(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 907 377**

(51) Int. CI.:

G10L 19/008 (2013.01)
G10L 19/16 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **01.10.2018 PCT/EP2018/076641**

(87) Fecha y número de publicación internacional: **11.04.2019 WO19068638**

(96) Fecha de presentación y número de la solicitud europea: **01.10.2018 E 18779381 (5)**

(97) Fecha y número de publicación de la concesión europea: **22.12.2021 EP 3692523**

(54) Título: **Aparato, procedimiento y programa informático para la codificación, la decodificación, el procesamiento de escenas y otros procedimientos relacionados con la codificación de audio espacial basada en DirAC**

(30) Prioridad:

04.10.2017 EP 17194816

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
25.04.2022

(73) Titular/es:

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

(72) Inventor/es:

FUCHS, GUILLAUME;
HERRE, JÜRGEN;
KÜCH, FABIAN;
DÖHLA, STEFAN;
MULTRUS, MARKUS;
THIERGART, OLIVER;
WÜBBOLT, OLIVER;
GHIDO, FLORIN;
BAYER, STEFAN y
JAEGERS, WOLFGANG

(74) Agente/Representante:

PONTI & PARTNERS, S.L.P.

DESCRIPCIÓN

Aparato, procedimiento y programa informático para la codificación, la decodificación, el procesamiento de escenas y otros procedimientos relacionados con la codificación de audio espacial basada en DirAC

Campo de la invención

[0001] La presente invención se refiere al procesamiento de señal de audio y, en particular, al procesamiento de señales de audio de descripciones de audio de escenas de audio.

Introducción y estado de la técnica:

[0002] La transmisión de una escena de audio en tres dimensiones requiere el manejo de múltiples canales que normalmente genera una gran cantidad de datos para transmitir. Además, el sonido 3D se puede representar en diferentes formas: sonido tradicional basado en canales en el que cada canal de transmisión está asociado con una posición de altavoz; sonido llevado a través de objetos de audio, que se puede posicionar en tres dimensiones de manera independiente de las posiciones de altavoces; y basado en escenas (o Ambisonics), donde la escena de audio está representada por un conjunto de señales de coeficientes que son los pesos lineales de funciones de base espacialmente ortogonal, por ej., armónicos esféricos. En contraste con la representación basada en canales, la representación basada en escenas es independiente de un altavoz específico de puesta a punto, y se puede reproducir en cualquier configuración de altavoz, a expensas de un proceso de renderización adicional en el decodificador.

[0003] Para cada uno de estos formatos, se desarrollaron esquemas de codificación específicos para el almacenamiento o la transmisión eficiente a bajas tasas de señales de audio. Por ejemplo, la envolvente MPEG es un esquema de codificación paramétrica para sonido envolvente basado en canales, mientras que la Codificación de Objeto de Audio Espacial MPEG (SAOC, por su sigla en inglés) es un procedimiento de codificación paramétrica dedicada a audio basado en objetos. Una técnica de codificación paramétrica para el orden superior de Ambisonics también se proporcionó en el reciente estándar de la fase 2de MPEG-H.

[0004] En este contexto, donde las tres representaciones de la escena de audio, audio basado en canales, basado en objetos y basado en escena se utilizan y necesitan ser soportados, existe la necesidad de diseñar un esquema universal que permita una codificación paramétrica eficiente de las tres representaciones de audio 3D. Además, hay una necesidad de ser capaces de codificar, transmitir y reproducir escenas de audio complejas compuestas por una mezcla de las diferentes representaciones de audio.

[0005] En el documento US 2016/0064005 A1 se describe una estrategia para combinar escenas de audio de diferentes formatos mediante la aplicación de una conversión de formato. La técnica de codificación de audio direccional (DirAC) [1] es una estrategia eficiente para el análisis y la reproducción del sonido espacial. DirAC utiliza una representación motivada perceptual del campo de sonido basada en la dirección de la llegada (DOA) y la difusión medida por banda de frecuencia. Se basa en el supuesto de que en un instante de tiempo y en una banda crítica, la resolución espacial del sistema auditivo se limita a decodificar una señal para la dirección y otra para la coherencia inter-audónica. El sonido espacial se representa entonces en dominio de frecuencia mediante fundido cruzado de dos corrientes: una corriente difusa no direccional y una corriente direccional no difusa.

[0006] DirAC fue pensado originalmente para el sonido en formato B grabado, pero también podría servir como un formato común para la mezcla de diferentes formatos de audio. DirAC ya se amplió para procesar el formato de sonido envolvente convencional 5.1 en [3]. También se propone la fusión de múltiples corrientes DirAC en [4]. Por otra parte, DirAC se extendió para soportar también las entradas de micrófono que no sean en formato B [6].

[0007] Sin embargo, falta un concepto universal para hacer DirAC una representación universal de escenas de audio en 3D que también es capaz de soportar el concepto de objetos de audio.

[0008] Se realizaron algunas consideraciones previamente para el manejo de objetos de audio en DirAC. DirAC se empleó en [5] como un extremo delantero acústico para el Codificador de Audio Espacial, SAOC, como una separación de fuente ciega para la extracción de varios transmisores de una mezcla de fuentes. Sin embargo, no se previó el uso de DirAC en sí mismo como el esquema de codificación de audio espacial y para procesar objetos de audio directamente junto con sus metadatos y para combinarlos potencialmente entre sí y con otras representaciones de audio.

[0009] Un objeto de la presente invención es proporcionar un concepto mejorado de manipulación y procesamiento de escenas de audio y descripciones de escenas de audio.

[0010] Este objeto se consigue por medio de un aparato para la generación de una descripción de una

escena de audio combinada de la reivindicación 1, un procedimiento para la generación de una descripción de una escena de audio combinada de la reivindicación 14, o un programa informático relacionado de la reivindicación 15

[0011] Las realizaciones de la invención se refieren a un esquema de codificación paramétrica universal para la escena de audio 3D en torno al paradigma de Codificación de Audio Direccional (DirAC), una técnica perceptivamente motivada para el procesamiento de audio espacial. Originalmente DirAC fue diseñada para analizar una grabación en formato B de la escena de audio. La presente invención tiene como objetivo ampliar su capacidad para procesar de manera eficiente cualquier formato de audio espacial tal como audio basado en canal, Ambisonics, objetos de audio, o una mezcla de ellos.

[0012] La reproducción DirAC puede ser sencillamente generada para los diseños de altavoces arbitrarios y auriculares. La presente invención también se extiende a esta capacidad de salida, además, Ambisonics, objetos de audio o una mezcla de un formato. Más importante aún, la invención permite la posibilidad para el usuario de manipular objetos de audio y para lograr, por ejemplo, una mejora del diálogo en el extremo del decodificador.

Contexto: vista general del sistema de un Codificador de Audio Espacial DirAC

[0013] A continuación, se presenta una vista general de un sistema de codificación de audio espacial novedoso basado en DirAC de inmersión diseñado para Servicios de Voz y Audio (IVAS, por su sigla en inglés). El objetivo de este sistema es que sea capaz de manejar diferentes formatos de audio espacial que representan la escena de audio y codificarlos a bajas tasas de bits y reproducir la escena de audio original lo más fielmente posible después de la transmisión.

[0014] El sistema puede aceptar como entrada diferentes representaciones de escenas de audio. La escena de audio de entrada puede ser capturada por señales de múltiples canales destinadas a ser reproducidas en las diferentes posiciones de los altavoces, los objetos auditivos junto con metadatos que describen las posiciones de los objetos a lo largo del tiempo, o un formato Ambisonics de primer orden o de orden superior que representa el campo de sonido en la posición de escucha o de referencia.

[0015] Con preferencia, el sistema se basa en los Servicios de Voz Mejorados (EVS, por su sigla en inglés) 3GPP, dado que se espera que la solución opere con baja latencia para habilitar servicios de conversación en redes móviles.

[0016] La Fig. 9 es el lado del codificador de la codificación de audio espacial basada en DirAC que soporta diferentes formatos de audio. De acuerdo con lo mostrado en la Fig. 9, el codificador (codificador de IVAS) es capaz de soportar diferentes formatos de audio presentados al sistema por separado o al mismo tiempo. Las señales de audio pueden ser de naturaleza acústica, recogidas por los micrófonos, o de naturaleza eléctrica, que se supone que debe ser transmitida a los altavoces. Los formatos de audio soportados pueden ser señal de múltiples canales, componentes Ambisonics de primer orden y de orden superior, y objetos de audio. Una escena de audio compleja también se puede describir por medio de la combinación de diferentes formatos de entrada. Todos los formatos de audio se transmiten a continuación al análisis DirAC 180, que extrae una representación paramétrica de la escena de audio completa. Una dirección de llegada y una difusividad medida por unidad de tiempo-frecuencia forma los parámetros. El análisis DirAC es seguido por un codificador de metadatos espaciales 190, que cuantifica y codifica los parámetros DirAC para obtener una representación paramétrica de baja tasa de bits.

[0017] Junto con los parámetros, una señal de mezcla descendente derivada 160 de las diferentes fuentes o señales de entrada de audio se codifica para la transmisión por un codificador de núcleo de audio convencional 170. En este caso, un codificador de audio basado en EVS se adopta para la codificación de la señal de mezcla descendente. La señal de mezcla descendente consiste en diferentes canales, llamados canales de transporte: la señal puede ser, por ej., las cuatro señales de coeficientes que componen una señal de formato B, un par estéreo o una mezcla descendente monofónica dependiendo de la tasa de bits objetivo. Los parámetros espaciales codificados y la corriente de bits de audio codificada se multiplexan antes de ser transmitidos a través del canal de comunicación.

[0018] La Fig. 10 es un decodificador de la codificación de audio espacial basada en DirAC que entrega diferentes formatos de audio. En el decodificador, que se muestra en la Fig. 10, los canales de transporte son decodificados por el decodificador de núcleo 1020, mientras que los metadatos DirAC primero se decodifican 1060 antes de ser transportados con los canales de transporte decodificados a la síntesis DirAC 220, 240. En esta etapa (1040), se pueden considerar diferentes opciones. Se puede solicitar reproducir la escena de audio directamente en las configuraciones de altavoces o auriculares como suele ser posible en un sistema DirAC convencional (MC en la Fig. 10). Además, también se puede solicitar renderizar la escena a un formato Ambisonics para otras manipulaciones adicionales, tales como la rotación, la reflexión o el movimiento de la escena (FOA/HOA en la Fig. 10). Finalmente, el decodificador puede entregar los objetos individuales tal como se presentaron en el lado del codificador (Objetos en la Fig. 10).

[0019] Los objetos de audio también pueden ser restituidos pero es más interesante para el oyente ajustar la mezcla dictada por la manipulación interactiva de los objetos. Las manipulaciones de objetos habituales son el ajuste de nivel, la ecualización o la ubicación espacial del objeto. La mejora del diálogo basado en objetos se vuelve, por ejemplo, una posibilidad propuesta por esta característica de interactividad. Por último, es posible dar salida a los formatos originales como se presentaron en la entrada del codificador.

[0020] En este caso, podría ser una combinación de canales de audio y objetos o Ambisonics y objetos. Con el fin de lograr la transmisión separada de múltiples canales y componentes Ambisonics, se podrían utilizar varias instancias del sistema descrito.

[0021] La presente invención es ventajosa en que, en especial de acuerdo con el primer aspecto, se establece un marco con el fin de combinar diferentes descripciones de la escena en una escena de audio combinada por medio de un formato común, que permite combinar las diferentes descripciones de escenas de audio.

[0022] Este formato común puede, por ejemplo, ser el formato B o puede ser el formato de representación de la señal de presión/velocidad, o, con preferencia, también puede ser el formato de representación de parámetros DirAC.

[0023] Este formato es un formato compacto que, además, permite una cantidad significativa de interacción por parte del usuario, por una parte, y, por otra parte, es útil con respecto a una tasa de bits requerida para la representación de una señal de audio.

[0024] De acuerdo con un aspecto adicional de la presente invención, una síntesis de una pluralidad de escenas de audio se puede llevar a cabo de manera ventajosa por medio de la combinación de dos o más descripciones DirAC diferentes. Ambas de estas diferentes descripciones DirAC se pueden procesar por medio de la combinación de las escenas en el dominio de parámetro o, de manera alternativa, por medio de la renderización por separado de cada escena de audio y, a continuación, por medio de la combinación de las escenas de audio que se han renderizado de las descripciones DirAC individuales en el dominio espectral o, de manera alternativa, ya en el dominio temporal.

[0025] Este procedimiento permite un procesamiento muy eficiente y sin embargo de alta calidad de diferentes escenas de audio que van a ser combinadas en una sola representación de escena y, en particular, una señal de audio de dominio temporal único.

[0026] Un aspecto adicional de la invención es ventajoso en que un conjunto de datos de audio en particular útil convertido para la conversión de metadatos de objetos en metadatos DirAC se deriva, donde este conversor de datos de audio se puede utilizar en el marco del primer, el segundo o el tercer aspecto o también se pueden aplicar de manera independiente unos de otros. El conversor de datos de audio permite convertir de manera eficiente los datos de objetos de audio, por ejemplo, una señal de forma de onda para un objeto de audio, y la posición correspondiente de datos, de manera típica, con respecto al tiempo para la representación de una cierta trayectoria de un objeto de audio dentro de una configuración de reproducción en una descripción de escenas de audio muy útil y compacta, y, en particular, el formato de descripción de escenas de audio DirAC. Aunque una descripción de objeto de audio típico con una señal de forma de onda de objeto de audio y metadatos de posición del objeto de audio está relacionada con una configuración de reproducción particular o, en general, está relacionada con un determinado sistema de coordenadas de reproducción, la descripción DirAC es en particular útil porque está relacionada con una posición de oyente o micrófono y está completamente libre de cualquier limitación con respecto a la configuración de un altavoz o una configuración de reproducción.

[0027] Por lo tanto, la descripción DirAC generada a partir de señales de metadatos de objetos de audio, además, permite una combinación muy útil y compacta y de alta calidad de objetos de audio diferente de otras tecnologías de combinación de objeto de audio tal como la codificación de objetos de audio espacial o por medio del paneo de amplitud de los objetos en una configuración de reproducción.

[0028] Un codificador de escenas de audio de acuerdo con un aspecto adicional de la presente invención es en particular útil para el suministro de una representación combinada de una escena de audio que tiene metadatos DirAC y, de manera adicional, un objeto de audio con metadatos de objetos de audio.

[0029] En particular, en esta situación, es en particular útil y ventajoso para una alta interactividad con el fin de generar una descripción combinada de metadatos que tienen metadatos DirAC por un lado y, en paralelo, metadatos de objeto por otro lado. De este modo, en este aspecto, los metadatos de objetos no se combinan con los metadatos DirAC, sino que se convierten en metadatos similares a DirAC de tal manera que los metadatos de objetos comprendan la dirección o, de manera adicional, una distancia y/o una difusividad del objeto individual junto con la señal de objeto. Por lo tanto, la señal de objeto se convierte en una representación similar a DirAC de tal manera que se permita un manejo muy flexible de una representación DirAC para una primera escena de audio y un objeto adicional dentro de esta primera escena de audio se hace posible. De este modo, por ejemplo, los objetos

específicos se pueden procesar de manera muy selectiva debido a que todavía está disponible su canal de transporte correspondiente, por una parte, y los parámetros de estilo DirAC, por otra parte.

[0030] De acuerdo con un aspecto adicional de la invención, un aparato o un procedimiento para la realización de una síntesis de datos de audio es en particular útil porque se proporciona un manipulador para la manipulación de una descripción DirAC de uno o más objetos de audio, una descripción DirAC de la señal de múltiples canales o una descripción DirAC de señales Ambisonics de primer orden o señales Ambisonics de orden superior. Y, la descripción DirAC manipulada se sintetiza a continuación, por el uso de un sintetizador DirAC.

[0031] Este aspecto tiene la ventaja particular de que cualquier manipulación específica con respecto a cualquier señal de audio se lleva a cabo de manera muy útil y eficiente en el dominio DirAC, es decir, por medio de la manipulación ya sea del canal de transporte de la descripción DirAC o de manera alternativa por medio de la manipulación de los datos paramétricos de la descripción DirAC. Esta modificación es sustancialmente más eficiente y más práctica para llevar a cabo en el dominio DirAC en comparación con la manipulación en otros dominios. En particular, las operaciones de ponderación dependientes de la posición como operaciones de manipulación preferidas se pueden llevar a cabo en particular en el dominio DirAC. Por lo tanto, en una realización específica, una conversión de una representación de la señal correspondiente en el dominio DirAC y, a continuación, la realización de la manipulación dentro del dominio DirAC es un escenario de aplicación en particular útil para el procesamiento y la manipulación de escenas de audio modernas.

[0032] Las realizaciones preferidas se discuten posteriormente con respecto a las figuras que se acompañan, en las cuales:

La Fig. 1a es un diagrama de bloques de una implementación preferida de un aparato o un procedimiento para la generación de una descripción de una escena de audio combinada de acuerdo con un primer aspecto de la invención;
La Fig. 1b es una implementación de la generación de una escena de audio combinada, donde el formato común es la representación de presión/velocidad;
La Fig. 1c es una implementación preferida de la generación de una escena de audio combinada, donde los parámetros DirAC y la descripción DirAC es el formato común;
La Fig. 1d es una implementación preferida del combinador en la Fig. 1c que ilustra dos alternativas diferentes para la implementación del combinador de parámetros DirAC de diferentes escenas de audio o descripciones de escenas de audio;
La Fig. 1e es una implementación preferida de la generación de una escena de audio combinada donde el formato común es el formato B como un ejemplo para una representación Ambisonics;
La Fig. 1f es una ilustración de un conversor de objeto de audio/DirAC útil en el contexto de, por ejemplo, la Fig. 1c o 1d o útil en el contexto del tercer aspecto relativo a un conversor de metadatos;
La Fig. 1g es una ilustración de ejemplo de una señal de múltiples canales 5.1 en una descripción DirAC;
La Fig. 1h es una ilustración adicional de la conversión de un formato de múltiples canales en el formato DirAC en el contexto de un codificador y un lado del decodificador;
La Fig. 2a ilustra una realización de un aparato o un procedimiento para la realización de una síntesis de una pluralidad de escenas de audio de acuerdo con un segundo aspecto de la presente invención;
La Fig. 2b ilustra una implementación preferida del sintetizador DirAC de la Fig. 2a;
La Fig. 2c ilustra una implementación adicional del sintetizador DirAC con una combinación de señales renderizadas;
La Fig. 2d ilustra una implementación de un manipulador selectivo conectado ya sea antes del combinador de escenas 221 de la Fig. 2b o antes del combinador 225 de la Fig. 2c;
La Fig. 3a es una implementación preferida de un aparato o un procedimiento para la realización y la conversión de datos de audio de acuerdo con un tercer aspecto de la presente invención;
La Fig. 3b es una implementación preferida del conversor de metadatos también ilustrado en la Fig. 1f;
La Fig. 3c es un diagrama de flujo para la realización de una implementación adicional de una conversión de datos de audio a través del dominio de presión/velocidad;
La Fig. 3d ilustra un diagrama de flujo para llevar a cabo una combinación dentro del dominio DirAC;
La Fig. 3e ilustra una implementación preferida para la realización de diferentes descripciones DirAC, por ejemplo, de acuerdo con lo ilustrado en la Fig. 1d con respecto al primer aspecto de la presente invención;
La Fig. 3f ilustra la conversión de un dato de posición del objeto en una representación paramétrica DirAC;
La Fig. 4a ilustra una implementación preferida de un codificador de escenas de audio de acuerdo con un cuarto aspecto de la presente invención para la generación de una descripción de metadatos combinada que comprende los metadatos DirAC y los metadatos de objetos;
La Fig. 4b ilustra una realización preferida con respecto al cuarto aspecto de la presente invención;
La Fig. 5a ilustra una implementación preferida de un aparato para la realización de una síntesis de datos de audio o un procedimiento correspondiente de acuerdo con un quinto aspecto de la presente invención;
La Fig. 5b ilustra una implementación preferida del sintetizador DirAC de la Fig. 5A;
La Fig. 5c ilustra una alternativa adicional del procedimiento del manipulador de la Fig. 5A;
La Fig. 5d ilustra un procedimiento adicional para la implementación del manipulador de la Fig. 5A;

La Fig. 6 ilustra un conversor de señales de audio para la generación, a partir de una señal mono y una dirección de la información de llegada, es decir, a partir de una descripción DirAC de ejemplo, donde la difusividad, por ejemplo, se ajusta en cero, una representación en formato B que comprende un componente omnidireccional y componentes direccionales en X, Y y Z;
La Fig. 7a ilustra una implementación de un análisis DirAC de una señal de formato B de micrófono;
La Fig. 7b ilustra una implementación de una síntesis DirAC de acuerdo con un procedimiento conocido;
La Fig. 8 ilustra un diagrama de flujo para la ilustración de realizaciones adicionales de, en particular, la realización de la Fig. 1a;
La Fig. 9 es el lado del codificador de la codificación de audio espacial basada en DirAC que soporta diferentes formatos de audio;
La Fig. 10 es un decodificador de la codificación de audio espacial basada en DirAC que entrega diferentes formatos de audio;
La Fig. 11 es una vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada en un formato B combinado;
La Fig. 12 es una vista general del sistema del codificador/decodificador basado en DirAC que combina en el dominio de presión/velocidad;
La Fig. 13 es una vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada en el dominio DirAC con la posibilidad de la manipulación de objetos en el lado del decodificador;
La Fig. 14 es una vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada en el lado del decodificador a través de un combinador de metadatos DirAC;
La Fig. 15 es una vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada en el lado del decodificador en la síntesis DirAC; y
La Fig. 16a-f ilustra varias representaciones de formatos de audio útiles en el contexto del primer al quinto aspecto de la presente invención.

[0033] La Fig. 1a ilustra una realización preferida de un aparato para la generación de una descripción de una escena de audio combinada. El aparato comprende una interfaz de entrada 100 para la recepción de una primera descripción de una primera escena en un primer formato y una segunda descripción de una segunda escena en un segundo formato, en el que el segundo formato es diferente del primer formato. El formato puede ser cualquier formato de escenas de audio tal como cualquiera de los formatos o descripciones de escenas ilustradas en las Figs. 16a a 16f.

[0034] La Fig. 16a, por ejemplo, ilustra una descripción de objeto que consiste, de manera típica, en un señal de forma de onda de un objeto (codificado) 1 tal como un canal mono y metadatos correspondientes relacionados con la posición del objeto 1, donde esta información de manera típica se da para cada marco de tiempo o un grupo de marcos de tiempo, y cuya señal de forma de onda del objeto 1 se codifica. Las representaciones correspondientes para un segundo o más objetos pueden ser incluidas de acuerdo con lo ilustrado en la Fig. 16a.

[0035] Otra alternativa puede ser una descripción de objeto que consiste en una mezcla descendente de objetos que es una señal mono, una señal estéreo con dos canales o una señal con tres o más canales y metadatos de objetos tales como energías de objeto, información de correlación por compartimento de tiempo/frecuencia y, de manera opcional, las posiciones del objeto. Sin embargo, las posiciones del objeto también se pueden dar en el lado del decodificador como información de renderización típica y, por lo tanto, pueden ser modificados por un usuario. El formato en la Fig. 16b se puede implementar, por ejemplo, como el formato SAOC (codificación de objeto de audio espacial) bien conocido.

[0036] Otra descripción de una escena se ilustra en la Fig. 16c como una descripción de múltiples canales que tiene una representación codificada y no codificada de un primer canal, un segundo canal, un tercer canal, un cuarto canal, o un quinto canal, donde el primer canal puede ser el canal izquierdo L, el segundo canal puede ser el canal derecho R, el tercer canal puede ser el canal central C, el cuarto canal puede ser el canal envolvente izquierdo LS y el quinto canal puede ser el canal envolvente derecho RS. Naturalmente, la señal de múltiples canales puede tener un número menor o mayor de canales, tal como sólo dos canales para un canal estéreo o seis canales para un formato de 5.1 u ocho canales para un formato 7.1, etc.

[0037] Una representación más eficiente de una señal de múltiples canales se ilustra en la Fig. 16d, en la que la mezcla descendente de canal tal como una mezcla descendente mono, o una mezcla descendente estéreo o una mezcla descendente con más de dos canales se asocia con información lateral paramétrica como metadatos de canal para, de manera típica, cada compartimento de tiempo y/o frecuencia. Tal representación paramétrica se puede implementar, por ejemplo, de acuerdo con el estándar de sonido envolvente MPEG.

[0038] Otra representación de una escena de audio puede, por ejemplo, ser el formato B que consiste en una señal de omnidireccional W, y componentes direccionales X, Y, Z de acuerdo con lo mostrado en la Fig. 16e. Esta sería una señal de primer orden o FoA. Una señal Ambisonics de orden superior, es decir, una señal HoA puede tener componentes adicionales de acuerdo con lo conocido en la técnica.

[0039] La representación de la Fig. 16e es, en contraste con la representación de la Fig. 16c y la Fig. 16d una representación que no es dependiente de una cierta configuración de altavoz, sino que describe un campo de sonido de acuerdo con lo experimentado en una posición determinada (micrófono o el oyente).

[0040] Otra de tal descripción del campo de sonido es el formato DirAC, por ejemplo, de acuerdo con lo ilustrado en la Fig. 16f. El formato DirAC de manera típica comprende una señal de mezcla descendente DirAC que es una señal de mezcla descendente mono o estéreo o cualesquiera o una señal de transporte y la correspondiente información lateral paramétrica. Esta información lateral paramétrica es, por ejemplo, una dirección de información de llegada por compartimento de tiempo/frecuencia y, de manera opcional, información de difusividad por compartimento de tiempo/frecuencia.

[0041] La entrada en la interfaz de entrada 100 de la Fig. 1a puede ser, por ejemplo, en cualquiera de esos formatos ilustrados con respecto a la Fig. 16a a la Fig. 16f. La interfaz de entrada 100 reenvía las descripciones de formato correspondientes a un conversor de formatos 120. El conversor de formatos 120 está configurado para convertir la primera descripción en un formato común y para convertir la segunda descripción en el mismo formato común, cuando el segundo formato es diferente del formato común. Cuando, sin embargo, el segundo formato ya está en el formato común, entonces el conversor de formatos solamente convierte la primera descripción en el formato común, dado que la primera descripción ya está en un formato diferente del formato común.

[0042] Por lo tanto, en la salida del conversor de formatos o, en general, en la entrada de un combinador de formatos, sí existe una representación de la primera escena en el formato común y la representación de la segunda escena en el mismo formato común. Debido al hecho de que ambas descripciones ahora están incluidas en uno y el mismo formato común, el combinador de formatos ahora puede combinar la primera descripción y la segunda descripción para obtener una escena de audio combinada.

[0043] De acuerdo con una realización ilustrada en la Fig. 1e, el conversor de formatos 120 está configurado para convertir la primera descripción en una primera señal de formato B, por ejemplo, de acuerdo con lo ilustrado en 127 en la Fig. 1e y para calcular la representación de formato B para la segunda descripción de acuerdo con lo ilustrado en la Fig. 1e en 128.

[0044] A continuación, el combinador de formatos 140 se implementa como un sumador de componentes de señal ilustrado en 146a para el componente sumador W, 146b para el componente sumador X, ilustrado en 146c para el componente sumador Y e ilustrado en 146d para el componente sumador Z.

[0045] Por lo tanto, en la realización de la Fig. 1e, la escena de audio combinada puede ser una representación de formato B y las señales en formato B a continuación, pueden funcionar como los canales de transporte y entonces se pueden codificar a través de un codificador del canal de transporte 170 de la Fig. 1a. Por lo tanto, la escena de audio combinada con respecto a la señal de formato B puede ser directamente de entrada en el codificador 170 de la Fig. 1a para generar una señal de formato B codificada, que podrían salir a continuación a través de la interfaz de salida 200. En este caso, no es necesario ningún metadato espacial, pero, al precio de una representación codificada de cuatro señales de audio, es decir, el componente omnidireccional W y los componentes direccionales X, Y, Z.

[0046] De manera alternativa, el formato común es el formato de presión/velocidad, de acuerdo con lo ilustrado en la Fig. 1b. Para este fin, el conversor de formatos 120 comprende un analizador de tiempo/frecuencia 121 para la primera escena de audio y el analizador de tiempo/frecuencia 122 para la segunda escena de audio o, en general, la escena de audio con el número N, donde N es un número entero.

[0047] A continuación, para cada dicha representación espectral generada por los conversores espectrales 121, 122, la presión y la velocidad se calculan de acuerdo con lo ilustrado en 123 y 124, y, el combinador de formatos a continuación, está configurado para calcular una señal de presión sumada por un lado, por medio de la suma de las correspondientes señales de presión generadas por los bloques 123, 124. Y, de manera adicional, una señal de velocidad individual se calcula también por cada uno de los bloques 123, 124 y las señales de velocidad se pueden sumar juntas con el fin de obtener una señal de presión/velocidad combinada.

[0048] Dependiendo de la implementación, los procedimientos en los bloques 142, 143 no necesariamente se tienen que llevar a cabo. En cambio, la señal de presión combinada o “sumada” y la señal de velocidad combinada o “sumada” se pueden codificar en una analogía de acuerdo con lo ilustrado en la Fig. 1e de la señal de formato B y esta representación de presión/velocidad se podría codificar si bien una vez más a través de ese codificador 170 de la Fig. 1a y se podría transmitir a continuación al decodificador sin ninguna información lateral adicional con respecto a los parámetros espaciales, dado que la representación de presión/velocidad combinada ya incluye la información espacial necesaria para la obtención de un campo de sonido de alta calidad finalmente renderizado en un lado del decodificador.

[0049] En una realización, sin embargo, se prefiere llevar a cabo un análisis DirAC a la representación de

presión/velocidad generada por el bloque 141. Con este fin, se calcula el vector de intensidad 142 y, en el bloque 143, se calculan los parámetros DirAC desde el vector de intensidad y, a continuación, los parámetros DirAC combinados se obtienen como una representación paramétrica de la escena de audio combinada. Con este fin, el analizador DirAC 180 de la Fig. 1a se implementa para llevar a cabo la funcionalidad del bloque 142 y 143 de la Fig. 1b. Y, con preferencia, los datos DirAC se someten además a una operación de codificación de metadatos en el codificador de metadatos 190. El codificador de metadatos 190 comprende de manera típica un cuantificador y codificador de entropía con el fin de reducir la tasa de bits requerida para la transmisión de los parámetros DirAC.

[0050] Junto con los parámetros DirAC codificados, también se transmite un canal de transporte codificado. El canal de transporte codificado se genera por el generador del canal de transporte 160 de la Fig. 1a que se puede implementar, por ejemplo, de acuerdo con lo ilustrado en la Fig. 1b por un primer generador de mezcla descendente 161 para la generación de una mezcla descendente de la primera escena de audio y un N-ésimo generador de mezcla descendente 162 para la generación de una mezcla descendente de la N-ésima escena de audio.

[0051] A continuación, los canales de mezcla descendente se combinan en el combinador 163 de manera típica por una adición directa y la señal de mezcla descendente combinada es entonces el canal de transporte que es codificado por el codificador 170 de la Fig. 1a. La mezcla descendente combinada puede ser, por ejemplo, un par estéreo, es decir, un primer canal y un segundo canal de una representación estéreo o puede ser un canal mono, es decir, una sola señal de canal.

[0052] De acuerdo con una realización adicional que se ilustra en la Fig. 1c, una conversión de formato en el conversor de formatos 120 se hace para convertir directamente cada uno de los formatos de audio de entrada en el formato DirAC como el formato común. Para este fin, el conversor de formatos 120 forma de nuevo una conversión de tiempo-frecuencia o un análisis de tiempo/frecuencia en los correspondientes bloques 121 para la primera escena y el bloque 122 para una segunda escena o una adicional. A continuación, los parámetros DirAC se derivan de las representaciones espectrales de las escenas de audio correspondientes ilustradas en 125 y 126. El resultado del procedimiento en los bloques 125 y 126 son parámetros DirAC que consisten en información de energía por mosaico de tiempo/frecuencia, una dirección de información de llegada e_{DOA} por mosaico de tiempo/frecuencia y una información de difusividad ψ de cada mosaico de tiempo/frecuencia. A continuación, el combinador de formatos 140 está configurado para llevar a cabo una combinación directamente en el dominio de parámetros DirAC con el fin de generar parámetros DirAC combinados Ψ para la difusividad y e_{DOA} para la dirección de llegada. En particular, la información de energía E_1 y E_N es requerida por el combinador 144, pero no forma parte de la representación paramétrica combinada final generada por el combinador de formatos 140.

[0053] Por lo tanto, la comparación de la Fig. 1c a la Fig. 1e revela que, cuando el combinador de formatos 140 ya lleva a cabo una combinación en el dominio de parámetros DirAC, el analizador DirAC 180 no es necesario y no se implementa. En lugar de ello, la salida del combinador de formatos 140 que es la salida del bloque 144 en la Fig. 1c se reenvía directamente al codificador de metadatos 190 de la Fig. 1a y desde allí a la interfaz de salida 200 de tal manera que los metadatos espaciales codificados y, en particular, los parámetros DirAC combinados codificados estén incluidos en la salida de la señal de salida codificada por la interfaz de salida 200.

[0054] Además, el generador del canal de transporte 160 de la Fig. 1a puede recibir, ya desde la interfaz de entrada 100, una representación de la señal de forma de onda para la primera escena y la representación de la señal de forma de onda para la segunda escena. Estas representaciones se introducen en los bloques generadores de mezcla descendente 161, 162 y los resultados se suman en el bloque 163 para obtener una mezcla descendente combinada de acuerdo con lo ilustrado con respecto a la Fig. 1b.

[0055] La Fig. 1d ilustra una representación similar con respecto a la Fig. 1c. Sin embargo, en la Fig. 1d, la forma de onda de objeto de audio se introduce en el conversor de representación de tiempo/frecuencia 121 para el objeto de audio 1 y 122 para el objeto de audio N. Además, los metadatos se introducen, junto con la representación espectral en los calculadores de parámetros DirAC 125, 126 de acuerdo con lo ilustrado también en la Fig. 1c.

[0056] Sin embargo, la Fig. 1D proporciona una representación más detallada con respecto a cómo operan las implementaciones preferidas del combinador 144. En una primera alternativa, el combinador lleva a cabo una suma ponderada de energía de la difusividad individual para cada objeto o escena individual y, un cálculo ponderado por energía correspondiente de una DoA combinada para cada mosaico de tiempo/frecuencia se lleva a cabo de acuerdo con lo ilustrado en la ecuación inferior de alternativa 1.

[0057] Sin embargo, también se pueden llevar a cabo otras implementaciones. En particular, otro cálculo muy eficiente establece la difusividad en cero para los metadatos DirAC combinados y para seleccionar, como la dirección de llegada para cada mosaico de tiempo/frecuencia la dirección de llegada calculada a partir de un objeto de audio determinado que tiene la energía más alta dentro del mosaico de tiempo/frecuencia específico. Con preferencia, el procedimiento en la Fig. 1d es más apropiado cuando la entrada en la interfaz de entrada son objetos de audio individuales que de manera correspondiente representa una forma de onda o señal mono para cada objeto y metadatos correspondientes, tal como información de posición ilustrada con respecto a la Fig. 16a o 16b.

[0058] Sin embargo, en la realización de la Fig. 1c, la escena de audio puede ser cualquier otra de las representaciones ilustradas en las Figs. 16c, 16d, 16e o 16f. Entonces, puede haber metadatos o no, es decir, los metadatos en la Fig. 1c son opcionales. Entonces, sin embargo, una difusividad típicamente útil se calcula para una descripción de escena determinada tal como una descripción de escena Ambisonics en la Fig. 16e y, a continuación, la primera alternativa de la manera en que se combinan los parámetros se prefiere sobre la segunda alternativa de la Fig. 1d. Por lo tanto, de acuerdo con la invención, el conversor de formatos 120 está configurado para convertir un formato Ambisonics de orden superior o Ambisonics de primer orden en el formato B, en el que el formato Ambisonics de orden superior se trunca antes de ser convertido en el formato B.

[0059] En una realización adicional, el conversor de formatos está configurado para proyectar un objeto o un canal en armónicos esféricos en la posición de referencia para obtener las señales proyectadas, y en el que el combinador de formatos está configurado para combinar las señales de proyección para obtener coeficientes en formato B, en los que el objeto o el canal está situado en el espacio en una posición especificada y tiene una distancia individual opcional desde una posición de referencia. Este procedimiento funciona bien en particular para la conversión de señales de objetos o señales de múltiples canales en señales Ambisonics de primer orden o de orden superior.

[0060] En una alternativa adicional, el conversor de formatos 120 está configurado para llevar a cabo un análisis DirAC que comprende un análisis de tiempo-frecuencia de los componentes en formato B y una determinación de los vectores de presión y velocidad y en el que el combinador de formatos está configurado entonces para la combinación de diferentes vectores de presión/velocidad y donde el combinador de formatos comprende además el analizador DirAC 180 para derivar metadatos DirAC de los datos de presión/velocidad combinados.

[0061] En una realización alternativa adicional, el conversor de formatos está configurado para extraer los parámetros DirAC directamente de los metadatos de objetos de un formato de objeto de audio como el primer o el segundo formato, en el que el vector de presión para la renderización DirAC es la señal de forma de onda de objeto y la dirección se deriva de la posición del objeto en el espacio o la difusividad está directamente dada en los metadatos de objetos o se establece en un valor predeterminado tal como el valor cero.

[0062] En una realización adicional, el conversor de formatos está configurado para convertir los parámetros DirAC derivados del formato de datos de objeto en los datos de presión/velocidad y el combinador de formatos está configurado para combinar los datos de presión/velocidad con los datos de presión/velocidad derivados de diferentes descripciones de uno o más objetos de audio diferentes.

[0063] Sin embargo, en una implementación preferida que se ilustra con respecto a la Fig. 1c y 1d, el combinador de formatos está configurado para combinar directamente los parámetros DirAC derivados por el conversor de formatos 120 de tal manera que la escena de audio combinada generada por el bloque 140 de la Fig. 1a sea ya el resultado final y un analizador DirAC 180 que se ilustra en la Fig. 1a no sea necesario, dado que la salida de datos por el combinador de formatos 140 ya está en el formato DirAC.

[0064] En una implementación adicional, el conversor de formatos 120 ya comprende un analizador DirAC para un formato de entrada Ambisonics de primer orden o Ambisonics de orden superior o un formato de señal de múltiples canales. Además, el conversor de formatos comprende un conversor de metadatos para la conversión de los metadatos de objetos en metadatos DirAC, y un conversor de tales metadatos se ilustra, por ejemplo, en la Fig. 1f en 150 que opera una vez más en el análisis de tiempo/frecuencia en el bloque 121 y calcula la energía por banda por marco de tiempo que se ilustra en 147, la dirección de llegada se ilustra en el bloque 148 de la Fig. 1f y la difusividad se ilustra en el bloque 149 de la Fig. 1f. Y, los metadatos son combinados por el combinador 144 para la combinación de las corrientes de metadatos DirAC individuales, con preferencia por medio de una suma ponderada de acuerdo con lo ilustrado a modo de ejemplo por una de las dos alternativas de la realización de la Fig. 1d.

[0065] Las señales de canal de múltiples canales se pueden convertir directamente al formato B. El formato B obtenido puede ser procesado entonces por un DirAC convencional. La Fig. 1g ilustra una conversión 127 a formato B y un posterior procesamiento DirAC 180.

[0066] La referencia [3] describe maneras de llevar a cabo la conversión de la señal de múltiples canales a formato B. En principio, la conversión de señales de audio de múltiples canales a formato B es simple: se definen altavoces virtuales para estar en diferentes posiciones de diseño de altavoces. Por ejemplo, para un diseño 5.0, los altavoces se posicionan en el plano horizontal en ángulos de azimut +/- 30 y +/- 110 grados. Un micrófono de formato B virtual se define a continuación, para estar en el centro de los altavoces, y se lleva a cabo una grabación virtual. Por lo tanto, el canal W se crea por medio de la suma de todos los canales de altavoces del archivo de audio 5.0. El proceso para obtener W y otros coeficientes en formato B se puede sintetizar a continuación:

$$W = \sum_{i=1}^{k} \sqrt{\frac{1}{2}} w_i s_i$$

$$X = \sum_{i=1}^{k} w_i s_i (\cos(\theta_i) \cos(\varphi_i))$$

$$Y = \sum_{i=1}^{k} w_i s_i (\sin(\theta_i) \cos(\varphi_i))$$

$$Z = \sum_{i=1}^{k} w_i s_i (\sin(\varphi_i))$$

donde s_i son las señales de múltiples canales situadas en el espacio en las posiciones de altavoces definidas por el ángulo de azimut θ_i y el ángulo de elevación φ_i, de cada altavoz y w_i son pesos en función de la distancia. Si la distancia no está disponible o simplemente se ignora, entonces $w_i = 1$. Sin embargo, esta técnica sencilla es limitada, dado que es un proceso irreversible. Además, dado que los altavoces por lo general están distribuidos de manera no uniforme, también hay un sesgo en la estimación realizada por un análisis DirAC posterior hacia la dirección con la densidad de altavoz más alta. Por ejemplo, en el diseño 5.1, habrá un sesgo hacia la parte delantera, dado que hay más altavoces en la parte delantera que en la parte posterior.

[0067] Para hacer frente a este problema, se propuso una técnica adicional en [3] para el procesamiento de la señal de múltiples canales 5.1 con DirAC. El esquema de codificación final tendrá un aspecto entonces de acuerdo con lo ilustrado en la Fig. 1h que muestra el conversor de formato B 127, el analizador DirAC 180 de acuerdo con lo descrito por lo general con respecto al elemento 180 en la Fig. 1, y los otros elementos 190, 1000, 160, 170, 1020, y/o 220, 240.

[0068] En una realización adicional, la interfaz de salida 200 está configurada para sumar, al formato combinado, una descripción de objeto separada para un objeto de audio, donde la descripción de objeto comprende al menos uno de una dirección, una distancia, una difusividad o cualquier otro atributo de objeto, donde este objeto tiene una sola dirección a través de todas las bandas de frecuencia y es ya estático o está en movimiento a un ritmo más lento que un umbral de velocidad.

[0069] Esta característica se elabora además en más detalle con respecto al cuarto aspecto de la presente invención descrito con respecto a la Fig. 4a y Fig. 4b.

Primera alternativa de codificación: combinación y procesamiento de diferentes representaciones de audio a través del formato B o una representación equivalente.

[0070] Una primera realización del codificador previsto se puede lograr por medio de la conversión de todos los formatos de entrada en un formato B combinado de acuerdo con lo representado en la Fig. 11.
Fig. 11: vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada en un formato B combinado

[0071] Dado que DirAC está diseñado originalmente para el análisis de una señal de formato B, el sistema convierte los distintos formatos de audio a una señal de formato B combinado. Los formatos se convierten primero de manera individual 120 en una señal de formato B antes de ser combinados juntos por medio de la suma de sus componentes en formato B W, X, Y, Z. Los componentes Ambisonics de Primer Orden (FOA) pueden ser normalizados y reordenados a un formato B. Suponiendo que FOA está en formato ACN/N3D, las cuatro señales de la entrada formato B se obtienen por medio de:

$$\begin{cases} W = Y_0^0 \\ X = \sqrt{\frac{2}{3}} Y_1^1 \\ Y = \sqrt{\frac{2}{3}} Y_1^{-1} \\ Z = \sqrt{\frac{2}{3}} Y_1^0 \end{cases}$$

[0072] Donde Y'_m denota el componente Ambisonics del orden *l* y el índice *m*, $-l \leq m \leq +l$. Dado que los componentes de FOA están totalmente contenidos en el formato Ambisonics de orden superior, el formato de HOA sólo necesita ser truncado antes de ser convertido en el formato B.

[0073] Dado que los objetos y los canales han determinado las posiciones en el espacio, es posible proyectar cada objeto individual y canal en armónicos esféricos (SH) en la posición central, tal como la grabación o la posición de referencia. La suma de las proyecciones permite combinar diferentes objetos y múltiples canales en un solo formato B y puede entonces ser procesada por el análisis DirAC. Los coeficientes en formato B (W, X, Y, Z) vienen dados a continuación por:

$$W = \sum_{i=1}^{k} \sqrt{\frac{1}{2}} w_i s_i$$

$$X = \sum_{i=1}^{k} w_i s_i (\cos(\theta_i) \cos(\varphi_i))$$

$$Y = \sum_{i=1}^{k} w_i s_i (\sin(\theta_i) \cos(\varphi_i))$$

$$Z = \sum_{i=1}^{k} w_i s_i (\sin(\varphi_i))$$

donde s_i son señales independientes situadas en el espacio en las posiciones definidas por el ángulo de azimut θ_i y el ángulo de elevación φ_i, y w_i son pesos en función de la distancia. Si la distancia no está disponible o simplemente se ignora, entonces $w_i = 1$. Por ejemplo, las señales independientes corresponden a objetos de audio que se encuentran en la posición dada o la señal asociada con un canal de altavoz en la posición especificada.

[0074] En aplicaciones donde se desea una representación Ambisonics de órdenes superiores al primer orden, la generación de coeficientes Ambisonics presentada con anterioridad para el primer orden se extiende por medio de la consideración adicional de componentes de orden superior.

[0075] El generador del canal de transporte 160 puede recibir directamente la señal de múltiples canales, señales de forma de onda de objeto, y componentes Ambisonics de orden superior. El generador del canal de transporte reducirá el número de canales de entrada que se van a transmitir por medio de la mezcla descendente de los mismos. Los canales se pueden mezclar juntos como en envolvente MPEG en una mezcla descendente mono o en estéreo, mientras que las señales de forma de onda de objeto se pueden sintetizar de una manera pasiva en una mezcla descendente mono. Además, a partir del Ambisonics de orden superior, es posible extraer una representación de orden inferior o crear por medio de formación de haces una mezcla descendente estéreo o cualquier otro seccionamiento del espacio. Si las mezclas descendentes obtenidas a partir del diferente formato de entrada son compatibles entre sí, se pueden combinar entre sí por medio de una simple operación de suma.

[0076] De manera alternativa, el generador de canal de transporte 160 puede recibir el mismo formato B combinado como el transmitido al análisis DirAC. En este caso, un subconjunto de los componentes o el resultado de una formación de haces (o de otro procesamiento) forman los canales de transporte que se van a codificar y

transmitir al decodificador. En el sistema propuesto, se requiere una codificación de audio convencional, que puede estar basada, pero no se limita, al códec estándar 3GPP EVS. 3GPP SVE es la elección de códec preferida debido a su capacidad para codificar señales de habla o música a bajas tasas de bits de alta calidad mientras que requiere un retraso relativamente bajo que permite la comunicación en tiempo real.

[0077] A una tasa de bits muy baja, el número de canales que se va a transmitir se debe limitar a uno y, por lo tanto, sólo se transmite la señal de micrófono omnidireccional W del formato B. Si la tasa de bits lo permite, el número de canales de transporte se puede aumentar por medio de la selección de un subconjunto de los componentes de formato B. De manera alternativa, las señales en formato B se pueden combinar en un formador de haces 160 dirigido a las particiones específicas del espacio. A modo de ejemplo dos cardioides se pueden diseñar para señalar en direcciones opuestas, por ejemplo, a la izquierda y a la derecha de la escena espacial:

$$\begin{cases} L = \sqrt{2}W + Y \\ R = \sqrt{2}W - Y \end{cases}$$

[0078] Estos dos canales estéreo L y R pueden ser entonces codificados de manera eficiente 170 por una codificación de estéreo conjunta. Las dos señales serán entonces explotadas de manera adecuada por la Síntesis DirAC en el lado del decodificador para la representación de la escena de sonido. Se puede prever otra formación de haces, por ejemplo, se puede dirigir un micrófono cardioide virtual hacia cualquier dirección desde un azimut θ y elevación φ dados:

$$C = \sqrt{2}W + \cos(\theta)\cos(\varphi)\,X + \sin(\theta)\cos(\varphi)\,Y + \sin(\varphi)Z$$

[0079] Se pueden prever otras maneras de formar canales de transmisión que llevan más información espacial de lo que haría un solo canal de transmisión monofónico.
De manera alternativa, los 4 coeficientes del formato B se pueden transmitir directamente. En ese caso, los metadatos DirAC se pueden extraer directamente en el lado del decodificador, sin la necesidad de transmitir información adicional para los metadatos espaciales.

[0080] La Fig. 12 muestra otro procedimiento alternativo para la combinación de los diferentes formatos de entrada. La Fig. 12 también es una vista general del sistema del codificador/decodificador basado en DirAC que se combina en el dominio de presión/velocidad.

[0081] Ambos componentes de señal de múltiples canales y Ambisonics se introducen en un análisis DirAC 123, 124. Para cada formato de entrada se lleva a cabo un análisis DirAC que consiste en un análisis de tiempo-frecuencia de los componentes en formato B $w^i(n), x^i(n), y^i(n), z^i(n)$ y la determinación de los vectores de presión y velocidad:

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

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

donde *i* es el índice de la entrada y, *k* y *n* los índices de tiempo y frecuencia del mosaico de tiempo-frecuencia, y $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ representan los vectores unitarios cartesianos.

[0082] *P(n,k)* y *U(n,k)* son necesarios para calcular los parámetros DirAC, a saber DOA y difusividad. El combinador de metadatos DirAC puede explotar que *N* fuentes que se reproducen juntas dan como resultado una combinación lineal de sus presiones y velocidades de las partículas que se miden cuando se reproducen solas. Las cantidades combinadas se derivan entonces por:

$$P(n,k) = \sum_{i=1}^{N} P^i(n,k)$$

$$U(n,k) = \sum_{i=1}^{N} U^i(n,k)$$

[0083] Los parámetros DirAC combinados se calculan 143 a través del cálculo del vector de intensidad combinada:

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

donde (.) denota una conjugación compleja. La difusividad del campo de sonido combinado está dada por:

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

donde E{.} designa el operador de promediado temporal, *c* la velocidad del sonido y *E(k,n)* la energía del campo de sonido dada por: $\mathrm{E}\{.\}cE(k,n)$

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

[0084] La dirección de llegada (DOA) se expresa por medio del vector unitario, $e_{DOA}(k,n)$ definido como

$$e_{DOA}(k,n) = -\frac{I(k,n)}{||I(k,n)||}$$

[0085] Si se introduce un objeto de audio, los parámetros DirAC se pueden extraer directamente de los metadatos de objetos mientras que el vector de presión $P^i(k,n)$ es la señal de esencia del objeto (forma de onda). Más precisamente, la dirección se deriva de manera directa a partir de la posición del objeto en el espacio, mientras que la difusividad está directamente dada en los metadatos de objetos o (si no está disponible) se puede ajustar por defecto en cero. A partir de los parámetros DirAC, los vectores de presión y velocidad están directamente dados por:

$$\widehat{P}^i(k,n) = \sqrt{1-\psi^i(k,n)}P^i(k,n)$$

$$\widehat{U}^i(k,n) = -\frac{1}{\rho_0 c}\widehat{P}^i(k,n).e^i_{DOA}(k,n)$$

[0086] La combinación de objetos o la combinación de un objeto con diferentes formatos de entrada a continuación, se obtiene por medio de la suma de los vectores de presión y velocidad de acuerdo con lo explicado con anterioridad.

[0087] En síntesis, la combinación de diferentes contribuciones de entrada (Ambisonics, canales, objetos) se lleva a cabo en el dominio de presión/velocidad y el resultado se convierte entonces posteriormente en parámetros DirAC de dirección/difusividad. La operación en el dominio de presión/velocidad es teóricamente equivalente a operar en formato B. El principal beneficio de esta alternativa en comparación con la anterior es la posibilidad de optimizar el análisis DirAC de acuerdo con cada formato de entrada de acuerdo con lo propuesto en [3] para el formato de sonido envolvente 5.1.

[0088] El principal inconveniente de tal fusión en un formato B combinado o un dominio de presión/velocidad es que la conversión que ocurre en el extremo frontal de la cadena de procesamiento ya es un cuello de botella para todo el sistema de codificación. En efecto, la conversión de las representaciones de audio Ambisonics de orden superior, objetos o canales a una señal de formato B (de primer orden) ya engendra una gran pérdida de resolución espacial que no puede ser recuperada después.

Segunda Alternativa de Codificación: combinación y procesamiento en el dominio DirAC

[0089] Para sortear las limitaciones de la conversión de todos los formatos de entrada en una señal de formato B combinado, la presente alternativa propone derivar los parámetros DirAC directamente desde el formato original y, a continuación, combinarlos posteriormente en el dominio de parámetros DirAC. La vista general de un sistema de este tipo se da en la Fig. 13. La Fig. 13 es una vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada en el dominio DirAC con la posibilidad de la manipulación de objetos en el lado del decodificador.

[0090] En lo sucesivo, también podemos considerar canales individuales de una señal de múltiples canales como una entrada de objeto de audio para el sistema de codificación. Los metadatos de objetos son entonces estáticos en el tiempo y representan la posición del altavoz y la distancia en relación con la posición del oyente.

[0091] El objetivo de esta solución alternativa es evitar la combinación sistemática de los diferentes formatos de entrada en un formato B combinado o una representación equivalente. El objetivo es calcular los parámetros DirAC antes de combinarlos. El procedimiento evita entonces cualquier sesgo en la dirección y estimación de

difusividad debido a la combinación. Además, se pueden explotar de manera óptima las características de cada representación de audio durante el análisis DirAC o durante la determinación de los parámetros DirAC.

[0092] La combinación de los metadatos DirAC se produce después de determinar 125, 126, 126a los parámetros DirAC, la difusividad, la dirección, para cada formato de entrada, así como la presión contenida en los canales de transporte transmitidos. El análisis DirAC puede estimar los parámetros de un formato B intermedio, obtenidos por medio de la conversión del formato de entrada de acuerdo con lo explicado con anterioridad. Como alternativa, los parámetros DirAC se pueden estimar de manera ventajosa sin pasar por el formato B, sino directamente desde el formato de entrada, lo que podría mejorar aún más la precisión de la estimación. Por ejemplo, en [7], se propone estimar la difusividad directa de Ambisonics de orden superior. En el caso de objetos de audio, un simple conversor de metadatos 150 en la Fig. 15 puede extraer la dirección de metadatos de objeto y la difusividad para cada objeto.

[0093] La combinación 144 de las varias corrientes de metadatos DirAC en una única corriente de metadatos DirAC combinados se puede lograr de acuerdo con lo propuesto en [4]. Para algunos contenidos, es mucho mejor estimar directamente los parámetros DirAC desde el formato original en lugar de la conversión a un formato B combinado primero antes de llevar a cabo un análisis DirAC. En efecto, los parámetros, la dirección y la difusividad, pueden estar sesgados cuando va a un formato B [3] o durante la combinación de las diferentes fuentes. Además, esta alternativa permite un…

[0094] Otra alternativa más simple puede promediar los parámetros de las diferentes fuentes por medio de la ponderación en función de sus energías:

$$\psi(k,n) = \frac{1}{\sum_{i=1}^{N} E^i(k,n)} \sum_{i=1}^{N} E^i(k,n)\, \psi^i(k,n)$$

$$e_{DOA}(k,n) = \frac{1}{\sum_{i=1}^{N} (1-\psi^i(k,n)) E^i(k,n)} \sum_{i=1}^{N} (1-\psi^i(k,n)) E^i(k,n) e^i_{DOA}(k,n)$$

[0095] Para cada objeto todavía existe la posibilidad de enviar su propia dirección y de manera opcional la distancia, la difusividad o cualquier otro atributo del objeto relevante como parte de la corriente de bits transmitida desde el codificador al decodificador (véanse, por ej., las Figs. 4a, 4b). Esta información lateral adicional enriquecerá los metadatos DirAC combinados y permitirá que el decodificador restituya y/o manipule el objeto por separado. Dado que un objeto tiene una sola dirección a lo largo de todas las bandas de frecuencia y se puede considerar ya sea estático o en movimiento a un ritmo lento, la información adicional requiere ser actualizada con menos frecuencia que otros parámetros DirAC y sólo engendrará una tasa de bits adicional muy baja.

[0096] En el lado del decodificador, el filtrado direccional se puede llevar a cabo de acuerdo con lo enseñado en [5] para la manipulación de objetos. El filtrado direccional se basa en una técnica de atenuación espectral a corto plazo. Se lleva a cabo en el dominio espectral por una función de ganancia de fase cero, que depende de la dirección de los objetos.

[0097] La dirección puede estar contenida en la corriente de bits si las direcciones de los objetos se transmiten como información lateral. De lo contrario, la dirección también se podría dar de forma interactiva por el usuario.

Tercera alternativa: combinación del lado del decodificador

[0098] De manera alternativa, la combinación se puede llevar a cabo en el lado del decodificador. La Fig. 14 es una vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada en el lado de decodificador a través de un combinador de metadatos DirAC. En la Fig. 14, el esquema de codificación basado en DirAC funciona a tasas de bits más elevadas que con anterioridad, pero permite la transmisión de los metadatos DirAC individuales. Las diferentes corrientes de metadatos DirAC se combinan 144, por ejemplo, de acuerdo con lo propuesto en [4] en el decodificador antes de la síntesis DirAC 220, 240. El combinador de metadatos DirAC 144 también puede obtener la posición de un objeto individual para la posterior manipulación del objeto en el análisis DirAC.

[0099] La Fig. 15 es una vista general del sistema del codificador/decodificador basado en DirAC que combina diferentes formatos de entrada del lado del decodificador en la síntesis DirAC. Si la tasa de bits permite, el sistema además se puede mejorar de acuerdo con lo propuesto en la Fig. 15 por medio del envío para cada componente de entrada (FOA/HOA, MC, Objeto), de su propia señal de mezcla descendente junto con sus metadatos DirAC asociados. Incluso de este modo, las diferentes corrientes de DirAC comparten una síntesis DirAC

común 220, 240 en el decodificador para reducir la complejidad.

[0100] La Fig. 2a ilustra un concepto para la realización de una síntesis de una pluralidad de escenas de audio de acuerdo con un segundo aspecto adicional de la presente invención. Un aparato ilustrado en la Fig. 2a comprende una interfaz de entrada 100 para la recepción de una primera descripción DirAC de una primera escena y para la recepción de una segunda descripción DirAC de una segunda escena y uno o más canales de transporte.

[0101] Además, se proporciona un sintetizador DirAC 220 para la síntesis de la pluralidad de escenas de audio en un dominio espectral para obtener una señal de audio en el dominio espectral que representa la pluralidad de escenas de audio. Además, se proporciona un conversor de tiempo espectral 214 que convierte la señal de audio en el dominio espectral en un dominio temporal con el fin de emitir una señal de audio de dominio temporal que se puede emitir por altavoces, por ejemplo. En este caso, el sintetizador DirAC está configurado para llevar a cabo la renderización de la señal de salida de los altavoces. De manera alternativa, la señal de audio podría ser una señal estéreo que se puede emitir a un auricular. Una vez más, de manera alternativa, la salida de la señal de audio por el conversor de tiempo espectral 214 puede ser una descripción del campo de sonido en formato B. Todas estas señales, es decir, las señales de altavoz para más de dos canales, las señales de los auriculares o las descripciones de los campos de sonido son señales en el dominio del tiempo para su posterior procesamiento, tal como la salida por los altavoces o los auriculares, o para la transmisión o el almacenamiento en el caso de descripciones de campos de sonido, tal como las señales Ambisonics de primer orden o las señales Ambisonics de orden superior.

[0102] Además, el dispositivo de la Fig. 2a comprende de manera adicional una interfaz de usuario 260 para el control del sintetizador DirAC 220 en el dominio espectral. Además, uno o más canales de transporte se pueden proporcionar a la interfaz de entrada 100 que se van a utilizar junto con la primera y la segunda descripción DirAC que son, en este caso, las descripciones paramétricas que proporcionan, para cada mosaico de tiempo/frecuencia, una información de dirección de llegada y, de manera opcional y adicional, una información de difusividad.

[0103] De manera típica, la entrada de dos descripciones DirAC diferentes en la interfaz 100 en la Fig. 2a describen dos escenas de audio diferentes. En este caso, el sintetizador DirAC 220 está configurado para llevar a cabo una combinación de estas escenas de audio. Una alternativa de la combinación se ilustra en la Fig. 2b. En este caso, un combinador de escenas 221 está configurado para combinar las dos descripciones DirAC en el dominio paramétrico, es decir, se combinan los parámetros para obtener una dirección combinada de parámetros de llegada (DoA) y los parámetros de difusividad de manera opcional en la salida del bloque 221. Estos datos se introducen entonces en el renderizador DirAC 222 que recibe, de manera adicional, los uno o más canales de transporte con el fin de obtener la señal de audio en el dominio espectral 222. La combinación de los datos paramétricos DirAC se lleva a cabo preferentemente de acuerdo con lo ilustrado en la Fig. 1d y, de acuerdo con lo descrito con respecto a esta figura y, en particular, con respecto a la primera alternativa.

[0104] En el caso de que al menos uno de la entrada de dos descripciones en el combinador de escenas 221 incluya valores de difusividad de cero o no haya valores de difusividad en absoluto, entonces, de manera adicional, se puede aplicar la segunda alternativa, también de acuerdo con lo discutido en el contexto de la Fig. 1d.

[0105] Otra alternativa se ilustra en la Fig. 2c. En este procedimiento, las descripciones DirAC individuales se renderizan por medio de un primer renderizador DirAC 223 para la primera descripción y un segundo renderizador DirAC 224 para la segunda descripción y en la salida de los bloques 223 y 224, están disponibles una primera y la segunda señal de audio de dominio espectral, y estas primera y segunda señales de audio de dominio espectral se combinan dentro del combinador 225 para obtener, en la salida del combinador 225, una señal de combinación de dominio espectral.

[0106] A modo de ejemplo, el primer renderizador DirAC 223 y el segundo renderizador DirAC 224 están configurados para generar una señal estéreo que tiene un canal izquierdo L y un canal derecho R. Entonces, el combinador 225 está configurado para combinar el canal izquierdo desde el bloque 223 y el canal izquierdo desde el bloque 224 para obtener un canal izquierdo combinado. Además, se añade el canal derecho desde el bloque 223 con el canal derecho desde el bloque 224, y el resultado es un canal derecho combinado en la salida del bloque 225.

[0107] Para los canales individuales de una señal de múltiples canales, se lleva a cabo el procedimiento análogo, es decir, los canales individuales se añaden de manera individual, de tal manera que se añada siempre el mismo canal desde un renderizador DirAC 223 en el mismo canal correspondiente del otro renderizador DirAC y así sucesivamente. El mismo procedimiento se lleva a cabo también para, por ejemplo, señales Ambisonics de orden superior o en formato B. Cuando, por ejemplo, el primer renderizador DirAC 223 emite las señales W, X, Y, Z, y el segundo renderizador DirAC 224 emite un formato similar, entonces el combinador combina las dos señales omnidireccionales para obtener una señal omnidireccional W combinada, y el mismo procedimiento se lleva a cabo también para los componentes correspondientes con el fin de obtener finalmente un componente X, Y y Z combinado.

[0108] Además, de acuerdo con lo descrito con anterioridad con respecto a la Fig. 2a, la interfaz de entrada

está configurada para recibir metadatos de objetos de audio adicionales para un objeto de audio. Este objeto de audio ya se puede incluir en la primera o la segunda descripción DirAC o está separado de la primera y la segunda descripción DirAC. En este caso, el sintetizador DirAC 220 está configurado para manipular de manera selectiva los metadatos de objetos de audio adicionales o datos de objeto relacionados con estos metadatos de objetos de audio adicionales para llevar a cabo, por ejemplo, un filtrado direccional con base en los metadatos de objetos de audio adicionales o con base en información de dirección dada por el usuario obtenida de la interfaz de usuario 260. De manera alternativa o de manera adicional, y de acuerdo con lo ilustrado en la Fig. 2d, el sintetizador DirAC 220 está configurado para llevar a cabo, en el dominio espectral, una función de ganancia de fase cero, la función de ganancia de fase cero depende de una dirección de un objeto de audio, en el que la dirección está contenida en una corriente de bits si las direcciones de los objetos se transmiten como información lateral, o en el que la dirección se recibe desde la interfaz de usuario 260. Los metadatos de objetos de audio adicionales que se introducen en la interfaz 100 como una característica opcional en la Fig. 2a reflejan la posibilidad de enviar aún, para cada objeto individual su propia dirección y de manera opcional la distancia, la difusividad y cualquier otro atributo de objeto relevante como parte de la corriente de bits transmitida desde el codificador al decodificador. Por lo tanto, los metadatos de objetos de audio adicionales se pueden relacionar con un objeto ya incluido en la primera descripción DirAC o en la segunda descripción DirAC o es un objeto adicional no incluido en la primera descripción DirAC y ya en la segunda descripción DirAC.

[0109] Sin embargo, se prefiere tener los metadatos de objetos de audio adicionales ya en un estilo DirAC, es decir, una dirección de la información de llegada y, de manera opcional, una información de difusividad aunque los objetos de audio típicos tienen una difusión de cero, es decir, o concentrados a su posición real que da como resultado una dirección concentrada y específica de llegada que es constante en todas las bandas de frecuencia y que es, con respecto a la tasa de marco, estática o está en movimiento a un ritmo lento. Por lo tanto, dado que tal objeto tiene una sola dirección a lo largo de todas las bandas de frecuencia y se puede considerar estática o en movimiento a un ritmo lento, la información adicional requiere ser actualizada con menos frecuencia que otros parámetros DirAC y, por lo tanto, sólo incurrirá en una muy baja tasa de bits adicional. A modo de ejemplo, aunque la primera y la segunda descripción DirAC tienen datos de DoA y datos de difusión para cada banda espectral y para cada marco, los metadatos de objetos de audio adicionales solo requieren datos de una sola DoA para todas las bandas de frecuencia y estos datos solo para cada segundo marco o, preferentemente, cada tercer, cuarto, quinto o incluso cada décimo marco en la realización preferida.

[0110] Además, con respecto a la filtración direccional llevada a cabo en el sintetizador DirAC 220 que se incluye de manera típica dentro de un decodificador en un lado del decodificador de un sistema de codificador/decodificador, el sintetizador DirAC puede, en la alternativa de la Fig. 2b, llevar a cabo el filtrado direccional dentro del dominio de parámetro antes de la combinación de escenas o llevar a cabo de nuevo el filtrado direccional posterior a la combinación de escenas. Sin embargo, en este caso, el filtrado direccional se aplica a la escena combinada en lugar de a las descripciones individuales.

[0111] Además, en el caso de que un objeto de audio no esté incluido en la primera o la segunda descripción, pero se incluya por sus propios metadatos de objetos de audio, el filtrado direccional de acuerdo con lo ilustrado por el manipulador selectivo se puede aplicar de manera selectiva sólo al objeto de audio adicional, para lo cual existen metadatos de objetos de audio adicionales sin afectar a la primera o la segunda descripción DirAC o a la descripción DirAC combinada. Para el objeto de audio en sí, allí tampoco existe un canal de transporte separado que representa la señal de forma de onda de objeto o la señal de formas de onda de objeto está incluida en el canal de transporte de mezcla descendente.

[0112] Una manipulación selectiva de acuerdo con lo ilustrado, por ejemplo, en la Fig. 2b puede, por ejemplo, proceder de tal manera que una cierta dirección de llegada esté dada por la dirección del objeto de audio introducido en la Fig. 2d incluido en la corriente de bits como información lateral o recibida desde una interfaz de usuario. A continuación, con base en la dirección dada por el usuario o información de control, el usuario puede exponer, por ejemplo, que desde una cierta dirección, los datos de audio se pretenden mejorar o se desean atenuar. Por lo tanto, el objeto (metadatos) para el objeto en cuestión se amplifica o atenúa.

[0113] En el caso de los datos de forma de onda real como los datos de objetos introducidos en el manipulador selectivo 226 desde la izquierda en la Fig. 2d, los datos de audio se atenuarían o mejorarían realmente en función de la información de control. Sin embargo, en el caso de datos de objetos que tienen, además de la dirección de llegada y de manera opcional difusividad o distancia, una información de energía adicional, entonces la información de energía para el objeto se reduciría en el caso de una atenuación requerida para el objeto o la información de energía se incrementaría en el caso de una amplificación necesaria de los datos de objeto.

[0114] Por lo tanto, el filtrado direccional se basa en una técnica de atenuación espectral a corto plazo, y se lleva a cabo que el dominio espectral por una función de ganancia de fase cero que depende de la dirección de los objetos. La dirección puede estar contenida en la corriente de bits si las direcciones de los objetos se transmiten como información lateral. De lo contrario, la dirección también se podría dar de forma interactiva por el usuario. Naturalmente, el mismo procedimiento no sólo se puede aplicar al objeto individual dado y reflejado por los

metadatos de objetos de audio adicionales proporcionados de manera típica por datos de DoA para todas las bandas de frecuencia y los datos de DoA con una baja proporción de actualización con respecto a la frecuencia de imagen y también propuesta por la información de energía para el objeto, sino que el filtrado direccional también se puede aplicar a la primera descripción DirAC independiente de la segunda descripción DirAC o viceversa o también se puede aplicar a la descripción combinada de DirAC, de acuerdo con el caso.

[0115] Además, cabe destacar que la característica con respecto a los datos de objetos de audio adicionales también se puede aplicar en el primer aspecto de la presente invención ilustrado con respecto a las Figs. 1a a 1f. A continuación, la interfaz de entrada 100 de la Fig. 1a recibe además los datos de objetos de audio adicionales de acuerdo con lo discutido con respecto a la Fig. 2a, y el combinador de formatos se puede implementar como el sintetizador DirAC en el dominio espectral 220 controlado por una interfaz de usuario 260.

[0116] Además, el segundo aspecto de la presente invención de acuerdo con lo ilustrado en la Fig. 2 es diferente del primer aspecto en que la interfaz de entrada recibe ya dos descripciones DirAC, es decir, las descripciones de un campo de sonido que se encuentran en el mismo formato y, por lo tanto, para el segundo aspecto, no se requiere necesariamente el conversor de formatos 120 del primer aspecto.

[0117] Por otro lado, cuando la entrada en el combinador de formatos 140 de la Fig. 1a consiste en dos descripciones DirAC, entonces el combinador de formatos 140 se puede implementar de acuerdo con lo discutido con respecto al segundo aspecto que se ilustra en la Fig. 2a, o, de manera alternativa, los dispositivos 220, 240 de la Fig. 2a, se pueden implementar de acuerdo con lo discutido con respecto al combinador de formatos 140 de la Fig. 1a del primer aspecto.

[0118] La Fig. 3a ilustra un conversor de datos de audio que comprende una interfaz de entrada 100 para la recepción de una descripción de objeto de un objeto de audio que tiene metadatos del objeto de audio. Además, la interfaz de entrada 100 está seguida por un conversor de metadatos 150 que también corresponde a los conversores de metadatos 125, 126 discutidos con respecto al primer aspecto de la presente invención para la conversión de los metadatos del objeto de audio en metadatos DirAC. La salida del conversor de audio de la Fig. 3a está constituida por una interfaz de salida 300 para la transmisión o el almacenamiento de los metadatos DirAC. La interfaz de entrada 100 puede recibir además una señal de forma de onda de acuerdo con lo ilustrado por la segunda entrada de flecha en la interfaz 100. Además, la interfaz de salida 300 se puede implementar para introducir, de manera típica una representación codificada de la señal de forma de onda en la salida de señal de salida por el bloque 300. Si el conversor de datos de audio está configurado para convertir solo una descripción de un solo objeto, incluidos los metadatos, la interfaz de salida 300 también proporciona una descripción DirAC de este objeto de audio único junto con la señal de forma de onda codificada de manera típica como el canal de transporte DirAC.

[0119] En particular, los metadatos de objetos de audio tienen una posición del objeto, y los metadatos DirAC tienen una dirección de llegada con respecto a una posición de referencia derivada de la posición del objeto. En particular, el conversor de metadatos 150, 125, 126 está configurado para convertir los parámetros DirAC derivados del formato de datos de objeto en los datos de presión/velocidad, y el conversor de metadatos está configurado para aplicar un análisis DirAC a estos datos de presión/velocidad como, por ejemplo, se ilustra por el diagrama de flujo de la Fig. 3c que consiste en el bloque 302, 304, 306. Para este propósito, los parámetros DirAC que salen por el bloque 306 tienen una mejor calidad que los parámetros DirAC derivados de los metadatos de objetos obtenidos por el bloque 302, es decir, son parámetros DirAC mejorados. La Fig. 3b ilustra la conversión de una posición para un objeto en la dirección de llegada con respecto a una posición de referencia para el objeto específico.

[0120] La Fig. 3f ilustra un diagrama esquemático para explicar la funcionalidad del conversor de metadatos 150. El conversor de metadatos 150 recibe la posición del objeto indicado por el vector P en un sistema de coordenadas. Además, la posición de referencia, con la que los metadatos DirAC se tienen que relacionar está dada por el vector R en el mismo sistema de coordenadas. Por lo tanto, la dirección del vector de llegada DoA se extiende desde la punta de vector R hasta la punta del vector B. Por lo tanto, el vector DoA real se obtiene por medio de la sustracción del vector de posición de referencia R del vector de posición del objeto P.

[0121] Con el fin de tener una información DoA normalizada indicada por el vector DoA, el vector de diferencia se divide por la magnitud o la duración del vector DoA. Además, y si esto fuera necesario y deseado, la longitud del vector DoA también pueden ser incluida en los metadatos generados por el conversor de metadatos 150 de tal manera que, de forma adicional, la distancia del objeto desde el punto de referencia se incluya también en los metadatos, de tal manera que una manipulación selectiva de este objeto también se pueda llevar a cabo con base en la distancia del objeto desde la posición de referencia. En particular, el bloque de dirección de extracto 148 de la Fig. 1f también puede funcionar de acuerdo con lo discutido con respecto a la Fig. 3f, aunque otras alternativas para el cálculo de la información DoA y, de manera opcional, la información de distancia también se pueden aplicar. Además, de acuerdo con lo discutido con anterioridad con respecto a la Fig. 3a, los bloques 125 y 126 ilustrados en la Fig. 1c o 1d pueden operar de manera similar a la descrita con respecto a la Fig. 3f.

[0122] Además, el dispositivo de la Fig. 3a puede estar configurado para recibir una pluralidad de descripciones de objetos de audio, y el conversor de metadatos está configurado para convertir cada descripción de metadatos directamente en una descripción DirAC y, a continuación, el conversor de metadatos está configurado para combinar las descripciones de metadatos DirAC individuales para obtener una descripción DirAC combinada como los metadatos DirAC que se ilustran en la Fig. 3a. En una realización, la combinación se lleva a cabo por medio del cálculo 320 de un factor de ponderación para una primera dirección de llegada por el uso de una primera energía y el cálculo 322 de un factor de ponderación para una segunda dirección de llegada por el uso de una segunda energía, donde la dirección de llegada es procesada por los bloques 320, 332 en relación con el mismo compartimento de tiempo/frecuencia. Entonces, en el bloque 324, se lleva a cabo una suma ponderada también de acuerdo con lo discutido con respecto al punto 144 en la Fig. 1d. De este modo, el procedimiento ilustrado en la Fig. 3a representa una realización de la primera alternativa de la Fig. 1d.

[0123] Sin embargo, con respecto a la segunda alternativa, el procedimiento sería que toda la difusividad se ponga en cero o a un valor pequeño y, para un compartimiento de tiempo/frecuencia, todas las direcciones diferentes de valores de llegada que se dan para este compartimento de tiempo/frecuencia se consideren y la dirección más larga de valor de llegada se seleccione para que sea la dirección combinada de valor de llegada para este compartimento de tiempo/frecuencia. En otras realizaciones, también se podría seleccionar el segundo con el valor más grande con la condición de que la información de energía para estas dos direcciones de valores de llegada no sea tan diferente. Se selecciona la dirección del valor de llegada, cuya energía es la energía más grande entre las energías de las diferentes contribuciones para este compartimento de tiempo/frecuencia o la segunda o tercera energía más alta.

[0124] Por lo tanto, el tercer aspecto, de acuerdo con lo descripto con respecto a las Figs. 3a a 3f es diferente del primer aspecto en que el tercer aspecto también es útil para la conversión de una sola descripción de objeto en un metadato DirAC. De manera alternativa, la interfaz de entrada 100 puede recibir varias descripciones de objetos que se encuentran en el mismo formato de objeto/metadatos. Por lo tanto, no se requiere ningún conversor de formatos de acuerdo con lo discutido con respecto al primer aspecto en la Fig. 1a. Por lo tanto, la realización de la Fig. 3a puede ser útil en el contexto de recibir dos descripciones de objetos diferentes por el uso de diferentes señales de forma de onda de objeto y diferentes metadatos de objetos como la primera descripción de escena y la segunda descripción como entrada en el combinador de formatos 140, y la salida del conversor de metadatos 150, 125, 126 o 148 puede ser una renderización DirAC con metadatos DirAC y, por lo tanto, tampoco se requiere el analizador DirAC 180 de la Fig. 1. Sin embargo, los otros elementos con respecto al generador del canal de transporte 160 que corresponden al mezclador descendente 163 de la Fig. 3a se pueden utilizar en el contexto del tercer aspecto, así como el codificador del canal de transporte 170, el codificador de metadatos 190 y, en este contexto, la interfaz de salida 300 de la Fig. 3a corresponde a la interfaz de salida 200 de la Fig. 1a. Por lo tanto, todas las descripciones correspondientes dadas con respecto al primer aspecto también se aplican al tercer aspecto.

[0125] Las Figs. 4a, 4b ilustran un cuarto aspecto de la presente invención en el contexto de un aparato para la realización de una síntesis de datos de audio. En particular, el aparato tiene una interfaz de entrada 100 para la recepción de una descripción DirAC de una escena de audio que tiene metadatos DirAC y, además, para la recepción de una señal de objeto que tiene metadatos de objetos. Este codificador de escenas de audio que se ilustra en la Fig. 4b comprende además el generador de metadatos 400 para la generación de una descripción de metadatos combinada que comprende los metadatos DirAC por un lado y los metadatos de objetos por otro lado. Los metadatos DirAC comprenden la dirección de llegada de los mosaicos de tiempo/frecuencia individuales y los metadatos de objetos comprenden una dirección o de manera adicional una distancia o una difusividad de un objeto individual.

[0126] En particular, la interfaz de entrada 100 está configurada para recibir, de manera adicional, una señal de transporte asociada con la descripción DirAC de la escena de audio de acuerdo con lo ilustrado en la Fig. 4b, y la interfaz de entrada está configurada además para la recepción de una señal de forma de onda de objeto asociada con la señal de objeto. Por lo tanto, el codificador de escenas comprende además un codificador de señales de transporte para la codificación de la señal de transporte y la señal de forma de onda de objeto, y el codificador de transporte 170 puede corresponder al codificador 170 de la Fig. 1a.

[0127] En particular, el generador de metadatos 400 que genera los metadatos combinados se puede configurar de acuerdo con lo descrito con respecto al primer aspecto, el segundo aspecto o el tercer aspecto. Y, en una realización preferida, el generador de metadatos 400 está configurado para generar, para los metadatos de objetos, una única dirección de banda ancha por tiempo, es decir, durante un cierto marco de tiempo, y el generador de metadatos está configurado para actualizar la única dirección de banda ancha por tiempo con menos frecuencia que los metadatos DirAC.

[0128] El procedimiento descrito con respecto a la Fig. 4b permite tener metadatos combinados que tienen metadatos para una descripción DirAC completa y que tienen, además, metadatos para un objeto de audio adicional, pero en el formato DirAC de manera que se pueda llevar a cabo una renderización DirAC muy útil, al mismo tiempo, que se puede llevar a cabo un filtrado direccional selectivo o modificación de acuerdo con lo discutido con

anterioridad con respecto al segundo aspecto.

[0129] De este modo, el cuarto aspecto de la presente invención y, en particular, el generador de metadatos 400 representan un conversor de formatos específico en el que el formato común es el formato DirAC, y la entrada es una descripción DirAC para la primera escena en el primer formato discutido con respecto a la Fig. 1a y la segunda escena es uno solo o un combinado tal como una señal de objeto SAOC. Por lo tanto, la salida del conversor de formatos 120 representa la salida del generador de metadatos 400, pero, en contraste con una combinación específica real de los metadatos por una de las dos alternativas, por ejemplo, de acuerdo con lo discutido con respecto a la Fig. 1d, los metadatos de objeto están incluidos en la señal de salida, es decir, los "metadatos combinados" separados de los metadatos para la descripción DirAC para permitir una modificación selectiva de los datos de objeto.

[0130] Por lo tanto, la "dirección/distancia/difusividad" indicada en el punto 2 en el lado derecho de la Fig. 4a corresponde a la entrada de metadatos del objeto de audio adicional en la interfaz de entrada 100 de la Fig. 2a, pero, en la realización de la Fig. 4a, para una sola descripción DirAC solamente. Por lo tanto, en cierto sentido, se podría decir que la Fig. 2a representa una implementación del lado del decodificador del codificador ilustrado en la Fig. 4a, 4b con la condición de que el lado del decodificador del dispositivo de la Fig. 2a reciba solamente una única descripción DirAC y los metadatos de objeto generados por el generador de metadatos 400 dentro de la misma corriente de bits que los "metadatos de objetos de audio adicionales".

[0131] Por lo tanto, se puede llevar a cabo una modificación completamente diferente de los datos de objetos adicionales cuando la señal de transporte codificado tiene una representación separada de la señal de forma de onda de objeto separada de la corriente de transporte de DirAC. Y, sin embargo, el codificador de transporte 170 mezcla en forma descendente ambos datos, es decir, el canal de transporte para la descripción DirAC y la señal de forma de onda desde el objeto, a continuación, la separación será menos perfecta, pero por medio de información de energía objeto adicional, incluso está disponible una separación de un canal de mezcla descendente combinado y una modificación selectiva del objeto con respecto a la descripción DirAC.

[0132] Las Figs. 5a a 5d representan un quinto aspecto adicional de la invención en el contexto de un aparato para la realización de una síntesis de datos de audio. Con este fin, se proporciona una interfaz de entrada 100 para la recepción de una descripción DirAC de uno o más objetos de audio y/o una descripción DirAC de una señal de múltiples canales y/o una descripción DirAC de una señal Ambisonics de primer orden y/o una señal Ambisonics de orden superior, en la que la descripción DirAC comprende información de posición de los uno o más objetos o una información lateral para la señal Ambisonics de primer orden o la señal Ambisonics de orden superior o una información de posición para la señal de múltiples canales como información lateral o desde una interfaz de usuario.

[0133] En particular, un manipulador 500 se configura para la manipulación de la descripción DirAC de los uno o más objetos de audio, la descripción DirAC de la señal de múltiples canales, la descripción DirAC de señales Ambisonics de primer orden o la descripción DirAC de señales Ambisonics de orden superior para obtener una descripción DirAC manipulada. Para sintetizar esta descripción DirAC manipulada, un sintetizador DirAC 220, 240 está configurado para la síntesis de esta descripción DirAC manipulada para obtener datos de audio sintetizados.

[0134] En una realización preferida, el sintetizador DirAC 220, 240 comprende un renderizador DirAC 222 de acuerdo con lo ilustrado en la Fig. 5b y el conversor de tiempo espectral posteriormente conectado 240 que emite la señal de dominio temporal manipulada. En particular, el manipulador 500 está configurado para llevar a cabo una operación de ponderación dependiente de la posición antes de la renderización DirAC.

[0135] En particular, cuando el sintetizador DirAC está configurado para dar salida a una pluralidad de objetos de una señal Ambisonics de primer orden o una señal Ambisonics de orden superior o una señal de múltiples canales, el sintetizador DirAC está configurado para utilizar un conversor de tiempo espectral separado para cada objeto o cada componente de las primeras o las señales Ambisonics de orden superior o para cada canal de la señal de múltiples canales de acuerdo con lo ilustrado en la Fig. 5D en los bloques 506, 508. De acuerdo con lo indicado en el bloque 510 entonces la salida de las correspondientes conversiones separadas se añaden juntas con la condición de que todas las señales estén en un formato común, es decir, en un formato compatible.

[0136] Por lo tanto, en el caso de la interfaz de entrada 100 de la Fig. 5a, tras la recepción de más de una, es decir, dos o tres representaciones, cada representación se podría manipular por separado, de acuerdo con lo ilustrado en el bloque 502 en el dominio de parámetro de acuerdo con lo discutido con anterioridad con respecto a las Figs. 2b o 2c, y, a continuación, se podría llevar a cabo una síntesis de acuerdo con lo indicado en el bloque 504 para cada descripción manipulada, y la síntesis se podría añadir entonces en el dominio temporal de acuerdo con lo discutido con respecto al bloque 510 en la Fig. 5d. De manera alternativa, el resultado de los procedimientos de síntesis DirAC individuales en el dominio espectral ya se podría sumar en el dominio espectral y, a continuación, también se podría utilizar una sola conversión de dominio temporal. En particular, el manipulador 500 se puede implementar como el manipulador discutido con respecto a la Fig. 2D o discutido con respecto a cualquier otro aspecto anterior.

[0137] Por lo tanto, el quinto aspecto de la presente invención proporciona una característica significativa con respecto al hecho de que, cuando se introducen las descripciones DirAC individuales de señales de sonido muy diferentes, y cuando se lleva a cabo una cierta manipulación de las descripciones individuales de acuerdo con lo discutido con respecto al bloque 500 de la Fig. 5a, donde una entrada en el manipulador 500 puede ser una descripción DirAC de cualquier formato, que incluye sólo un único formato, mientras que el segundo aspecto se concentraba en la recepción de al menos dos descripciones DirAC diferentes o cuando el cuarto aspecto, por ejemplo, estaba relacionado con la recepción de una descripción DirAC por un lado y una descripción de la señal de objeto por otro lado.

[0138] Posteriormente, se hace referencia a la Fig. 6. La Fig. 6 ilustra otra implementación para la realización de una síntesis diferente del sintetizador DirAC. Cuando, por ejemplo, un analizador de campo de sonido genera, para cada señal de fuente, una señal mono separada S y una dirección original de llegada y cuando, dependiendo de la información de traslación, se calcula una nueva dirección de llegada, a continuación, el generador de señal Ambisonics 430 de la Fig. 6, por ejemplo, se utilizaría para generar una descripción de campo de sonido para la señal de fuente de sonido, es decir, la señal mono S pero para la nueva dirección de llegada de datos (DoA) que consiste en un ángulo horizontal θ o un ángulo de elevación θ y un ángulo de azimut φ. Entonces, un procedimiento llevado a cabo por el calculador de campo de sonido 420 de la Fig. 6 sería generar, por ejemplo, una representación de campo de sonido Ambisonics de primer orden para cada fuente de sonido con la nueva dirección de llegada y, a continuación, se podría llevar a cabo una modificación adicional por fuente de sonido por el uso de un factor de escala que depende de la distancia del campo de sonido a la nueva ubicación de referencia y, a continuación, todos los campos de sonido de las fuentes individuales se podrían superponer entre sí para obtener finalmente el campo de sonido modificado, una vez más, por ejemplo, en una representación Ambisonics relacionada con una cierta nueva ubicación de referencia.

[0139] Cuando se interpreta que cada compartimento de tiempo/frecuencia procesado por el analizador DirAC 422 representa una cierta fuente de sonido (ancho de banda limitado), entonces el generador de señal Ambisonics 430 se podría utilizar, en lugar del sintetizador DirAC 425, para generar, para cada compartimento de tiempo/frecuencia, una representación Ambisonics completa por el uso de la señal de mezcla descendente o la señal de presión o componente omnidireccional para este compartimento de tiempo/frecuencia como la “señal mono S” de la Fig. 6. A continuación, una conversión de frecuencia-tiempo individual en el conversor de frecuencia-tiempo 426 para cada uno de los componentes W, X, Y, Z daría como resultado entonces una descripción del campo de sonido diferente de lo que se ilustra en la Fig. 6.

[0140] Posteriormente, se dan más explicaciones sobre un análisis DirAC y una síntesis DirAC de acuerdo con lo conocido en la técnica. La Fig. 7a ilustra un analizador DirAC de acuerdo con lo descrito originalmente, por ejemplo, en la referencia “Directional Audio Coding” de IWPASH de 2009. El analizador DirAC comprende un banco de filtros de banda 1310, un analizador de energía 1320, un analizador de intensidad 1330, un bloque promedio temporal 1340 y un calculador de difusividad 1350 y el calculador de dirección 1360. En DirAC, tanto el análisis como la síntesis se llevan a cabo en el dominio de la frecuencia. Existen varios procedimientos para la división del sonido en bandas de frecuencia, cada una dentro de propiedades distintas. Las transformadas de frecuencia más comúnmente utilizadas incluyen transformada de Fourier de tiempo corto (STFT, por su sigla en inglés), y el banco de filtros de espejo en Cuadratura (QMF, por su sigla en inglés). Además de estos, hay una plena libertad para diseñar un banco de filtros con filtros arbitrarios que están optimizados para fines específicos. El objetivo del análisis direccional es estimar en cada banda de frecuencia la dirección de llegada del sonido, junto con una estimación de si el sonido está llegando a partir de una o múltiples direcciones al mismo tiempo. En principio, esto se puede llevar a cabo con una serie de técnicas, sin embargo, se ha encontrado que el análisis energético del campo de sonido es adecuado, el cual se ilustra en la Fig. 7a. El análisis energético se puede llevar a cabo cuando la señal de presión y las señales de velocidad en una, dos o tres dimensiones son capturadas desde una única posición. En las señales en formato B de primer orden, la señal omnidireccional se llama señal W, que se ha reducido por la raíz cuadrada de dos. La presión sonora se puede estimar como, $S = \sqrt{2} * W$ expresado en el dominio STFT.

[0141] Los canales X, Y y Z tienen el patrón direccional de un dipolo dirigido a lo largo del eje cartesiano, que forman juntos un vector U = [X, Y, Z]. El vector estima el vector de velocidad del campo de sonido, y se expresa también en el dominio STFT. Se calcula la energía E del campo de sonido. La captura de las señales en formato B se puede obtener ya sea con el posicionamiento coincidente de micrófonos direccionales, o con un conjunto estrechamente espaciado de micrófonos omnidireccionales. En algunas aplicaciones, las señales del micrófono se pueden formar en un dominio computacional, es decir, simulado. La dirección del sonido se define para que sea la dirección opuesta del vector de intensidad I. La dirección se denota como valores de azimut y elevación angulares correspondientes en los metadatos transmitidos. La difusividad del campo de sonido también se calcula por el uso de un operador de expectativa del vector de intensidad y la energía. El resultado de esta ecuación es un número de valor real entre cero y uno, que caracteriza si la energía del sonido está llegando desde una única dirección (la difusividad es cero), o desde todas las direcciones (la difusividad es uno). Este procedimiento es apropiado en el caso en que la información completa de velocidad 3D o de menos dimensiones está disponible.

[0142] La Fig. 7b ilustra una síntesis DirAC, que tiene de nuevo un banco de filtros de banda 1370, un bloque de micrófono virtual 1400, un bloque de sintetizador directo/difuso 1450, y una cierta configuración de altavoz o una configuración de altavoz virtual prevista 1460. De manera adicional, se utilizan un transformador de ganancia de difusividad 1380, un bloque de tabla de ganancia de paneo de amplitud basado en vectores (VBAP, por su sigla en inglés) 1390, un bloque de compensación de micrófono 1420, un bloque promedio de ganancia del altavoz 1430 y un distribuidor 1440 para otros canales. En esta síntesis DirAC con altavoces, la versión de alta calidad de síntesis DirAC que se muestra en la Fig. 7b recibe todas las señales en formato B, para las que se calcula una señal de micrófono virtual para cada dirección de altavoz de la configuración de altavoz 1460. El patrón direccional utilizado de manera típica es un dipolo. Las señales de micrófono virtuales se modifican entonces de manera no lineal, dependiendo de los metadatos. La versión de baja tasa de bits de DirAC no se muestra en la Fig. 7b, sin embargo, en esta situación, sólo un canal de audio se transmite de acuerdo con lo ilustrado en la Fig. 6. La diferencia en el procesamiento es que todas las señales de micrófono virtuales se sustituirían por el único canal de audio que se recibe. Las señales de micrófono virtuales se dividen en dos corrientes: las corrientes difusas y no difusas, que se procesan por separado.

[0143] El sonido no difuso se reproduce como fuentes puntuales por el uso de paneo de amplitud de base vectorial (VBAP). En el paneo, una señal de sonido monofónico se aplica a un subconjunto de los altavoces después de la multiplicación con factores de ganancia de altavoz específico. Los factores de ganancia se calculan por el uso de la información de una configuración de altavoz, y la dirección de paneo especificada. En la versión de baja tasa de bits, la señal de entrada simplemente se panea a las direcciones implicadas por los metadatos. En la versión de alta calidad, cada señal de micrófono virtual se multiplica por el correspondiente factor de ganancia, que produce el mismo efecto que con el paneo, sin embargo, es menos propenso a los artefactos no lineales.

[0144] En muchos casos, los metadatos de dirección están sujetos a cambios temporales abruptos. Para evitar los artefactos, los factores de ganancia para los altavoces calculados con VBAP son suavizados por la integración temporal con constantes de tiempo dependientes de la frecuencia que equivalen a aproximadamente 50 períodos de ciclo en cada banda. Esto elimina de manera eficaz los artefactos, sin embargo, los cambios de dirección no se perciben como más lentos que sin un promedio en la mayoría de los casos. El objetivo de la síntesis del sonido difuso es crear la percepción del sonido que rodea al oyente. En la versión de baja tasa de bits, la corriente difusa es reproducida por medio de la descorrelación de la señal de entrada y la reproducción desde cada altavoz. En la versión de alta calidad, las señales de los micrófonos virtuales de corriente difusa ya son incoherentes en cierto grado, y tienen que ser correlacionados sólo ligeramente. Esta estrategia proporciona una mejor calidad espacial de reverberación envolvente y sonido ambiente que la versión de baja tasa de bits. Para la síntesis DirAC con auriculares, DirAC está formulado con una cierta cantidad de altavoces virtuales alrededor del oyente para la corriente no difusa y un cierto número de altavoces para la corriente difusa. Los altavoces virtuales se implementan como convolución de las señales de entrada con funciones de transferencia relacionadas con cabezales medidos (HRTF, por su sigla en inglés).

[0145] Posteriormente, se da una relación general de manera adicional con respecto a los diferentes aspectos y, en particular, con respecto a otras implementaciones del primer aspecto de acuerdo con lo discutido con respecto a la Fig. 1a. En general, la presente invención se refiere a la combinación de diferentes escenas en diferentes formatos por el uso de un formato común, donde el formato común puede ser, por ejemplo, el dominio en formato B, el dominio de presión/velocidad o el dominio de metadatos de acuerdo con lo discutido, por ejemplo, en los puntos 120, 140 de la Fig. 1a.

[0146] Cuando la combinación no se lleva a cabo directamente en el formato común DirAC, a continuación, un análisis DirAC 802 se lleva a cabo en una de las alternativas antes de la transmisión en el codificador de acuerdo con lo discutido con anterioridad con respecto al punto 180 de la Fig. 1a.

[0147] Entonces, con posterioridad al análisis DirAC, el resultado se codifica de acuerdo con lo discutido con anterioridad con respecto al codificador 170 y el codificador de metadatos 190 y el resultado codificado se transmite a través de la señal de salida codificada generada por la interfaz de salida 200. Sin embargo, en una alternativa adicional, el resultado podría estar directamente renderizado por un dispositivo de la Fig. 1a cuando la salida del bloque 160 de la Fig. 1a y la salida del bloque 180 de la Fig. 1a se reenvía a un renderizador DirAC. De este modo, el dispositivo de la Fig. 1a no sería un dispositivo de codificador específico, sino que sería un analizador y un renderizador correspondiente.

[0148] Una alternativa adicional se ilustra en la rama derecha de la Fig. 8, donde se lleva a cabo una transmisión desde el codificador al decodificador y, de acuerdo con lo ilustrado en el bloque 804, el análisis DirAC y la síntesis DirAC se llevan a cabo con posterioridad a la transmisión, es decir, en el lado del decodificador. Este procedimiento sería el caso cuando se utiliza la alternativa de la Fig. 1a, es decir, que la señal de salida codificada es una señal de formato B sin metadatos espaciales. Después del bloque 808, el resultado se podría renderizar para la reproducción o, de manera alternativa, el resultado podría incluso ser codificado y transmitido de nuevo. Por lo tanto, se hace evidente que los procedimientos de la invención de acuerdo con lo definido y descrito con respecto a los diferentes aspectos son altamente flexibles y se pueden adaptar muy bien a casos de uso específicos.

Primer aspecto de la invención: codificación/renderización de audio espacial basada en DirAC universal

[0149] Un codificador de audio espacial basado en DirAC que puede codificar señales de múltiples canales, formatos Ambisonics y objetos de audio por separado o de manera simultánea.

Beneficios y ventajas sobre el estado de la técnica

[0150]

- Esquema de codificación de audio espacial basado en DirAC universal para los formatos de entrada de audio de inmersión más relevantes
- Renderización de audio universal de diferentes formatos de entrada en diferentes formatos de salida

Segundo aspecto de la invención: combinación de dos o más descripciones DirAC en un decodificador

[0151] El segundo aspecto de la invención se relaciona con la combinación y la renderización de dos o más descripciones DirAC en el dominio espectral.

Beneficios y ventajas sobre el estado de la técnica

[0152]

- Combinación de corrientes DirAC eficiente y precisa
- Permite el uso de DirAC que representa universalmente cualquier escena y combina de manera eficiente diferentes corrientes en el dominio de parámetro o el dominio espectral
- Manipulación de escenas eficaz e intuitiva de escenas DirAC individuales o de la escena combinada en el dominio espectral y posterior conversión en el dominio temporal de la escena combinada manipulada.

Tercer aspecto de la invención: conversión de objetos de audio en el dominio DirAC

[0153] El tercer aspecto de la invención está relacionado con la conversión de metadatos de objetos y de manera opcional señales de forma de onda de objeto directamente en el dominio DirAC y en una realización la combinación de varios objetos en una representación de objeto.

Beneficios y ventajas sobre el estado de la técnica

[0154]

- Estimación de metadatos DirAC eficiente y precisa por medio de un simple transcodificador de metadatos de los metadatos de objetos de audio
- Permite a DirAC codificar escenas de audio complejas que incluyen uno o más objetos de audio
- Procedimiento eficiente para la codificación de objetos de audio a través de DirAC en una única representación paramétrica de la escena de audio completa.

Cuarto aspecto de la invención: combinación de metadatos de objetos y metadatos DirAC regulares

[0155] El tercer aspecto de la invención se refiere a la enmienda de los metadatos DirAC con las direcciones y, de manera óptima, la distancia o la difusividad de los objetos individuales que componen la escena de audio combinada representada por los parámetros DirAC. Esta información adicional se codifica con facilidad, dado que consiste principalmente en una sola dirección de banda ancha por unidad de tiempo y se puede actualizar con menos frecuencia que los otros parámetros DirAC dado que se puede suponer que los objetos son estáticos o están en movimiento a un ritmo lento.

Beneficios y ventajas sobre el estado de la técnica

[0156]
DirAC permite codificar una escena de audio compleja que implica uno o más objetos de audio

- Una estimación de metadatos DirAC eficiente y precisa por medio del simple transcodificador de metadatos de los metadatos de objetos de audio.
- Procedimiento más eficiente para la codificación de objetos de audio a través de DirAC por medio de la combinación eficiente de sus metadatos en el dominio DirAC
- Procedimiento eficiente para la codificación de objetos de audio y a través de DirAC por medio de la combinación eficiente de sus representaciones de audio en una única representación paramétrica de la escena de audio.

Quinto aspecto de la invención: manipulación de escenas de objetos MC y FOA/HOA C en la síntesis DirAC

[0157] El cuarto aspecto está relacionado con el lado del decodificador y aprovecha las posiciones conocidas de objetos de audio. Las posiciones pueden ser dadas por el usuario a través de una interfaz interactiva y también se pueden incluir como información lateral adicional dentro de la corriente de bits.

[0158] El objetivo es ser capaz de manipular una escena de audio de salida que comprende un número de objetos por medio del cambio individual de atributos de los objetos tales como los niveles, la ecualización y/o las posiciones espaciales. También se puede prever filtrar por completo el objeto o restituir los objetos individuales de la corriente combinada.

[0159] La manipulación de la escena de audio de salida se puede lograr por medio del procesamiento conjunto de los parámetros espaciales de los metadatos DirAC, los metadatos de los objetos, la entrada del usuario interactivo si está presente y las señales de audio transportadas en los canales de transporte.

Beneficios y ventajas sobre el estado de la técnica

[0160]

- Permite a DirAC dar salida a los objetos de audio del lado del decodificador de acuerdo con lo que se presenta en la entrada del codificador.
- Permite la reproducción de DirAC para manipular objetos de audio individuales por medio de la aplicación de las ganancias, rotación o...
- La capacidad requiere un mínimo esfuerzo computacional adicional dado que sólo requiere una operación de ponderación dependiente de la posición antes de la representación y un banco de filtros de síntesis al final de la síntesis DirAC (las salidas de objetos adicionales sólo requerirán un banco de filtros de síntesis adicional por salida de objeto).

Referencias que se incorporan en su totalidad como referencia:

[0161]

[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, Nov. 2009, Zao; Miyagi, Japón.
[2] Ville Pulkki. “Virtual source positioning using vector base amplitude panning”. J. Audio Eng. Soc., 45(6): 456 a 466, junio de 1997.
[3] M. V. Laitinen and V. Pulkki, "Converting 5.1 audio recordings to B-format for directional audio coding reproduction," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)), Praga, 2011, págs. 61 a 64.
[4] G. Del Galdo, F. Kuech, M. Kallinger and R. Schultz-Amling, "Efficient merging of multiple audio streams for spatial sound reproduction in Directional Audio Coding," 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, 2009 págs. 265 a 268.
[5] Jürgen HERRE, CORNELIA FALCH, DIRK MAHNE, GIOVANNI DEL GALDO, MARKUS KALLINGER, AND OLIVER THIERGART, “Interactive Teleconferencing Combining Spatial Audio Object Coding and DirAC Technology”, J. Audio Eng. Soc., Vol. 59, Núm. 12, diciembre de 2011.
[6] R. Schultz-Amling, F. Kuech, M. Kallinger, G. Del Galdo, J. Ahonen, V. Pulkki, “Planar Microphone Array Processing for the Analysis and Reproduction of Spatial Audio using Directional Audio Coding,” Audio Engineering Society Convention 124, Ámsterdam, Países Bajos, 2008.
[7] Daniel P. Jarrett and Oliver Thiergart and Emanuel A. P. Habets and Patrick A. Naylor, “Coherence-Based Diffuseness Estimation in the Spherical Harmonic Domain”, IEEE 27th Convention of Electrical and Electronics Engineers in Israel (IEEEI), 2012.
[8] Patente de Estados Unidos 9.015.051.

[0162] La presente invención proporciona, en realizaciones adicionales, y en particular con respecto al primer aspecto y también con respecto a los otros aspectos, diferentes alternativas. Estas alternativas son las siguientes:

En primer lugar, la combinación de diferentes formatos en el dominio formato B y, o bien la realización del análisis DirAC en el codificador o la transmisión de los canales combinados a un decodificador y la realización del análisis y la síntesis DirAC allí.
En segundo lugar, la combinación de diferentes formatos en el dominio de presión/velocidad y la realización del análisis DirAC en el codificador. De manera alternativa, los datos de presión/velocidad se transmiten al decodificador y el análisis DirAC se lleva a cabo en el decodificador y la síntesis también se lleva a cabo en el decodificador.
En tercer lugar, la combinación de diferentes formatos en el dominio de metadatos y la transmisión de una única

corriente DirAC o la transmisión de varias corrientes DirAC a un decodificador antes de combinarlos y hacer la combinación en el decodificador.

[0163] Además, las realizaciones o aspectos de la presente invención están relacionados con los siguientes aspectos:

En primer lugar, la combinación de diferentes formatos de audio de acuerdo con las tres alternativas anteriores.
En segundo lugar, se lleva a cabo una recepción, una combinación y una renderización de dos descripciones DirAC ya en el mismo formato.
En tercer lugar, se implementa un objeto específico al conversor DirAC con una “conversión directa” de datos de objeto a los datos DirAC.
En cuarto lugar, los metadatos de objetos además de metadatos DirAC normales y una combinación de ambos metadatos; también los datos que son existentes en el lado a lado de la corriente de bits, pero también los objetos de audio se describen también por el estilo de metadatos DirAC.
En quinto lugar, los objetos y la corriente DirAC se transmiten por separado a un decodificador y los objetos son manipulados de manera selectiva dentro del decodificador antes de convertir las señales de audio de salida (altavoz) en el dominio temporal.

[0164] Cabe mencionar aquí que todas las alternativas o aspectos de acuerdo con lo discutido con anterioridad y todos los aspectos de acuerdo con lo definido por medio de las reivindicaciones independientes en las reivindicaciones siguientes se pueden utilizar de manera individual, es decir, sin ninguna otra alternativa u objeto que la alternativa, el objeto o la reivindicación independiente contempladas. Sin embargo, en otras realizaciones, dos o más de las alternativas o los aspectos o las reivindicaciones independientes se pueden combinar entre sí y, en otras realizaciones, todos los aspectos, o alternativas y todas las reivindicaciones independientes se pueden combinar entre sí.

[0165] Una señal de audio codificada de acuerdo con la invención se puede almacenar en un medio de almacenamiento digital o un medio de almacenamiento no transitorio o se puede transmitir sobre un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable, tal como Internet.

[0166] Aunque algunos aspectos se han descrito en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o un dispositivo corresponde a una etapa del procedimiento o una característica de una etapa de procedimiento. De manera análoga, los aspectos descritos en el contexto de una etapa del procedimiento también representan una descripción de un bloque correspondiente o un elemento o característica de un aparato correspondiente.

[0167] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede llevar a cabo por el uso de un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un CD, una memoria ROM, una memoria PROM, una memoria EPROM, una memoria EEPROM o una memoria FLASH, que tienen señales de control legibles de manera electrónica almacenadas en el mismo, que cooperan (o son capaces de cooperar) con un sistema informático programable de tal manera que se lleve a cabo el procedimiento respectivo.

[0168] Algunas realizaciones de acuerdo con la invención comprenden un soporte de datos con señales de control legibles de manera electrónica, que son capaces de cooperar con un sistema informático programable, de tal manera que se lleve a cabo uno de los procedimientos descritos en esta invención.

[0169] Por lo general, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, el código de programa es operativo para llevar a cabo uno de los procedimientos, cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede estar almacenado en un soporte legible por máquina, por ejemplo.

[0170] Otras realizaciones comprenden el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por máquina o un medio de almacenamiento no transitorio.

[0171] En otras palabras, una realización del procedimiento de acuerdo con la invención es, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los procedimientos descritos en esta invención, cuando el programa informático se ejecuta en un ordenador.

[0172] Una realización adicional de los procedimientos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en el mismo, el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención.

[0173] Una realización adicional del procedimiento de acuerdo con la invención es, por lo tanto, una corriente

de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención. La corriente de datos o la secuencia de señales pueden estar, por ejemplo, configuradas para ser transferidas a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

[0174] Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los procedimientos descritos en esta invención.

[0175] Una realización comprende además un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención.

[0176] En algunas realizaciones, un dispositivo lógico programable (por ejemplo, una matriz de puertas programables por campo) se puede utilizar para llevar a cabo algunas o todas las funcionalidades de los procedimientos descritos en esta memoria. En algunas realizaciones, una matriz de puertas programable por campo podrá cooperar con un microprocesador con el fin de llevar a cabo uno de los procedimientos descritos en esta invención. Por lo general, los procedimientos se llevan a cabo con preferencia por cualquier aparato de hardware.

[0177] Las realizaciones descritas con anterioridad son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en esta invención serán evidentes para los expertos en la técnica. Por lo tanto, la intención es estar limitado sólo por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de esta invención.

REIVINDICACIONES

1. Un aparato para la generación de una descripción de una escena de audio combinada, que comprende:

una interfaz de entrada (100) para la recepción de una primera descripción de una primera escena en un primer formato y una segunda descripción de una segunda escena en un segundo formato, en el que el segundo formato es diferente del primer formato;
un conversor de formatos (120) para la conversión de la primera descripción en un formato común y para la conversión de la segunda descripción en el formato común, cuando el segundo formato es diferente del formato común; y
un combinador de formatos (140) para la combinación de la primera descripción en el formato común y la segunda descripción en el formato común para obtener la escena de audio combinada.

2. El aparato de la reivindicación 1,

en el que el primer formato y el segundo formato se seleccionan de un grupo de formatos que comprende un formato Ambisonics de primer orden, un formato Ambisonics de orden superior, un formato DirAC, un formato de objeto de audio y un formato de múltiples canales, y
en el que el segundo formato se selecciona de un grupo de formatos que comprenden un formato Ambisonics de primer orden, un formato Ambisonics de orden superior, el formato común, un formato Dirac, un formato de objeto de audio y un formato multicanal.

3. El aparato de la reivindicación 1 o 2,

en el que el conversor de formatos (120) está configurado para convertir la primera descripción en una primera representación de la señal de formato B y para convertir la segunda descripción en una segunda representación de la señal de formato B, y
en el que el combinador de formatos (140) está configurado para combinar la primera y la segunda representación de la señal de formato B por medio de la combinación individual de los componentes individuales de la primera y la segunda representación de la señal de formato B.

4. El aparato de una de las reivindicaciones anteriores,

en el que en el conversor de formatos (120) está configurado para convertir la primera descripción en una primera representación de la señal de presión/velocidad y para convertir la segunda descripción en una segunda representación de la señal de presión/velocidad, y
en el que el combinador de formatos (140) está configurado para combinar la primera y la segunda representación de la señal de presión/velocidad por medio de la combinación individual de los componentes individuales de las representaciones de la señal de presión/velocidad para obtener una representación de la señal de presión/velocidad combinada.

5. El aparato de una de las reivindicaciones anteriores,

en el que el conversor de formatos (120) está configurado para convertir la primera descripción en una primera representación de parámetros DirAC y para convertir la segunda descripción en una segunda representación de parámetros DirAC, cuando la segunda descripción es diferente de la representación de parámetros DirAC, y
en el que el combinador de formatos (140) está configurado para combinar la primera y la segunda representaciones de parámetros DirAC por medio de la combinación individual de los componentes individuales de la primera y la segunda representaciones de parámetros DirAC para obtener una representación de parámetros DirAC combinada para la escena de audio combinada.

6. El aparato de la reivindicación 5,

en el que el combinador de formatos (140) está configurado para generar la dirección de los valores de llegada para los mosaicos de tiempo-frecuencia o la dirección de los valores de llegada y los valores de difusividad para los mosaicos de tiempo-frecuencia que representan la escena de audio combinada.

7. El aparato de una de las reivindicaciones anteriores,

que comprende además un analizador DirAC (180) para el análisis de la escena de audio combinada para derivar parámetros DirAC para la escena de audio combinada,
en el que los parámetros DirAC comprenden la dirección de los valores de llegada para los mosaicos de tiempo-frecuencia o la dirección de los valores de llegada y los valores de difusividad para los mosaicos de tiempo-frecuencia que representan la escena de audio combinada.

8. El aparato de una de las reivindicaciones anteriores,

que comprende además un generador del canal de transporte (160) para la generación de una señal del canal de transporte de la escena de audio combinada o desde la primera escena y la segunda escena, y
un codificador del canal de transporte (170) para la codificación del núcleo de la señal del canal de transporte, o
en el que el generador del canal de transporte (160) está configurado para generar una señal estéreo a partir de la primera escena o la segunda escena que está en un formato Ambisonics de primer orden o Ambisonics de orden superior por el uso de un formador de haces dirigido a una posición izquierda o la posición derecha, respectivamente, o
en el que el generador del canal de transporte (160) está configurado para generar una señal estéreo a partir de la primera escena o la segunda escena que está en una representación de múltiples canales por medio de la mezcla descendente de tres o más canales de la representación de múltiples canales, o
en el que el generador del canal de transporte (160) está configurado para generar una señal estéreo a partir de la primera escena o la segunda escena que está en una representación de objeto de audio por medio del paneo de cada objeto por el uso de una posición del objeto o por medio de la mezcla descendente de objetos en una mezcla descendente en estéreo por el uso de la información que indica qué objeto se encuentra en qué canal estéreo, o
en el que el generador del canal de transporte (160) está configurado para sumar sólo el canal izquierdo de la señal estéreo al canal de transporte de mezcla descendente izquierdo y sumar sólo el canal derecho de la señal estéreo para obtener un canal de transporte derecho, o
en el que el formato común es el formato B, y en el que el generador del canal de transporte (160) está configurado para procesar una representación en formato B combinada para derivar la señal del canal de transporte, en el que el procesamiento comprende la realización de una operación de formación de haces o la extracción de un subconjunto de componentes de la señal de formato B tal como el componente omnidireccional como el canal de transporte mono, o
en el que el procesamiento comprende la formación de haces por el uso de la señal omnidireccional y el componente Y con signos opuestos del formato B para calcular canales izquierdo y derecho, o
en el que el procesamiento comprende una operación de formación de haces por medio de los componentes del formato B y el ángulo de azimut dado y el ángulo de elevación dado, o
en el que el generador del canal de transporte (160) está configurado para proporcionar las señales en formato B de la escena de audio combinada al codificador del canal de transporte, en el que los metadatos espaciales no están incluidos en la salida de escena de audio combinada por el combinador de formatos (140).

9. El aparato de una de las reivindicaciones anteriores, que comprende además:

un codificador de metadatos (190)
para la codificación de metadatos DirAC descritos en la escena de audio combinada para obtener metadatos DirAC codificados, o
para la codificación de metadatos DirAC derivados de la primera escena para obtener primeros metadatos DirAC codificados y para la codificación de metadatos DirAC derivados de la segunda escena para obtener segundos metadatos DirAC codificados.

10. El aparato de una de las reivindicaciones anteriores, que comprende además:

una interfaz de salida (200) para la generación de una señal de salida codificada que representa la escena de audio combinada, la señal de salida que comprende metadatos DirAC codificados y uno o más canales de transporte codificados.

11. El aparato de una de las reivindicaciones anteriores,

en el que el conversor de formatos (120) está configurado para convertir un formato Ambisonics de orden superior o Ambisonics de primer orden en el formato B, en el que el formato Ambisonics de orden superior se trunca antes de ser convertido en el formato B, o
en el que el conversor de formatos (120) está configurado para proyectar un objeto o un canal en armónicos esféricos en una posición de referencia para obtener señales proyectadas, y en el que el combinador de formatos (140) está configurado para combinar las señales de proyección para obtener coeficientes en formato B, en el que el objeto o el canal está situado en el espacio en una posición especificada y tiene una distancia individual opcional desde una posición de referencia, o
en el que el conversor de formatos (120) está configurado para llevar a cabo un análisis DirAC que comprende un análisis de tiempo-frecuencia de los componentes en formato B y una determinación de los vectores de presión y velocidad, y en el que el combinador de formatos (140) está configurado para combinar diferentes vectores de presión/velocidad y en el que el combinador de formatos (140) comprende además un analizador DirAC para derivar metadatos DirAC de los datos de presión/velocidad combinados, o
en el que el conversor de formatos (120) está configurado para extraer parámetros DirAC de metadatos de

objetos de un formato de objeto de audio como el primer o el segundo formato, en el que el vector de presión es la señal de forma de onda de objeto y la dirección se deriva de la posición del objeto en el espacio o la difusividad está directamente dada en los metadatos de objetos o está ajustada a un valor predeterminado, tal como un valor de 0, o

en el que el conversor de formatos (120) está configurado para convertir los parámetros DirAC derivados del formato de datos de objeto en los datos de presión/velocidad y el combinador de formatos (140) está configurado para combinar los datos de presión/velocidad con datos de presión/velocidad derivados de una descripción diferente de uno o más objetos de audio diferentes o

en el que el conversor de formatos (120) está configurado para derivar directamente parámetros DirAC, y en el que el combinador de formatos (140) está configurado para combinar los parámetros DirAC para obtener la escena de audio combinada.

12. El aparato de una de las reivindicaciones anteriores, en el que el conversor de formatos (120) comprende:

un analizador DirAC (180) para un formato de entrada Ambisonics de primer orden o Ambisonics de orden superior o un formato de señal de múltiples canales;

un conversor de metadatos (150, 125, 126, 148) para la conversión de metadatos de objetos en metadatos DirAC o para la conversión de una señal de múltiples canales que tiene una posición invariable en el tiempo en los metadatos DirAC; y

un combinador de metadatos (144) para la combinación de las corrientes de metadatos DirAC individuales o la combinación de dirección de los metadatos de llegada de varias corrientes por medio de una suma ponderada, la ponderación de la suma ponderada se lleva a cabo de conformidad con las energías de energías de señal de presión asociadas, o para la combinación de los metadatos de difusividad de varias corrientes por una suma ponderada, la ponderación de la suma ponderada se lleva a cabo de conformidad con las energías de energías de señal de presión asociadas, o

en el que el combinador de metadatos (144) está configurado para calcular, para un compartimento de tiempo/frecuencia de la primera descripción de la primera escena, un valor de energía, y la dirección del valor de llegada, y para calcular, para el compartimento de tiempo/frecuencia de la segunda descripción de la segunda escena, un valor de energía y una dirección del valor de llegada, y en el que el combinador de formatos (140) está configurado para multiplicar la primera energía a la primera dirección de valor de entrada y sumar un resultado de la multiplicación del segundo valor de energía y la segunda dirección del valor de llegada para obtener la dirección combinada de valor de llegada o, de manera alternativa, para seleccionar la dirección del valor de llegada entre la primera dirección del valor de llegada y la segunda dirección del valor de llegada que está asociado con la energía más alta que la dirección combinada del valor de llegada.

13. El aparato de una de las reivindicaciones anteriores,

que comprende además una interfaz de salida (200, 300) para la suma al formato combinado, de una descripción de objeto separado para un objeto de audio, comprendiendo la descripción de objeto al menos uno de una dirección, una distancia, una difusividad o cualquier otro atributo de objeto, en el que el objeto tiene una única dirección a través de todas las bandas de frecuencia y que es estático o se mueve más lento que un umbral de velocidad.

14. Un procedimiento para la generación de una descripción de una escena de audio combinada, que comprende:

la recepción de una primera descripción de una primera escena en un primer formato y la recepción de una segunda descripción de una segunda escena en un segundo formato, en el que el segundo formato es diferente del primer formato;

la conversión de la primera descripción en un formato común y la conversión de la segunda descripción en el formato común, cuando el segundo formato es diferente del formato común; y

la combinación de la primera descripción en el formato común y la segunda descripción en el formato común para obtener la descripción de la escena de audio combinada.

15. Un programa informático configurado para la realización, cuando se ejecuta en un ordenador o un procesador, del procedimiento de la reivindicación 14.

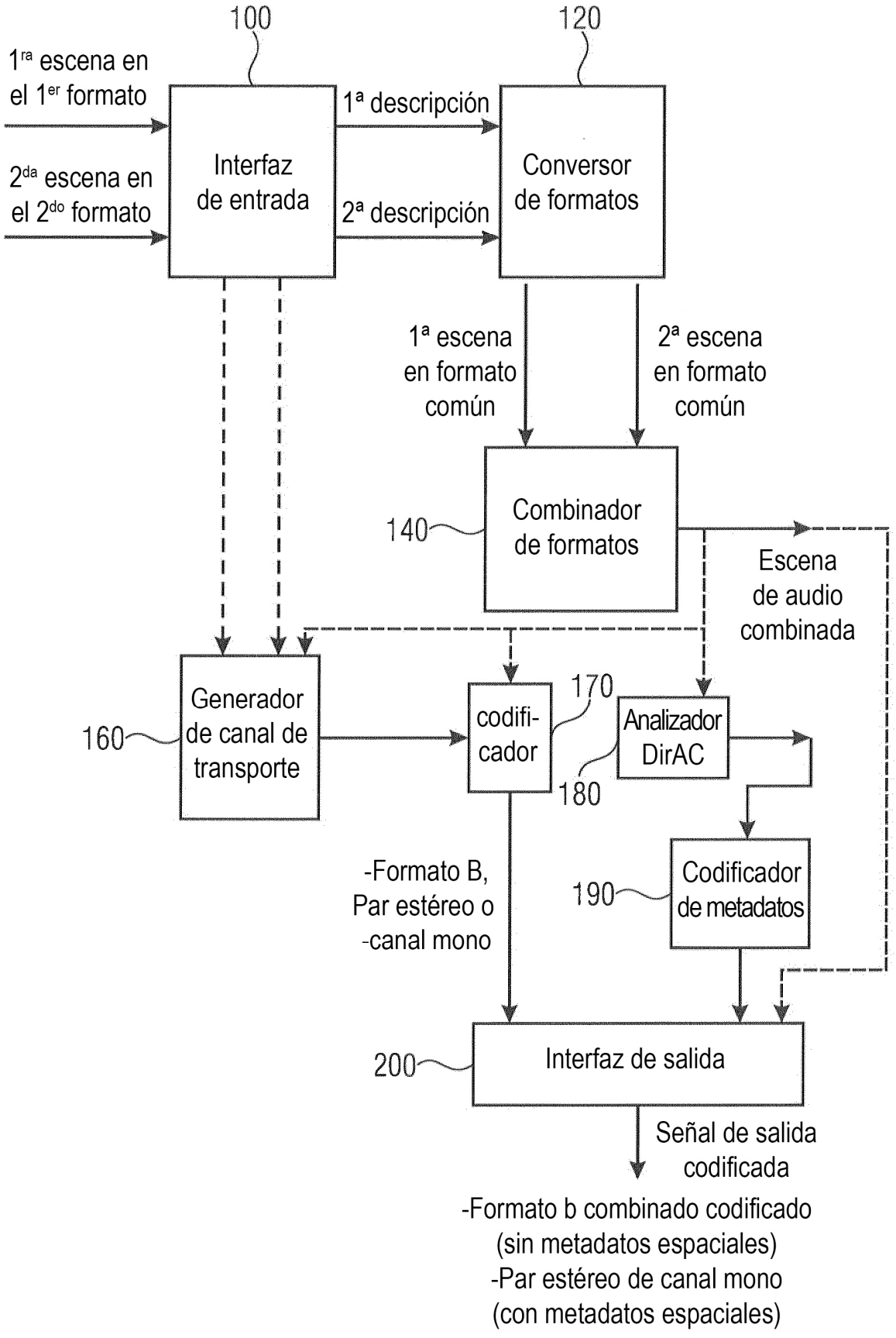
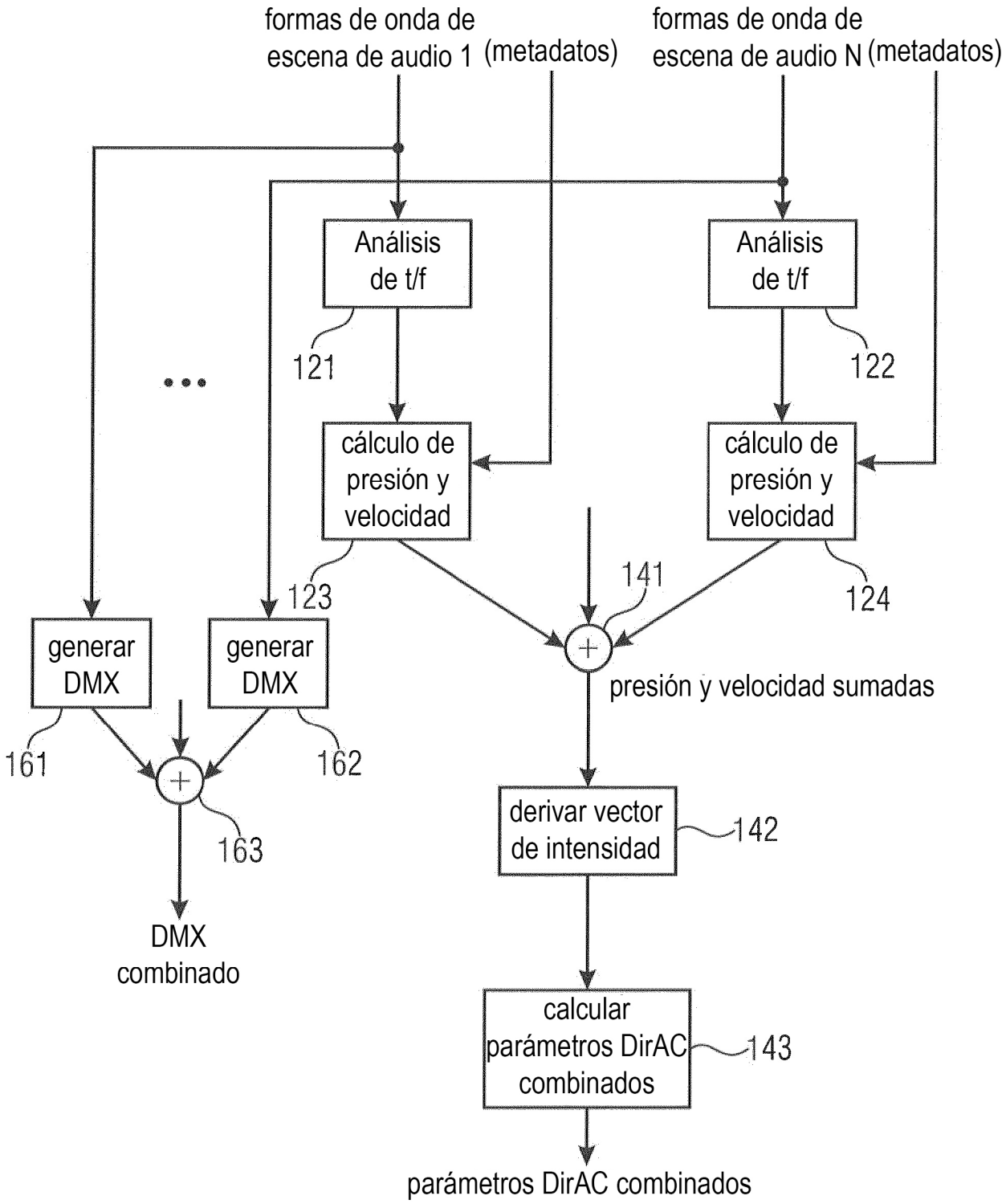
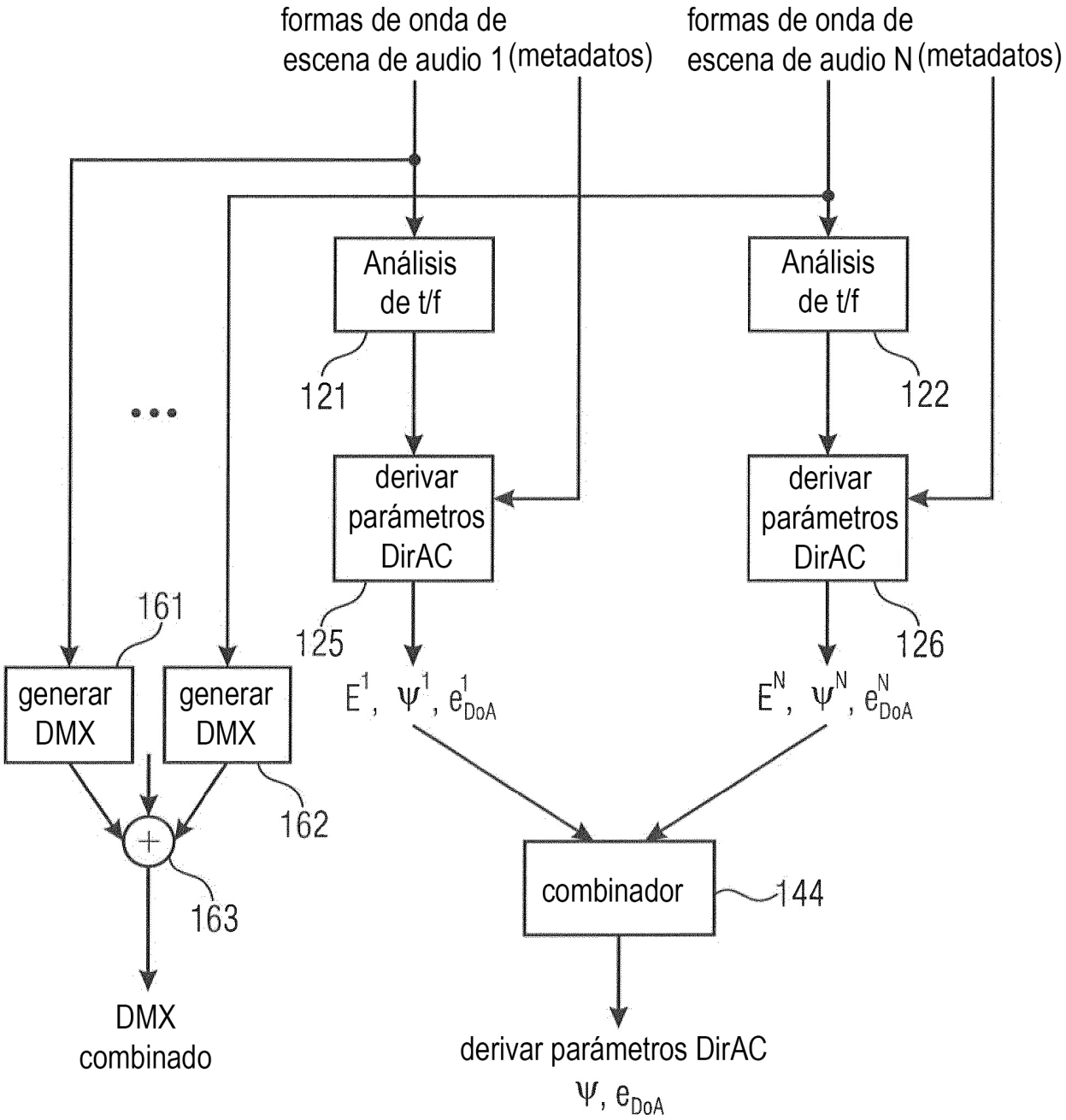

100
120
1ra escena en el 1er formato
2da escena en el 2do formato
Interfaz de entrada
1ª descripción
2ª descripción
Conversor de formatos
1ª escena en formato común
2ª escena en formato común
140
Combinador de formatos
Escena de audio combinada
160
Generador de canal de transporte
codifi-cador
170
Analizador DirAC
180
-Formato B, Par estéreo o -canal mono
190
Codificador de metadatos
200
Interfaz de salida
Señal de salida codificada
-Formato b combinado codificado (sin metadatos espaciales)
-Par estéreo de canal mono (con metadatos espaciales)

Fig. 1a



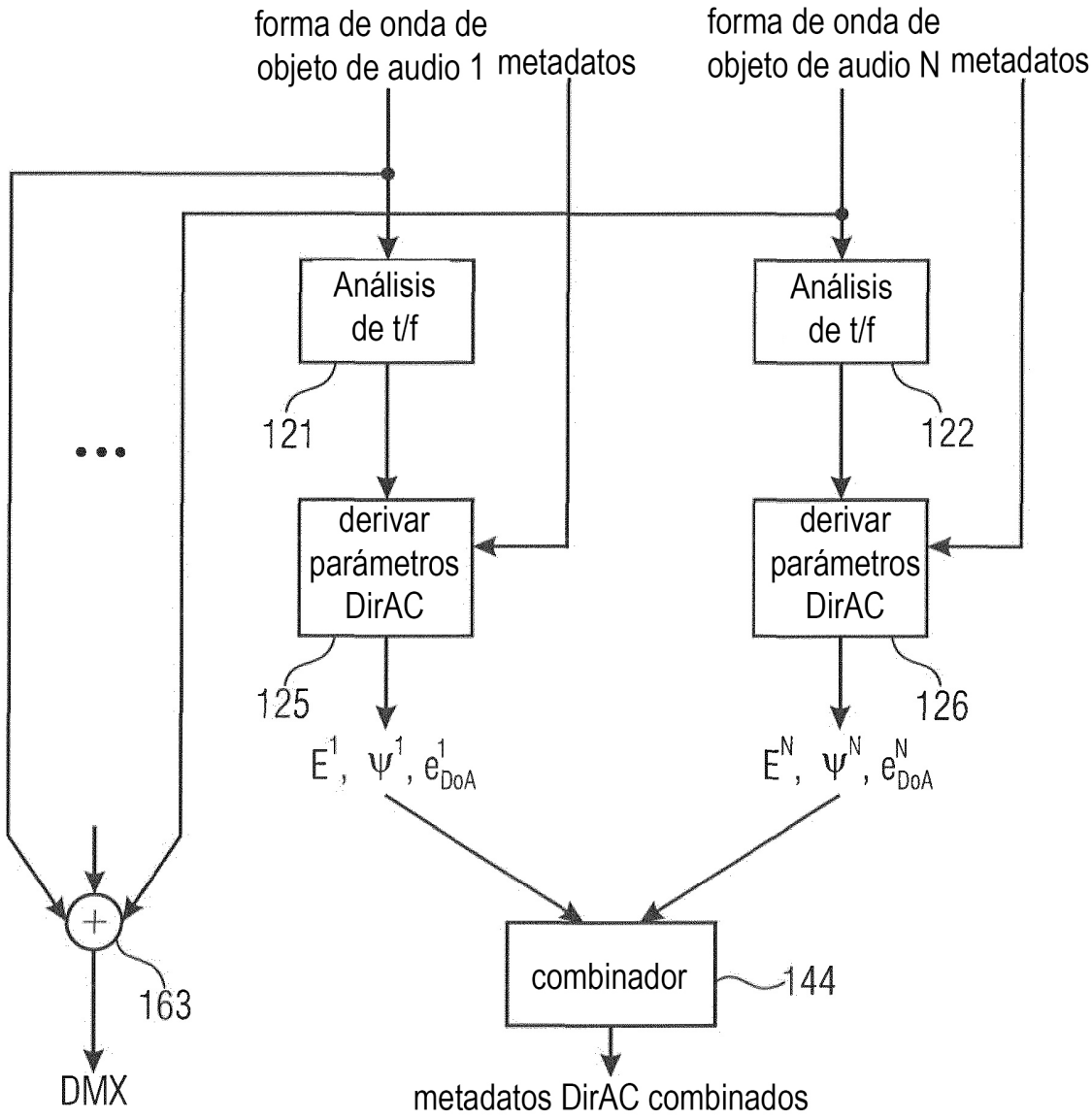
-diagrama de bloques para la combinación de diferentes escenas de audio en diferentes formatos en el dominio de presión/velocidad

Fig. 1b



- diagrama de bloques para la combinación de ≠ escenas de audio en diferentes formatos en el dominio de parámetros DirAC

Fig. 1c



el combinador podría ser:

alt. ①

$$\psi = \frac{1}{\sum E^i} \sum_{i=1}^{N} E^i \psi^i$$

$$e_{DoA} = \frac{1}{\sum (1-\psi^i E^i)} \sum_{i=1}^{N} (1-\psi^i E^i) E^i e^i_{DoA}$$

alt. ②

$\psi = 0$ (ya que los objetos de audio por lo general no tienen difusividad)

$$e_{DoA} = e_{DoA}^{argmax(E^i)}$$

Fig. 1d

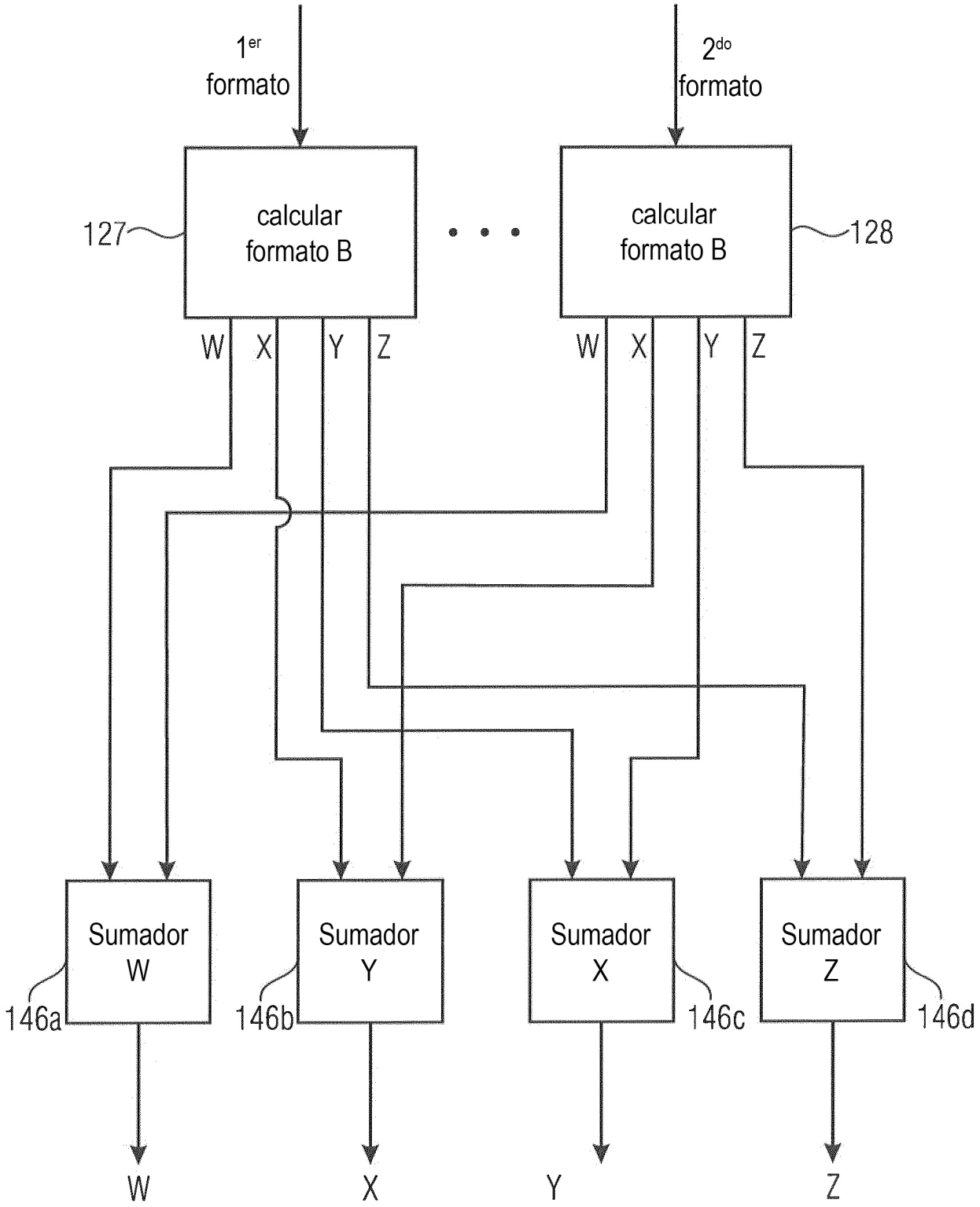
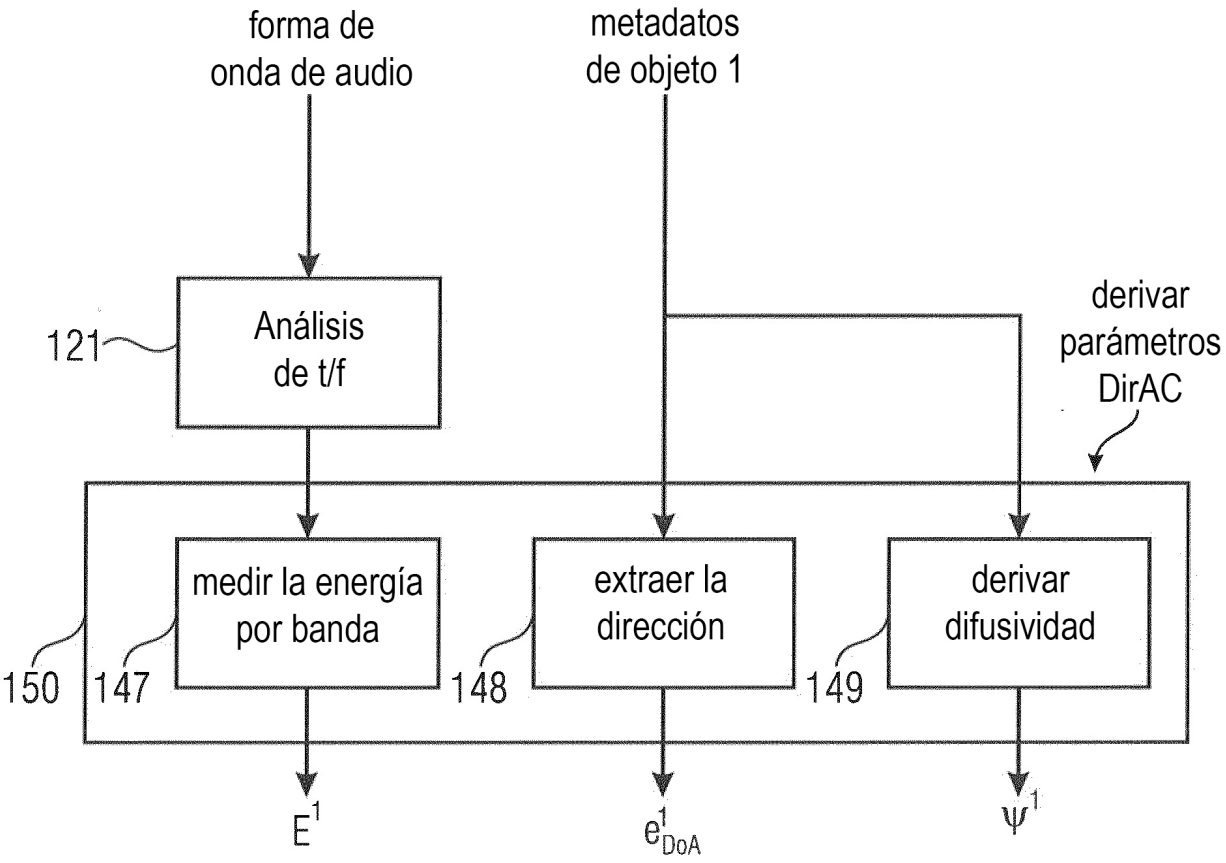

1er formato
2do formato
calcular formato B
127
calcular formato B
128
W X Y Z
W X Y Z
Sumador W
Sumador Y
Sumador X
Sumador Z
146a
146b
146c
146d
W
X
Y
Z

Fig. 1e



-diagrama de bloques para la derivación de parámetros DirAC a partir de un objeto de audio

Fig. 1f

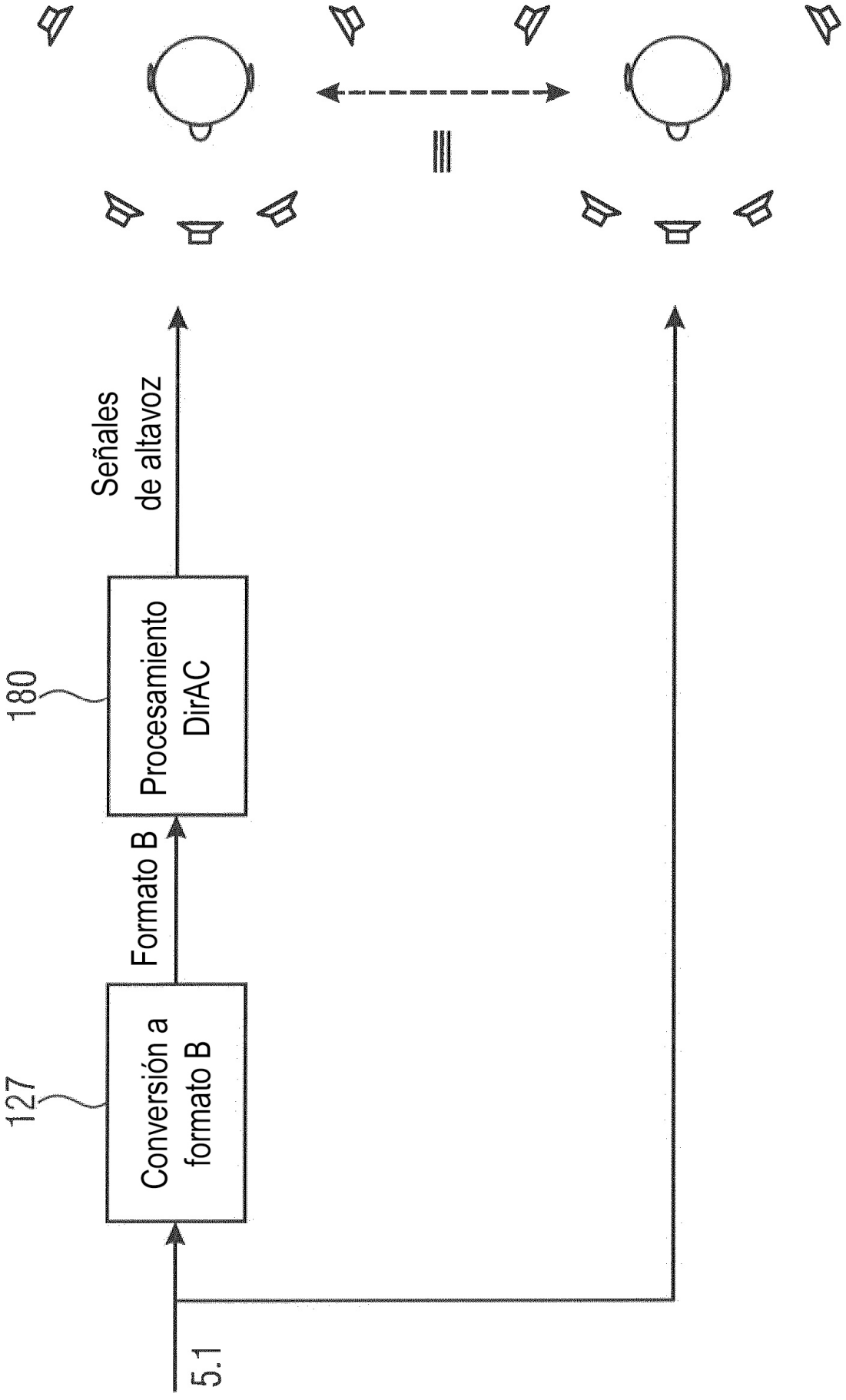

5.1
127
Conversión a formato B
Formato B
180
Procesamiento DirAC
Señales de altavoz

Fig. 1g

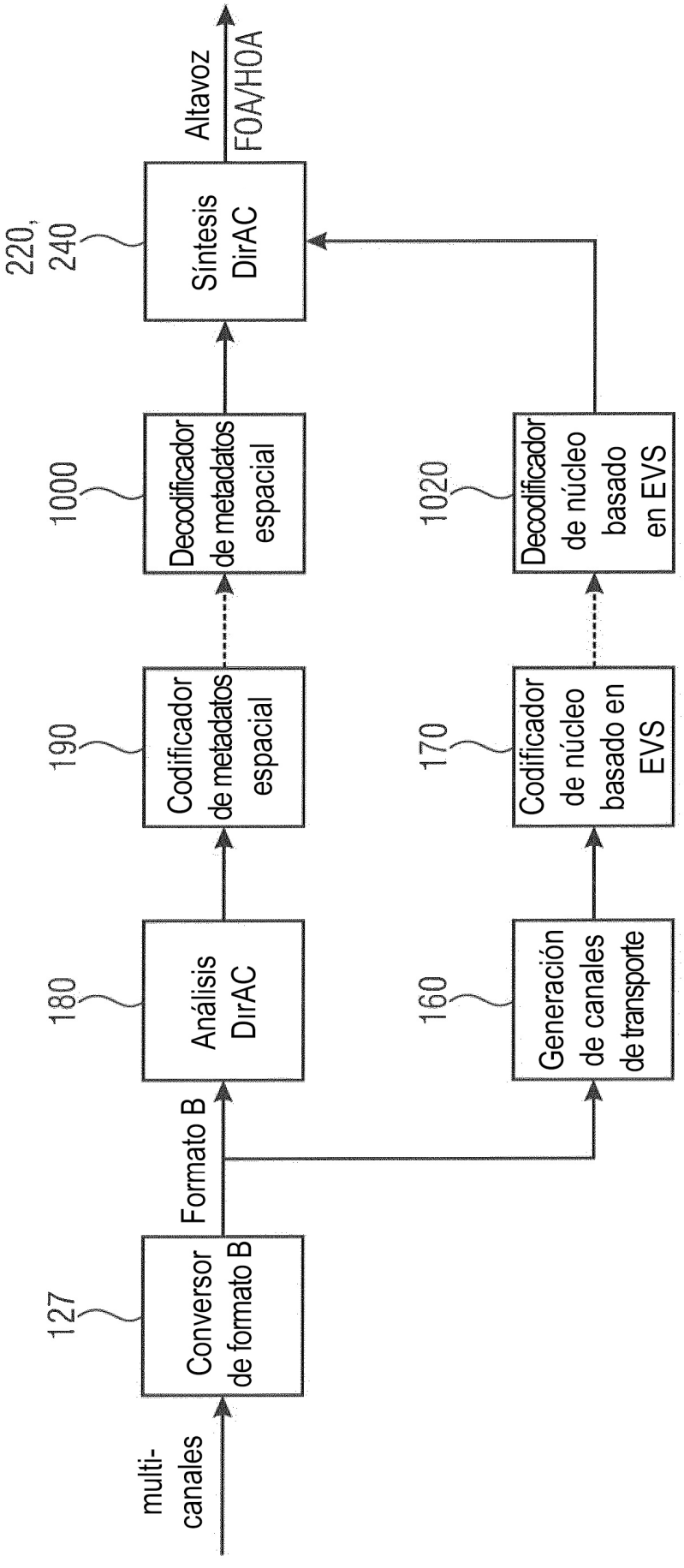

multi-canales
127
Conversor de formato B
Formato B
180
Análisis DirAC
190
Codificador de metadatos espacial
1000
Decodificador de metadatos espacial
160
Generación de canales de transporte
170
Codificador de núcleo basado en EVS
1020
Decodificador de núcleo basado en EVS
220, 240
Síntesis DirAC
Altavoz FOA/HOA

Fig. 1h

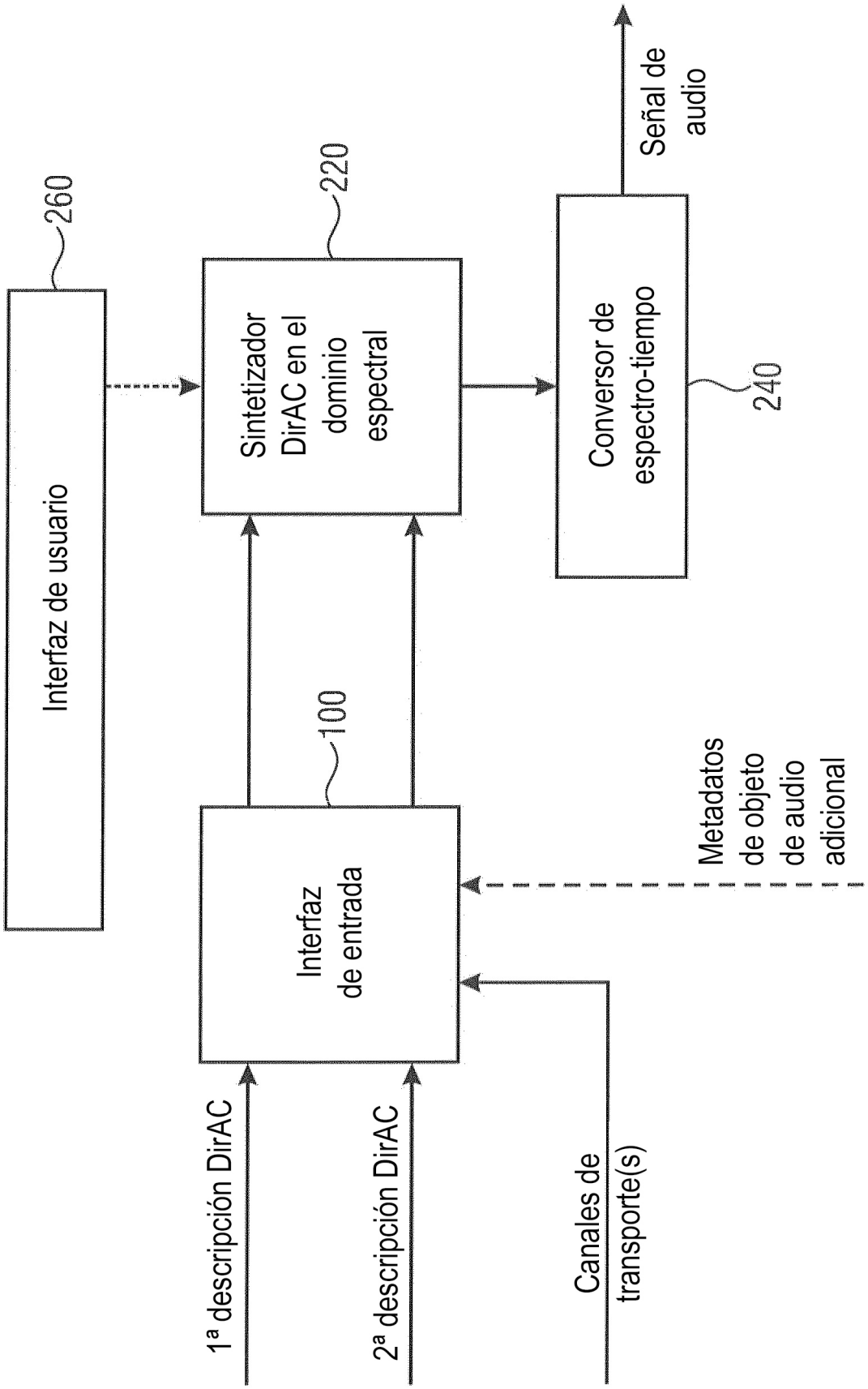
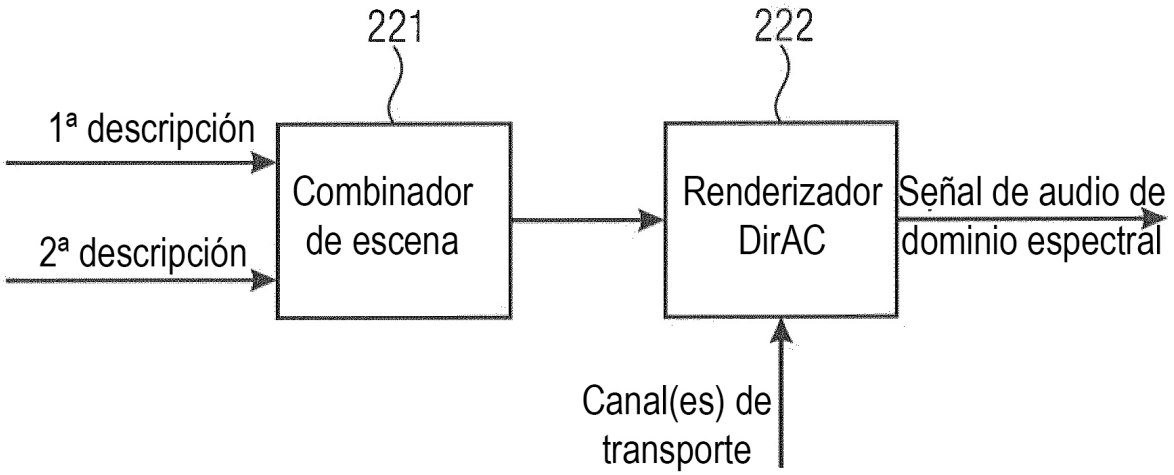
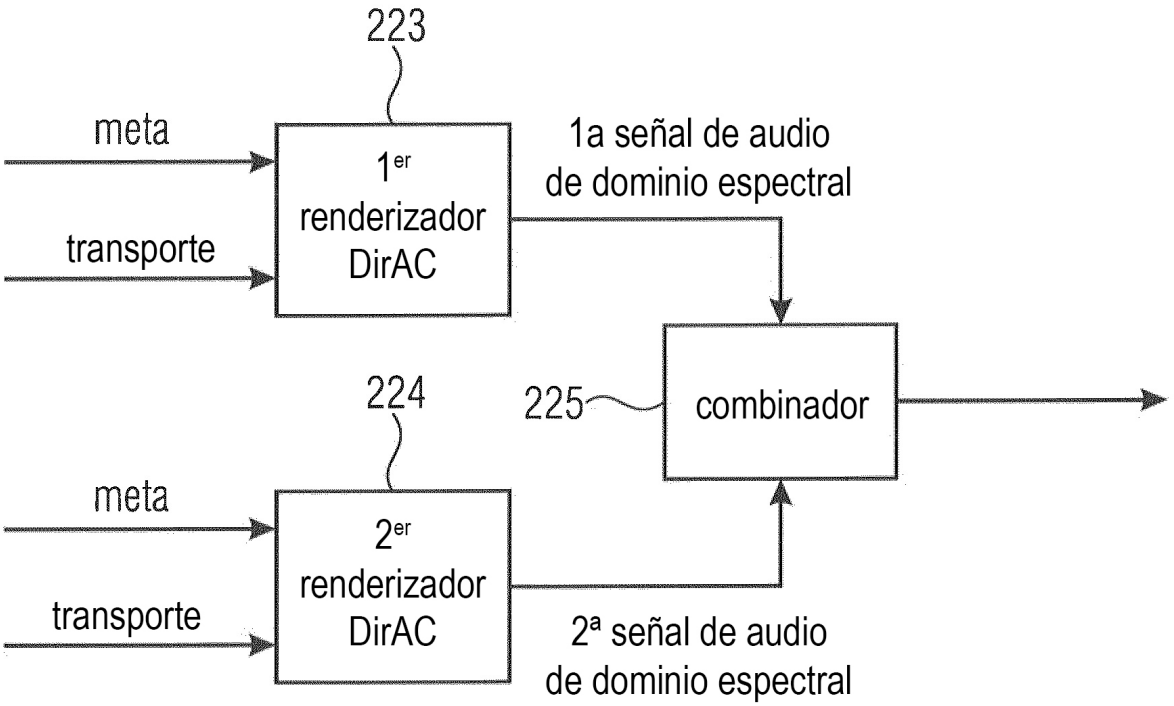

Interfaz de usuario
260
Sintetizador DirAC en el dominio espectral
220
Conversor de espectro-tiempo
240
Señal de audio
Interfaz de entrada
100
1ª descripción DirAC
2ª descripción DirAC
Canales de transporte(s)
Metadatos de objeto de audio adicional

Fig. 2a



Sintetizador DirAC

Fig. 2b



Sintetizador DirAC

Fig. 2c

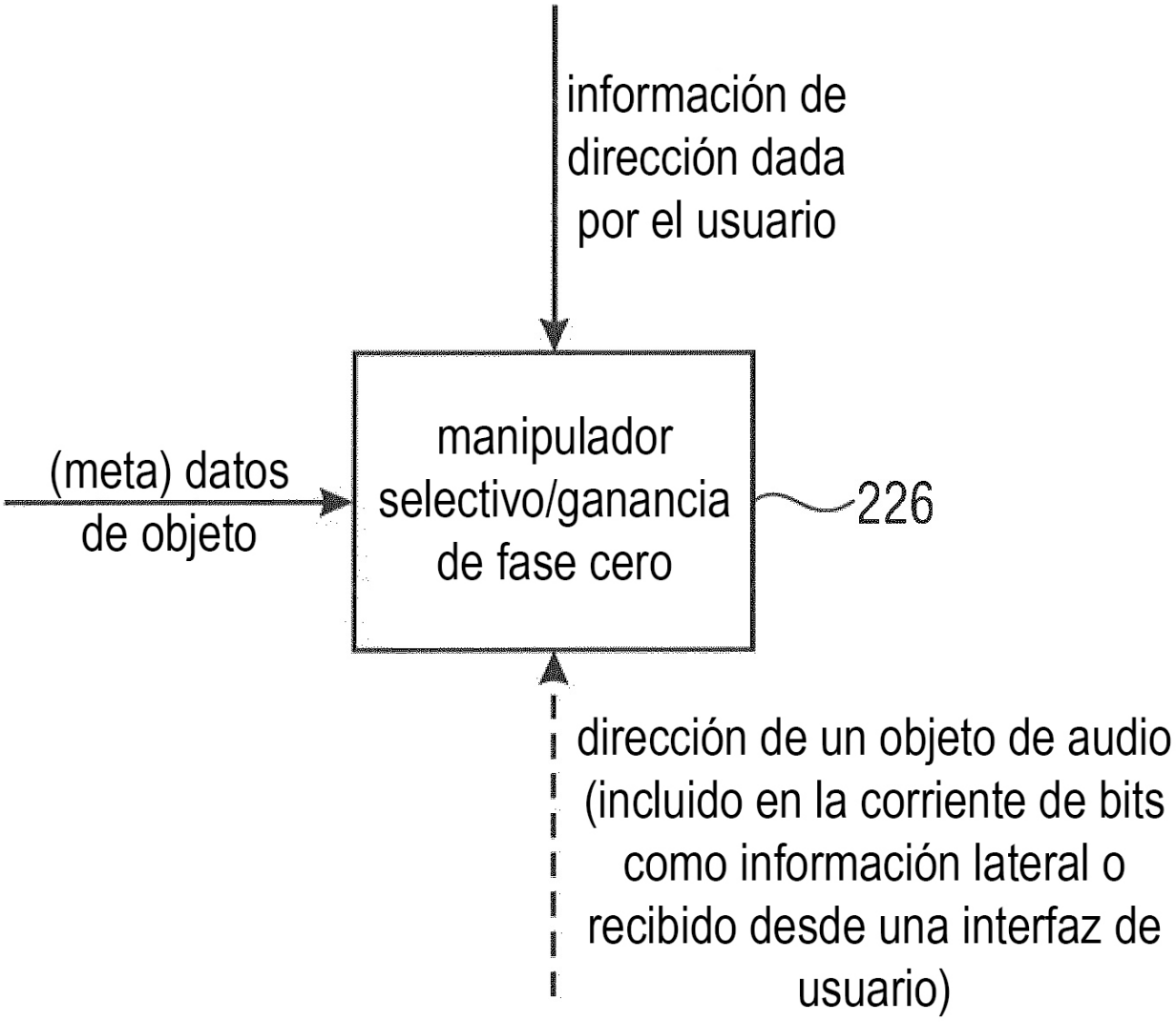

información de dirección dada por el usuario
(meta) datos de objeto
manipulador selectivo/ganancia de fase cero
226
dirección de un objeto de audio (incluido en la corriente de bits como información lateral o recibido desde una interfaz de usuario)

Fig. 2d

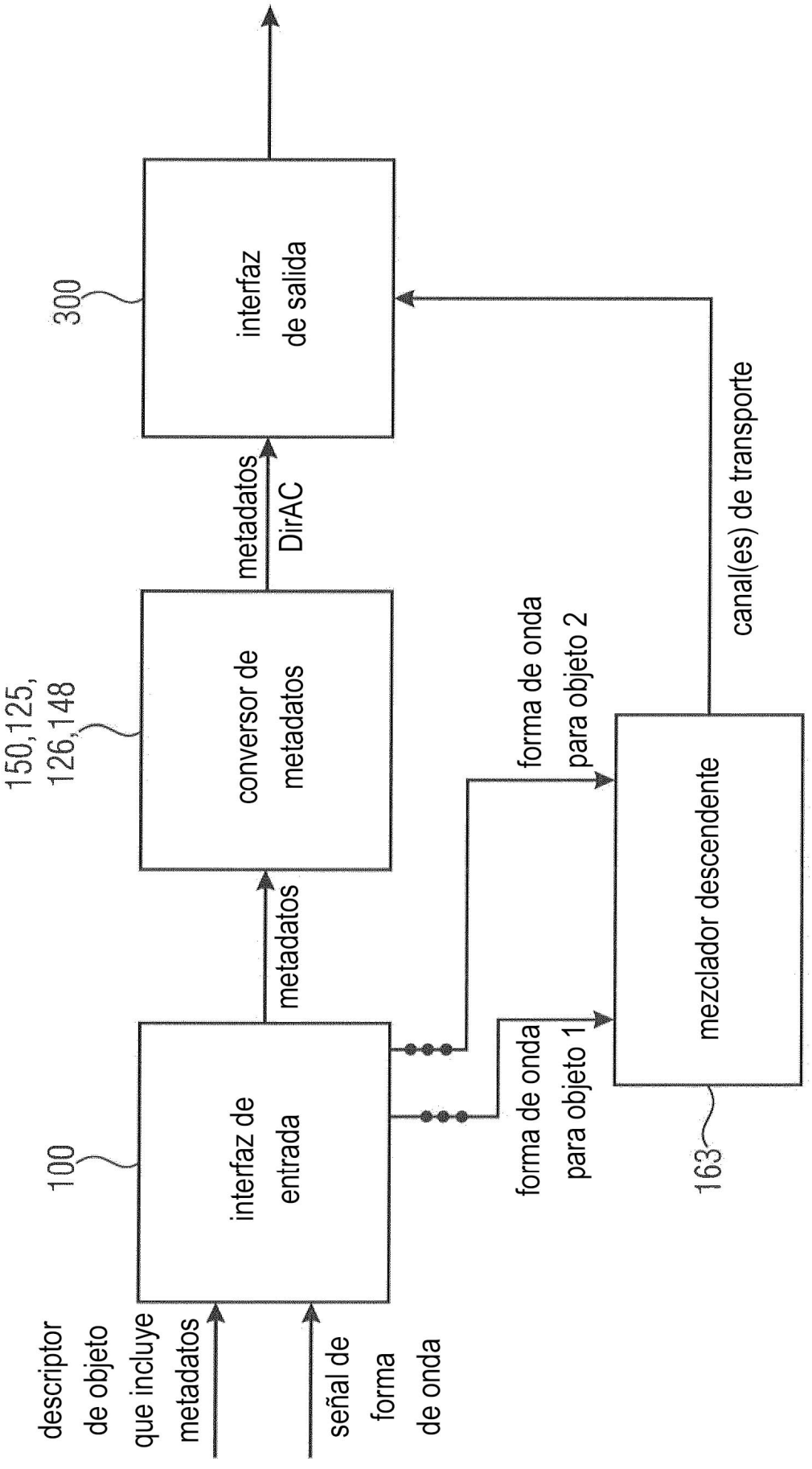

descriptor de objeto que incluye metadatos
señal de forma de onda
100
interfaz de entrada
metadatos
150,125, 126,148
conversor de metadatos
metadatos
DirAC
300
interfaz de salida
forma de onda para objeto 1
forma de onda para objeto 2
163
mezclador descendente
canal(es) de transporte

Fig. 3a

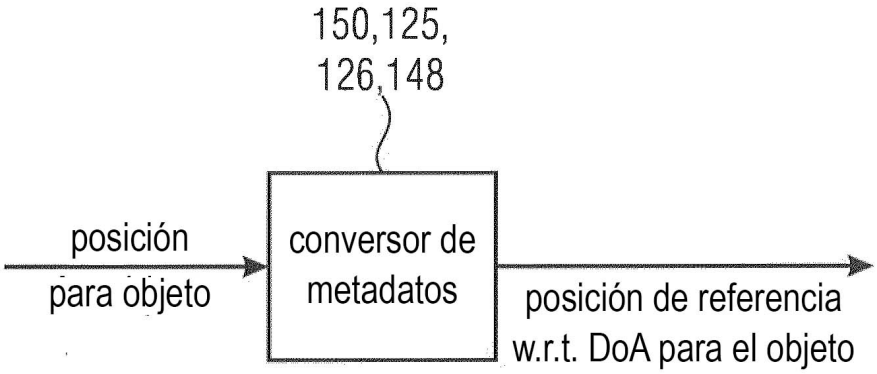

150,125,
126,148
posición
para objeto
conversor de
metadatos
posición de referencia
w.r.t. DoA para el objeto

Fig. 3b

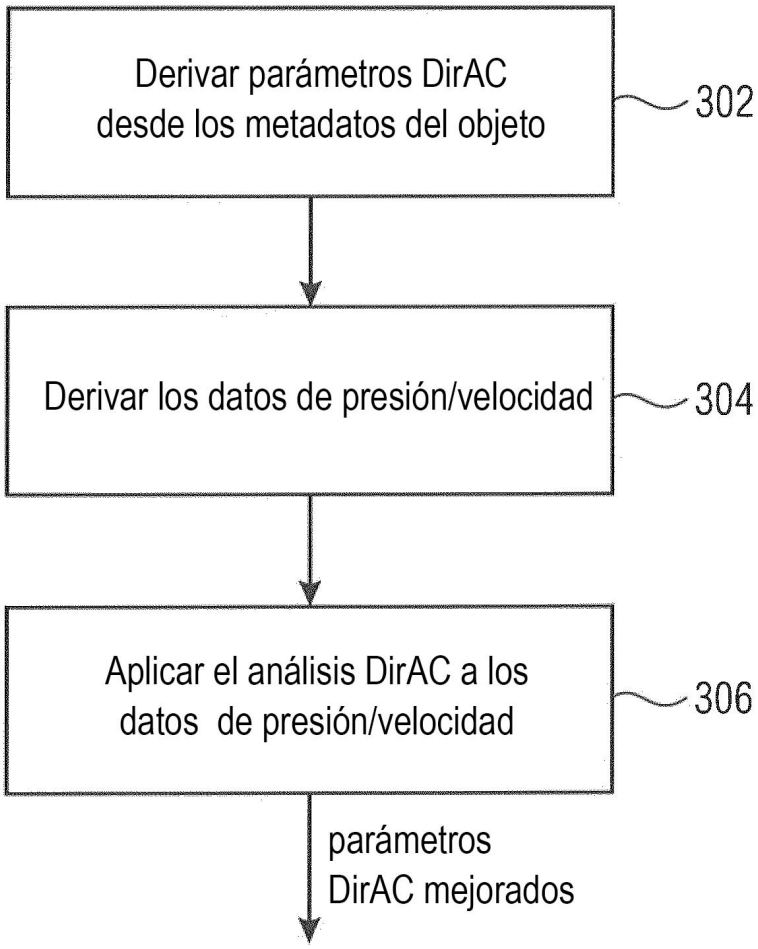

Derivar parámetros DirAC
desde los metadatos del objeto
302
Derivar los datos de presión/velocidad
304
Aplicar el análisis DirAC a los
datos de presión/velocidad
306
parámetros
DirAC mejorados

Fig. 3c

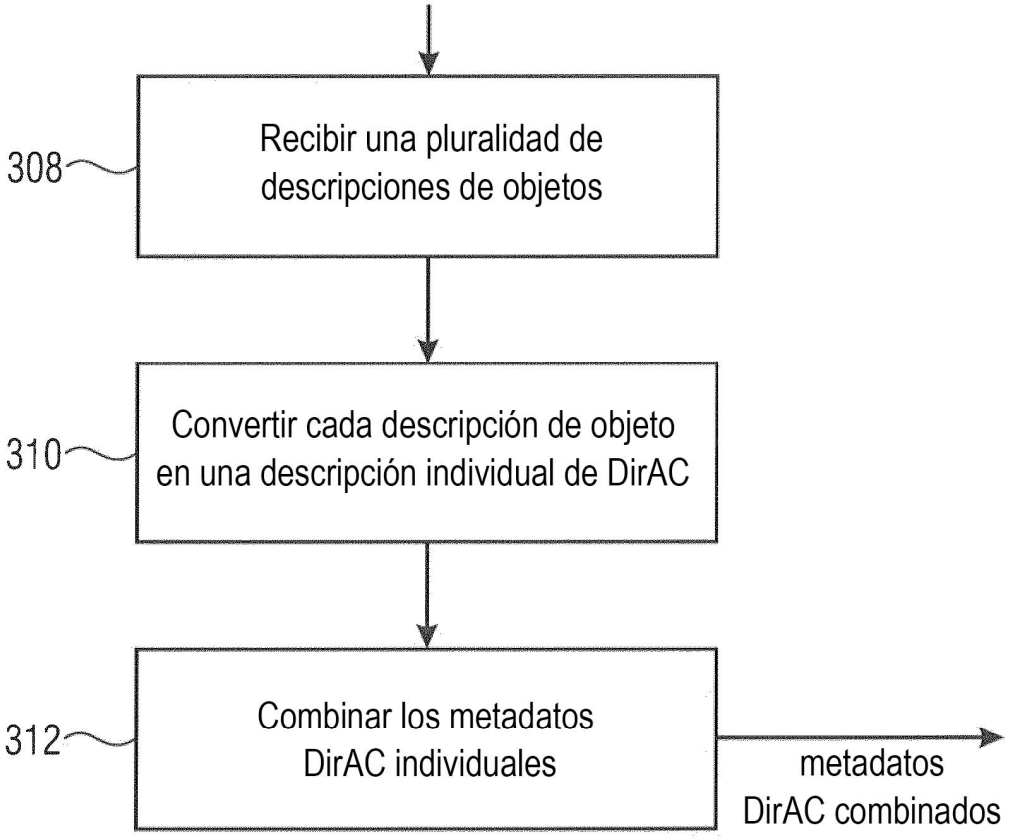

308
Recibir una pluralidad de descripciones de objetos
310
Convertir cada descripción de objeto en una descripción individual de DirAC
312
Combinar los metadatos DirAC individuales
metadatos DirAC combinados

Fig. 3d

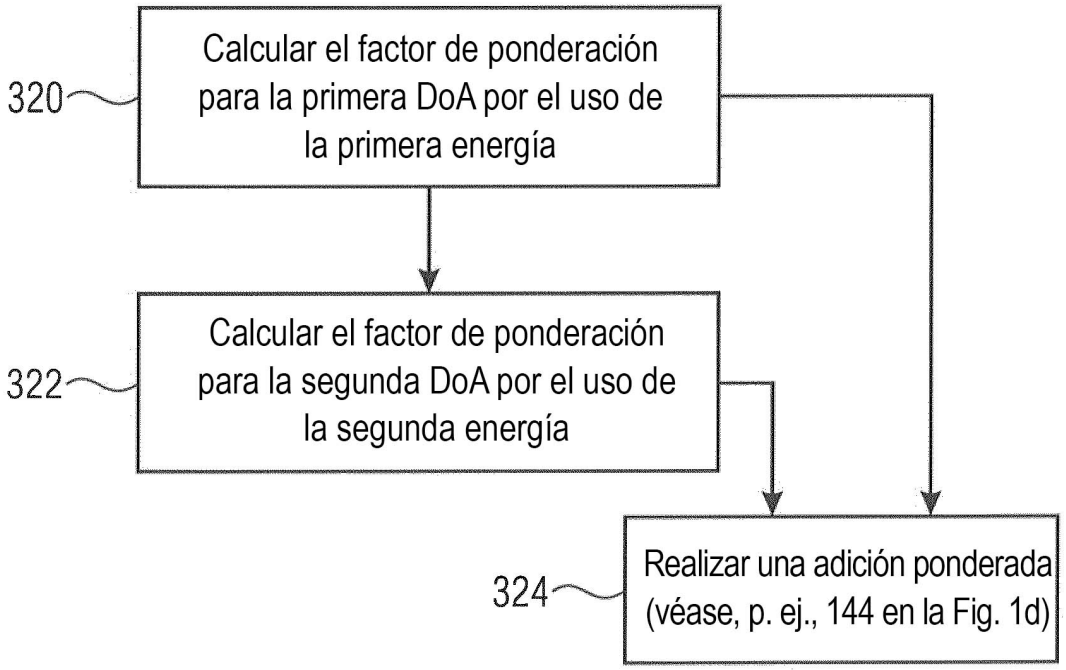
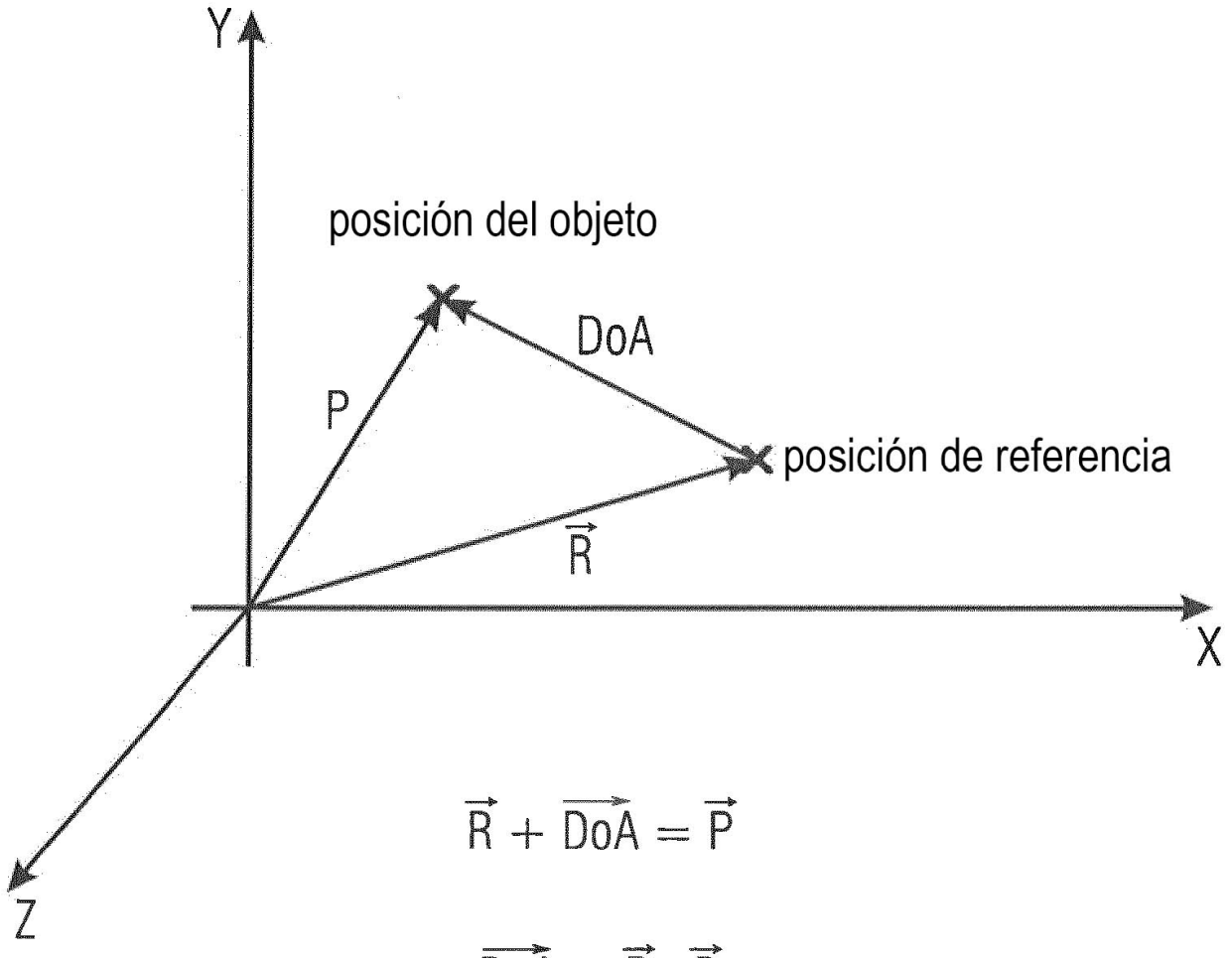

320
Calcular el factor de ponderación para la primera DoA por el uso de la primera energía
322
Calcular el factor de ponderación para la segunda DoA por el uso de la segunda energía
324
Realizar una adición ponderada (véase, p. ej., 144 en la Fig. 1d)

Fig. 3e



$$\vec{R} + \overrightarrow{DoA} = \vec{P}$$

$$\overrightarrow{DoA} = \vec{P} - \vec{R}$$

$$\overrightarrow{doa} = \frac{\vec{P} - \vec{R}}{|\overrightarrow{DoA}|}$$

Fig. 3f

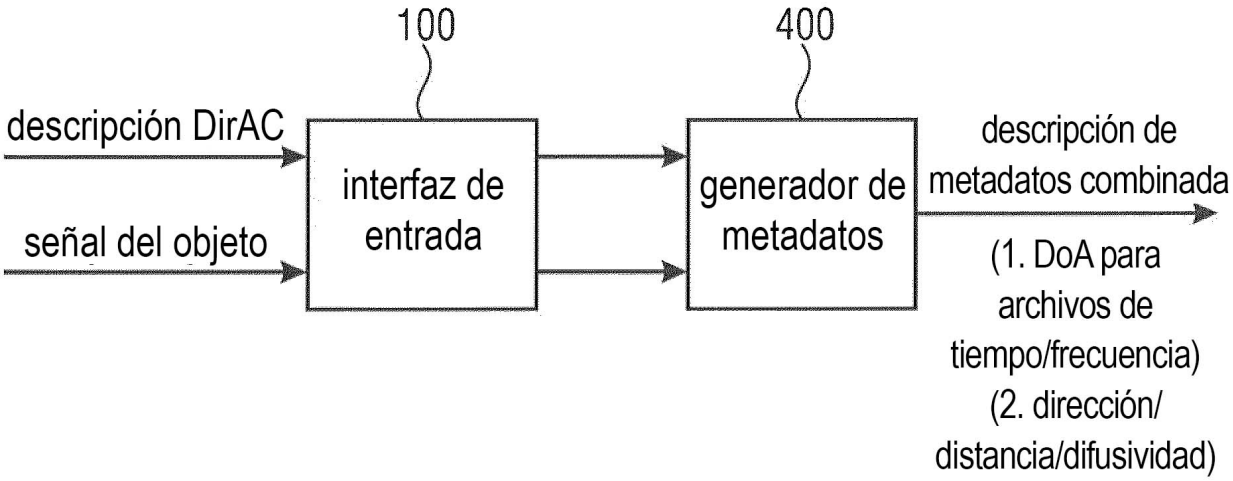

100
400
descripción DirAC
señal del objeto
interfaz de entrada
generador de metadatos
descripción de metadatos combinada
(1. DoA para archivos de tiempo/frecuencia)
(2. dirección/ distancia/difusividad)

Fig. 4a

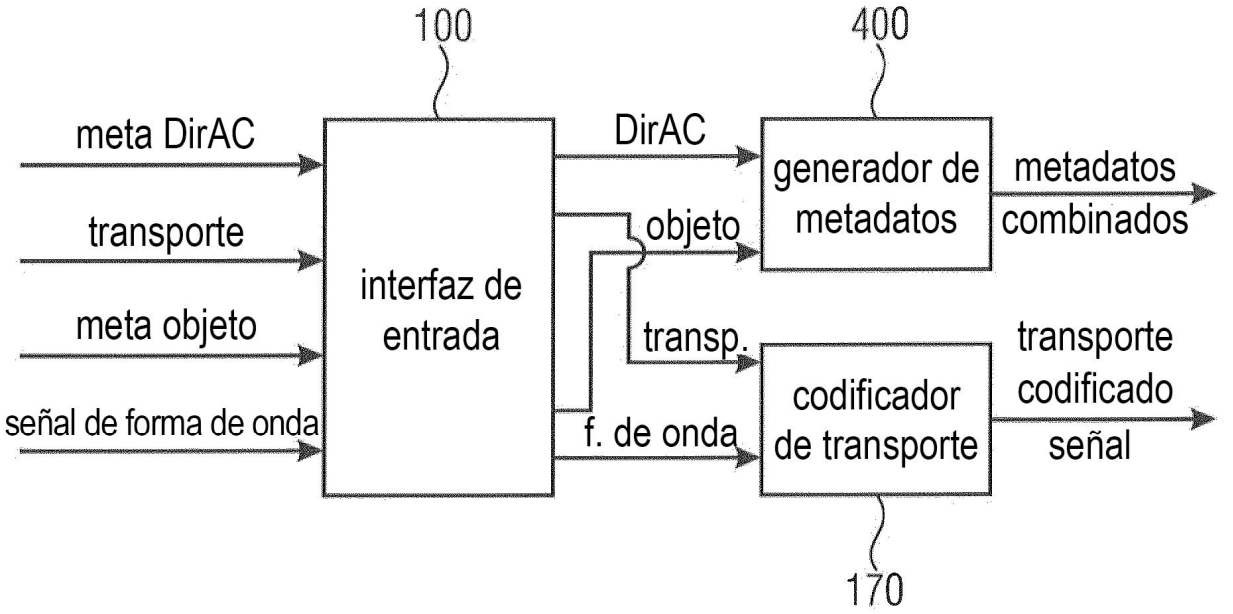

100
400
meta DirAC
transporte
meta objeto
señal de forma de onda
interfaz de entrada
DirAC
objeto
transp.
f. de onda
generador de metadatos
metadatos combinados
codificador de transporte
transporte codificado señal
170

Fig. 4b

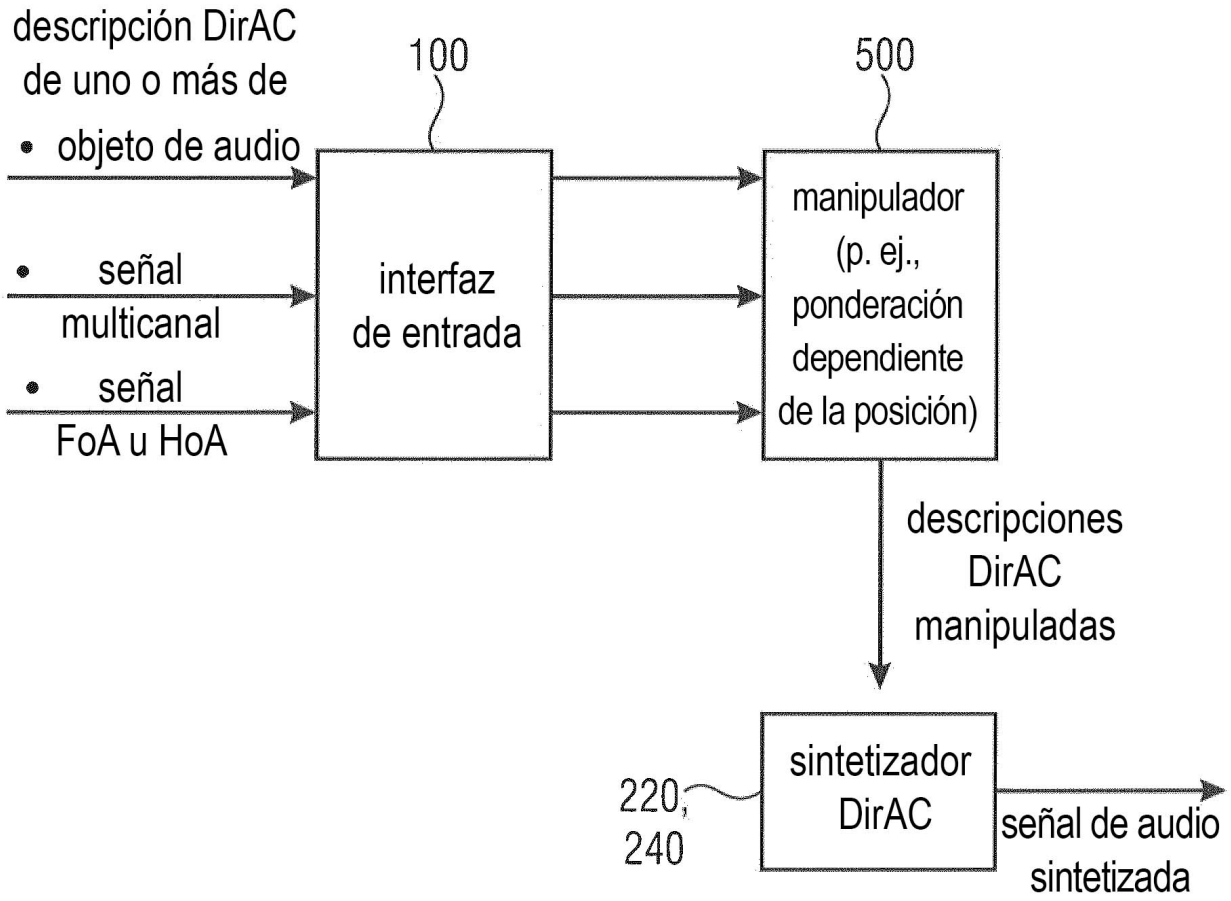

descripción DirAC
de uno o más de
• objeto de audio
• señal multicanal
• señal FoA u HoA
100
interfaz de entrada
500
manipulador (p. ej., ponderación dependiente de la posición)
descripciones DirAC manipuladas
220, 240
sintetizador DirAC
señal de audio sintetizada

Fig. 5a

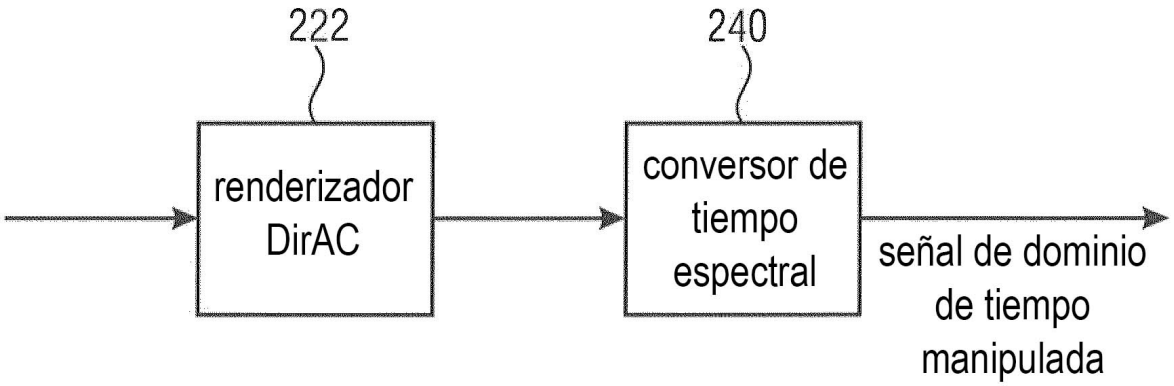

222
renderizador DirAC
240
conversor de tiempo espectral
señal de dominio de tiempo manipulada

Fig. 5b

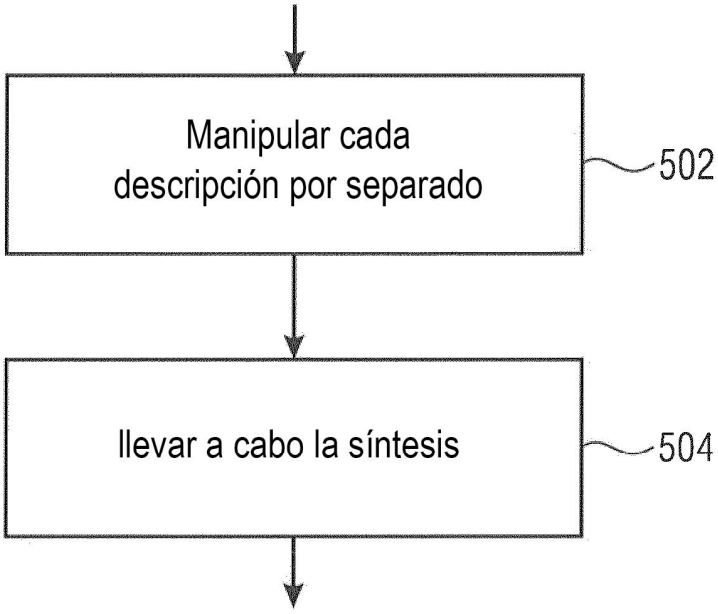

Manipular cada descripción por separado
502
llevar a cabo la síntesis
504

Fig. 5c

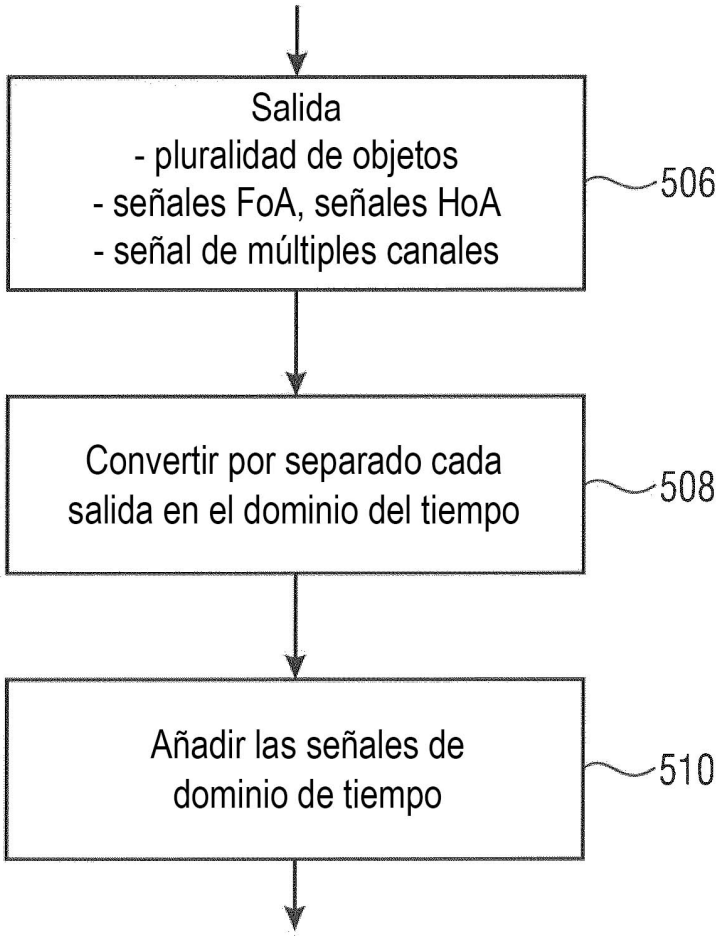
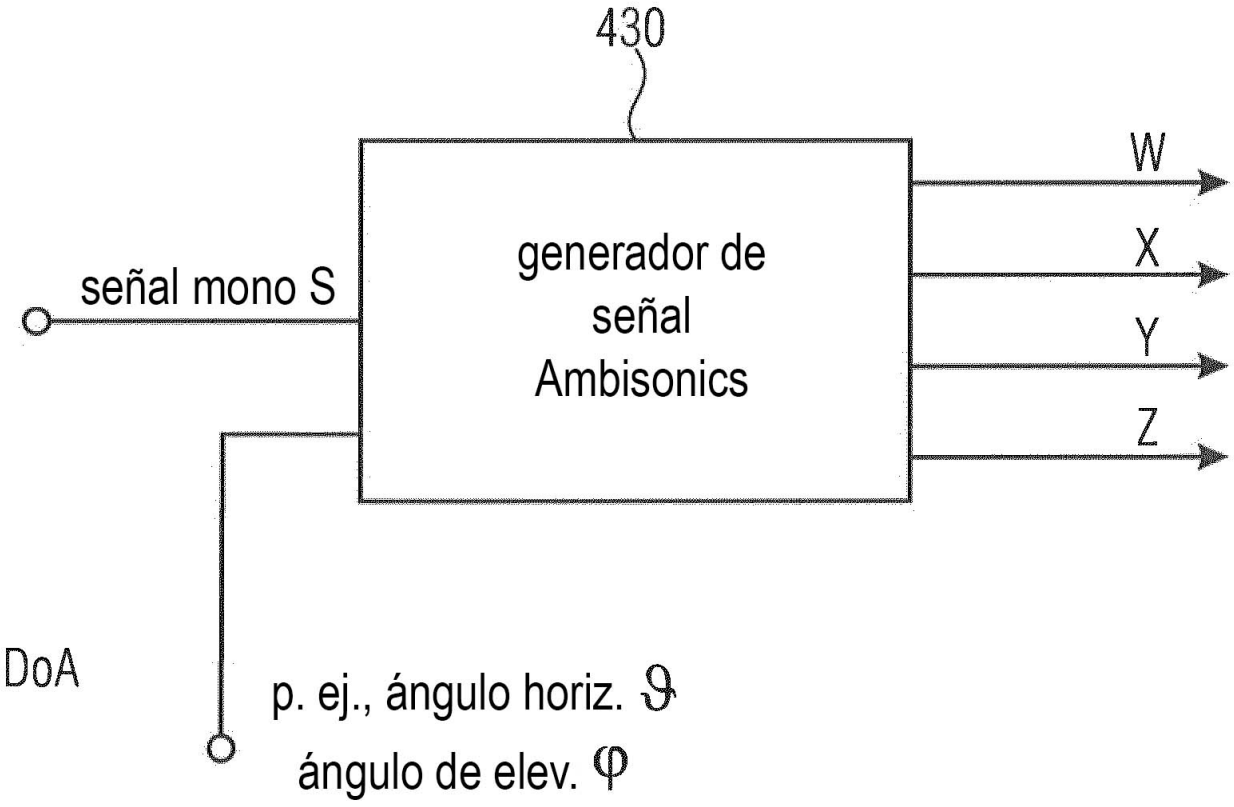

Salida
- pluralidad de objetos
- señales FoA, señales HoA
- señal de múltiples canales
506
Convertir por separado cada salida en el dominio del tiempo
508
Añadir las señales de dominio de tiempo
510

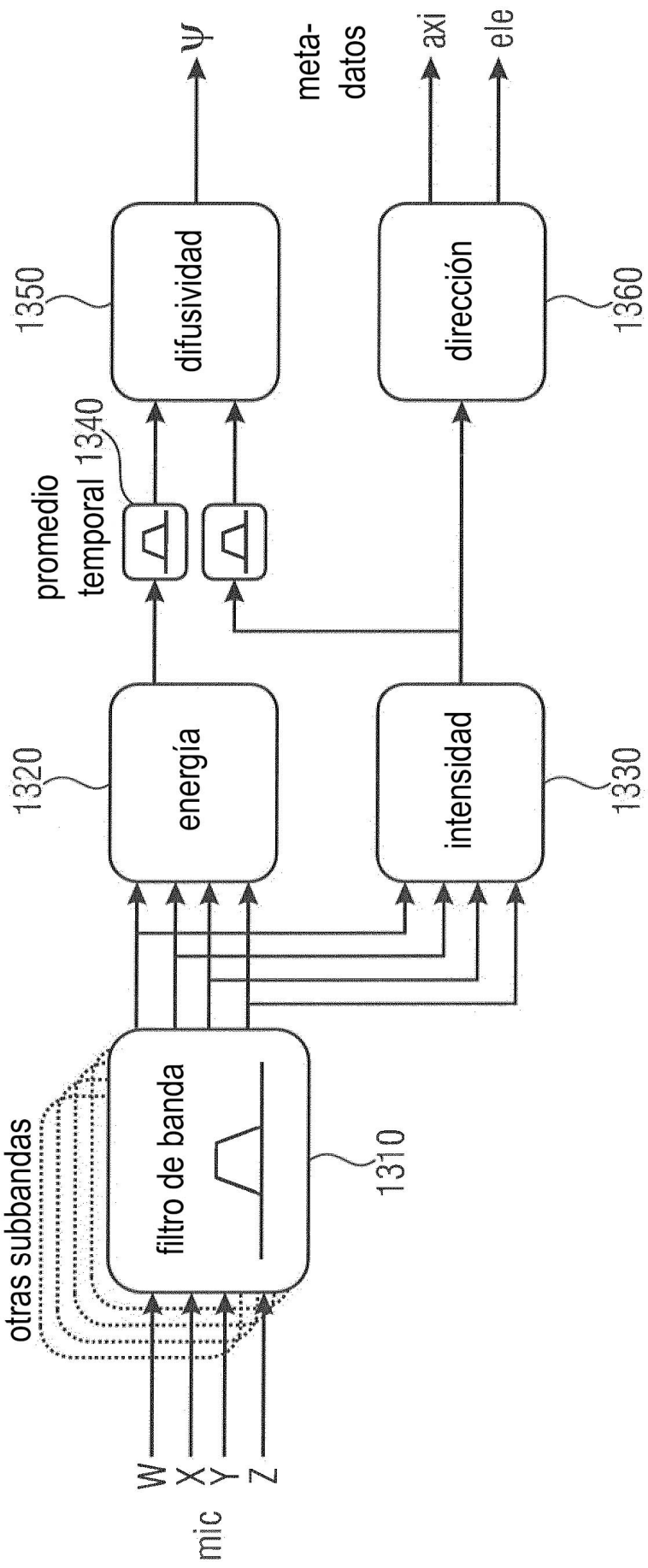
Fig. 5d



$$\text{componente omnidireccional } W = S \cdot \frac{1}{\sqrt{2}}$$

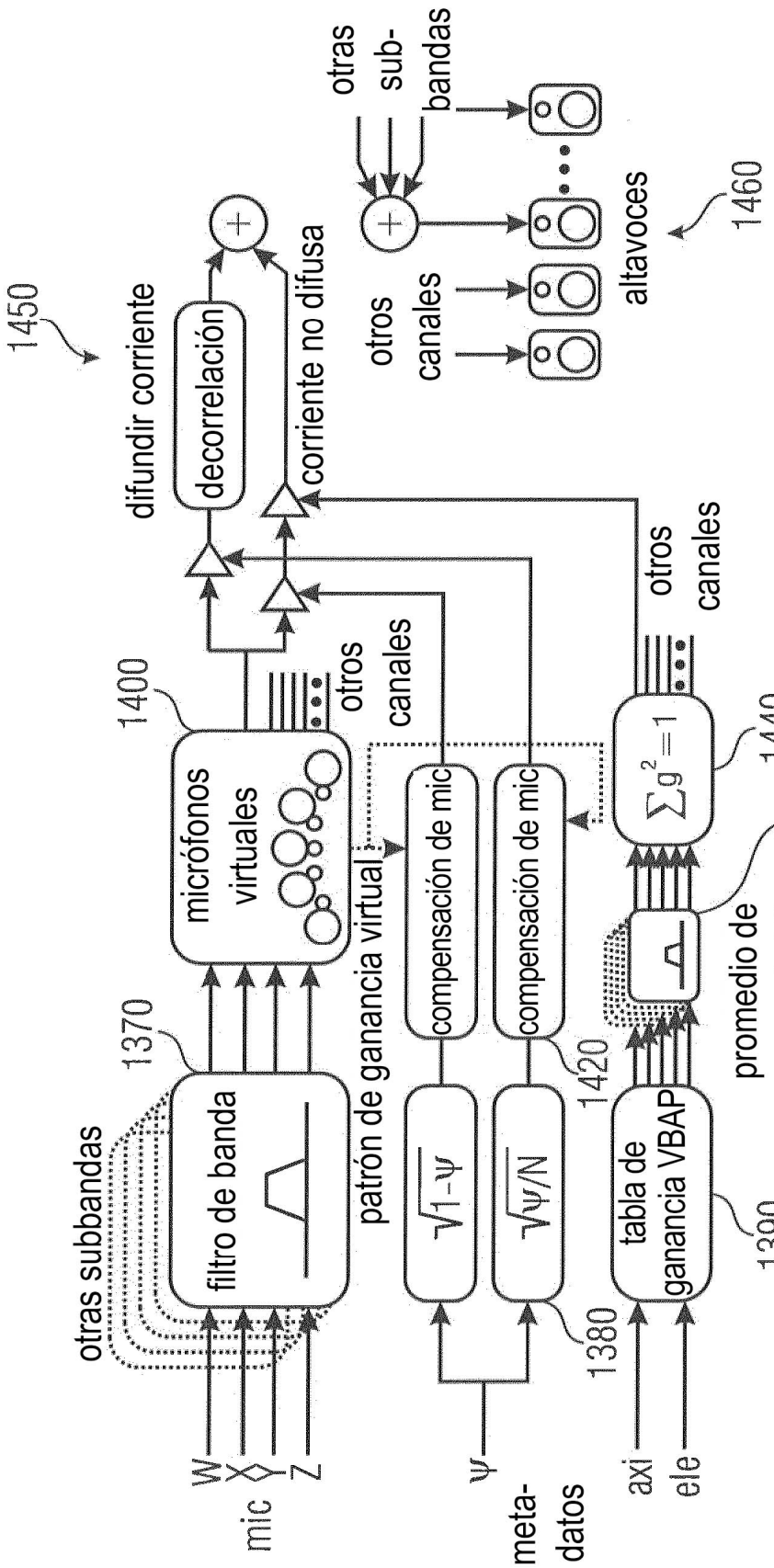
$$\text{componentes direccionales} \begin{cases} X = S \cdot \cos\vartheta \cos\varphi \\ Y = S \cdot \sin\vartheta \cos\varphi \\ Z = S \cdot \sin\varphi \end{cases}$$

Fig. 6



Análisis DirAC

Fig. 7a
(TÉCNICA ANTERIOR)



Análisis DirAC

Fig. 7b
(TÉCNICA ANTERIOR)

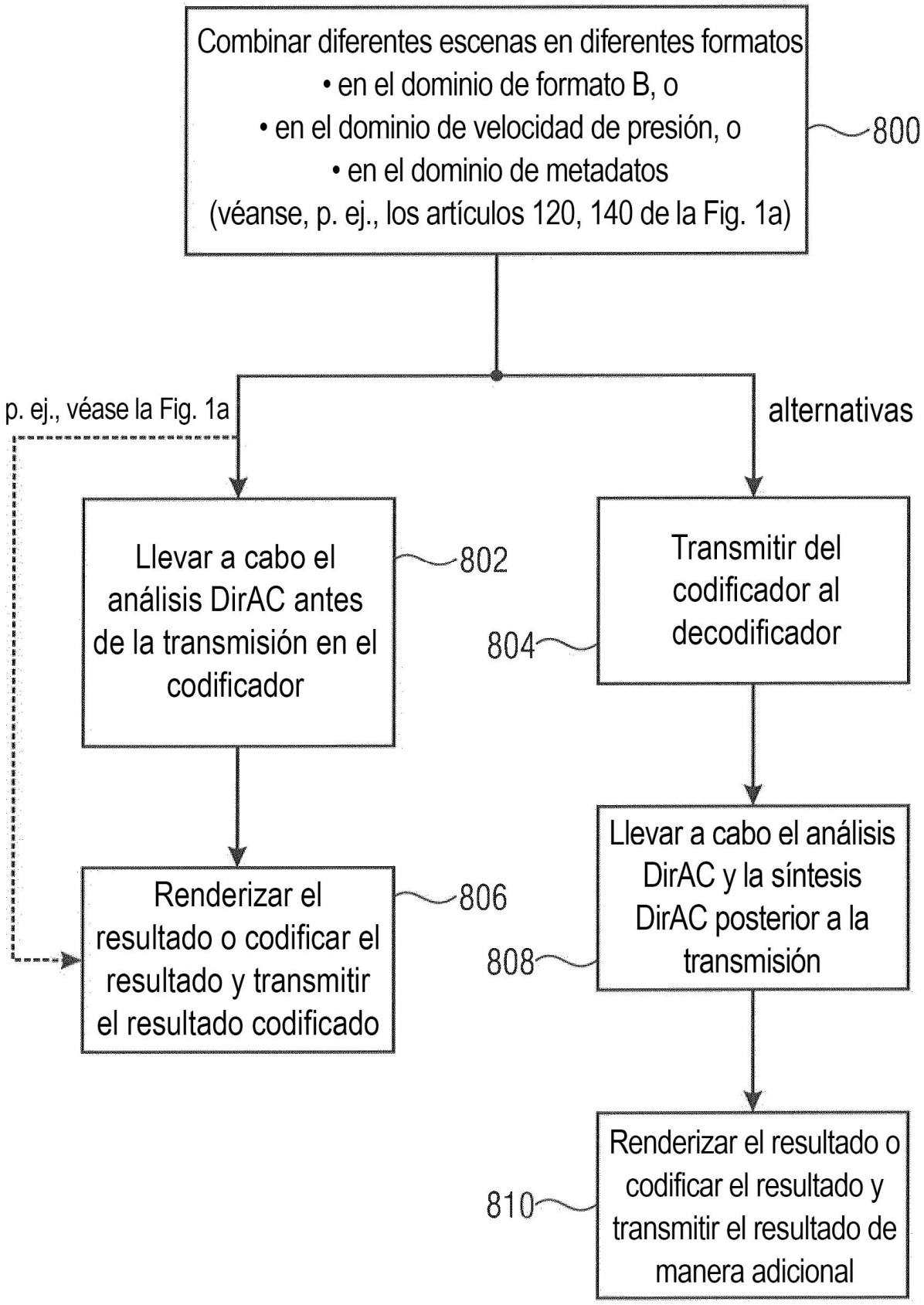

Combinar diferentes escenas en diferentes formatos
• en el dominio de formato B, o
• en el dominio de velocidad de presión, o
• en el dominio de metadatos
(véanse, p. ej., los artículos 120, 140 de la Fig. 1a)
800
p. ej., véase la Fig. 1a
alternativas
Llevar a cabo el análisis DirAC antes de la transmisión en el codificador
802
Transmitir del codificador al decodificador
804
Renderizar el resultado o codificar el resultado y transmitir el resultado codificado
806
Llevar a cabo el análisis DirAC y la síntesis DirAC posterior a la transmisión
808
Renderizar el resultado o codificar el resultado y transmitir el resultado de manera adicional
810

Fig. 8

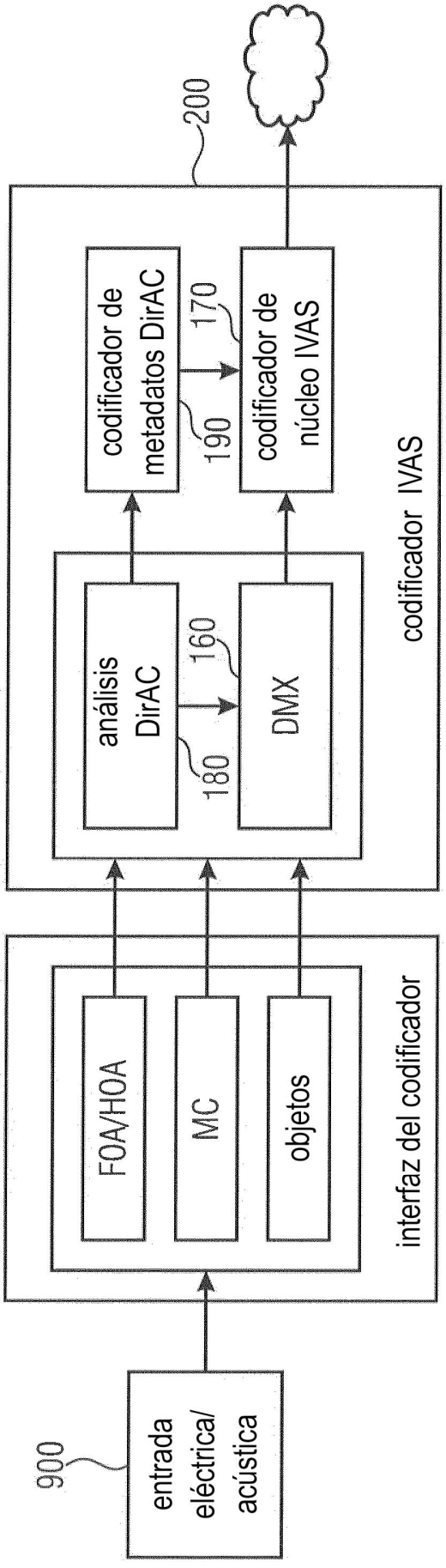

900
entrada eléctrica/ acústica
interfaz del codificador
FOA/HOA
MC
objetos
codificador IVAS
análisis DirAC
180
160
DMX
codificador de metadatos DirAC
190
170
codificador de núcleo IVAS
200

Fig. 9

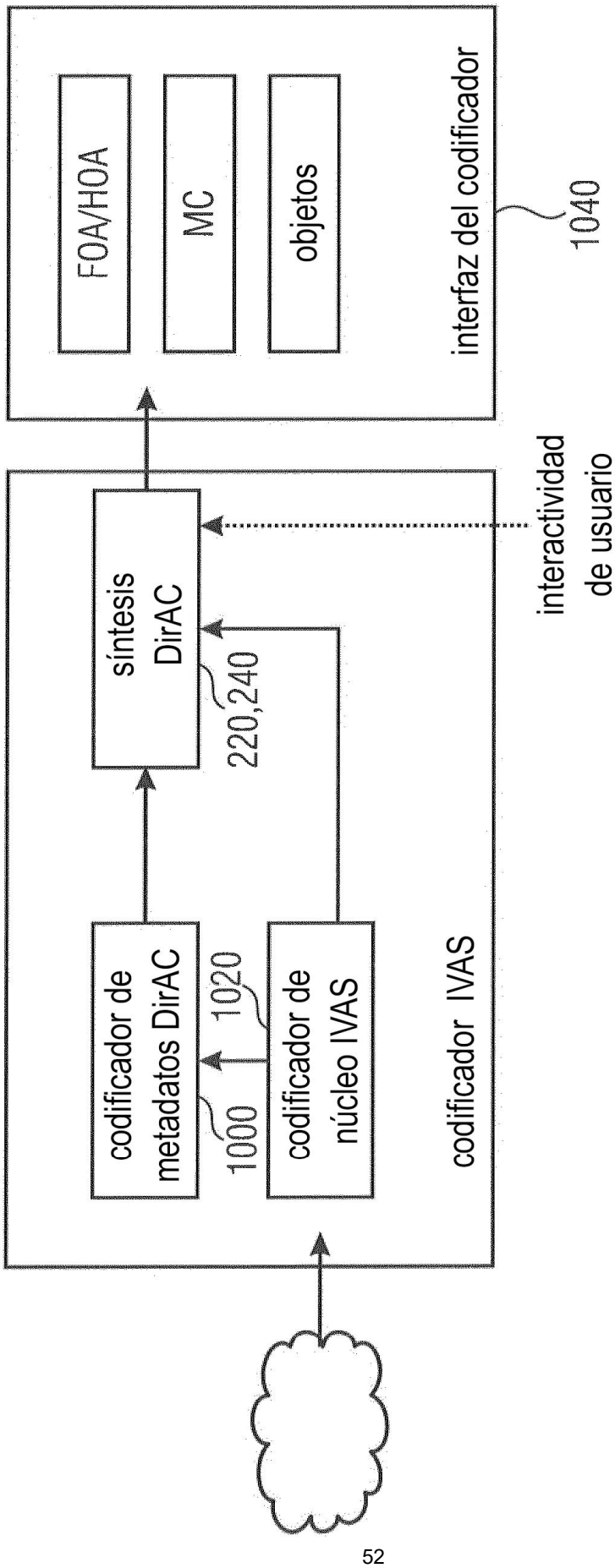

FOA/HOA
MC
objetos
interfaz del codificador
1040
interactividad de usuario
síntesis DirAC
220,240
codificador de metadatos DirAC
1000
1020
codificador de núcleo IVAS
codificador IVAS

Fig. 10

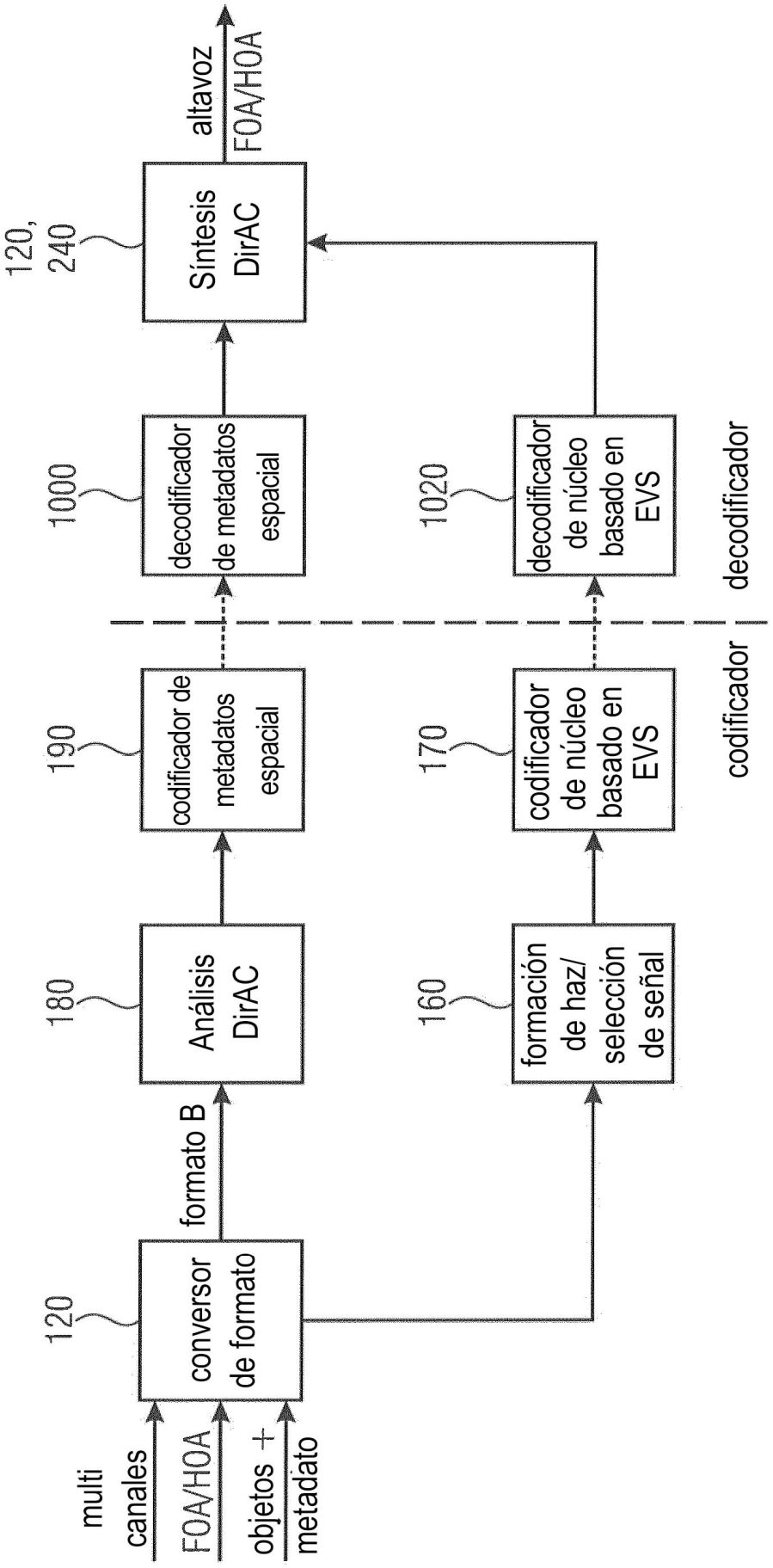

multi canales
FOA/HOA
objetos + metadato
120
conversor de formato
formato B
180
Análisis DirAC
190
codificador de metadatos espacial
1000
decodificador de metadatos espacial
120, 240
Síntesis DirAC
altavoz FOA/HOA
160
formación de haz/ selección de señal
170
codificador de núcleo basado en EVS
1020
decodificador de núcleo basado en EVS
codificador
decodificador

Fig. 11

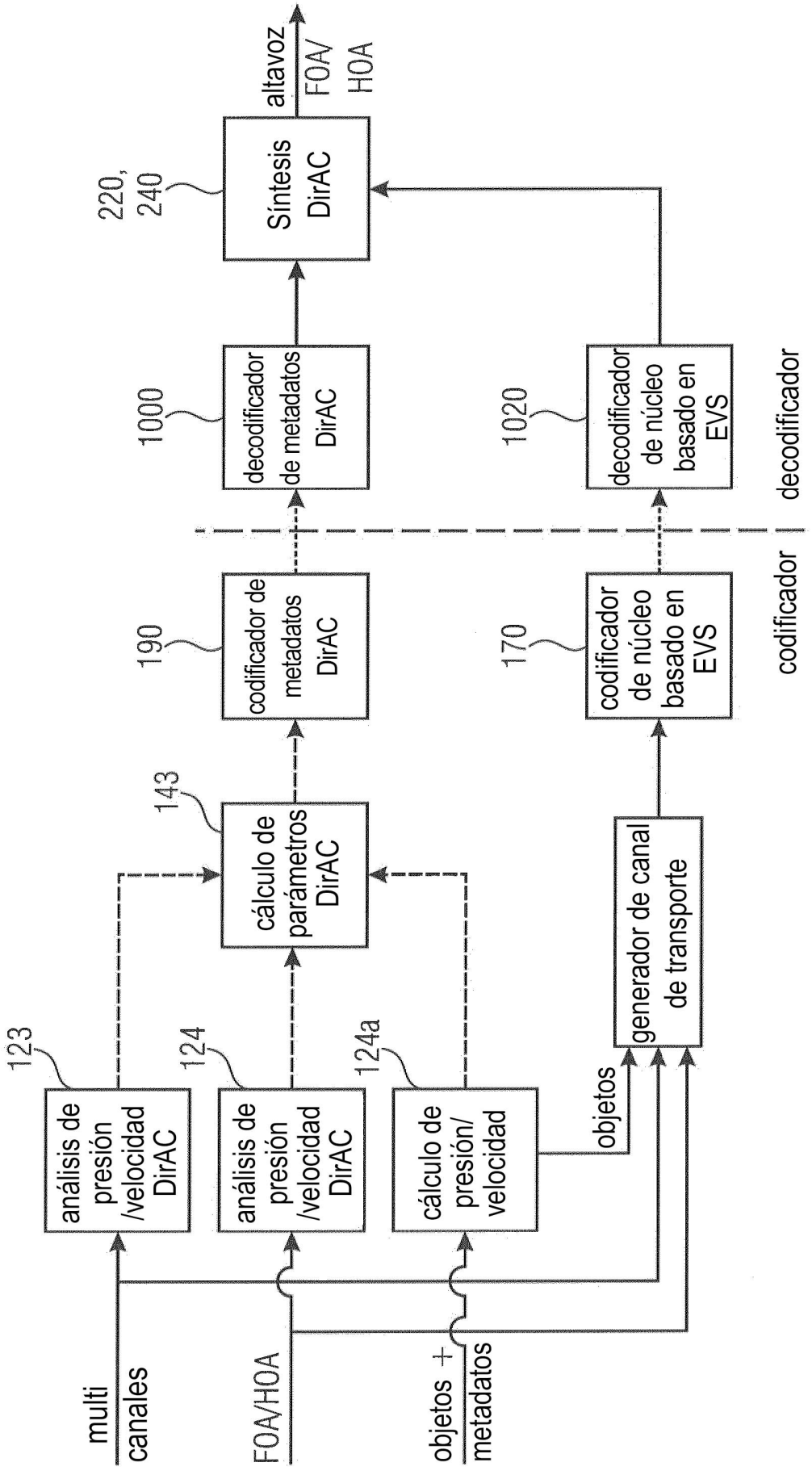

multi canales
F0A/H0A
objetos + metadatos
123
124
124a
143
190
170
1000
1020
220, 240
análisis de presión /velocidad DirAC
análisis de presión /velocidad DirAC
cálculo de presión/ velocidad
cálculo de parámetros DirAC
codificador de metadatos DirAC
codificador de núcleo basado en EVS
decodificador de metadatos DirAC
decodificador de núcleo basado en EVS
Síntesis DirAC
objetos
generador de canal de transporte
altavoz F0A/ H0A
codificador
decodificador

Fig. 12

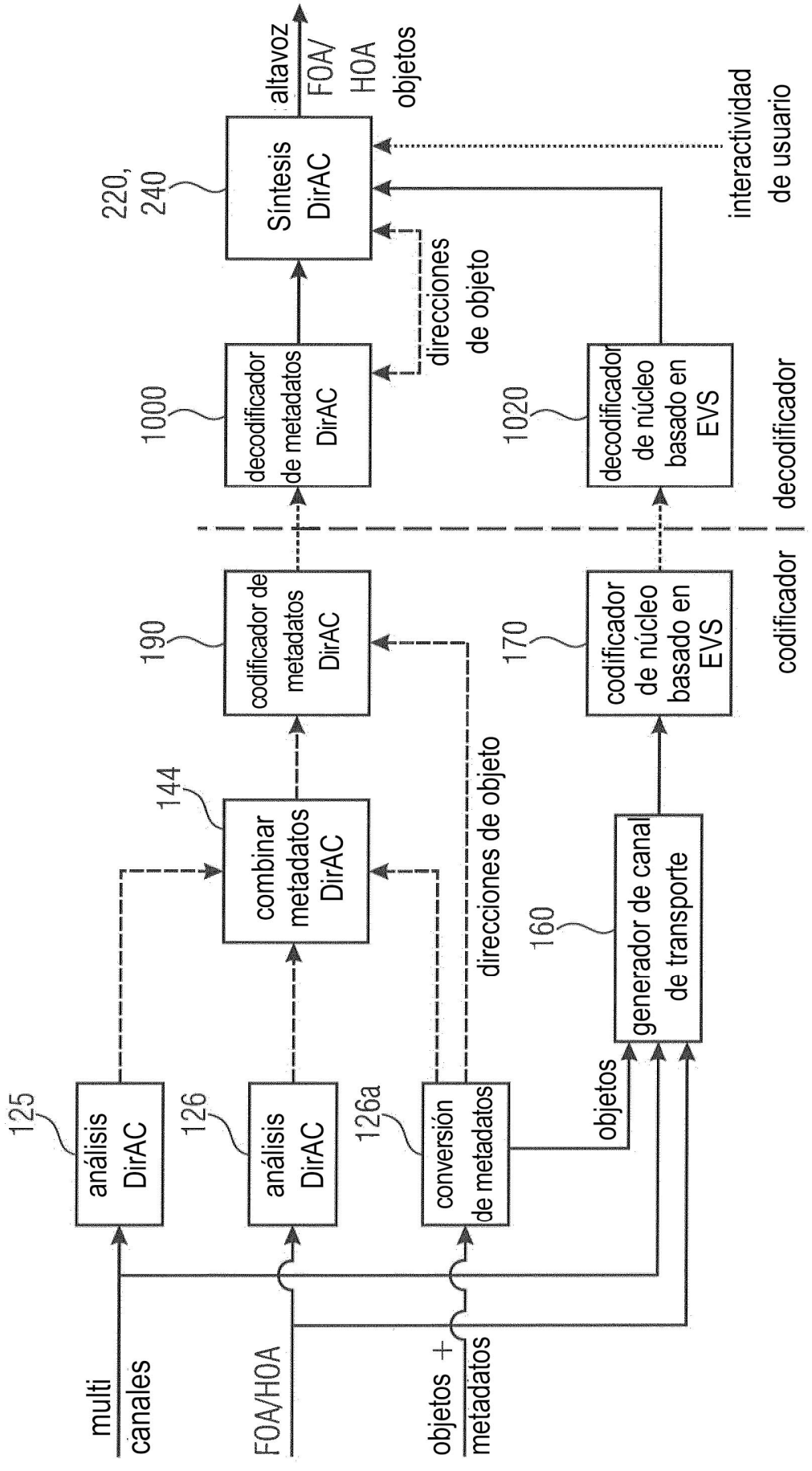

multi canales
F0A/H0A
objetos + metadatos
125
análisis DirAC
126
análisis DirAC
126a
conversión de metadatos
144
combinar metadatos DirAC
190
codificador de metadatos DirAC
160
generador de canal de transporte
170
codificador de núcleo basado en EVS
codificador
direcciones de objeto
objetos
1000
decodificador de metadatos DirAC
1020
decodificador de núcleo basado en EVS
decodificador
direcciones de objeto
220, 240
Síntesis DirAC
altavoz F0A/ H0A objetos
interactividad de usuario

Fig. 13

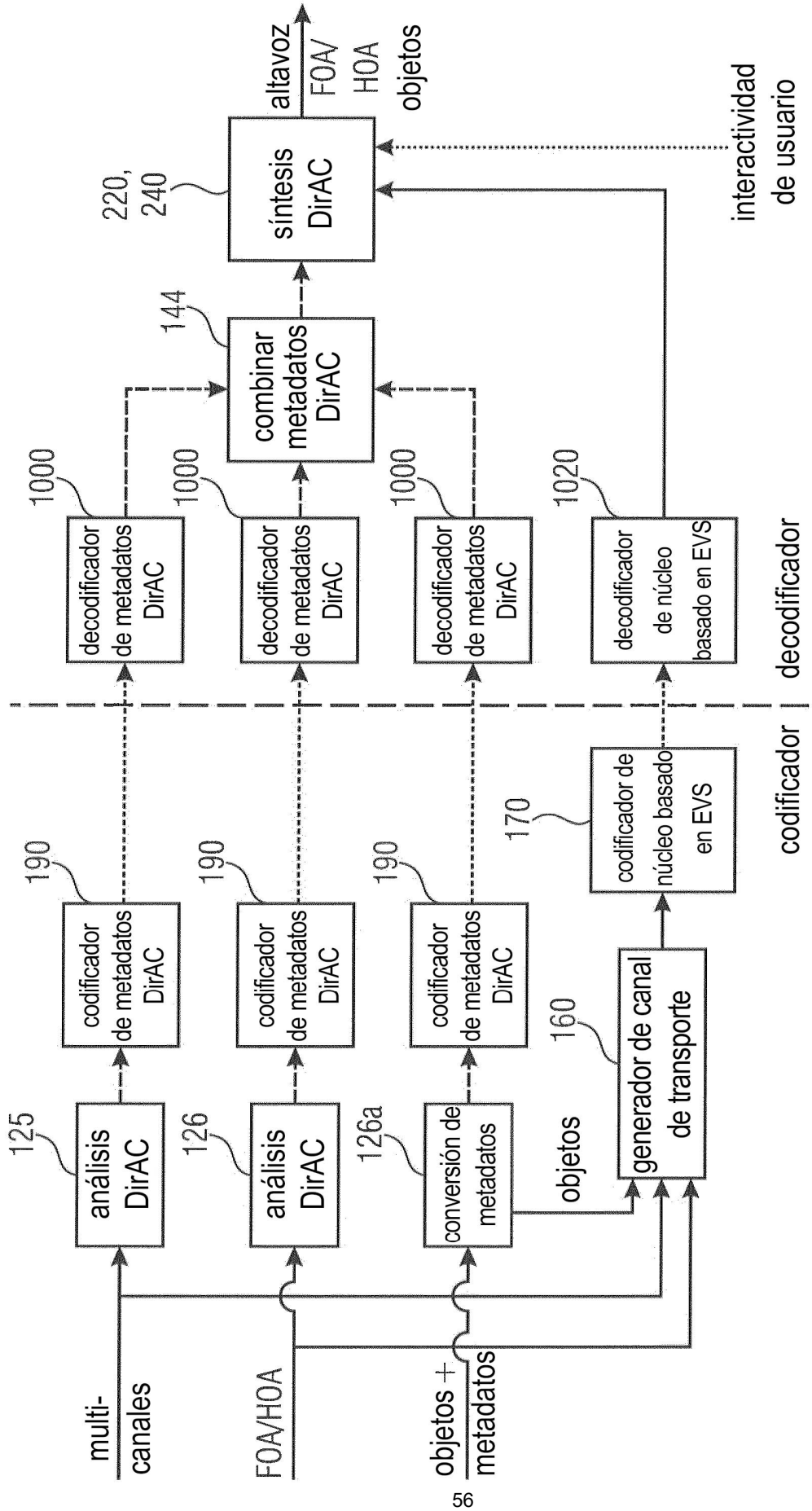

multi-canales
FOA/HOA
objetos + metadatos
125
análisis DirAC
126
análisis DirAC
126a
conversión de metadatos
objetos
160
generador de canal de transporte
190
codificador de metadatos DirAC
190
codificador de metadatos DirAC
190
codificador de metadatos DirAC
170
codificador de núcleo basado en EVS
codificador
1000
decodificador de metadatos DirAC
1000
decodificador de metadatos DirAC
1000
decodificador de metadatos DirAC
1020
decodificador de núcleo basado en EVS
decodificador
144
combinar metadatos DirAC
220, 240
síntesis DirAC
altavoz
FOA/ HOA
objetos
interactividad de usuario

Fig. 14

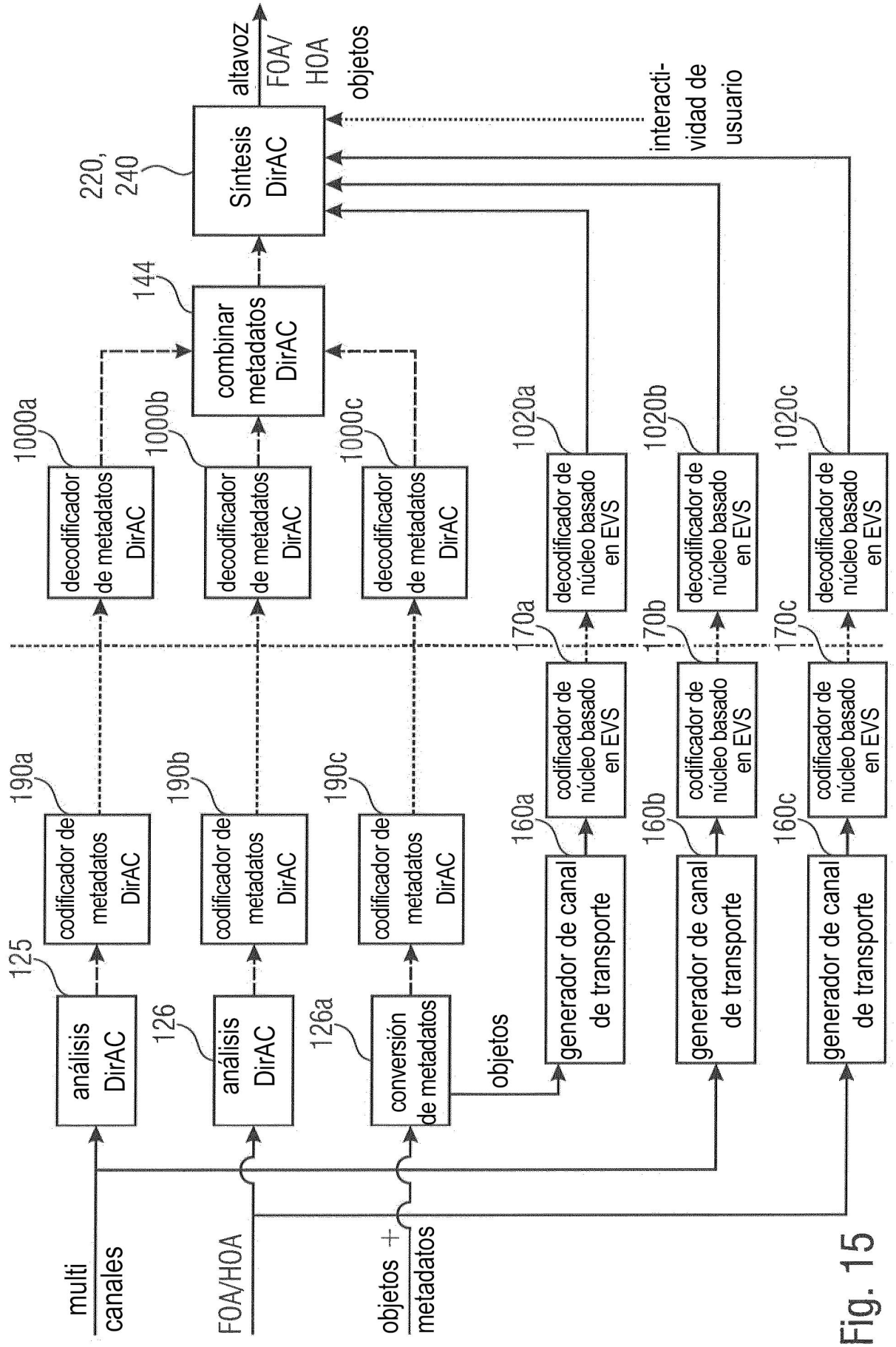

multi canales
FOA/HOA
objetos + metadatos
análisis DirAC
125
análisis DirAC
126
conversión de metadatos
126a
objetos
codificador de metadatos DirAC
190a
codificador de metadatos DirAC
190b
codificador de metadatos DirAC
190c
generador de canal de transporte
160a
generador de canal de transporte
160b
generador de canal de transporte
160c
codificador de núcleo basado en EVS
170a
codificador de núcleo basado en EVS
170b
codificador de núcleo basado en EVS
170c
decodificador de metadatos DirAC
1000a
decodificador de metadatos DirAC
1000b
decodificador de metadatos DirAC
1000c
decodificador de núcleo basado en EVS
1020a
decodificador de núcleo basado en EVS
1020b
decodificador de núcleo basado en EVS
1020c
combinar metadatos DirAC
144
Síntesis DirAC
220, 240
altavoz FOA/ HOA objetos
interacti-vidad de usuario

Fig. 15

Fig. 16a

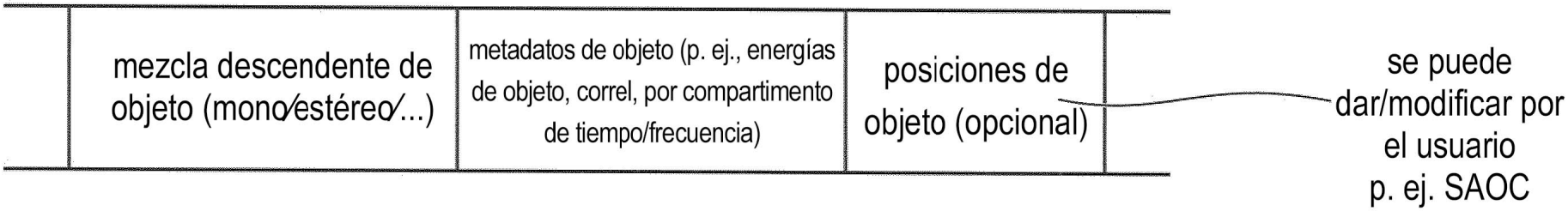


Fig. 16b

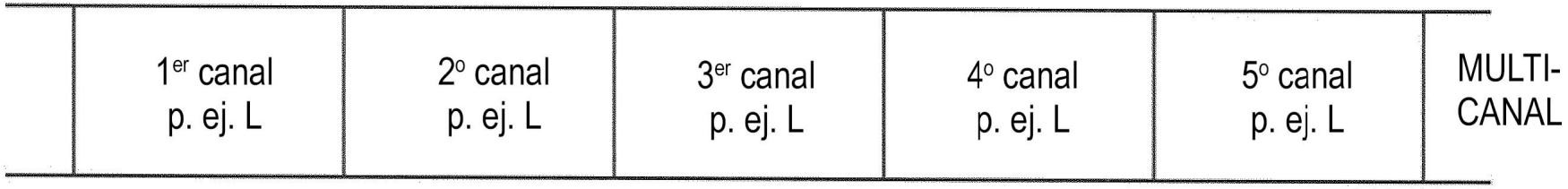
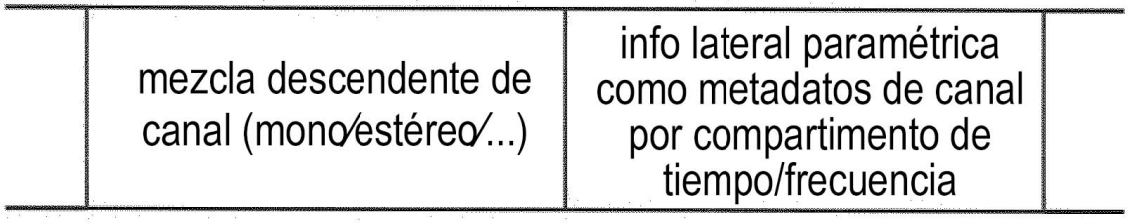


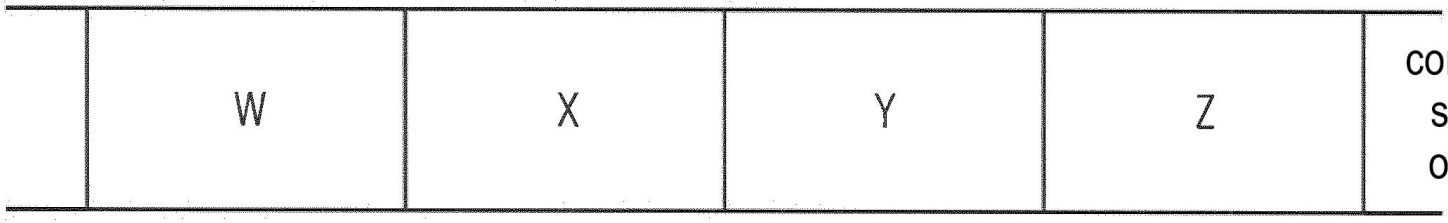
Fig. 16c

ES 2 907 377 T3



p. ej., ENVOLVENTE MPEG

Fig. 16d



FoA
HoA

Fig. 16e

DirAC

Fig. 16f

ES 2 907 377 T3