(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 980 463**

(51) Int. Cl.:

G06F 3/16 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **20.08.2019 PCT/EP2019/072195**

(87) Fecha y número de publicación internacional: **05.03.2020 WO20043539**

(96) Fecha de presentación y número de la solicitud europea: **20.08.2019 E 19753392 (0)**

(97) Fecha y número de publicación de la concesión europea: **03.04.2024 EP 3844606**

(54) Título: **Aparato de audio y método de procesamiento de audio**

(30) Prioridad:

28.08.2018 EP 18191241

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
01.10.2024

(73) Titular/es:

KONINKLIJKE PHILIPS N.V. (100.0%)
High Tech Campus 52
5656 AG Eindhoven, NL

(72) Inventor/es:

DE BRUIJN, WERNER PAULUS JOSEPHUS y
SOUVIRAA-LABASTIE, NATHAN

(74) Agente/Representante:

ISERN JARA, Jorge

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 de audio y método de procesamiento de audio

Campo de la invención

La invención se refiere a un aparato de audio y a un método de procesamiento de audio, y en particular, pero no exclusivamente, al uso de los mismos para soportar una aplicación de Realidad Aumentada/Virtual.

Antecedentes de la invención

La variedad y el alcance de las experiencias basadas en contenidos audiovisuales han aumentado sustancialmente en los últimos años, con el desarrollo y la introducción continuos de nuevos servicios y formas de utilizar y consumir dichos contenidos. En concreto, se están desarrollando numerosos servicios, aplicaciones y experiencias espaciales e interactivas para ofrecer a los usuarios una experiencia más envolvente e inmersiva.

Ejemplos de tales aplicaciones son la Realidad Virtual (VR), la Realidad Aumentada (AR) y la Realidad Mixta (MR), que se están convirtiendo rápidamente en la corriente principal, con una serie de soluciones dirigidas al mercado de consumo. Varios organismos de normalización también están elaborando una serie de normas. Estas actividades de normalización están desarrollando activamente normas para los diversos aspectos de los sistemas de VR/AR/MR, que incluyen la transmisión, la difusión, el renderizado, etc.

Las aplicaciones de VR tienden a proporcionar experiencias de usuario correspondientes a que el usuario se encuentra en un mundo/entorno/escena diferente, mientras que las aplicaciones de AR (incluyendo la Realidad Mixta MR) tienden a proporcionar experiencias de usuario correspondientes a que el usuario se encuentra en el entorno actual, pero con información adicional u objetos virtuales o información añadida. Así, las aplicaciones de VR tienden a proporcionar un mundo/escena sintético totalmente inmersivo, mientras que las aplicaciones de AR tienden a proporcionar un mundo/escena parcialmente sintético que se superpone a la escena real en la que el usuario está físicamente presente. Sin embargo, los términos se utilizan a menudo indistintamente y tienen un alto grado de solapamiento. En lo sucesivo, el término Realidad Virtual/VR se utilizará para designar tanto la Realidad Virtual como la Realidad Aumentada.

Como ejemplo, un servicio cada vez más popular es la provisión de imágenes y audio de tal forma que el usuario sea capaz de interactuar activa y dinámicamente con el sistema para cambiar los parámetros de la renderización de tal forma que ésta se adapte al movimiento y a los cambios en la posición y orientación del usuario. Una característica muy atractiva en muchas aplicaciones es la posibilidad de cambiar la posición efectiva de visión y la dirección de visión del espectador, tal como por ejemplo permitir que el espectador se mueva y "mire a su alrededor" en la escena que se presenta.

Dicha característica puede permitir específicamente proporcionar una experiencia de realidad virtual a un usuario. Esto puede permitir al usuario moverse (relativamente) libremente en un entorno virtual y cambiar dinámicamente su posición y hacia dónde mira. Normalmente, estas aplicaciones de realidad virtual se basan en un modelo tridimensional de la escena que se evalúa dinámicamente para proporcionar la vista específica solicitada. Este enfoque es bien conocido, por ejemplo, en las aplicaciones de juegos, tal como los tiradores en primera persona, para ordenadores y consolas.

También es deseable, en particular para aplicaciones de realidad virtual, que la imagen presentada sea una imagen tridimensional. De hecho, para optimizar la inmersión del espectador, normalmente se prefiere que el usuario experimente la escena presentada como una escena tridimensional. De hecho, es preferible que una experiencia de realidad virtual permita al usuario seleccionar su propia posición, punto de vista de la cámara y momento en el tiempo en relación con un mundo virtual.

Típicamente, las aplicaciones de realidad virtual están inherentemente limitadas al estar basadas en un modelo predeterminado de la escena, y típicamente en un modelo artificial de un mundo virtual. En algunas aplicaciones, se puede proporcionar una experiencia de realidad virtual basada en la captura del mundo real. En muchos casos, este enfoque tiende a basarse en un modelo virtual del mundo real que se construye a partir de las capturas del mundo real. La experiencia de realidad virtual se genera entonces evaluando este modelo.

Muchos de los enfoques actuales tienden a ser subóptimos y suelen requerir una gran cantidad de recursos informáticos o de comunicación y/o proporcionan una experiencia de usuario subóptima con, por ejemplo, una calidad reducida o una libertad restringida.

Como ejemplo de aplicación, se han introducido en el mercado gafas de realidad virtual que permiten a los espectadores experimentar vídeo capturado en 360° (panorámico) o 180°. Estos vídeos de 360° se suelen capturar previamente con equipos de cámaras en los que las imágenes individuales se unen en un único mapa esférico. Los formatos estéreo habituales para vídeo de 180° o 360° son arriba/abajo e izquierda/derecha. Al igual que en el vídeo

estéreo no panorámico, las imágenes del ojo izquierdo y el derecho se comprimen, por ejemplo, como parte de un único flujo de vídeo H.264.

Además de la renderización visual, la mayoría de las aplicaciones de VR/AR ofrecen una experiencia de audio. En muchas aplicaciones, el audio proporciona preferentemente una experiencia de audio espacial donde se percibe que las fuentes de audio llegan desde posiciones que corresponden a las posiciones de los objetos correspondientes en la escena visual. De este modo, las escenas de audio y vídeo se perciben preferiblemente de forma coherente y ambas proporcionan una experiencia espacial completa.

En cuanto al audio, hasta ahora la atención se ha centrado principalmente en la reproducción con auriculares utilizando tecnología de renderización de audio binaural. En muchos escenarios, la reproducción con auriculares permite una experiencia altamente inmersiva y personalizada para el usuario. Gracias al seguimiento de cabeza, el renderizado puede responder a los movimientos de la cabeza del usuario, lo que aumenta enormemente la sensación de inmersión. Un ejemplo de un sistema de audio que utiliza múltiples renderizadores diferentes para renderizar audio se divulga en el documento GB2550877A.

Recientemente, tanto en el mercado como en los debates sobre estándares, se están empezando a proponer casos de uso que implican un aspecto "social" o "compartido" de la VR (y la AR), es decir, la posibilidad de compartir una experiencia junto con otras personas. Puede tratarse de personas en distintos lugares, pero también de personas en el mismo lugar (o una combinación de ambos). Por ejemplo, varias personas en la misma sala pueden compartir la misma experiencia de VR con una proyección (audio y vídeo) de cada participante presente en el contenido/escena de VR. Por ejemplo, en un juego en el que participan varias personas, cada jugador puede tener una ubicación diferente en la escena del juego y, en consecuencia, una proyección diferente de la escena de audio y vídeo.

Como ejemplo específico, MPEG intenta estandarizar un flujo de bits y un decodificador para experiencias de AR/VR realistas e inmersivas con seis grados de libertad. La VR social es una característica importante y permite a los usuarios interactuar en un entorno compartido (juegos, conferencias telefónicas, compras en línea, etc.). El concepto de VR social también facilita que una experiencia de VR se convierta en una actividad más social para los usuarios que se encuentran físicamente en el mismo lugar, pero en los que, por ejemplo, una pantalla montada en la cabeza u otro casco de VR proporciona un aislamiento perceptivo del entorno físico.

Una desventaja de la reproducción de auriculares en tales casos de uso "social" o "compartido" de AR (o VR) es que con cada usuario usando auriculares individuales, los usuarios que están en la misma ubicación (por ejemplo, habitación) están al menos parcialmente aislados acústicamente unos de otros, lo que disminuye la parte "social" de la experiencia (por ejemplo, se hace difícil o incómodo para las personas que están una al lado de la otra tener una conversación natural).

Esto puede solucionarse utilizando altavoces en lugar de auriculares para la reproducción de audio. Sin embargo, esto tiene el inconveniente de que la reproducción de audio no se puede adaptar y personalizar tan libremente a cada usuario. Por ejemplo, dificulta la adaptación dinámica de la reproducción de audio a los movimientos de la cabeza y, en concreto, a los cambios de orientación de la cabeza de cada usuario. Este efecto es muy importante para una experiencia inmersiva, por lo que los altavoces tienden a ser poco óptimos para generar una experiencia de usuario optimizada.

Por lo tanto, sería ventajoso un enfoque mejorado para el procesamiento de audio, en particular para una experiencia/aplicación de realidad virtual aumentada/mixta, sería ventajosa. En particular, sería ventajoso un enfoque que permita un funcionamiento mejorado, una mayor flexibilidad, una complejidad reducida, una implementación facilitada, una experiencia de audio mejorada, una percepción más coherente de una escena de audio y visual, una personalización mejorada, una experiencia de realidad virtual mejorada y/o un rendimiento y/o funcionamiento mejorados.

Resumen de la invención

En consecuencia, la Invención busca preferiblemente mitigar, aliviar o eliminar una o más de las desventajas antes mencionadas individualmente o en cualquier combinación.

De acuerdo con un aspecto de la invención, se proporciona un aparato de audio de acuerdo con la reivindicación 1.

El enfoque puede proporcionar una experiencia de usuario mejorada en muchas realizaciones y puede proporcionar específicamente una experiencia de usuario mejorada para muchas aplicaciones de realidad virtual (incluyendo realidad aumentada y mixta), incluyendo específicamente experiencias sociales o compartidas. El enfoque puede proporcionar un rendimiento mejorado utilizando el renderizado híbrido. Por ejemplo, en muchas realizaciones, puede permitir a los usuarios en la misma habitación hablar directamente con mayor facilidad, al tiempo que proporciona una renderización específica y personalizada de la escena de audio.

El enfoque puede reducir la complejidad y los requisitos de recursos.

En algunas realizaciones, el aparato puede comprender un primer controlador para controlar el conjunto de altavoces a partir del primer conjunto de señales de audio y un segundo controlador para controlar los auriculares a partir del segundo conjunto de señales de audio. El primer conjunto de señales de audio puede ser específicamente un conjunto de señales envolventes y el segundo conjunto de señales de audio puede ser específicamente una señal estéreo binaural.

El primer indicador de propiedad de renderización de audio es indicativo de una propiedad del renderizado que se va a aplicar al primer elemento de audio o una propiedad del primer elemento de audio.

De acuerdo con una característica de la invención, el aparato comprende además un receptor de pose del oyente para recibir una pose del oyente indicativa de una pose de un oyente, y el primer renderizador está dispuesto para generar el primer conjunto de señales de audio independientemente de la pose del oyente y el segundo renderizador está dispuesto para generar el segundo conjunto de señales de audio en respuesta a la pose del oyente.

El aparato de audio puede proporcionar una experiencia de usuario muy ventajosa y flexible que permite una estrecha coherencia entre, por ejemplo, el movimiento del oyente y la escena de audio percibida. Una pose puede referirse a datos de posición y/u orientación, y también puede denominarse colocación. Una pose del oyente puede ser una indicación de posición para un oyente, una indicación de orientación para un oyente, o una indicación combinada de posición y orientación para un oyente. Una pose/ubicación puede estar representada por uno o más valores que proporcionan una indicación de una posición y/o dirección.

De acuerdo con una característica de la invención, el aparato está dispuesto para generar señales de audio para una pluralidad de oyentes en el que el primer renderizador está dispuesto para generar el primer conjunto de señales de audio como un conjunto común de señales de audio para la pluralidad de oyentes; y el segundo renderizador está dispuesto para generar el segundo conjunto de señales de audio para auriculares para un primer oyente de la pluralidad de oyentes y para generar un tercer conjunto de señales de audio para auriculares para un segundo oyente de la pluralidad de oyentes.

El aparato de audio puede proporcionar un soporte ventajoso para múltiples usuarios. En muchas aplicaciones, se puede lograr un soporte mejorado de baja complejidad y uso de recursos, pero proporcionando una experiencia de usuario atractiva, a menudo con una percepción coherente y natural de la etapa de audio.

El segundo conjunto de señales de audio puede generarse en respuesta a una primera pose de oyente para el primer oyente y el tercer conjunto de señales de audio puede generarse en respuesta a una segunda pose de oyente para el segundo oyente. El primer conjunto de señales de audio puede generarse independientemente de la postura del oyente.

De acuerdo con una característica opcional de la invención, la primera parte es un subintervalo de frecuencia del primer elemento de audio.

Esto puede proporcionar un rendimiento mejorado en muchas realizaciones.

De acuerdo con una característica de la invención, el selector está dispuesto para seleccionar diferentes renderizadores del primer renderizador y del segundo renderizador para la primera parte del primer elemento de audio y para una segunda parte del primer elemento de audio.

Esto puede proporcionar una experiencia de usuario mejorada en muchas realizaciones. El selector puede estar específicamente dispuesto para seleccionar diferentes renderizadores para diferentes intervalos de frecuencia del primer elemento de audio.

Esto puede proporcionar un enfoque eficiente en muchas aplicaciones. El indicador de la propiedad de renderización de audio puede ser indicativo de si el primer elemento de audio es diegético o no.

De acuerdo con una característica opcional de la invención, el indicador de propiedad de renderización de audio es indicativo de un formato de audio del primer elemento de audio.

Esto puede proporcionar una experiencia de usuario mejorada en muchas realizaciones. El indicador de propiedad de renderización de audio puede ser indicativo de un formato de audio de un conjunto de formatos de audio que incluye al menos un formato de audio del grupo de: un formato de objeto de audio; un formato de audio Ambisonics de Orden Superior; y un formato de audio de señal de canal de audio.

De acuerdo con una característica opcional de la invención, el indicador de propiedad de renderización de audio es indicativo de un tipo de fuente de audio para el primer elemento de audio.

Esto puede proporcionar una experiencia de usuario mejorada en muchas realizaciones. El indicador de propiedad de renderización de audio puede ser indicativo de un tipo de fuente de audio de un conjunto de tipos de fuente de audio que incluye al menos un tipo de fuente de audio del grupo de: audio de voz; audio de música; audio de primer plano; audio de fondo; voz sobre audio; y audio de narrador.

De acuerdo con una característica opcional de la invención, en la que el indicador de propiedad de renderización de audio es indicativo de una propiedad de renderización de orientación para el renderizado del renderizado del primer elemento de audio.

Esto puede proporcionar una mejor experiencia de usuario y/o rendimiento en muchas realizaciones.

De acuerdo con una característica de la invención, el indicador de propiedad de reproducción de audio es indicativo de si la primera parte del primer elemento de audio está destinada a reproducirse a través de altavoces o auriculares.

Esto puede proporcionar una mejor experiencia de usuario y/o rendimiento en muchas realizaciones.

De acuerdo con una característica opcional de la invención, el receptor se dispone además para recibir datos visuales indicativos de una escena virtual correspondiente a la escena de audio, y el indicador de propiedad de renderización de audio es indicativo de si el primer elemento de audio representa una fuente de audio correspondiente a un objeto de escena visual.

Esto puede proporcionar una mejor experiencia de usuario y/o rendimiento en muchas realizaciones.

En algunas realizaciones, el indicador de propiedad de renderización de audio puede ser indicativo de si el primer elemento de audio representa una fuente de audio correspondiente a un objeto de escena que está dentro de una ventana gráfica determinada para una pose de oyente actual.

De acuerdo con una característica opcional de la invención, el aparato comprende además una entrada de usuario para recibir una entrada de usuario y en donde el selector está dispuesto para seleccionar entre el primer renderizador y el segundo renderizador para renderizar al menos la primera parte del primer elemento de audio en respuesta a la entrada de usuario.

Esto puede proporcionar una experiencia de usuario mejorada en muchas realizaciones.

De acuerdo con una característica opcional de la invención, el selector está dispuesto para determinar una propiedad de audio del primer elemento de audio y para seleccionar entre el primer renderizador y el segundo renderizador para el renderizado de al menos la primera parte del primer elemento de audio en respuesta a la propiedad de audio.

Esto puede proporcionar una mejor experiencia de usuario y/o rendimiento en muchas realizaciones.

De acuerdo con un aspecto de la invención se proporciona un método de acuerdo con la reivindicación 13.

Estos y otros aspectos, características y ventajas de la invención serán evidentes y se dilucidarán con referencia a la(s) realización(es) descrita(s) a continuación.

Breve descripción de los dibujos

Las realizaciones de la invención se describirán, a modo de ejemplo únicamente, con referencia a los dibujos, en los que

La FIG. 1 ilustra un ejemplo de sistema de realidad virtual basado en cliente-servidor; y
La FIG. 2 ilustra un ejemplo de elementos de un aparato de audio de acuerdo con algunas realizaciones de la invención.

Descripción detallada de algunas realizaciones de la invención

Las experiencias de realidad virtual (incluyendo las de realidad aumentada y mixta) que permiten a un usuario moverse en un mundo virtual o aumentado son cada vez más populares y se están desarrollando servicios para satisfacer tales demandas. En muchos de estos enfoques, los datos visuales y sonoros pueden generarse dinámicamente para reflejar la postura actual del usuario (o espectador).
En este campo, los términos colocación y pose se utilizan como un término común para posición y/o dirección/orientación. La combinación de la posición y la dirección/orientación de, por ejemplo, un objeto, una cámara, una cabeza o una vista puede denominarse pose o colocación. Así, una indicación de posición o pose puede comprender hasta seis valores/componentes/grados de libertad con cada valor/componente describiendo típicamente una propiedad individual de la posición/ubicación o la orientación/dirección del objeto correspondiente. Por supuesto, en muchas situaciones, una colocación o pose puede representarse con menos componentes, por ejemplo, si uno o

más componentes se consideran fijos o irrelevantes (por ejemplo, si se considera que todos los objetos están a la misma altura y tienen una orientación horizontal, cuatro componentes pueden proporcionar una representación completa de la pose de un objeto). En lo sucesivo, el término pose se utiliza para referirse a una posición y/u orientación que puede estar representada por uno a seis valores (correspondientes al máximo de grados de libertad posibles).

Muchas aplicaciones de VR se basan en una pose que tiene el máximo de grados de libertad, es decir, tres grados de libertad de cada uno de la posición y la orientación resultando en un total de seis grados de libertad. Así, una pose puede representarse mediante un conjunto o vector de seis valores que representan los seis grados de libertad y, por lo tanto, un vector de pose puede proporcionar una posición tridimensional y/o una indicación de dirección tridimensional. Sin embargo, se apreciará que, en otras realizaciones, la pose puede estar representada por menos valores.

Un sistema o entidad basado en proporcionar el máximo grado de libertad para el espectador es típicamente referido como teniendo 6 Grados de Libertad (6DoF). Muchos sistemas y entidades sólo proporcionan una orientación o posición y se suele decir que tienen 3 grados de libertad (3DoF).

Típicamente, la aplicación de realidad virtual genera una salida tridimensional en forma de imágenes de vista separadas para los ojos izquierdo y derecho. A continuación, pueden presentarse al usuario por medios adecuados, tal como las pantallas individuales de los ojos izquierdo y derecho de un casco de VR. En otras realizaciones, una o más imágenes de vista pueden, por ejemplo, presentarse en una pantalla autoestereoscópica o, de hecho, en algunas realizaciones sólo puede generarse una única imagen bidimensional (por ejemplo, utilizando una pantalla bidimensional convencional).

De forma similar, para una pose dada de espectador/usuario/oyente, puede proporcionarse una representación de audio de la escena. La escena sonora se renderiza normalmente para proporcionar una experiencia espacial en la que se percibe que las fuentes de audio se originan en las posiciones deseadas. Como las fuentes de audio pueden estar estáticas en la escena, los cambios en la pose del usuario provocarán un cambio en la posición relativa de la fuente de audio con respecto a la pose del usuario. En consecuencia, la percepción espacial de la fuente de audio debe cambiar para reflejar la nueva posición relativa al usuario. La renderización de audio puede adaptarse en función de la postura del usuario.

La entrada de la pose del espectador o usuario puede determinarse de diferentes maneras en diferentes aplicaciones. En muchas realizaciones, el movimiento físico de un usuario puede rastrearse directamente. Por ejemplo, una cámara que vigila una zona del usuario puede detectar y seguir la cabeza del usuario (o incluso los ojos (seguimiento ocular)). En muchas realizaciones, el usuario puede llevar un casco de VR que puede ser rastreado por medios externos y/o internos. Por ejemplo, los auriculares pueden incluir acelerómetros y giroscopios que proporcionan información sobre el movimiento y la rotación de los auriculares y, por tanto, de la cabeza. En algunos ejemplos, el casco de realidad virtual puede transmitir señales o comprender identificadores (por ejemplo, visuales) que permiten a un sensor externo determinar la posición del casco de realidad virtual.

En algunos sistemas, la pose del espectador puede ser proporcionada por medios manuales, por ejemplo, por el usuario controlando manualmente una palanca de mando o una entrada manual similar. Por ejemplo, el usuario puede mover manualmente el visor virtual por la escena virtual controlando una primera palanca de mando analógica con una mano y controlar manualmente la dirección en la que mira el visor virtual moviendo manualmente una segunda palanca de mando analógica con la otra mano.

En algunas aplicaciones se puede utilizar una combinación de enfoques manuales y automatizados para generar la pose de entrada del espectador. Por ejemplo, un auricular puede seguir la orientación de la cabeza y el movimiento/posición del espectador en la escena puede ser controlado por el usuario mediante una palanca de mando.

En algunos sistemas, la aplicación de VR puede ser proporcionada localmente a un espectador por, por ejemplo, un dispositivo autónomo que no utiliza, o incluso no tiene ningún acceso a, ningún dato o procesamiento remoto de VR. Por ejemplo, un dispositivo tal como una consola de juegos puede comprender un almacén para guardar los datos de la escena, una entrada para recibir/generar la pose del espectador y un procesador para generar las imágenes correspondientes a partir de los datos de la escena.

En otros sistemas, la aplicación de VR puede implementarse y ejecutarse a distancia del espectador. Por ejemplo, un dispositivo local al usuario puede detectar/recibir datos de movimiento/postura que se transmiten a un dispositivo remoto que procesa los datos para generar la pose del espectador. A continuación, el dispositivo remoto puede generar imágenes de vista adecuadas para la pose del espectador basándose en los datos de la escena que la describen. A continuación, las imágenes vistas se transmiten al dispositivo local del espectador donde se presentan. Por ejemplo, el dispositivo remoto puede generar directamente un flujo de vídeo (normalmente un flujo de vídeo estéreo/3D) que es presentado directamente por el dispositivo local.

De forma similar, el dispositivo remoto puede generar una escena de audio que refleje el entorno de audio virtual. En muchas realizaciones, esto puede hacerse generando elementos de audio que correspondan a la posición relativa de

diferentes fuentes de audio en el entorno de audio virtual, y que se rendericen para percibirse en las posiciones correspondientes.

Por ejemplo, un dispositivo remoto puede generar datos de audio que representen una escena de audio y puede transmitir componentes/objetos/señales de audio u otros elementos de audio correspondientes a diferentes fuentes de audio en la escena de audio junto con información de posición indicativa de la posición de estos (que puede, por ejemplo, cambiar dinámicamente para objetos en movimiento). Los elementos de audio pueden incluir elementos asociados a posiciones específicas, pero también pueden incluir elementos para fuentes de audio más distribuidas o difusas. Por ejemplo, se pueden proporcionar elementos de audio que representen sonido de fondo genérico (no localizado), sonido ambiente, reverberación difusa, etc.

El dispositivo local de VR puede entonces renderizar los elementos de audio apropiadamente, por ejemplo, aplicando un procesamiento binaural apropiado que refleje la posición relativa de las fuentes de audio para los componentes de audio.

Para la parte de audio de un servicio de VR, un servidor central puede, en algunas realizaciones, generar datos de audio que representen una escena de audio, y puede representar específicamente esta escena de audio mediante una serie de elementos de audio que pueden ser renderizados por el cliente/dispositivo local.

La FIG. 1 ilustra un ejemplo de sistema de VR en el que un servidor 101 central está en contacto con varios clientes 103 remotos, por ejemplo, a través de una red 105, tal como Internet. El servidor 101 central puede estar preparado para atender simultáneamente a un número potencialmente elevado de clientes 103 remotos.

Un enfoque de este tipo puede proporcionar en muchos escenarios una compensación mejorada, por ejemplo, entre la complejidad y las demandas de recursos para diferentes dispositivos, requisitos de comunicación, etc. Por ejemplo, la pose del espectador y los datos de la escena correspondientes pueden transmitirse con intervalos mayores, mientras que el dispositivo local procesa localmente la pose del espectador y los datos de la escena recibidos para ofrecer una experiencia en tiempo real con poco retardo. Esto puede, por ejemplo, reducir sustancialmente el ancho de banda de comunicación requerido, proporcionando al mismo tiempo una experiencia de baja latencia y permitiendo que los datos de la escena se almacenen, generen y mantengan de forma centralizada. Por ejemplo, puede ser adecuado para aplicaciones en las que se proporciona una experiencia de VR a una pluralidad de dispositivos remotos.

La FIG. 2 ilustra elementos de un aparato de audio que proporciona una reproducción de audio mejorada en muchas aplicaciones y escenarios. En particular, el aparato de audio proporciona una renderización mejorada para muchas aplicaciones de VR, y el aparato de audio está específicamente dispuesto para realizar el procesamiento de audio y la renderización para un cliente 103 de VR de la FIG. 1.

El aparato de audio de la FIG. 2 está preparado para renderizar la escena de audio generando un conjunto híbrido de señales de salida con un primer (sub)conjunto de señales de salida generadas para ser renderizadas por un conjunto de altavoces y un segundo (sub)conjunto de señales de salida generadas para ser renderizadas por auriculares. El primer conjunto de señales de audio es específicamente un conjunto de señales de sonido envolvente para su renderización en un conjunto de altavoces de sonido envolvente. El segundo conjunto de señales de audio es específicamente una señal estéreo binaural para su reproducción en un par de auriculares.

El aparato de audio de la FIG. 2 puede formar parte de un sistema híbrido de reproducción de audio para VR/AR que utiliza una combinación de reproducción por auriculares y altavoces para ofrecer una presentación de una escena de audio.

Este enfoque puede proporcionar un funcionamiento ventajoso en muchas realizaciones. Por ejemplo, en muchos casos, el uso de una combinación de altavoces y auriculares en lugar de uno de los dos puede proporcionar una experiencia de AR (o VR/MR) que sea altamente envolvente para cada usuario y, al mismo tiempo, no obstaculice el aspecto "social" o "compartido" de la experiencia. Por ejemplo, puede permitir que el audio reproducido se adapte a usuarios individuales y al contexto actual del usuario. Por ejemplo, puede permitir que las posiciones de las fuentes de audio se adapten con precisión a los movimientos/rotaciones de la cabeza del usuario. Al mismo tiempo, puede reducir la complejidad requerida, por ejemplo, para el procesamiento binaural, ya que partes sustanciales de la escena de audio pueden representarse mediante un procesamiento de canales de audio/sonido envolvente de menor complejidad. Además, puede basarse, por ejemplo, en el uso de auriculares con baja atenuación del sonido externo, facilitando así, por ejemplo, la interacción directa entre usuarios en el mismo entorno/sala.
La siguiente descripción se centrará en las realizaciones en las que el sistema reproduce la escena de audio utilizando una combinación de altavoces envolventes (por ejemplo, un sistema 5.1 o 7.1) que es común a todos los usuarios locales, y auriculares individuales (abiertos o semiabiertos) para los usuarios individuales (donde "auriculares individuales" significa: auriculares que reproducen una señal que se ha generado o adaptada para el usuario que lleva esos auriculares).

El aparato se describirá específicamente con referencia a un caso de uso de un aspecto "social" o "compartido" de la aplicación VR/AR/MR con múltiples personas compartiendo una experiencia. Pueden estar en lugares diferentes, pero

lo más interesante para el ejemplo es que también pueden estar en el mismo lugar (por ejemplo, en la misma habitación). Un ejemplo de uso específico es el de varias personas en la misma habitación, compartiendo la misma experiencia de AR que se "proyecta" dentro de su entorno real compartido. Por ejemplo, una pareja sentada junta en un sofá, viendo una película inmersiva proyectada virtualmente en la pared de su salón. Pueden llevar gafas transparentes que les permitan verse entre sí y a su entorno, así como auriculares abiertos que permitan tanto una reproducción personalizada dedicada como escuchar el audio del entorno, incluyendo el generado por una configuración de sonido envolvente.

El aparato de la FIG. 2 incluye específicamente un receptor 201 que está dispuesto para recibir datos que describen la escena virtual. Los datos pueden incluir datos que proporcionen una descripción visual de la escena e incluir datos que proporcionen una descripción sonora de la escena. De este modo, los datos recibidos pueden proporcionar una descripción sonora de la escena y una descripción visual de la misma.

El receptor 201 está acoplado a un renderizador 203 visual que procede a renderizar imágenes correspondientes a la pose de visualización actual de un espectador. Por ejemplo, los datos pueden incluir datos de imágenes 3D espaciales (por ejemplo, imágenes y profundidad o una descripción de modelo de la escena) y, a partir de ellos, el renderizador 203 visual puede generar imágenes estereoscópicas (imagen para los ojos izquierdo y derecho de un usuario), como sabrá el experto. Las imágenes pueden presentarse al usuario, por ejemplo, a través de las pantallas individuales de los ojos izquierdo y derecho de un casco de VR.

Los datos recibidos comprenden datos de audio que describen la escena. Los datos de audio comprenden específicamente datos de audio para un conjunto de elementos de audio correspondientes a fuentes de audio en la escena. Algunos elementos de audio pueden representar fuentes de audio localizadas en la escena que están asociadas con una posición específica en la escena (la posición puede, por supuesto, cambiar dinámicamente para un objeto en movimiento). A menudo, un elemento de audio puede representar audio generado por un objeto de escena específico en la escena virtual y, por lo tanto, puede representar una fuente de audio en una posición correspondiente a la del objeto de escena (por ejemplo, un ser humano hablando).

Otros elementos pueden representar fuentes de audio más distribuidas o difusas, tales como por ejemplo ruido ambiental o de fondo que puede ser difuso. Como otro ejemplo, algunos elementos de audio pueden representar total o parcialmente componentes de audio no localizados espacialmente procedentes de fuentes de audio localizadas, tal como por ejemplo una reverberación difusa procedente de una fuente de audio espacialmente bien definida.

Los elementos de audio pueden ser datos de audio codificados, tales como señales de audio codificadas. Los elementos de audio pueden ser de diferentes tipos, incluyendo diferentes tipos de señales y componentes, y de hecho en muchas realizaciones el primer receptor 201 puede recibir datos de audio que definen diferentes tipos/formatos de audio. Por ejemplo, los datos de audio pueden incluir audio representado por señales de canal de audio, objetos de audio individuales, Ambisonics de Orden Superior (HOA), etc.

El audio puede, por ejemplo, representarse como audio codificado para un componente de audio dado que va a renderizarse. Los datos de audio pueden incluir además datos de posición que indiquen una posición de la fuente del componente de audio. Los datos de posición pueden incluir, por ejemplo, datos de posición absoluta que definen una posición de la fuente de audio en la escena.

El aparato comprende además dos renderizadores 205, 207.

Un primer renderizador 205 está dispuesto para renderizar elementos de audio sobre un conjunto de altavoces. Específicamente, el primer renderizador 205 genera un primer conjunto de señales de audio para un conjunto de altavoces donde el primer conjunto de señales de audio es, por ejemplo, un conjunto de señales de sonido envolvente para una configuración de altavoces de sonido envolvente.

El primer renderizador 205 genera así un conjunto de señales de audio destinadas que van a ser reproducidas por una configuración específica de altavoces espaciales. El primer renderizador 205 genera una señal para cada altavoz de una configuración de sonido envolvente, y por tanto para renderizar desde una ubicación específica correspondiente a la posición del altavoz en la configuración.

El primer renderizador 205 puede estar dispuesto para generar las señales de audio de tal manera que un elemento de audio dado sea renderizado de tal manera que el efecto combinado conduzca a una impresión de que el elemento de audio está siendo renderizado desde la posición deseada. Típicamente, los datos recibidos pueden para al menos algunos elementos de audio incluir indicaciones de posición específicas y el primer renderizador 205 puede renderizar los elementos de audio de tal manera que se perciban originados desde la posición indicada. Otros elementos de audio pueden, por ejemplo, distribuirse y difundirse, y pueden renderizarse como tales.

Se apreciará que muchos algoritmos y enfoques para la renderización de audio espacial utilizando altavoces, y específicamente en sistemas de sonido envolvente, serán conocidos por el experto y que cualquier enfoque adecuado puede utilizarse sin desvirtuar la invención.

Por ejemplo, el primer renderizador 205 puede generar señales de audio para cinco altavoces en una configuración de sonido envolvente con un altavoz central, un altavoz frontal izquierdo, un altavoz frontal derecho, un altavoz envolvente izquierdo y un altavoz envolvente derecho. El primer renderizador 205 puede generar un conjunto de señales de audio que comprenda una señal de audio para cada altavoz. A continuación, las señales pueden amplificarse para generar señales de accionamiento para cada altavoz.

En algunas realizaciones, un elemento de audio que está siendo renderizado usando los altavoces puede ser recibido como, por ejemplo, un mezcla estéreo descendente y el primer renderizador 205 puede realizar una mezcla ascendente para generar las señales envolventes que en algunos casos pueden renderizarse directamente. Este enfoque puede ser útil, por ejemplo, para elementos de audio que representen un sonido difuso que no esté directamente relacionado con la postura del usuario. Por ejemplo, un elemento de audio que represente audio ambiental difuso genérico puede proporcionarse como una mezcla estéreo descendente que puede mezclarse ascendentemente directamente para proporcionar los canales de audio de sonido envolvente adecuados. Cada una de las señales mezcladas ascendentemente resultantes puede combinarse con señales para los altavoces correspondientes generadas a partir de otros elementos de audio para generar el conjunto de señales de salida.

Algunos elementos de audio que son renderizados a través de la configuración del altavoz pueden, por ejemplo, proporcionarse en forma de objetos de audio. Dicho objeto de audio puede estar representado por datos de audio que describen el audio específico y datos de posición asociados que describen la posición de la fuente de audio. Basándose en los datos de posición y las posiciones de los altavoces (ya sean posiciones reales o posiciones nominales para la configuración de altavoces de sonido envolvente), el primer renderizador 205 puede determinar coeficientes para una matriz o vector que mapea la señal de audio a los diferentes canales de sonido envolvente.

En algunas realizaciones, el primer renderizador 205 puede además estar dispuesto para adaptar las señales de audio generadas basándose en datos del entorno acústico. Por ejemplo, si se proporcionan datos que indican que el entorno actual es un entorno altamente reflectante (por ejemplo, un cuarto de baño o un entorno acústico similar con un alto grado de reflexiones), entonces el primer renderizador 205 puede generar y aplicar un filtro que tenga una respuesta al impulso correspondiente a la función de transferencia de la sala para el entorno (primeras reflexiones, etc.). En algunas realizaciones, el filtro puede aplicarse a cada una de las señales de audio generadas para los canales de sonido envolvente individuales o, en algunas realizaciones, puede aplicarse al elemento de audio antes de la mezcla ascendente a los diferentes canales de audio.

En algunas realizaciones, el primer renderizador 205 puede alternativamente o adicionalmente estar dispuesto para añadir reverberación que específicamente puede estar basada en datos de entorno recibidos con el elemento de audio. Por ejemplo, el primer renderizador 205 puede aplicar un reverberador sintético, como un reverberador Jot, con parámetros que se establecen en función de los datos del entorno acústico (por ejemplo, con un sostenido de reverberación según lo indicado por los datos). El reverberador suele aplicarse al elemento de audio antes de cualquier mezcla ascendente o asignación a los canales de sonido envolvente. El segundo renderizador 207 está dispuesto para generar un segundo conjunto de señales de audio para un auricular. El segundo conjunto de señales de audio puede ser específicamente una señal estéreo binaural.

En muchas realizaciones, el renderizado de audio por el segundo renderizador 207 es un proceso de renderización binaural que utiliza funciones de transferencia binaural adecuadas para proporcionar el efecto espacial deseado para un usuario que lleva auriculares. Por ejemplo, el segundo renderizador 207 está preparado para generar un componente de audio que se perciba como procedente de una posición específica utilizando procesamiento binaural.

Se sabe que el procesamiento binaural se utiliza para proporcionar una experiencia espacial mediante el posicionamiento virtual de las fuentes de sonido utilizando señales individuales para los oídos del oyente. Con un procesamiento binaural adecuado, se pueden calcular las señales necesarias en los tímpanos para que el oyente perciba el sonido desde cualquier dirección deseada, y las señales se pueden renderizar de forma que proporcionen el efecto deseado. A continuación, estas señales se recrean en el tímpano mediante auriculares o un método de cancelación de la diafonía (adecuado para la reproducción a través de altavoces muy próximos entre sí). La reproducción binaural puede considerarse un método de generación de señales para los oídos de un oyente que induce al sistema auditivo humano a pensar que un sonido procede de las posiciones deseadas.
La renderización binaural se basa en funciones de transferencia binaurales que varían de una persona a otra debido a las propiedades acústicas de la cabeza, los oídos y las superficies reflectantes, tal como los hombros. Por ejemplo, se pueden utilizar filtros binaurales para crear una grabación binaural que simule múltiples fuentes en diveras ubicaciones. Esto puede realizarse convolviendo cada fuente de sonido con el par de, por ejemplo, respuestas al impulso relacionadas con la cabeza (HRIR) que corresponden a la posición de la fuente de sonido.

Un método bien conocido para determinar las funciones de transferencia binaural es la grabación binaural. Es un método de grabación de sonido que utiliza un micrófono específico y está pensado para reproducirse con auriculares. La grabación se realiza colocando micrófonos en el conducto auditivo de un sujeto o utilizando una cabeza ficticia con micrófonos incorporados, un busto que incluye pabellones auriculares (orejas externas). El uso de este tipo de cabeza

ficticia con pabellones auriculares proporciona una impresión espacial muy similar a la que se obtendría si la persona que escucha las grabaciones estuviera presente durante la grabación.

Midiendo, por ejemplo, las respuestas de una fuente de sonido en una ubicación específica en el espacio 2D o 3D a micrófonos colocados en o cerca de los oídos humanos, se pueden determinar los filtros binaurales apropiados. A partir de esas mediciones, pueden generarse filtros binaurales que reflejen las funciones de transferencia acústica a los oídos del usuario. Los filtros binaurales pueden utilizarse para crear una grabación binaural que simule múltiples fuentes en diversas ubicaciones. Esto puede realizarse, por ejemplo, convolviendo cada fuente sonora con el par de respuestas al impulso medidas para una posición deseada de la fuente sonora. Para crear la ilusión de que una fuente sonora se mueve alrededor del oyente, suele ser necesario un gran número de filtros binaurales con una resolución espacial adecuada, por ejemplo, 10 grados.

Las funciones de transferencia binaural relacionadas con la cabeza pueden ser representadas, por ejemplo, como Respuestas al Impulso Relacionadas con la Cabeza (HRIR), o equivalentemente como Funciones de Transferencia Relacionadas con la Cabeza (HRTFs) o, Respuestas Binaurales al Impulso de la Habitación (BRIRs), o Funciones Binaurales de Transferencia de la Habitación (BRTFs). La función de transferencia (por ejemplo, estimada o supuesta) desde una posición determinada a los oídos (o tímpanos) del oyente puede darse, por ejemplo, en el dominio de la frecuencia, en cuyo caso suele denominarse HRTF o BRTF, o en el dominio del tiempo, en cuyo caso suele denominarse HRIR o BRIR. En algunos escenarios, las funciones de transferencia binaural relacionadas con la cabeza se determinan para incluir aspectos o propiedades del entorno acústico y, específicamente, de la sala en la que se realizan las mediciones, mientras que en otros ejemplos sólo se tienen en cuenta las características del usuario. Ejemplos del primer tipo de funciones son las BRIR y las BRTF.

El segundo renderizador 207 puede, en consecuencia, comprender un almacén con funciones de transferencia binaural para un número, típicamente alto, de posiciones diferentes con cada función de transferencia binaural proporcionando información de cómo una señal de audio debe procesarse/filtrada para percibirse como originada desde esa posición. La aplicación individual de procesamiento binaural a una pluralidad de señales/fuentes de audio y la combinación del resultado puede utilizarse para generar una escena de audio con una serie de fuentes de audio situadas en posiciones adecuadas en la etapa sonora.

El segundo renderizador 207 puede para un elemento de audio dado que debe percibirse como originado desde una posición dada relativa a la cabeza del usuario, seleccionar y recuperar la función de transferencia binaural almacenada que más se aproxime a la posición deseada (o en algún caso puede generar esto interpolando entre una pluralidad de funciones de transferencia binaurales cercanas). A continuación, puede aplicar la función de transferencia binaural seleccionada a la señal de audio del elemento de audio, generando así una señal de audio para el oído izquierdo y una señal de audio para el oído derecho.

La señal estéreo de salida generada en forma de señal de oído izquierdo y derecho es entonces adecuada para la renderización de auriculares y puede amplificarse para generar señales de accionamiento que se alimentan a los auriculares de un usuario. El usuario percibirá entonces que el elemento de audio procede de la posición deseada.

Se apreciará que, en algunas realizaciones, el elemento de audio también puede procesarse para, por ejemplo, añadir efectos de entorno acústico. Por ejemplo, como se ha descrito para el primer renderizador 205, el elemento de audio puede procesarse para añadir reverberación o, por ejemplo, decorrelación/difusión. En muchas realizaciones, este procesamiento puede llevarse a cabo en la señal binaural generada en lugar de directamente en la señal del elemento de audio.

Así, el segundo renderizador 207 está dispuesto para generar las señales de audio de tal manera que un elemento de audio dado se renderiza de tal manera que un usuario que lleva los auriculares percibe que el elemento de audio se recibe desde la posición deseada. Típicamente, el segundo renderizador 207 puede renderizar elementos de audio de tal manera que se perciban originados desde la posición indicada en los datos posicionales incluidos con los datos de audio. Otros elementos de audio pueden, por ejemplo, distribuirse y difundirse, y pueden renderizarse como tales.

En consecuencia, el aparato puede formar parte de un cliente 103 que recibe datos que incluyen datos de audio que describen una escena de audio desde un servidor 101 central. En muchas aplicaciones, el servidor 101 central puede proporcionar una serie de elementos de audio en forma de objetos de audio, canales de audio, componentes de audio, HOA, señales de audio, etc. En muchas situaciones, algunos de los elementos de audio pueden corresponder a una única fuente de audio con una posición específica. Otros elementos de audio pueden corresponder a fuentes de audio más difusas, menos definidas y más distribuidas.

Se apreciará que muchos algoritmos y enfoques para la renderización de audio espacial utilizando auriculares, y específicamente para la renderización binaural, serán conocidos por el experto y que cualquier enfoque adecuado puede utilizarse sin desvirtuar la invención.

El aparato de la FIG. 2 puede utilizarse en un cliente 103 para procesar los datos de audio recibidos y renderizar la escena de audio deseada. En concreto, puede procesar cada elemento de audio en función de los datos de posición deseados (cuando proceda) y, a continuación, combinar los resultados.

El aparato de la FIG. 2 utiliza en consecuencia dos técnicas de renderización diferentes para generar el audio que representa la escena. Las diferentes técnicas de renderización pueden tener diferentes propiedades y el aparato de la FIG. 2 comprende un selector 209 dispuesto para seleccionar qué elementos de audio son renderizados por el primer renderizador 205 y qué elementos de audio son renderizados por el segundo renderizador 207. Específicamente, para un primer elemento de audio dado, el selector 211 selecciona qué renderizador 205, 207 debe utilizarse para la renderización. En consecuencia, el selector 209 recibe el primer elemento de audio y lo envía al primer renderizador 205 o al segundo renderizador 207 en función de la selección.

En el sistema, el receptor 201, además de los datos de audio (y posiblemente los datos visuales), está preparado para recibir metadatos que comprenden indicadores de propiedades de renderización de audio para al menos uno de los elementos de audio y, a menudo, para la mayoría o incluso la totalidad del elemento de audio. Específicamente, se incluye al menos un primer indicador de propiedad de renderización de audio para el primer elemento de audio.

El selector 209 está dispuesto para seleccionar que renderizador usar dependiendo de los metadatos recibidos y los indicadores de propiedades de renderización de audio. Específicamente, el selector 209 está dispuesto para considerar el primer indicador de propiedad de renderización de audio y decidir si el primer elemento de audio debe ser renderizado por el primer renderizador 205 o por el segundo renderizador 207, es decir, si debe renderizarse utilizando los altavoces o los auriculares.

Como un ejemplo de baja complejidad, los datos pueden para cada elemento de audio incluir datos de audio codificados así como metadatos que comprenden una indicación de posición (típicamente la posición de la fuente de audio correspondiente al elemento de audio) y un indicador de propiedad de renderización de audio para el elemento de audio, donde el indicador de propiedad de renderización de audio en el ejemplo específico simplemente puede ser una indicación binaria de si el elemento de audio debería ser renderizado por el primer renderizador 205 o por el segundo renderizador 207. El selector 209 puede entonces evaluar esta indicación binaria y seleccionar el renderizador 205, 207 indicado. El renderizador 205, 207 puede entonces generar las señales de salida apropiadas para los altavoces y auriculares respectivamente, de forma que se perciba que los elementos de audio llegan desde una posición como la indicada por el indicador de posición. La contribución de cada elemento de audio para el cual la indicación es que deben renderizarse usando el primer renderizador 205 puede entonces combinarse para generar un primer conjunto de señales de audio para los altavoces y la contribución de cada elemento de audio para el cual la indicación es que deben renderizarse usando el segundo renderizador 207 puede entonces combinarse para generar un segundo conjunto de señales de audio para los auriculares.

De este modo, el aparato de audio de la FIG. 2 puede renderizar la escena de audio a través de un sistema híbrido de renderización de audio que incluya tanto altavoces como auriculares. Además, la distribución de los elementos de audio a través de los auriculares y altavoces puede controlarse/guiarse a distancia. Por ejemplo, el proveedor de la experiencia de VR también puede controlar y decidir cómo deben renderizarse los elementos de audio. Dado que el proveedor puede disponer normalmente de información adicional sobre la naturaleza específica de la fuente de audio para cada elemento de audio, esto puede permitir que la selección de cómo renderizar cada elemento de audio se controle basándose en información y conocimientos adicionales que pueden no estar disponibles en el cliente. El enfoque puede proporcionar un mejor renderizado en muchas situaciones y puede proporcionar una experiencia de usuario mejorada en muchos escenarios. Este enfoque puede, por ejemplo, proporcionar una renderización precisa y natural de la escena sonora y permitir, por ejemplo, que las personas que se encuentran en la misma habitación hablen entre sí con mayor naturalidad.

Así, en muchas realizaciones, el indicador de propiedad de renderización de audio proporciona una guía al cliente y al aparato de audio sobre cómo deberían renderizarse los datos de audio recibidos. El indicador de propiedad de renderización de audio puede ser indicativo de una propiedad de renderización de guía para el renderizado del primer elemento de audio. En muchas realizaciones, la propiedad de renderización de guía puede ser una propiedad de renderización preferida, sugerida o nominal que se recomienda que utilice el renderizador local. Así, la propiedad de renderización de orientación puede ser un dato de control que el cliente puede utilizar para establecer un parámetro de renderización.

En algunas realizaciones, la propiedad de renderización guía puede ser una propiedad de renderización obligatoria que debe ser usada cuando se renderiza el elemento de audio, pero en otras realizaciones la propiedad de renderización guía puede ser una propiedad sugerida que puede ser usada o no por el cliente. Así, en muchas realizaciones, el aparato de audio puede elegir entre adaptar su renderizado para que coincida con la propiedad de renderización de guiado o puede elegir emplear un valor diferente. Sin embargo, el planteamiento proporciona un enfoque que permite al aparato de audio adaptar su funcionamiento bajo la guía del servidor/proveedor remoto. Esto puede mejorar el rendimiento en muchos casos, ya que el servidor/proveedor remoto puede disponer de información adicional. Por ejemplo, también puede permitir una optimización o análisis manual centralizado para mejorar potencialmente el renderizado, al tiempo que permite al cliente conservar la libertad y flexibilidad en el renderizado.

En el ejemplo específico mencionado anteriormente, el indicador de propiedad de reproducción de audio es indicativo de si el primer elemento de audio está destinado a reproducirse a través de altavoces o si está destinado a reproducirse a través de auriculares. Para un primer elemento de audio, el selector 209 puede estar dispuesto para seleccionar el primer renderizador 205 para la renderización si un primer indicador de renderización para el primer elemento de audio es indicativo de que el primer elemento de audio está destinado a ser renderizado por altavoces y para seleccionar el segundo renderizador 207 para la renderización del primer elemento de audio si el primer indicador de renderización es indicativo de que el primer elemento de audio está destinado a ser renderizado por auriculares. El selector 209 lo proporciona entonces al renderizador 205, 207 seleccionado para su renderización.

Así, en muchas realizaciones, el indicador de propiedad de renderización de audio es indicativo de una propiedad del renderizado que se va a aplicar al primer elemento de audio, y específicamente el indicador de renderización para un elemento de audio puede ser indicativo de si el elemento de audio está destinado a ser renderizado por altavoces o por auriculares.

En algunas realizaciones, puede señalarse explícitamente mediante metadatos en el flujo de contenido si un elemento de audio debe reproducirse a través de los altavoces o a través de los auriculares en el caso de que se utilice un sistema de reproducción híbrido. Esto puede ser una elección artística explícita realizada por el productor de contenidos y, por lo tanto, puede proporcionar un mejor control/guía para el renderizado.

En el aparato de la FIG. 2, el renderizado de audio depende (al igual que el renderizado visual) de la pose del espectador. Específicamente, el aparato comprende un receptor 211 de pose de oyente que está dispuesto para recibir una pose del oyente indicativa de una pose del oyente. La pose del oyente está representada específicamente por la pose del casco, por ejemplo, determinada por el seguimiento del casco de VR que lleva el usuario/oyente. Se apreciará que cualquier método adecuado para generar, estimar, recibir y proporcionar una pose de oyente puede utilizarse sin menoscabo de la invención.

El receptor 211 de pose de oyente se conecta al renderizador 203 visual y se utiliza para generar la salida visual correspondiente a la pose específica. Además, el receptor 211 de pose de oyente está acoplado al segundo renderizador 207 y se utiliza en el renderizado de los elementos de audio para el auricular. Así, el segundo renderizador 207 está preparado para generar el segundo conjunto de señales de audio en respuesta a la pose del oyente.

El segundo renderizador 207 puede realizar específicamente una renderización binaural de tal manera que los elementos de audio sean renderizados para percibirse como originados en las posiciones apropiadas con respecto a la orientación y posición actual del oyente. Por ejemplo, para el primer elemento de audio, el segundo renderizador 207 puede determinar primero la posición en el espacio de escena indicada por la indicación de posición recibida para el primer elemento de audio en el flujo de datos. La posición relativa del primer elemento de audio con respecto al usuario puede determinarse entonces analizando la pose actual del oyente y la pose correspondiente en el espacio de la escena. El segundo renderizador 207 puede entonces recuperar HRTFs correspondientes a esta posición relativa y filtrar la primera señal de audio usando los HRTFs recuperados para generar componentes de señal estéreo binaural para el primer elemento de audio. A continuación, los componentes pueden añadirse a los componentes correspondientes generados a partir de otros elementos de audio para generar señales estéreo binaurales de salida.

Se apreciará que se conocen muchos enfoques diferentes para generar señales de auriculares (y específicamente señales binaurales) correspondientes a fuentes de audio en posiciones espaciales y que el segundo renderizador 207 puede utilizar cualquier enfoque o algoritmo adecuado.

En contraste con el segundo renderizador 207, el renderizado por el primer renderizador 205 (es decir, el renderizado para los altavoces) no depende de la pose del oyente y por lo tanto el primer renderizador 205 es en el ejemplo de la FIG. 2 dispuesto para generar el primer conjunto de señales de audio independientemente de la postura del oyente.

El primer renderizador 205 considera específicamente la indicación de posición para un elemento de audio que va a ser renderizado por el primer renderizador 205 y mapea esto a una posición en el espacio de renderización de los altavoces. A continuación, el primer renderizador 205 genera las señales para que los altavoces proporcionen una percepción espacial del elemento de audio correspondiente a la posición determinada.
Se apreciará que se conocen muchos enfoques diferentes para generar señales de altavoces (y específicamente señales de sonido envolvente) correspondientes a fuentes de audio en posiciones espaciales y que cualquier enfoque o algoritmo adecuado puede ser utilizado por el primer renderizador 205.

Así, en el ejemplo, las señales de los auriculares se generan continuamente para reflejar el movimiento y las rotaciones de la cabeza del oyente, proporcionando así una experiencia de usuario continua y consistente. Al mismo tiempo, la renderización mediante los altavoces no varía con respecto a los movimientos y la rotación de la cabeza del oyente, lo que también proporciona un enfoque coherente. El enfoque puede proporcionar un enfoque en el que los diferentes enfoques de renderización proporcionan una representación coherente de la escena de audio con respecto a un oyente no estático.

Los ejemplos anteriores se han centrado en una situación en la que el aparato genera una representación de la escena de audio para un único usuario. Sin embargo, en muchas realizaciones, el aparato puede generar una representación de la escena de audio para una pluralidad de usuarios, tal como específicamente para dos o más usuarios ubicados en la misma habitación.

En tal caso, el primer renderizador 205 puede estar dispuesto para generar un conjunto común de señales de audio para la pluralidad de usuarios mientras que el segundo renderizador 207 está dispuesto para generar señales de auriculares individuales para cada usuario.

Así, para los elementos de audio que son seleccionados para ser renderizados por el primer renderizador 205, solo un único conjunto de señales de salida es generado para todos los usuarios, por ejemplo, solo una única señal de altavoz es generada para cada altavoz en la configuración y estas pueden típicamente no depender de ninguna propiedad específica del usuario. En concreto, el primer conjunto de señales de audio generadas para ser reproducidas por los altavoces se genera sin tener en cuenta ninguna postura del oyente. Se genera la misma renderización de la escena de audio para todos los usuarios.

Sin embargo, para elementos de audio que son renderizados por el segundo renderizador 207, un conjunto diferente de señales de audio puede generarse para cada usuario. En concreto, se puede generar una señal estéreo binaural para cada usuario. Estas señales individuales pueden generarse para reflejar propiedades o características específicas para el oyente individual y pueden generarse específicamente para reflejar la pose de oyente del oyente individual. Así, pueden generarse señales binaurales que reflejen la posición y orientación actuales de los usuarios.

El aparato puede así, en particular, proporcionar un soporte muy eficiente para escenarios multiusuario. El procesamiento de audio necesario para dar soporte a múltiples oyentes puede reducirse sustancialmente. Por ejemplo, el procesamiento binaural suele ser relativamente complejo y consumir muchos recursos, y el número de señales de audio que deben generarse utilizando el procesamiento binaural puede reducirse sustancialmente, reduciendo así la complejidad y la carga computacional en muchas realizaciones.

Así, en un ejemplo en el que el aparato soporta dos usuarios en la misma habitación, el primer renderizador 205 está dispuesto para generar un primer conjunto común de señales de audio para renderizar usando altavoces y el segundo renderizador 207 está dispuesto para generar un segundo conjunto de señales de audio para auriculares para un primer oyente y para generar un tercer conjunto de señales de audio para auriculares para un segundo oyente. El primer conjunto de señales de audio se genera independientemente de la postura del primer y segundo oyente, y el segundo conjunto de señales de audio se genera en respuesta a la postura del primer oyente y el tercer conjunto de señales de audio se genera en respuesta a la postura del segundo oyente.

El indicador de propiedad de renderización de audio proporcionado en el flujo de datos recibido puede, en diferentes realizaciones, representar diferentes datos.

El indicador de propiedad de renderización de audio es indicativo de si la primera parte del primer elemento de audio está asociada con una posición dependiente de la pose del oyente o con una posición no dependiente de la pose del oyente. El indicador de la propiedad de reproducción de audio puede indicar específicamente si el primer elemento de audio es diegético o no.

Como ejemplo específico, en algunas realizaciones, el selector 209 está dispuesto para distribuir los elementos de audio entre el primer renderizador 205 y el segundo renderizador 207 basándose en si un indicador de propiedad de renderización de audio para el elemento de audio indica que está "fijado a la orientación de la cabeza" o "no fijado a la orientación de la cabeza" de acuerdo con la terminología MPEG.

Un elemento de audio indicado por el indicador de propiedad de renderización de audio como "fijo a la cabeza" es un elemento de audio que está destinado a tener una ubicación fija relativa a la cabeza del usuario. Tales elementos de audio pueden renderizarse usando el segundo renderizador 207 y pueden renderizarse independientemente de la pose del oyente. Por lo tanto, la renderización de tales elementos de audio no tiene en cuenta (los cambios en) la orientación de la cabeza del usuario, en otras palabras, tales elementos de audio son elementos de audio para los que la posición relativa no cambia cuando el usuario gira la cabeza (por ejemplo, audio no espacial tal como el ruido ambiente o, por ejemplo, música que está destinada a seguir al usuario sin cambiar una posición relativa).

Un elemento de audio indicado por el indicador de propiedad de renderización de audio como "No-fijo a la cabeza" es un elemento de audio que está destinado a tener una localización fija en el entorno (virtual o real), y por tanto su renderizado se adapta dinámicamente a (cambios en) la orientación de la cabeza del usuario. En muchas realizaciones, esto puede ser más realista cuando dicho elemento de audio se renderiza como una señal binaural de auriculares que se adapta con base en la postura actual del oyente. Por ejemplo, la percepción de la posición de una fuente de audio renderizada por una configuración de altavoces de sonido envolvente puede depender de la posición y orientación del usuario y, por lo tanto, la renderización de un elemento de audio indicado como "No fijo a la cabeza" por dicha configuración de altavoces puede dar lugar a una fuente de audio que se percibe como en movimiento cuando el usuario mueve la cabeza.

Así, en algunas realizaciones, los elementos "no-fijados a la orientación de la cabeza" son renderizados sobre los auriculares de los usuarios, con sus posiciones adaptadas para cada usuario individual de acuerdo con la orientación de la cabeza rastreada de ese usuario. En cambio, los elementos "fijados a la orientación de la cabeza" se renderizan sobre los altavoces y no se adaptan a los movimientos de cabeza de los usuarios.

La ventaja de una realización de este tipo es que los elementos "fijados a la orientación de la cabeza" que ahora están presentes principalmente a través de los altavoces (y no a través de los auriculares) son los principales responsables del aislamiento acústico que se experimenta cuando todos los elementos se renderizan a través de los auriculares. El razonamiento aquí es que los sonidos "fijados a la orientación de la cabeza" (sobre todo música y sonidos atmosféricos como, por ejemplo, multitudes, viento, lluvia, truenos, etc.) suelen ser continuos y espacialmente omnipresentes por naturaleza, lo que resulta en una "manta" de sonido que aísla al usuario de su entorno físico. En cambio, los elementos "no fijados a la orientación de la cabeza" suelen estar más localizados y ser más dispersos en el espacio y el tiempo, por lo que "enmascaran" mucho menos el entorno acústico físico del usuario.

En algunas realizaciones prácticas, la percepción por parte del usuario de los sonidos "fijados a la orientación de la cabeza" que se renderizan a través de los altavoces puede ser algo diferente de cómo se perciben típicamente cuando se reproducen a través de auriculares. Sin embargo, esto no suele ser un problema, ya que los sonidos "fijados a la orientación de la cabeza" que emiten los altavoces no suelen ser direccionales ni críticos en términos de localización espacial.

Qué elementos de audio están "no fijados a la orientación de la cabeza" y cuáles están "fijados a la orientación de la cabeza" puede señalarse explícitamente mediante metadatos en el flujo de contenido de audio.

En el contexto de la reproducción de audio de AR (y VR), el término "diegético" también se utiliza comúnmente para describir si un elemento de audio debe ser "fijado a la orientación de la cabeza" o no. "Diegético" describe elementos que deben permanecer en la misma posición virtual cuando un usuario mueve la cabeza (lo que significa que la posición renderizada relativa a la cabeza del usuario debe modificarse). "No diegético" describe elementos para los que esto no es importante, o incluso puede ser preferible que sus posiciones no tengan en cuenta los movimientos de la cabeza del usuario (lo que significa que se moverán con la cabeza del usuario o estarán "pegados" a ella).

En algunas realizaciones, el indicador de propiedad de renderización de audio para un elemento de audio puede ser indicativo de un formato de audio del elemento de audio. El selector 209 puede estar dispuesto para seleccionar si se utiliza el primer renderizador 205 o el segundo renderizador 207 para renderizar un elemento de audio basándose en el formato de audio del elemento de audio. El indicador de la propiedad de renderización de audio puede, por ejemplo, indicar que el elemento de audio es un formato de audio del grupo de: un formato de objeto de audio; un formato de audio Ambisonics de Orden Superior; y un formato de audio de señal de canal de audio.

En algunas realizaciones, el selector 209 puede estar preparado para distinguir entre elementos que van a ser reproducidos por los auriculares o los altavoces basándose en el formato de los elementos de audio.

Por ejemplo: los elementos basados en canales o Ambisonics de Orden Superior (HOA), que a menudo se utilizan para transmitir sonidos de fondo como música y sonidos atmosféricos, pueden ser renderizados a través de los altavoces, mientras que los elementos objeto, que se utilizan típicamente para transmitir los principales elementos de audio de una escena (a menudo representando fuentes de audio con posiciones bien definidas), pueden ser renderizados a través de auriculares para cada usuario individualmente. Esto también permite al usuario no sólo cambiar la orientación de su cabeza, sino también interactuar con los objetos de audio individuales (si el productor de contenidos ha previsto que los objetos sean interactivos).

Esta realización puede verse como una alternativa o adición a la provisión de indicadores de propiedades de renderización de audio que definen directamente que renderizador debe usarse. Por ejemplo, en situaciones en las que no se incluye una señalización explícita de si un elemento de audio es un elemento "no fijado a la orientación de la cabeza"/"fijado a la orientación de la cabeza", el selector 209 puede evaluar el formato de audio para determinar qué renderizador 205, 207 debe utilizarse.

Los enfoques y diferentes indicadores de propiedades de renderización de audio pueden ser combinados, por ejemplo, canal-, HOA-, y elementos que son explícitamente señalados como "fijos a la orientación de la cabeza" son renderizados sobre los altavoces, mientras que los objetos y elementos "no-fijos a la orientación de la cabeza" son renderizados sobre los auriculares.

En algunas realizaciones, el indicador de propiedad de renderización de audio puede ser indicativo de un tipo de fuente de audio para el primer elemento de audio. Por ejemplo, el indicador de propiedad de renderización de audio puede indicar si el elemento de audio es un tipo de fuente de audio de un conjunto que incluye, por ejemplo, uno o más de los siguientes: audio de voz; audio de música; audio de primer plano; audio de fondo; voz sobre audio; y audio de narrador.

En algunas realizaciones, la distribución de elementos de audio sobre altavoces y auriculares puede basarse en indicaciones en el flujo de contenido de tipos de fuente para los elementos de audio, por ejemplo, metadatos como "habla" o "música" o "sonidos de primer plano" o "sonidos de fondo". En este ejemplo, las fuentes de "habla" podrían reproducirse a través de los auriculares, mientras que las fuentes de "música" y "fondo" podrían reproducirse a través de los altavoces. Un caso especial podría ser el habla marcada como "voz en apagado" o "narrador", que podría renderizarse mejor a través de los altavoces (ya que no se pretende que tenga una ubicación específica en el espacio, sino que sea "omnipresente").

En algunas realizaciones, el receptor 201 puede, como se ha descrito anteriormente, recibir también datos visuales indicativos de una escena virtual correspondiente a la escena de audio. Estos datos pueden enviarse al renderizador 203 visual para que los renderice utilizando una técnica de renderización adecuada, por ejemplo, generando imágenes estereoscópicas correspondientes a la pose actual del usuario.

En algunas realizaciones, el indicador de propiedad de renderización de audio para un elemento de audio puede ser indicativo de si el primer elemento de audio representa una fuente de audio correspondiente a un objeto de escena visual. El objeto de la escena visual puede ser un objeto para el que los datos visuales comprenden una representación visual.

En un ejemplo en el que los datos visuales proporcionan datos visuales para una ventana gráfica, el indicador de propiedad de renderización de audio puede indicar si el elemento de audio está enlazado a un objeto dentro de la ventana gráfica.

Si el indicador de propiedad de renderización de audio indica que el objeto correspondiente al elemento de audio es visible en la escena, el selector 209 puede decidir renderizarlo usando auriculares y en caso contrario puede renderizar el elemento de audio usando altavoces. En algunas realizaciones, el indicador de la propiedad de renderización de audio puede indicar directamente si el objeto es visible. Sin embargo, en otras realizaciones, el indicador de propiedad de representación de audio puede proporcionar una indicación indirecta de si el elemento de audio corresponde a un objeto de escena visible.

Por ejemplo, el indicador de propiedad de renderización de audio puede comprender una indicación de un objeto de escena que es representado por los datos visuales recibidos. El selector 209 puede entonces proceder a evaluar si el objeto vinculado al elemento de audio es visible para la pose actual del oyente. En caso afirmativo, puede proceder a renderizarlo mediante auriculares y, en caso contrario, el objeto puede renderizarse mediante altavoces.

En algunas realizaciones, la distribución de elementos de audio sobre los altavoces y auriculares puede basarse en una indicación en el flujo de contenido recibido de si un elemento de audio está enlazado a un elemento/objeto visual en el flujo de contenido. Si el indicador lo indica, el elemento de audio se reproduce a través de los auriculares. Si el indicador indica que no es así, los elementos de audio se reproducen por los altavoces.

En los ejemplos anteriores, el selector 209 se ha dispuesto para seleccionar el renderizador apropiado 205, 207 basándose únicamente en los datos recibidos. Sin embargo, se apreciará que en muchas realizaciones pueden tenerse en cuenta otras consideraciones y específicamente otros datos.

En muchas realizaciones, el aparato puede incluir una función de entrada de usuario que puede recibir una entrada de usuario. El selector 209 puede, en tales realizaciones, estar dispuesto además para seleccionar entre el primer renderizador 205 y el segundo renderizador 207 basándose en la entrada del usuario. La entrada del usuario puede ser, por ejemplo, una indicación directa de una renderización preferida, tal como, por ejemplo, una indicación explícita de que un elemento de audio específico debe renderizarse a través de auriculares en lugar de altavoces. En otras realizaciones, la entrada del usuario puede ser más indirecta y puede, por ejemplo, modificar un criterio de selección o sesgar la selección hacia uno de los renderizadores 205, 207. Por ejemplo, una entrada de usuario puede indicar que se desea que más elementos de audio sean reproducidos por auriculares y el selector 209 puede cambiar un criterio de decisión para lograrlo.

Así, en algunas realizaciones, el usuario puede ser capaz de impactar directamente en la distribución de los elementos sobre los altavoces y auriculares. Un ejemplo es ofrecer a los usuarios la posibilidad de designar manualmente elementos individuales para su reproducción a través de los auriculares o los altavoces.

Otro ejemplo de control de la distribución por parte del usuario consiste en proporcionarle dos, o unos pocos, modos entre los que pueda elegir; por ejemplo, un modo de "experiencia individual" y otro de "experiencia compartida". En el caso en el que el usuario seleccione el modo "experiencia compartida", cualquiera de las formas descritas anteriormente para determinar qué elementos de audio deben reproducirse a través de los altavoces y auriculares, respectivamente, se puede utilizar en cualquier combinación.

En algunas realizaciones, el selector 209 puede estar preparado para analizar los elementos de audio y determinar qué renderizador 205, 207 utilizar basándose en este análisis. Por ejemplo, si no se recibe ningún indicador de propiedad de renderización de audio para un elemento de audio dado, el selector 209 puede proceder a analizar los

elemento(s) de audio para determinar una propiedad de audio, tal como por ejemplo el número de elementos de audio en la escena, el número de canales por elemento de audio, la posición de los elementos de audio, las distancias al/los oyente(s) (o a cada altavoz) de los elementos de audio o el movimiento de los elementos de audio. El selector 209 puede entonces proceder a decidir qué renderizador 205, 207 utilizar basándose en esta propiedad de audio o en una pluralidad de ellas.

En un ejemplo específico de configuración, en adelante referido como configuración X, el selector 209 puede seleccionar el renderizador para cada elemento de audio para producir la representación espacial más precisa de la escena de audio. Por ejemplo, si un elemento de audio se encuentra en una posición virtual relativamente cercana a la posición de uno de los altavoces físicos, entonces podría renderizarse en ese altavoz específico. Por el contrario, si un elemento de audio cae en una zona no cubierta por ningún altavoz, puede reproducirse a través de los auriculares. El hecho de que un elemento de audio tenga la misma dirección que un altavoz (desde el punto de vista de un oyente) también puede utilizarse del mismo modo para un único oyente, y también para varios oyentes, pero con la condición de que todos se alineen con el altavoz. Sin embargo, esto no suele ser práctico, ya que los usuarios pueden cambiar de posición a lo largo del tiempo. En esta configuración específica X, el selector 209 podría tener en cuenta la precisión angular del renderizador de auriculares (binaurales) 207 para tomar esta decisión.

Así, en algunas realizaciones, que no forman parte de la invención reivindicada, la selección del renderizador apropiado 205, 207 puede basarse adicionalmente en un análisis de las señales de audio. Por ejemplo, se puede utilizar un estimador de una propiedad acústica de las señales de audio para determinar propiedades tales como la distancia (o la velocidad) del objeto/fuente de audio (especialmente en el caso de señales multicanal) o el tiempo de reverberación. También pueden utilizarse clasificadores de señales de audio, tales como clasificadores de habla/música, clasificadores de géneros musicales o clasificadores de eventos de audio. Un tipo particular de clasificadores también podría utilizarse para determinar qué tipo de micrófonos (HOA, micrófono de solapa, omnidireccional, XY ...) se ha utilizado para grabar una señal determinada. También puede utilizarse un análisis de la distribución de frecuencias de la señal de audio para decidir qué sistema de audio (auriculares o altavoces) es más adecuado para reproducir todo el elemento de audio.

En los ejemplos anteriores, el selector 209 ha sido dispuesto para seleccionar el primer renderizador 205 o el segundo renderizador 207 sobre una base de elemento de audio por elemento de audio. Sin embargo, se entiende que esto no es necesario ni esencial. Por ejemplo, en algunas realizaciones, el selector 209 puede estar dispuesto para seleccionar qué renderizador 205, 207 utilizar para grupos de elementos de audio.

También, en algunas realizaciones, el selector 209 puede estar dispuesto para seleccionar separadamente entre los renderizadores 205, 207 para diferentes partes de un único elemento de audio. Por ejemplo, para algunos elementos de audio una parte puede ser renderizada por el primer renderizador 205 y otra parte puede ser renderizada por el segundo renderizador 207.

Se apreciará que un elemento de audio puede dividirse en diferentes partes de diferentes maneras dependiendo de los requisitos y preferencias de la realización individual. Por ejemplo, en algunas realizaciones, el elemento de audio puede recibirse como una combinación o colección de diferentes partes y el selector 209 puede seleccionar individualmente un renderizador 207 para cada parte. Por ejemplo, un elemento de audio puede representar una fuente de audio específica mediante un primer componente que representa una fuente de audio con una posición bien definida (por ejemplo, correspondiente al audio directo) y un segundo componente que representa un sonido más difuso y distribuido (por ejemplo, correspondiente al sonido reverberante). En este caso, el selector 209 puede estar preparado para renderizar el primer componente mediante auriculares y el segundo componente mediante altavoces.

En otras realizaciones, el selector 209 puede estar dispuesto para dividir el elemento de audio en diferentes partes para su renderización. Por ejemplo, un elemento de audio recibido puede corresponder a una señal de audio que puede analizarse para dividirla en diferentes partes que luego pueden renderizarse por separado.

Específicamente, en muchas realizaciones, diferentes partes del elemento de audio pueden corresponder a diferentes intervalos de frecuencia. Por ejemplo, el selector 209 puede, para una primera parte dada correspondiente a una gama de frecuencias específica, estar dispuesto para seleccionar qué renderizador 205, 207 utilizar. Puede proceder a hacer lo mismo para una gama de frecuencias diferente y, por tanto, puede dar lugar a que se utilicen diferentes renderizadores 205, 207 para la primera y la segunda gama de frecuencias.

En algunas realizaciones, que no forman parte de la invención reivindicada, pueden proporcionarse diferentes indicadores de propiedad de renderización de audio para diferentes partes del elemento de audio, y el selector 209 puede considerar el indicador de propiedad de renderización de audio específico para la parte dada cuando decide cómo renderizarlo. En otras realizaciones, se puede proporcionar un indicador de propiedad de reproducción de audio para todo el elemento de audio, pero con diferentes criterios de decisión que se utilizan para diferentes partes. Por ejemplo, para un intervalo de frecuencias medio-alto, la selección entre auriculares y altavoces se realiza en función del indicador de propiedad de renderización de audio recibido para el elemento de audio, mientras que para un intervalo de frecuencias muy bajo, el primer renderizador 205 se utiliza para renderizar la señal a través de los

altavoces independientemente de lo que indique el indicador de propiedad de renderización de audio (lo que refleja que las bajas frecuencias tienden a proporcionar señales espaciales mucho menos significativas).

Por ejemplo, la señal puede ser separada en una parte de baja frecuencia y una parte de alta frecuencia, usando filtrado de paso bajo y paso alto, donde la parte de baja frecuencia es enviada a los altavoces y la parte de alta frecuencia a los auriculares dependiendo del indicador de propiedad de renderización de audio. En algunas realizaciones de este tipo, puede utilizarse una separación avanzada de fuentes de audio (por ejemplo, dividiendo cada punto de tiempo-frecuencia entre renderizadores).

El uso de un filtrado que preserve la energía en cada punto de frecuencia temporal puede permitir a un sistema de renderización híbrido físico atenuar los posibles errores generados por el filtrado.

El enfoque descrito puede proporcionar una serie de efectos ventajosos incluyendo, como se ha descrito anteriormente, permitir una renderización espacial precisa percibida de una escena de audio al tiempo que permite/facilita a los usuarios en la misma ubicación interactuar directamente.

El enfoque puede reducir la complejidad y el uso de recursos en muchos escenarios debido a que se requiere una cantidad potencialmente reducida de procesamiento binaural. Otra ventaja que puede lograrse a menudo es una reducción de la energía utilizada por el sistema de reproducción de los auriculares, por ejemplo, en términos de potencia del amplificador y/o carga de procesamiento para el renderizador integrado, lo que puede ser crítico en el caso de auriculares no conectados (por ejemplo, auriculares alimentados por pilas).

Otra propiedad interesante del sistema híbrido de reproducción de audio para aplicaciones de VR es que tiende a proporcionar una seguridad mejorada. De hecho, al contrario que, si llevaran auriculares cerrados, los asistentes no estarían totalmente aislados del peligro potencial del entorno real que les rodea. Esto puede ser un factor importante en muchas situaciones prácticas.

Otra ventaja de un sistema híbrido tal como los descritos es el hecho de que parte del contenido de audio se reproduce a través del conjunto de altavoces comunes, lo que tiende a aumentar la sensación de experiencia compartida de los usuarios. El enfoque tiende a proporcionar una experiencia de usuario mejorada.

Se apreciará que la descripción anterior para mayor claridad ha descrito realizaciones de la invención con referencia a diferentes circuitos funcionales, unidades y procesadores. Sin embargo, es evidente que cualquier distribución adecuada de la funcionalidad entre los diferentes circuitos funcionales, unidades o procesadores se puede utilizar sin menoscabo de la invención. Por ejemplo, la funcionalidad ilustrada para ser realizada por procesadores o controladores separados puede ser realizada por el mismo procesador o controladores. Por lo tanto, las referencias a unidades funcionales o circuitos específicos sólo deben considerarse referencias a medios adecuados para proporcionar la funcionalidad descrita y no indicativas de una estructura u organización lógica o física estricta.

La invención puede implementarse de cualquier forma adecuada, incluyendo hardware, software, firmware o cualquier combinación de los mismos. La invención puede implementarse opcionalmente, al menos en parte, como software informático que se ejecuta en uno o más procesadores de datos y/o procesadores de señales digitales. Los elementos y componentes de una realización de la invención pueden implementarse física, funcional y lógicamente de cualquier forma adecuada. De hecho, la funcionalidad puede implementarse en una sola unidad, en una pluralidad de unidades o como parte de otras unidades funcionales. Como tal, la invención puede implementarse en una sola unidad o puede distribuirse física y funcionalmente entre diferentes unidades, circuitos y procesadores.

Aunque la presente invención se ha descrito en relación con algunas realizaciones, no pretende limitarse a la forma específica expuesta en el presente documento. Más bien, el alcance de la presente invención está limitado únicamente por las reivindicaciones que la acompañan. Además, aunque una característica pueda parecer descrita en relación con realizaciones particulares, un experto en la técnica reconocería que diversas características de las realizaciones descritas pueden combinarse de acuerdo con la invención. En las reivindicaciones, el término que comprende no excluye la presencia de otros elementos o pasos.

Además, aunque enumerados individualmente, una pluralidad de medios, elementos, circuitos o pasos del método pueden ser implementados, por ejemplo, por un único circuito, unidad o procesador. Además, aunque las características individuales pueden incluirse en diferentes reivindicaciones, éstas pueden combinarse ventajosamente, y la inclusión en diferentes reivindicaciones no implica que una combinación de características no sea factible y/o ventajosa. Asimismo, la inclusión de una característica en una categoría de reivindicaciones no implica una limitación a esta categoría, sino que indica que la característica es igualmente aplicable a otras categorías de reivindicaciones, según proceda.

Además, las referencias singulares no excluyen una pluralidad. Así, las referencias a "un", "una", "primero", "segundo", etc. no excluyen una pluralidad. Los signos de referencia en las reivindicaciones se proporcionan meramente como ejemplo clarificador y no deben interpretarse como limitativos del alcance de las reivindicaciones en modo alguno.

REIVINDICACIONES

1. Un aparato de audio que comprende:

un receptor (201) para recibir datos que describen una escena de audio, los datos que comprenden datos de audio para un conjunto de elementos de audio correspondientes a fuentes de audio en la escena y metadatos que comprenden al menos un primer indicador de propiedad de renderización de audio para un primer elemento de audio del conjunto de elementos de audio;
un primer renderizador (205) para renderizar elementos de audio generando un primer conjunto de señales de audio para un conjunto de altavoces;
un segundo renderizador (207) para renderizar elementos de audio generando un segundo conjunto de señales de audio para un auricular; y
un selector (209) dispuesto para seleccionar entre el primer renderizador (205) y el segundo renderizador (207) para la renderización de al menos una primera parte del primer elemento de audio en respuesta al primer indicador de propiedad de renderización de audio;
caracterizado porque
el primer indicador de propiedad de renderización de audio es indicativo de si la primera parte del primer elemento de audio representa una fuente de audio con una propiedad espacial que tiene una orientación fija con respecto a la cabeza y está destinada a tener una ubicación fija con respecto a la cabeza de un usuario o representa una fuente de audio con una propiedad espacial que tiene una orientación no fija con respecto a la cabeza y está destinada a tener una ubicación fija en un entorno.

2. El aparato de la reivindicación 1 comprende además un receptor (211) de pose de oyente para recibir una pose del oyente indicativa de la pose de un oyente, y el primer renderizador (205) está dispuesto para generar el primer conjunto de señales de audio independientemente de la pose del oyente y el segundo renderizador (207) está dispuesto para generar el segundo conjunto de señales de audio en respuesta a la pose del oyente.

3. El aparato de la reivindicación 1 dispuesto para generar señales de audio para una pluralidad de oyentes en el que el primer renderizador (205) está dispuesto para generar el primer conjunto de señales de audio como un conjunto común de señales de audio para la pluralidad de oyentes; y el segundo renderizador (207) está dispuesto para generar el segundo conjunto de señales de audio para auriculares para un primer oyente de la pluralidad de oyentes y para generar un tercer conjunto de señales de audio para auriculares para un segundo oyente de la pluralidad de oyentes.

4. El aparato de cualquier reivindicación anterior, en el que la primera parte es un subintervalo de frecuencia del primer elemento de audio.

5. El aparato de cualquier reivindicación anterior en el que el selector (209) está dispuesto para seleccionar diferentes renderizadores (205, 207) del primer renderizador (205) y del segundo renderizador (207) para la primera parte del primer elemento de audio y para una segunda parte del primer elemento de audio.

6. El aparato de la reivindicación 1, en el que el indicador de propiedad de renderización de audio es indicativo de un formato de audio del primer elemento de audio.

7. El aparato de la reivindicación 1, en el que el indicador de propiedad de renderización de audio es indicativo de un tipo de fuente de audio para el primer elemento de audio.

8. El aparato de cualquier reivindicación anterior, en el que el indicador de propiedad de renderización de audio es indicativo de una propiedad de renderización de orientación para el renderizado del renderizado del primer elemento de audio.

9. El aparato de la reivindicación 8, en el que el indicador de propiedad de reproducción de audio es indicativo de si la primera parte del primer elemento de audio está destinada a reproducirse a través de altavoces o auriculares.

10. El aparato de cualquier reivindicación anterior, en el que el receptor (201) está dispuesto además para recibir datos visuales indicativos de una escena virtual correspondiente a la escena de audio, y el indicador de propiedad de renderización de audio es indicativo de si el primer elemento de audio representa una fuente de audio correspondiente a un objeto de escena visual.

11. El aparato de cualquier reivindicación anterior comprende además una entrada de usuario para recibir una entrada de usuario y en el que el selector (211) está dispuesto para seleccionar entre el primer renderizador (205) y el segundo renderizador (207) para renderizar al menos la primera parte del primer elemento de audio en respuesta a la entrada de usuario.

12. El aparato de cualquier reivindicación anterior en el que el selector (209) está dispuesto para determinar una propiedad de audio del primer elemento de audio y para seleccionar entre el primer renderizador (205) y el segundo

renderizador (207) para el renderizado de al menos la primera parte del primer elemento de audio en respuesta a la propiedad de audio.

13. Un método de procesamiento de audio que comprende:

recibir datos que describen una escena de audio, los datos que comprenden datos de audio para un conjunto de elementos de audio correspondientes a fuentes de audio en la escena y metadatos que comprenden al menos un primer indicador de propiedad de renderización de audio para un primer elemento de audio del conjunto de elementos de audio;
renderización de elementos de audio mediante la generación de un primer conjunto de señales de audio para un conjunto de altavoces;
renderización de elementos de audio mediante la generación de un segundo conjunto de señales de audio para un auricular; y
seleccionar entre la renderización de al menos una primera parte del primer elemento de audio para el conjunto de altavoces y para el auricular en respuesta al primer indicador de propiedad de renderización de audio
caracterizado porque
el indicador de propiedad de renderización de audio es indicativo de si la primera parte del primer elemento de audio representa una fuente de audio con una propiedad espacial que tiene una orientación fija con respecto a la cabeza y está destinada a tener una ubicación fija con respecto a la cabeza de un usuario o representa una fuente de audio con una propiedad espacial que tiene una orientación no fija con respecto a la cabeza y está destinada a tener una ubicación fija en un entorno.

14. Un producto de programa de ordenador que comprende medios de código de programa de ordenador adaptados para realizar todos los pasos de la reivindicación 13 cuando dicho programa se ejecuta en un ordenador.

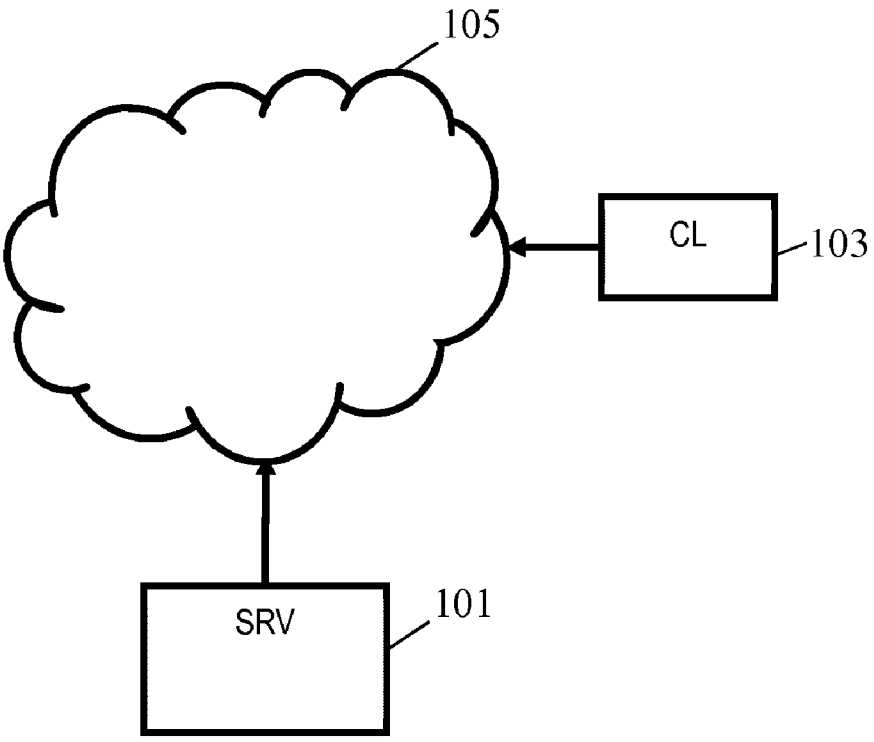

FIG. 1

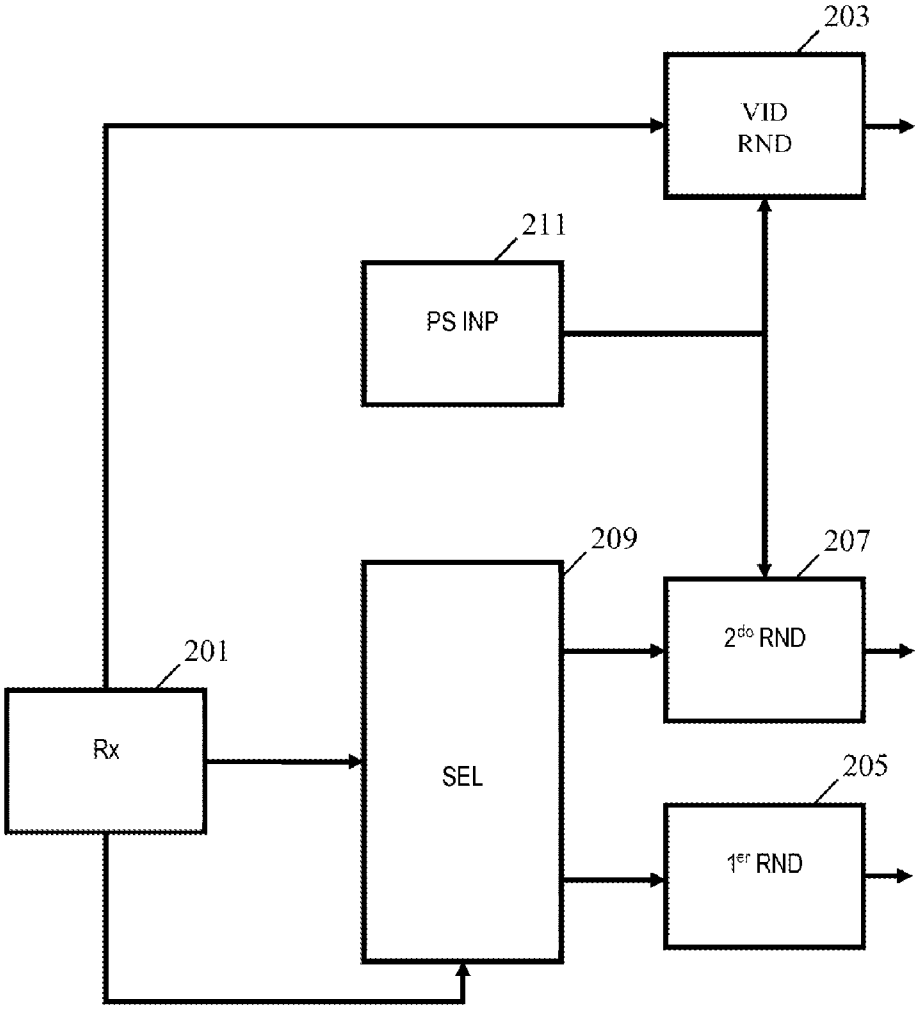

FIG. 2