

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 760 873**

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: **18.07.2014 PCT/EP2014/065534**

87 Fecha y número de publicación internacional: **29.01.2015 WO15011055**

96 Fecha de presentación y número de la solicitud europea: **18.07.2014 E 14741304 (1)**

97 Fecha y número de publicación de la concesión europea: **18.09.2019 EP 3025520**

54 Título: **Procedimiento de tratamiento de una señal de audio, unidad de tratamiento de señales, renderizador binaural, codificador de audio y decodificador de audio**

30 Prioridad:

22.07.2013 EP 13177361

18.10.2013 EP 13189255

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

18.05.2020

73 Titular/es:

**FRAUNHOFER GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27C
80686 München, DE**

72 Inventor/es:

**FÜG, SIMONE y
PLOGSTIES, JAN**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 760 873 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

Procedimiento de tratamiento de una señal de audio, unidad de tratamiento de señales, renderizador binaural, codificador de audio y decodificador de audio

5

[0001] La presente invención se refiere al campo de la codificación/decodificación de audio, especialmente a la codificación de audio espacial y codificación de objetos de audio espacial, por ejemplo, el campo de los sistemas códec de audio 3D. Las realizaciones de la invención se refieren a un procedimiento para tratar una señal de audio según una respuesta al impulso del ambiente, a una unidad de tratamiento de señales, un renderizador binaural, un codificador de audio y un decodificador de audio.

10

[0002] Las herramientas de codificación espacial de audio son bien conocidas en la técnica y están normalizadas, por ejemplo, en la norma MPEG-envolvente. La codificación de audio espacial comienza a partir de una pluralidad de entradas originales, por ejemplo, cinco o siete canales de entrada, que se identifican por su colocación en una configuración de reproducción, por ejemplo, como un canal izquierdo, un canal central, un canal derecho, un canal envolvente izquierdo, un canal envolvente derecho y un canal de realce de las bajas frecuencias. Un codificador de audio espacial puede deducir uno o más canales de submezcla de los canales originales y, adicionalmente, puede deducir datos paramétricos con relación a las señales espaciales tales como diferencias de nivel entre canales en los valores de coherencia de canales, diferencias de fase entre canales, diferencias de tiempo entre canales, etc. Los uno o más canales de submezcla se transmiten junto con la información lateral paramétrica que indica las señales espaciales a un decodificador de audio espacial para la decodificación de los canales de submezcla y los datos paramétricos asociados con el fin de obtener finalmente los canales de salida que son una versión aproximada de los canales de entrada originales. La colocación de los canales en la configuración de salida puede ser fija, por ejemplo, un formato de 5.1, un formato 7.1, etc.

25

[0003] Además, las herramientas de codificación de objetos de audio espacial son bien conocidas en la técnica y están normalizadas, por ejemplo, en la norma MPEG SAOC (SAOC = codificación de objetos de audio espacial). A diferencia de la codificación de audio espacial que comienza a partir de los canales originales, la codificación de objetos de audio espacial comienza a partir de objetos de audio que no están automáticamente dedicados para una configuración de reproducción renderizada. Al contrario, la colocación de los objetos de audio en la escena de reproducción es flexible y puede ser configurada por un usuario, por ejemplo, mediante la introducción de cierta información de renderización en un decodificador de codificación de objetos de audio espacial. De forma alternativa o adicional, la información de renderización puede transmitirse como información lateral adicional o metadatos; la información de renderización puede incluir información en cuya posición en la configuración de reproducción se colocará un cierto objeto de audio (por ejemplo, en el tiempo). Con el fin de obtener una cierta compresión de datos, se codifica una serie de objetos de audio mediante un codificador SAOC que calcula, a partir de los objetos de entrada, uno o más canales de transporte por la submezcla de los objetos según cierta información de submezcla. Además, el codificador SAOC calcula la información lateral paramétrica que representa las señales entre objetos tales como las diferencias de nivel de objeto (OLD), valores de coherencia de objeto, etc. Al igual que en SAC (SAC = Codificación de Audio Espacial), los datos paramétricos entre objetos se calculan para figuras individuales de tiempo/frecuencia. Para una cierta trama (por ejemplo, de 1.024 o 2.048 muestras) de la señal de audio se considera una pluralidad de bandas de frecuencia (por ejemplo 24, 32 o 64 bandas) de manera que se proporcionen datos paramétricos para cada trama y cada banda de frecuencia. Por ejemplo, cuando una pieza de audio tiene 20 tramas y cuando cada trama está subdividida en 32 bandas de frecuencia, el número de figuras de tiempo/frecuencia es 640.

45

[0004] En los sistemas de audio en 3D puede ser deseable proporcionar una impresión espacial de una señal de audio como si la señal de audio fuera escuchada en un ambiente específico. En tal situación, se proporciona una respuesta al impulso del ambiente del ambiente específico, por ejemplo basándose en una medida del mismo, y se usa para tratar la señal de audio tras la presentación a un oyente. Se puede desear tratar el sonido directo y las reflexiones tempranas en dicha presentación separada de la reverberación tardía.

50

[0005] El documento WO-2012/116934-A1 describe un aparato para determinar una medida de un nivel percibido de reverberación en una señal de mezclado consistente en un componente de señal directa y un componente de señal de reverberación. El aparato tiene un procesador de modelo de sonoridad que tiene una etapa de filtro perceptual para filtrar el componente de señal en seco del componente de señal de reverberación o la señal de mezclado, en el que la etapa de filtro perceptual está configurada para modelizar un mecanismo de percepción de un auditorio de una entidad para obtener una señal directa filtrada, una señal de reverberación filtrada o una señal de mezclado filtrada. El aparato tiene además un estimador de sonoridad para estimar una primera medida de sonoridad usando la señal directa filtrada y para estimar una segunda medida de sonoridad usando la señal de reverberación filtrada o la señal de mezclado filtrada, en el que la señal de filtrado mezclada se deriva de una superposición del componente de señal directa y el componente de señal de reverberación. El aparato tiene también un combinador para combinar las medidas de sonoridad primera y segunda con el fin de obtener una medida para el nivel percibido de reverberación.

60

[0006] El documento US-5.371.799-A describe un sistema para tratar una señal de audio para su reproducción en auriculares en los que la fuente de sonido aparente está situada fuera de la cabeza del oyente y trata la señal como si estuviera formada por una parte de onda directa, una parte de reflexiones tempranas y una parte de reverberaciones. La parte de onda directa de la señal es tratada en filtros cuyos coeficientes de filtro se eligen basándose en el acimut deseado de la posición de la fuente de sonido virtual. La parte de reflexión temprana se hace pasar a través de un banco de filtros conectados en paralelo cuyos coeficientes se eligen basándose en cada acimut de reflexión. Las salidas de estos filtros se hacen pasar a través de escalares para ajustar la amplitud de manera que estimule un intervalo deseado de la fuente de sonido virtual. La parte de reverberación se trata sin ninguna información de posición de la fuente de sonido, usando un generador de números aleatorios, por ejemplo, y la salida se atenúa en un atenuador exponencial que se desvanece. Las salidas de los escalares y los atenuadores se suman a continuación para producir señales de auriculares izquierdo y derecho para su reproducción en los transductores de los auriculares respectivos.

[0007] El documento US-2005/0.100.171-A1 describe un procedimiento, un aparato y un producto de software para tratar una pluralidad de señales de audio de entrada. El aparato acepta una pluralidad de señales de entrada e incluye un reverberador de múltiples entradas y múltiples salidas dispuesto para generar un conjunto de señales de salida que incluye componentes de reverberación retrasados que simulan las reverberaciones que es probable que escuche un oyente en un entorno de audición. El aparato incluye además un filtro de múltiples entradas y dos salidas que acepta las salidas del reverberador y la pluralidad de terminales de entrada, proporcionando salidas para los oídos izquierdo y derecho, y configurado para implementar un conjunto de funciones de transferencia relacionadas con la cabeza correspondientes a un entorno de audición. El aparato es tal que un oyente que audición las salidas a través de los auriculares tiene la sensación de escuchar la pluralidad de señales de audio de entrada como si procedieran de una pluralidad de altavoces dispuestos espacialmente en el entorno de audición en una pluralidad de direcciones correspondientes.

[0008] El documento WO-2010/012.478-A2 describe un dispositivo para generar una señal binaural basándose en una señal de múltiples canales que representa una pluralidad de canales y está destinada a su reproducción por una configuración de altavoces que tiene una posición de fuente de sonido virtual asociada con cada canal. Comprende un reductor de correlación para tratar de forma diferencial, y con ello reducir, una correlación entre al menos uno de entre un canal izquierdo y un canal derecho de la pluralidad de canales, un canal delantero y un canal trasero de la pluralidad de canales y un canal central y un canal no central de la pluralidad de canales, con el fin de obtener un conjunto reducido de canales con semejanza entre ellos; una pluralidad de direccional filtros, un primer mezclador para mezclar salidas de los filtros direccionales que modelizan la transmisión acústica al canal del primer oído del oyente, y un segundo mezclador para mezclar salidas de los filtros direccionales que modelizan la transmisión acústica al canal del segundo oído del oyente.

[0009] El documento EP-1.768.107-A1 describe la codificación de señales de audio de múltiples canales mediante la generación de una señal de reverberación que usa un filtro pasatodo, una segmentación de una señal en los dominios del tiempo y la frecuencia con el fin de realizar ajuste de nivel y el mezclado de una señal binaural codificada con una señal original codificada hasta una frecuencia de paso fija.

[0010] El objeto subyacente de la presente invención consiste en proporcionar un enfoque aprobado para tratar por separado la señal de audio con una parte temprana y una reverberación tardía de la respuesta al impulso del ambiente que permita lograr un resultado que sea perceptualmente lo más idéntico posible al resultado de una convolución de la señal de audio con la respuesta al impulso completa.

[0011] Este objeto se consigue por medio un procedimiento según la reivindicación 1, una unidad de tratamiento de señal según la reivindicación 11, un renderizador binaural según la reivindicación 15, un codificador de audio según la reivindicación 16 y un decodificador de audio según la reivindicación 17.

[0012] La presente invención se basa en los hallazgos del autor de la invención de que en los enfoques convencionales existe un problema según el cual tras el tratamiento de la señal de audio según la respuesta al impulso del ambiente, el resultado de tratar la señal de audio por separado con respecto a la parte temprana y la reverberación tardía se desvía de un resultado obtenido cuando se lleva a cabo la aplicación de una convolución con una respuesta al impulso completa. La invención se basa además en los hallazgos del autor de la invención de que un nivel adecuado de reverberación depende tanto de la señal de audio de entrada como de la respuesta al impulso, debido a que la influencia de la señal de audio de entrada en la reverberación no se conserva completamente cuando, por ejemplo, se usa un enfoque de reverberación sintético. La influencia de la respuesta al impulso puede considerarse usando las características de reverberación conocidas como parámetro de entrada. La influencia de la señal de entrada puede considerarse mediante un cambio de escala dependiente de la señal para adaptar el nivel de reverberación que se determina basándose en la señal de audio de entrada. Se ha descubierto que mediante este enfoque, el nivel percibido de la reverberación coincide mejor con el nivel de reverberación cuando se usa el enfoque de convolución completo para la renderización binaural.

[0013] La presente invención proporciona un renderizador binaural, que comprende la unidad de tratamiento de señales de la invención.

5 **[0014]** La presente invención proporciona un codificador de audio para codificar señales de audio, que comprende la unidad de tratamiento de señales de la invención o el renderizador binaural de la invención para el tratamiento de señales de audio antes de la codificación.

[0015] La presente invención proporciona un decodificador de audio para decodificar señales de audio
10 codificadas, que comprende la unidad de tratamiento de señales de la invención o el renderizador binaural de la invención para el tratamiento de las señales de audio decodificadas.

[0016] Las realizaciones de la presente invención se describirán con respecto a las figuras adjuntas, en el que:

15 la Fig. 1 ilustra una visión general de un codificador de audio 3D de un sistema de audio 3D;
la Fig. 2 ilustra una visión general de un decodificador de audio 3D de un sistema de audio 3D;
la Fig. 3 ilustra un ejemplo para implementar un convertidor de formato que puede ser implementado en el decodificador de audio 3D de la Fig. 2;

20 la Fig. 4 ilustra una realización de un renderizador binaural que puede ser implementada en el decodificador de audio 3D de la Fig. 2;

la Fig. 5 ilustra un ejemplo de una respuesta al impulso del ambiente $h(t)$;

la Fig. 6 ilustra diferentes posibilidades para tratar una señal de entrada de audio con una respuesta al impulso del ambiente, en el que la Fig. 6(a) muestra el tratamiento de la señal de audio completa, según la respuesta al impulso
25 del ambiente, y la Fig. 6(b) muestra el tratamiento separado de la parte temprana y la parte de reverberación tardía;

la Fig. 7 ilustra un diagrama de bloques de una unidad de tratamiento de señales, como un renderizador binaural, que funciona según las enseñanzas de la presente invención;

la Fig. 8 ilustra esquemáticamente el tratamiento binaural de señales de audio en un renderizador binaural según una realización de la presente invención; y

30 la Fig. 9 ilustra esquemáticamente el tratamiento en el reverberador del dominio de la frecuencia del renderizador binaural de la Fig. 8 según una realización de la presente invención.

[0017] A continuación se describirán realizaciones útiles del enfoque de la invención. La siguiente descripción se iniciará con una visión general del sistema de un sistema de códec de audio 3D en el que se puede implementar
35 el procedimiento de la invención.

[0018] Las Fig. 1 y 2 muestran los bloques algorítmicos de un sistema de audio 3D según las realizaciones. Más en concreto, la Fig. 1 muestra una visión general de un codificador de audio 3D 100. El codificador de audio 100
40 recibe en un circuito de pre-renderizador/mezclador 102, que puede proporcionarse opcionalmente, señales de entrada, más en concreto una pluralidad de canales de entrada que proporcionan al codificador de audio 100 una pluralidad de señales de canales 104, una pluralidad de señales de objetos 106 y metadatos de objetos correspondientes 108. Las señales de objetos 106 tratadas por el pre-renderizador/mezclador 102 (véanse las
45 señales 110) se pueden proporcionar a un codificador SAOC 112 (SAOC = Codificación de Objeto de Audio Espacial). El codificador SAOC 112 genera los canales de transporte SAOC 114 proporcionados a un codificador USAC 116 (USAC = Codificación de Discurso y Audio Unificada). Además, la señal de SAOC-SI 118 (SAOC-SI = información lateral SAOC) también se proporciona al codificador USAC 116. El codificador USAC 116 recibe además
50 señales de objeto 120 directamente desde el pre-renderizador/mezclador, así como señales de canal y señales de objeto pre-renderizadas 122. La información de metadatos de objeto 108 se aplica a un codificador OAM 124 (OAM = metadatos de objeto) que proporciona la información de metadatos de objeto 126 comprimido al codificador USAC. El codificador USAC 116, basándose en las señales de entrada antes mencionadas, genera una señal de salida MP4 comprimida, como se muestra en 128.

[0019] La Fig. 2 muestra una visión general de un decodificador de audio 3D 200 del sistema de audio 3D. La señal 128 (mp4) codificada generada por el codificador de audio 100 de la Fig. 1 es recibida en el decodificador de
55 audio 200, más en concreto en un decodificador USAC 202. El decodificador USAC 202 decodifica la señal recibida 128 en las señales de canal 204, las señales de objeto pre-renderizadas 206, las señales de objeto 208 y las señales del canal de transporte SAOC 210. Además, a la información de metadatos de objeto comprimidos 212 y a la señal SAOC-SI 214 se les da salida por el decodificador USAC 202. Las señales de objeto 208 se proporcionan a un renderizador de objetos 216 para producir las señales de objetos renderizados 218. Las señales de los canales
60 de transporte SAOC 210 se suministran al decodificador 220 SAOC para producir las señales de objetos renderizados 222. La metainformación de objeto comprimido 212 se suministra al decodificador OAM 224 para producir las señales de control respectivas al renderizador de objeto 216 y al decodificador SAOC 220 para generar las señales de objetos renderizados 218 y las señales de objetos renderizados 222. El decodificador comprende además un mezclador 226 que recibe, como se muestra en la Fig. 2, las señales de entrada 204, 206, 218 y 222

para producir las señales de canal 228. A las señales de canal puede dársele salida directamente a un altavoz, por ejemplo, un altavoz de canal 32, como se indica en 230. Las señales 228 pueden ser proporcionadas a un circuito de conversión de formato 232 que recibe como entrada de control una señal de la disposición de reproducción que indica la forma en que se convertirán las señales de los canales 228. En la realización representada en la Fig. 2, se supone que la conversión se ha de llevar a cabo de una manera que las señales se puedan proporcionar a un sistema de altavoces 5.1 tal como se indica en 234. Además, los canales de señales 228 pueden ser proporcionados a un renderizador binaural 236 para generar dos señales de salida, por ejemplo, para auriculares, como se indica en 238.

10 **[0020]** El sistema de codificación/decodificación que se representa en las Fig. 1 y 2 puede basarse en el códec MPEG-D USAC para codificar señales de canal y objeto (véanse las señales 104 y 106). Para aumentar la eficiencia de codificación de una gran cantidad de objetos, puede usarse la tecnología MPEG SAOC. Existen tres tipos de renderizadores que pueden llevar a cabo las tareas de renderización de objetos a canales, renderizando los canales a auriculares o renderizando los canales a una configuración de altavoces diferente (véase la Fig. 2, referencias 230, 234 y 238). Cuando las señales de objetos se transmiten de manera explícita o se codifican paramétricamente usando SAOC, la información de metadatos objeto correspondiente 108 se comprime (véase la señal 126) y se multiplexa en el tren de bits de audio 3D 128.

15 **[0021]** A continuación se describirán con más detalle los bloques de algoritmos para el sistema de audio 3D global mostrados en las Fig. 1 y 2.

20 **[0022]** El pre-renderizador/mezclador 102 se puede proporcionar opcionalmente para convertir una escena de entrada de canal más objetos en una escena de canal antes de la codificación. Funcionalmente, es idéntico al renderizador/mezclador de objetos que se describirá a continuación. La pre-renderización de objetos puede ser deseable para asegurar una entropía de señal determinista en la entrada del codificador que sea básicamente independiente de la cantidad de señales de objetos activos simultáneamente. Con la pre-renderización de los objetos, no se requiere ningún objeto de transmisión de metadatos. Se representan señales de objetos discretos en el diseño del canal que el codificador está configurado para usar. Las ponderaciones de los objetos para cada canal se obtuvieron a partir de los metadatos de objeto asociados (OAM).

25 **[0023]** El codificador USAC 116 es el códec principal para las señales del altavoz-canal, señales de objetos discretos, señales de submezcla de objeto y señales pre-renderizadas. Se basa en la tecnología MPEG-D USAC. Maneja la codificación de las señales superiores mediante la creación de información de correspondencia de canal y objeto basada en la información geométrica y semántica del canal de entrada y la asignación de objetos. Esta información de correspondencia describe la correspondencia entre los canales de entrada y los objetos con los elementos de canales USAC, tales como elementos pares de canales (CPE), elementos de un solo canal (SCE), efectos de baja frecuencia (LFE) y elementos de canal cuádruple (QCE) y CPE, SCE y LFE, y la información correspondiente se transmite al decodificador. Todas las cargas adicionales como datos SAOC 114, 118 o metadatos de objeto 126 se consideran en el control de la velocidad del codificador. La codificación de objetos es posible en diferentes formas, dependiendo de los requisitos de velocidad/distorsión y los requisitos de interactividad para el renderizador. Según las realizaciones, son posibles las siguientes variantes de codificación de objeto:

- Objetos pre-renderizados: las señales de objetos son pre-renderizadas y mezcladas con las señales del canal 22.2 antes de la codificación. La cadena de codificación tardía ve 22.2 señales de canal.
- 45 • Formas de onda de objetos discretos: los objetos se suministran como formas de onda monofónicas al codificador. El codificador usa los elementos de canal individuales (SCE) para transmitir los objetos además de las señales de canal. Los objetos decodificados se renderizan y se mezclan en el lado del receptor. La información de metadatos de objeto comprimidos se transmite al receptor/renderizador.
- 50 • Formas de onda de objetos paramétricos: se describen propiedades de los objetos y su relación entre sí mediante parámetros SAOC. La submezcla de las señales de objetos se codifica con USAC. La información paramétrica se transmite junto a esta. El número de canales de submezcla se elige en función del número de objetos y la velocidad de datos global. La información de metadatos de objeto comprimidos se transmite al renderizador SAOC.

55 **[0024]** El codificador SAOC 112 y el decodificador SAOC 220 para señales de objeto pueden estar basados en la tecnología MPEG SAOC. El sistema es capaz de recrear, modificar y renderizar una serie de objetos de audio basado en un número menor de canales de transmisión y datos paramétricos adicionales, tales como OLD, IOC (coherencia entre objetos), DMG (ganancias de submezcla). Los datos paramétricos adicionales presentan una velocidad de datos significativamente menor que la requerida para la transmisión de todos los objetos individualmente, haciendo la codificación muy eficiente. El codificador SAOC 112 toma como entrada las señales de objeto/canal en formas de onda monofónicas y produce la información paramétrica (que se empaqueta en el tren de bits de audio 3D 128) y los canales de transporte SAOC (que son codificados usando elementos de un solo canal y transmitidos). El decodificador SAOC 220 reconstruye las señales de objeto/canal de los canales de transporte SAOC decodificados 210 y la información paramétrica 214, y genera la escena de audio de salida basándose en el diseño de la reproducción, la información de metadatos de objeto descomprimido y, opcionalmente, en la

información de la interacción del usuario.

[0025] El códec de metadatos de objeto (véase el codificador OAM 124 y el decodificador OAM 224) se proporciona de manera que, para cada objeto, los metadatos asociados que especifican la posición geométrica y el volumen de los objetos en el espacio 3D se codifican de manera eficiente por la cuantificación de las propiedades del objeto en tiempo y espacio. Los metadatos de objeto comprimidos cOAM 126 se transmiten al receptor 200 como información lateral.

[0026] El renderizador de objeto 216 usa los metadatos de objeto comprimidos para generar formas de onda de objetos según el formato de reproducción dado. Cada objeto se renderiza a un canal de salida determinado en función de sus metadatos. La salida de este bloque procede de la suma de los resultados parciales. Si se decodifican tanto el contenido basado en el canal, como también los objetos discretos/paramétricos, las formas de onda basadas en el canal y las formas de onda de los objetos renderizados son mezcladas por el mezclador 226 antes de producir las formas de onda resultantes 228 o antes de suministrar a un módulo post-renderizador como el renderizador binaural 236 o el módulo de posprocesador de altavoz 232.

[0027] El módulo de renderizador binaural 236 produce una submezcla binaural del material de audio de canales múltiples de manera que cada canal de entrada esté representado por una fuente de sonido virtual. El tratamiento se realiza trama a trama en el dominio QMF (Banco de Filtro de Espejo de Cuadratura), y la binauralización se basa en las respuestas de impulso a espacio binaurales medidas.

[0028] El renderizador de altavoz 232 convierte entre la configuración del canal de transmisión 228 y el formato de reproducción deseado. También puede denominarse “convertidor de formato”. El convertidor de formato realiza conversiones para reducir el número de canales de salida, es decir, crea las submezclas.

[0029] La Fig. 3 muestra un ejemplo para implementar un convertidor de formato 232. El convertidor de formato 232, también denominado renderizador de altavoz, convierte entre la configuración del canal transmisor y el formato de reproducción deseado. El convertidor de formato 232 realiza conversiones a un menor número de canales de salida, es decir, realiza un proceso de submezcla (DMX) 240. El submezclador 240, que opera preferentemente en el dominio QMF, recibe las señales de salida del mezclador 228 y produce las señales del altavoz 234. Puede proporcionarse un configurador 242, también denominado controlador, que recibe, como una entrada de control, una señal 246 indicativa de la distribución de salidas del mezclador, es decir, la disposición para la que se determinan los datos representados por la señal de salida del mezclador 228, y la señal 248 indicativa de la disposición de reproducción deseada. Basándose en esta información, el controlador 242, preferentemente de forma automática, genera matrices de submezcla optimizadas para la combinación dada de formatos de entrada y de salida y aplica estas matrices para el submezclador 240. El convertidor de formato 232 permite configuraciones de altavoces estándar, así como configuraciones aleatorias con posiciones de altavoces no estándar.

[0030] La Fig. 4 ilustra una realización del renderizador binaural 236 de la Fig. 2. El módulo de renderizador binaural puede proporcionar una submezcla binaural del material de audio de canales múltiples. La binauralización puede basarse en las respuestas al impulso del ambiente binaural medidas. Las respuestas al impulso del ambiente pueden considerarse una “huella digital” de las propiedades acústicas de un ambiente real. Las respuestas al impulso del ambiente se miden y se almacenan, y pueden proporcionarse señales acústicas arbitrarias con esta “huella digital”, permitiendo de ese modo al oyente una simulación de las propiedades acústicas del ambiente asociadas con la respuesta al impulso del ambiente. El renderizador binaural 236 puede configurarse o programarse para la renderización de los canales de salida en dos canales binaurales usando funciones de transferencia relacionadas con la cabeza o respuestas al impulso del ambiente binaurales (BRIR). Por ejemplo, para los dispositivos móviles se desea renderización binaural para auriculares o altavoces conectados a dichos dispositivos móviles. En dichos dispositivos móviles, debido a las limitaciones puede ser necesario limitar el decodificador y la complejidad de la renderización. Además de la omisión de decorrelación en dichos escenarios de tratamiento, puede preferirse llevar a cabo primero una submezcla usando un submezclador 250 a una señal de submezcla intermedia 252, es decir, a un menor número de canales de salida que se traduce en un menor número de canales de entrada para el convertidor binaural real 254. Por ejemplo, un material de 22.2 canales puede submezclarse mediante el submezclador 250 a una submezcla intermedia 5.1 o, alternativamente, la submezcla intermedia puede ser calculada directamente por el decodificador SAOC 220 en la Fig. 2 en una especie de modo de “acceso abreviado”. Entonces, la renderización binaural sólo tiene que aplicar diez HRTF (Funciones de Transferencia Relacionadas con la Cabeza) o funciones BRIR para la renderización de los cinco canales individuales en diferentes posiciones en contraste con la aplicación de 44 funciones HRTF o BRIR si los canales de entrada de 22.2 debieran ser renderizados directamente. Las operaciones de convolución necesarias para la renderización binaural requieren una gran cantidad de energía de tratamiento y, por lo tanto, reducir esta energía de tratamiento, obteniendo aún una calidad de audio aceptable, particularmente útil para dispositivos móviles. El renderizador binaural 236 produce una submezcla binaural 238 del material de audio de canales múltiples 228, de manera que cada canal de entrada (con exclusión de los canales LFE) esté representado por una fuente de sonido virtual. El tratamiento puede llevarse a cabo trama a trama en el dominio QMF. La binauralización se basa en las respuestas al impulso del ambiente

binaurales medidas, y el sonido directo y las reflexiones tempranas pueden imprimirse en el material de audio a través de un enfoque de convolución en un dominio pseudo-FFT usando una convolución rápida en la parte superior del dominio QMF, mientras que la reverberación tardía puede tratarse por separado.

5 **[0031]** La Fig. 5 muestra un ejemplo de una respuesta al impulso del ambiente $h(t)$ 300. La respuesta al impulso del ambiente comprende tres componentes, el sonido directo 301, las reflexiones tempranas 302 y la reverberación tardía 304. Por lo tanto, la respuesta al impulso del ambiente describe el comportamiento de las reflexiones de un espacio acústico reverberante cerrado cuando se reproduce un impulso. Las reflexiones tempranas 302 son reflexiones discretas con aumento de la densidad, y la parte de la respuesta al impulso en el que las reflexiones individuales ya no pueden ser discriminadas se denomina reverberación tardía 304. El sonido directo 301 se puede identificar fácilmente en la respuesta al impulso del ambiente y se puede separar de las reflexiones tempranas, sin embargo, la transición de la reflexión temprana 302 a la reverberación tardía 304 es menos evidente.

15 **[0032]** Como se ha descrito anteriormente, en un renderizador binaural, por ejemplo, un renderizador binaural tal como se representa en la Fig. 2, se conocen enfoques diferentes para el tratamiento de una señal de entrada de audio de canales múltiples según una respuesta al impulso del ambiente.

[0033] La Fig. 6 muestra diferentes posibilidades para tratar una señal de entrada de audio con una respuesta al impulso del ambiente. La Fig. 6(a) muestra el tratamiento de la señal de audio completa, según la respuesta al impulso del ambiente, y la Fig. 6(b) muestra el tratamiento separado de la parte temprana y la parte de reverberación tardía. Como se muestra en la Fig. 6(a) una señal de entrada 400, por ejemplo, una señal de entrada de audio de canales múltiples, se recibe y se aplica a un procesador 402 que está configurado o programado para permitir una convolución completa de la señal de entrada de audio de canales múltiples 400 con la respuesta al impulso del ambiente (véase la Fig. 5) que, en la realización representada, obtiene la señal de salida de audio de 2 canales 404. Como se mencionó anteriormente, este enfoque se considera desventajoso dado que el uso de la convolución para la totalidad de la respuesta al impulso es computacionalmente muy costoso. Por lo tanto, según otro enfoque, como se representa en la Fig. 6 (b), en lugar de tratar toda la señal de entrada de audio de canales múltiples mediante la aplicación de una convolución completa con una respuesta al impulso del ambiente como se ha descrito con respecto a la Fig. 6(a), el tratamiento se separa con respecto a las partes tempranas 301, 302 (véase la Fig. 5) de la respuesta al impulso del ambiente 300, y la parte de reverberación tardía 302. Más en concreto, como se muestra en la Fig. 6(b), se recibe la señal de entrada de audio de canales múltiples 400, sin embargo, la señal se aplica en paralelo a un primer procesador 406 para el tratamiento de la parte temprana, es decir, para el tratamiento de la señal de audio según el sonido directo 301 y las reflexiones tempranas 302 en la respuesta al impulso del ambiente 300 como se muestra en la Fig. 5. La señal de entrada de audio de canales múltiples 400 también se aplica a un procesador 408 para tratar la señal de audio según la reverberación tardía 304 de la respuesta al impulso del ambiente 300. En la realización representada en la Fig. 6(b) la señal de entrada de audio de canales múltiples también se puede aplicar a un submezclador 410 para la submezcla de la señal de canales múltiples 400 a una señal que tiene un menor número de canales. La salida del submezclador 410 se aplica entonces al procesador 408. Las salidas de los procesadores 406 y 408 se combinan en 412 para generar la señal de salida de audio de 2 canales 404'.

[0034] En un renderizador binaural, como se mencionó anteriormente, se puede desear tratar el sonido directo y las reflexiones tempranas por separado de la reverberación tardía, principalmente a causa de la reducción de la complejidad computacional. El tratamiento del sonido directo y las reflexiones tempranas puede, por ejemplo, 45 imprimirse en la señal de audio por un enfoque convolucional llevado a cabo por el procesador 406 (véase la Fig. 6 (b)) mientras que la reverberación tardía puede ser sustituido por una reverberación sintética proporcionada por el procesador 408. La señal de salida global binaural 404' es entonces una combinación del resultado convolucional proporcionado por el procesador 406 y la señal reverberada sintética proporcionada por el procesador 408.

50 **[0035]** Este tratamiento también se describe en la referencia de la técnica anterior [1]. El resultado del enfoque anteriormente descrito debe ser perceptualmente lo más idéntico posible idéntico al resultado de una convolución de la respuesta al impulso completa, el enfoque de conversión completa descrito con respecto a la Fig. 6(a). Sin embargo, si una señal de audio o, más en general, un material de audio se convolucionan con el sonido directo y una parte de reflexión temprana de la respuesta al impulso, los diferentes canales resultantes se suman 55 para formar una señal de sonido general que se asocia con la señal de reproducción a un oído del oyente. La reverberación, sin embargo, no se calcula a partir de esta señal general, sino que es en general una señal reverberada de un canal o de la submezcla de la señal de audio de entrada original. Por tanto, los autores de la presente invención han determinado que la reverberación tardía no se ajusta adecuadamente al resultado de la convolución proporcionada por el procesador 406. Se ha descubierto que el nivel adecuado de reverberación 60 depende tanto de la señal de audio de entrada como de las respuestas al impulso del ambiente 300. La influencia de las respuestas de impulso se consigue mediante el uso de características de reverberación como parámetro de entrada de un reverberador que puede ser parte del procesador 408, y estos parámetros de entrada se obtienen de un análisis de respuestas de impulsos medidas como, por ejemplo, el tiempo de reverberación dependiente de la frecuencia y la medida de la energía dependiente de la frecuencia. Estas medidas, en general, pueden determinarse

a partir de una única respuesta al impulso, por ejemplo mediante el cálculo de la energía y el tiempo de reverberación RT60 en un análisis de banco de filtros de octava, o son valores medios de los resultados de los análisis de respuesta al impulso múltiple.

5 **[0036]** Sin embargo, se ha descubierto que a pesar de estos parámetros de entrada proporcionados al reverberador, la influencia de la señal de audio de entrada en la reverberación no se conserva completamente cuando se usa un enfoque de reverberación sintética como se describe con respecto a la Fig. 6(b). Por ejemplo, debido a la submezcla usada para la generación de la cola de reverberación sintética, se pierde la influencia de la señal de audio de entrada. Por consiguiente, el nivel resultante de la reverberación no es perceptualmente idéntico
10 al resultado del enfoque de convolución completo, especialmente en el caso en que la señal de entrada comprende múltiples canales.

[0037] Hasta el momento, no existen enfoques conocidos que comparen la cantidad de reverberación tardía con los resultados del enfoque de convolución completa o coincidan con el resultado de convolución. Existen
15 algunas técnicas que tratan de evaluar la calidad de la reverberación tardía o cuán natural parece. Por ejemplo, en un procedimiento se define una medida de sonoridad para la reverberación de sonido natural, que predice la sonoridad percibida de reverberación usando un modelo de sonoridad. Este enfoque se describe en la referencia de la técnica anterior [2], y el nivel puede ajustarse a un valor objeto. La desventaja de este enfoque es que se basa en un modelo del oído humano que es complicado e inexacto. También es necesario un volumen objeto para
20 proporcionar un factor de cambio de escala para la reverberación tardía que se podría encontrar usando el resultado de convolución completo.

[0038] En otro procedimiento descrito en la referencia de la técnica anterior [3] se usa un criterio de correlación cruzada para pruebas de calidad de reverberación artificial. Sin embargo, esto sólo es aplicable para
25 probar diferentes algoritmos de reverberación, pero no para el audio de canales múltiples, ni para el audio binaural ni para la calificación del cambio de escala de reverberación tardía.

[0039] Otro enfoque posible consiste en usar el número de canales de entrada en el oído considerado como un factor de cambio de escala, sin embargo, esto no produce un cambio de escala perceptualmente correcto, porque
30 la amplitud percibida de la señal de sonido en general depende de la correlación de los diferentes canales de audio y no sólo del número de canales.

[0040] Por lo tanto, se proporciona un procedimiento de cambio de escala dependiente de la señal que se adapta al nivel de reverberación según la señal de audio de entrada. Como se mencionó anteriormente, se desea
35 que el nivel percibido de la reverberación coincida con el nivel de reverberación cuando se usa el enfoque de convolución completo para la renderización binaural, y por lo tanto la determinación de una medida para un nivel adecuado de reverberación es importante para lograr una buena calidad de sonido. Según las realizaciones, una señal de audio se trata por separado con una parte temprana y una reverberación tardía de la respuesta al impulso del ambiente, en el que el tratamiento de la reverberación tardía comprende la generación de una señal reverberada
40 con cambio de escala, en el que el cambio de escala depende de la señal de audio. La parte temprana tratada de la señal de audio y la señal reverberada con cambio de escala se combinan en la señal de salida. Según una realización, el cambio de escala depende de la condición de los uno o más canales de entrada de la señal de audio (por ejemplo, el número de canales de entrada, el número de canales de entrada activos y/o la actividad en el canal de entrada). Según otra realización el cambio de escala depende de una medida de correlación predefinida o
45 calculada para la señal de audio. En realizaciones alternativas puede realizarse el cambio de escala basándose en una combinación de la condición de los uno o más canales de entrada y la medida de correlación predefinida o calculada.

[0041] Según las realizaciones la señal reverberada con cambio de escala puede ser generada mediante la
50 aplicación de un factor de ganancia que se determina basándose en la condición de los uno o más canales de entrada de la señal de audio, o basándose en la medida de correlación predefinida o calculada para la señal de audio, o basándose en una combinación de los mismos.

[0042] Según las realizaciones, el tratamiento por separado de la señal de audio comprende el tratamiento de
55 la señal de audio con la parte de reflexión temprana 301, 302 de la respuesta al impulso del ambiente 300 durante un primer proceso, y el tratamiento de la señal de audio con la reverberación difusa 304 de la respuesta al impulso del ambiente 300 durante un segundo proceso que es diferente y separado del primer proceso. El cambio del primer proceso al segundo proceso se produce en el momento de transición. Según realizaciones adicionales, en el segundo proceso la reverberación difusa (tardía) 304 puede ser sustituida por una reverberación sintética. En este
60 caso la respuesta al impulso del ambiente aplicada al primer proceso contiene sólo la parte de reflexión temprana 300, 302 (véase la Fig. 5) y la reverberación difusa tardía 304 no está incluida.

[0043] A continuación se describirá una realización del enfoque de la invención con más detalle según el cual el factor de ganancia se calcula basándose en un análisis de correlación de la señal de audio de entrada. La Fig. 7

muestra un diagrama de bloques de una unidad de tratamiento de señales, como un renderizador binaural. El renderizador binaural 500 comprende una primera rama que incluye el procesador 502 que recibe desde una entrada 504 la señal de audio $x[k]$ que incluye los N canales. El procesador 502, cuando forma parte de un renderizador binaural, trata la señal de entrada 504 para generar la señal de salida 506 $x_{conv}[k]$. Más en concreto, el
 5 procesador 502 causa una convolución de la señal de entrada de audio 504 con un sonido directo y las reflexiones tempranas de la respuesta al impulso del ambiente que se pueden proporcionar al procesador 502 a partir de una base de datos externa 508 que contiene una serie de respuestas al impulso del ambiente binaurales grabadas. El procesador 502, como se ha mencionado, puede operar basándose en las respuestas al impulso del ambiente binaural proporcionadas por la base de datos 508, produciendo de este modo la señal de salida 502 que tiene sólo
 10 dos canales. La señal de salida 506 se proporciona desde el procesador 502 a un sumador 510. La señal de entrada 504 se proporciona además a una rama de reverberación 512 que incluye el procesador del reverberador 514 y un submezclador 516. La señal de entrada de submezcla se proporciona al reverberador 514 que basándose en los parámetros del reverberador, como la reverberación RT60 y la energía de reverberación mantenida en las bases de datos 518 y 520, respectivamente, genera una señal reverberada $r[k]$ en la salida del reverberador 514 que puede
 15 incluir sólo dos canales. Los parámetros almacenados en las bases de datos 518 y 520 se pueden obtener a partir de las respuestas al impulso del ambiente binaural almacenadas por un análisis apropiado 522 como se indica en líneas de puntos en la Fig. 7.

[0044] La rama de reverberación 512 incluye además un procesador de análisis de correlación 524 que
 20 recibe la señal de entrada 504 y genera un factor de ganancia g en su salida. Además, se proporciona una etapa de ganancia 526 que está acoplada entre el reverberador 514 y el sumador 510. La etapa de ganancia 526 es controlada por el factor de ganancia g , generando de este modo en la salida de la etapa de ganancia 526 la señal reverberada con cambio de escala $r_g[k]$ que se aplica al sumador 510. El sumador 510 combina la parte tratada temprana y la señal reverberada para proporcionar la señal de salida $y[k]$ que también incluye dos canales.
 25 Opcionalmente, la rama de reverberación 512 puede comprender un filtro de paso bajo 528 acoplado entre el procesador 524 y la etapa de ganancia para suavizar el factor de ganancia en una serie de tramas de audio. Opcionalmente, también puede proporcionarse un elemento de retardo 530 entre la salida de la etapa de ganancia 526 y el sumador 510 para retrasar la señal reverberada con cambio de escala de manera que coincida con una transición entre la reflexión temprana y la reverberación tardía en la respuesta al impulso del ambiente.

[0045] Como se describió anteriormente, la Fig. 7 es un diagrama de bloques de un renderizador binaural que trata el sonido directo y las reflexiones por separado de la reverberación tardía. Como se puede ver, la señal de entrada $x[k]$ que se trata con los reflexiones directas y tempranas de los resultados de respuesta al impulso del ambiente binaurales da lugar a una señal $x_{conv}[k]$. Esta señal, como se muestra, se envía al sumador 510 para
 35 añadirla a un componente de señal de reverberante $r_g[k]$. Esta señal es generada por el suministro a una submezcla, por ejemplo, una submezcla estéreo, de la señal de entrada $x[k]$ al reverberador 514 seguido por el multiplicador o etapa de ganancia 526 que recibe una señal reverberada $r[k]$ de la submezcla y el factor de ganancia g . El factor de ganancia g se obtiene por un análisis de correlación de la señal de entrada $x[k]$ llevado a cabo por el procesador 524, y como se mencionó anteriormente se puede suavizar en el tiempo por medio del filtro de paso bajo 528. El
 40 componente reverberante con cambio de escala o ponderado puede ser retrasado opcionalmente por el elemento de retardo 530 para que su comienzo coincida con el punto de transición de las reflexiones tempranas a reverberación tardía de modo que en la salida del sumador 510 se obtenga la señal de salida $y[k]$.

[0046] El renderizador binaural de canales múltiples representado en la Fig. 7 introduce una reverberación
 45 tardía sintética de 2 canales y para superar los inconvenientes anteriormente expuestos de los enfoques convencionales y según el enfoque de la invención, la reverberación tardía sintética se cambia de escala mediante el factor de ganancia g para que coincida la percepción con un resultado de un enfoque de convolución completa. La superposición de múltiples canales (por ejemplo, hasta 22.2) en el oído de un oyente depende de la correlación. Por este motivo, la reverberación tardía puede cambiarse de escala según la correlación del canal de señal de entrada, y
 50 las realizaciones del enfoque de la invención proporcionan un procedimiento de cambio de escala dependiente del tiempo basado en la correlación que determina una amplitud adecuada de la reverberación tardía.

[0047] Para el cálculo de los factores de cambio de escala, se introduce una medida de correlación que se basa en el coeficiente de correlación y según las realizaciones se define en un dominio del tiempo-frecuencia de dos
 55 dimensiones, por ejemplo, el dominio QMF. Se calcula un valor de correlación entre -1 y 1 para cada trama de audio multidimensional, cada trama de audio se define por un número de bandas de frecuencia N, un número de intervalos de tiempo M por trama y un número de canales de audio A. Se obtiene un factor de cambio de escala por trama por cada oído.

[0048] A continuación, se describirá con más detalle una realización. En primer lugar, se hace referencia a la medida de correlación usada en el procesador de análisis de correlación 524 de la Fig. 7. La medida de correlación, según esta realización, se basa en el Coeficiente del Producto-Momento de Pearson (también denominado coeficiente de correlación) que se calcula dividiendo la covarianza de dos variables X , Y por el producto de sus desviaciones típicas:

$$\rho_{\{X,Y\}} = \frac{E\{(X - \bar{X}) \cdot (Y - \bar{Y})\}}{\sigma_X \cdot \sigma_Y}$$

en el que

5

$E\{\cdot\}$ = operador de valor esperado

$\rho_{\{X,Y\}}$ = coeficiente de correlación,

σ_X, σ_Y = desviaciones típicas de las variables X, Y

- 10 **[0049]** Este tratamiento según la realización descrita se transfiere a dos dimensiones en un dominio del tiempo-frecuencia, por ejemplo, el dominio QMF. Las dos dimensiones son los intervalos de tiempo y las bandas de QMF. Este enfoque es razonable, porque los datos a menudo se codifican y transmiten también en el dominio del tiempo-frecuencia. El operador de expectativa se sustituye con una operación media sobre varias muestras de tiempo y/o frecuencia de modo que la medida de correlación de tiempo-frecuencia entre dos variables cero-media
- 15 x_m, x_n en el intervalo de (0, 1) se define de la siguiente manera:

$$\rho[m,n] = \left| \frac{1}{(N-1)} \cdot \frac{\sum_i \sum_j x_m[i,j] \cdot x_n[i,j]^*}{\sum_j \sigma(x_m[j]) \cdot \sigma(x_n[j])} \right|$$

en el que

20

$\rho[m, n]$ = coeficiente de correlación,

$\sigma(x_m[j])$ = desviación típica en un intervalo de tiempo j del canal m ,

$\sigma(x_n[j])$ = desviación típica en un intervalo de tiempo j del canal n ,

x_m, x_n = variables de media cero,

25 $i \forall [1, N]$ = bandas de frecuencia,

$j \forall [1, M]$ = intervalos de tiempo,

$m, n \forall [1, K]$ = canales,

* = conjugado complejo.

- 30 **[0050]** Después del cálculo de este coeficiente para una pluralidad de combinaciones de canales (m,n) de una trama de audio, los valores de $\rho[m,n,t_i]$ se combinan para una medida de correlación única $\rho_m(t_i)$ tomando la media de (o promediando) una pluralidad de valores de correlación $\rho[m,n,t_i]$. Se observa que la trama de audio puede comprender 32 intervalos de tiempo QMF, y t_i indica la trama de audio respectiva. El tratamiento anterior se puede resumir para una trama de audio de la siguiente manera:

35

(i) En primer lugar, se calcula el valor medio global $\bar{x}(k)$ para cada uno de los canales k de la trama de audio o datos x que tiene un tamaño de $[N,M,K]$, en el que según las realizaciones todos los canales k se submezclan en un canal de entrada del reverberador.

(ii) Se calcula una trama de media cero o datos restando los valores $\bar{x}(k)$ de los canales correspondientes.

- 40 (iii) Para una pluralidad de combinación de canales (m,n) se calcula el coeficiente de correlación o valor de correlación definido c .

(iv) Se calcula un valor de correlación medio c_m como la media de una pluralidad de valores de correlación $\rho[m,n]$ (excluyendo los valores calculados erróneamente, por ejemplo una división por cero).

- 45 **[0051]** Según la realización descrita anteriormente el cambio de escala se determinó basándose en la medida de correlación calculada para la señal de audio. Esto es ventajoso, a pesar de los recursos computacionales adicionales necesarios, por ejemplo, cuando se desea obtener la medida de correlación de la señal de audio que se está tratando individualmente.

- 50 **[0052]** Sin embargo, la presente invención no se limita a dicho enfoque. Según otras realizaciones, en vez de calcular la medida de correlación también puede usarse una medida de correlación predefinida. El uso de una medida de correlación predefinida es ventajoso ya que reduce la complejidad computacional en el proceso. La medida de correlación predefinida puede tener un valor fijo, por ejemplo, de 0,1 a 0,9, que se puede determinar empíricamente basándose en un análisis de una pluralidad de señales de audio. En tal caso, el análisis de
- 55 correlación 524 puede ser omitido y la ganancia de la etapa de ganancia puede establecerse por medio de una señal de control apropiada.

[0053] Según otras realizaciones, el cambio de escala puede depender de la condición de los uno o más

canales de entrada de la señal de audio (por ejemplo, el número de canales de entrada, el número de canales de entrada activos y/o la actividad en el canal de entrada). Esto es ventajoso debido a que el cambio de escala se puede determinar fácilmente a partir de la señal de audio de entrada con una sobrecarga de cálculo reducida. Por ejemplo, el cambio de escala se puede determinar simplemente determinando el número de canales en la señal de audio original que se mezclan en un canal de submezcla actualmente considerado que incluye un número reducido de canales cuando se compara con la señal de audio original. Alternativamente, el número de canales activos (canales que muestran algún tipo de actividad en una trama de audio actual) submezclados en el canal de submezcla actualmente considerado pueden formar la base para cambiar de escala la señal reverberada. Esto se puede llevar a cabo en el bloque 524.

[0054] A continuación, se describirá en detalle una realización que determina el cambio de escala de la señal reverberada basándose en la condición de los uno o más canales de entrada de la señal de audio y basándose en una medida de correlación (ya sea fija o calculada como se describe anteriormente). Según dicha realización, el factor de ganancia o factor de ganancia o de cambio de escala g se define de la siguiente manera:

$$g = c_u + \rho \cdot (c_c - c_u)$$

$$c_u = 10^{\frac{10 \cdot \log_{10}(K_{in})}{20}} = \sqrt{K_{in}}$$

$$c_c = 10^{\frac{20 \cdot \log_{10}(K_{in})}{20}} = K_{in}$$

en el que

ρ = coeficiente de correlación predefinida o calculada para la señal de audio,

c_u, c_c = factores indicativos de la condición de los uno o más canales de entrada de la señal de audio, refiriéndose c_u a los canales totalmente no correlacionados y c_c a los canales totalmente correlacionados, K_{in} = número de canales de submezcla activos distintos de cero o fijos.

c_u es el factor que se aplica si los canales de submezcla son totalmente no correlacionados (sin dependencias entre canales). Si se usa sólo la condición de los uno o más canales de entrada $g = c_u$ y el coeficiente de correlación fijo predefinido se establece en cero, c_c es el factor que se aplica si los canales submezclados están totalmente correlacionados (las señales son versiones ponderadas (más desfase y desplazamiento) de cada uno). Si se usa sólo la condición de los uno o más canales de entrada $g = c_c$, el coeficiente de correlación fijo predefinido se establece en uno. Estos factores describen el cambio de escala mínimo y máximo de la reverberación tardía en la trama de audio (dependiendo del número de canales activos).

[0055] El "número de canal" K_{in} se define, según las realizaciones, de la siguiente manera: una señal de audio de canales múltiples se submezcla con una submezcla estéreo usando una matriz de submezcla \mathbf{Q} que define qué canales de entrada se incluyen en cada canal de submezcla (tamaño $M \times 2$, siendo M el número de canales de entrada del material de entrada de audio, por ejemplo, 6 canales para una configuración de 5.1).

[0056] Un ejemplo para la matriz de submezcla \mathbf{Q} puede ser el siguiente:

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.7071 & 0.7071 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

[0057] Para cada uno de los dos canales de submezcla el coeficiente de cambio de escala se calcula de la siguiente manera:

$$g = f(c_c, c_u, \rho_{avg}) = c_u + \rho_{avg} \cdot (c_c - c_u)$$

en el que ρ_{avg} es el valor promedio/media de todos los coeficientes de correlación $\rho [m, n]$ para una serie de combinaciones de canales $K_{in} \cdot K_{in} [m, n]$ y en el que c_c, c_u dependen del número de canal K_{in} , que puede ser de la siguiente manera:

$\bullet K_{in}$ puede ser el número de canales que se submezclan al canal de submezcla actualmente considerado $k \in [1,2]$

(el número de filas en la matriz de submezcla \mathbf{Q} en la columna k que contiene valores no iguales a cero). Este número es invariante en el tiempo debido a que la matriz de submezcla \mathbf{Q} está predefinida para una configuración de canal de entrada y no cambia en la longitud de una señal de entrada de audio.

Por ejemplo, cuando se considera una señal de entrada de 5.1 se aplica lo siguiente:

5

- los canales 1, 3, 4 se submezclan al canal de submezcla 1 (véase la matriz \mathbf{Q} más arriba),
- $K_n = 3$ en cada trama (3 canales)

• K_n puede ser el número de canales activos que se submezclan en el canal de submezcla actualmente considerado
 10 $k \in [1,2]$ (número de canales de entrada en el que hay actividad en la trama de audio actual y en el que la fila correspondiente de la matriz de submezcla \mathbf{Q} en la columna k contiene un valor no igual a cero \rightarrow número de canales en la intersección de canales activos y elementos no iguales en la columna k de \mathbf{Q}). Este número puede ser variable en el tiempo en la longitud de una señal de entrada de audio, porque incluso si \mathbf{Q} permanece igual, la actividad de la señal puede variar con el tiempo.

15 Por ejemplo, cuando se considera una señal de entrada de 5.1 se aplica lo siguiente:

- los canales 1, 3, 4 se submezclan al canal de submezcla 1 (véase la matriz \mathbf{Q} más arriba),
- En la trama n :

20

- los canales activos son los canales 1, 2, 4,
- K_n es el número de canales en la intersección {1, 4},
- $K_n(n) = 2$

- En la trama $n + 1$:

25

- los canales activos son los canales 1, 2, 3, 4
- K_n es el número de canales en la intersección {1, 3, 4},
- $K_n(n+1) = 3$.

30 **[0058]** Un canal de audio (en la trama predefinida) puede considerarse activo en caso de que tenga una amplitud o una energía dentro de la trama predefinida que sea superior a un valor de umbral predeterminado, por ejemplo, según las realizaciones, una actividad en un canal de audio (en una trama predefinida) se puede definir de la siguiente manera:

- 35 • la suma o el valor máximo de las amplitudes absolutas de la señal (en el dominio del tiempo, dominio QMF, etc.) en la trama es mayor que cero, o
- la suma o el valor máximo de la energía de la señal (valor absoluto cuadrado de las amplitudes en el dominio del tiempo o dominio QMF) en la trama es mayor que cero.

40 **[0059]** En lugar de cero también puede usarse otro umbral (relativo a la energía o amplitud máxima) mayor que cero, por ejemplo, un umbral de 0,01.

[0060] Según las realizaciones, se proporciona un factor de ganancia para cada oído que depende del número de canales activos (que varían en el tiempo) o el número fijo de canales incluidos (matriz de submezcla no igual a cero) K_n en el canal de submezcla. Se supone que el factor aumenta linealmente entre el caso totalmente no correlacionado y el caso totalmente correlacionado. Totalmente no correlacionado significa sin dependencias entre canales (el valor de correlación es cero) y totalmente correlacionado significa que las señales son versiones ponderadas de cada uno (con diferencia de fase de desplazamiento, el valor de correlación es uno).

50 **[0061]** Como se mencionó anteriormente, la ganancia o factor de cambio de escala g puede suavizarse en las tramas de audio por medio del filtro de paso bajo 528. El filtro de paso bajo 528 puede tener una constante de tiempo de t_s que da lugar a un factor de ganancia suavizado de $g_s(t)$ para un tamaño de trama k de la siguiente manera:

55

$$g_s(t_i) = c_{s,old} \cdot g_s(t_i - 1) + c_{s,new} \cdot g$$

$$c_{s,old} = e^{-\left(\frac{1}{f_s \cdot t_s \cdot k}\right)}$$

$$c_{s,new} = 1 - c_{s,old}$$

60

en el que

t_s = constante de tiempo del filtro de paso bajo en [s]

t_i = trama de audio en la trama t_i

5 g_s = factor de ganancia suavizado

k = tamaño de la trama, y

f_s = frecuencia de muestreo en [Hz]

[0062] El tamaño de la trama k puede ser el tamaño de una trama de audio en muestras en el dominio del
10 tiempo, por ejemplo, 2.048 muestras.

[0063] La señal reverberada del canal izquierdo de la trama de audio $x(t_i)$ se somete a un cambio de escala por medio del factor $g_{s,left}(t_i)$, y la señal reverberada del canal derecho se somete a un cambio de escala por medio del factor $g_{s,right}(t_i)$. El factor de cambio de escala se calcula una vez con K_{in} como el número de canales (activos no
15 cero o número total de canales) que están presentes en el canal izquierdo de la submezcla estéreo que se suministra al reverberador resultante en el factor de cambio de escala $g_{s,left}(t_i)$. Entonces el factor de cambio de escala se calcula una vez más con K_{in} como el número de canales (activos no cero o número total de canales) que están presentes en el canal izquierdo de la submezcla estéreo que se suministra al reverberador resultante en el factor de cambio de escala $g_{s,right}(t_i)$. El reverberador devuelve una versión reverberada estéreo de la trama de audio.
20 El canal izquierdo de la versión reverberada (o el canal izquierdo de la entrada del reverberador) se somete a un cambio de escala con $g_{s,left}(t_i)$ y el canal derecho de la versión reverberada (o el canal derecho de la entrada de la reverberador) se somete a un cambio de escala con $g_{s,right}(t_i)$.

[0064] La reverberación tardía (sintética) artificial con cambio de escala se aplica al sumador 510 que se
25 añade a la señal 506 que ha sido tratada con el sonido directo y las reflexiones tempranas.

[0065] Como se mencionó anteriormente, el enfoque de la invención, según las realizaciones puede usarse en un procesador binaural para el tratamiento binaural de señales de audio. A continuación se describirá una
30 realización del tratamiento binaural de señales de audio. El tratamiento binaural puede llevarse a cabo como un proceso de decodificador que convierte la señal decodificada en una señal de submezcla binaural que proporciona una experiencia de sonido envolvente cuando se escucha con auriculares.

[0066] La Fig. 8 muestra una representación esquemática de un renderizador binaural 800 para el tratamiento binaural de señales de audio según una realización de la presente invención. La Fig. 8 también proporciona una
35 visión general del tratamiento del dominio QMF en el renderizador binaural.

[0067] En una entrada 802 el renderizador binaural 800 recibe la señal de audio que se tratará, por ejemplo, una señal de entrada que incluye N canales y 64 bandas QMF. Además, el renderizador binaural 800 recibe una serie de parámetros de entrada para controlar el tratamiento de la señal de audio. Los parámetros de entrada
40 incluyen la respuesta al impulso del ambiente binaural (BRIR) 804 para los 2xN canales y 64 bandas QMF, una indicación K_{max} 806 de la banda máxima que se usa para la convolución de la señal de entrada de audio con la parte de reflexión temprana de BRIR 804, y los parámetros del reverberador 808 y 810 mencionados anteriormente (RT60 y la energía de reverberación). El renderizador binaural 800 comprende un procesador de convolución rápida 812 para tratar la señal de audio de entrada 802 con la parte temprana de las BRIR recibidas 804. El procesador 812
45 genera en una salida la señal temprana tratada 814 que incluye dos canales y K_{max} bandas QMF. El renderizador binaural 800 comprende, además de la rama de tratamiento temprano que tiene el procesador de convolución rápida 812, también una rama de reverberación que incluye dos reverberadores 816a y 816b en los que cada uno recibe como parámetro de entrada la información RT60 808 y la información de energía de reverberación 810. La rama de reverberación incluye además un procesador de submezcla estéreo 818 y un procesador de análisis de correlación
50 así como también la recepción de la señal de audio de entrada 802. Además, se proporcionan dos etapas de ganancia 821a y 821b entre el procesador de submezcla estéreo 818 y los respectivos reverberadores 816a y 816b para controlar la ganancia de una submezcla de la señal 822 proporcionada por el procesador de submezcla estéreo 818. El procesador de submezcla estéreo 818 proporciona basándose en la señal de entrada 802 la señal submezclada 822 que tiene dos bandas y 64 bandas QMF. La ganancia de las etapas de ganancia de 821a y 821b
55 está controlada por señales de control respectivas 824a y 824b proporcionadas por el procesador de análisis de correlación 820. La señal submezclada controlada por ganancia se introduce en los reverberadores respectivos 816a y 816b para generar señales reverberadas respectivas 826a, 826b. La señal tratada temprana 814 y las señales reverberadas 826a, 826b son recibidas por un mezclador 828 que combina las señales recibidas en la señal de audio de salida 830 que tiene dos canales y 64 bandas QMF. Además, el procesador de convolución rápida 812 y
60 los reverberadores 816a y 816b reciben un parámetro de entrada adicional 832 que indica la transición en la respuesta al impulso del ambiente 804 de la parte temprana a la reverberación tardía determinada como se expuso anteriormente.

[0068] El módulo renderizador binaural 800 (por ejemplo, el renderizador binaural 236 de la Fig. 2 o la Fig. 4)

tiene como entrada 802 el tren de datos decodificados. La señal es tratada por un banco de filtros de análisis QMF como se indica en la norma ISO/IEC 14496-3:2009, subcláusula 4.B.18.2 con las modificaciones indicadas en la norma ISO/IEC 14496-3:2009, subcláusula 8.6.4.2. El módulo renderizador 800 también puede tratar datos de entrada en el dominio QMF; en este caso el banco de filtros de análisis puede ser omitido. Las respuestas al impulso del ambiente binaurales (BRIR) 804 se renderizan como filtros complejos en el dominio QMF. La conversión de las respuestas al impulso del ambiente binaurales en el dominio del tiempo a la representación del filtro QMF complejo se describe en la norma ISO/IEC FDIS 23003-1:2006, Anexo B. Las BRIR 804 se limitan a un cierto número de intervalos de tiempo en el dominio QMF complejo, que contienen sólo la parte de reflexión temprana 301, 302 (véase la Fig. 5) y la reverberación difusa tardía 304 no está incluida. El punto de transición 832 de las reflexiones tempranas de reverberación tardía se determina como se ha descrito anteriormente, por ejemplo, por un análisis de la BRIR 804 en un paso de pretratamiento del tratamiento binaural. Las señales de audio en el dominio QMF 802 y las BRIR en el dominio QMF 804 son entonces tratadas por una convolución rápida por bandas 812 para realizar el tratamiento binaural. Un reverberador en el dominio QMF 816a, 816b se usa para generar una reverberación tardía de 2 canales en el dominio QMF 826a, 826b. El módulo de reverberación 816a, 816b usa un conjunto de tiempos de reverberación dependientes de la frecuencia 808 y valores de energía 810 para adaptar las características de la reverberación. La forma de onda de la reverberación se basa en una submezcla estéreo 818 de la señal de entrada de audio 802 y se somete a un cambio de escala de forma adaptativa 821a, 821b en amplitud dependiendo de un análisis de correlación 820 de la señal de audio de canales múltiples 802. El resultado convolucional en el dominio QMF de 2 canales 814 y la reverberación en el dominio QMF de 2 canales 816a, 816b se combinan en 828 y, finalmente, dos bancos de filtros de síntesis QMF calculan las señales de salida en el dominio del tiempo binaural 830 como se indica en la norma ISO/IEC 14496-3:2009, subcláusula 4.6.18.4.2. El renderizador también puede producir datos de salida en el dominio QMF; por tanto, el banco de filtros de síntesis se omite.

DEFINICIONES

[0069] Las señales de audio 802 que se introducen en el módulo renderizador binaural 800 se denominan en lo sucesivo *señales de entrada*. Las señales de audio 830 que son el resultado del tratamiento binaural se denominan *señales de salida*. Las señales de entrada 802 del módulo renderizador binaural 800 son señales de salida de audio del decodificador principal (véanse, por ejemplo, las señales 228 en la Fig. 2). Se usan las siguientes definiciones de variables:

N_{in}	Número de canales de entrada
N_{out}	Número de canales de salida, $N_{out} = 2$
M_{DMX}	Matriz de submezcla que contiene coeficientes de submezcla no negativos con valores reales (ganancias de submezcla). M_{DMX} es de dimensión $N_{out} \times N_{in}$
L	Longitud de la trama medida en muestras de audio en el dominio del tiempo.
v	Índice de muestra en el dominio del tiempo
n	Índice de intervalos de tiempo QMF (índice muestra sub-banda)
L_n	Longitud de la trama medida en intervalos de tiempo QMF
F	Índice de trama (número de trama)
K	Número de bandas de frecuencia QMF, $K = 64$
k	Índice de banda QMF (1 .. 64)
A, B, ch	Índices de canales (número de canal de configuraciones de canal)
L_{trans}	Longitud de la parte de reflexión temprana de la BRIR en muestras en el dominio del tiempo
$L_{trans,n}$	Longitud de la parte de reflexión temprana de la BRIR en intervalos de tiempo QMF
N_{BRIR}	Número de pares de BRIR en un conjunto de datos BRIR
L_{FFT}	Duración de transformada FFT
$\Re(\cdot)$	Parte real de una señal de valor complejo
$\Im(\cdot)$	Parte imaginaria de una señal de valor complejo
\mathbf{m}_{conv}	Vector que señala qué canal de señal de entrada pertenece a cada par de BRIR en el conjunto de datos BRIR
f_{max}	Frecuencia máxima usada para el tratamiento binaural
$f_{max,decoder}$	Frecuencia máxima de la señal que está presente en la señal de salida de audio del decodificador
K_{max}	Banda máxima que se usa para la convolución de la señal de entrada de audio con la parte de reflexión temprana de las BRIR
a	Coficiente de matriz de submezcla
$C_{eq,k}$	Factor de ecualización de energía por bandas
ϵ	Constante numérica, $\epsilon = 10^{-20}$
d	Retardo en los intervalos de tiempo en el dominio QMF
$\tilde{\mathbf{y}}_{ch}^{n,k}$	Representación de la señal en el dominio pseudo-FFT en la banda de frecuencia k

(continuación)

n'	Índice de frecuencia pseudo-FFT
$\tilde{\mathbf{h}}^{n',k}$	Representación en el dominio pseudo-FFT de BRIR en la banda de frecuencia k
$\tilde{\mathbf{z}}_{\text{ch,conv}}^{n',k}$	Resultado de convolución en el dominio pseudo-FFT en la banda de frecuencia k
$\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k}$	Señal intermedia: resultado convolucional de 2 canales en el dominio QMF
$\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k}$	Señal intermedia: reverberación de 2 canales en el dominio QMF
K_{ana}	Número de bandas de frecuencia de análisis (usado para el reverberador)
$f_{c,\text{ana}}$	Frecuencias centrales de las bandas de frecuencia de análisis
$N_{\text{DMX,act}}$	Número de canales que se submezclan en un canal de la submezcla estéreo y están activos en la trama de señal real
c_{corr}	Coefficiente de correlación global para una trama de señal
$c_{\text{corr}}^{A,B}$	Coefficiente de correlación para la combinación de canales A, B
$\sigma_{\hat{y}_{\text{ch,A}}^n}$	Desviación típica para el intervalo de tiempo n de la señal $\hat{y}_{\text{ch,A}}^n$
$\mathbf{c}_{\text{scale}}$	Vector de factor de cambio de escala dos
$\tilde{\mathbf{c}}_{\text{scale}}$	Vector de factor de cambio de escala dos, suavizado en el tiempo

TRATAMIENTO

5 **[0070]** A continuación se describe el tratamiento de la señal de entrada. El módulo renderizador binaural opera en tramas contiguas no superpuestas de longitud $L = 2.048$ muestras en dominio del tiempo de las señales de audio de entrada y produce una trama de muestras L por trama de entrada tratada de longitud L .

(1) Inicialización y pretratamiento

10 **[0071]** La inicialización del bloque de tratamiento binaural se realiza antes de que se lleve a cabo el tratamiento de las muestras de audio suministradas por el decodificador principal (véase, por ejemplo, el decodificador de 200 en la Fig. 2). La inicialización se compone de varias etapas de tratamiento.

(a) Lectura de valores de análisis

15 **[0072]** El módulo reverberador 816a, 816b tiene un conjunto dependiente de la frecuencia de tiempos de reverberación 808 y valores de energía 810 como parámetros de entrada. Estos valores se leen a partir de una interfaz en la inicialización del módulo de tratamiento binaural 800. Además se lee el tiempo de transición 832 de las reflexiones tempranas a reverberación tardía en muestras en el dominio del tiempo. Los valores se pueden almacenar en un archivo binario escrito con 32 bits por muestra, valores flotantes, ordenamiento de tipo little-endian.
20 Los valores de lectura que se necesitan para el tratamiento se indican en la siguiente tabla:

Descripción del valor	Número	Tipo de datos
Longitud de transición L_{trans}	1	Entero
Número de bandas de frecuencia K_{ana}	1	Entero
Frecuencias centrales $f_{c,\text{ana}}$ de bandas de frecuencia	K_{ana}	Flotante
Tiempos de reverberación RT60 en segundos	K_{ana}	Flotante
Valores de energía que representan la energía (amplitud al cuadrado) de la parte de reverberación tardía de una BRIR	K_{ana}	Flotante

(b) Lectura y pretratamiento de BRIR

25 **[0073]** Las respuestas al impulso del ambiente binaural 804 se leen a partir de dos archivos dedicados que almacenan individualmente las BRIR del oído izquierdo y derecho. Las muestras en el dominio del tiempo de las BRIR se almacenan en archivos de onda enteros con una resolución de 24 bits por muestra y 32 canales. El orden de BRIR en el archivo es como se indica en la siguiente tabla:

30

Número de canal	Etiqueta de altavoz
1	CH_M_L045
2	CH_M_R045
3	CH_M_000

(continuación)

4	CH_LFE1
5	CH_M_L135
6	CH_M_R135
7	CH_M_L030
8	CH_M_R030
9	CH_M_180
10	CH_LFE2
11	CH_M_L090
12	CH_M_R090
13	CH_U_L045
14	CH_U_R045
15	CH_U_000
16	CH_T_000
17	CH_U_L135
18	CH_U_R135
19	CH_U_L090
20	CH_U_R090
21	CH_U_180
22	CH_L_000
23	CH_L_L045
24	CH_L_R045
25	CH_M_L060
26	CH_M_R060
27	CH_M_L110
28	CH_M_R110
29	CH_U_L030
30	CH_U_R030
31	CH_U_L110
32	CH_U_R110

[0074] Si no hay BRIR medida en una de las posiciones de los altavoces, el canal correspondiente en el archivo de onda contiene valores cero. Los canales LFE no se usan para el tratamiento binaural.

5 **[0075]** Como un paso de pretratamiento, el conjunto dado de respuestas al impulso del ambiente binaurales (BRIR) se transforma de filtros en el dominio del tiempo a filtros en el dominio QMF de valor complejo. La aplicación de los filtros en el dominio del tiempo dados en el dominio QMF de valor complejo se realiza según la norma ISO/IEC FDIS 23003-1:2006, Anexo B. Los coeficientes del filtro prototipo para la conversión de filtro se usan según ISO/IEC FDIS 23003 -1:2006, Anexo B, Tabla B.1. La representación del dominio del tiempo $\tilde{\mathbf{h}}_{\text{ch}}^v = \left[\tilde{h}_1^v \cdots \tilde{h}_{N_{\text{BRIR}}}^v \right]$ con

10 $1 \leq v \leq L_{\text{trans}}$ se trata para obtener un filtro en el dominio QMF con valor complejo $\hat{\mathbf{h}}_{\text{ch}}^{n,k} = \left[\hat{h}_1^{n,k} \cdots \hat{h}_{N_{\text{BRIR}}}^{n,k} \right]$ con $1 \leq n \leq L_{\text{trans},n}$.

(2) Tratamiento de la señal de audio

15 **[0076]** El bloque de tratamiento de audio del módulo renderizador binaural 800 obtiene muestras de audio en el dominio del tiempo 802 para N_{in} canales de entrada desde el decodificador principal y genera una señal de salida binaural 830 que consiste en $N_{\text{out}} = 2$ canales.

[0077] El tratamiento toma como entrada

- 20
- los datos de audio decodificados 802 desde el decodificador principal,
 - la representación del dominio QMF del complejo de la parte de reflexión temprana de la BRIR fijada 804, y
 - el parámetro dependiente de la frecuencia fijo 808, 810, 832 que es usado por el reverberador en el dominio QMF 816a, 816b para generar la reverberación tardía 826a, 826b.

(a) Análisis de QMF de la señal de audio

5 **[0078]** Como la primera etapa de tratamiento, el módulo renderizador binaural transforma $L = 2.048$ muestras en el dominio del tiempo de la señal de entrada del dominio del tiempo de N_{in} canales (proveniente del decodificador principal) $\left[\tilde{y}_{ch,1}^v \cdots \tilde{y}_{ch,N_{in}}^v \right] = \tilde{\mathbf{y}}_{ch}^v$ en una representación de la señal en el dominio QMF de N_{in} canales 802 de dimensión $L_n = 32$ intervalos de tiempo de QMF (índice de intervalo n) y $K = 64$ bandas de frecuencia (índice de banda k).

10 **[0079]** Se realiza un análisis de QMF como se indica en la norma ISO/IEC 14496-3:2009, 4.B.18.2 subcláusula con las modificaciones establecidas en la norma ISO/IEC 14496-3:2009, subsección 8.6.4.2. en una trama de la señal en el dominio del tiempo $\tilde{\mathbf{y}}_{ch}^v$ para obtener una trama de la señal en el dominio QMF $\left[\hat{y}_{ch,1}^{n,k} \cdots \hat{y}_{ch,N_{in}}^{n,k} \right] = \hat{\mathbf{y}}_{ch}^{n,k}$ con $1 \leq v \leq L$ y $1 \leq n \leq L_n$.

15 (b) Convolución rápida del dominio de señal de audio QMF y las BRIR del dominio QMF

[0080] A continuación, se lleva a cabo una convolución rápida de banda 812 para tratar la señal de audio en el dominio QMF 802 y las BRIR en el dominio QMF 804. Puede realizarse un análisis FFT para cada banda de frecuencia QMF k para cada canal de la señal de entrada 802 y cada BRIR 804.

20

[0081] Debido a los valores complejos en el dominio QMF se realiza un análisis FFT en la parte real de la representación de la señal en el dominio QMF y un análisis FFT en las partes imaginarias de la representación de la señal en el dominio QMF. A continuación, se combinan los resultados para formar la señal en el dominio pseudo-FFT con valor complejo por bandas final

25

$$\tilde{\mathbf{y}}_{ch}^{n',k} = \text{FFT}(\hat{\mathbf{y}}_{ch}^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{y}}_{ch}^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{y}}_{ch}^{n',k}))$$

y las BRIR con valor complejo por bandas

30 $\tilde{\mathbf{h}}_1^{n',k} = \text{FFT}(\hat{\mathbf{h}}_1^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{h}}_1^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{h}}_1^{n',k}))$ para el oído izquierdo

$$\tilde{\mathbf{h}}_2^{n',k} = \text{FFT}(\hat{\mathbf{h}}_2^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{h}}_2^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{h}}_2^{n',k}))$$
 para el oído derecho.

35 **[0082]** La longitud de la transformada FFT se determina según la longitud de los filtros de BRIR en el dominio QMF con valor complejo $L_{trans,n}$ y la longitud de la trama en intervalos de tiempo en el dominio QMF L_n de manera que

$$L_{\text{FFT}} = L_{\text{trans},n} + L_n - 1.$$

40 **[0083]** Las señales en el dominio pseudo-FFT con valor complejo se multiplican por los filtros de BRIR en el dominio pseudo-FFT con valor para formar los resultados de convolución rápida. Se usa un vector \mathbf{m}_{conv} para indicar qué canal de la señal de entrada corresponde a cada par BRIR en el conjunto de datos BRIR.

[0084] Esta multiplicación se realiza por bandas para todas las k bandas de frecuencia QMF con
45 $1 \leq k \leq K_{\max}$. La banda máxima K_{\max} se determina por la banda QMF que representa una frecuencia de 18 kHz o la frecuencia de señal máxima que está presente en la señal de audio del decodificador principal

$$f_{\max} = \min(f_{\max, \text{decoder}}, 18\text{kHz}).$$

50 **[0085]** Los resultados de la multiplicación de cada canal de entrada de audio con cada par BRIR se suman en cada banda de frecuencia QMF k con $1 \leq k \leq K_{\max}$ para producir una señal en el dominio pseudo-FFT K_{\max} de 2

canales intermedia. $\tilde{z}_{ch,1,conv}^{n,k} = \sum_{ch=1}^{ch=N_{in}} \tilde{y}_{ch,ch}^{n,k} \cdot \tilde{h}_{1,m_{conv}[ch]}^{n,k}$ y $\tilde{z}_{ch,2,conv}^{n,k} = \sum_{ch=1}^{ch=N_{in}} \tilde{y}_{ch,ch}^{n,k} \cdot \tilde{h}_{2,m_{conv}[ch]}^{n,k}$ son el resultado de convolución de pseudo-FFT $\tilde{z}_{ch,conv}^{n,k} = \left[\tilde{z}_{ch,1,conv}^{n,k}, \tilde{z}_{ch,2,conv}^{n,k} \right]$ en la banda de frecuencia en el dominio QMF k .

[0086] A continuación, se lleva a cabo una síntesis de FFT por bandas para transformar el resultado de convolución nuevamente al dominio QMF que produce una señal en el dominio QMF intermedia de banda K_{max} de 2 canales con L_{FFT} intervalos de tiempo $\hat{z}_{ch,conv}^{n,k} = \left[\hat{z}_{ch,1,conv}^{n,k}, \hat{z}_{ch,2,conv}^{n,k} \right]$ con $1 \leq n \leq L_{FFT}$ y $1 \leq k \leq K_{max}$.

[0087] Para cada trama de señal de entrada en el dominio QMF $L = 32$ intervalos de tiempo se devuelve una trama de señal de resultado de convolución con $L = 32$ intervalos de tiempo. Los $L_{FFT} - 32$ intervalos de tiempo restantes se almacenan y se lleva a cabo un tratamiento de superposición-suma en la o las tramas siguientes.

(c) Generación de reverberación tardía

[0088] Como segunda señal intermedia 826a, 826B se genera una señal de reverberación denominada $\hat{z}_{ch,rev}^{n,k} = \left[\hat{z}_{ch,1,rev}^{n,k}, \hat{z}_{ch,2,rev}^{n,k} \right]$ por un módulo reverberador en el dominio de la frecuencia 816a, 816b. El reverberador en el dominio de la frecuencia 816a, 816b toma como entrada

- una submezcla estéreo en el dominio QMF 822 de una trama de la señal de entrada,
- un conjunto de parámetros que contiene tiempos de reverberación dependientes de la frecuencia 808 y valores de energía 810.

[0089] El reverberador en el dominio de la frecuencia 816a, 816b devuelve una cola de reverberación tardía en el dominio QMF de 2 canales.

[0090] El número de banda usado máximo del conjunto de parámetros dependientes de la frecuencia se calcula en función de la frecuencia máxima.

[0091] En primer lugar se lleva a cabo una submezcla estéreo de un dominio QMF 818 de una trama de la señal de entrada $\hat{y}_{ch}^{n,k}$ para formar la entrada del reverberador por una suma ponderada de los canales de señal de entrada. Las ganancias de ponderación están contenidas en la matriz de submezcla \mathbf{M}_{DMX} . Tienen valor real y no son negativas y la matriz de submezcla tiene como dimensión $N_{out} \times N_{in}$. Contiene un valor distinto a cero, en el que un canal de la señal de entrada se hace corresponder con uno de los dos canales de salida.

[0092] Los canales que representan las posiciones de los altavoces en el hemisferio izquierdo se hacen corresponder con el canal de salida izquierdo y los canales que representan los altavoces situados en el hemisferio derecho se hacen corresponder con el canal de salida derecho. Las señales de estos canales se ponderan con un coeficiente de 1. Los canales que representan los altavoces situados en el plano medio se hacen corresponder con los dos canales de salida de la señal binaural. Las señales de entrada de estos canales se ponderan con un coeficiente

$$a = 0.7071 \approx \frac{1}{\sqrt{2}}.$$

[0093] Además, se lleva a cabo una etapa de equalización de energía en la submezcla. Se adapta la energía por bandas de un canal de submezcla para que sea igual a la suma de la energía por bandas de los canales de señal de entrada que están contenidos en este canal de submezcla. Esta equalización de energía se lleva a cabo mediante una multiplicación por bandas con un coeficiente de valor real

$$c_{eq,k} = \sqrt{\frac{P_{in}^k}{P_{out}^k + \varepsilon}}.$$

[0094] El factor $c_{eq,k}$ está limitado a un intervalo de $[0,5, 2]$. La constante numérica ε se introduce para evitar una división por cero. La submezcla también se limita en banda a la frecuencia f_{max} ; los valores en todas las bandas de frecuencia más altas se ajustan a cero.

[0095] La Fig. 9 representa esquemáticamente el tratamiento en el reverberador en el dominio de la frecuencia 816a, 816b del renderizador binaural 800.

5 **[0096]** En el reverberador en el dominio de la frecuencia, se calcula una submezcla mono de la entrada estéreo usando un mezclador de entrada 900. Esto se hace aplicando incoherentemente un cambio de fase de 90° en el segundo canal de entrada.

[0097] Esta señal mono se suministra a un bucle de retardo de realimentación 902 en cada banda de frecuencia k , que crea una secuencia de debilitamiento de los impulsos. Se sigue de elementos de decorrelación FIR paralelos que distribuyen la energía de la señal de una manera debilitadora en los intervalos entre los impulsos y crean incoherencia entre los canales de salida. Se aplica una densidad de filtro de debilitamiento para crear el debilitamiento de la energía. Las operaciones de fase de filtro de reducción se limitan a cuatro opciones para implementar un elemento de correlación escaso y sin multiplicadores.

10 **[0098]** Tras el cálculo de la reverberación se incluye una corrección de coherencia entre canales (ICC) 904 en el módulo reverberador para cada banda de frecuencia QMF. En la etapa de corrección ICC se usan ganancias directas dependientes de la frecuencia $g_{directa}$ y ganancias de mezcla cruzada $g_{cruzada}$ para adaptar la ICC.

20 **[0099]** La cantidad de energía y los tiempos de reverberación para las diferentes bandas de frecuencia están contenidos en el conjunto de parámetros de entrada. Los valores se dan en una serie de puntos de frecuencia que se hacen corresponder internamente con las $K = 64$ bandas de frecuencia QMF.

[0100] Se usan dos instancias del reverberador en el dominio de la frecuencia para calcular la señal intermedia final $\hat{\mathbf{z}}_{ch,rev}^{n,k} = \left[\hat{z}_{ch,1,rev}^{n,k}, \hat{z}_{ch,2,rev}^{n,k} \right]$. La señal $\hat{z}_{ch,1,rev}^{n,k}$ es el primer canal de salida de la primera instancia del reverberador, y $\hat{z}_{ch,2,rev}^{n,k}$ es el segundo canal de salida de la segunda instancia del reverberador. Se combinan con la trama de señal de reverberación final que tiene la dimensión de 2 canales, 64 bandas y 32 intervalos de tiempo.

[0101] La submezcla estéreo 822 se somete a un cambio de escala de dos veces 821a,b según una medida de correlación 820 de la trama de la señal de entrada para asegurar el cambio de escala correcto de la salida del reverberador. El factor de cambio de escala se define como un valor en el intervalo de $\left[\sqrt{N_{DMX,act}}, N_{DMX,act} \right]$ linealmente en función de un coeficiente de correlación c_{corr} entre 0 y 1 con

$$c_{corr} = \frac{1}{N_{in}^2} \cdot \sum_{A=1}^{A=N_{DMX,act}} \sum_{B=1}^{B=N_{DMX,act}} c_{corr}^{A,B}$$

35 y

$$c_{corr}^{A,B} = \left| \frac{1}{K-1} \cdot \frac{\sum_k \sum_n \hat{y}_{ch,A}^{n,k} \cdot \hat{y}_{ch,B}^{n,k*}}{\sum_n \sigma_{\hat{y}_{ch,A}^n} \cdot \sigma_{\hat{y}_{ch,B}^n}} \right|$$

40 en el que $\sigma_{\hat{y}_{ch,A}^n}$ significa la desviación típica en un intervalo de tiempo n de canal A , el operador $\{*\}$ denota el conjugado complejo e \hat{y} es la versión de media cero de la señal en el dominio QMF \hat{y} en la trama de la señal real.

[0102] c_{corr} se calcula dos veces: una vez para la pluralidad de canales A,B que están activos en la trama de la señal real F y se incluyen en el canal izquierdo de la submezcla estéreo y una vez para la pluralidad de canales A, B que están activos en la trama de la señal real F y que se incluyen en el canal derecho de la submezcla estéreo. $N_{DMX,act}$ es el número de canales de entrada que reducen la mezcla a un canal de submezcla A (número de elemento de matriz en la A -ésima fila de la matriz de submezcla \mathbf{M}_{DMX} que son diferentes a cero) y que están activos en la trama actual.

50 **[0103]** Entonces, los factores de cambio de escala son

$$\mathbf{c}_{\text{scale}} = \begin{bmatrix} c_{\text{scale},1} \\ c_{\text{scale},2} \end{bmatrix} \\ = \begin{bmatrix} \sqrt{N_{\text{DMX,act},1}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},1} - \sqrt{N_{\text{DMX,act},1}}) \\ \sqrt{N_{\text{DMX,act},2}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},2} - \sqrt{N_{\text{DMX,act},2}}) \end{bmatrix}$$

[0104] Los factores de cambio de escala se suavizan mediante tramas de señal de audio mediante un filtro de paso bajo de primer orden que da lugar a factores de cambio de escala suavizados $\tilde{\mathbf{c}}_{\text{scale}} = \begin{bmatrix} \tilde{c}_{\text{scale},1} \\ \tilde{c}_{\text{scale},2} \end{bmatrix}$.

5

[0105] Los factores de cambio de escala se inicializan en la primera trama de datos de entrada de audio por medio de un análisis de correlación en el dominio del tiempo con las mismas medias.

[0106] La entrada de la primera instancia del reverberador se somete a un cambio de escala con el factor de cambio de escala $\tilde{c}_{\text{scale},1}$ y la entrada de la segunda instancia del reverberador se somete a un cambio de escala con el factor de cambio de escala $\tilde{c}_{\text{scale},2}$.

(d) *Combinación de resultados convolucionales y reverberación tardía*

15 **[0107]** A continuación, el resultado convolucional 814, $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} = \begin{bmatrix} \hat{z}_{\text{ch},1,\text{conv}}^{n,k} \\ \hat{z}_{\text{ch},2,\text{conv}}^{n,k} \end{bmatrix}$, y la salida del reverberador 826a, 826B, $\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k} = \begin{bmatrix} \hat{z}_{\text{ch},1,\text{rev}}^{n,k} \\ \hat{z}_{\text{ch},2,\text{rev}}^{n,k} \end{bmatrix}$, para una trama de entrada de audio en el dominio QMF se combinan mediante un proceso de mezcla 828 que suma por bandas las dos señales. Téngase en cuenta que las bandas superiores mayores que K_{max} son cero en $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k}$ dado que la convolución sólo se lleva a cabo en las bandas hasta K_{max} .

20

[0108] La salida de la reverberación tardía se retrasa en una cantidad de intervalos de tiempo $d = ((L_{\text{trans}} - 20 \cdot 64 + 1) / 64 + 0,5) + 1$ en el proceso de mezcla.

[0109] El retardo d tiene en cuenta el tiempo de transición desde las reflexiones tempranas a las reflexiones tardías en las BRIR y un retardo inicial del reverberador de 20 intervalos de tiempo QMF, así como también un retardo de análisis de 0,5 intervalos de tiempo QMF para el análisis QMF de las BRIR con el fin de garantizar la inserción de la reverberación tardía en un intervalo de tiempo razonable. La señal combinada $\hat{\mathbf{z}}_{\text{ch}}^{n,k}$ en un intervalo de tiempo n se calcula por $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} + \hat{\mathbf{z}}_{\text{ch,rev}}^{n-d,k}$.

30 (e) *Síntesis QMF de señal en el dominio QMF binaural*

[0110] Una trama de 2 canales de 32 intervalos de tiempo de la señal de salida en el dominio QMF $\hat{\mathbf{z}}_{\text{ch}}^{n,k}$ se transforma en una trama de señal en el dominio del tiempo de 2 canales con longitud L por la síntesis QMF según ISO/IEC 14496-3:2009, subcláusula 4.6.18.4.2. para producir la señal de salida en el dominio del tiempo final L 830, $\tilde{\mathbf{z}}_{\text{ch}}^v = \begin{bmatrix} \tilde{z}_{\text{ch},1}^v \\ \dots \\ \tilde{z}_{\text{ch},2}^v \end{bmatrix}$. La reverberación tardía sintética o artificial se somete a un cambio de escala teniendo en cuenta las características de la señal de entrada, mejorando así la calidad de la señal de salida mientras se aprovecha de la reducción de la complejidad computacional obtenida por el tratamiento separado. Además, como se puede observar en la descripción anterior, no se requieren modelos adicionales de audición o sonoridad de reverberación objeto.

40

[0111] Aunque la realización anterior se ha descrito en combinación con el dominio QMF, se indica que también pueden usarse otros dominios de tiempo y frecuencia, por ejemplo, el dominio STFT. Además, el factor de cambio de escala puede calcularse de una manera dependiente de la frecuencia de modo que la correlación no se calcule en todo el número de bandas de frecuencia, es decir $\mathcal{N}[1, M]$, sino que se calcule en un número de S subconjuntos definidos de la siguiente manera:

45

$$i_1 \forall [1, N_1], i_2 \forall [N_1 + 1, N_2], \dots, i_S \forall [N_{S-1} + 1, N]$$

[0112] Además, la suavización puede aplicarse en todas las bandas de frecuencias o pueden combinarse 50 bandas según una regla específica, por ejemplo, según la resolución de la frecuencia de la audición. La suavización puede adaptarse a diferentes constantes de tiempo, por ejemplo, dependientes del tamaño de la trama o de la

preferencia del oyente.

[0113] El enfoque de la invención también puede aplicarse para diferentes tamaños de trama, incluso es posible un tamaño de trama de un solo intervalo de tiempo en el dominio del tiempo-frecuencia.

5

[0114] Según las realizaciones, pueden usarse diferentes matrices de submezcla para la submezcla, por ejemplo, matrices de submezcla simétricas o matrices asimétricas.

[0115] La medida de correlación puede deducirse de los parámetros que se transmiten en el tren de bits de audio, por ejemplo, de la coherencia entre canales en MPEG envolvente o SAOC. También, según las realizaciones, es posible excluir algunos valores de la matriz del cálculo del valor medio, por ejemplo valores calculados erróneamente o valores de la diagonal principal, los valores de autocorrelación, si es necesario.

[0116] El proceso puede realizarse en el codificador en vez de usarlo en el renderizador binaural en el lado del decodificador, por ejemplo, cuando se aplica un perfil binaural de baja complejidad. Esto da lugar a que algunas representaciones de los factores de cambio de escala, por ejemplo los propios factores de cambio de escala, con una medida de correlación entre 0 y 1 y similares, y estos parámetros, se transmitan en el tren de bits del codificador al decodificador para una matriz en sentido descendente fija.

[0117] Además, aunque la realización descrita anteriormente se describe por la aplicación de la ganancia después del reverberador 514, se indica que según otras realizaciones la ganancia también se puede aplicar antes del reverberador 514 o dentro del reverberador, por ejemplo mediante la modificación de las ganancias dentro del reverberador 514. Esto es ventajoso ya que puede requerirse un menor número de cálculos.

[0118] Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que estos aspectos también representan una descripción del procedimiento correspondiente, en el que un bloque o dispositivo corresponde a una etapa de un procedimiento o a una función de una etapa de un procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa de un procedimiento también representan una descripción de un bloque o elemento o función correspondiente de un aparato correspondiente. Algunas o todas las etapas del procedimiento pueden ejecutarse por (o mediante) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, una o más de las etapas del procedimiento más importantes pueden ser realizadas por un aparato de este tipo.

[0119] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede llevarse a cabo usando un medio de almacenamiento no transitorio tal como un medio de almacenamiento digital, por ejemplo un disquete, DVD, disco Blu-Ray, CD, ROM, PROM, EPROM y EEPROM o una memoria FLASH, que tiene señales de control legibles electrónicamente almacenadas, que cooperan (o son capaces de cooperar) con un sistema informático programable de manera que se lleve a cabo el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

[0120] Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de manera que se lleve a cabo uno de los procedimientos descritos en la presente memoria.

45

[0121] En general, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, estando el código de programa operativo para llevar a cabo uno de los procedimientos cuando el producto del programa informático se ejecuta en un ordenador. El código de programa puede, por ejemplo, almacenarse en un soporte legible en una máquina.

50

[0122] Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en la presente memoria, almacenado en un soporte legible en una máquina.

[0123] En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en la presente memoria, cuando el programa informático se ejecuta en un ordenador.

[0124] Una realización adicional del procedimiento de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los procedimientos descritos en la presente memoria. El soporte de datos, el medio de almacenamiento digital o el medio de grabado son normalmente tangibles y/o no transitorios.

[0125] Una realización adicional del procedimiento de la invención es, por lo tanto, un tren de datos o una secuencia de señales que representa el programa informático para realizar uno de los procedimientos descritos en la

presente memoria. El tren de datos o la secuencia de señales pueden, por ejemplo, configurarse para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

5 **[0126]** Una realización adicional comprende un medio de tratamiento, por ejemplo, un ordenador o un dispositivo lógico programable, configurados o programados para realizar uno de los procedimientos descritos en la presente memoria.

10 **[0127]** Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los procedimientos descritos en la presente memoria.

15 **[0128]** Una realización adicional según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, por medios electrónicos u ópticos) un programa informático para realizar uno de los procedimientos descritos en la presente memoria a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similares. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa informático al receptor.

20 **[0129]** En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo, una matriz de puertas programables en campo) para llevar a cabo todas o algunas de las funcionalidades de los procedimientos descritos en la presente memoria. En algunas realizaciones, una matriz de puertas programables en campo puede cooperar con un microprocesador para llevar a cabo uno de los procedimientos descritos en la presente memoria. En general, los procedimientos son realizados ventajosamente por cualquier aparato de hardware.

25 **[0130]** Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se comprende que las modificaciones y variaciones de las disposiciones y los detalles descritos en la presente serán evidentes para los expertos en la materia. Por lo tanto, se pretende que sólo esté limitada por el alcance de las reivindicaciones de la patente en tramitación y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en la presente memoria.

Bibliografía

30

[0131]

35 [1] M. R. Schroeder, "Digital Simulation of Sound Transmission in Reverberant Spaces", The Journal of the Acoustical Society of America, VoS. 47, pp. 424-431 (1970) and enhanced in JA. Moorer, "About This Reverberation Business", Computer Music Journal, Vol. 3, no. 2, pp. 13-28, MIT Press (1979).

[2] Uhle, Christian; Paulus, Jouni; Herre, Jürgen: "Predicting the Perceived Level of Late Reverberation Using Computational Models of Loudness" Proceedings, 17th International Conference on Digital Signal Processing (DSP), July 6 – 8, 2011, Corfu, Greece.

40 [3] Czyzewski, Andrzej: "A Method of Artificial Reverberation Quality Testing" J. Audio Eng. Soc., Vol. 38, No 3, 1990.

REIVINDICACIONES

1. Un procedimiento para el tratamiento de una señal de audio (504, 802) según una respuesta al impulso del ambiente (300), comprendiendo el procedimiento:

5 el tratamiento por separado (502, 514, 812, 816a, 816b) de la señal de audio (504, 802) con una parte temprana (301, 302) y una reverberación tardía (304) de la respuesta al impulso del ambiente (300), en el que el tratamiento de la reverberación tardía (304) comprende la generación de una señal reverberada con cambio de escala; y
 10 la combinación de la señal de audio tratada con la parte temprana de la respuesta al impulso del ambiente y la señal reverberada con cambio de escala, en el que la señal de audio (504, 802) comprende una pluralidad de canales de entrada, y en el que la generación de la señal reverberada con cambio de escala comprende la aplicación de un factor de ganancia a la señal de audio tratada con la reverberación tardía (304) de la respuesta al impulso del ambiente (300), estando el factor de ganancia determinado basándose en una medida de correlación predefinida o en una medida de correlación calculada de los canales de entrada;
 15 en el que la medida de correlación predefinida tiene un valor fijo de 0,1 a 0,9, y en el que el factor de ganancia se determina del modo siguiente:

$$20 \quad g = c_u + \rho \cdot (c_c - c_u)$$

en el que

25 ρ = medida de correlación predefinida o calculada para la señal de audio (504, 802),
 c_u , c_c = factores indicativos de una condición de la pluralidad de canales de entrada de la señal de audio (504, 802), con c_u refiriéndose a canales totalmente no correlacionados, y c_c refiriéndose a canales totalmente correlacionados,

30 en el que c_u y c_c se determinan del modo siguiente:

$$c_u = 10^{\frac{10 \cdot \log_{10}(K_{in})}{20}} = \sqrt{K_{in}}$$

$$c_c = 10^{\frac{20 \cdot \log_{10}(K_{in})}{20}} = K_{in}$$

35 en el que

K_{in} = número de canales de entrada activos de la señal de audio.

2. El procedimiento según la reivindicación 1, en el que el factor de ganancia es filtrado en paso bajo sobre una pluralidad de tramas de audio.

3. El procedimiento según la reivindicación 2, en el que el factor de ganancia es filtrado en paso bajo del modo siguiente:

$$45 \quad g_s(t_i) = c_{s,old} \cdot g_s(t_i - 1) + c_{s,new} \cdot g$$

$$c_{s,old} = e^{-\left(\frac{1}{f_s \cdot t_s}\right)}$$

$$c_{s,new} = 1 - c_{s,old}$$

50 en el que

55 t_s = constante de tiempo del filtro de paso bajo
 t_i = trama de audio en la trama t_i ,
 g_s = factor de ganancia suavizado
 k = tamaño de trama, y
 f_s = frecuencia de muestreo.

4. El procedimiento según una de las reivindicaciones 1 a 3, en el que un análisis de correlación de la señal de audio (504, 802) comprende la determinación para una trama de audio de la señal de audio (504, 802) de una medida de correlación combinada, y en el que la medida de correlación combinada se calcula mediante la combinación de coeficientes de correlación para una pluralidad de combinaciones de canales de una trama de audio, comprendiendo cada trama de audio uno o más intervalos de tiempo.
5. El procedimiento según la reivindicación 4, en el que la combinación de los coeficientes de correlación comprende el promediado de una pluralidad de coeficientes de correlación de la trama de audio.
- 10 6. El procedimiento según la reivindicación 4 o 5, en el que la determinación de la medida de correlación combinada comprende:
- (i) el cálculo de un valor medio global para cada canal de la una trama de audio,
 - (ii) el cálculo de una trama de audio de media cero restando los valores medios globales de los canales correspondientes,
 - (iii) el cálculo para una pluralidad de combinación de canales del coeficiente de correlación, y
 - (iv) el cálculo de la medida de correlación combinada como la media de una pluralidad de coeficientes de correlación.
- 20 7. El procedimiento según una de las reivindicaciones 1 a 6, en el que el coeficiente de correlación para una combinación de canales se calcula del modo siguiente:

$$\rho[m,n] = \left| \frac{1}{(N-1)} \cdot \frac{\sum_i \sum_j x_m[i,j] \cdot x_n[i,j]^*}{\sum_j \sigma(x_m[j]) \cdot \sigma(x_n[j])} \right|$$

25 en el que

- $\rho[m, n]$ = coeficiente de correlación,
- $\sigma(x_m[j])$ = desviación típica en un intervalo de tiempo j del canal m ,
- $\sigma(x_n[j])$ = desviación típica en un intervalo de tiempo j del canal n ,
- 30 x_m, x_n = variables de media cero,
- $i \in [1, N]$ = bandas de frecuencia,
- $j \in [1, M]$ = intervalos de tiempo,
- $m, n \in [1, K]$ = canales,
- * = conjugado complejo.

- 35 8. El procedimiento según una de las reivindicaciones 1 a 7, que comprende el retardo de la señal reverberada con cambio de escala para hacer coincidir su inicio en el punto de transición desde las reflexiones tempranas con la reverberación tardía (304) en la respuesta al impulso del ambiente (300).
- 40 9. El procedimiento según una de las reivindicaciones 1 a 8, en el que el tratamiento de la reverberación tardía comprende la aplicación de la señal de audio a un submezclador (410) para la submezcla de la señal de audio (504, 802) con una señal que tiene un número de canales inferior y la aplicación de la señal de audio submezclada a un reverberador.
- 45 10. Un producto informático no tangible que incluye un medio legible por ordenador que almacena instrucciones para hacer que un ordenador realice el procedimiento según una de las reivindicaciones 1 a 9 cuando es ejecutado por el ordenador.
- 50 11. Una unidad de tratamiento de señales, que comprende:
- una entrada para recibir una señal de audio (504, 802),
 - un procesador de parte temprana para el tratamiento de la señal de audio recibida (504, 802) según una parte temprana (301, 302) de una respuesta al impulso del ambiente (300),
 - un procesador de reverberación tardía para el tratamiento de la señal de audio recibida (504, 802) según una
 - 55 reverberación tardía de la respuesta al impulso del ambiente (300), con el procesador de reverberación tardía configurado para generar una señal reverberada con cambio de escala; y
 - una salida para la combinación de la parte temprana tratada de la señal de audio recibida (504, 802) y la señal reverberada con cambio de escala en una señal de audio de salida,
 - en el que la señal de audio (504, 802) comprende una pluralidad de canales de entrada, y
 - 60 en el que la señal reverberada con cambio de escala es generada por la aplicación de un factor de ganancia a la señal de audio tratada con la reverberación tardía (304) de la respuesta al impulso del ambiente (300), estando el

factor de ganancia determinado basándose en una medida de correlación predefinida o en una medida de correlación calculada de los canales de entrada;
 en el que la medida de correlación predefinida tiene un valor fijo de 0,1 a 0,9, y
 en el que el factor de ganancia se determina del modo siguiente:

5

$$g = c_u + \rho \cdot (c_c - c_u)$$

en el que

10

ρ = medida de correlación predefinida o calculada para la señal de audio (504, 802),
 c_u, c_c = factores indicativos de una condición de la pluralidad de canales de entrada de la señal de audio (504, 802), con
 c_u refiriéndose a canales totalmente no correlacionados, y c_c refiriéndose a canales totalmente correlacionados,

15

en el que c_u y c_c se determinan del modo siguiente:

$$c_u = 10^{\frac{10 \cdot \log_{10}(K_{in})}{20}} = \sqrt{K_{in}}$$

$$c_c = 10^{\frac{20 \cdot \log_{10}(K_{in})}{20}} = K_{in}$$

20

en el que

K_{in} = número de canales de entrada activos de la señal de audio.

25 12.

La unidad de tratamiento de señales según la reivindicación 11, en la que el procesador de reverberación tardía comprende:

un reverberador que recibe la señal de audio (504, 802) y que genera una señal reverberada; y
 una etapa de ganancia acoplada a una entrada o a una salida del reverberador y controlada por el factor de ganancia.

30

13. La unidad de tratamiento de señales según la reivindicación 11 o 12, que comprende un analizador de correlación que genera el factor de ganancia dependiendo de la señal de audio (504, 802).

35 14.

La unidad de tratamiento de señales según una de las reivindicaciones 11 a 13, que comprende además al menos uno de entre:

un filtro de paso bajo acoplado a la etapa de ganancia, y
 un elemento de retardo acoplado entre la etapa de ganancia y un sumador, con el sumador acoplado además al procesador de parte temprana y a la salida.

40

15. Un renderizador binaural, que comprende una unidad de tratamiento de señales según una de las reivindicaciones 11 a 14.

45 16.

Un codificador de audio para codificar señales de audio, que comprende: una unidad de tratamiento de señales según una de las reivindicaciones 11 a 14 o un renderizador binaural según la reivindicación 15 para el tratamiento de las señales de audio antes de la codificación.

50 17.

Un decodificador de audio para decodificar señales de audio codificadas, que comprende: una unidad de tratamiento de señales según una de las reivindicaciones 11 a 14 o un renderizador binaural según la reivindicación 15 para el tratamiento de las señales de audio decodificadas.

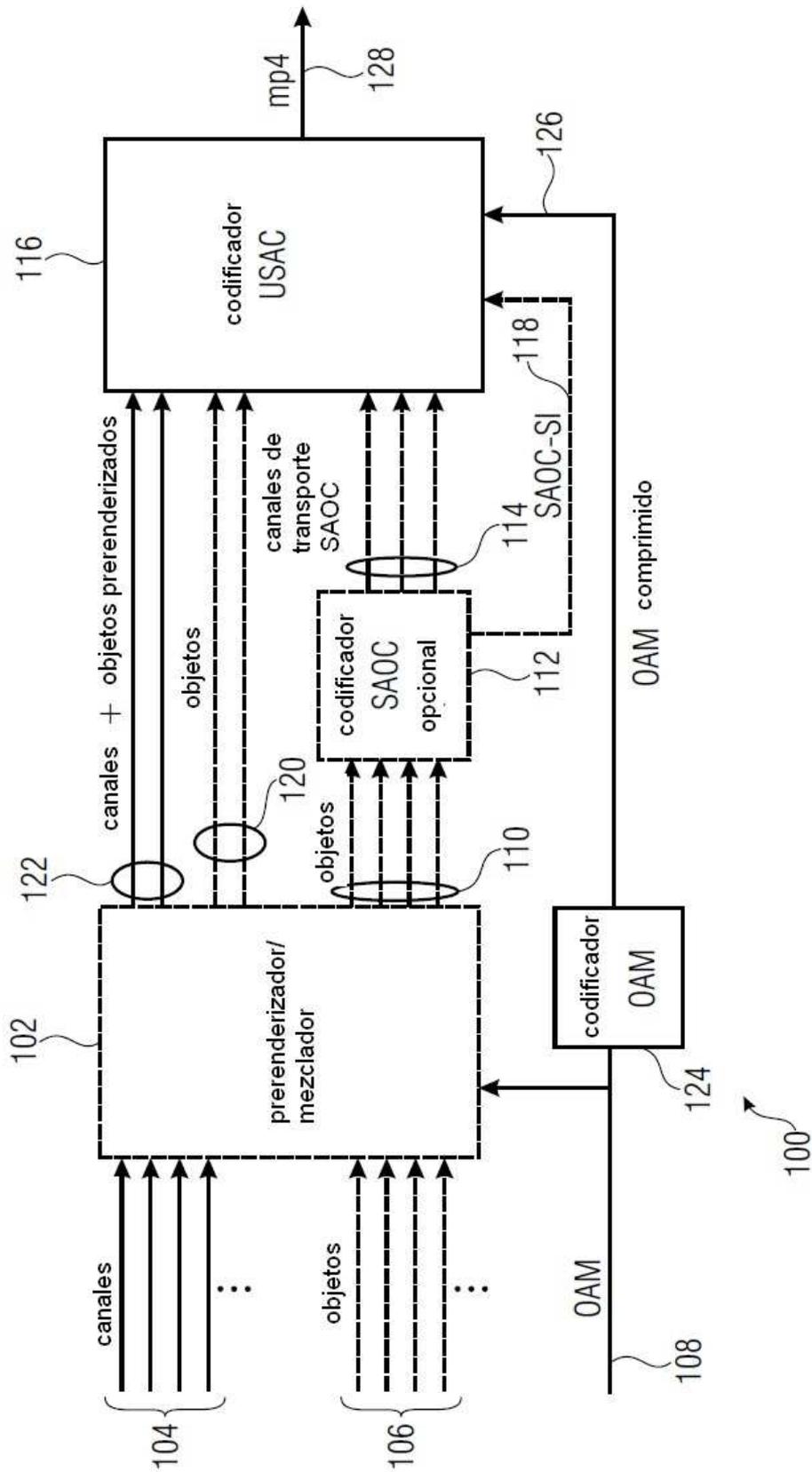


FIGURA 1

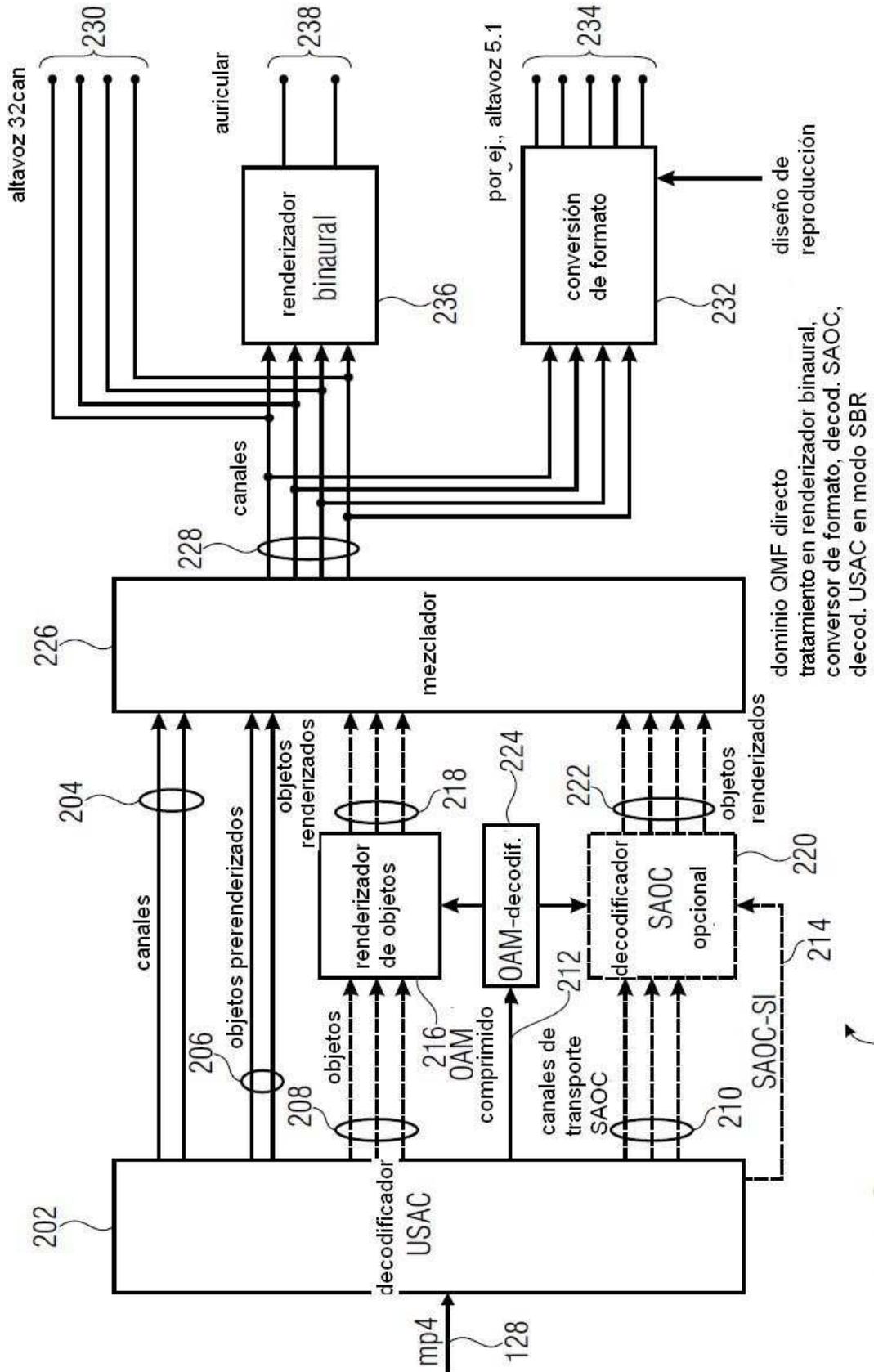


FIGURA 2 200

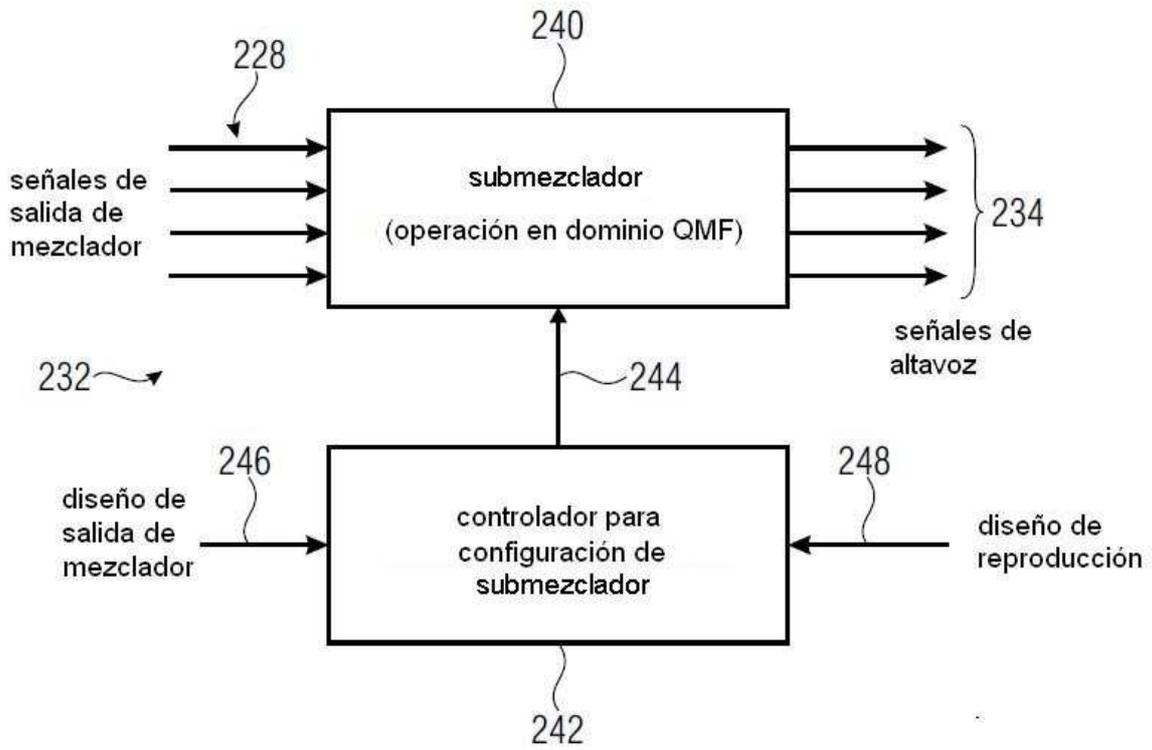


FIGURA 3

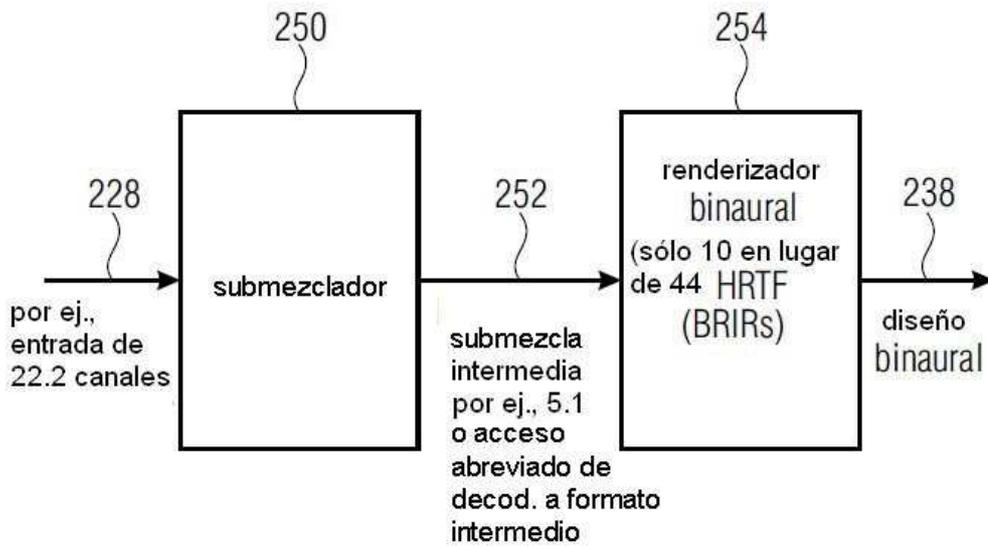


FIGURA 4

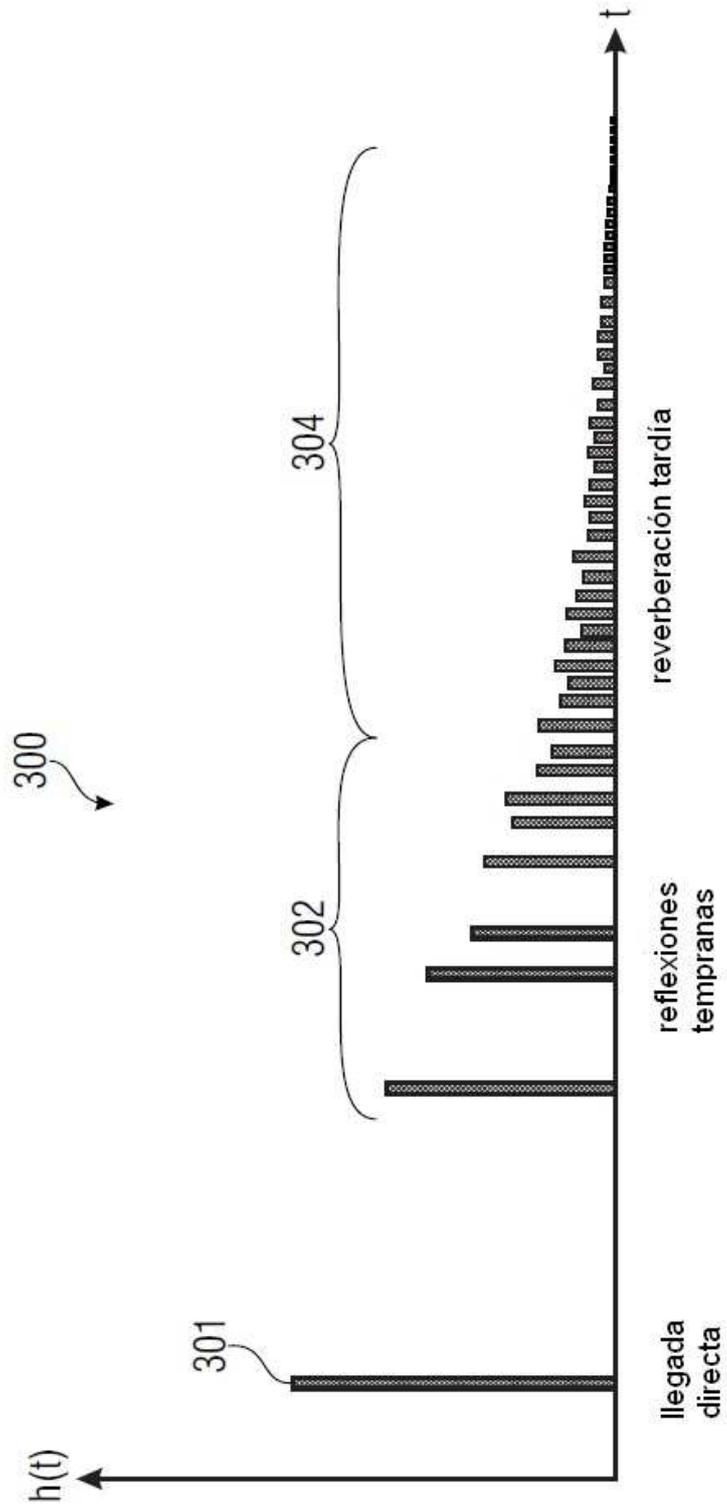


FIGURA 5

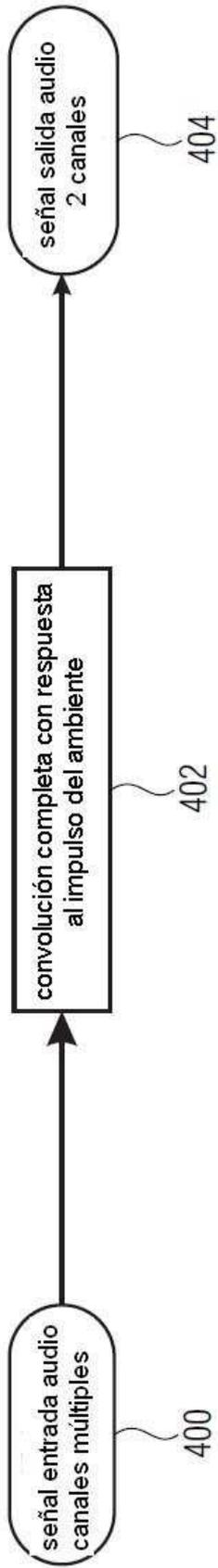


FIGURA 6A

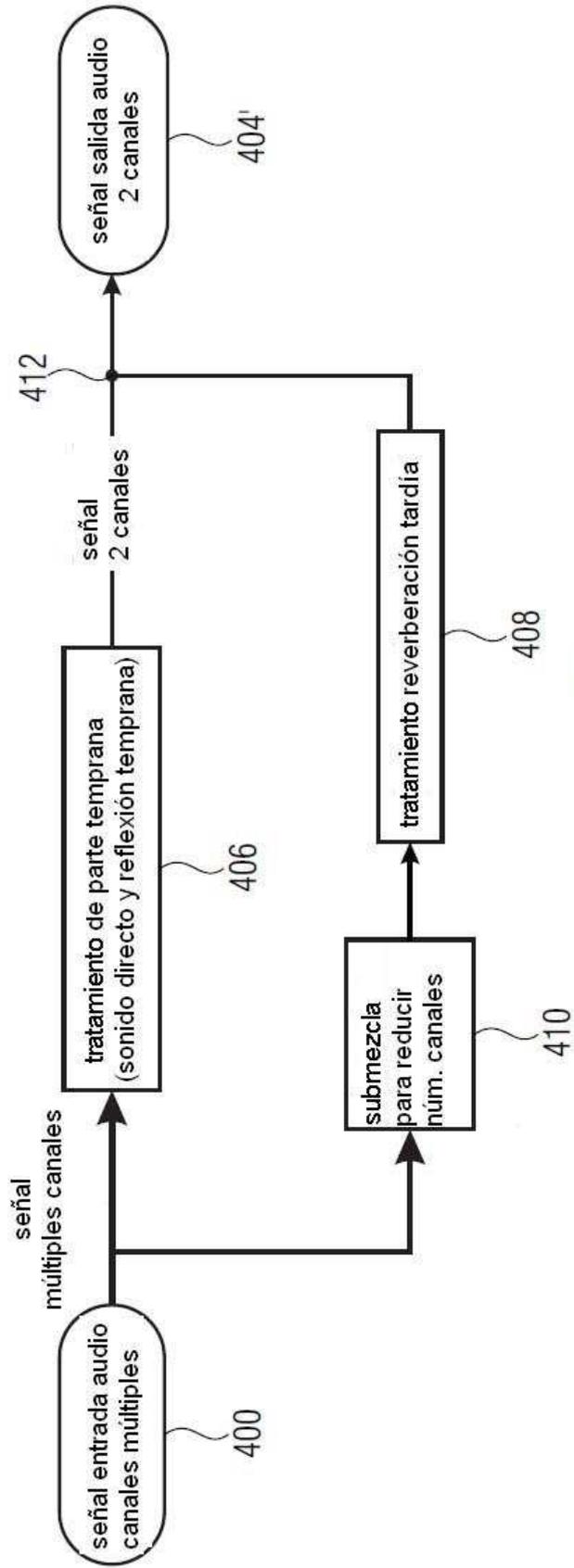


FIGURA 6B

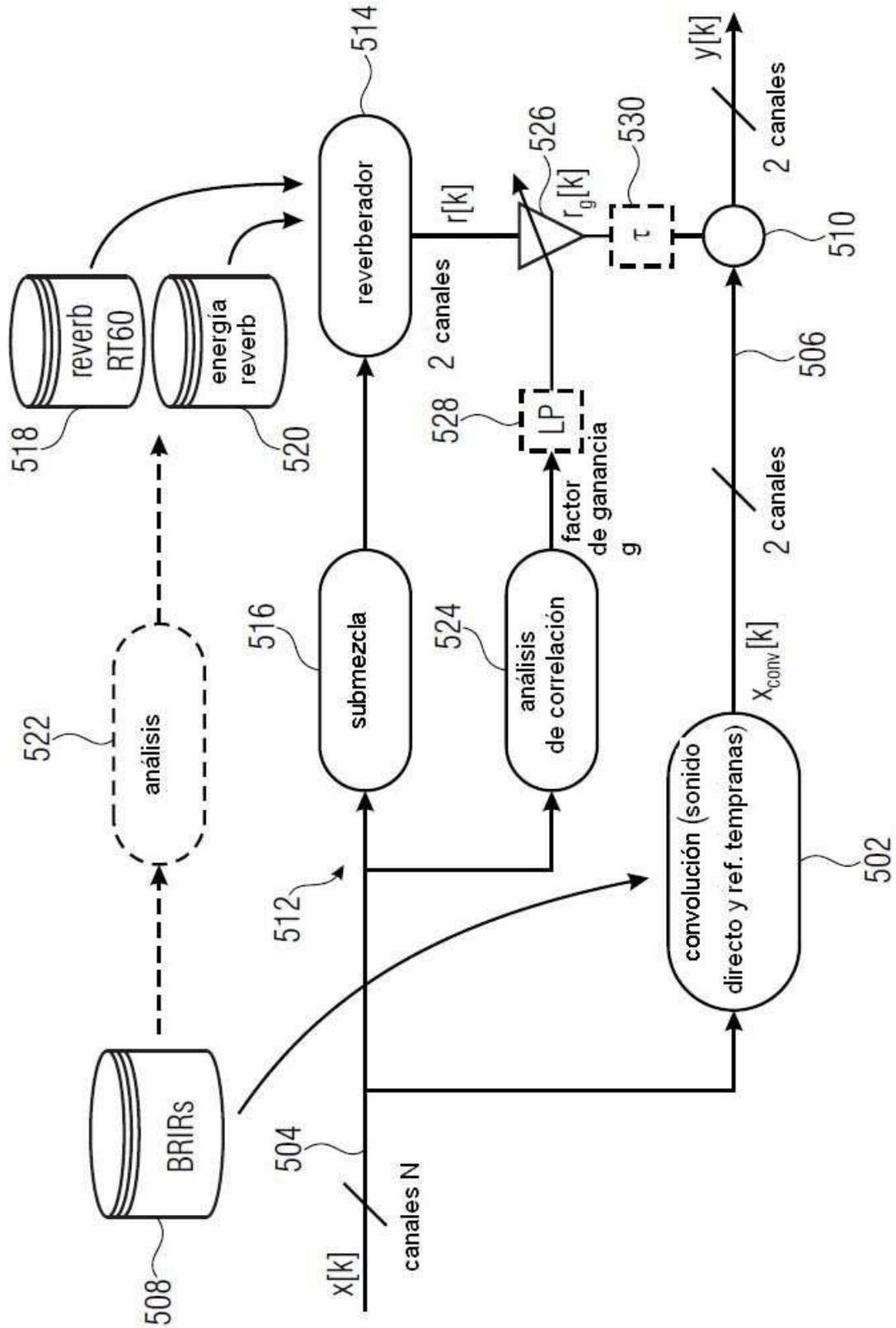


FIGURA 7

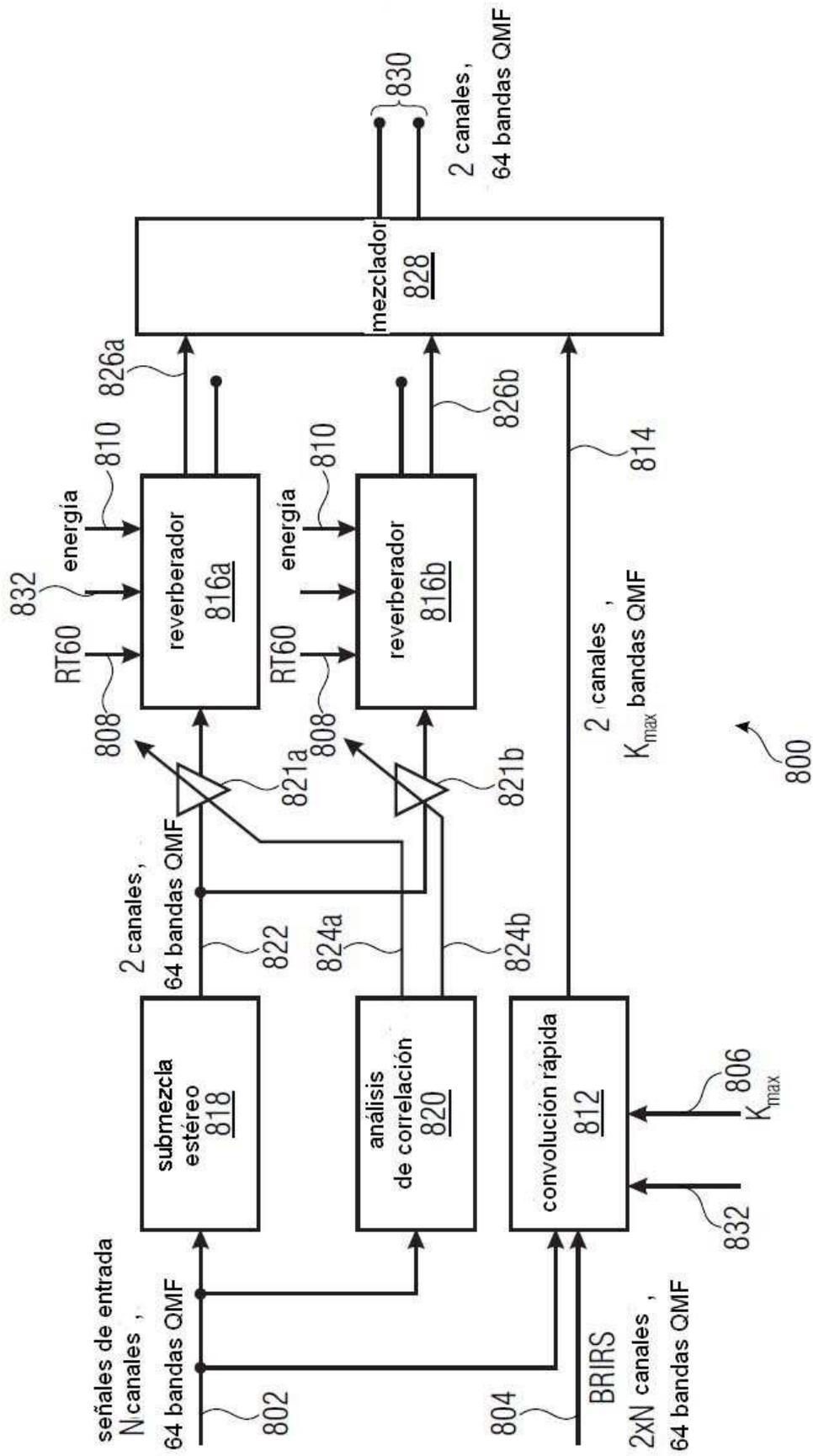


FIGURA 8

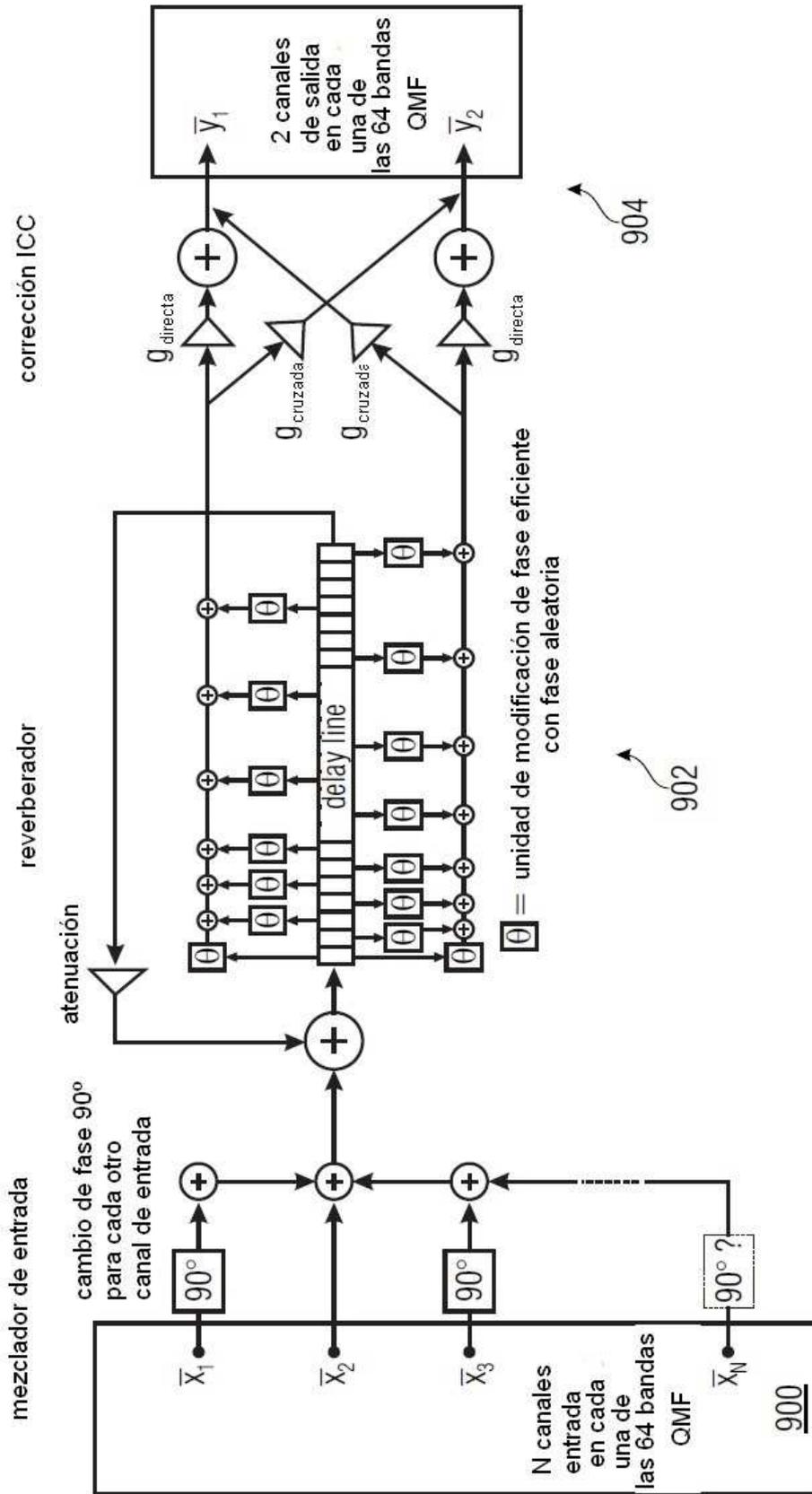


FIGURA 9