(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 773 293**

(51) Int. CI.:

H04S 7/00 (2006.01)
H04S 3/00 (2006.01)
G10L 19/08 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **04.03.2015 PCT/EP2015/054514**

(87) Fecha y número de publicación internacional: **01.10.2015 WO15144409**

(96) Fecha de presentación y número de la solicitud europea: **04.03.2015 E 15709657 (9)**

(97) Fecha y número de publicación de la concesión europea: **25.12.2019 EP 3123747**

(54) Título: **Aparato y método para la renderización de audio empleando una definición de distancia geométrica**

(30) Prioridad:

26.03.2014 EP 14161823
08.12.2014 EP 14196765

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
10.07.2020

(73) Titular/es:

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

(72) Inventor/es:

PLOGSTIES, JAN;
FÜG, SIMONE;
NEUENDORF, MAX;
HERRE, JÜRGEN y
GRILL, BERNHARD

(74) Agente/Representante:

ARIZTI ACHA, Monica

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

DESCRIPCIÓN

Aparato y método para la renderización de audio empleando una definición de distancia geométrica

La presente invención se refiere al procesamiento de señales de audio, en particular, a un aparato y un método para la renderización de audio y, más particularmente, a un aparato y un método para la renderización de audio empleando una definición de distancia geométrica.

Con el creciente consumo de contenido multimedia en la vida diaria, aumenta constantemente la demanda de soluciones multimedia sofisticadas. En este contexto, el posicionamiento de los objetos de audio desempeña un papel importante. Sería deseable un posicionamiento óptimo de los objetos de audio para una configuración de altavoces existentes.

En la técnica anterior, se conocen los objetos de audio. Los objetos de audio pueden considerarse, por ejemplo, como pistas de sonido con metadatos asociados. Los metadatos pueden describir, por ejemplo, las características de los datos de audio sin procesar, por ejemplo, la posición de reproducción deseado o el nivel de volumen. Una ventaja del audio basado en objetos es que puede reproducirse un movimiento predefinido mediante un procedimiento de renderización especial del lado de la reproducción de la mejor manera posible para todas las disposiciones de altavoces para la reproducción.

Pueden usarse metadatos geométricos para definir dónde debe renderizarse un objeto de audio, por ejemplo, los ángulos en acimut o la elevación o las posiciones absolutas con respecto a un punto de referencia, por ejemplo, el oyente. Los metadatos se almacenan o se transmiten junto con las señales de audio de objetos.

En el contexto de MPEG-H, en la 105ª reunión de MPEG el grupo de audio revisó los requisitos y los plazos de diferentes normas de aplicación (MPEG = Grupo de Expertos en Imágenes en Movimiento). Según esa revisión, sería esencial cumplir con ciertos puntos de tiempo y requisitos específicos para un sistema de transmisión de próxima generación. Según eso, un sistema debería poder aceptar objetos de audio en la entrada del codificador. Además, el sistema debería admitir la señalización, envío y renderización de objetos de audio y debería habilitar el control del usuario de los objetos, por ejemplo, para la mejora de diálogos, pistas de lenguajes alternativos y lenguaje de descripción de audio.

En la técnica anterior, se conocen diferentes conceptos. Un primer concepto es la renderización de sonido reflejado para el audio basado en objetos (véase [2]). Se incluye información de ubicación del ajuste a los altavoces en una definición de metadatos como información de renderización útil. Sin embargo, en [2], no se proporciona información sobre cómo se usa la información en el procedimiento de reproducción. Además, no se proporciona información sobre cómo se determina una distancia entre dos posiciones.

Se describe otro concepto de la técnica anterior, el sistema y las herramientas para la creación y renderización de audio 3D mejorada en [5]. La figura 6B del documento [5] es un diagrama que ilustra cómo puede realizarse algorítmicamente un “ajuste” a un altavoz. En detalle, según el documento [5] si se determina ajustar la posición del objeto de audio a la ubicación de un altavoz (véase el bloque 665 de la figura 6B del documento [5]), se mapeará la posición del objeto de audio a una ubicación del altavoz (véase el bloque 670 de la figura 6B del documento [5]), generalmente el que esté más cerca de la posición pretendida (x,y,z) recibida para el objeto de audio. Según [5], el “ajuste” puede aplicarse a un pequeño grupo de altavoces de reproducción y/o a un altavoz de reproducción individual. Sin embargo, [5] emplea coordenadas cartesianas (x,y,z) en lugar de coordenadas esféricas. Además, el comportamiento del renderizador se describe solo como la posición del objeto de audio mapeada a la ubicación del altavoz; si la bandera de ajuste es una, no se proporciona una descripción detallada. Además, no se proporcionan detalles de cómo se determina el altavoz más cercano.

Según otra técnica anterior, el Sistema y Método para la Generación de Señales de Audio Adaptativas, Codificación y Renderización descritos en el documento [1], la información de metadatos (elementos de metadatos) especifica que “se renderizan uno o más componentes de sonido a la alimentación de un altavoz para la reproducción a través de un altavoz más cercano a una ubicación de reproducción pretendida del componente de sonido, tal como indican los metadatos de posición”. Sin embargo, no se proporciona información sobre cómo se determina el altavoz más cercano.

En una técnica anterior adicional, un modelo de definición de audio, descrito en el documento [4], se define una bandera de metadatos denominada “channelLock”. Si se ajusta a 1, un renderizador puede bloquear el objeto al canal o altavoz más cercano, en lugar de la renderización normal. Sin embargo, no se describe la determinación del canal más cercano.

En otra técnica anterior, se describe la mezcla ascendente de audio basado en objetos (véase [3]). El documento [3] describe un método para el uso de una medida de distancia de los altavoces en un campo de aplicación diferente. Aquí se usa para la mezcla ascendente de material de audio basado en objetos. El sistema de renderización está

configurado para determinar, de un programa de audio basado en objetos (y el conocimiento de las posiciones de los altavoces que se han de emplear para ejecutar el programa), la distancia entre cada posición de una fuente de audio indicada por el programa y la posición de cada uno de los altavoces. Además, el sistema de renderización de [3] está configurado para determinar, para cada posición de fuente real (por ejemplo, cada posición de fuente a lo largo de una trayectoria de fuente) indicada por el programa, una subserie completa de altavoces (una subserie "primaria") que consiste en esos altavoces completa (o el altavoz completa) que están más cerca de la posición de la fuente real, en el que "más cerca" se define, en este contexto, en algún sentido definido de manera razonable. Sin embargo, no se proporciona información sobre cómo debe calcularse la distancia.

El objetivo de la presente invención es proporcionar conceptos mejorados para la renderización de audio. El objetivo de la presente invención se resuelve mediante un aparato según la reivindicación 1, mediante un dispositivo decodificador según la reivindicación 4, mediante un método según la reivindicación 5 y mediante un programa informático según la reivindicación 6.

Se proporciona un aparato para reproducir un objeto de audio asociado a una posición. El aparato comprende un calculador de distancia para calcular distancias de la posición a los altavoces o para leer las distancias desde la posición hasta los altavoces. El calculador de distancia está configurado para adoptar una solución con la menor distancia. El aparato está configurado para reproducir el objeto de audio usando el altavoz que corresponde a la solución.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces o para leer las distancias desde la posición hasta los altavoces solo si está habilitada una bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout), recibida por el aparato. Además, el calculador de distancia puede estar configurado, por ejemplo, para adoptar una solución con una menor distancia solo si está habilitada la bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout). Además, el aparato puede estar configurado, por ejemplo, para reproducir el objeto de audio usando el altavoz que corresponde a la solución solo si está habilitada la bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout).

En una realización, el aparato puede estar configurado, por ejemplo, para no llevar a cabo ninguna renderización del objeto de audio, si está habilitada la bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout).

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve una distancia euclidiana ponderada o una distancia de gran arco.

En una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve diferencias absolutas ponderadas en los ángulos acimutal y de elevación.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve diferencias absolutas ponderadas a la potencia p, en la que p es un número. En una realización, p puede ajustarse, por ejemplo, a p = 2.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve una diferencia angular ponderada.

En una realización, una función de distancia puede definirse, por ejemplo, según

$$\text{diffAngle} = \text{acos}(\cos(\text{azDiff}) * \cos(\text{elDiff})),$$

en el que azDiff indica una diferencia de dos ángulos acimutales, en el que elDiff indica una diferencia de dos ángulos de elevación, y en el que diffAngle indica la diferencia angular ponderada.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición a uno de los altavoces se calcula según

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, y β_2 indica un ángulo de elevación de dicho uno de los altavoces. O α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de

elevación de dicho uno de los altavoces, y β_2 indica un ángulo de elevación de la posición.

En una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1,P_2)=|\beta_1-\beta_2|+|\alpha_1-\alpha_2|+|r_1-r_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, r_1 indica un radio de la posición y r_2 indica un radio de dicho uno de los altavoces. O α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces, β_2 indica un ángulo de elevación de la posición, r_1 indica un radio de dicho uno de los altavoces y r_2 indica un radio de la posición.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1,P_2)=\ b\cdot|\beta_1-\ \beta_2|+a\cdot|\alpha_1-\ \alpha_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, a es un primer número y b es un segundo número. O α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces, β_2 indica un ángulo de elevación de la posición, a es un primer número y b es un segundo número.

En una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1,P_2)=\ b\cdot|\beta_1-\ \beta_2|+a\cdot|\alpha_1-\ \alpha_2|+\ c\cdot|r_1-r_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, r_1 indica un radio de la posición, r_2 indica un radio de dicho uno de los altavoces, a es un primer número y b es un segundo número. O, α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces y β_2 indica un ángulo de elevación de la posición, r_1 indica un radio de dicho uno de los altavoces y r_2 indica un radio de la posición, a es un primer número, b es un segundo número y c es un tercer número.

Según una realización, se proporciona un dispositivo decodificador. El dispositivo decodificador comprende un codificador USAC para decodificar un flujo de bits para obtener uno o más canales de entrada de audio, para obtener uno o más objetos de audio de entrada, para obtener metadatos de objeto comprimidos y para obtener uno o más canales de transporte para SAOC. Además, el dispositivo decodificador comprende un decodificador SAOC para decodificar el uno o más canales de transporte para SAOC para obtener un grupo de uno o más objetos de audio renderizados. Además, el dispositivo decodificador comprende un decodificador de metadatos de objeto para decodificar los metadatos de objeto comprimidos para obtener metadatos descomprimidos. Además, el dispositivo decodificador comprende un conversor de formato para convertir el uno o más canales de entrada de audio para obtener uno o más canales convertidos. Además, el dispositivo decodificador comprende un mezclador para mezclar el uno o más objetos de audio renderizados del grupo de uno o más objetos de audio renderizados, el uno o más objetos de audio de entrada y el uno o más canales convertidos para obtener uno o más canales de audio decodificados. El decodificador de metadatos de objeto y el mezclador juntos forman un aparato según una de las realizaciones descritas anteriormente. El decodificador de metadatos de objeto comprende el calculador de distancia del aparato según una de las realizaciones descritas anteriormente, en la que el calculador de distancia está configurado, para cada objeto de audio de entrada del uno o más objetos de audio de entrada, para calcular distancias de la posición asociada a dicho objeto de audio de entrada hasta los altavoces o para leer las distancias de la posición asociada a dicho objeto de audio de entrada hasta los altavoces y para adoptar una solución con la menor distancia. El mezclador está configurado para emitir cada objeto de audio de entrada del uno o más objetos de audio de entrada dentro de uno del uno o más canales de audio decodificados al altavoz que corresponde a la solución determinada por el calculador de distancia del aparato según una de las realizaciones descritas anteriormente para dicho objeto de audio de entrada.

A continuación, se describen realizaciones de la presente invención en más detalle con referencia a las figuras, en las que:

La figura 1 es un aparato según una realización,

La figura 2 ilustra un renderizador de objetos según una realización,

La figura 3 ilustra un procesador de metadatos de objeto según una realización,

La figura 4 ilustra una visión general de un codificador de audio 3D,

La figura 5 ilustra una visión general de un decodificador de audio 3D según una realización, y

La figura 6 ilustra una estructura de un conversor de formato.

La figura 1 ilustra un aparato 100 para reproducir un objeto de audio asociado a una posición proporcionada.

El aparato 100 comprende un calculador 110 de distancia para calcular las distancias de la posición hasta los altavoces o para leer las distancias desde la posición hasta los altavoces. El calculador 110 de distancia está configurado para adoptar una solución con la distancia más corta.

El aparato 100 está configurado para reproducir el objeto de audio usando el altavoz que corresponde a la solución.

Por ejemplo, para cada altavoz, se determina una distancia entre la posición (la posición del objeto de audio) y dicho altavoz (la ubicación de dicho altavoz).

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces o para leer las distancias desde la posición hasta los altavoces solo si está habilitada una bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout), recibida por el aparato 100. Además, el calculador de distancia puede estar configurado, por ejemplo, para adoptar una solución con la menor distancia solo si está habilitada la bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout). Además, el aparato 100 puede estar configurado, por ejemplo, para reproducir el objeto de audio usando el altavoz que corresponde a la solución solo si está habilitada la bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout).

En una realización, el aparato 100 puede estar configurado, por ejemplo, para no llevar a cabo ninguna renderización del objeto de audio, si está habilitada la bandera de reproducción del altavoz más cercano (mdae_closestSpeakerPlayout).

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve una distancia euclidiana ponderada o una distancia de gran arco.

En una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve diferencias absolutas ponderadas en los ángulos acimutales y de elevación.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve diferencias absolutas ponderadas a la potencia p, en la que p es un número. En una realización, puede ajustarse p, por ejemplo, a p = 2.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias dependiendo de una función de distancia que devuelve una diferencia angular ponderada.

En una realización, puede definirse una función de distancia, por ejemplo, según

$$\text{diffAngle} = \text{acos}(\cos(\text{azDiff}) * \cos(\text{elDiff})),$$

en el que azDiff indica una diferencia de dos ángulos acimutales, en el que elDiff indica una diferencia de dos ángulos de elevación y en el que diffAngle indica la diferencia angular ponderada.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición hasta uno de los altavoces

se calcula según

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición y β_2 indica un ángulo de elevación de dicho uno de los altavoces. O, α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces y β_2 indica un ángulo de elevación de la posición.

En una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, r_1 indica un radio de la posición y r_2 indica un radio de dicho uno de los altavoces. O α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces, β_2 indica un ángulo de elevación de la posición, r_1 indica un radio de dicho uno de los altavoces y r_2 indica un radio de la posición.

Según una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, a es un primer número y b es un segundo número. O α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces, β_2 indica un ángulo de elevación de la posición, a es un primer número y b es un segundo número.

En una realización, el calculador de distancia puede estar configurado, por ejemplo, para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1,P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|$$

α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, r_1 indica un radio de la posición, r_2 indica un radio de dicho uno de los altavoces, a es un primer número, b es un segundo número y c es un tercer número. O, α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces y β_2 indica un ángulo de elevación de la posición, r_1 indica un radio de dicho uno de los altavoces y r_2 indica un radio de la posición, a es un primer número, b es un segundo número y c es un tercer número.

A continuación, se describen realizaciones de la presente invención. Las realizaciones proporcionan conceptos para el uso de una definición de distancia geométrica para la renderización de audio.

Pueden usarse metadatos de objeto para definir una de las siguientes:

1) en qué lugar del espacio debe renderizarse un objeto, o

2) qué altavoz debe usarse para reproducir el objeto.

Si la posición del objeto indicada en los metadatos no recae en un único altavoz, el renderizador de objetos crearía la señal de salida basándose en el uso de múltiples altavoces y reglas de panoramización definidas. La panoramización es subóptima en cuanto a la localización de los sonidos o del color del sonido.

Por lo tanto, puede ser deseable para el productor de contenido basado en objetos definir que un determinado sonido debe provenir de un único altavoz desde una determinada dirección.

Puede suceder que este altavoz no exista en la configuración de altavoces del usuario. En ese caso, se establece una bandera en los metadatos que hace que el sonido sea reproducido por el altavoz más cercano disponible sin renderización.

La invención describe cómo puede encontrarse el altavoz más cercano permitiendo que cierta ponderación explique una desviación tolerable dese la posición deseada del objeto.

La figura 2 ilustra un renderizador de objetos según una realización.

En los formatos de audio basados en objetos se almacenan o se transmiten metadatos junto con las señales de objeto. Los objetos de audio se renderizan del lado de la reproducción usando los metadatos e información acerca del entorno de reproducción. Dicha información es, por ejemplo, el número de altavoces o el tamaño de la pantalla.

Tabla 1 – Ejemplo de metadatos:

	ObjectID
OAM Dinámicos	Acimut Elevación Ganancia Distancia
Interactividad	AllowOnOff AllowPositionInteractivity AllowGainInteractivity DefaultOnOff DefaultGain InteractivityMinGain InteractivtiyMaxGain InteractivityMinAzOffset InteractivityMaxAzOffset InteractivityMinElOffset InteractivityMaxElOffset InteractivityMinDist InteractivityMaxDist ObjectID
Reproducción	IsSpeakerRelatedGroup SpeakerConfig3D AzimuthScreenRelated ElevationScreenRelated ClosestSpeakerPlayout
Contenido	ContentKind ContentLanguage
Grupo	GroupID GroupDescription GroupNumMembers GroupMembers Priority
Cambio de grupo	SwitchGroupID SwitchGroupDescription SwitchGroupDefault SwitchGroupNumMembers SwitchGroupMembers
Escena de audio	NumGroupsTotal IsMainScene NumGroupsPresent NumSwitchGroups

Para los objetos pueden usarse metadatos geométricos para definir cómo deben renderizarse, por ejemplo, los ángulos en acimut o la elevación o las posiciones absolutas con respecto a un punto de referencia, por ejemplo, el oyente. El renderizador calcula las señales de altavoces a partir de los datos geométricos y los altavoces disponibles

y su posición.

Si un objeto de audio (señal de audio con una posición en el espacio 3D, por ejemplo, el dado de acimut, elevación y distancia) no debe renderizarse en su posición asociada, sino que en su lugar debe reproducirse mediante un altavoz que existe en la configuración de altavoces local, una manera sería definir el altavoz en el que el objeto debe reproducirse mediante metadatos.

Sin embargo, hay casos en los que el productor no quiere que el contenido de objetos se reproduzca mediante un altavoz específico, sino más bien mediante el siguiente altavoz disponible, es decir, el altavoz “geométricamente más cercano”. Esto permite una reproducción discreta sin necesidad de definir qué altavoz corresponde a qué señal de audio o hacer la renderización entre múltiples altavoces.

Las realizaciones según la presente invención surgen de lo anterior de la siguiente manera.

Campos de metadatos:

ClosestSpeakerPlayout	el objeto debe reproducirse mediante el altavoz geométricamente más cercano, sin renderización (solo para objetos dinámicos (IsSpeakerRelatedGroup == 0))

Tabla 2 — Sintaxis de GroupDefinition():

Sintaxis	N.º de bits	Mnemotecnia

mdae_GroupDefinition(numGroups)
{
para (grp = 0; grp < numGroups; grp++) {
 	
mdae_groupID[grp];
	7	**uimsbf**

… **mdae_groupPriority**[grp]; **mdae_closestSpeakerPlayout**[grp];	**3** **1**	**uimsbf** **bslbf**
… } }		

mdae_closestSpeakerPlayout Esta bandera define que los elementos del grupo de elementos de metadatos no debe renderizarse sino reproducirse directamente por los altavoces que estén más cerca de la posición geométrica de los elementos.

El remapeo se hace en un procesador de metadatos de objeto que tiene en cuenta la configuración de altavoces local y realiza un enrutamiento de las señales hasta los correspondientes renderizadores con información específica de por qué altavoz o desde qué dirección debe renderizarse un sonido.

La figura 3 ilustra un procesador de metadatos de objeto según una realización.

A continuación, se describe una estrategia para el cálculo de la distancia:

- si está ajustada la bandera de metadatos de altavoz más cercana, el sonido se reproduce en el altavoz más cercano
- para ello, se calcula la distancia hasta los siguientes altavoces (o se lee de una tabla previamente almacenada)
- se adopta la solución con la menor distancia
- la función de distancia puede ser, por ejemplo (pero sin limitación a):
    - la distancia euclidiana o de gran arco ponderada
    - las diferencias absolutas ponderadas en el ángulo acimutal y de elevación

- las diferencias absolutas ponderadas a la potencia p (p=2 => solución de cuadrados mínimos)
- la diferencia angular ponderada, por ejemplo, diffAngle = acos(cos(azDiff)*cos(elDiff))

A continuación, se exponen ejemplos del cálculo del altavoz más cercano.

Si está habilitada la bandera *mdae_closestSpeakerPlayout* de un grupo de elementos de audio, cada uno de los elementos de audio del grupo se reproducirán por el altavoz que está más cerca de la posición dada del elemento de audio. No se aplica renderización.

La distancia de dos posiciones P_1 y P_2 en un sistema de coordenadas esférico se define como la diferencia absoluta de sus ángulos acimutales α y sus ángulos de elevación β .

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

Esta distancia debe calcularse para todas las posiciones conocidas P_1 a P_N de los N altavoces de salida con respecto a la posición deseada del elemento de audio $P_{deseada}$.

La posición del altavoz más cercano conocido es esa en la que la distancia hasta la posición deseada del elemento de audio llega a ser mínima

$$P_{siguiente} = \min(\Delta(P_{deseada}, P_1), \Delta(P_{deseada}, P_2), .., \Delta(P_{deseada}, P_N))$$

Con esta fórmula, es posible añadir ponderaciones a la elevación, el acimut y/o el radio. De esa manera es posible declarar que una desviación del acimut debe ser menos tolerable que una desviación de la elevación mediante la ponderación de la desviación del acimut en un número alto:

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|$$

Un ejemplo se refiere a un cálculo del altavoz más cercano para la renderización binaural.

Si debe reproducirse el contenido de audio como señal estéreo binaural por los auriculares o una configuración de altavoces estéreo, tradicionalmente se combina matemáticamente cada canal del contenido de audio con una respuesta al impulso binaural de sala o la respuesta al impulso relacionada con la cabeza.

La posición para la medición de esta respuesta al impulso tiene que corresponder a la dirección desde la que debe percibirse el contenido de audio del canal asociado. En los sistemas de audio multicanal o de audio basado en objetos existe el caso en el que el número de posiciones definibles (ya sea por un altavoz o por una posición del objeto) es mayor que el número de respuestas al impulso disponible. En ese caso, también debe elegirse una respuesta al impulso apropiada si no hay una especial disponible para la posición del canal o la posición del objeto. Para impartir solo cambios de posición mínimos en la percepción, la respuesta al impulso elegida debe ser la respuesta al impulso “geométricamente más cercana”.

En ambos casos es necesario determinar cuál de la lista de posiciones conocidas (es decir, altavoces de reproducción o BRIR) es la siguiente a la posición deseada (BRIR = Respuesta al Impulso Binaural de Sala). Por lo tanto, tiene que definirse una “distancia” entre diferentes posiciones.

La distancia entre diferentes posiciones se define aquí como la diferencia absoluta de sus ángulos acimutales y de elevación.

Se usa la siguiente fórmula para calcular una distancia de dos posiciones P_1 , P_2 en un sistema de coordenadas que se define por la elevación α y el acimut β :

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2|$$

Es posible añadir el radio *r* como una tercera variable:

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|$$

La posición más cercana conocida es esa en la que la distancia hasta la posición deseada llega a ser mínima

$$P_{siguiente} = \min(\Delta(P_{deseada}, P_1), \Delta(P_{deseada}, P_2), .., \Delta(P_{deseada}, P_N))$$

En una realización, pueden añadirse ponderaciones, por ejemplo, a la elevación, el acimut y/o el radio:

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|$$

Según algunas realizaciones, puede determinarse el altavoz más cercano, por ejemplo, de la siguiente manera:

La distancia entre dos posiciones P_1 y P_2 en un sistema de coordenadas esférico puede definirse, por ejemplo, como la diferencia absoluta de sus ángulos acimutales φ y sus ángulos de elevación θ.

$$\Delta(P_1, P_2) = |\theta_1 - \theta_2| + |\varphi_1 - \varphi_2|$$

Tiene que calcularse esta distancia para todas las posiciones conocidas P_1 a P_N de los N canales de salida con respecto a la posición deseada del elemento de audio $P_{deseada}$.

La posición del altavoz más cercana conocida es esa en la que la distancia hasta la posición deseada del elemento de audio llega a ser mínima:

$$P_{siguiente} = \min(\Delta(P_{deseada}, P_1), \Delta(P_{deseada}, P_2), .., \Delta(P_{deseada}, P_N))$$

Por ejemplo, según algunas realizaciones, el procesamiento de reproducción del altavoz más cercano según algunas realizaciones puede llevarse a cabo determinando la posición del altavoz existente más cercano para cada elemento del grupo de objetos de audio, si la bandera ClosestSpeakerPlayout es igual a uno.

El procesamiento de reproducción del altavoz más cercano puede ser, por ejemplo, especialmente importante para grupos de elementos con datos de posición dinámicos. La posición del altavoz conocido más cercano puede ser esa, por ejemplo, en la que la distancia hasta la posición buscada/deseada del elemento de audio llega a ser mínima.

A continuación, se proporciona una visión general de sistema de un sistema de códec de audio 3D. Las realizaciones de la presente invención pueden emplearse en dicho sistema de códec de audio 3D. El sistema de códec de audio 3D puede basarse, por ejemplo, en un Códec MPEG-D USAC para la codificación de señales de canales y objetos.

Según las realizaciones, para aumentar la eficiencia en la codificación de una gran cantidad de objetos, se ha adaptado la tecnología MPEG SAOC (SAOC = Codificación de Objetos de Audio Espacial). Por ejemplo, según algunas realizaciones, tres tipos de renderizadores pueden ejecutar, por ejemplo, tareas de renderización de objetos a canales, renderización de canales a auriculares o renderización de canales a una configuración de altavoces diferente.

Cuando las señales de objeto se transmiten explícitamente o se codifican paramétricamente usando SAOC, se comprime y se multiplexa la correspondiente información de metadatos de objeto en el flujo de bits de audio 3D.

La figura 4 y la figura 5 ilustran los diferentes bloques algorítmicos del sistema de audio 3D. En particular, la figura 4 ilustra una visión general de un codificador de audio 3D. La figura 5 ilustra una visión general de un decodificador de audio 3D según una realización.

Se describen ahora posibles realizaciones de los módulos de la figura 4 y la figura 5.

En la figura 4, se ilustra un prerenderizador 810 (también denominado mezclador). En la configuración de la figura 4, el prerenderizador 810 (mezclador) es opcional. El prerenderizador 810 también puede usarse de manera opcional para convertir una escena de entrada de Canal + Objeto a una escena de canal antes de la codificación. Funcionalmente, el prerenderizador 810 del lado del codificador puede estar relacionado, por ejemplo, con la funcionalidad del renderizador/mezclador 920 de objetos del lado del decodificador, que se describe a continuación. La prerenderización de objetos garantiza una entropía de la señal determinística a la entrada del codificador que es básicamente independiente del número de señales de objeto activas simultáneamente. Con la prerenderización de objetos, no se requiere transmisión de metadatos de objeto. Se renderizan señales de objeto discretas a la disposición

de canales que el codificador está configurado para usar. Las ponderaciones de los objetos para cada canal se obtienen de los metadatos de objeto asociados (OAM).

El códec de núcleos para las señales de canales de los altavoces, las señales de objeto discretas, las señales de mezcla a la baja de objetos y las señales prerenderizadas se basa en la tecnología MPEG-D USAC (Códec de Núcleos de USAC). El codificador 820 USAC (por ejemplo, el ilustrado en la figura 4) se encarga de la codificación de la multitud de señales mediante la creación de información de mapeo de canales y objetos basada en la información geométrica y semántica del canal de entrada y la asignación de objetos. Esta información de mapeo describe cómo se mapean los canales y objetos de entrada a los elementos de canales de USAC (CPE, SCE, LFE) y cómo se transmite la información correspondiente al decodificador.

Todas las cargas útiles adicionales como datos o metadatos de objeto de SAOC se han pasado por elementos de extensión y se las puede considerar, por ejemplo, en el control de la tasa del codificador USAC.

La codificación de objetos es posible de diferentes maneras, dependiendo de los requisitos de tasa/distorsión y los requisitos de interactividad para el renderizador. Son posibles las siguientes variantes de codificación de objetos:

- Objetos prerenderizados: se prerenderizan y se mezclan las señales de objeto con las señales de 22.2 canales antes de la codificación. La cadena de codificación posterior ve señales de 22.2 canales.

- Formas de onda de objetos discretos: se suministran objetos como formas de onda monofónicas al codificador 820 USAC. El codificador 820 USAC usa elementos de canales únicos SCEs para transmitir los objetos además de las señales de canal. Los objetos decodificados se renderizan y se mezclan del lado del receptor. Se transmite información de metadatos de objeto comprimidos al receptor/renderizador de forma paralela.

- Formas de onda paramétricas de objetos: se describen propiedades de los objetos y su relación entre sí mediante parámetros de SAOC. La mezcla a la baja de las señales de objeto se codifica con USAC por el codificador 820 USAC. La información paramétrica se transmite de forma paralela. El número de canales de mezcla a la baja se elige dependiendo del número de objetos y de la tasa de datos general. Se transmite información de metadatos de objeto comprimida al renderizador de SAOC.

Del lado del decodificador, un decodificador 910 USAC lleva a cabo la decodificación de USAC.

Además, según las realizaciones, se proporciona un decodificador, véase la figura 5. El decodificador comprende un decodificador 910 USAC para decodificar un flujo de bits para obtener uno o más canales de entrada de audio, para obtener uno o más objetos de audio, para obtener metadatos de objeto comprimidos y para obtener uno o más canales de transporte de SAOC.

Además, el decodificador comprende un decodificador 915 SAOC para decodificar el uno o más canales de transporte de SAOC para obtener un primer grupo de uno o más objetos de audio renderizados.

Además, el decodificador comprende un convertidor 922 de formato para convertir el uno o más canales de entrada de audio para obtener uno o más canales convertidos.

Además, el decodificador comprende un mezclador 930 para mezclar los objetos de audio del primer grupo de uno o más objetos de audio renderizados, el objeto de audio del segundo grupo de uno o más objetos de audio renderizados y el uno o más canales convertidos para obtener uno o más canales de audio decodificados.

En la figura 5 se ilustra una realización particular de un decodificador. El codificador 815 SAOC (el codificador 815 SAOC es opcional, véase la figura 4) y el decodificador 915 SAOC (véase la figura 5) para las señales de objeto se basan en la tecnología de MPEG SAOC. El sistema es capaz de recrear, modificar y renderizar un número de objetos de audio basándose en un número menor de canales transmitidos y datos paramétricos adicionales (OLD, IOC, DMG) (OLD = diferencia de nivel de los objetos, IOC = correlación entre los objetos, DMG = ganancia de mezcla a la baja). Los datos paramétricos adicionales muestran una tasa de datos significativamente más baja que la necesaria para transmitir todos los objetos individualmente, lo que confiere mucha eficiencia a la codificación.

El codificador 815 SAOC toma como entrada señales de objeto/canales como formas de onda monofónicas y emite la información paramétrica (que está incluida en el flujo de bits de audio 3D) y los canales de transporte de SAOC (que se codifican y se transmiten usando elementos de canales únicos).

El decodificador 915 SAOC reconstruye las señales de objeto/canales de los canales de transporte de SAOC decodificados e información paramétrica, y genera la escena de audio de salida basándose en la disposición de la reproducción, la información de metadatos de objeto descomprimida y opcionalmente en la información de interacción

con el usuario.

En lo que respecta al códec de metadatos de objeto, para cada objeto, los metadatos asociados que especifican la posición geométrica y la expansión del objeto en el espacio 3D se codifican de manera eficiente mediante la cuantificación de las propiedades del objeto en tiempo y espacio, por ejemplo, mediante el codificador 818 de metadatos de la figura 4. Los metadatos de objeto comprimidos cOAM (cOAM = metadatos de objeto de audio comprimidos) se transmiten al receptor como información secundaria. En el receptor, se decodifican los cOAM mediante el decodificador 918 de metadatos.

Por ejemplo, en la figura 5, el decodificador 918 de metadatos puede implementar por ejemplo, el calculador 110 de distancia de la figura 1 según una de las realizaciones descritas anteriormente.

Un renderizador de objetos, por ejemplo, el renderizador 920 de objetos de la figura 5, utiliza los metadatos de objeto comprimidos para generar formas de onda de objetos según el formato de reproducción dado. Se renderiza cada objeto a ciertos canales de salida según sus metadatos. La salida de este bloque es el resultado de la suma de los resultados parciales. En algunas realizaciones, si se lleva a cabo la determinación del altavoz más cercano, el renderizador 920 de objetos puede pasar, por ejemplo, los objetos de audio recibidos del decodificador 910 3D de USAC, sin renderizarlos, al mezclador 930. El mezclador 930 puede pasar, por ejemplo, los objetos de audio al altavoz que determinara el calculador de distancia (por ejemplo, el implementado dentro del decodificador 918 de metadatos) a los altavoces. De esta manera según una realización, el decodificador 918 de metadatos que puede comprender, por ejemplo, un calculador de distancia, el mezclador 930 y, opcionalmente, el renderizador 920 de objetos pueden implementar juntos el aparato 100 de la figura 1.

Por ejemplo, el decodificador 918 de metadatos comprende un calculador de distancia (no se muestra) y dicho calculador de distancia o el decodificador 918 de metadatos pueden señalizar, por ejemplo, mediante una conexión (no se muestra) al mezclador 930, el altavoz más cercano para cada objeto de audio del uno o más objetos de audio recibidos del decodificador 3D de USAC. Entonces, el mezclador 930 puede emitir el objeto de audio dentro de un canal de altavoz solo al altavoz más cercano (determinado por el calculador de distancia) de la pluralidad de altavoces.

En algunas realizaciones adicionales, el altavoz más cercano solo se señaliza para uno o más de los objetos de audio por el calculador de distancia o el decodificador 918 de metadatos al mezclador 930.

Si se decodifican tanto el contenido basado en los canales como los objetos discretos/paramétricos, se mezclan las formas de onda basadas en los canales y las formas de onda de objetos renderizadas antes de emitir las formas de onda resultantes, por ejemplo, por el mezclador 930 de la figura 5 (o antes de alimentarlas a un módulo de postprocesador como el renderizador binaural o el módulo renderizador de altavoces).

Un módulo 940 renderizador binaural puede producir, por ejemplo, una mezcla a la baja binaural del material de audio multicanal, de tal manera que cada canal de entrada esté representado por una fuente de sonido virtual. El procesamiento se lleva a cabo trama por trama en el dominio QMF La binauralización se puede basar, por ejemplo, en respuestas al impulso binaurales de sala medidas.

Un renderizador 922 de altavoces puede convertir, por ejemplo, entre la configuración de canales transmitida y el formato de reproducción deseado. Así, a continuación se le denomina convertidor 922 de formato. El convertidor 922 de formato realiza conversiones para reducir los números de canales de salida, por ejemplo, genera mezclas a la baja. El sistema genera automáticamente matrices optimizadas de mezcla a la baja para la combinación dada de formatos de entrada y salida y aplica estas matrices en un procedimiento de mezcla a la baja. El convertidor 922 de formato admite configuraciones de altavoces estándares, así como para configuraciones aleatorias con posiciones de altavoces no estándares.

Según las realizaciones, se proporciona un dispositivo decodificador. El dispositivo decodificador comprende un decodificador 910 USAC para decodificar un flujo de bits para obtener uno o más canales de entrada de audio, para obtener uno o más objetos de audio de entrada, para obtener metadatos de objeto comprimidos y para obtener uno o más canales de transporte para SAOC.

Además, el dispositivo decodificador comprende un decodificador 915 de SAOC para decodificar el uno o más canales de transporte para SAOC para obtener un grupo de uno o más objetos de audio renderizados.

Además, el dispositivo decodificador comprende un decodificador 918 de metadatos de objeto para decodificar los metadatos de objeto comprimidos para obtener metadatos no comprimidos.

Además, el dispositivo decodificador comprende con convertidor 922 de formato para convertir el uno o más canales de entrada de audio para obtener uno o más canales convertidos.

Además, el dispositivo decodificador comprende un mezclador 930 para mezclar el uno o más objetos de audio renderizados del grupo de uno o más objetos de audio renderizados, el uno o más objetos de audio de entrada y el uno o más canales convertidos para obtener uno o más canales de audio decodificados.

El decodificador 918 de metadatos de objeto y el mezclador 930 juntos conforman un aparato 100 según una de las realizaciones descritas anteriormente, por ejemplo, según la realización de la figura 1.

El decodificador 918 de metadatos de objeto comprende el calculador 110 de distancia del aparato 100 según una de las realizaciones descritas anteriormente, en el que el calculador 110 de distancia está configurado, para cada objeto de audio de entrada del uno o más objetos de audio de entrada, para calcular las distancias de la posición asociada con dicho objeto de audio de entrada hasta los altavoces o para leer las distancias de la posición asociada con dicho objeto de audio de entrada hasta los altavoces y para adoptar una solución con la menor distancia.

El mezclador 930 está configurado para emitir cada objeto de audio de entrada del uno o más objetos de audio de entrada dentro de uno del uno o más canales de audio decodificados al altavoz que corresponde a la solución determinada por el calculador 110 de distancia del aparato 100 según una de las realizaciones descritas anteriormente para dicho objeto de audio de entrada.

En esas realizaciones, el renderizador 920 de objetos puede, por ejemplo, ser opcional. En algunas realizaciones, el renderizador 920 de objetos puede estar presente, pero puede renderizar solo objetos de audio de entrada si la información de metadatos indica que está desactivada la reproducción del altavoz más cercano. Si la información de metadatos indica que está activada la reproducción del canal más cercano, entonces el renderizador 920 de objetos puede pasar, por ejemplo, los objetos de audio de entrada directamente al mezclador sin renderizar los objetos de audio de entrada.

La figura 6 ilustra una estructura de un convertidor de formato. La figura 6 ilustra un configurador 1010 de mezcla a la baja y un procesador de mezcla a la baja para procesar la mezcla a la baja en el dominio QMF (dominio QMF = dominio de filtros espejo en cuadratura).

A continuación, se describen realizaciones y conceptos de realizaciones adicionales de la presente invención.

En las realizaciones, los objetos de audio pueden, por ejemplo, renderizarse por un renderizador de objetos, del lado de la reproducción usando los metadatos e información acerca del entorno de reproducción. Dicha información puede ser, por ejemplo, el número de altavoces o el tamaño de la pantalla. El renderizador de objetos puede calcular, por ejemplo, las señales de altavoces a partir de los datos geométricos y los altavoces disponibles y sus posiciones.

El control de los objetos por el usuario puede realizarse, por ejemplo, mediante metadatos descriptivos, por ejemplo, mediante información acerca de la existencia de un objeto dentro del flujo de bits y pueden realizarse propiedades de alto nivel de los objetos, por ejemplo, mediante metadatos restrictivos, o, por ejemplo, información sobre cómo es posible la interacción o cómo está habilitada por el creador de contenidos.

Según las realizaciones, la señalización, entrega y renderización de objetos de audio puede realizarse, por ejemplo, mediante metadatos de posición, por ejemplo, por metadatos estructurales, por ejemplo, el agrupamiento y jerarquía de los objetos, por ejemplo, por la capacidad de renderizar a un altavoz específico y de señalizar el contenido de canales como objetos y, por ejemplo, para adaptar la escena de objetos al tamaño de la pantalla.

Por lo tanto, se han desarrollado nuevos campos de metadatos además de la posición geométrica ya definida y el nivel del objeto en el espacio 3D.

En general, la posición de un objeto se define por una posición en el espacio 3D que se indica en los metadatos.

Este altavoz de reproducción puede ser un altavoz específico que existe en la configuración de altavoces local. En este caso, el altavoz deseado puede definirse directamente mediante metadatos.

Sin embargo, hay casos en los que el productor no quiere que el contenido de objetos se reproduzca por un altavoz específico, sino en su lugar por el siguiente altavoz disponible, por ejemplo, el altavoz “geométricamente más cercano”. Esto permite una reproducción discreta sin necesidad de definir qué altavoz corresponde a qué señal de audio. Esto es útil, ya que el productor puede desconocer la disposición de altavoces de reproducción, de manera que puede no saber entre qué altavoces puede elegir.

Las realizaciones proporcionan una definición sencilla de una función de distancia que no necesita ninguna operación de raíz cuadrada ni funciones de coseno/seno. En las realizaciones, una función de distancia funciona en el dominio angular (acimut, elevación, distancia), por lo que no es necesaria la transformación a ningún otro sistema de coordenadas (cartesiano, longitud/latitud). Según las realizaciones, hay ponderaciones en la función que proporcionan

una posibilidad de cambiar el foco entre la desviación acimutal, la desviación de elevación y la desviación radial. Las ponderaciones en la función podrían ajustarse, por ejemplo, a las capacidades de la audición humana (por ejemplo, ajustar las ponderaciones según la diferencia apenas perceptible en la dirección acimutal y de elevación). La función podría no solo aplicarse a la determinación del altavoz más cercano, sino también para elegir una respuesta al impulso binaural de sala o una respuesta al impulso relacionada con la cabeza para la renderización binaural. No es necesaria la interpolación de las respuestas al impulso en este caso, en su lugar puede usarse la respuesta al impulso "más cercana".

Según una realización, una bandera "ClosestSpeakerPlayout" denominada *mae_closestSpeakerPlayout* puede definirse, por ejemplo, en los metadatos basados en objetos que hacen que la reproducción del sonido por el altavoz disponible más cercano sin renderización. Un objeto puede marcarse, por ejemplo, para la reproducción por el altavoz más cercano si se ajusta su bandera "ClosestSpeakerPlayout" a uno. La bandera "ClosestSpeakerPlayout" puede definirse, por ejemplo, en un nivel de un "grupo" de objetos. Un grupo de objetos es un concepto de una reunión de objetos relacionados que deben renderizarse o modificarse en conjunto. Si se ajusta esta bandera a uno, es aplicable a todos los elementos del grupo.

Según las realizaciones, para determinar el altavoz más cercano, si está habilitada la bandera *mae_closestSpeakerPlayout* de un grupo, por ejemplo, un grupo de objetos de audio, cada uno de los elementos del grupo se reproducirá por el altavoz que esté más cerca de la posición dada del objeto. No se aplica renderización. Si está habilitada "ClosestSpeakerPlayout" para un grupo, entonces se lleva a cabo el siguiente procesamiento:

Para cada uno de los elementos del grupo, se determina la posición geométrica del elemento (de los metadatos de objetos dinámicos (OAM)) y se determina el altavoz más cercano, ya sea mediante la búsqueda en una tabla previamente almacenada o mediante el cálculo con ayuda de una medición de distancia. Se calcula la distancia de la posición del elemento a cada uno (o solo a una subserie) de los altavoces existentes. El altavoz que da la distancia mínima se define como el altavoz más cercano y el elemento se encamina a su altavoz más cercano. Cada uno de los elementos del grupo se reproduce por su altavoz más cercano.

Tal como ya se ha descrito, las medidas de distancia para la determinación del altavoz más cercano pueden implementarse, por ejemplo, como:

- Las diferencias absolutas ponderadas en el ángulo acimutal y de elevación
- Las diferencias absolutas ponderadas en el acimut, la elevación y el radio/la distancia y por ejemplo (pero sin limitación a):
- Las diferencias absolutas ponderadas a la potencia p (p=2 => Solución de Cuadrados Mínimos)
- Teorema de Pitágoras / Distancia euclidiana (ponderados)

La distancia *d* para las coordenadas cartesianas puede realizarse, por ejemplo, empleando la fórmula

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

donde x_1, y_1, z_1 son los valores de las coordenadas x-, y - y z de una primera posición, donde x_2, y_2, z_2 son los valores de las coordenadas x-, y- y z de una segunda posición y donde d es la distancia entre la primera y la segunda posición.

Una medida de la distancia d para las coordenadas polares puede realizarse, por ejemplo, empleando la fórmula:

$$d = \sqrt{a \cdot (\alpha_1 - \alpha_2)^2 + b \cdot (\beta_1 - \beta_2)^2 + c \cdot (r_1 - r_2)^2}$$

donde α_1, β_1 y r_1 son las coordenadas polares de una primera posición, donde α_2, β_2 y r_2 son las coordenadas polares de una segunda posición y donde d es la distancia entre la primera y la segunda posición.

La diferencia angular ponderada puede definirse, por ejemplo, según

$$\text{diffAngle} = \text{acos}(\cos(\alpha_1 - \alpha_2) \cdot \cos(\beta_1 - \beta_2))$$

En lo que respecta a la distancia ortodrómica, la distancia de gran arco o la distancia de gran círculo, la distancia medida a lo largo de la superficie de una esfera (a diferencia de una línea recta a través del interior de la esfera). Pueden emplearse, por ejemplo, operaciones de raíz cuadrada y funciones trigonométricas. Las coordenadas pueden transformarse, por ejemplo, a latitud y longitud.

Volviendo a la fórmula presentada anteriormente:

$$\Delta(P_1,P_2)=|\beta_1-\beta_2|+|\alpha_1-\alpha_2|+|r_1-r_2|$$,

la fórmula puede verse como una *geometría Taxicab* modificada usando coordenadas polares en lugar de coordenadas cartesianas como en la definición de geometría taxicab original

$$\Delta(P_1,P_2)=|x_1-x_2|+|y_1-y_2|$$.

Con esta fórmula, es posible añadir ponderaciones a la elevación, el acimut y/o el radio. De esa manera es posible declarar que la desviación de acimut deber ser menos tolerable que una desviación de elevación mediante la ponderación de la desviación de acimut con un número elevado:

$$\Delta(P_1,P_2)=b\cdot|\beta_1-\beta_2|+a\cdot|\alpha_1-\alpha_2|+c\cdot|r_1-r_2|$$.

Como un comentario secundario adicional, debe observarse que, en las realizaciones, puede considerarse que el "audio de objeto renderizado" de la figura 2, por ejemplo, es "audio basado en objetos renderizado". En la figura 2, se usan usacConfigExtention con respecto a metadatos de objeto estáticos y usacExtension solo como ejemplos de realizaciones particulares.

Con respecto a la figura 3, debe observarse que, en algunas realizaciones, los metadatos de objetos dinámicos de la figura 3 pueden ser, por ejemplo, OAM posicionales (metadatos de objeto de audio, datos posicionales + ganancia). En algunas realizaciones, pueden llevarse a cabo las "señales de ruta", por ejemplo, enrutando las señales a un convertidor de formato o a un renderizador de objetos.

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

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

Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo, un disco blando, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales de control legibles electrónicamente, que cooperan (o son capaces de cooperar) con un sistema informático programable de tal manera que se realiza el método respectivo.

Algunas realizaciones según la invención comprenden un portador de datos no transitorio que tiene señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de tal manera que se realice uno de los métodos descritos en el presente documento.

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

Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenado en un portador legible por una máquina.

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

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

Una realización adicional del método de la invención es, por lo tanto, un flujo de datos o una secuencia de señales

que representa el programa informático para realizar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para transferirse mediante una conexión de comunicación de datos, por ejemplo, mediante Internet.

Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurada o adaptada para realizar uno de los métodos descritos en el presente documento.

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

En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo, una formación de puertas programable en campo) para realizar algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una formación de puertas programable en campo puede cooperar con un microprocesador para realizar uno de los métodos descritos en el presente documento. Generalmente, los métodos se realizan preferentemente mediante cualquier aparato de hardware.

Las realizaciones descritas anteriormente son meramente ilustrativas para los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para otros expertos en la técnica. Se pretende, por lo tanto, que solo estén limitadas por el alcance de las reivindicaciones de patente pendientes y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

Literatura

[1] “System and Method for Adaptive Audio Signal Generation, Coding and Rendering”, solicitud de patente número: US20140133683 A1 (Reivindicación 48)

[2] “Reflected sound rendering for object-based audio”, solicitud de patente número: WO2014036085 A1 (Capítulo Aplicaciones de Reproducción)

[3] “Upmixing object based audio”, solicitud de patente número: US20140133682 A1 (BREVE DESCRIPCIÓN DE LAS REALIZACIONES A MODO DE EJEMPLO + Reivindicación 71 b))

[4] “Audio Definition Model”, EBU-TECH 3364, https://tech.ebu.ch/docs/tech/tech3364.pdf

[5] “System and Tools for Enhanced 3D Audio Authoring and Rendering”, solicitud de patente número: US20140119581 A1

REIVINDICACIONES

1. Aparato (100) para reproducir un objeto de audio asociado con una posición, que comprende:

un calculador (110) de distancia para calcular distancias de la posición hasta los altavoces,

en el que el calculador (110) de distancia está configurado para adoptar una solución con la distancia más corta, y

en el que el aparato (100) está configurado para reproducir el objeto de audio usando el altavoz que corresponde a la solución, caracterizado porque el

calculador (110) de distancia está configurado para calcular las distancias dependiendo de una función de distancia que devuelve una diferencia angular ponderada dependiendo de la diferencia entre dos ángulos acimutales y dependiendo de la diferencia entre dos ángulos de elevación, en el que la función de distancia se define según

diffAngle = acos(cos(azDiff) * cos(elDiff)),

en el que azDiff indica la diferencia entre los dos ángulos acimutales, en el que elDiff indica la diferencia entre los dos ángulos de elevación, en el que diffAngle indica la diferencia angular ponderada.

o

el calculador (110) de distancia está configurado para calcular las distancias desde la posición hasta los altavoces, de manera que cada distancia $\Delta(P_1, P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| \quad ,$$

o según

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2| \quad ,$$

o según

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| \quad ,$$

o según

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2| \quad ,$$

en el que α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, r_1 indica un radio de la posición y r_2 indica un radio de dicho uno de los altavoces, a es un primer número, b es un segundo número y c es un tercer número, o

en el que α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces, y β_2 indica un ángulo de elevación de la posición, r_1 indica un radio de dicho uno de los altavoces y r_2 indica un radio de la posición, α es un primer número, b es un segundo número y c es un tercer número.

2. Aparato (100) según la reivindicación 1,

en el que el calculador (110) de distancia está configurado para calcular las distancias desde la posición hasta los altavoces solo si está habilitada la bandera de reproducción del altavoz más cercano *mdae_closestSpeakerPlayout*, recibida por el aparato (100),

en el que el calculador (110) de distancia está configurado para adoptar una solución con la distancia más corta solo si está habilitada la bandera de reproducción del altavoz más cercano *mdae_closestSpeakerPlayout*, y

en el que el aparato (100) está configurado para reproducir el objeto de audio usando el altavoz que corresponde

a la solución solo si está habilitada la bandera de reproducción del altavoz más cercano *mdae_closestSpeakerPlayout*.

3. Aparato (100) según la reivindicación 2, en el que el aparato (100) está configurado para no llevar a cabo ninguna renderización en el objeto de audio, si está habilitada la bandera de reproducción del altavoz más cercano *mdae_closestSpeakerPlayout*.

4. Dispositivo decodificador que comprende:

un decodificador (910) USAC para decodificar un flujo de bits para obtener uno o más canales de entrada de audio, para obtener uno o más objetos de audio de entrada, para obtener metadatos de objeto comprimidos y para obtener uno o más canales de transporte para SAOC,

un decodificador (915) SAOC para decodificar el uno o más canales de transporte para SAOC para obtener un grupo de uno o más objetos de audio renderizados,

un decodificador (918) de metadatos de objeto, para decodificar los metadatos de objeto comprimidos para obtener metadatos descomprimidos,

un convertidor (922) de formato para convertir el uno o más canales de entrada de audio para obtener uno o más canales convertidos y

un mezclador (930) para mezclar el uno o más objetos de audio renderizados del grupo de uno o más objetos de audio renderizados, el uno o más objetos de audio de entrada y el uno o más canales convertidos para obtener uno o más canales de audio decodificados,

en el que el decodificador (918) de metadatos de objeto comprende el calculador (110) de distancia del aparato (100) según una de las reivindicaciones anteriores, en el que el calculador (110) de distancia está configurado, para cada objeto de audio de entrada del uno o más objetos de audio de entrada, para calcular distancias de la posición asociada con dicho objeto de audio de entrada hasta los altavoces y para adoptar una solución con la menor distancia y

en el que el mezclador (930) está configurado para emitir cada objeto de audio de entrada del uno o más objetos de audio de entrada dentro de uno del uno o más canales de audio decodificados al altavoz que corresponde a la solución determinada por el calculador (110) de distancia del aparato (100) según una de las reivindicaciones anteriores para dicho objeto de audio de entrada.

5. Método para reproducir un objeto de audio asociado con una posición, que comprende:

calcular distancias de la posición hasta los altavoces,

adoptar una solución con una menor distancia y

reproducir el objeto de audio usando el altavoz que corresponde a la solución, caracterizado porque:

calcular las distancias se lleva a cabo dependiendo de una función de distancia que devuelve una diferencia angular ponderada dependiendo de la diferencia entre dos ángulos acimutales y dependiendo de la diferencia entre dos ángulos de elevación, en el que la función de distancia se define según

$$\text{diffAngle} = \text{acos}(\cos(\text{azDiff}) * \cos(\text{elDiff})),$$

en el que azDiff indica la diferencia entre los dos ángulos acimutales, en el que elDiff indica la diferencia entre los dos ángulos de elevación, en el que diffAngle indica la diferencia angular ponderada.

o

calcular las distancias desde la posición hasta los altavoces se lleva a cabo de manera que cada distancia $\Delta(P_1, P_2)$ de la posición hasta uno de los altavoces se calcula según

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| \quad ,$$

o según

$$\Delta(P_1, P_2) = |\beta_1 - \beta_2| + |\alpha_1 - \alpha_2| + |r_1 - r_2|,$$

o según

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2|,$$

o según

$$\Delta(P_1, P_2) = b \cdot |\beta_1 - \beta_2| + a \cdot |\alpha_1 - \alpha_2| + c \cdot |r_1 - r_2|,$$

en el que α_1 indica un ángulo acimutal de la posición, α_2 indica un ángulo acimutal de dicho uno de los altavoces, β_1 indica un ángulo de elevación de la posición, β_2 indica un ángulo de elevación de dicho uno de los altavoces, r_1 indica un radio de la posición, r_2 indica un radio de dicho uno de los altavoces, a es un primer número, b es un segundo número y c es un tercer número, o

en el que α_1 indica un ángulo acimutal de dicho uno de los altavoces, α_2 indica un ángulo acimutal de la posición, β_1 indica un ángulo de elevación de dicho uno de los altavoces, y β_2 indica un ángulo de elevación de la posición, r_1 indica un radio de dicho uno de los altavoces y r_2 indica un radio de la posición, a es un primer número, b es un segundo número y c es un tercer número.

6. Producto de programa informático que comprende instrucciones que, cuando se implementan en un ordenador o procesador de señales, provocará que dicho ordenador o procesador de señales realice el método de la reivindicación 5.

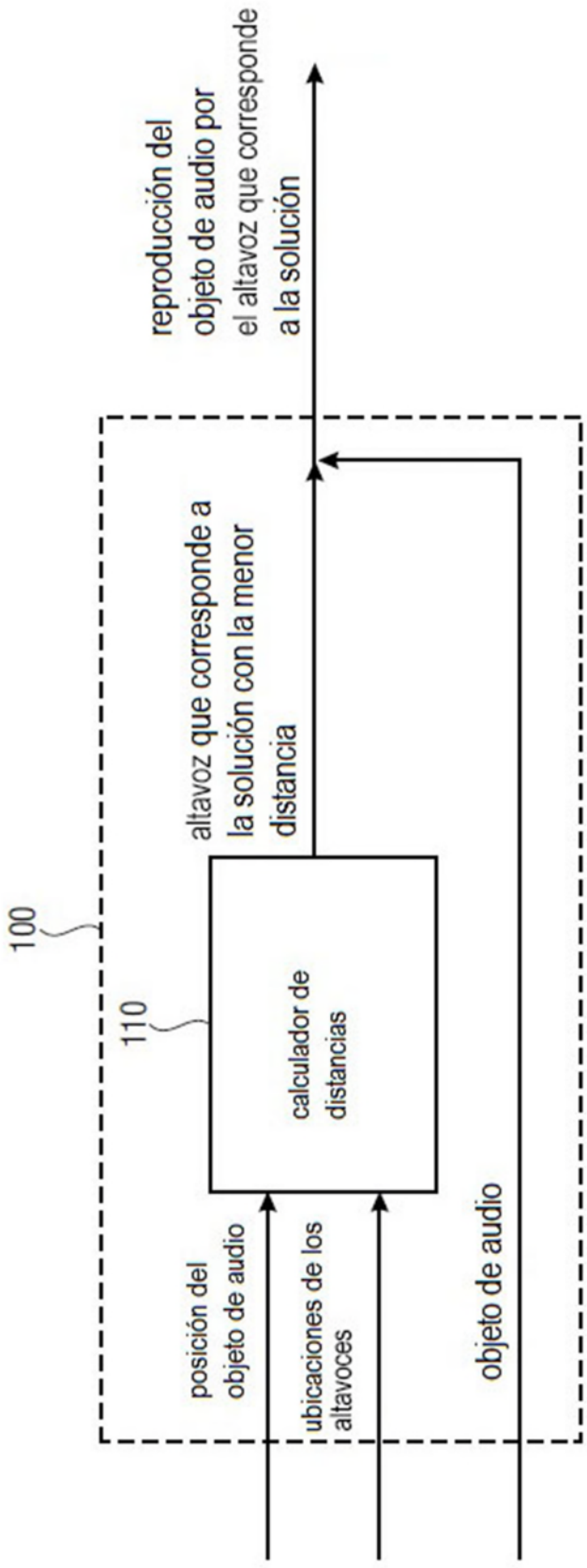
100
110
calculador de distancias
posición del objeto de audio
ubicaciones de los altavoces
objeto de audio
altavoz que corresponde a la solución con la menor distancia
reproducción del objeto de audio por el altavoz que corresponde a la solución

FIG 1

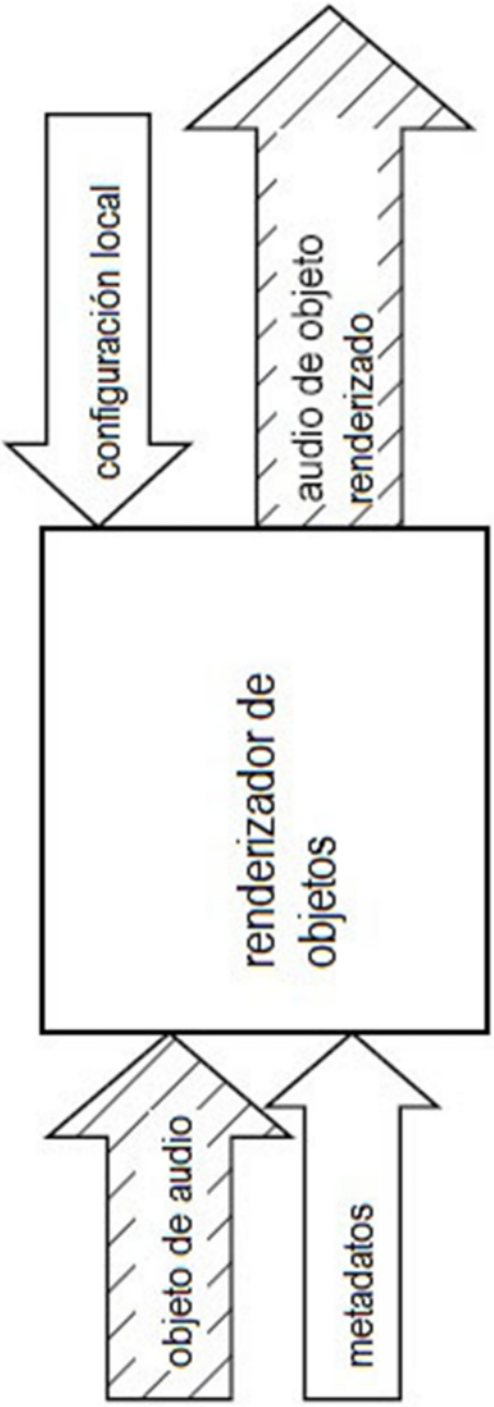
configuración local
audio de objeto renderizado
renderizador de objetos
objeto de audio
metadatos

FIG 2

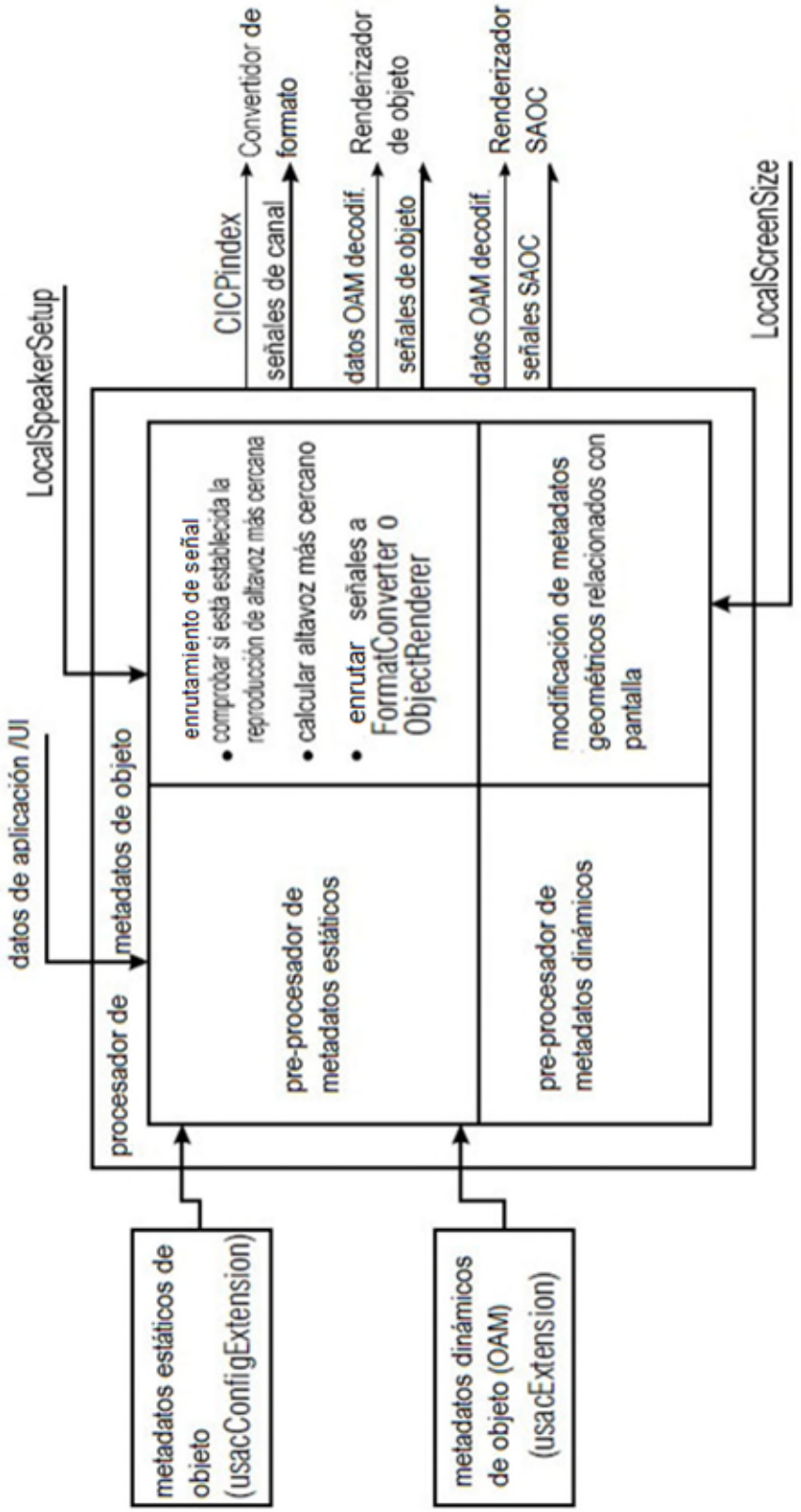
metadatos estáticos de objeto (usacConfigExtension)
metadatos dinámicos de objeto (OAM) (usacExtension)
procesador de metadatos de objeto
datos de aplicación /UI
LocalSpeakerSetup
pre-procesador de metadatos estáticos
enrutamiento de señal
• comprobar si está establecida la reproducción de altavoz más cercana
• calcular altavoz más cercano
• enrutar señales a FormatConverter o ObjectRenderer
pre-procesador de metadatos dinámicos
modificación de metadatos geométricos relacionados con pantalla
CICPindex
señales de canal
Convertidor de formato
datos OAM decodif.
señales de objeto
Renderizador de objeto
datos OAM decodif.
señales SAOC
Renderizador SAOC
LocalScreenSize

FIG 3

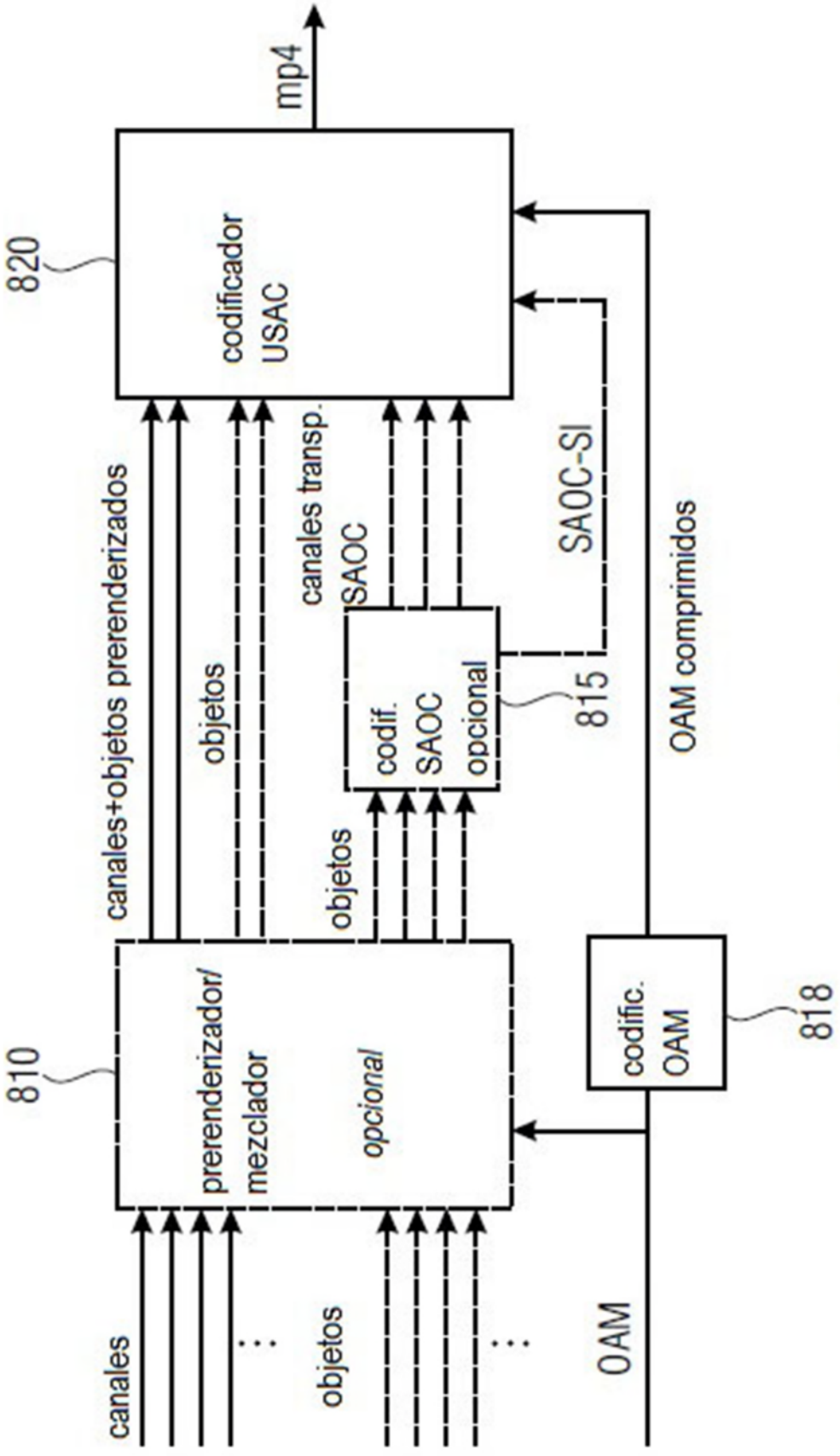
mp4
820
codificador
USAC
canales+objetos prerenderizados
objetos
canales transp.
SAOC
SAOC-SI
OAM comprimidos
codif.
SAOC
opcional
815
objetos
810
prerenderizador/
mezclador
opcional
codific.
OAM
818
canales
objetos
OAM

FIG 4

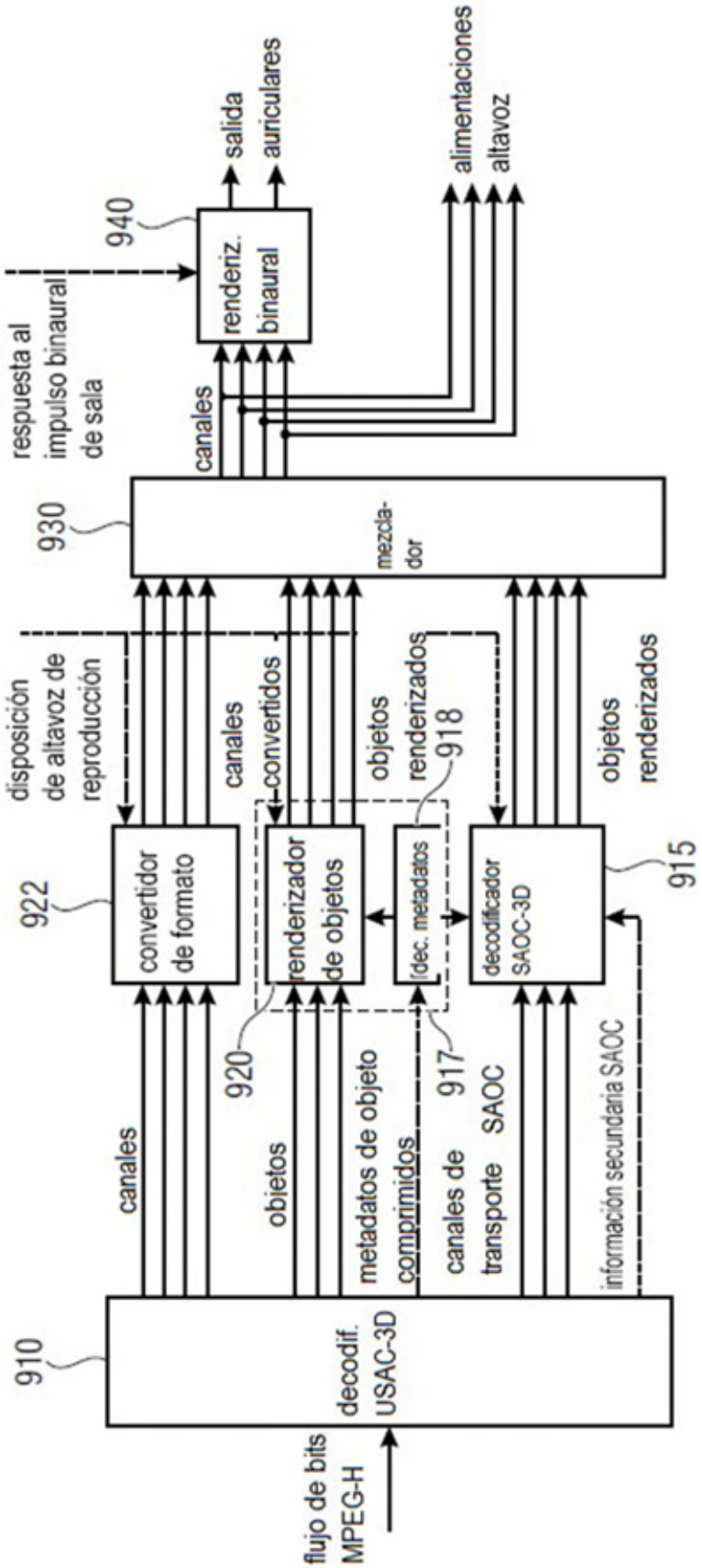
flujo de bits MPEG-H
910
decodif. USAC-3D
canales
objetos
metadatos de objeto comprimidos
canales de transporte SAOC
información secundaria SAOC
920
917
922
convertidor de formato
renderizador de objetos
dec. metadatos
decodificador SAOC-3D
915
disposición de altavoz de reproducción
canales convertidos
objetos renderizados
918
objetos renderizados
930
mezcla-dor
respuesta al impulso binaural de sala
canales
940
renderiz. binaural
salida auriculares
alimentaciones altavoz

FIG 5

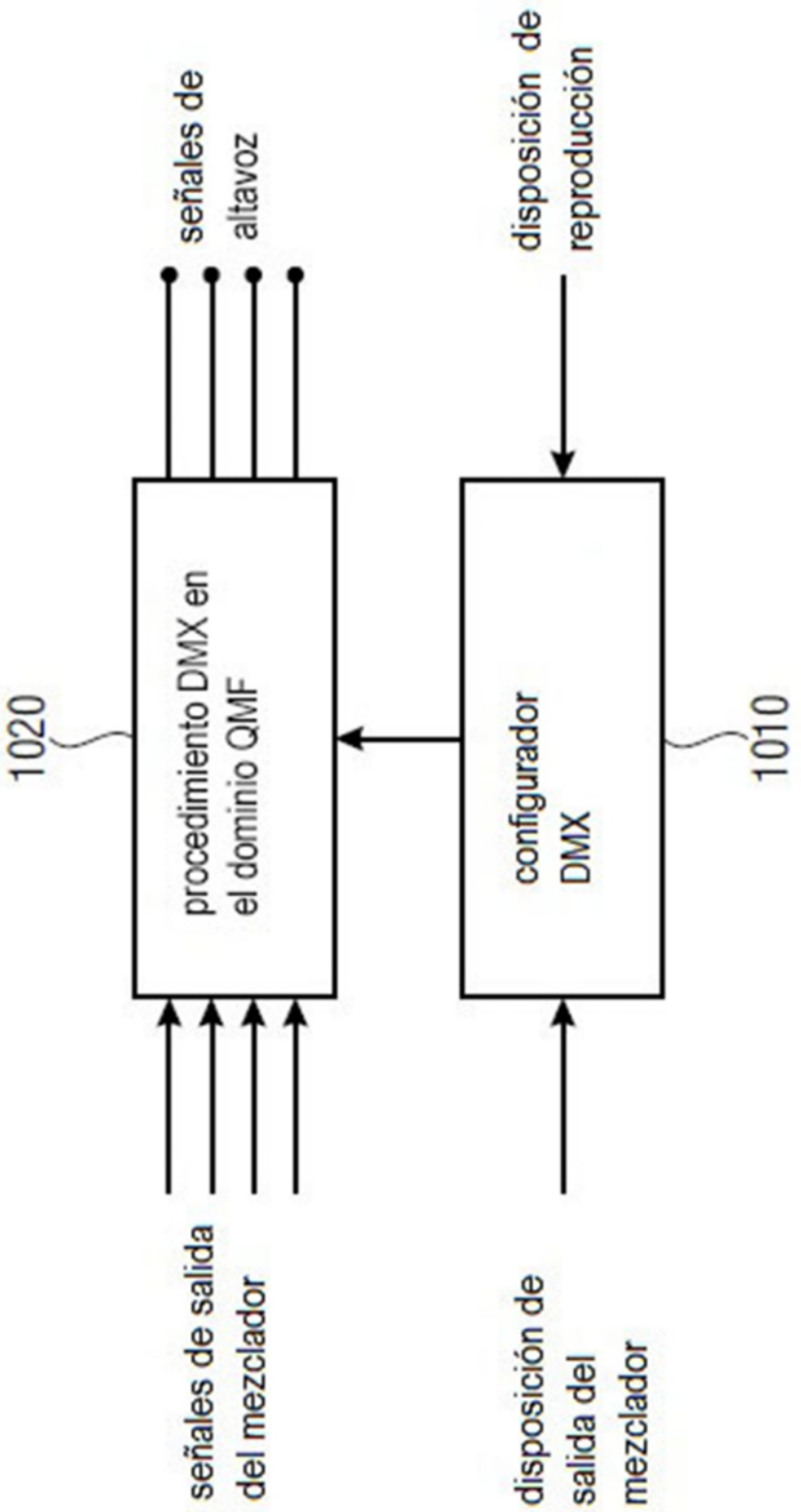
señales de salida del mezclador
disposición de salida del mezclador
procedimiento DMX en el dominio QMF
configurador DMX
1020
1010
señales de altavoz
disposición de reproducción

FIG 6