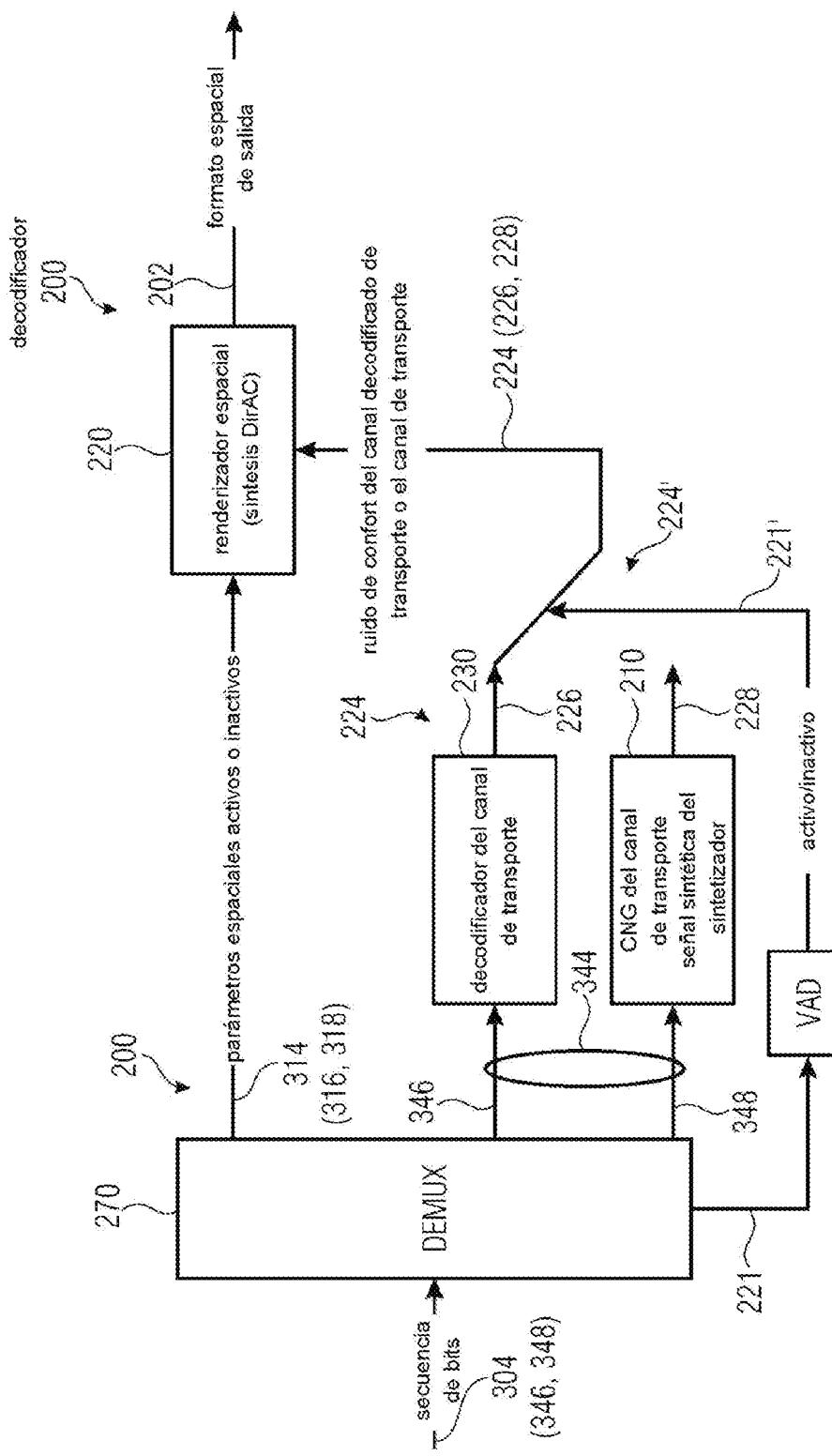


FIGURA 6

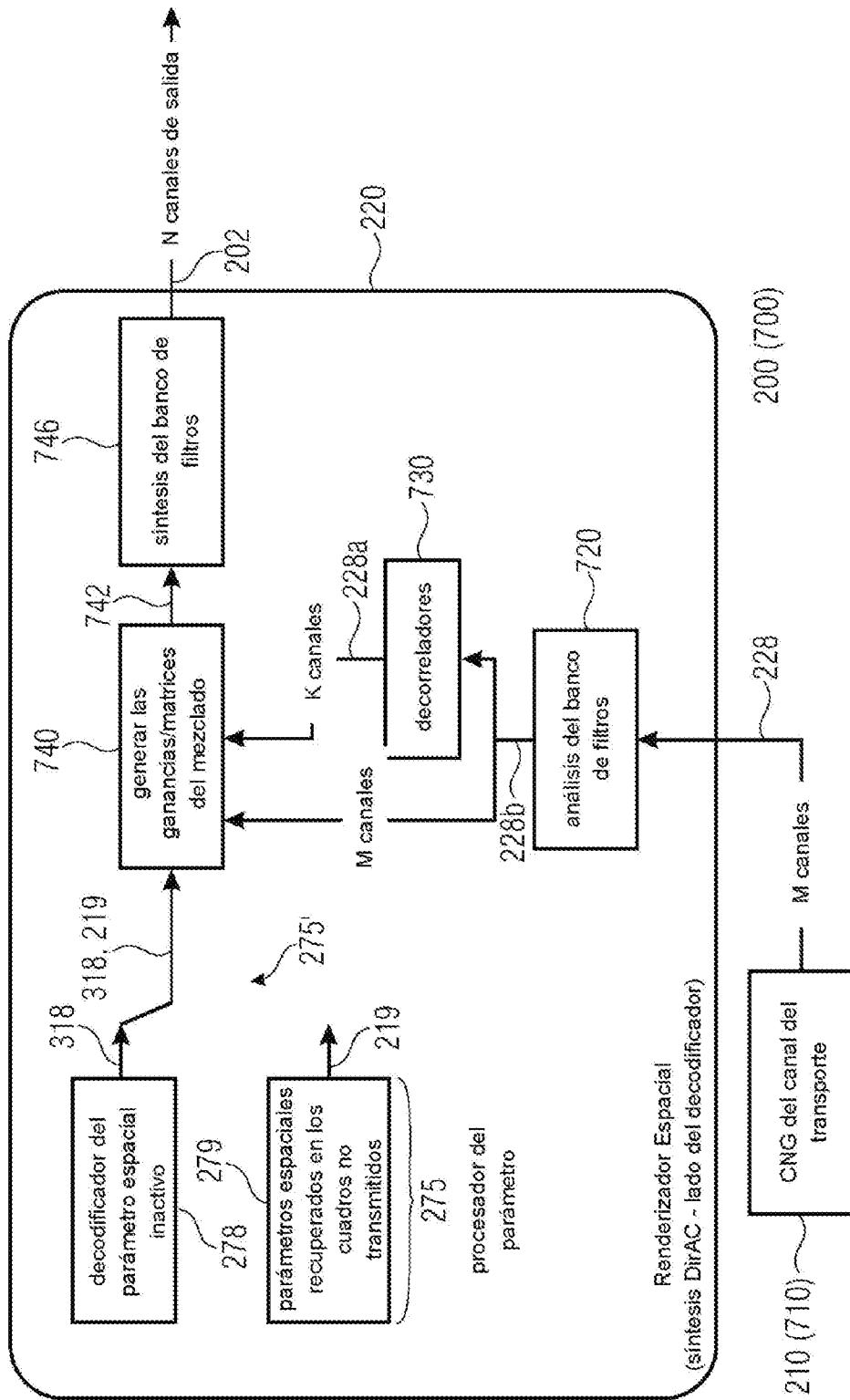


FIGURA 7

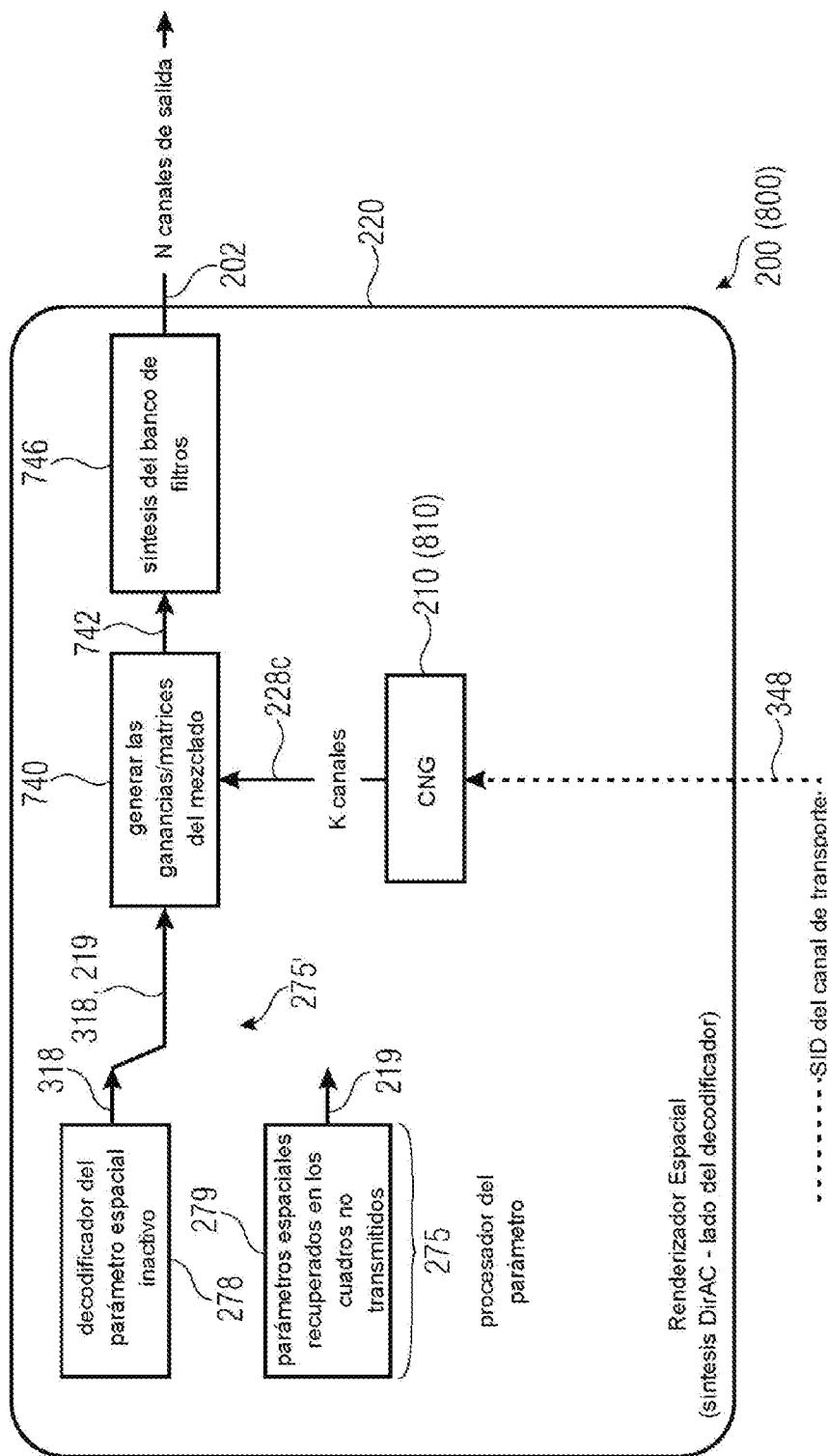


FIGURA 8

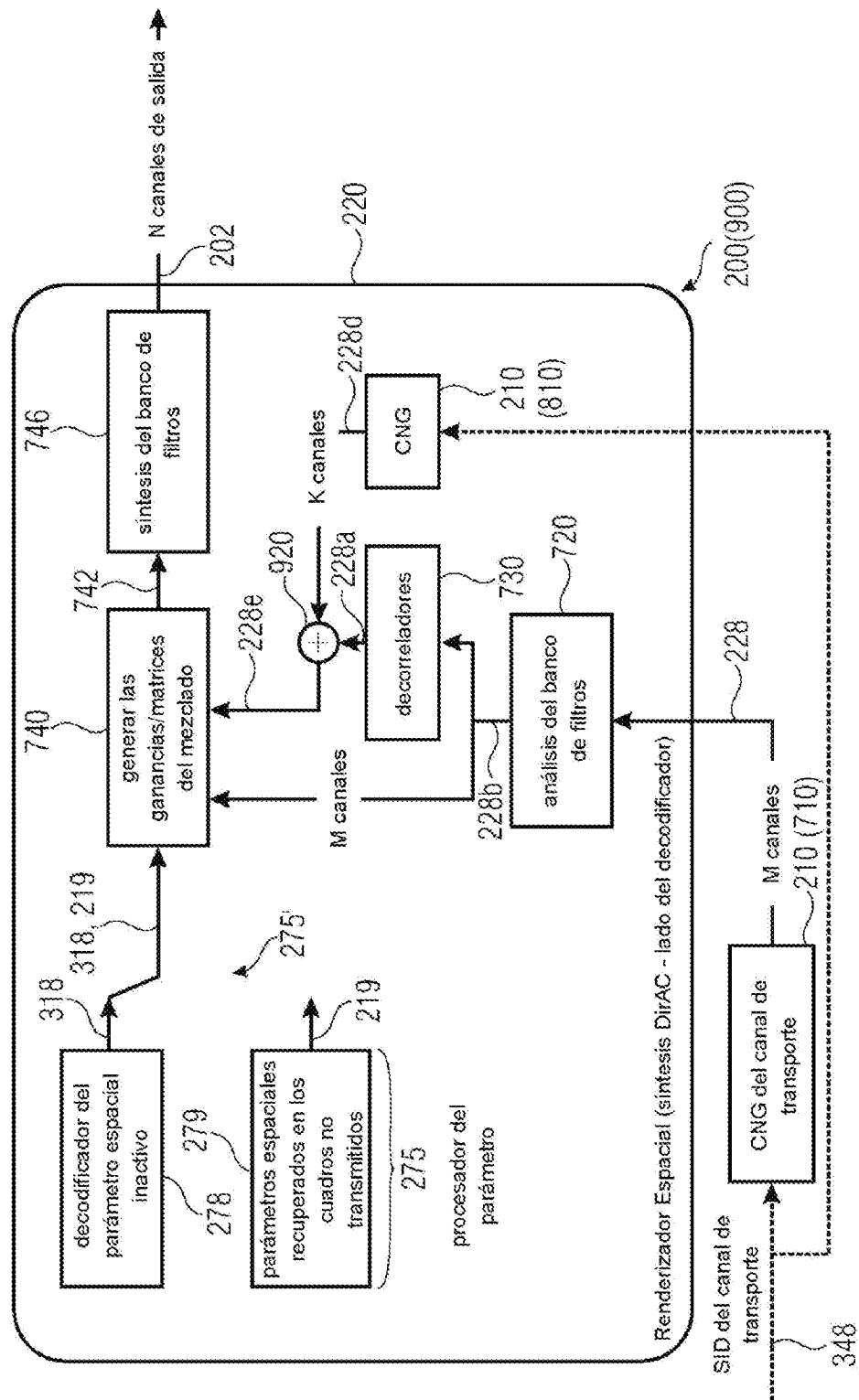


FIGURA 9

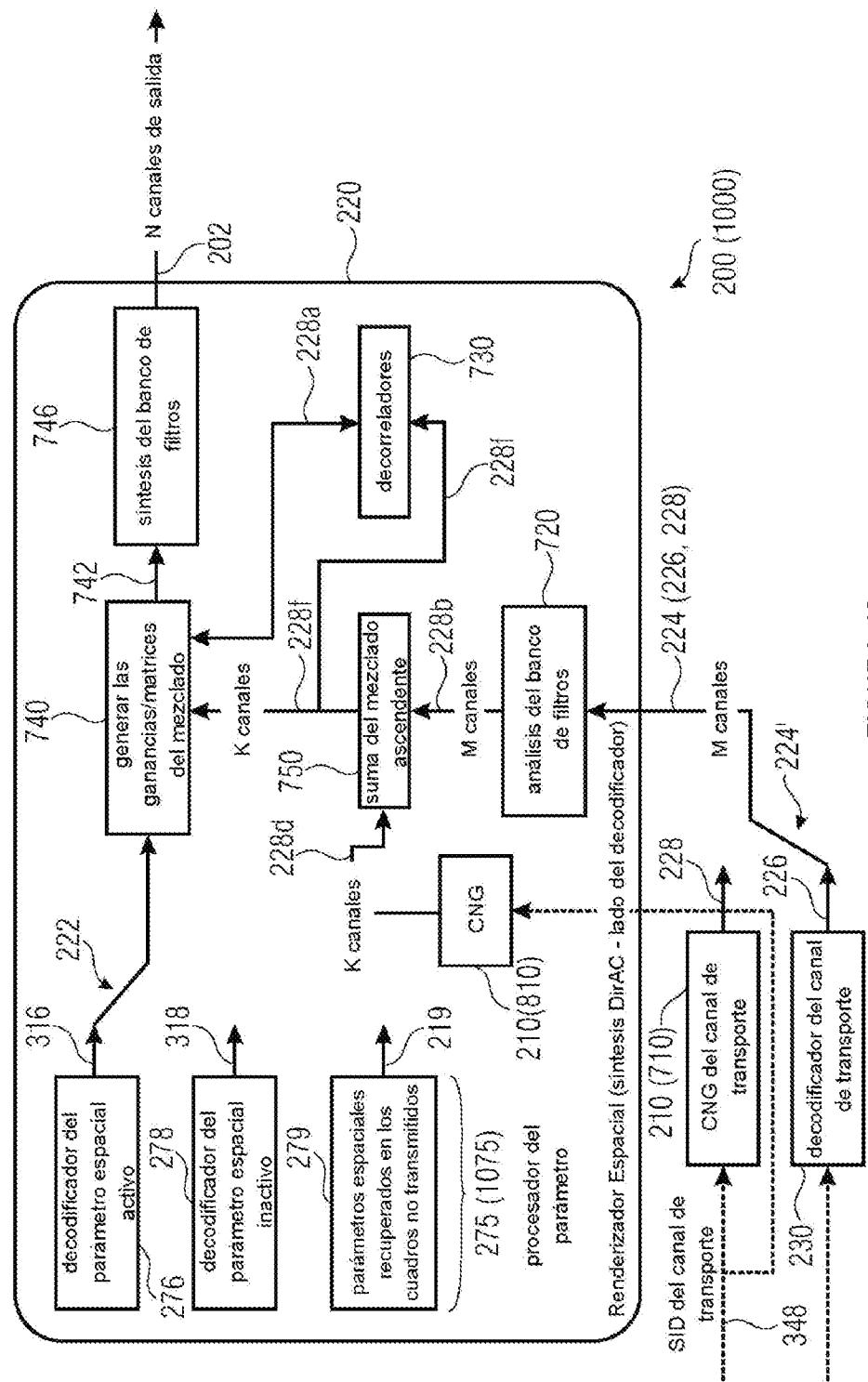


FIGURA 10

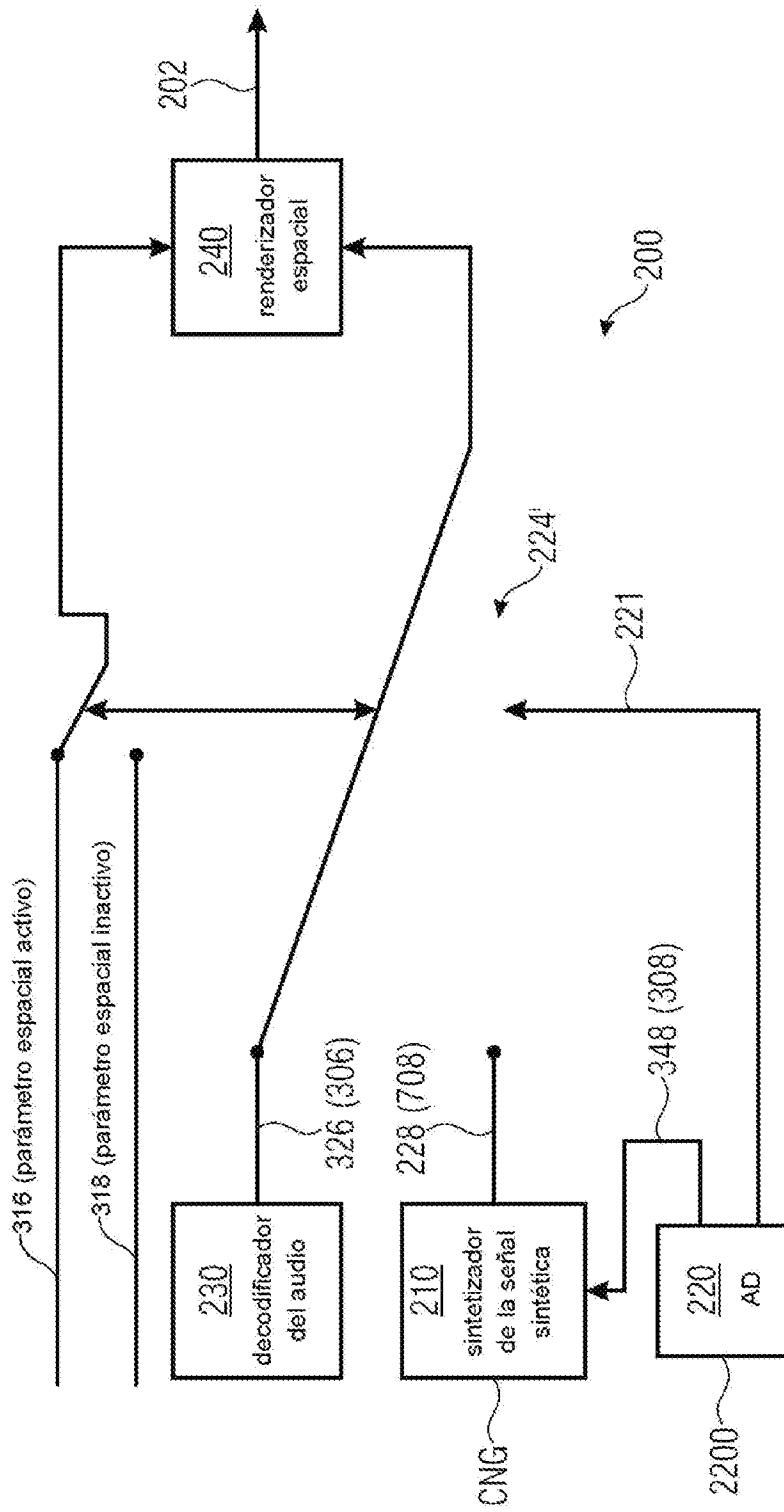


FIGURA 11