

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 974 833**

51 Int. Cl.:

H04S 7/00 (2006.01)

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **21.06.2021** **PCT/EP2021/066763**

87 Fecha y número de publicación internacional: **30.12.2021** **WO21259829**

96 Fecha de presentación y número de la solicitud europea: **21.06.2021** **E 21732929 (1)**

97 Fecha y número de publicación de la concesión europea: **20.12.2023** **EP 4169267**

54 Título: **Aparato y método para generar una señal de reverberación difusa**

30 Prioridad:

22.06.2020 EP 20181351

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

01.07.2024

73 Titular/es:

KONINKLIJKE PHILIPS N.V. (100.0%)

High Tech Campus 52

5656 AG Eindhoven, NL

72 Inventor/es:

KOPPENS, JEROEN, GERARDUS, HENRICUS y
KECHICHIAN, PATRICK

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 974 833 T3

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

DESCRIPCIÓN

Aparato y método para generar una señal de reverberación difusa

5 Campo de la invención

La invención se refiere a un aparato y método que procesa datos de audio, y en particular, pero no exclusivamente, para el procesamiento para generar señales de reverberación difusa para aplicaciones de realidad aumentada/mixta/virtual.

10

Antecedentes de la invención

La variedad y el rango de las experiencias basadas en contenidos audiovisuales han aumentado sustancialmente en los últimos años, con el desarrollo y la introducción continua de nuevos servicios y formas de utilizar y consumir dichos contenidos. En particular, se están desarrollando muchos servicios, aplicaciones y experiencias espaciales e interactivas para proporcionar a los usuarios una experiencia más involucrada e inmersiva.

Ejemplos de dichas aplicaciones son las aplicaciones de Realidad Virtual (VR), Realidad Aumentada (AR) y Realidad Mixta (MR), que se están convirtiendo rápidamente en la corriente principal, con una serie de soluciones dirigidas al mercado de consumo. Una serie de organismos de estandarización también están elaborando una serie de estándares. Dichas actividades de estandarización están desarrollando activamente estándares para los diversos aspectos de los sistemas de VR/AR/MR, que incluyen, por ejemplo, retransmisión, radiodifusión, renderización, etc.

Las aplicaciones de VR tienden a proporcionar experiencias de usuario correspondientes a que el usuario se encuentre en un mundo/entorno/escena diferente, mientras que las aplicaciones de AR (que incluyen la Realidad Mixta MR) tienden a proporcionar experiencias de usuario correspondientes al usuario que se encuentra en el entorno actual, pero con información adicional u objetos virtuales o información que se agrega. Por lo tanto, las aplicaciones de VR tienden a proporcionar un mundo/escena totalmente inmersivo generado sintéticamente, 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 a menudo se utilizan indistintamente y tienen un alto grado de superposición. A continuación, se utilizará el término Realidad Virtual/VR para denotar tanto la Realidad Virtual como la Realidad Aumentada/Mixta.

Como un ejemplo, un servicio cada vez más popular es la provisión de imágenes y audio de tal manera que un usuario es capaz de interactuar activa y dinámicamente con el sistema para cambiar los parámetros de la renderización de tal manera 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 capacidad de cambiar la posición de visualización efectiva y la dirección de visualización 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 que se proporcione 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 está mirando. Normalmente, dichas aplicaciones de realidad virtual se basan en un modelo tridimensional de la escena en el que el modelo se evalúa dinámicamente para proporcionar la vista específica solicitada. Este enfoque es bien conocido, por ejemplo, en aplicaciones de juegos, tal como en la categoría de juegos de disparos en primera persona, para ordenadores y consolas.

También es deseable, en particular para aplicaciones de realidad virtual, que la imagen que se presenta sea una imagen tridimensional, normalmente presentada utilizando una pantalla estereoscópica. De hecho, para optimizar la inmersión del espectador, normalmente se prefiere que el usuario experimente la escena presentada como una escena tridimensional. De hecho, una experiencia de realidad virtual debería permitir preferiblemente al usuario seleccionar su propia posición, punto de vista y momento en el tiempo con respecto a un mundo virtual.

Además de la renderización visual, la mayoría de las aplicaciones de realidad virtual/realidad aumentada proporcionan una experiencia de audio correspondiente. En muchas aplicaciones, el audio proporciona preferiblemente una experiencia de audio espacial en la que se percibe que las fuentes de audio llegan desde posiciones que corresponden a las posiciones de los objetos correspondientes en la escena visual. Por lo tanto, las escenas de audio y video se perciben preferiblemente como consistentes y ambas proporcionan una experiencia espacial completa.

Por ejemplo, muchas experiencias inmersivas son proporcionadas por una escena de audio virtual que se genera mediante la reproducción de auriculares utilizando tecnología de renderización de audio binaural. En muchos escenarios, dicha reproducción de auriculares se puede basar en el seguimiento de la cabeza, de tal manera que la renderización pueda responder a los movimientos de la cabeza del usuario, lo que aumenta en gran medida la sensación de inmersión.

65

Una característica importante para muchas aplicaciones es la de cómo generar y/o distribuir audio que pueda proporcionar una percepción natural y realista del entorno de audio. Por ejemplo, al generar audio para una aplicación de realidad virtual, es importante que no solo se generen las fuentes de audio deseadas, sino que también se modifiquen para proporcionar una percepción realista del entorno de audio, que incluye la amortiguación, reflexión, la coloración, etc.

En el caso de la acústica de habitaciones, o más generalmente de la acústica de entornos, las reflexiones de las ondas de sonido en las paredes, suelo, techo, objetos, etc. de un entorno hacen que las versiones retardadas y atenuadas (normalmente dependientes de la frecuencia) de la señal de la fuente de sonido lleguen al oyente (es decir, al usuario de un sistema VR/AR) a través de diferentes rutas. El efecto combinado se puede modelar mediante una respuesta al impulso, que en lo sucesivo se puede denominar Respuesta al Impulso de la Habitación (RIR) (aunque el término sugiere un uso específico para un entorno acústico en forma de habitación, se tiende a utilizar de forma más general con respecto a un entorno acústico, ya sea o no que este corresponda a una habitación).

Como se ilustra en la FIG. 1, una respuesta al impulso de una habitación consiste normalmente en un sonido directo que depende de la distancia de la fuente de sonido al oyente, seguido de una porción reverberante que caracteriza las propiedades acústicas de la habitación. El tamaño y la conformación de la habitación, la posición de la fuente de sonido y del oyente en la habitación y las propiedades reflectantes de las superficies de la habitación desempeñan una función importante en las características de esta porción reverberante.

La porción reverberante se puede descomponer en dos regiones temporales, usualmente superpuestas. La primera región contiene las llamadas reflexiones tempranas, que representan reflexiones aisladas de la fuente de sonido sobre paredes u obstáculos dentro de la habitación antes de llegar al oyente. A medida que aumenta el desfase de tiempo, aumenta el número de reflexiones presentes en un intervalo de tiempo fijo y las rutas pueden incluir reflexiones secundarias o de orden superior (por ejemplo, las reflexiones pueden estar en varias paredes o en ambas paredes y techo, etc.).

La segunda región en la porción reverberante es la parte donde la densidad de estas reflexiones aumenta hasta un punto en el que ya no se pueden aislar por el cerebro humano. Esta región se denomina normalmente reverberación difusa, reverberación tardía o cola de reverberación.

La porción reverberante contiene pistas que dan al sistema auditivo información sobre la distancia de la fuente, el tamaño y las propiedades acústicas de la habitación. La energía de la porción reverberante con respecto a la de la porción anecoica determina en gran medida la distancia percibida de la fuente de sonido. El nivel y el retardo de las primeras reflexiones pueden proporcionar pistas sobre qué tan cerca está la fuente de sonido de una pared, y el filtrado por antropometrías puede fortalecer la evaluación de la pared, piso o techo específicos.

La densidad de las reflexiones (tempranas) contribuye al tamaño percibido de la habitación. El tiempo que tardan las reflexiones en caer 60 dB en el nivel de energía, indicado por el tiempo de reverberación T_{60} , es una medida utilizada con frecuencia para la rapidez con la que se disipan las reflexiones en la habitación. El tiempo de reverberación proporciona información sobre las propiedades acústicas de la habitación; tal como específicamente si las paredes son muy reflectantes (por ejemplo, el baño) o si hay mucha absorción de sonido (por ejemplo, el dormitorio con muebles, alfombras y cortinas).

Además, los RIR pueden depender de las propiedades antropométricas de un usuario cuando forman parte de una respuesta de impulso binaural de la habitación (BRIR), debido a que la RIR se filtra por la cabeza, orejas y hombros; es decir, las respuestas de impulso relacionadas con la cabeza (HRIR).

Como las reflexiones en la reverberación tardía no pueden ser diferenciadas y aisladas por un oyente, a menudo se simulan y representan paramétricamente con, por ejemplo, un reverberador paramétrico utilizando una red de retardo de retroalimentación, como el reverberador Jot bien conocido.

Para las reflexiones tempranas, la dirección de incidencia y los retardos dependientes de la distancia son pistas importantes para que los humanos extraigan información acerca de la habitación y la posición relativa de la fuente de sonido. Por lo tanto, la simulación de las reflexiones tempranas debe ser más explícita que la reverberación tardía. Por lo tanto, en los algoritmos de renderización acústico eficientes, las reflexiones tempranas se simulan de manera diferente a la reverberación tardía. Un método bien conocido para las reflexiones tempranas es reflejar las fuentes de sonido en cada uno de los límites de la habitación para generar una fuente de sonido virtual que represente la reflexión.

Para las reflexiones tempranas, la posición del usuario y/o de la fuente de sonido con respecto a los límites (paredes, techo, suelo) de una habitación es relevante, mientras que para la reverberación tardía, la respuesta acústica de la habitación es difusa y, por lo tanto, tiende a ser más homogénea en toda la habitación. Esto permite que la simulación de la reverberación tardía sea a menudo más eficiente desde el punto de vista computacional que las reflexiones tempranas.

Dos propiedades principales de la reverberación tardía que están definidas por la habitación son el valor T60 y el nivel de reverberación. En términos de la respuesta al impulso de reverberación difusa, estos valores representan la pendiente y la amplitud de la respuesta al impulso. Ambos normalmente dependen en gran medida de la frecuencia en las habitaciones naturales.

El parámetro T60 es importante para proporcionar una impresión de la reflectividad y el tamaño de la habitación, mientras que el nivel de reverberación es indicativo del efecto compuesto de múltiples reflexiones en los límites de la habitación. El nivel de reverberación y su comportamiento de frecuencia dependen del retardo previo, lo que indica dónde se hace la distinción entre reflexiones tempranas y reverberación tardía (véase FIG. 2).

El nivel de reverberación tiene su principal relevancia psicoacústica con respecto al sonido directo. La diferencia de nivel entre los dos es una indicación de la distancia entre la fuente de sonido y el usuario (o punto de medición RIR). Una mayor distancia provocará una mayor atenuación en el sonido directo, mientras que el nivel de la reverberación tardía se mantiene igual (es el mismo en toda la habitación). Del mismo modo, para las fuentes cuya directividad depende de dónde se encuentre el usuario con respecto a la fuente, la directividad influye en la respuesta directa a medida que el usuario se mueve alrededor de la fuente, pero no en el nivel de reverberación.

Un desafío importante y una consideración para muchos sistemas, tales como las aplicaciones de realidad virtual, es el de cómo representar y distribuir eficientemente el entorno de audio. A menudo, el audio de un entorno se representa y distribuye al proporcionar señales que representan las señales de origen individuales junto con datos que pueden describir paramétricamente las propiedades de la fuente de audio y del entorno acústico. Este desafío dista mucho de ser una cuestión trivial y se pueden considerar un rango de cuestiones.

Se ha propuesto separar las descripciones de las rutas directas y la reverberación difusa. Sin embargo, la cuestión de cómo representar, distribuir y renderizar/sintetizar la reverberación difusa está recibiendo actualmente un gran interés.

Se ha propuesto proporcionar una indicación del nivel de reverberación que no está en relación con el sonido directo, sino más bien mediante una propiedad más genérica. Se ha presentado una propuesta específica como parte de los preparativos para la Convocatoria de Propuestas de Audio I MPEGI (CfP) en la que se ha definido un Formato de Entrada de Codificador (EIF) (sección 3.9 del documento de salida MPEG N19211, "MPEG-I 6DoF Audio Encoder Input Format", MPEG 130). El EIF define el nivel de reverberación mediante un retardo previo y una Relación Directa a Difusa (DDR). La DDR se define como la relación entre la energía de reverberación difusa y la energía de la fuente emitida después del retardo previo:

$$DDR = \frac{\text{energía de reverberación difusa total}}{\text{energía total emitida}}$$

Sin embargo, si bien dicho parámetro puede ser útil, hay muchas cuestiones sustanciales que se deben abordar. Por ejemplo, actualmente no existe ninguna propuesta sobre cómo deben definirse o determinarse los parámetros específicos. Tampoco se tiene en cuenta cómo se puede utilizar una indicación DDR para renderizar audio y, específicamente, cómo se puede utilizar para generar señales de reverberación difusa.

El documento EP3402222 divulga métodos de virtualización para generar una señal binaural en respuesta a canales de una señal de audio multicanal, que aplican una respuesta de impulso de habitación binaural (BRIR) a cada canal, incluso al utilizar al menos una red de retardo de retroalimentación (FDN) para aplicar una reverberación tardía común a una mezcla descendente de los canales.

Por lo tanto, los enfoques y propuestas actuales sobre cómo representar y generar audio y específicamente la reverberación difusa tienden a ser subóptimos o insuficientes y/o incompletos. Esto es particularmente el caso, por ejemplo, de las aplicaciones de realidad virtual, en las que la posición para la que se debe generar el audio puede cambiar sustancialmente.

Por lo tanto, sería ventajoso un enfoque para generar señales de reverberación difusa. En particular, sería ventajoso un enfoque que permitiera mejorar la operación, aumentar la flexibilidad, reducir la complejidad, facilitar la implementación, mejorar la experiencia de audio, mejorar la calidad de audio, reducir la carga computacional, mejorar la idoneidad para diversas posiciones, mejorar el rendimiento para aplicaciones de realidad virtual/mixta/aumentada, mejorar las pistas perceptivas para la reverberación difusa y/o mejorar el rendimiento y/o la operación.

Resumen de la invención

De acuerdo con lo anterior, la Invención busca preferiblemente mitigar, aliviar o eliminar una o más de las desventajas mencionadas anteriormente individualmente o en cualquier combinación.

De acuerdo con un aspecto de la invención se proporciona un aparato de audio para generar una señal de reverberación difusa para un entorno; el aparato comprende: un receptor dispuesto para recibir una pluralidad de señales de audio que representan fuentes de sonido en el entorno; un receptor de metadatos dispuesto para recibir

metadatos para la pluralidad de señales de audio, los metadatos comprenden: una relación entre la señal de reverberación difusa y la señal total indicativa de un nivel de sonido de reverberación difusa con respecto al sonido total emitido en el entorno, y para cada señal de audio: una indicación del nivel de señal; datos de directividad indicativos de la directividad de la radiación de sonido de la fuente de sonido representada por la señal de audio; Un circuito dispuesto de tal manera que, para cada una de las pluralidades de señales de audio, se determine: una indicación de la energía total emitida basada en la indicación del nivel de señal y los datos de directividad, y un coeficiente de mezcla descendente basado en la indicación de la energía total emitida y la relación entre la señal de reverberación difusa y la señal total; un mezclador descendente dispuesto para generar una señal de mezcla descendente al combinar los componentes de la señal de cada señal de audio generada al aplicar el coeficiente de mezcla descendente para cada señal de audio a la señal de audio; Un reverberador para generar la señal de reverberación difusa para el entorno a partir del componente de señal de mezcla descendente.

La invención puede proporcionar una determinación mejorada y/o facilitada de una señal de reverberación difusa en muchas realizaciones. La invención puede, en muchas realizaciones y escenarios, generar una señal de reverberación difusa con un sonido más natural que proporciona una percepción mejorada del entorno acústico. La generación de la señal de reverberación difusa a menudo se puede generar con baja complejidad y bajos requisitos de recursos computacionales. El enfoque permite que el sonido de reverberación difuso en un entorno acústico se represente eficazmente mediante relativamente pocos parámetros, lo que también proporciona una representación eficiente de las fuentes individuales y de la propagación del sonido de ruta individual a partir de ellas, y específicamente para una propagación de la ruta directa.

En muchas realizaciones, el enfoque puede permitir que se genere una señal de reverberación difusa independientemente de las posiciones de la fuente y/o del oyente. Esto puede permitir la generación eficiente de señales de reverberación difusa para aplicaciones dinámicas en las que cambian las posiciones, tal como para muchas aplicaciones de realidad virtual y realidad aumentada.

La relación entre la señal de reverberación difusa y la señal total también se puede denominar relación entre el nivel de la señal de reverberación difusa y el nivel de la señal total o la relación entre el nivel de reverberación difusa y el nivel total o relación entre la energía de la fuente emitida con la energía de reverberación difusa (o variaciones/permutaciones de la misma).

El aparato de audio se puede implementar en un solo dispositivo o en una sola unidad funcional, o se puede distribuir a través de diferentes dispositivos o funcionalidades. Por ejemplo, el aparato de audio se puede implementar como parte de una unidad funcional del decodificador o se puede distribuir con algunos elementos funcionales que se realizan en el lado del decodificador y otros elementos que se realizan en el lado del codificador.

De acuerdo con una característica opcional de la invención, la directividad de la radiación de sonido depende de la frecuencia y el circuito se dispone para generar una energía total emitida dependiente de la frecuencia y coeficientes de mezcla descendente dependientes de la frecuencia.

El enfoque puede proporcionar una operación particularmente eficiente para generar señales de reverberación difusa que reflejen dependencias de frecuencia.

De acuerdo con una característica opcional de la invención, la relación entre la señal de reverberación difusa y la señal total depende de la frecuencia y el circuito se dispone para generar coeficientes de mezcla descendente dependientes de la frecuencia.

El enfoque puede proporcionar una operación particularmente eficiente para generar señales de reverberación difusa dependientes de la frecuencia que reflejen dependencias de frecuencia.

De acuerdo con una característica opcional de la invención, la relación entre la señal de reverberación difusa y la señal total comprende una parte dependiente de la frecuencia y una parte no dependiente de la frecuencia, y en la que el circuito se dispone para generar los coeficientes de mezcla descendente en dependencia de la parte no dependiente de la frecuencia y para adaptar el reverberador en dependencia de la parte dependiente de la frecuencia.

El enfoque puede proporcionar una operación particularmente eficiente para generar señales de reverberación difusa que reflejen dependencias de frecuencia, y puede reducir específicamente la complejidad y/o el uso de recursos. Por ejemplo, el enfoque puede permitir que la dependencia de la frecuencia se refleje mediante un solo filtrado de la señal de mezcla descendente.

De acuerdo con una característica opcional de la invención, el circuito se dispone para determinar la indicación de energía total emitida para una primera señal de audio de la pluralidad de señales de audio en respuesta a un escalado de la indicación del nivel de señal para la primera señal de audio por un valor determinado al integrar un patrón de directividad de la fuente de sonido representada por la primera señal de audio.

- 5 Esto puede proporcionar una operación particularmente ventajosa en muchas realizaciones. El escalado puede ser cualquier función aplicada a la indicación del nivel de señal con respecto a la determinación de los coeficientes de mezcla descendente. Normalmente, la función puede aumentar monótonamente como una función de la indicación de la energía total emitida. El escalado puede ser lineal o no lineal.
- 10 El escalado puede ser independiente de la variación temporal de la señal y, por lo tanto, puede que no sea necesario actualizarla con niveles instantáneos de la señal de audio y que sólo sea necesario volver a calcularla cuando cambie la indicación del nivel de señal o el patrón de directividad.
- 15 De acuerdo con una característica opcional de la invención, la indicación del nivel de señal para una primera señal de audio de la pluralidad de señales de audio comprende una distancia de referencia, la distancia de referencia indica una distancia desde la fuente de audio representada por la primera señal de audio para una ganancia de referencia de distancia para la primera señal de audio.
- 20 Esto puede proporcionar una operación particularmente ventajosa en muchas realizaciones. La ganancia de referencia de distancia puede ser un valor predeterminado y, normalmente puede ser común al menos a algunas y, a menudo, a todas las fuentes y señales de audio. En muchas realizaciones, la ganancia de referencia de distancia puede ser de 0 dB.
- 25 De acuerdo con una característica opcional de la invención, la integración se realiza para una distancia que es la distancia de referencia desde la fuente de audio representada por la primera señal de audio.
- Esto puede proporcionar un enfoque particularmente eficiente y puede facilitar la operación.
- 30 De acuerdo con una característica opcional de la invención, la relación entre la señal de reverberación difusa y la señal total es indicativa de una energía del sonido de reverberación difusa con respecto a una energía del sonido total emitido en el entorno.
- Esto puede proporcionar una operación particularmente ventajosa en muchas realizaciones.
- 35 De acuerdo con una característica opcional de la invención, la relación entre la señal difusa y la señal total es indicativa de una amplitud inicial del sonido difuso con respecto a una energía del sonido total emitido en el entorno.
- Esto puede proporcionar una operación particularmente ventajosa en muchas realizaciones.
- 40 De acuerdo con una característica opcional de la invención, el coeficiente de mezcla descendente determinado para una primera señal de audio de la pluralidad de señales de audio es independiente de una posición de una primera fuente de audio representada por la primera señal de audio.
- 45 Esto puede proporcionar una operación particularmente ventajosa en muchas realizaciones y, en particular, puede facilitar la operación para aplicaciones dinámicas con fuentes de sonido que cambian de posición, tales como para aplicaciones de realidad virtual.
- De acuerdo con una característica opcional de la invención, el coeficiente de mezcla descendente determinado para una primera señal de audio de la pluralidad de señales de audio es independiente de la posición de un oyente.
- 50 Esto puede proporcionar una operación particularmente ventajosa en muchas realizaciones, y puede, en particular, facilitar la operación para aplicaciones dinámicas con posiciones cambiantes, tales como para aplicaciones de realidad virtual.
- 55 En algunas realizaciones, el procesamiento del aparato de audio es independiente de las posiciones de la fuente de audio. En algunas realizaciones, el procesamiento del aparato de audio es independiente de la posición del oyente.
- En algunas realizaciones, el procesamiento del aparato de audio es sólo independiente de la posición del oyente dentro de una región para la que se aplica la relación entre la señal difusa y la señal total.
- 60 En algunas realizaciones, una tasa de actualización para los coeficientes de mezcla descendente es menor que una tasa de actualización para las posiciones de una primera fuente de audio representada por la primera señal de audio. En algunas realizaciones, una tasa de actualización para los coeficientes de mezcla descendente es menor que una tasa de actualización para las posiciones de un oyente. Los coeficientes de mezcla descendente se pueden calcular a una tasa temporal mucho más baja que la tasa de actualización de la posición del oyente/posición de la fuente de audio.
- 65 De acuerdo con una característica opcional de la invención, la indicación del nivel de señal para una primera señal de audio de la pluralidad de señales de audio comprende además una indicación de ganancia para la primera señal de audio, la indicación de ganancia es indicativa de una ganancia que se aplicará a la primera señal de audio cuando se

renderiza el sonido desde una primera fuente de audio representada por la primera señal de audio, y en el que el circuito se dispone para determinar el coeficiente de mezcla descendente para la primera señal de audio en respuesta a la indicación de ganancia.

5 De acuerdo con una característica opcional de la invención, el aparato de audio comprende además un circuito de renderización directa dispuesto para generar una señal de audio de ruta directa para una primera señal de audio de la pluralidad de señales de audio en respuesta a una indicación del nivel de señal y datos de directividad para la primera señal de audio.

10 Esto puede proporcionar una operación particularmente ventajosa en muchas realizaciones.

De acuerdo con una característica opcional de la invención, los metadatos comprenden además una indicación de retardo y la relación entre la señal difusa y la señal total (DSR) es indicativa de una energía de sonido de reverberación difusa en el entorno que tiene un retardo mayor que un retardo indicado por la indicación de retardo con respecto a una energía del sonido total emitido.

15 La energía del sonido de reverberación difusa en el entorno que tiene un retardo mayor que la indicación de retardo puede reflejar/estar determinada por/como contribuciones de respuesta al impulso de la habitación que ocurren al menos un cierto retardo después de la emisión del sonido correspondiente en una fuente de audio, donde el cierto retardo está indicado por la indicación de retardo.

En algunas realizaciones, la relación entre la señal difusa y la señal total (la DSR) es indicativa de una energía de sonido de reverberación difusa con respecto a una energía de sonido total emitido en el entorno, en el que la energía del sonido de reverberación difusa está determinada por las contribuciones de respuesta de la habitación que ocurren al menos un cierto retardo después de la emisión del sonido correspondiente en una fuente de audio.

25 De acuerdo con otro aspecto de la invención, se proporciona un método para generar una señal de reverberación difusa para un entorno, el método comprende: recibir una pluralidad de señales de audio que representan fuentes de sonido en el entorno; recibir metadatos para la pluralidad de señales de audio, los metadatos comprenden: una relación entre la señal de reverberación difusa y la señal total indicativa de un nivel de sonido de reverberación difusa con respecto al sonido total emitido en el entorno, y para cada señal de audio: una indicación del nivel de señal; datos de directividad indicativos de la directividad de la radiación de sonido de la fuente de sonido representada por la señal de audio; para cada una de las pluralidades de señales de audio, que determina: una indicación de la energía total emitida basada en la indicación del nivel de señal y los datos de directividad, y un coeficiente de mezcla descendente basado en la indicación de la energía total emitida y la relación entre la señal de reverberación difusa y la señal total; generar una señal de mezcla descendente al combinar los componentes de la señal de cada señal de audio generada al aplicar el coeficiente de mezcla descendente para cada señal de audio a la señal de audio; generar de la señal de reverberación difusa para el entorno a partir del componente de señal de mezcla descendente.

40 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

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

La FIG. 1 ilustra un ejemplo de una respuesta al impulso de una habitación;

50 La FIG. 2 ilustra un ejemplo de una respuesta al impulso de una habitación;

La FIG. 3 ilustra un ejemplo de elementos del sistema de realidad virtual;

55 La FIG. 4 ilustra un ejemplo de un aparato de audio para generar una salida de audio de acuerdo con algunas realizaciones de la invención;

La FIG. 5 ilustra un ejemplo de un aparato de reverberación de audio para generar una señal de reverberación difusa de acuerdo con algunas realizaciones de la invención;

60 La FIG. 6 ilustra un ejemplo de una respuesta al impulso de una habitación; y

La FIG. 7 ilustra un ejemplo de un reverberador.

Descripción detallada de algunas realizaciones de la invención

La siguiente descripción se centrará en el procesamiento y la generación de audio para una aplicación de realidad virtual, pero se apreciará que los principios y conceptos descritos se pueden utilizar en muchas otras aplicaciones y realizaciones.

- 5 Las experiencias virtuales que permiten a un usuario moverse alrededor de un mundo virtual se están volviendo cada vez más populares y se están desarrollando servicios para satisfacer dicha demanda.

En algunos sistemas, la aplicación de VR se puede proporcionar localmente a un espectador mediante, por ejemplo, un dispositivo independiente que no utiliza, o incluso no tiene acceso a, ningún dato o procesamiento de VR remoto. Por ejemplo, un dispositivo tal como una consola de juegos puede comprender un almacén para almacenar 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.

- 10 En otros sistemas, la aplicación VR se puede implementar y realizar de forma remota por el espectador. Por ejemplo, un dispositivo local para el usuario puede detectar datos de movimiento/pose de recepción 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 visualización adecuadas y las señales de audio correspondientes para la pose del usuario en base a los datos de la escena que describen la escena. A continuación, las imágenes de visualización y las señales de audio correspondientes 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) y el correspondiente flujo de audio que es presentada directamente por el dispositivo local. Por lo tanto, en dicho ejemplo, el dispositivo local no puede realizar ningún procesamiento VR, excepto para transmitir datos de movimiento y presentar datos de vídeo recibidos.

- 25 En muchos sistemas, la funcionalidad se puede distribuir a través de un dispositivo local y un dispositivo remoto. Por ejemplo, el dispositivo local puede procesar la entrada recibida y los datos del sensor para generar poses de usuario que se transmiten continuamente al dispositivo VR remoto. El dispositivo VR remoto luego puede generar las imágenes de vista correspondientes y señales de audio correspondientes y transmitirlos al dispositivo local para presentación. En otros sistemas, es posible que el dispositivo de realidad virtual remoto no genere directamente las imágenes de vista y las señales de audio correspondientes, sino que seleccione datos de escena relevantes y los transmita al dispositivo local, que luego puede generar las imágenes de vista y las señales de audio correspondientes que se presentan. Por ejemplo, el dispositivo VR remoto puede identificar el punto de captura más cercano y extraer los datos de la escena correspondiente (por ejemplo, un conjunto de fuentes de objetos y sus metadatos de posición) y transmitirlos al dispositivo local. A continuación, el dispositivo local puede procesar los datos de la escena recibidos para generar las imágenes y las señales de audio para la pose específica y actual del usuario. La pose del usuario normalmente corresponderá a la pose de la cabeza, y las referencias a la pose del usuario normalmente se pueden considerar equivalentemente correspondientes a las referencias a la pose de la cabeza.

- 30 En muchas aplicaciones, especialmente para servicios de radiodifusión, una fuente puede transmitir o retransmitir datos de escena en forma de imagen (que incluyen vídeo) y representación de audio de la escena que es independiente de la pose del usuario. Por ejemplo, las señales y los metadatos correspondientes a las fuentes de audio dentro de los confines de una determinada habitación virtual se pueden transmitir o retransmitir a una pluralidad de clientes. A continuación, los clientes individuales pueden sintetizar localmente las señales de audio correspondientes a la pose del usuario actual. Del mismo modo, la fuente puede transmitir una descripción general del entorno de audio, que incluye describir las fuentes de audio en el entorno y las características acústicas del entorno. A continuación, se puede generar una representación de audio localmente y presentarla al usuario, por ejemplo, utilizando la renderización y procesamiento binaurales.

- 40 La FIG. 3 ilustra un ejemplo de un sistema de VR en el que un dispositivo 301 de cliente de VR remoto se enlaza con un servidor 303 de VR, por ejemplo, a través de una red 305, tal como Internet. El servidor 303 se puede disponer para admitir simultáneamente un número potencialmente grande de dispositivos 301 de cliente.

- El servidor 303 VR puede, por ejemplo, admitir una experiencia de radiodifusión al transmitir una señal de imagen que comprende una representación de imagen en forma de datos de imagen que pueden ser utilizados por los dispositivos cliente para sintetizar localmente imágenes de vista correspondientes a las poses apropiadas del usuario (una pose se refiere a una posición y/u orientación). Del mismo modo, el servidor 303 VR puede transmitir una representación de audio de la escena, lo que permite sintetizar localmente el audio para las poses del usuario. Específicamente, a medida que el usuario se mueve alrededor del entorno virtual, la imagen y el audio sintetizados y presentados al usuario se actualizan para reflejar la posición y orientación actuales (virtuales) del usuario en el entorno (virtual).

- 60 En muchas aplicaciones, como la de la FIG. 3, puede ser deseable modelar una escena y generar una representación eficiente de imagen y audio que se pueda incluir de manera eficiente en una señal de datos que luego se pueda transmitir o retransmitir a varios dispositivos que puedan sintetizar localmente vistas y audio para diferentes poses que las poses de captura.

En algunas realizaciones, un modelo que representa una escena por ejemplo, se puede almacenar localmente y se puede utilizar localmente para sintetizar imágenes y audio apropiados. Por ejemplo, un modelo de audio de una habitación puede incluir una indicación de las propiedades de las fuentes de audio que se pueden escuchar en la habitación, así como las propiedades acústicas de la habitación. Los datos del modelo se pueden utilizar para sintetizar el audio apropiado para una posición específica.

Es una cuestión crítica cómo se representa la escena de audio y cómo se utiliza esta representación para generar audio. La renderización de audio destinada a proporcionar efectos naturales y realistas a un oyente normalmente incluye la renderización de un entorno acústico. Para muchos entornos, esto incluye la representación y renderización de la reverberación difusa presente en el entorno, tal como en una habitación. Se ha comprobado que la renderización y representación de dicha reverberación difusa tiene un efecto significativo en la percepción del entorno, tal como en si el audio se percibe como un entorno natural y realista. A continuación, se describirán los enfoques ventajosos para representar una escena de audio y para renderizar audio, y en particular audio de reverberación difusa, basados en esta representación.

El enfoque se describirá con referencia a un aparato de audio como se ilustra en la FIG. 4. El aparato de audio se dispone para generar una señal de salida de audio que representa el audio en un entorno acústico. Específicamente, el aparato de audio puede generar audio que representa el audio percibido por un usuario que se mueve alrededor de un entorno virtual con varias fuentes de audio y con determinadas propiedades acústicas. Cada fuente de audio está representada por una señal de audio que representa el sonido de la fuente de audio, así como metadatos que pueden describir las características de la fuente de audio (tal como proporcionar una indicación de nivel para la señal de audio). Además, se proporcionan metadatos para caracterizar el entorno acústico.

El aparato de audio comprende un renderizador 401 de ruta para cada fuente de audio. Cada renderizador 401 de ruta se dispone para generar un componente de señal de ruta directa que representa la ruta directa desde la fuente de audio hasta el oyente. El componente de señal de ruta directo se genera en función de las posiciones del oyente y de la fuente de audio y puede generar específicamente el componente de señal directa al escalar la señal de audio, potencialmente dependiente de la frecuencia, para la fuente de audio dependiente de la distancia y, por ejemplo, la ganancia relativa de la fuente de audio en la dirección específica para el usuario (por ejemplo, para fuentes no omnidireccionales).

En muchas realizaciones, el renderizador 401 también puede generar la señal de ruta directa basada en elementos de oclusión o difracción (virtuales) que se encuentran entre las posiciones de origen y usuario.

En muchas realizaciones, el renderizador 401 de ruta también puede generar componentes de señal adicionales para rutas individuales donde estas incluyen una o más reflexiones. Esto se puede hacer, por ejemplo, al evaluar las reflexiones de las paredes, techo, etc., según lo sepa el experto. Los componentes de ruta directa y ruta reflejada se pueden combinar en una única señal de salida para cada renderizador de ruta y, por lo tanto, se puede generar una única señal que represente la ruta directa y reflexiones tempranas/discretas para cada fuente de audio.

En algunas realizaciones, la señal de audio de salida para cada fuente de audio puede ser una señal binaural y, por lo tanto, cada señal de salida puede incluir tanto una (sub)señal de oído izquierdo como una de oído derecho.

Las señales de salida de los renderizadores 401 de ruta se proporcionan a un combinador 403 que combina las señales de los diferentes renderizadores 401 de ruta para generar una sola señal combinada. En muchas realizaciones, se puede generar una señal de salida binaural y el combinador puede realizar una combinación, tal como una combinación ponderada, de las señales individuales de los renderizadores 401 de ruta, es decir, todas las señales del oído derecho de los renderizadores 401 de ruta se pueden sumar para generar las señales combinadas del oído derecho y todas las señales del oído izquierdo de los renderizadores 401 de ruta se pueden sumar para generar las señales combinadas del oído izquierdo.

Los renderizadores y combinadores de rutas se pueden implementar de cualquier manera adecuada, incluyendo normalmente como código ejecutable para procesamiento en un recurso computacional adecuado, tal como un microcontrolador, microprocesador, procesador de señal digital o unidad central de procesamiento, que incluye circuitos de soporte tales como memoria, etc. Se apreciará que la pluralidad de renderizadores de ruta se pueden implementar como unidades funcionales paralelas, tales como, por ejemplo, un banco de unidades de procesamiento dedicadas, o se pueden implementar como operaciones repetidas para cada fuente de audio. Normalmente, se ejecuta el mismo algoritmo/código para cada fuente/señal de audio.

Además de los componentes de audio de ruta individuales, el aparato de audio se dispone para generar un componente de señal que representa la reverberación difusa en el entorno. La señal de reverberación difusa se genera (de manera eficiente) al combinar las señales de la fuente en una señal de mezcla descendente y luego al aplicar un algoritmo de reverberación a la señal de mezcla descendente para generar la señal de reverberación difusa.

El aparato de audio de la FIG. 4 comprende un mezclador 405 descendente que recibe las señales de audio para una pluralidad de fuentes de sonido (normalmente todas las fuentes dentro del entorno acústico para las cuales el

reverberador está simulando la reverberación difusa), y las combina en una mezcla descendente. De acuerdo con lo anterior, la mezcla descendente refleja todo el sonido generado en el entorno. La mezcla descendente se alimenta a un reverberador 407 que se dispone para generar una señal de reverberación difusa basada en la mezcla descendente. El reverberador 407 puede ser específicamente un reverberador paramétrico tal como un reverberador Jot. El reverberador 407 se acopla al combinador 403 al que se alimenta la señal de reverberación difusa. A continuación, el combinador 403 procede a combinar la señal de reverberación difusa con las señales de ruta que representan las rutas individuales para generar una señal de audio combinada que representa el sonido combinado en el entorno tal y como lo percibe el oyente.

La generación de la señal de reverberación difusa se describirá con más detalle con referencia a un aparato de reverberación de audio como se ilustra en la FIG. 5. El aparato de reverberación de audio se puede incluir en el aparato de audio de la FIG. 4 y puede implementar específicamente el mezclador 405 descendente y el reverberador 407.

El aparato de reverberación de audio comprende un receptor 501 que se dispone para recibir datos de escenas de audio que representan audio. Los datos de la escena de audio incluyen específicamente una pluralidad de señales de audio en las que cada una de las señales de audio representa una fuente de audio (y, por lo tanto, una señal de audio describe el sonido de una fuente de audio). Además, el receptor 501 recibe metadatos para cada una de las fuentes de audio. Estos metadatos incluyen una indicación del nivel de señal (relativa) para la fuente de audio, donde la indicación del nivel de señal puede ser indicativa de un nivel/energía/amplitud de la fuente de sonido representada por la señal de audio. Los metadatos de una fuente incluyen además datos de directividad indicativos de la directividad de la radiación de sonido de la fuente de sonido. Los datos de directividad de una señal de audio pueden, por ejemplo, describir un patrón de ganancia y pueden describir específicamente la ganancia relativa/densidad de energía de la fuente de audio en diferentes direcciones desde la posición de la fuente de audio.

El receptor 501 recibe además metadatos indicativos del entorno acústico. Específicamente, el receptor 501 recibe una relación entre la señal de reverberación difusa y la señal total y, específicamente, una relación entre la señal de reverberación difusa y la señal total (también se puede denominar relación entre el nivel de la señal de reverberación difusa y el nivel de la señal total o, en algunos casos, la relación entre el nivel de la señal de reverberación difusa y la energía total de la señal, o la relación entre la energía emitida y la energía de reverberación difusa) que es indicativa de un nivel de sonido de reverberación difusa con respecto al sonido total emitido en el entorno acústico. La relación entre la señal de reverberación difusa y la señal total también se denominará, en lo sucesivo, por razones de brevedad, la Relación Difusa a la Fuente DSR o, equivalentemente, la Relación entre la Fuente y la Difusa, SDR (la siguiente descripción utilizará predominantemente la primera).

Se apreciará que una relación y una relación inversa pueden proporcionar la misma información, es decir, que cualquier relación se puede expresar como la relación inversa. Por lo tanto, la relación entre la señal de reverberación difusa y la señal total se puede expresar mediante una fracción de un valor que refleje el nivel del sonido de reverberación difusa dividido por un valor que refleje el sonido total emitido, o equivalentemente por una fracción del valor que refleje el sonido total emitido dividido por un valor que refleje el nivel del sonido de reverberación difusa. También se apreciará que se pueden introducir varias modificaciones de los valores estimados, por ejemplo, se puede aplicar una función no lineal (por ejemplo, una función logarítmica).

Cualquier indicación de una relación entre la señal de reverberación difusa y la señal total que sea indicativa de un nivel de sonido de reverberación difusa con respecto al sonido total emitido en el entorno acústico se puede utilizar y proporcionar en los metadatos. La siguiente descripción se centrará en la relación representada por una relación entre un nivel de la señal de reverberación difusa y un nivel (por ejemplo, energía o densidad de energía) de la relación de señal total. Por lo tanto, la descripción se centrará en un ejemplo de una relación entre la señal de reverberación difusa y la señal total que también se denominará como DSR.

El receptor 501 se puede implementar de cualquier manera adecuada, que incluye, por ejemplo, el uso de electrónica discreta o dedicada. El receptor 501, por ejemplo, se puede implementar como un circuito integrado, tal como un Circuito Integrado de Aplicación Específica (ASIC). En algunas realizaciones, el circuito se puede implementar como una unidad de procesamiento programada, tal como por ejemplo como firmware o software que se ejecuta en un procesador adecuado, tal como una unidad central de procesamiento, una unidad de procesamiento de señales digitales o un microcontrolador, etc. Se apreciará que en dichas realizaciones, la unidad de procesamiento puede incluir memoria integrada o externa, circuitos de accionamiento de reloj, circuitos de interfaz, circuitos de interfaz de usuario, etc. Dichos circuitos se pueden implementar además como parte de la unidad de procesamiento, como circuitos integrados y/o como circuitos electrónicos discretos.

El receptor 501 puede recibir los datos de la escena de audio de cualquier fuente adecuada y en cualquier forma adecuada, que incluye, por ejemplo, como parte de una señal de audio. Los datos se pueden recibir desde una fuente interna o externa. El receptor 401 se puede disponer, por ejemplo, para recibir los datos de la habitación a través de una conexión de red, una conexión de radio o cualquier otra conexión adecuada a una fuente interna. En muchas realizaciones, el receptor puede recibir los datos de una fuente local, como una memoria local. En muchas realizaciones, el receptor 501, por ejemplo, se puede disponer para recuperar los datos de habitación de la memoria local, tal como la memoria RAM local o la memoria ROM.

El receptor 501 se puede acoplar a los renderizadores 401 de ruta y puede reenviar los datos de la escena de audio a éstos para la generación de los componentes de la señal de ruta (ruta directa y reflexiones tempranas) como se describió anteriormente.

El aparato de reverberación de audio comprende además el mezclador 405 descendente que también se alimenta de los datos de la escena de audio. El mezclador 405 descendente consiste en un circuito /procesador 505 de energía, un circuito /procesador 507 de coeficiente y un circuito/procesador 509 de mezcla descendente.

El mezclador 405 descendente y, de hecho, cada uno de los circuitos /procesador 505 de energía, circuito /procesador 507 de coeficiente y circuito /procesador 509 de mezcla descendente se pueden implementar de cualquier manera adecuada, que incluye, por ejemplo, el uso de electrónica discreta o dedicada. El receptor 501, por ejemplo, se puede implementar como un circuito integrado, tal como un Circuito Integrado de Aplicación Específica (ASIC). En algunas realizaciones, un circuito/procesador se puede implementar como una unidad de procesamiento programada, tal como por ejemplo como firmware o software que se ejecuta en un procesador adecuado, tal como una unidad central de procesamiento, una unidad de procesamiento de señales digitales o un microcontrolador, etc. Se apreciará que en tales realizaciones, la unidad de procesamiento puede incluir memoria integrada o externa, circuitos de accionamiento de reloj, circuitos de interfaz, circuitos de interfaz de usuario, etc. Dichos circuitos se pueden implementar además como parte de la unidad de procesamiento, como circuitos integrados y/o como circuitos electrónicos discretos.

El procesador 507 de coeficiente se dispone para determinar los coeficientes de mezcla descendente para al menos algunas de las señales de audio recibidas. El coeficiente de mezcla descendente para una señal de audio puede corresponder a una ponderación de esa señal de audio en la mezcla descendente. El coeficiente de mezcla descendente puede ser un peso para la señal de audio en una combinación ponderada que genera la señal de mezcla descendente. Por lo tanto, los coeficientes de mezcla descendente pueden ser pesos relativos para las señales de audio cuando se combinan para generar la señal de mezcla descendente (que en muchas realizaciones es una señal mono), por ejemplo, pueden ser los pesos de una suma ponderada.

El procesador 507 de coeficiente se dispone para generar los coeficientes de mezcla descendente basados en la relación entre la señal de reverberación difusa recibida y la señal total, es decir, la Relación entre Difuso y Fuente, DSR.

Los coeficientes se determinan además en respuesta a una indicación determinada de energía total emitida que es indicativa de la energía total emitida desde la fuente de audio. Mientras que la DSR normalmente es común para algunas señales de audio, y normalmente para todas, la indicación de energía total emitida normalmente es específica de cada fuente de audio.

La indicación de energía total emitida es normalmente indicativa de una energía total emitida normalizada. La misma normalización se puede aplicar a todas las fuentes de audio y a los componentes de ruta directa y reflejada. Por lo tanto, la indicación de la energía total emitida puede ser un valor relativo con respecto a las indicaciones de energía total emitida para otras fuentes/señales de audio o con respecto a los componentes individuales de la ruta o con respecto a un valor de muestra a escala completa de una señal de audio.

La indicación de energía total emitida, cuando se combina con la DSR, puede proporcionar para cada fuente de audio un coeficiente de mezcla descendente que refleja la contribución relativa al sonido de reverberación difusa de esa fuente de audio. Por lo tanto, la determinación del coeficiente de mezcla descendente en función de la DSR y la indicación de la energía total emitida puede proporcionar coeficientes de mezcla descendente que reflejen la contribución relativa al sonido difuso. Por lo tanto, utilizar los coeficientes de mezcla descendente para generar una señal de mezcla descendente puede dar como resultado una señal de mezcla descendente que refleja el sonido total generado en el entorno con cada una de las fuentes de sonido ponderadas adecuadamente y con el entorno acústico modelado con precisión.

En muchas realizaciones, el coeficiente de mezcla descendente en función de la DSR y la indicación de energía total emitida combinada con un escalado en respuesta a las propiedades del reverberador (407) puede proporcionar coeficientes de mezcla descendente que reflejan el nivel relativo apropiado del sonido de reverberación difusa con respecto a los componentes de la señal de ruta correspondientes.

El procesador 505 de energía se acopla al procesador 507 de coeficiente y se dispone para determinar las indicaciones de energía total emitida a partir de los metadatos recibidos para las fuentes de audio.

Los metadatos recibidos incluyen un nivel de referencia de señal para cada fuente que proporciona una indicación de un nivel de audio. El nivel de referencia de la señal es normalmente un valor normalizado o relativo que proporciona una indicación del nivel de referencia de la señal con respecto a otras fuentes de audio o con respecto a un nivel de referencia normalizado. Por lo tanto, el nivel de referencia de la señal normalmente no indica el nivel de sonido absoluto de una fuente, sino más bien un nivel relativo con respecto a otras fuentes de audio.

En el ejemplo específico, el nivel de referencia de la señal puede incluir una indicación en forma de distancia de referencia que proporciona una distancia para la cual se debe aplicar una atenuación de distancia a la señal de audio de 0 dB. Por lo tanto, para una distancia entre la fuente de audio y el oyente igual a la distancia de referencia, la señal de audio recibida se puede utilizar sin ningún escalado dependiente de la distancia. Para una distancia menor que la distancia de referencia, la atenuación es menor y, por lo tanto, se debe aplicar una ganancia mayor de 0 dB al determinar el nivel de sonido en la posición de escucha. Para una distancia mayor de la distancia de referencia, la atenuación es mayor y, por lo tanto, se debe aplicar una atenuación mayor de 0 dB al determinar el nivel de sonido en la posición de escucha. De manera equivalente, para una distancia dada entre la fuente de audio y la posición de escucha, se aplicará una ganancia más alta a una señal de audio asociada con una distancia de referencia más alta que una asociada con una distancia de referencia más corta. Dado que la señal de audio se normaliza normalmente para representar una distancia de referencia significativa o para explotar toda la gama dinámica (por ejemplo, un motor a reacción y un grillo estarán representados por señales de audio que explotan todo el rango dinámico de las palabras de datos utilizadas), la distancia de referencia proporciona una indicación del nivel de referencia de la señal para la fuente de audio específica.

En el ejemplo, el nivel de referencia de la señal se indica además mediante una ganancia de referencia denominada ganancia previa. La ganancia de referencia se proporciona para cada fuente de audio y proporciona una ganancia que se debe aplicar a la señal de audio al determinar los niveles de audio renderizados. Por lo tanto, la ganancia previa se puede utilizar para indicar aún más las variaciones de nivel entre las diferentes fuentes de audio.

Los metadatos incluyen además datos de directividad que son indicativos de la directividad de la radiación de sonido de la fuente de sonido representada por la señal de audio. Los datos de directividad de cada fuente de audio pueden ser indicativos de una ganancia relativa, con respecto al nivel de referencia de la señal, en direcciones diferentes de la fuente de audio. Los datos de directividad pueden, por ejemplo, proporcionar una función completa o una descripción de un patrón de radiación procedente de la fuente de audio que defina la ganancia en cada dirección. Como otro ejemplo, se puede utilizar una indicación simplificada, tal como, por ejemplo, un único valor de datos que indique un patrón predeterminado. Como otro ejemplo, los datos de directividad pueden proporcionar valores de ganancia individuales para un rango de diferentes intervalos de dirección (por ejemplo, segmentos de una esfera).

Por lo tanto, los metadatos, junto con las señales de audio, pueden permitir la generación de niveles de audio. Específicamente, los renderizadores de ruta pueden determinar el componente de la señal para la ruta directa al aplicar una ganancia a la señal de audio donde la ganancia es una combinación de la ganancia previa, una ganancia de distancia determinada en función de la distancia entre la fuente de audio y el oyente y la distancia de referencia, y la ganancia de directividad en la dirección desde la fuente de audio hasta el oyente.

Con respecto a la generación de la señal de reverberación difusa, los metadatos se utilizan para determinar una indicación de energía total emitida (normalizada) para una fuente de audio basada en el nivel de referencia de la señal y los datos de directividad para la fuente de audio.

Específicamente, la indicación de energía total emitida se puede generar al integrar la ganancia de directividad en todas las direcciones (por ejemplo, integrando sobre una superficie de una esfera centrada en la posición de la fuente de audio) y escalar por el nivel de referencia de la señal, y específicamente por la ganancia de distancia y la ganancia previa.

La indicación de energía total emitida determinada se alimenta al procesador 507 de coeficiente, donde se procesa con la DSR para generar los coeficientes de mezcla descendente.

Los coeficientes de mezcla descendente se utilizan por el procesador 509 de mezcla descendente para generar la señal de mezcla descendente. Específicamente, la señal de mezcla descendente se puede generar como una combinación, y específicamente una suma, de las señales de audio, con cada señal de audio ponderada por el coeficiente de mezcla descendente para la señal de audio correspondiente.

La mezcla descendente se genera normalmente como una señal mono que luego se alimenta al reverberador 407 que procede a generar la señal de reverberación difusa.

Se debe tener en cuenta que, mientras que la renderización y generación de los componentes individuales de la señal de ruta por los renderizadores 401 de ruta depende de la posición, por ejemplo, con respecto a la determinación de la ganancia de distancia y la ganancia de directividad, luego la generación de la señal de reverberación difusa puede ser independiente de la posición tanto de la fuente como del oyente.

La indicación de la energía total emitida se puede determinar en base al nivel de referencia de la señal y de los datos de directividad sin tener en cuenta las posiciones de la fuente y el oyente. Específicamente, la ganancia previa y la distancia de referencia para una fuente se pueden utilizar para determinar un nivel de referencia de señal no dependiente de la directividad a una distancia nominal de la fuente (la distancia nominal es la misma para todas las señales/fuentes de audio) y que se normaliza con respecto a, por ejemplo, una muestra a escala completa de las señales de audio. La integración de las ganancias de directividad en todas las direcciones se puede realizar para una

esfera normalizada, como por ejemplo, para una esfera a la distancia de referencia. Por lo tanto, la indicación de energía total emitida será independiente de la fuente y de la posición del oyente (lo que refleja que el sonido de reverberación difusa tiende a ser homogéneo en un entorno tal como una habitación). A continuación, la indicación de energía total emitida se combina con la DSR para generar los coeficientes de mezcla descendente (en muchas realizaciones también se pueden considerar otros parámetros, tales como los parámetros del reverberador). Como la DSR también es independiente de las posiciones, al igual que la mezcla descendente y el procesamiento de reverberación, la señal de reverberación difusa se puede generar sin tener en cuenta las posiciones específicas de la fuente y el oyente.

Dicho enfoque puede proporcionar un alto rendimiento y una percepción de audio que suene de forma natural sin requerir un recurso computacional excesivo. Puede ser especialmente adecuado para, por ejemplo, aplicaciones de realidad virtual en las que un usuario (y las fuentes) se pueden mover en el entorno y, por lo tanto, en las que las posiciones relativas del oyente (y posiblemente algunas o todas las fuentes de audio) pueden cambiar dinámicamente.

En los siguientes aspectos específicos de varias realizaciones del enfoque de las FIG. 4 y 5 se describirán con más detalle.

En muchas realizaciones, los metadatos pueden comprender además una indicación de cuándo debe comenzar la señal de reverberación difusa, es decir, puede indicar un retardo de tiempo asociado con la señal de reverberación difusa. La indicación de retardo de tiempo puede estar específicamente en forma de retardo previo.

El retardo previo puede representar el retardo/desfase en una RIR y se puede definir como el umbral entre las reflexiones tempranas y la reverberación difusa y tardía. Dado que este umbral se produce normalmente como parte de una transición suave de reflexiones (más o menos) discretas en una mezcla de reflexiones de orden superior totalmente interferentes, se puede seleccionar un umbral adecuado utilizando un proceso de evaluación/decisión adecuado. La determinación se puede hacer automática en base al análisis de la RIR, o calcular en base a las dimensiones de la habitación y/o las propiedades del material.

Alternativamente, se puede elegir un umbral fijo, como, por ejemplo, 80 ms en la RIR. El retardo previo se puede indicar en segundos, milisegundos o muestras. En la siguiente descripción, se supone que el retardo previo se elige para estar en un punto después del cual la reverberación es realmente difusa. Sin embargo, el método descrito puede funcionar lo suficiente si este no es el caso.

Por lo tanto, el retardo previo es indicativo del inicio de la respuesta de reverberación difusa desde el inicio de la emisión de la fuente. Por ejemplo, para un ejemplo como el mostrado en la FIG. 6, si una fuente comienza a emitir en t_0 (por ejemplo, $t_0 = 0$), el sonido directo llega al usuario en $t_1 > t_0$, la primera reflexión llega al usuario en $t_2 > t_1$ y el umbral definido entre las reflexiones tempranas y la reverberación difusa llega al usuario en $t_3 > t_2$. Entonces el retardo previo es $t_3 - t_0$.

En el sistema, la relación entre la señal de reverberación difusa y la señal total, es decir, la Relación Difusa a Fuente, DSR, se puede utilizar para expresar la cantidad de energía de reverberación difusa o el nivel de una fuente recibida por un usuario como una relación de la energía total emitida de esa fuente. Se puede expresar de tal manera que la energía de reverberación difusa esté condicionada adecuadamente para la calibración del nivel de las señales que se van a renderizar y los metadatos correspondientes (por ejemplo, ganancia previa).

Expresarlo de esta manera puede asegurar que el valor sea independiente de las posiciones y orientaciones absolutas del oyente y la fuente en el entorno, independiente de la posición y orientación relativas del usuario con respecto a la fuente y viceversa, independiente de algoritmos específicos para renderizar la reverberación, y que exista un vínculo significativo con los niveles de señal utilizados en el sistema.

El enfoque descrito calcula los coeficientes de mezcla descendente que tienen en cuenta tanto los patrones de directividad para imponer los niveles relativos correctos entre las señales de la fuente, como la DSR para lograr el nivel correcto en la salida del reverberador 407.

La DSR puede representar la relación entre la energía de la fuente emitida y una propiedad de reverberación difusa, tal como específicamente la energía o el nivel (inicial) de la señal de reverberación difusa.

La descripción se centrará principalmente en una DSR indicativa de la energía de reverberación difusa con respecto a la energía total:

$$DSR = \frac{\text{energía de reverberación difusa}}{\text{energía total emitida}}$$

La energía de reverberación difusa se puede considerar como la energía creada por la respuesta de la habitación desde el inicio de la sección difusa, por ejemplo, puede ser la energía de la RIR desde un tiempo indicado por retardo previo hasta el infinito. Tenga en cuenta que las excitaciones posteriores de la habitación se sumarán a la energía de

reverberación, por lo que normalmente solo se puede medir directamente mediante la excitación con un pulso de Dirac. Alternativamente, se puede derivar de una RIR medida.

La energía de reverberación representa la energía en un solo punto en el espacio de campo difuso en lugar de estar integrada en todo el espacio.

Una alternativa particularmente ventajosa a la anterior sería utilizar una DSR que sea indicativa de una amplitud inicial de sonido difuso con respecto a una energía del sonido total emitido en el entorno. Específicamente, la DSR puede indicar la amplitud de la reverberación en el momento indicado por el retardo previo.

La amplitud en el retardo previo puede ser la mayor excitación de la respuesta al impulso de la habitación en o muy cerca después del retardo previo. Por ejemplo, dentro de los 5, 10, 20 o 50 ms después del retardo previo. La razón para elegir la excitación más grande en un cierto rango es que en el tiempo de retardo previo, la respuesta al impulso de la habitación puede estar casualmente en una parte baja de la respuesta. Dado que la tendencia general es una amplitud decreciente, la excitación más grande en un intervalo corto después del retardo previo normalmente también es la excitación más grande de toda la respuesta de reverberación difusa.

El uso de una DSR indicativa de una amplitud inicial (dentro de un intervalo de, por ejemplo, 10 mseg) hace que sea más fácil y robusto mapear la DSR a los parámetros en muchos algoritmos de reverberación. Por lo tanto, en algunas realizaciones, la DSR puede darse como:

$$DSR = \frac{\text{amplitud inicial de reverberación difusa}}{\text{energía total emitida}}$$

Los parámetros de la DSR se expresan con respecto a la misma referencia de nivel de señal de fuente.

Esto se puede lograr, por ejemplo, al medir (o simular) la RIR de la habitación de interés con un micrófono dentro de ciertas condiciones conocidas (tales como la distancia entre la fuente y el micrófono y el patrón de directividad de la fuente). La fuente debe emitir una cantidad calibrada de energía en la habitación, por ejemplo, un impulso de Dirac con energía conocida.

Un factor de calibración para conversiones eléctricas en el equipo de medición y conversión de analógico a digital se puede medir o derivar de especificaciones. También se puede calcular a partir de la respuesta directa de la ruta en la RIR, que es predecible a partir del patrón de directividad de la fuente y la distancia fuente-micrófono. La respuesta directa tiene una cierta energía en el dominio digital y representa la energía emitida multiplicada por la ganancia de directividad para la dirección del micrófono y una ganancia de distancia que puede depender de la superficie del micrófono con respecto al área total de la superficie de la esfera con un radio igual a la distancia fuente-micrófono.

Ambos elementos deben utilizar la misma referencia de nivel digital. Por ejemplo, un seno de escala completa de 1 kHz corresponde a 100 dB SPL.

Medir la energía de reverberación difusa de la RIR y compensarla con el factor de calibración da la energía apropiada en el mismo dominio que la energía emitida conocida. Junto con la energía emitida, se puede calcular la DSR adecuada.

La distancia de referencia puede indicar la distancia a la que la ganancia de distancia que se debe aplicar a la señal es de 0 dB, es decir, cuando no se debe aplicar ganancia o atenuación para compensar la distancia. La ganancia de distancia real que deben aplicar los renderizadores 401 de ruta se puede calcular al tener en cuenta la distancia real con respecto a la distancia de referencia.

La representación del efecto de la distancia a la propagación del sonido se realiza con referencia a una distancia dada. Una duplicación de la distancia reduce la densidad de energía (energía por unidad de superficie) en 6 dB. Una reducción a la mitad de la distancia induce la densidad de energía (energía por unidad de superficie) en 6 dB.

Para determinar la ganancia de distancia a una distancia dada, se debe conocer la distancia correspondiente a un nivel dado para que se pueda determinar la variación relativa de la distancia actual, es decir, para determinar cuánto se ha reducido o aumentado la densidad.

Ignorando la absorción en el aire y asumiendo que no hay reflexiones o elementos de oclusión presentes, la energía emitida de una fuente es constante en cualquier esfera con cualquier radio centrado en la posición de la fuente. La relación entre la superficie correspondiente a la distancia real y la distancia de referencia indica la atenuación de la energía. La ganancia de amplitud de la señal lineal a la distancia de renderización d se puede representar b:

$$g_{\text{dist}} = \sqrt{\frac{A_{\text{ref}}}{A_{\text{rend}}}} = \sqrt{\frac{4\pi r_{\text{ref}}^2}{4\pi d^2}} = \frac{r_{\text{ref}}}{d}$$

donde r_{ref} es la distancia de referencia.

- 5 Como un ejemplo, esto da como resultado una atenuación de la señal de aproximadamente 6 dB (o una ganancia de -6 dB) si la distancia de referencia es de 1 metro y la distancia de renderización es de 2 metros.

La indicación de energía total emitida puede representar la energía total que emite una fuente de sonido. Normalmente, las fuentes de sonido irradian en todas las direcciones, pero no por igual en todas las direcciones. Una integración de densidades de energía sobre una esfera alrededor de una fuente puede proporcionar la energía total emitida. En el caso de un altavoz, la energía emitida a menudo se puede calcular con el conocimiento del voltaje aplicado a los terminales y los coeficientes del altavoz que describen la impedancia, las pérdidas de energía y la transferencia de energía eléctrica a las ondas de presión de sonido.

15 El procesador 505 de energía se dispone para determinar la indicación de energía total emitida considerando los datos de directividad para la fuente de audio. Cabe señalar que es importante utilizar la energía total emitida en lugar de sólo el nivel de la señal o el nivel de referencia de la señal al determinar las señales de reverberación difusa para fuentes que pueden tener una directividad de fuente variable. Por ejemplo, considérese una directividad de la fuente correspondiente a un haz muy estrecho con un coeficiente de directividad 1, y con un coeficiente de 0 para todas las demás direcciones (es decir, la energía sólo se transmite en el haz muy estrecho). En este caso, la energía de la fuente emitida puede ser muy similar a la energía de la señal de audio y al nivel de referencia de la señal, ya que es representativa de la energía total. Si, en cambio, se considera otra fuente con una señal de audio con la misma energía y nivel de referencia de señal, pero con una directividad omnidireccional, la energía emitida por esta fuente será mucho mayor que la energía de la señal de audio y los niveles de referencia de la señal. Por lo tanto, con ambas fuentes activas al mismo tiempo, la señal de la fuente omnidireccional se debe representar mucho más fuerte en la señal de reverberación difusa y, por lo tanto, en la mezcla descendente, que la fuente muy direccional.

Como se mencionó, el procesador 505 de energía puede determinar la energía emitida al integrar la densidad de energía sobre la superficie de una esfera que rodea la fuente de audio. Ignorando la ganancia de distancia, es decir, integrando sobre una superficie para un radio donde la ganancia de distancia es de 0 dB (es decir, con un radio correspondiente a la distancia de referencia), se puede determinar una indicación de energía total emitida a partir de:

$$E = \iint_S (g \cdot p \cdot x)^2 dS$$

35 donde, g es la función de ganancia de directividad, p es la ganancia previa asociada con la señal/fuente de audio y x indica el nivel de la señal de audio en sí mismo.

Dado que p es independiente de la dirección, también se puede mover fuera de la integral. Del mismo modo, la señal x es independiente de la dirección (la ganancia de directividad refleja esa variación). (Se puede multiplicar más tarde, ya que:

$$g_1^2 * x^2 + g_2^2 * x^2 + g_3^2 * x^2 + \dots = (g_1^2 + g_2^2 + g_3^2 + \dots) * x^2$$

y así la integral se independiza de la señal).

45 Un enfoque específico para determinar esta integral se describe con más detalle a continuación.

Se desea integrar las ganancias de directividad sobre una esfera.

$$E_{\text{escala}} = \iint g^2 dS$$

50 El uso de una esfera con un radio igual a la distancia de referencia (r) significa que la ganancia de distancia es de 0 dB y, por lo tanto, se puede ignorar la ganancia/atenuación de la distancia.

55 En este ejemplo se elige una esfera porque proporciona un cálculo ventajoso, pero la misma energía se puede determinar a partir de cualquier superficie cerrada de cualquier conformación que encierre la posición de la fuente. Siempre y cuando se utilice la ganancia de distancia y la ganancia de directividad apropiadas en la integral, y se

considere la superficie efectiva que está orientada hacia la posición de la fuente (es decir, con un vector normal en línea con la posición de la fuente).

Una integral de superficie debe definir una pequeña superficie dS . Por lo tanto, definir una esfera con dos parámetros, acimut (a) y elevación (e) proporciona las dimensiones para hacerlo. Utilizando el sistema de coordenadas para nuestra solución obtenemos:

$$f(a, e, r) = r * \cos(e) * \cos(a) * u_x + r * \cos(e) * \sin(a) * u_y + r * \sin(e) * u_z$$

donde u_x , u_y y u_z son vectores base unitarios del sistema de coordenadas.

La pequeña superficie dS es la magnitud del producto cruzado de las derivadas parciales de la superficie de la esfera con respecto a los dos parámetros, multiplicado por los diferenciales de cada parámetro:

$$dS = |f_a \times f_e| da de$$

Las derivadas determinan los vectores tangenciales a la esfera en el punto de interés.

$$f_a = -r * \sin(e) * \sin(a) * u_x + r * \sin(e) * \cos(a) * u_y + 0 * u_z$$

$$f_e = -r * \cos(e) * \cos(a) * u_x - r * \cos(e) * \sin(a) * u_y + r * \sin(e) * u_z$$

El producto cruzado de las derivadas es un vector perpendicular a ambos.

$$f_a \times f_e = (r^2 * \sin(e) * \cos(a) * \cos(e) + 0 * \sin(e) * \sin(a)) * u_x + (-0 * \sin(e) * \cos(a) + r^2 * \sin(e) * \sin(a) * \cos(e)) * u_y + (r^2 * \sin(e) * \sin(a) * \sin(e) + r^2 * \cos(e) * \cos(a) * \sin(e) * \cos(a)) * u_z$$

$$\begin{aligned} &= r^2 * \cos^2(e) * \cos(a) * u_x + r^2 * \cos^2(e) * \sin(a) * u_y + (r^2 * \cos(e) * \sin(e) * (\sin^2(a) + \cos^2(a))) * u_z \\ &= r^2 * \cos^2(e) * \cos(a) * u_x + r^2 * \cos^2(e) * \sin(a) * u_y + (r^2 * \cos(e) * \sin(e) * 1) * u_z \\ &= r^2 * \cos^2(e) * \cos(a) * u_x + r^2 * \cos^2(e) * \sin(a) * u_y + r^2 * \cos(e) * \sin(e) * u_z \end{aligned}$$

La magnitud del producto cruzado es el área de la superficie del paralelogramo abarcada por los vectores f_a y f_e , y por lo tanto el área de la superficie en la esfera:

$$\begin{aligned} |f_a \times f_e| &= \sqrt{(r^2 * \cos^2(e) * \cos(a))^2 + (r^2 * \cos^2(e) * \sin(a))^2 + (r^2 * \cos(e) * \sin(e))^2} \\ &= \sqrt{r^4 * \cos^4(e) * \cos^2(a) + r^4 * \cos^4(e) * \sin^2(a) + r^4 * \cos^2(e) * \sin^2(e)} \\ &= \sqrt{r^4 * \cos^4(e) * (\cos^2(a) + \sin^2(a)) + r^4 * \cos^2(e) * \sin^2(e)} \\ &= \sqrt{r^4 * \cos^4(e) + r^4 * \cos^2(e) * \sin^2(e)} \\ &= \sqrt{r^4 * \cos^2(e) * (\cos^2(e) + \sin^2(e))} \\ &= \sqrt{r^4 * \cos^2(e)} \\ &= \text{abs}(r^2 * \cos(e)) = r^2 * \cos(e) \text{ cuando } e = [-0.5 * \pi, 0.5 * \pi] \end{aligned}$$

Resultando en:

$$dS = r^2 * \cos(e) * da * de$$

donde los dos primeros términos definen un área de superficie normalizada, y con la multiplicación por da y de se convierte en una superficie real, basada en el tamaño de los segmentos da y de . La doble integral sobre la superficie se puede expresar en términos de acimut y elevación. La superficie dS se expresa, según lo anterior, en términos de a y e . Las dos integrales se pueden realizar sobre acimut = 0 ... $2 * \pi$ (integral interna) y elevación = $-0.5 * \pi$... $0.5 * \pi$ (integral externa).

$$E_{\text{escala}} = \int_{-0.5\pi}^{0.5\pi} \int_0^{2\pi} g^2(a, e) \cdot r^2 \cdot \cos(e) \cdot da \cdot de$$

donde $g(a, e)$ es la directividad en función del acimut y la elevación. Por lo tanto, si $g(a, e) = 1$, el resultado debe ser la superficie de la esfera. (Calcular la integral analíticamente como una demostración da como resultado $4 \cdot \pi \cdot r^2$ como se esperaba).

En muchas realizaciones prácticas, el patrón de directividad puede no proporcionarse como una función integrable, sino, por ejemplo, como un conjunto discreto de puntos de muestreo. Por ejemplo, cada ganancia de directividad muestreada se asocia con un acimut y una elevación. Normalmente, estos ejemplos representarán una cuadrícula en una esfera. Un enfoque para manejar esto es convertir las integrales en sumas, es decir, se puede realizar una integración discreta. En este ejemplo, la integración se puede implementar como una suma de puntos de la esfera para los que se dispone de una ganancia de directividad. Esto da los valores de $g(a, e)$, pero requiere que a y e se elijan correctamente, de tal manera que no den lugar a grandes errores debido a superposiciones o huecos.

En otras realizaciones, el patrón de directividad se puede proporcionar como un número limitado de puntos no uniformemente espaciados en el espacio. En este caso, el patrón de directividad se puede interpolar y volver a muestrearse uniformemente en el rango de acimuts y elevaciones de interés.

Una solución alternativa puede ser asumir que $g(a, e)$ es constante alrededor de su punto definido y resolver la integral localmente analíticamente. Por ejemplo, para un pequeño rango de acimut y elevación. Por ejemplo, a medio camino entre los puntos definidos vecinos. Esto utiliza la integral anterior, pero con diferentes rangos de a y e , y $g(a, e)$ considerado constante.

Los experimentos muestran que con la suma directa, los errores son pequeños, incluso con una resolución bastante aproximada de la directividad. Además, los errores son independientes del radio. Para un espaciado lineal del acimut entre 10 puntos y 10 puntos de elevación espaciados linealmente, se obtiene un error relativo de -20 dB.

La integral, como se expresó anteriormente, proporciona un resultado que escala con el radio de la esfera. Por lo tanto, se escala con la distancia de referencia. Esta dependencia del radio se debe a que no tenemos en cuenta el efecto inverso de la 'ganancia de distancia' entre los dos radios diferentes. Si el radio se duplica, la energía que 'fluye' a través de una superficie fija (por ejemplo, 1 cm²) es 6 dB menor. Por lo tanto, se podría decir que la integración debe tener en cuenta la ganancia de distancia. Sin embargo, la integración se realiza a la distancia de referencia, que se define como la distancia a la que la ganancia de distancia se refleja en la señal. En otras palabras, el nivel de señal indicado por la distancia de referencia no se incluye como un escalado del valor que se está integrando, sino que se refleja en la superficie sobre la que se realiza la integración que varía con la distancia de referencia (ya que la integración se realiza sobre una esfera con un radio igual a la distancia de referencia).

Como resultado, la integral, como se ha descrito anteriormente, refleja un factor de escalado de energía de la señal de audio (que incluye cualquier ganancia previa o ajuste de calibración similar), porque la señal de audio representa la energía de reproducción de la señal correcta en un área de superficie fija en la esfera con un radio igual a la distancia de referencia (sin ganancia de directividad).

Esto significa que si la distancia de referencia es mayor, sin cambiar la señal, el factor de escalado de energía total de la señal también es mayor. Esto se debe a que la señal correspondiente representa una fuente de sonido que es relativamente más fuerte que una con la misma energía de señal, pero a una distancia de referencia más pequeña.

En otras palabras, al realizar la integración sobre la superficie de una esfera con un radio igual a la distancia de referencia, se tiene en cuenta automáticamente la indicación del nivel de señal proporcionada por la distancia de referencia. Una distancia de referencia más alta dará como resultado una mayor superficie y, por lo tanto, una indicación de energía total emitida más grande. La integración se realiza específicamente directamente a distancia para la que la ganancia de distancia es 1.

La integral anterior da como resultado valores que se normalizan a la unidad de superficie utilizada y a la unidad utilizada para indicar la distancia de referencia r . Si la distancia de referencia r se expresa en metros, entonces el resultado de la integral se proporciona en la unidad de m².

Para relacionar el valor estimado de la energía emitida con la señal, se debe expresar en una unidad de superficie que corresponda a la señal. Dado que el nivel de la señal representa el nivel al que debe reproducirse para un usuario a una distancia de referencia, el área de superficie de un oído humano puede ser más adecuada. A la distancia de referencia, esta superficie relativa a toda la superficie de la esfera estaría relacionada con la porción de energía de fuente que se percibe.

Por lo tanto, una indicación de energía total emitida que representa la energía de la fuente emitida normalizada para muestras a escala completa en la señal de audio se puede indicar por:

$$E_{norm} = \frac{E_{dir,r} * p^2}{S_{oído}}$$

5 donde $E_{dir,r}$ indica la energía determinada mediante la integración de la ganancia de directividad sobre la superficie de una esfera con un radio igual a la distancia de referencia, p es la ganancia previa y $S_{oído}$ es un factor de escala de normalización (para relacionar la energía determinada con el área de un oído humano).

10 Con la DSR que caracteriza las propiedades acústicas difusas de un espacio y la energía de fuente emitida calculada derivada de los metadatos de directividad, ganancia previa y distancia de referencia, se puede calcular la energía de reverberación correspondiente.

15 Normalmente, la DSR se puede determinar con los mismos niveles de referencia utilizados por ambos componentes. Esto puede o no ser lo mismo que la indicación de energía total emitida. En cualquier caso, cuando dicha DSR se combina con la indicación de energía total emitida, la energía de reverberación resultante también se expresa como una energía que se normaliza para muestras a escala completa en la señal de audio, cuando se utiliza la energía total emitida determinada por la integración anterior. En otras palabras, todas las energías consideradas están esencialmente normalizadas a los mismos niveles de referencia, de tal manera que se pueden combinar directamente sin necesidad de ajustes de nivel. Específicamente, la energía total emitida determinada se puede utilizar directamente con la DSR para generar una indicación de nivel para la reverberación difusa generada por cada fuente, donde la indicación de nivel indica directamente el nivel apropiado con respecto a la reverberación difusa para otras fuentes de audio y con respecto a los componentes individuales de la señal de ruta.

25 Como ejemplo específico, los niveles relativos de la señal para los componentes de la señal de reverberación difusa para las diferentes fuentes se pueden obtener directamente al multiplicar la DSR por la indicación de energía total emitida.

30 En el sistema descrito, la adaptación de las contribuciones de diferentes fuentes de audio a la señal de reverberación difusa se realiza, al menos parcialmente, al adaptar los coeficientes de mezcla descendente utilizados para generar la señal de mezcla descendente. Por lo tanto, los coeficientes de mezcla descendente se pueden generar de tal manera que las contribuciones/niveles de energía relativos del sonido difuso de cada fuente de audio reflejen la energía de reverberación difusa determinada para las fuentes.

35 Como ejemplo específico, si la DSR indica un nivel de amplitud inicial, se puede determinar que los coeficientes de mezcla descendente son proporcionales (o iguales) a la DSR multiplicada por la indicación de energía total emitida. Si la DSR indica un nivel de energía, se puede determinar que los coeficientes de mezcla descendente son proporcionales (o iguales) a la raíz cuadrada de la DSR multiplicada por la indicación de energía total emitida.

40 Como ejemplo específico, un coeficiente de mezcla descendente d_x para proporcionar un ajuste apropiado para la señal con índice x de la pluralidad de señales de entrada se puede calcular por:

$$d_x = p \cdot \sqrt{E_{norm,x} * DSR}$$

45 donde p denota la ganancia previa y $E_{norm,x} = \frac{E_{escala}}{S_{oído}}$ la energía de fuente emitida normalizada para la señal x , antes de la ganancia previa. La DSR representa la relación entre la energía de reverberación difusa y la energía de la fuente emitida. Cuando se aplica el coeficiente de mezcla descendente d_x a la señal de entrada x , la señal resultante representa un nivel de señal que, cuando se filtra por un reverberador que tiene una respuesta de reverberación de energía unitaria, proporciona la energía de reverberación difusa correcta para la señal x con respecto a la renderización directa de la ruta de la señal x y con respecto a las rutas directas y las energías de reverberación difusa de otras fuentes $j \neq x$.

Alternativamente, se puede calcular un coeficiente de mezcla descendente d_x de acuerdo con:

55
$$d_x = E_{norm,x} * DSR$$

60 donde $E_{norm,x} = \frac{E_{dir,r} * p^2}{S_{oído}}$ denota la energía de la fuente embotada normalizada para la señal x y DSR representa la relación entre la energía de reverberación difusa y la amplitud de respuesta de reverberación inicial. Cuando se aplica el coeficiente de mezcla descendente d_x a la señal de entrada x , la señal resultante representa un nivel de señal que corresponde al nivel inicial de la señal de reverberación difusa y puede ser procesada por un reverberador que tiene una respuesta de reverberación que comienza con la amplitud 1. Como resultado, la salida del reverberador proporciona la energía de reverberación difusa correcta para la señal x con respecto a la renderización de la ruta directa de la señal x y con respecto a las rutas directas y las energías de reverberación difusa de otras fuentes $j \neq x$.

En muchas realizaciones, los coeficientes de mezcla descendente se determinan parcialmente al combinar la DSR con la indicación de energía total emitida. Ya sea que la DSR indique la relación entre la energía total emitida y la energía de reverberación difusa o una amplitud inicial para la respuesta de reverberación difusa, a menudo es necesaria una adaptación adicional de los coeficientes de mezcla descendente para adaptarse al algoritmo de reverberador específico utilizado que escala las señales de tal manera que la salida del procesador de reverberación refleje la energía deseada o la amplitud inicial. Por ejemplo, la densidad de las reflexiones en un algoritmo de reverberación tiene una fuerte influencia sobre la energía de reverberación producida, mientras que el nivel de entrada permanece igual. Como otro ejemplo, la amplitud inicial del algoritmo de reverberación puede no ser igual a la amplitud de su excitación. Por lo tanto, puede ser necesario un ajuste específico a algoritmo o específico a algoritmo y configuración. Esto se puede incluir en los coeficientes de mezcla descendente y, normalmente es común a todas las fuentes. Para algunas realizaciones, estos ajustes se pueden aplicar a la mezcla descendente o incluir en el algoritmo del reverberador.

Una vez que se generan los coeficientes de mezcla descendente, el procesador 509 de mezcla descendente puede generar la señal de mezcla descendente, por ejemplo, mediante una combinación o suma ponderada directa.

Una ventaja del enfoque descrito es que puede utilizar un reverberador convencional. Por ejemplo, el reverberador 407 se puede implementar por una red de retardo de retroalimentación, como por ejemplo implementar en un reverberador Jot estándar.

Como se ilustra en la FIG. 7, el principio de las redes de retardo de retroalimentación utiliza uno o más (normalmente más de uno) bucles de retroalimentación con diferentes retardos. Una señal de entrada, en este caso la señal de mezcla descendente, se alimenta a bucles donde las señales se retroalimentan con ganancias de retroalimentación adecuadas. Las señales de salida se extraen al combinar señales en los bucles. Por lo tanto, las señales se repiten continuamente con diferentes retardos. El uso de retardos que son mutuamente primos y tener una matriz de retroalimentación que mezcla señales entre bucles puede crear un patrón similar a la reverberación en espacios reales.

El valor absoluto de los elementos en la matriz de retroalimentación debe ser menor que 1 para lograr una respuesta de impulso estable y decreciente. En muchas implementaciones, se incluyen ganancias o filtros adicionales en los bucles. Estos filtros pueden controlar la atenuación en lugar de la matriz. El uso de filtros tiene la ventaja de que la respuesta de decaimiento puede ser diferente para diferentes frecuencias.

En algunas realizaciones en las que la salida del reverberador se renderiza binauralmente, la reverberación estimada se puede filtrar por las HRTF (Funciones de Transferencia Relacionadas con la Cabeza) promedio para los oídos izquierdo y derecho, respectivamente, con el fin de producir las señales de reverberación de los canales izquierdo y derecho. Cuando las HRTF están disponibles para más de una distancia a intervalos uniformemente espaciados en una esfera alrededor del usuario, se puede apreciar que las HRTF promedio para los oídos izquierdo y derecho se generan utilizando el conjunto de HRTF con la mayor distancia. El uso de HRTF promedio se puede basar/reflejar en la consideración de que la reverberación es isotrópica y proviene de todas las direcciones. Por lo tanto, en lugar de incluir un par de HRTF para una dirección determinada, se puede utilizar un promedio de todos los HRTF. El promedio se puede realizar una vez para el oído izquierdo y otro para el derecho, y los filtros resultantes se pueden utilizar para procesar la salida del reverberador para la renderización binaural.

En algunos casos, el reverberador puede introducir coloración de las señales de entrada, lo que lleva a una salida que no tiene la energía de señal difusa de salida deseada como se describe en la DSR. Por lo tanto, los efectos de este proceso también pueden igualarse. Esta ecualización se puede realizar en base a filtros que se determinan analíticamente como la inversa de la respuesta de frecuencia de la operación del reverberador. En algunas realizaciones, la función de transferencia se puede estimar utilizando técnicas de aprendizaje de estimación automática como la regresión lineal, el ajuste de líneas, etc.

En algunas realizaciones, el mismo enfoque se puede aplicar uniformemente a toda la banda de frecuencias. Sin embargo, en otras realizaciones, se puede realizar un procesamiento dependiente de la frecuencia. Por ejemplo, uno o varios de los parámetros de metadatos proporcionados pueden depender de la frecuencia. En dicho ejemplo, el aparato se puede disponer para dividir las señales en diferentes bandas de frecuencia correspondientes a la dependencia de la frecuencia y el procesamiento descrito anteriormente se puede realizar individualmente en cada una de las bandas de frecuencia.

Específicamente, en algunas realizaciones, la relación entre la señal de reverberación difusa y la señal total, DSR, depende de la frecuencia. Por ejemplo, se pueden proporcionar diferentes valores de DSR para un rango de bandas/intervalos de frecuencias discretos o la DSR se puede proporcionar en función de la frecuencia. En dicha realización, el aparato se puede disponer para generar coeficientes de mezcla descendente dependientes de la frecuencia que reflejen la dependencia de la frecuencia de la DSR. Por ejemplo, se pueden generar coeficientes de mezcla descendente para bandas de frecuencia individuales. Del mismo modo, se puede generar en consecuencia una mezcla descendente dependiente de la frecuencia y una señal de reverberación difusa.

Para una DSR dependiente de la frecuencia, los coeficientes de mezcla descendente se pueden complementar en otras realizaciones con filtros que filtran la señal de audio como parte de la generación de la mezcla descendente. Como otro ejemplo, el efecto DSR se puede separar en un componente independiente de la frecuencia (banda ancha) que se utiliza para generar coeficientes de mezcla descendente independientes de la frecuencia que se utilizan para escalar las señales de audio individuales al generar la señal de mezcla descendente y un componente dependiente de la frecuencia que se puede aplicar a la mezcla descendente, por ejemplo, al aplicar un filtro dependiente de la frecuencia a la mezcla descendente. En algunas realizaciones, dicho filtro se puede combinar con otros filtros de coloración, por ejemplo, como parte del algoritmo del reverberador. La FIG. 7 ilustra un ejemplo con filtros de correlación (u , v) y coloración (h_L , h_R). Se trata de una Red de Retardo de Retroalimentación específica para la salida binaural, conocida como reverberador Jot.

Por lo tanto, en algunas realizaciones, la DSR puede comprender una parte componente dependiente de la frecuencia y una parte componente no dependiente de la frecuencia y el procesador 507 de coeficiente se puede disponer para generar los coeficientes de mezcla descendente en dependencia de la parte componente no dependiente de la frecuencia (e independientemente de la parte dependiente de la frecuencia). El procesamiento de la mezcla descendente se puede adaptar en base a la parte del componente dependiente de la frecuencia, es decir, el reverberador se puede adaptar en dependencia de la parte dependiente de la frecuencia.

En algunas realizaciones, la directividad de la radiación de sonido de una o más de las fuentes de audio puede ser dependiente de la frecuencia y, en dicho escenario, el procesador 505 de energía se puede disponer para generar una energía total emitida dependiente de la frecuencia que, cuando se combina con la DSR (que puede ser dependiente de la frecuencia o independiente) puede dar como resultado coeficientes de mezcla descendente dependientes de la frecuencia.

Esto se puede lograr, por ejemplo, al realizar un procesamiento individual en bandas de frecuencia discretas. A diferencia del procesamiento de una DSR dependiente de la frecuencia, la dependencia de la frecuencia para la directividad se debe realizar normalmente antes (o como parte de) la generación de la señal de mezcla descendente. Esto refleja que normalmente se requiere una mezcla descendente dependiente de la frecuencia para incluir los efectos de la directividad dependientes de la frecuencia, ya que estos normalmente son diferentes para diferentes fuentes. Después de la integración, es posible que el efecto neto tenga una variación significativa a lo largo de la frecuencia, es decir, la indicación de energía total emitida para una fuente dada puede tener una dependencia sustancial de la frecuencia, siendo ésta diferente para las diferentes fuentes. Por lo tanto, dado que las diferentes fuentes normalmente tienen diferentes patrones de directividad, la indicación de la energía total emitida para las diferentes fuentes normalmente también tiene diferentes dependencias de frecuencia.

A continuación se describirá un ejemplo específico de un posible enfoque. Proporcionar una DSR que caracterice las propiedades acústicas difusas de un espacio y determinar una fuente de energía emitida a partir de metadatos de directividad, ganancia previa y distancia de referencia, permite calcular la energía de reverberación deseada correspondiente. Por ejemplo, esto se puede determinar como:

$$E_{norm} * DSR$$

Cuando los componentes para calcular la DSR utilizan el mismo nivel de referencia (por ejemplo, relacionado con la escala completa de la señal), la energía de reverberación resultante también será una energía normalizada para muestras a escala completa en la señal PCM, cuando se utilice E_{norm} como se calculó anteriormente para la energía de la fuente emitida y, por lo tanto, corresponde a la energía de una Respuesta de Impulso (IR) para la reverberación difusa que se podría aplicar a la señal de entrada correspondiente para proporcionar el nivel correcto de reverberación en la representación de la señal utilizada.

Estos valores de energía se pueden utilizar para determinar los parámetros de configuración de un algoritmo de reverberación, un coeficiente de mezcla descendente o un filtro de mezcla descendente antes del algoritmo de reverberación.

Hay diferentes formas de generar reverberación. Los algoritmos basados en la Red de Retardo de Retroalimentación (FDN), como el reverberador Jot, son enfoques adecuados de baja complejidad. Alternativamente, se puede dar conformación a una secuencia de ruido para que tenga un decaimiento y una conformación espectral apropiados (dependientes de la frecuencia). En ambos ejemplos, una IR prototípica (con al menos T60 apropiado) se puede ajustar de tal manera que se corrija su nivel (dependiente de la frecuencia).

Los algoritmos del reverberador se pueden ajustar de tal manera que produzcan respuestas de impulso con energía unitaria (o la amplitud inicial unitaria de DSR puede estar en relación con una amplitud inicial) o el algoritmo del reverberador puede incluir su propia compensación, por ejemplo, en los filtros de coloración de un reverberador Jot. Alternativamente, la mezcla descendente se puede modificar con un ajuste (potencialmente dependiente de la frecuencia), o se pueden modificar los coeficientes de mezcla descendente producidos por el procesador 507 de coeficiente.

La compensación se puede determinar al generar una respuesta de impulso sin ningún ajuste de este tipo, pero con todas las demás configuraciones aplicadas (como el tiempo de reverberación apropiado (T60) y la densidad de reflexión (por ejemplo, valores de retardo en FDN) y al medir la energía de esa IR.

$$E_{protRev} = \sum_{n=0}^{N_{max}} h_{rev}(n)$$

La compensación puede ser la inversa de esa energía. Para su inclusión en los coeficientes de mezcla descendente, normalmente se aplica una raíz cuadrada. Por ejemplo:

$$d_x = \sqrt{\frac{E_{norm,x} * DSR}{E_{protRev}}}$$

En muchas otras realizaciones, la compensación se puede derivar de los parámetros de configuración. Por ejemplo, cuando la DSR es relativa a una amplitud de reverberación inicial, la primera reflexión se puede derivar de su configuración. Los filtros de correlación son preservadores de energía por definición, y también los filtros de coloración se pueden diseñar para serlo.

Suponiendo que no haya aumento neto o atenuación por los filtros de coloración, el reverberador puede, por ejemplo, resultar en una amplitud inicial (A_0) que depende de T60 y del valor de retardo más pequeño minDelay:

$$A_0 = 10^{\frac{-3 * \text{minDelay}}{T60 * fs}} * \sqrt{T60}$$

La predicción de la energía de reverberación también se puede hacer heurísticamente.

Como modelo general para la energía de reverberación difusa, se puede considerar una función exponencial $A(t)$:

$$\alpha = \frac{\ln\left(10^{-\frac{60}{T60}}\right)}{-T60}$$

$$A(t) = A_0 \cdot e^{-\alpha \cdot (t-t_3)}$$

para $t \geq t_3$ = retardo previo. Con α el factor de decaimiento controlado por T60, y A_0 la amplitud en el retardo previo.

Calculando la energía acumulada de una función como esta, se aproximará asintóticamente a algún valor de energía final. El valor de energía final tiene una relación lineal casi perfecta con T60.

El factor de la relación lineal depende de la dispersión de la función A (establecer cada segundo valor en 0 da como resultado aproximadamente la mitad de la energía), el valor inicial A_0 (la energía escala linealmente con A_0^2) y la tasa de muestra (escala linealmente con cambios en fs). La cola difusa se puede modelar de forma fiable con dicha función, utilizando T60, la densidad de reflexión (derivada de los retardos FDN) y la tasa de muestra. A_0 para el modelo se puede calcular como se mostró anteriormente, para que sea igual a la de la FDN.

Al generar múltiples reverberaciones paramétricas con valores de banda ancha T60 en el rango de 0.1-2 s, la energía de la IR es cercana a la lineal con el modelo. El factor de escala entre la energía real y los promedios del modelo de ecuaciones exponenciales se determina por la dispersión de la respuesta FDN. Esta dispersión se vuelve menor hacia el final de la IR, pero tiene más impacto al principio. Se ha encontrado, al probar lo anterior con múltiples configuraciones de los valores de retardo, que existe una relación casi lineal entre el factor de reducción del modelo y la diferencia más pequeña entre los retardos configurados en la FDN.

Por ejemplo, para una cierta implementación de un reverberador Jot, esto puede equivaler a un factor de escala SF calculado por:

$$SF = 7.0208 * \text{MinDelayDiff} + 214.1928$$

La energía del modelo se calcula al integrar desde $t = 0$ hasta el infinito. Esto se puede hacer analíticamente y da como resultado:

$$\text{revModelEnergy} = \frac{A_0^2 * fs}{2 * \alpha}$$

Combinando lo anterior, obtenemos la siguiente predicción para la energía de reverberación.

$$revModelEnergy = \frac{A_0^2 * fs}{2 * alfa} * \frac{1}{0.8776 * MinDelayDiff + 26.7741}$$

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

La invención se puede implementar en cualquier forma adecuada, que incluye hardware, software, firmware o cualquier combinación de estos. Opcionalmente, la invención se puede implementar, al menos parcialmente, 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 se pueden implementar física, funcional y lógicamente de cualquier manera adecuada. De hecho, la funcionalidad se puede implementar en una sola unidad, en una pluralidad de unidades o como parte de otras unidades funcionales. Como tal, la invención se puede implementar en una sola unidad o se puede distribuir física y funcionalmente entre diferentes unidades, circuitos y procesadores.

Aunque la presente invención ha sido descrita con respecto a algunas realizaciones, no está destinada a limitarse a la forma específica establecida en el presente documento. Más bien, el alcance de la presente invención está limitado únicamente por las reivindicaciones que la acompañan. Adicionalmente, aunque una característica pueda parecer descrita con respecto a realizaciones particulares, un experto en la técnica reconocería que varias características de las realizaciones descritas se pueden combinar de acuerdo con la invención. En las reivindicaciones, el término que comprender no excluye la presencia de otros elementos o etapas.

Además, aunque se enumeran individualmente, una pluralidad de medios, elementos, circuitos o pasos de método se pueden implementar, por ejemplo, por un solo circuito, unidad o procesador. Adicionalmente, aunque las características individuales se pueden incluir en diferentes reivindicaciones, es posible que éstas se combinen ventajosamente, y la inclusión en diferentes reivindicaciones no implica que una combinación de características no sea factible y/o ventajosa. También, 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. Por lo tanto, las referencias a "un", "una", "primero", "segundo", etc., no excluyen una pluralidad. Los signos de referencia en las reivindicaciones se proporcionan simplemente como ejemplo aclaratorio y no se interpretarán en el sentido de limitar en modo alguno el alcance de las reivindicaciones.

REIVINDICACIONES

1. Un aparato de audio para generar una señal de reverberación difusa para un entorno; el aparato comprende:
5 un receptor (501) dispuesto para recibir una pluralidad de señales de audio que representan fuentes de sonido en el entorno;
10 un receptor (501) de metadatos dispuesto para recibir metadatos para la pluralidad de señales de audio, los metadatos comprenden:
una medida de la relación entre la señal de reverberación difusa y la señal total indicativa de un nivel de sonido de reverberación difusa con respecto al sonido total emitido en el entorno,
15 y para cada señal de audio:
una indicación del nivel de señal;
20 datos de directividad indicativos de la directividad de la radiación de sonido de la fuente de sonido representada por la señal de audio;
un circuito (505, 507) dispuesto de tal manera que, para cada una de las señales de audio, se determine:
25 una indicación de la energía total emitida basada en la indicación del nivel de señal y los datos de directividad, y un coeficiente de mezcla descendente basado en la indicación de la energía total emitida y la medida de la relación entre la señal de reverberación difusa y la señal total;
un mezclador (509) descendente dispuesto para generar una señal de mezcla descendente al combinar los componentes de la señal de cada señal de audio generada al aplicar el coeficiente de mezcla descendente de cada
30 señal de audio a la señal de audio;
un reverberador (407) para generar la señal de reverberación difusa para el entorno a partir del componente de señal de mezcla descendente.
35 2. El aparato de audio de la reivindicación 1 en el que la directividad de la radiación de sonido depende de la frecuencia y el circuito se dispone para determinar una indicación de energía total emitida dependiente de la frecuencia y coeficientes de mezcla descendente dependientes de la frecuencia.
40 3. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que la medida de la relación entre la señal de reverberación difusa y la señal total depende de la frecuencia y el circuito (505, 507) se dispone para determinar los coeficientes de mezcla descendente dependientes de la frecuencia.
45 4. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que la medida de la relación entre la señal de reverberación difusa y la señal total comprende una parte dependiente de la frecuencia y una parte no dependiente de la frecuencia, y en el que el circuito (505, 507) se dispone para determinar los coeficientes de mezcla descendente en dependencia de la parte no dependiente de la frecuencia y para adaptar el reverberador (407) en dependencia de la parte dependiente de la frecuencia.
50 5. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que el circuito se dispone para determinar la indicación de energía total emitida para una primera señal de audio de la pluralidad de señales de audio en respuesta a un escalado de la indicación del nivel de señal para la primera señal de audio por un valor determinado al integrar un patrón de directividad de la fuente de sonido representada por la primera señal de audio.
55 6. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que la indicación del nivel de señal para una primera señal de audio de la pluralidad de señales de audio comprende una distancia de referencia, la distancia de referencia que indica una distancia desde la fuente de audio representada por la primera señal de audio asociada con una ganancia de referencia de distancia para la primera señal de audio.
60 7. El aparato de audio de la reivindicación 6 como dependiente de la reivindicación 5 en el que la integración se realiza a una distancia que es la distancia de referencia desde la fuente de audio representada por la primera señal de audio.
8. El aparato de audio de cualquier reivindicación anterior en el que la medida de la relación entre la señal de reverberación difusa y la señal total es indicativa de una energía del sonido de reverberación difusa con respecto a una energía del sonido total emitido en el entorno.
65

9. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que la medida de la relación entre la señal difusa y la señal total es indicativa de una amplitud inicial dentro de un intervalo de sonido difuso con respecto a una energía del sonido total emitido en el entorno.

5 10. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que el coeficiente de mezcla descendente determinado para una primera señal de audio de la pluralidad de señales de audio es independiente de una posición de una primera fuente de audio representada por la primera señal de audio.

10 11. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que el coeficiente de mezcla descendente determinado para una primera señal de audio de la pluralidad de señales de audio es independiente de la posición de un oyente.

15 12. El aparato de audio de cualquiera de las reivindicaciones anteriores en el que la indicación del nivel de señal para una primera señal de audio de la pluralidad de señales de audio comprende además una indicación de ganancia para la primera señal de audio, la indicación de ganancia es indicativa de una ganancia que se aplicará a la primera señal de audio cuando se renderiza el sonido desde una primera fuente de audio representada por la primera señal de audio, y en el que el circuito (505, 507) se dispone para determinar el coeficiente de mezcla descendente para la primera señal de audio en respuesta a la indicación de ganancia.

20 13. El aparato de audio de cualquiera de las reivindicaciones anteriores que comprende además un circuito (401) de renderización directo dispuesto para generar una señal de audio de ruta directo para una primera señal de audio de la pluralidad de señales de audio en respuesta a la indicación del nivel de señal y a los datos de directividad para la primera señal de audio.

25 14. El aparato de audio de cualquiera de las reivindicaciones anteriores, en el que los metadatos comprenden además una indicación de retardo y la medida de la relación entre la señal difusa y la señal total es indicativa de una energía de sonido de reverberación difusa que tiene un retardo mayor que la indicación de retardo con respecto a una energía del sonido total emitido en el entorno.

30 15. Un método para generar una señal de reverberación difusa para un entorno, el método comprende:

recibir una pluralidad de señales de audio que representan fuentes de sonido en el entorno;

recibir metadatos para la pluralidad de señales de audio, los metadatos comprenden:

35 una medida de la relación entre la señal de reverberación difusa y la señal total indicativa de un nivel de sonido de reverberación difusa con respecto al sonido total emitido en el entorno,

y para cada señal de audio:

40 una indicación del nivel de señal;

datos de directividad indicativos de la directividad de la radiación de sonido de la fuente de sonido representada por la señal de audio;

45 para cada una de la pluralidad de señales de audio, determinar:

una indicación de la energía total emitida basada en la indicación del nivel de señal y los datos de directividad, y

50 un coeficiente de mezcla descendente basado en la indicación de la energía total emitida y la medida de la relación entre la señal de reverberación difusa y la señal total;

generar una señal de mezcla descendente al combinar los componentes de la señal de cada señal de audio generada al aplicar el coeficiente de mezcla descendente para cada señal de audio a la señal de audio;

55 generar la señal de reverberación difusa para el entorno a partir del componente de señal de mezcla descendente.

16. Un producto de programa informático que comprende código de programa informático que, cuando se ejecuta por un ordenador, hace que el ordenador lleve a cabo todas las etapas de la reivindicación 15.

60

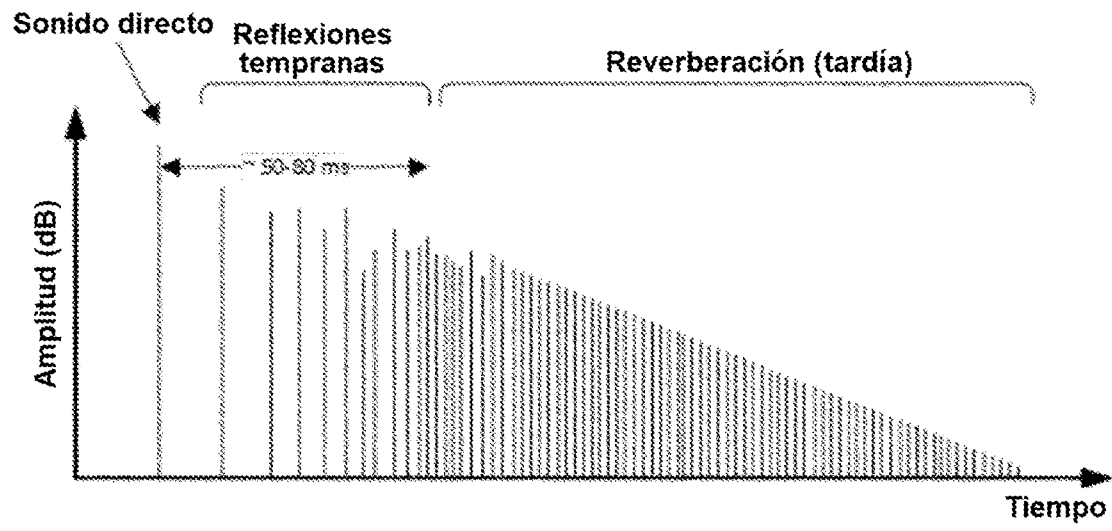


FIG. 1

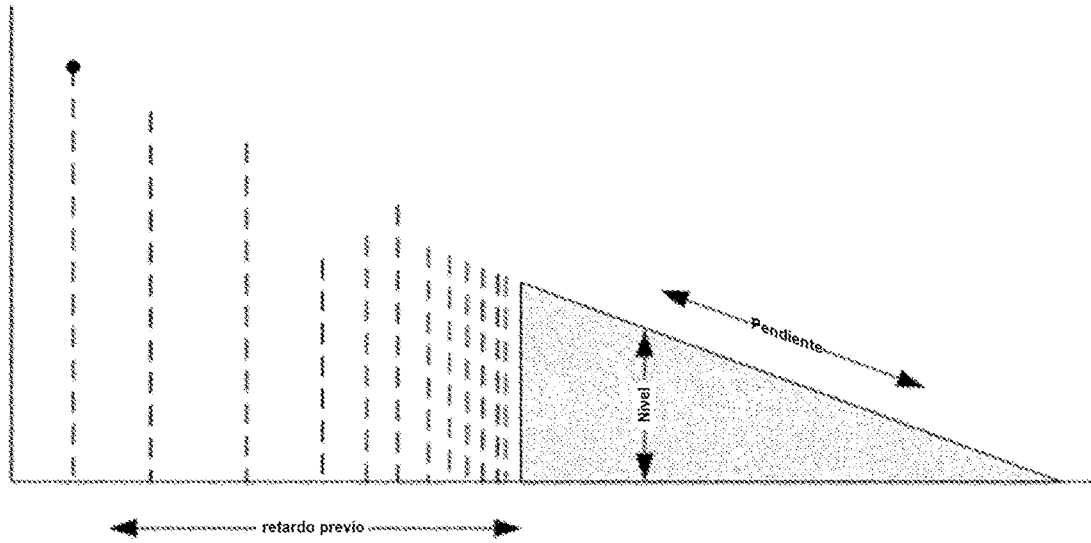


FIG. 2

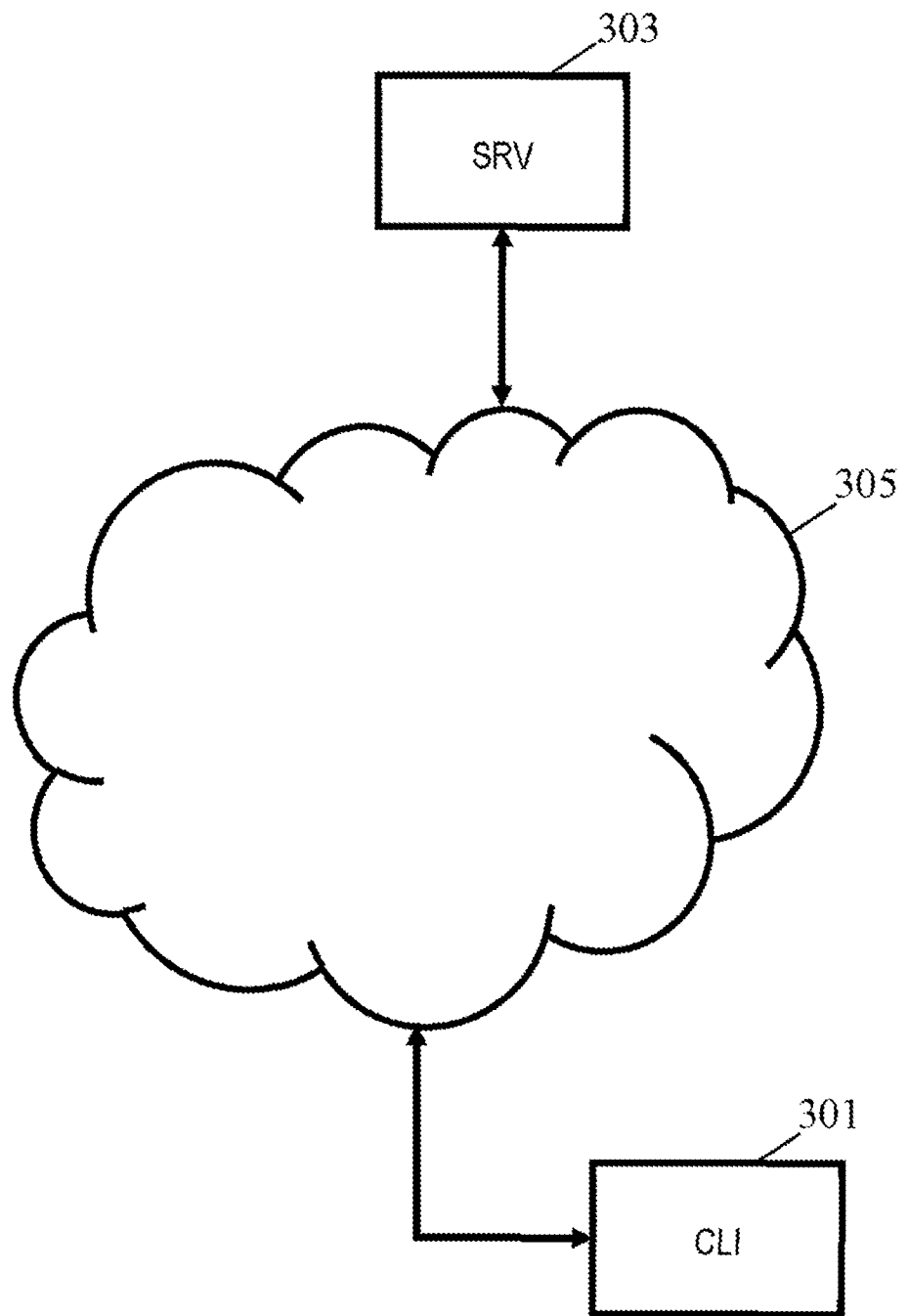


FIG. 3

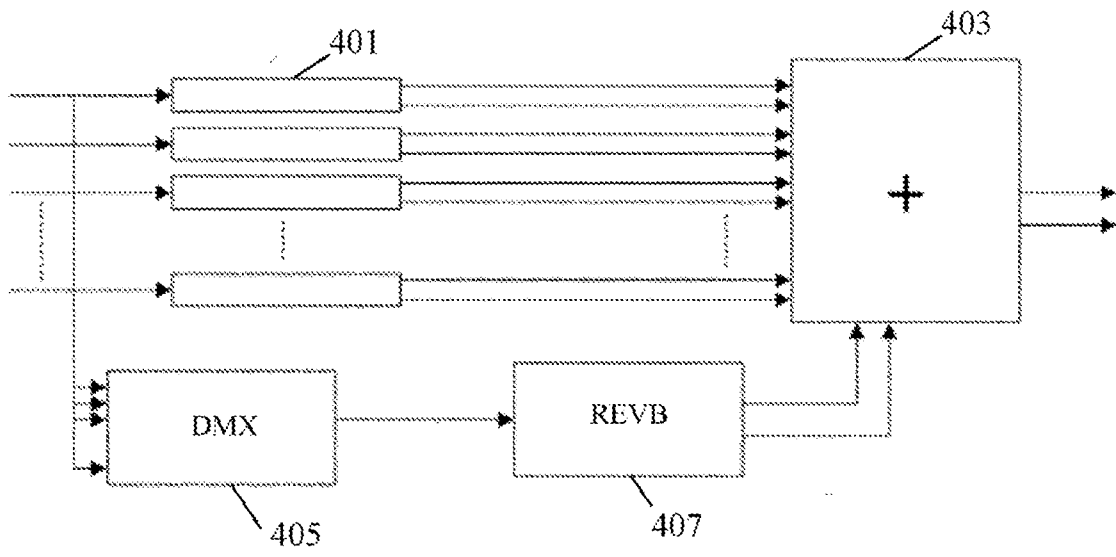


FIG. 4

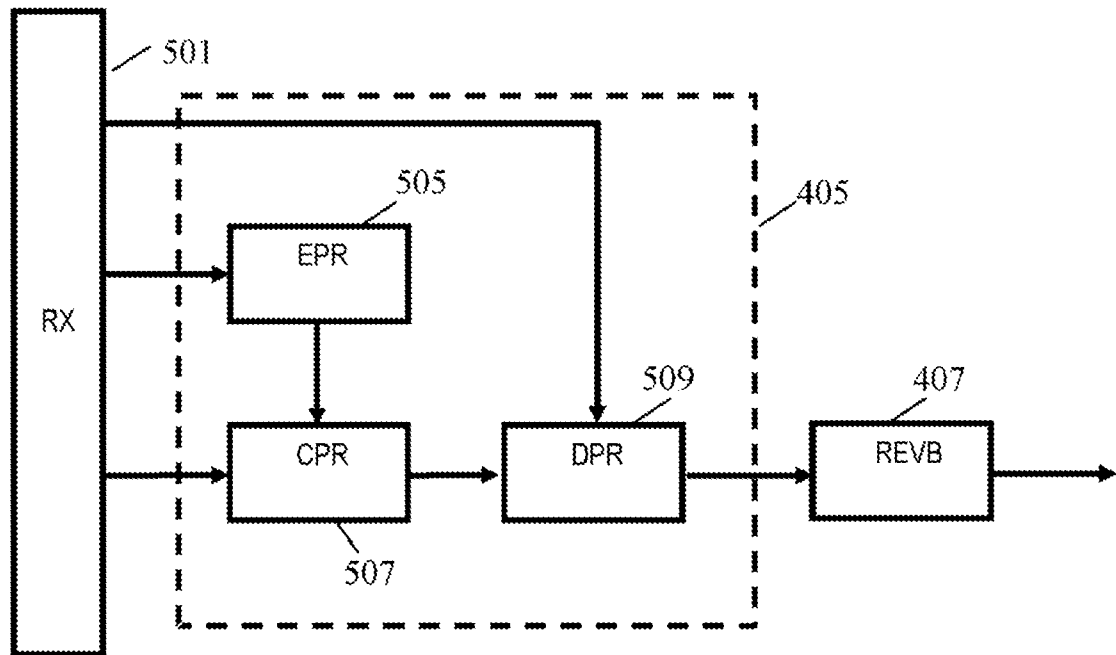


FIG. 5

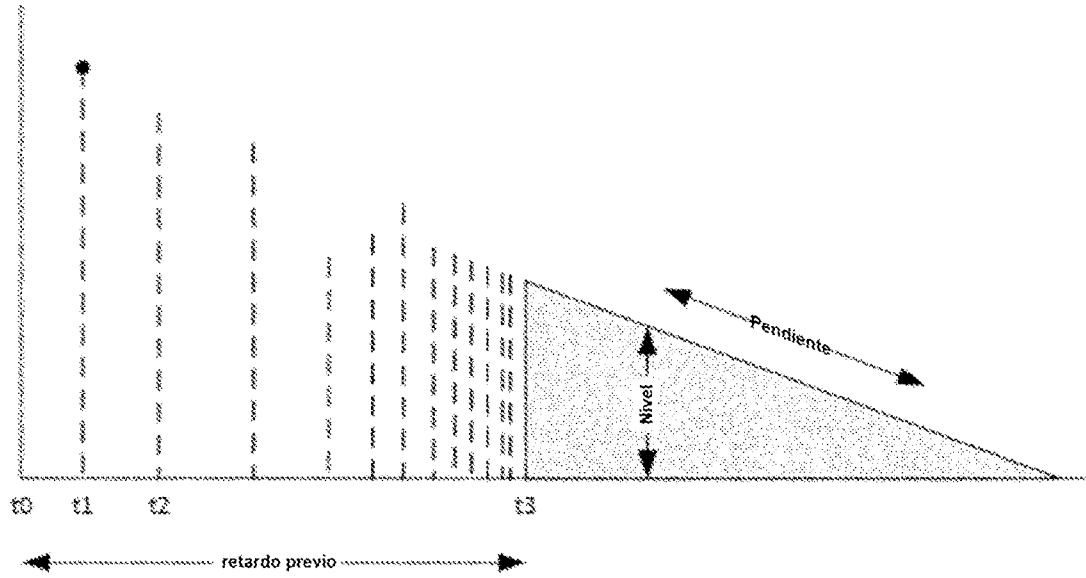


FIG. 6

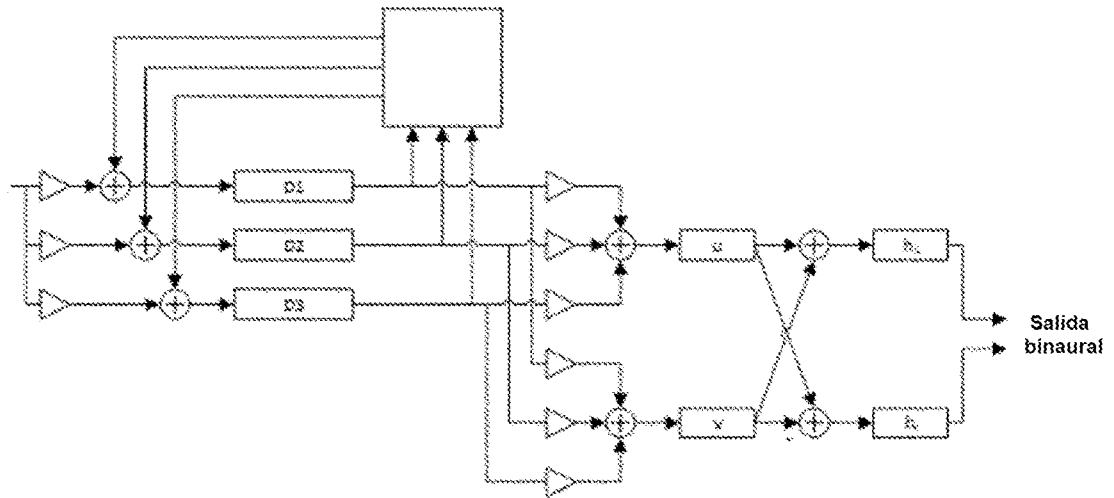


FIG. 7