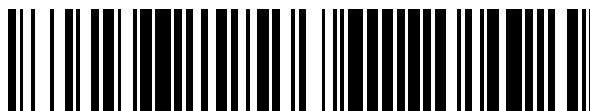


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 758 757**

51 Int. Cl.:

H04S 3/00 (2006.01)

H04S 7/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.07.2014 PCT/EP2014/065227**

87 Fecha y número de publicación internacional: **29.01.2015 WO15010983**

96 Fecha de presentación y número de la solicitud europea: **16.07.2014 E 14747856 (4)**

97 Fecha y número de publicación de la concesión europea: **02.10.2019 EP 3025327**

54 Título: **Método para el procesamiento de una señal de audio de acuerdo con una respuesta al impulso de la habitación, una unidad de procesamiento de señales, un codificador de audio, un decodificador de audio y un renderizador binaural**

30 Prioridad:

22.07.2013 EP 13177362

18.10.2013 EP 13189230

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

06.05.2020

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:

ARIZTI ACHA, Monica

ES 2 758 757 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

Método para el procesamiento de una señal de audio de acuerdo con una respuesta al impulso de la habitación, una unidad de procesamiento de señales, un codificador de audio, un decodificador de audio y un renderizador binaural

5

La presente invención se refiere al campo de la codificación/decodificación de audio, en especial a la codificación de audio espacial y a la codificación de objetos de audio espacial, por ejemplo, el campo de los sistemas de codecs de audio 3D. Las realizaciones de la invención se refieren a enfoques para el procesamiento de una señal de audio de acuerdo con una respuesta al impulso de la habitación y para la determinación en tal respuesta al impulso de la habitación de una transición de las reflexiones tempranas de reverberación tardía.

10

Las herramientas de codificación de audio espacial son muy conocidas en la técnica y están estandarizadas, por ejemplo, en el estándar de MPEG-envolvente. La codificación de audio espacial comienza a partir de una pluralidad de entrada original, 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 mejora de bajas frecuencias. Un codificador de audio espacial puede derivar uno o más canales de mezclado descendente de los canales originales y, en forma adicional, puede derivar datos paramétricos en relación a las señales espaciales tales como diferencias de nivel entre canales en los valores de coherencia de canal, diferencias de fase entre canales, diferencias de tiempo entre canales, etc. El uno o más canales de mezclado descendente 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 mezclado descendente 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 de 7.1, etc.

15

20

25

Además, las herramientas de codificación de objetos de audio espacial son muy conocidas en la técnica y están estandarizadas, por ejemplo, en el estándar MPEG SAOC (SAOC = codificación de objetos de audio espacial). A diferencia de la codificación de audio espacial a partir de los canales originales, la codificación de objetos de audio espacial comienza a partir de objetos de audio que no se dedican de forma automática durante una cierta instalación de reproducción renderizada. Más bien, la colocación de los objetos de audio en la escena de reproducción es flexible y puede ser configurada por un usuario, por ejemplo, por medio de la introducción de cierta información de representación en un decodificador de codificación de objetos de audio espacial. En forma alternativa o en forma adicional, la información de representación puede transmitirse como información lateral adicional o metadatos; la prestación de información puede incluir información en cuya posición en la configuración de la reproducción de un cierto objeto de audio se ha de colocar (por ejemplo, a lo largo del tiempo). Con el fin de obtener una cierta compresión de datos, un número de objetos de audio se codifica por medio de un codificador SAOC que calcula, a partir de los objetos de entrada, uno o más canales de transporte por medio del mezclado descendente de la reproducción de los objetos de acuerdo con cierta información de mezclado descendente. Además, el codificador SAOC calcula la información lateral paramétrica que representa las señales inter-objetos tales como las diferencias de nivel de objeto (OLD), los valores de coherencia de objeto, etc. Al igual que en SAC (SAC = codificación de audio espacial), los datos entre objetos paramétricos se calculan para tiempos individuales/mosaicos de frecuencia. Para un cierto marco (por ejemplo, 1024 o 2048 muestras) de la señal de audio de una pluralidad de bandas de frecuencia (por ejemplo, 24, 32, o 64 bandas) se consideran de manera tal que se proporcionan datos paramétricos para cada marco y cada banda de frecuencia. Por ejemplo, cuando una pieza de audio tiene 20 marcos y cuando cada marco está subdividido en 32 bandas de frecuencia, el número de mosaicos de tiempo/frecuencia es 640.

30

35

40

45

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 una habitación específica. En tal situación, se proporciona una respuesta al impulso de la habitación de la habitación específica, por ejemplo sobre la base de una medición de los mismos, y se utiliza para el procesamiento de la señal de audio al presentarla a un oyente. Se puede desear el procesamiento del sonido directo y las reflexiones en tal presentación separada de la reverberación tardía. Para ello es necesario determinar dónde terminan las reflexiones tempranas y dónde comienza la reverberación tardía.

50

El documento US 2008/175396/A1 describe un aparato para y un método de externalización de una salida de imagen sonido a auriculares. El método incluye localizar la imagen sonido de una señal de entrada en una zona predeterminada ante un oyente, y procesar una señal a un componente de señal izquierdo y un componente de señal derecho de la señal de entrada con valores de retardo y valores de ganancia diferentes, respectivamente.

55

El documento US 6,188,769 B1 describe un método y un aparato para procesar fuentes de sonido para estimular efectos ambientales. Los bloques de canal fuente incluyen bloques de reflexión directa, temprana y de reverberación tardía para el acondicionamiento de la alimentación fuente para incluir retardos, cambios espectrales y atenuaciones dependiendo de la posición, la orientación y la direccionalidad de las fuentes de sonido, la posición y orientación del oyente y la posición y la transmisión de sonido y propiedades de reflexión de obstáculos y paredes en un ambiente modelado. Las salidas de los bloques de canal fuente están combinadas y provistas para reverberación única de bloques que genera ambas las reflexiones tempranas y la reverberación tardía para todas las fuentes de sonido.

60

65

5 El documento WO 99/49574 A1 describe un método de procesar una serie de señales de audio de entrada que
representa una serie de fuentes de sonido de audio virtual situadas en posiciones predeterminadas alrededor de un
oyente para producir un conjunto de señales de salida de audio reducidas para reproducir en dispositivos de altavoz
situados alrededor del oyente. El método incluye (a) para cada señal de audio de entrada y para cada señal de salida de
audio: (i) convolucionar las señales de audio de entrada con una porción de cabezal inicial de una respuesta al impulso
correspondiente mapeando sustancialmente el sonido inicial y las reflexiones tempranas para una respuesta de impulso
de una fuente de audio virtual correspondiente a un dispositivo de altavoz correspondiente de modo que se forma una
serie de respuestas iniciales, (b) para cada señal de audio de entrada y para señal de salida de audio: (i) formar una
mezcla combinada a partir de las señales de entrada de audio; (ii) formar una cola de convolución combinada de las
respuestas de impulso correspondiente, (iii) convolucionar la mezcla combinada con la cola de convolución combinada
para formar una respuesta de cola combinada; y (c) para cada señal de salida de audio: (i) combinar una serie
correspondiente de respuestas iniciales y una respuesta de cola combinada correspondiente para formar la señal de
salida de audio.

15 El documento EP 2 028 884 A1 describe un método para modelar una respuesta al impulso de la habitación que incluye
recibir una señal de presión de sonido que se obtiene por un micrófono cuando una fuente de sonido de tipo impulso se
excita y detectar una respuesta al impulso de la habitación; determinar límites entre una pluralidad de intervalos de la
respuesta al impulso de la habitación de modo que la respuesta al impulso de la habitación se divide en la pluralidad de
intervalos en un dominio de tiempo; y modelar la respuesta al impulso de la habitación utilizando un método de
modelado diferente para cada pluralidad de intervalos divididos.

20 Un objetivo de la presente invención es proporcionar un enfoque mejorado para el procesamiento de una señal de audio
de acuerdo con una respuesta al impulso de la habitación.

25 Este objetivo se consigue por medio de un método de acuerdo con la reivindicación 1, una unidad de procesamiento de
señales de acuerdo con la reivindicación 12, un codificador de audio de acuerdo con la reivindicación 14, un
decodificador de audio de acuerdo con la reivindicación 16, y un renderizador binaural de acuerdo con la reivindicación
20.

30 La presente invención está basada en los hallazgos de los inventores de que en los enfoques convencionales existe el
problema de que hay situaciones en que la determinación de la transición de las reflexiones tempranas y la
reverberación tardía es demasiado pronto porque una correlación utilizado para juzgar la aparición de la transición ya
alcanza un umbral antes de que la primera reflexión incluso se produjo o incidía. Sobre la base de estos hallazgos y
dado que se sabe que el momento de transición debe ser mayor que el tiempo de llegada de la primera reflexión, debido
a que la primera reflexión es claramente distinta y puede asegurarse que no es la reverberación tardía difusa, los
inventores hallaron que es necesario evitar el uso de un umbral fijo, más bien, de acuerdo con el enfoque de la
invención, el umbral se define de manera tal que es dependiente de la correlación en el momento de incisión de una de
las reflexiones tempranas. Esto asegura que la primera reflexión siempre se encuentra antes de que el momento de
transición.

35 (1) La presente invención proporciona un método como se define en la reivindicación 1. El enfoque de la invención es
ventajoso ya que permite un procesamiento mejorado de la señal de audio sobre la base de un punto de transición
robusto. El enfoque de la invención es independiente de la habitación, de si se utiliza o no un enfoque binaural y del
ángulo de incidencia. En comparación con los enfoques de la técnica anterior, el enfoque de la invención es más
ventajoso porque no es fuertemente dependiente del ángulo azimutal de una respuesta al impulso binaural y la relación
entre las amplitudes de sonido directo y la primera reflexión de incisión.

45 (2) De acuerdo con las realizaciones la medida de correlación describe con respecto a la respuesta al impulso de la
habitación de la similitud de la decadencia en la energía acústica que incluye el estado inicial y de la decadencia
partiendo en energía acústica en cualquier momento después del estado inicial en un intervalo de frecuencia
predefinida.

50 (3) De acuerdo con las realizaciones la determinación de la transición comprende la determinación de una distribución
de energía acústica con base en la respuesta al impulso de la habitación, y determinar una pluralidad de medidas de
correlación que indican para una pluralidad de porciones de la distribución determinada una correlación entre la energía
acústica en la porción respectiva de la distribución determinada y la energía acústica en un estado inicial.

55 (4) De acuerdo con las realizaciones la determinación de la distribución comprende la determinación de una distribución
de tiempo-frecuencia de la energía acústica, en la que una porción de la distribución comprende un bloque de tiempo de
una longitud predefinida, el estado inicial está definido por el primero de la pluralidad de bloques de tiempo de la
distribución de tiempo-frecuencia. Esto es ventajoso ya que permite analizar la distribución de frecuencias en diferentes
momentos, lo que de ese modo proporciona una representación de las características de la distribución de tiempo-
frecuencia.

65

ES 2 758 757 T3

(5) De acuerdo con las realizaciones la determinación de la distribución comprende el cálculo del alivio del decaimiento de energía (*EDR*) de la respuesta al impulso de la habitación, en el que el *EDR* se calcula de acuerdo con lo presentado a continuación:

$$E(t, \omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

donde

- 5
- $E(t, \omega)$ = alivio del decaimiento de energía,
 - $h(\tau)$ = respuesta al impulso de la habitación,
 - ω = $2\pi F$.

10 (6) De acuerdo con las realizaciones, la respuesta al impulso de la habitación tiene una longitud efectiva predefinida, y en el que la determinación de la distribución de tiempo-frecuencia comprende el cálculo del espectro de FFT de la respuesta al impulso de la habitación por el uso de una ventana que tiene una longitud correspondiente a la longitud efectiva de la respuesta al impulso de la habitación. Esto es ventajoso, dado que la FFT/DFT está bien definida y hay algoritmos eficaces para calcular los valores espectrales. Si se conocen los valores en la ventana, la FFT/DFT se puede calcular de una manera no complicada.

15

(7) De acuerdo con las realizaciones, la energía acústica en el estado inicial se determina al tomar la longitud efectiva completa de la respuesta al impulso de la habitación, al calcular el espectro de FFT y al tomar el cuadrado de los valores absolutos, y la energía acústica de un bloque de tiempo se determina por medio del desplazamiento de la ventana por el tiempo asociado con el bloque de tiempo, el relleno con ceros de las muestras de ventana a la longitud efectiva, el cálculo del espectro de FFT y al tomar el cuadrado de los valores absolutos. Esto es una ventaja, ya que no se requiere ningún banco de filtros adicionales o similares para el cálculo de banda estrecha del *EDR*; sólo se requiere un desplazamiento de la ventana.

20

(8) De acuerdo con las realizaciones, la medida de correlación es una medida de correlación que describe la similitud de la decadencia en la energía acústica que incluye el estado inicial y la decadencia en la energía acústica partiendo en cualquier momento después el estado inicial. La medida de correlación se puede calcular de acuerdo con lo presentado a continuación:

25

$$\rho(t) = \frac{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))^2}}$$

donde

- 30
- $\rho(t)$ = medida de correlación,
 - $E(1, \omega)$ = alivio del decaimiento de energía de intervalo en la frecuencia f ,
 - $\bar{E}(1, \omega)$ = valor medio a lo largo de todas las frecuencias del alivio del decaimiento de energía inicial,
 - $E(t, \omega)$ = alivio del decaimiento de energía en la frecuencia f a partir de un tiempo t ,
 - 35 $\bar{E}(t, \omega)$ = valor medio a lo largo de todas las frecuencias del alivio del decaimiento de energía de intervalo completo a partir del tiempo t ,
 - ω = $2\pi F$.

Esto es ventajoso, dado que la fórmula se refiere al coeficiente de correlación de Pearson muy conocido (correlación de producto-momento de Pearson). El coeficiente de correlación se puede calcular directamente a partir del *EDR*.

40

(9) De acuerdo con las realizaciones, el umbral se determina en base a un valor constante y la medida de correlación para la seleccionada de las reflexiones tempranas. El umbral se puede definir de acuerdo con lo presentado a continuación:

45

$$\rho(t) = c \cdot \rho(t_F)$$

donde

- 45
- $\rho(t_F)$ = medida de correlación para la seleccionada de las reflexiones tempranas,
 - t_F = índice de tiempo en el que el seleccionada de las reflexiones tempranas después de la incisión del sonido directo,
 - 50 c = el valor constante que está basado en $\frac{1}{e}$, siendo e el número de Euler.

Esto es ventajoso, dado que el umbral no es constante, sino que depende de la reflexión temprana seleccionada para asegurar que la correlación no cae demasiado pronto por debajo del umbral.

55

(10) De acuerdo con las realizaciones, el momento de la seleccionada de las reflexiones tempranas se determina, por ejemplo, por medio de un operador de curtosis en funcionamiento, por una detección de umbral o por una detección de ataques. Esto es ventajoso, dado que el momento de incisión de la reflexión se puede calcular directamente y de forma automática a partir de las muestras de dominio de tiempo de la respuesta al impulso.

(11) De acuerdo con las realizaciones, la seleccionada de las reflexiones tempranas es la primera reflexión tras el sonido directo.

(12) La presente invención proporciona una unidad de procesamiento de señal como se define en la reivindicación 12.

(13) La presente invención proporciona un codificador de audio para codificar una señal de audio como se define en la reivindicación 14.

(14) La presente invención proporciona un decodificador de audio para decodificar una señal de audio codificada como se define en la reivindicación 16.

(15) La presente invención proporciona un renderizador binaural que comprende la unidad de procesamiento de señal de la invención.

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

La figura 1 ilustra una visión general de un codificador de audio 3D de un sistema de audio 3D;

la figura 2 ilustra una visión general de un decodificador de audio 3D de un sistema de audio 3D;

la figura 3 ilustra un ejemplo para implementar un conversor de formatos que se puede implementar en el decodificador de audio 3D de la Fig. 2;

la figura 4 ilustra una realización de un renderizador binaural que se puede implementar en el decodificador de audio 3D de la Fig. 2;

la figura 5 ilustra un ejemplo de una respuesta al impulso de la habitación $h(t)$;

la figura 6 (A) ilustra un diagrama de bloques de una unidad de procesamiento de señales (por ejemplo, en un renderizador binaural de la Fig. 4) para el procesamiento por separado de una señal de audio con una parte temprana y una reverberación tardía del impulso de la habitación de acuerdo con una realización de la invención;

la figura 6 (B) ilustra un diagrama de bloques de otra unidad de procesamiento de señales (por ejemplo, en un renderizador binaural de la Fig. 4) para el procesamiento por separado de una señal de audio con una parte temprana y una reverberación tardía del impulso de la habitación de acuerdo con una realización adicional de la invención;

la figura 7 ilustra un diagrama de flujo de un método para determinar un momento de transición entre las reflexiones tempranas y la reverberación tardía en una respuesta al impulso de la habitación, de acuerdo con una realización de la invención;

la figura 8 ilustra el alivio del decaimiento de energía (*EDR*) para lograr una respuesta al impulso determinada de acuerdo con un enfoque basado en FFT;

la figura 9 ilustra la determinación del momento de transición de acuerdo con una realización de la invención;

la figura 10 ilustra los tiempos de transición para un canal izquierdo y un canal derecho para una respuesta al impulso de la habitación binaural medida determinada por el uso de un método convencional;

la figura 11 ilustra los tiempos de transición para un canal izquierdo y un canal derecho para una respuesta al impulso de la habitación binaural medida determinada por el uso del método de la invención;

la figura 12 ilustra en forma esquemática el procesamiento binaural de señales de audio en un renderizador binaural de acuerdo con una realización de la presente invención; y

la figura 13 ilustra en forma esquemática el procesamiento en el reverberador del dominio de frecuencia del renderizador binaural de la Fig. 12 de acuerdo con una realización de la presente invención.

A continuación se describirán realizaciones del enfoque inventivo para el procesamiento de una señal de audio de acuerdo con una respuesta al impulso de la habitación y para la determinación en una respuesta al impulso de la habitación de una transición de las reflexiones tempranas de reverberación tardía. 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 el enfoque inventivo.

Las figuras 1 y 2 muestran los bloques algorítmicos de un sistema de audio 3D de acuerdo con las realizaciones. En forma más específica, la Fig. 1 muestra una visión general de un codificador 100 de audio 3D. El codificador 100 de

audio recibe en un circuito 102 pre-renderizador/mezclador, que puede estar provisto en forma opcional, por señales de entrada, en forma más específica una pluralidad de canales de entrada que proporciona al codificador 100 de audio una pluralidad de señales 104 de canales, una pluralidad de señales 106 de objetos y metadatos 108 de objetos correspondientes. Las señales de objetos 106 procesadas por el pre-renderizador/mezclador 102 (véase las señales 110) se pueden proporcionar a un codificador 112 SAOC (SAOC = Codificación de Objetos de Audio Espacial). El codificador 112 SAOC genera los canales 114 de transporte SAOC proporcionados a un 116 codificador USAC (USAC = Codificación Unificada de Voz y Audio). Además, la señal 118 de SAOC-SI (SAOC-SI = información lateral SAOC) también se proporciona para el codificador 116 USAC. El codificador 116 USAC además recibe señales 120 de objetos directamente desde el pre-renderizador/mezclador, así como las señales de canales y señales 122 objeto de pre-renderizadas. La información de metadatos 108 de objetos se aplica a un codificador 124 OAM (OAM = metadatos de objeto) que proporciona la información de metadatos de objeto comprimida 126 del codificador USAC. El codificador 116 USAC, sobre la base de las señales de entrada mencionadas con anterioridad, genera una señal de salida MP4 comprimido, de acuerdo con lo mostrado en 128.

La Fig. 2 muestra una visión general de un decodificador 200 de audio 3D del sistema de audio 3D. La señal 128 (mp4) codificada generada por el codificador 100 de audio de la Fig. 1 es recibida en el decodificador 200 de audio, en forma más específica en un codificador 202 USAC. El codificador 202 USAC decodifica la señal recibida 128 en las señales 204 de canales, las señales 206 de objetos pre-renderizadas, las señales 208 de objetos, y las señales 210 de canales de transporte SAOC. Además, la información de metadatos de objetos 212 comprimidos y la señal 214 SAOC-SI se emite por medio del codificador 202 USAC. Las señales 208 de objetos se proporcionan a un renderizador 216 de objetos que da salida a las señales de objetos 218 renderizados. Las señales de canales 210 de transporte SAOC se suministran al decodificador 220 SAOC que da salida a las señales de objetos 222 renderizados. La información de metadatos de objetos 212 comprimidos se suministra al decodificador 224 OAM que da salida a las señales de control respectivas para el renderizador 216 de objetos y el decodificador 220 SAOC para generar las señales de objetos 218 renderizados y las señales de objetos 222 renderizados. El decodificador comprende además un mezclador 226 que recibe, de acuerdo con lo mostrado en la Fig. 2, las señales 204, 206, 218 y 222 de entrada para dar salida a las señales 228 de canales. Las señales de canales pueden ser directamente la salida a un altavoz, por ejemplo, un canal 32 de altavoz, de acuerdo con lo indicado en 230. Las señales 228 se pueden proporcionar a un circuito 232 de conversión de formato que recibe como una entrada de control una señal de la disposición de reproducción que indica la forma en que se convertirán las señales 228 de canales. En la realización representada en la Fig. 2, se supone que la conversión se ha de hacer de una manera tal que las señales se pueden proporcionar a un sistema de altavoces 5.1 de acuerdo con lo indicado en 234. Además, las señales 228 de canales se pueden proporcionar a un renderizador 236 binaural la generación de dos señales de salida, por ejemplo, para un auricular, de acuerdo con lo indicado en 238.

En una realización de la presente invención, el sistema de codificación/decodificación se representa en las Figs. 1 y 2 está basado en el códec MPEG-D USAC para la codificación de señales de canales y objetos (véanse las señales 104 y 106). Para aumentar la eficiencia de codificación de una gran cantidad de objetos, se puede utilizar la tecnología MPEG SAOC. Tres tipos de renderizadores pueden llevar a cabo las tareas de representación de objetos a los canales, renderizando los canales a los auriculares o renderizando los canales a una configuración de altavoces diferente (véase la Fig. 2, los signos de referencia 230, 234 y 238). Cuando las señales de objetos se transmiten de manera explícita o se codifican en forma paramétrica por el uso de SAOC, la información de metadatos de objeto 108 correspondiente se comprime (véase la señal 126) y se multiplexa en el flujo de bits 128 de audio 3D.

Los bloques del algoritmo del sistema de audio 3D general que se muestran en las Figs. 1 y 2 se describirán en detalle adicional a continuación.

El pre-renderizador/mezclador 102 se puede proporcionar en forma opcional para convertir una escena de entrada de canal más objetos antes de la codificación. Funcionalmente, es idéntico al renderizador/mezclador de objetos que se describirá más adelante. Puede ser deseable la pre-renderización de objetos para asegurar una entropía de señal determinista en la entrada del codificador que es, básicamente, independiente del número de señales de objetos simultáneamente activas. Con la pre-renderización de los objetos, no se requiere ningún objeto de transmisión de metadatos. Las señales de objetos discretos se representan en el diseño del canal que el codificador está configurado para utilizar. Los pesos de los objetos para cada canal se obtuvieron a partir de los metadatos de objeto asociado (OAM).

El codificador 116 USAC es el códec de núcleo para las señales de altavoz de los canales, señales de objetos discretos, señales de mezclado descendente de objetos y señales pre-renderizadas. Está basado en la tecnología MPEG-D USAC. Se ocupa de la codificación de las señales anteriores por medio de la creación de información de canal y mapeo de objetos basado en la información geométrica y semántica del canal de entrada y la asignación de objetos. Esta información de asignación se describe cómo los canales de entrada y los objetos se asignan a USA-elementos de canal, como elementos de par de canales (EPP), elementos individuales de canal (SCE), efectos de baja frecuencia (LFE) y elementos de canal cuádruple (QCE) y CPE, CPE y LFE, y la información correspondiente se transmite al decodificador. Todas las cargas adicionales como datos 114, 118 SAOC o un objeto 126 de metadatos se consideran en el control de la frecuencia del codificador. La codificación de objetos es posible en diferentes formas, dependiendo de los requisitos

de tasa/distorsión y los requisitos de interactividad para el procesador. De acuerdo con las realizaciones, son posibles las siguientes variantes de codificación de objetos:

- 5 • Objetos pre-renderizados: Las señales de objetos son pre-renderizados y se mezclan con las señales de canales de 22.2 antes de la codificación. La cadena de codificación posterior ve 22.2 señales de canal.
- 10 • Las formas de onda de objetos discretos: Los objetos se suministran como formas de onda monofónicas al codificador. El codificador utiliza los elementos de canal individuales (CPE) para transmitir los objetos además de las señales de canal. Los objetos decodificados se representan y se mezclan en el lado del receptor. La información de metadatos de objeto comprimido se transmite al receptor/renderizador.
- 15 • Las formas de onda de objetos paramétricos: Las propiedades de los objetos y su relación entre sí, se describen por medio de parámetros SAOC. El mezclado descendente de las señales de objetos se codifica con la USAC. La información paramétrica se transmite a lo largo. Se elige el número de canales de mezclado descendente en función del número de objetos y la velocidad de datos global. La información de metadatos de objeto comprimido se transmite al renderizador SAOC.

20 El codificador 112 SAOC y el decodificador 220 SAOC para señales de objetos pueden estar basados en la tecnología MPEG SAOC. El sistema es capaz de volver a crear, modificar y renderizar un número de objetos de audio basado en un número menor de canales de transmisión y datos paramétricos adicionales, tales como OLD, IOC (Coherencia entre los Objetos), DMGS (Ganancias de Mezclado Descendente). Los datos paramétricos adicionales exhiben una velocidad de datos significativamente menor que la requerida para la transmisión de todos los objetos individualmente, lo que hace la codificación muy eficiente. El codificador 112 SAOC toma como entrada las señales de objetos/canal en forma de ondas monofónicas y emite la información paramétrica (que se empaqueta en el flujo de bits 128 de audio 3D) y los canales de transporte SAOC (que son codificados por el uso de elementos de un solo canal y se transmiten). El decodificador 220 SAOC reconstruye las señales de objeto/de canal de los canales 210 de transporte SAOC descodificados y la información 214 paramétrica, y genera la escena de audio de salida con base en el diseño de la reproducción, la información de metadatos de objetos descomprimidos y, en forma opcional, sobre la base de la información de la interacción del usuario.

30 El códec de metadatos de objeto (véase el codificador 124 OAM y el decodificador 224 OAM) se proporciona de manera tal que, para cada objeto, los metadatos asociados que especifican la posición geométrica y el volumen de los objetos en el espacio 3D se codifica de manera eficiente por la cuantificación de las propiedades del objeto en tiempo y espacio. El COAM 126 de metadatos de objeto comprimido se transmite al receptor 200 como información lateral.

35 El renderizador 216 de objetos utiliza los metadatos de objeto comprimido para generar formas de onda de objetos de acuerdo con el formato de reproducción dado. Cada objeto se representa a un canal de salida determinada en función de sus metadatos. La salida de este bloque resulta de la suma de los resultados parciales. Si tanto el contenido basado en canal, así como se decodifican objetos discretos/paramétricas, las formas de onda basadas canal y las formas de onda de los objetos renderizados son mezclados por el mezclador 226 antes de la salida de las formas 228 de onda de resultantes o antes de alimentar a un módulo post-procesador como el renderizador 236 binaural o el módulo renderizador 232 de altavoz.

45 El módulo 236 de renderizador binaural produce un mezclado descendente binaural del material de audio multicanal de manera tal que cada canal de entrada está representada por una fuente de sonido virtual. El procesamiento se realiza marco a marco en el dominio MGC (Banco de Filtro en Espejo de Cuadratura) y la binauralización está basada en las respuestas al impulso de la habitación binaurales medidas.

50 El renderizador 232 de altavoz convierte entre la configuración del canal 228 de transmisión y el formato de reproducción deseado. También puede ser llamado "convertor de formatos". El convertor de formatos realiza conversiones para reducir el número de canales de salida, es decir, crea mezclas descendentes.

55 La Fig. 3 muestra un ejemplo para implementar un convertor 232 de formatos. El convertor 232 de formatos, también conocido como renderizador de altavoz, convierte entre la configuración del canal transmisor y el formato de la reproducción deseada. El convertor 232 de formatos realiza conversiones a un menor número de canales de salida, es decir, se realiza un proceso de mezclado 240 descendente (DMX). El elemento de mezclado 240 descendente, que opera preferiblemente en el dominio de QMF, recibe las señales de salida 228 del mezclador y da salida a las señales 234 del altavoz. Se puede proporcionar un configurador 242, también denominado como controlador, que recibe, como una entrada de control, una señal 246 indicativa de la distribución de salidas del mezclador, es decir, la disposición para que los datos representados por la señal de salida 228 del mezclador se determina, y la señal 248 indicativa de la disposición de reproducción deseado. Con base en esta información, el controlador 242, preferiblemente en forma automática, genera matrices de mezclado descendente optimizados para la combinación dada de formatos de entrada y de salida y aplica estas matrices para el elemento de mezclado 240 descendente. El convertor 232 de formatos permite configuraciones de altavoces estándar, así como para configuraciones aleatorias con posiciones de los altavoces que no son estándar.

La Fig. 4 ilustra una realización del renderizador 236 binaural de la Fig. 2. El módulo de renderizador binaural puede proporcionar un mezclado descendente binaural del material de audio multicanal. La binauralización puede estar basada en una respuesta al impulso de la habitación binaural medida. La respuesta al impulso de la habitación puede ser considerada como una "huella digital" de las propiedades acústicas de una habitación real. La respuesta al impulso de la habitación se mide y se almacena, y se pueden proporcionar señales acústicas arbitrarias con esta "huella digital", permitiendo de ese modo la escucha de una simulación de las propiedades acústicas de la habitación asociadas con la respuesta al impulso de la habitación. El renderizador 236 binaural puede ser programado o configurado para la prestación de los canales de salida en dos canales binaurales por el uso de funciones de transferencia relacionadas con los cabezales o las respuestas al impulso de la habitación binaurales (BRIR). Por ejemplo, para los dispositivos móviles se desea una renderización binaural para auriculares o altavoces conectados a tales dispositivos móviles. En tales dispositivos móviles, debido a las limitaciones, puede ser necesario limitar el decodificador y la complejidad de representación. Además de la omisión de descorrelación en tales escenarios de procesamiento, puede ser preferible llevar a cabo primero un mezclado descendente por el uso de un elemento de mezclado 250 descendente a una señal de mezclado 252 descendente intermedia, es decir, a un menor número de canales de salida que se traduce en un menor número de canal de entrada para el conversor 254 binaural real. Por ejemplo, un material de 22.2 canales se puede mezclar en forma descendente por el elemento de mezclado 250 descendente a un mezclado descendente intermedio 5.1 o, en forma alternativa, el mezclado descendente intermedio puede calcularse directamente por el decodificador 220 SAOC en la Fig. 2 en una especie de modo de "acceso directo". La renderización binaural entonces sólo tiene que aplicar diez HRTF (Funciones de Transferencia de Cabezera) o funciones BRIR para la prestación de los cinco canales individuales en diferentes posiciones en contraste con la aplicación de 44 funciones HRTF o BRIR si los 22.2 canales de entrada debían ser renderizados directamente. Las operaciones de convolución necesarias para la renderización binaural requieren una gran cantidad de potencia de procesamiento y, por lo tanto, reducen esta potencia de procesamiento, si bien la obtención de una calidad de audio aceptable todavía es particularmente útil para dispositivos móviles. El renderizador 236 binaural produce un mezclado 238 descendente binaural del material de audio 228 multicanal, de manera tal que cada canal de entrada (con exclusión de los canales LFE) está representado por una fuente de sonido virtual. La transformación puede llevarse a cabo marco a marco a marco de dominio QMF. La binauralización está basada en las respuestas al impulso de la habitación binaurales medidas, y el sonido directo y las reflexiones puede ser impresas en el material de audio a través de un enfoque de convolución en un dominio pseudo-FFT por el uso de una convolución rápida en la parte superior del dominio QMF, mientras que la reverberación tardía se puede procesar por separado.

La Fig. 5 muestra un ejemplo de una respuesta 300 al impulso de la habitación $h(t)$. La respuesta al impulso de la habitación comprende tres componentes, el sonido 301 directo, las reflexiones 302 tempranas y la reverberación 304 tardía. Por lo tanto, la respuesta al impulso de la habitación describe el comportamiento de reflexión de un espacio acústico cerrado reverberante cuando se reproduce un impulso. Las reflexiones 302 tempranas son reflejos discretos 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 llama reverberación 304 tardía. El sonido 301 directo puede ser fácilmente identificado en la respuesta al impulso de la habitación y se puede separar de las reflexiones tempranas, sin embargo, la transición de reflexión 302 temprana a la reverberación 304 tardía es menos evidente.

En las siguientes realizaciones del enfoque de la invención se describirá con más detalle. De acuerdo con las realizaciones de la invención, una señal de audio se procesa por separado con una parte temprana y una reverberación tardía de una respuesta al impulso de la habitación. La señal de audio procesada con la parte temprana de la respuesta al impulso de la habitación y la señal reverberada se combinan y se emite como la señal de audio de salida. Para la tramitación separada de la transición en la respuesta al impulso de la habitación desde la parte temprana a la reverberación tardía necesita ser conocida. La transición está determinada por una medida de correlación que alcanza un umbral, en el que el umbral se establece dependiendo de la medida de correlación para una seleccionada de las reflexiones tempranas en la parte temprana de la respuesta al impulso de la habitación. La medida de correlación se puede describir con respecto a la respuesta al impulso de la habitación de la similitud de la decadencia en la energía acústica que incluye el estado inicial y la decadencia en la energía acústica partiendo en cualquier momento después del estado inicial en un intervalo de frecuencia predefinida.

De acuerdo con las realizaciones, el procesamiento independiente de la señal de audio comprende el procesamiento de la señal de audio con la reflexión 301, 302 parte temprana de la respuesta al impulso de la habitación durante un primer proceso, y procesar la señal de audio con la reverberación 304 difusa de la respuesta al impulso de la habitación 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. De acuerdo con las realizaciones adicionales, en el segundo proceso de reverberación 304 difusa (tardía) puede ser sustituido por una reverberación sintética. En este caso, la respuesta al impulso de la habitación siempre puede contener sólo la parte temprana de reflexión 301, 302 (véase la Fig. 5) y la reverberación 304 tardía difusa no está incluida.

La Fig. 6(A) muestra un diagrama de bloques que ilustra una primera unidad de procesamiento de señal representativa para el procesamiento por separado de una señal de audio con una parte temprana y una reverberación tardía del impulso de la habitación de acuerdo con una realización de la invención. El procesamiento de la señal de audio de

acuerdo con las diferentes partes de la respuesta al impulso de la habitación puede llevarse a cabo en un renderizador 236 binaural que se ha descrito con anterioridad. La señal 400 de entrada de audio puede ser un material de audio no reverberante, por ejemplo, una señal de entrada de audio multicanal, que se convoluciona con la respuesta al impulso de la habitación, por ejemplo, una respuesta al impulso de la habitación medida por el uso de una cabeza artificial o micrófonos en la oreja. Esta convolución permite obtener una impresión espacial del material de audio no reverberante original como si el material de audio se escuchara en la habitación asociada con respuesta al impulso de la habitación. Por ejemplo, en el renderizador 236 binaural mencionado con anterioridad, puede ser deseable el procesamiento de la señal de audio con el sonido 301 directo y la reflexión 302 temprana en la respuesta al impulso de la habitación y el procesamiento de la señal de audio con la reverberación 304 tardía por separado. Para el procesamiento de la señal 400 de audio de entrada, se proporcionan un bloque 402 para el procesamiento de sonido directo, un bloque 404 para el procesamiento de las reflexiones tempranas y un bloque 406 para el procesamiento de reverberación tardía. Las señales 408 y 410 de salida de los respectivos bloques 402 a 406 se combinan por medio de un primer sumador 412 para generar una señal 414 temprana procesada. La señal 414 temprana procesada y la 416 señal reverberada proporcionada por el procesador 406 se combinan por un segundo sumador 418 para la generación de la señal de salida de audio 420 que proporciona a un oyente la impresión como si la señal de audio fuera escuchada en la habitación asociada con las respuestas al impulso de la habitación.

El procesamiento de la reverberación 302 tardía separado del sonido directo y las reflexiones tempranas es ventajoso debido a la complejidad computacional reducida. En forma más específica, el uso de una convolución para toda la respuesta al impulso es computacionalmente muy costoso. Por lo tanto, los algoritmos de reverberación con menor complejidad normalmente se utilizan para el procesamiento de señales de audio con el fin de simular la reverberación tardía. El sonido y las reflexiones tempranas que son parte directa de la respuesta al impulso se calcula más exactamente, por ejemplo por medio de una convolución. Una ventaja adicional es la posibilidad de control de la reverberación. Esto permite que la reverberación tardía sea modificada dependiente de, por ejemplo, una entrada de usuario, un parámetro ambiente medido o dependiente de los contenidos de la señal de audio. Para lograr las ventajas por encima de la transición (por ejemplo, el punto en el tiempo), donde las reflexiones 302 tempranas finalizan y donde la reverberación 304 tardía arranca debe ser conocido. Cuando el procesamiento de reverberación tardía comienza demasiado pronto, la señal de audio puede ser de menor calidad dado que el oído humano puede detectar las reflexiones tempranas distintas faltantes. Por otro lado, si se detecta el momento de transición demasiado tarde, no se explota la eficiencia computacional, como el procesamiento de las reflexiones tempranas en forma típica es más costoso que el procesamiento de reverberación tardía. La transición, por ejemplo, en muestras de dominio de tiempo, se puede alimentar al renderizador binaural como un parámetro de entrada que luego, dependiendo de la transición recibida, controla los procesadores 402 a 406 para el procesamiento por separado de la señal de audio.

La Fig. 6 (B) ilustra un diagrama de bloques de otra unidad de procesamiento de señales representativa para el procesamiento por separado de una señal de audio con una parte temprana y una reverberación tardía del impulso de la habitación de acuerdo con otra realización de la invención. La señal 400 de entrada, por ejemplo, una señal de entrada de audio multicanal, se recibe y se aplica a un primer procesador 422 para el procesamiento de la parte temprana, es decir, para el procesamiento de la señal de audio de acuerdo con el sonido 301 directo y las reflexiones 302 tempranas en la respuesta 300 al impulso de la habitación mostrado en la Fig. 5. La señal 400 de entrada de audio multicanal también se aplica a un segundo procesador 424 para el procesamiento de la señal de audio de acuerdo con la reverberación 304 tardía de la respuesta al impulso de la habitación. En un renderizador binaural, de acuerdo con lo mencionado con anterioridad, se puede desear el procesamiento del sonido directo y las reflexiones tempranas separado de la reverberación tardía, principalmente a causa de la complejidad computacional reducida. El procesamiento del sonido directo y las reflexiones tempranas puede, por ejemplo, imprimir a la señal de audio un enfoque convolucional llevado a cabo por el primer procesador 422, mientras que la reverberación tardía puede ser sustituida por una reverberación sintética proporcionada por el segundo procesador 424. La señal de salida 420 global binaural es entonces una combinación del resultado 428 convolucional proporcionado por el procesador 422 y la señal 430 sintética reverberada proporcionada por el procesador 424. De acuerdo con las realizaciones, las señales 428 y 430 son combinadas por un sumador 432 que da salida a la señal 420 de salida binaural global.

De acuerdo con lo mencionado, el primer procesador 422 puede provocar una convolución de la señal 400 de entrada de audio con un sonido directo y las reflexiones tempranas de la respuesta al impulso de la habitación que se puede proporcionar al primer procesador 422 de una base 434 de datos externa sosteniendo una pluralidad de respuestas al impulso de la habitación binaural registradas. El segundo procesador o reverberador 424 puede operar sobre la base de parámetros del reverberador, como la reverberación RT60 y la energía de reverberación, que se puede obtener a partir de las respuestas al impulso de la habitación binaural almacenados por un análisis 436. Se observa que el análisis 436 no es necesariamente parte del procesador, en cambio esto es para indicar que a partir de las respuestas respectivas almacenadas en la base de datos se pueden derivar los parámetros 434 de reverberación respectivos; esto se puede hacer en forma externa. Los parámetros del reverberador se pueden determinar, por ejemplo, por medio del cálculo de la energía y el tiempo de reverberación RT60 en una octava o un tercio de octava de filtros de análisis, o pueden ser valores medios de los resultados de los análisis de respuesta al impulso múltiple.

Además, ambos procesadores 422 y 424 también reciben de la base 434 de datos, directamente o a través del análisis 436, como parámetro de entrada, información acerca de la transición en la respuesta al impulso de la habitación desde

la parte temprana a la reverberación tardía. La transición puede ser determinada de una manera de acuerdo con lo descrito con más detalle a continuación.

5 De acuerdo con las realizaciones, el análisis de transición se puede utilizar para separar las reflexiones tempranas y la reverberación tardía. Se puede alimentar al renderizador binaural como parámetro de entrada (por ejemplo, se puede leer de un archivo/interfaz dedicada junto con valores RT60 y valores de energía que se utilizan para configurar el reverberador). El análisis puede estar basado en un conjunto de respuestas al impulso de la habitación binaurales (un conjunto de pares BRIR para una multitud de ángulos de azimut y de elevación). El análisis puede ser un paso de procesamiento previo que se lleva a cabo por separado para cada respuesta al impulso y luego el promedio de todos los valores de transición se toma como un valor global de transición del conjunto BRIR. Este valor global de transición se puede utilizar entonces para separar las reflexiones tempranas de la reverberación tardía en el cálculo de la señal de salida binaural.

15 Varios enfoques para la determinación de la transición son conocidos, sin embargo, estos enfoques tienen desventajas de acuerdo con lo descrito a continuación. En referencia de la técnica anterior [1] se describe un método que utiliza el alivio del decaimiento de energía (*EDR*) y una medida de correlación para determinar el momento de transición desde las reflexiones tempranas de reverberación tardía. Sin embargo, el enfoque descrito en la referencia de la técnica anterior [1] es desventajoso.

20 1. El enfoque es fuertemente dependiente del ángulo azimutal de la respuesta al impulso binaural y la relación entre las amplitudes de sonido directo y la primera reflexión que incide.

2. El momento de transición se calcula en bandas de frecuencias arbitrarias. No hay conocimiento general acerca de cuál de las bandas de frecuencia da el momento de transición correcto a utilizarse para la respuesta al impulso global.

25 3. No hay información sobre el paso de correlación esencial del enfoque.

Otro enfoque conocido es describir las reflexiones tempranas por la dispersión de los ecos en un espacio, por ejemplo, por el número promedio de reflexiones por segundo, y para determinar el comienzo de la reverberación tardía cuando este número supera un umbral predefinido (véase la referencia de la técnica anterior [2]). Este enfoque está basado en la característica de la habitación, es decir, el volumen de la habitación, que a menudo es desconocido. El volumen de la habitación no puede ser fácilmente extraído de una respuesta al impulso medida. Por lo tanto, este método no es aplicable para el cálculo de la transición de respuestas al impulso medidas. También, no hay conocimiento común cómo densa las reflexiones tienen que ser para ser denominado reverberación tardía.

35 Otra posibilidad, descrita en la referencia de la técnica anterior [3], es comparar la distribución real en un momento en una ventana de respuesta al impulso de una distribución gaussiana en el dominio del tiempo. Se supone que la reverberación tardía tiene una distribución normal. En una distribución normal, aproximadamente un tercio (exactamente $1/e$) de las muestras se encuentran fuera de una desviación estándar de la media y dos tercios de las muestras están dentro de una desviación estándar de la media. Las reflexiones tempranas distintas tienen más muestras dentro de una desviación estándar y menos fuera. La proporción de muestras fuera de una desviación estándar en comparación con las muestras dentro de una desviación estándar se puede utilizar para definir el momento de transición. Sin embargo, la desventaja de este enfoque es que la transición es difícil de definir con esta medida, debido a que la relación a veces fluctúa en torno al umbral. La medida también depende en gran medida del tamaño y del tipo de la ventana deslizante en el que se calcula la proporción.

45 Además de los enfoques mencionados con anterioridad, también la curtosis (la más alta orden acumulante de una señal estocástica) se puede utilizar para determinar el momento de transición. Ésta disminuye rápidamente al acercarse hacia la última parte de la respuesta al impulso, de acuerdo con lo descrito en la referencia de la técnica anterior [4]. Sin embargo, la definición del umbral para la transición (ya sea el uso de una rápida disminución o el momento en el que primero llega a cero) no está clara.

55 Hay todavía otro enfoque que no está basado en el análisis de una respuesta al impulso medida, pero en el volumen de la habitación, de acuerdo con lo descrito en [2]. Este enfoque asume que el momento de transición sólo depende del volumen, pero no tiene en cuenta las propiedades de difusión o los límites. Por lo tanto, el resultado sólo puede ser una aproximación del momento de transición y no es tan preciso como sea necesario para evitar las desventajas mencionadas con anterioridad cuando no determinan con precisión el momento de transición. Además, el volumen de una habitación a menudo no se conoce y no puede ser fácilmente extraído de una respuesta al impulso medida.

60 Otros enfoques conocidos ignoran completamente el medio ambiente y definen el momento de transición a ser simplemente 80ms, véase por ejemplo en la referencia de la técnica anterior [5]. Este número, sin embargo, está totalmente separado de las características de la habitación o una respuesta al impulso medida y, por lo tanto, es demasiado inexacto para el propósito de separar la reverberación tardía del recordatorio de la respuesta al impulso.

La presente invención, de acuerdo con las realizaciones, ofrece, además de la mejora de procesamiento de señales de audio también un enfoque mejorado para determinar el momento de transición entre las reflexiones y la reverberación tardía en una respuesta al impulso de la habitación dando una determinación más precisa del momento de transición. Las realizaciones, de acuerdo con lo descrito más adelante, proporcionan una posibilidad simple y eficaz para calcular el momento de transición de una respuesta al impulso medida por el uso de un análisis de FFT.

La Fig. 7 muestra un diagrama de flujo de un método para determinar un momento de transición entre las reflexiones tempranas y la reverberación tardía en una respuesta al impulso de la habitación, de acuerdo con una realización de la invención. Para determinar el momento de transición de las reflexiones tempranas a la reverberación tardía, en un primer paso 500 se determina una distribución de tiempo-frecuencia de la energía acústica. Por ejemplo, de acuerdo con las realizaciones el alivio del decaimiento de energía ($E(t, f)$, EDR) se puede calcular en el paso 500. El EDR se puede calcular directamente a partir de una (por ejemplo, binaural) respuesta al impulso medida y se puede ser interpretado como una expansión de dominio de frecuencia de la curva de caída de energía comúnmente utilizado (integración Schroeder, EDC (d)) que muestra la energía restante en la respuesta al impulso después de un tiempo t . En lugar de utilizar la respuesta al impulso de banda ancha, el EDR se deriva de una representación tiempo-frecuencia y se pueden utilizar muchas representaciones de tiempo-frecuencia diferentes para este propósito. Una vez que la distribución de tiempo-frecuencia de la energía acústica se ha determinado en el paso 500, en el paso 502 se determina una medida de correlación entre la energía acústica en un bloque de tiempo de la distribución de tiempo-frecuencia y la energía acústica en general en un estado inicial. En el paso 504 se determina si la medida de correlación alcanza un umbral definido (por ejemplo, cae por debajo del umbral definido) o no. Si no alcanza el umbral, el método pasa al paso 506 donde se selecciona el siguiente bloque de tiempo y la distribución siguiente el bloque de tiempo actual y los pasos 502 y 504 se repiten para el siguiente bloque de tiempo. Así, de acuerdo con los pasos 502 a 506, se utiliza una medida de correlación para calcular el valor de correlación entre cada bloque de tiempo del EDR determinado en el paso 500 con la energía total en el estado inicial. Se ha alcanzado el momento de transición cuando la medida de correlación alcanza el umbral definido (por ejemplo, cae por debajo del umbral definido). En otras palabras, cuando se determina en el paso 504 que un bloque de tiempo actual de la medida de correlación es inferior al umbral, el método pasa al paso 508 donde el tiempo del bloque de tiempo actual es la salida como el momento de transición.

En lo que sigue, se describirá con más detalle una realización del enfoque de la invención. Inicialmente, una respuesta al impulso binaural medido se puede tomar como una entrada para el cálculo del momento de transición. Entonces, se emplea una distribución Page o Levin para el cálculo del alivio del decaimiento de energía (EDR). La distribución de Page hace referencia a la derivada del último intervalo de funcionamiento y la distribución de Page de la señal en tiempo inverso se llama la distribución Levin (véase también la referencia de la técnica anterior [2]). Esta distribución describe un espectro de potencia instantánea, y el EDR de la respuesta al impulso $h(t)$ (véase, por ejemplo, la Fig. 5.) se calcula de acuerdo con lo presentado a continuación:

$$E(t, \omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

donde
 $E(t, \omega)$ = alivio del decaimiento de energía,
 $h(\tau)$ = respuesta al impulso de la habitación,
 ω = $2\pi F$.

El cálculo de acuerdo con la ecuación anterior se inicia en el sonido 301 directo (véase la Fig. 5), y al aumentar el tiempo el alivio del decaimiento de energía contiene reflexiones menos distintas y una reverberación más estocástica. De acuerdo con la realización descrita, el alivio del decaimiento de energía se calcula para bloques de tiempo que tienen una longitud de 1 ms para facilidad de cálculo. Por medio de la funcionalidad descrita con anterioridad, la distribución de tiempo-frecuencia de la energía acústica se determina de acuerdo con lo descrito con respecto al paso 500 en la Fig. 7.

Después de esto, de acuerdo con lo descrito con respecto a los pasos 502 a 506 en la Fig. 7, se determina la medida de correlación $\rho(T)$ que está basada en la correlación de producto-momento de Pearson (también conocido como coeficiente de correlación). En forma más específica, se determina la correlación de la energía acústica para cada bloque de tiempo con la energía total en el estado inicial, de acuerdo con las realizaciones, de acuerdo con lo presentado a continuación:

$$\rho(t) = \frac{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))^2}}$$

donde

$E(1, \omega)$ = alivio del decaimiento de energía en la frecuencia f ,
 $\bar{E}(1, \omega)$ = valor medio a lo largo de todas las frecuencias del alivio del decaimiento de energía inicial,
 $E(t, \omega)$ = alivio del decaimiento de energía en la frecuencia f a partir de un tiempo t ,
 $\bar{E}(t, \omega)$ = valor medio a lo largo de todas las frecuencias del alivio del decaimiento de energía de intervalo completo a partir del tiempo t ,

$$\omega = 2\pi f.$$

La correlación anterior describe la similitud de la decadencia que incluye el estado inicial y el decaimiento partiendo en cualquier tiempo t . Se calcula a partir del *EDR* de banda ancha, por el uso del intervalo de frecuencias del *EDR* para el cálculo, lo que de ese modo compara la situación energética inicial completa con la situación en el tiempo t .

La presente invención no se limita al cálculo de la correlación a lo largo de todas las frecuencias. Más bien, la correlación también se puede calcular en un intervalo de frecuencia predefinida. El intervalo de frecuencia se puede determinar a partir de la señal de audio a ser procesada. Por ejemplo, para señales de audio específicas, el intervalo de frecuencias puede estar limitado a un intervalo predefinido, por ejemplo, el intervalo de frecuencias audibles. De acuerdo con las realizaciones, el intervalo de frecuencia puede ser de 20 Hz a 20 kHz. Se hace notar que también se pueden seleccionar otros intervalos, por ejemplo, por medio de estudios empíricos.

De acuerdo con una realización, se puede utilizar una implementación eficaz basada en FFT del *EDR*. Se aplica una ventana que tiene una longitud efectiva de la respuesta al impulso medida, y se supone que una respuesta al impulso medida tiene una longitud efectiva de 2^{13} que es igual a 8192 contenedores de frecuencia. Durante el cálculo, esta ventana se desplaza por la longitud discreta de un solo bloque de tiempo, y el final de la ventana es rellenado con ceros. De acuerdo con las realizaciones, se utiliza una longitud de bloque de tiempo de 1 ms, y por un cálculo simple y eficaz del *EDR* se aplica el siguiente enfoque:

(1) Se toma la longitud entera eficaz de la respuesta al impulso medida para calcular el espectro de FFT, y los valores absolutos se elevan al cuadrado para dar $E(t, \omega)$.

(2) Hasta que se alcanza el final de la respuesta al impulso, la ventana se mueve por la longitud de bloque de tiempo discreto de 1 ms hacia el final de la respuesta al impulso, las muestras en la ventana se rellenan con ceros hasta la longitud efectiva (es decir, las muestras más allá de la longitud efectiva se hacen cero), y luego se calcula el espectro de FFT, que da $E(t, \omega)$.

El enfoque anterior es ventajoso, ya que no se requiere ningún banco de filtros adicional o similar para el cálculo de banda estrecha del *EDR*; sólo se requiere un desplazamiento de la ventana. La Fig. 8 muestra un ejemplo para un alivio del decaimiento de energía logrado para una respuesta al impulso de acuerdo con el enfoque basado en FFT descrito con anterioridad.

De acuerdo con lo descrito en la Fig. 7 con respecto a los pasos 504 y 508, la correlación determinada en la forma descrita con anterioridad será comparada a un umbral predefinido. Cuanto menor sea el umbral, más el momento de transición se mueve hacia el final de la respuesta al impulso. Por ejemplo, para las respuestas al impulso binaural, si el umbral se elige para que sea $1/e \approx 0,3679$ (véase también la referencia de la técnica anterior [2]), la transición es demasiado temprana en algunos ángulos azimutales, dado que la correlación cae por debajo del umbral ya antes de que ocurriera o incidiera la primera reflexión. Sin embargo, dado que se sabe que el momento de transición debe ser más tarde de la hora de llegada de la primera reflexión, debido a que la primera reflexión es claramente distinta y puede no ser segura para la reverberación difusa tardía, de acuerdo con las realizaciones, el umbral no se define como un umbral fijo. Más bien, de acuerdo con el enfoque de la invención, el umbral se define de manera tal que es dependiente de la correlación en el momento que incide de la primera reflexión. Con esta definición, se asegura que la primera reflexión siempre se encuentre antes que el momento de transición. De acuerdo con las realizaciones, el momento de transición, de acuerdo con lo mostrado en el paso 508, se considera que se alcanza cuando se aplica lo siguiente:

$$\rho(t) = c \cdot \rho(t_F)$$

donde

$\rho(t_F)$ = medida de correlación para la seleccionada de las reflexiones tempranas,
 t_F = índice de tiempo en el que la seleccionada de las reflexiones tempranas después de que incide el sonido directo,
 c = el valor constante que está basado en $\frac{1}{e}$, siendo e el número de Euler.

De acuerdo con las realizaciones, el valor constante puede ser $\frac{1}{e}$, sin embargo, la presente invención no se limita a este valor. De acuerdo con las realizaciones, el valor constante se puede aproximar por $\frac{1}{e}$, por ejemplo, por medio del redondeo o truncando de $\frac{1}{e}$ con respecto a un lugar decimal predefinido (véase a continuación).

En la realización descrita, t_F es el índice del bloque de tiempo donde la primera reflexión después de que incide el sonido directo.

La Fig. 9 representa la determinación del momento de transición de acuerdo con el enfoque de la invención en el que el umbral se calcula dependiendo de la respuesta al impulso por la multiplicación de la correlación en el punto de incisión de la primera reflexión y un valor fijo o constante de $1/e$. La amplitud de la respuesta 600 al impulso de la habitación se

muestra sobre el número de muestras, y también se indica una primera reflexión 602. La forma 604 de onda indica los valores de correlación obtenidos por aplicación de la ecuación (2). En 606 se muestra el valor de correlación en la primera reflexión que, en el ejemplo representado, tiene un valor de 0,58. Además, el umbral fijado convencionalmente utilizado de $1/e$ se muestra en 608. El valor 606 de correlación para la primera reflexión y el valor fijo original de $1/e$, se aplican a un multiplicador 610 que genera el nuevo umbral que depende del valor de correlación en la primera reflexión y, en la realización descrita tiene un valor de 0,21 de acuerdo con lo mostrado en 612. Por lo tanto, en comparación con los enfoques convencionales, el punto 614 de transición se mueve además hacia la derecha de modo que todas las muestras siguientes el punto 614 de transición se consideran ahora reverberación 304 tardía y todas las muestras anteriores se consideran reflexión 302 temprana. Se puede observar que el tiempo 614 de decisión resultante es más robusto. Por ejemplo, en una respuesta al impulso de la habitación binaural esto significa que el momento de transición calculada es mucho más estable a lo largo del ángulo azimutal. Esto se puede observar a partir de una comparación de las Figs. 10 y 11. La Fig. 10 muestra los tiempos de transición cuando se aplica el enfoque descrito en la referencia de la técnica anterior [1] para el canal 700 izquierdo y el canal 702 derecho para una respuesta al impulso de la habitación binaural medida por el uso de la aplicación EDC descrita con anterioridad pero con un umbral fijo de $1/e$. Una dependencia de la oreja y el ángulo azimutal es claramente visible, así como la profundidad en el momento de transición a menos de 10 ms que se deben al hecho de que la correlación $p(t)$ cae por debajo del umbral antes de que incida la primera reflexión. La Fig. 11 muestra el momento de transición para el canal 700 izquierdo y el canal 702 derecho cuando se calcula de acuerdo con el enfoque de la invención. Se puede observar que el momento de transición resultante es mucho menos dependiente de la oreja y el ángulo azimutal en comparación con el enfoque convencional explicó con respecto a la Fig. 10.

De acuerdo con las realizaciones, se considera que el momento de transición se alcanza cuando la correlación cae por debajo de o es igual al valor umbral para la primera vez y no aumenta de nuevo por encima del umbral después. El valor de tiempo que está asociado con este ejemplo en la función de correlación calculado es el tiempo en el que se considera la reverberación tardía de la respuesta al impulso para comenzar. De acuerdo con el enfoque de la invención, el tiempo de incisión de la primera reflexión se puede determinar por un operador de curtosis en funcionamiento, de acuerdo con lo descrito en la referencia de la técnica anterior [6]. En forma alternativa, la primera reflexión se puede detectar por otros métodos, por ejemplo, por una detección de umbral o por una detección de ataques tal como se describe, por ejemplo, en la referencia de la técnica anterior [7].

De acuerdo con las realizaciones, $e^{-1} = 0,3679$ se utiliza como un valor para indicar una baja correlación en los procesos estocásticos como se indica, por ejemplo, también en la referencia de la técnica anterior [1]. Según las realizaciones, este valor se utiliza con cuatro dígitos decimales de manera tal que e^{-1} se aproxima como 0,3679. De acuerdo con otras realizaciones, también se pueden utilizar más o menos dígitos decimales y se ha observado que el momento de transición detectada cambia en consecuencia con la desviación desde el número exacto de e^{-1} . Por ejemplo, cuando se utiliza el valor de 0,368 esto se traduce sólo en cambios mínimos en el momento de transición de 1 ms de abajo.

De acuerdo con las realizaciones adicionales, la respuesta al impulso puede ser de banda limitada, y en este caso, el *EDR* se puede calcular en un intervalo de frecuencia limitado y también la correlación puede ser calculada sobre el intervalo de frecuencias limitado del *EDR*. También se pueden utilizar transformadas de frecuencia alternativa o bancos de filtros, por ejemplo, enfoques que operan completamente en el dominio de FFT, con el consiguiente ahorro de transformaciones adicionales, por ejemplo cuando se utiliza con base de FFT de filtrado/convolución.

Se observa que en la descripción anterior de las realizaciones, se ha hecho referencia a un valor del valor de correlación para la primera reflexión. Sin embargo, otras realizaciones pueden utilizar un valor de correlación calculado para otra de las reflexiones tempranas.

De acuerdo con lo mencionado con anterioridad, el enfoque de la invención, de acuerdo con las realizaciones se puede utilizar en un renderizador binaural para el procesamiento de señales de audio binaurales. A continuación se describirá una realización del procesamiento binaural de señales de audio. El procesamiento binaural se puede llevar a cabo como un proceso descodificador de la conversión de la señal decodificada en una señal de mezclado descendente binaural que proporciona una experiencia de sonido envolvente cuando se escucha con auriculares.

La Fig. 12 muestra una representación esquemática de un renderizador 800 binaural para el procesamiento binaural de señales de audio de acuerdo con una realización de la presente invención. La Fig. 12 también proporciona una visión general del procesamiento de dominio de QMF en el renderizador binaural. En una entrada 802, el renderizador 800 binaural recibe la señal de audio a ser procesada, por ejemplo, una señal de entrada que incluye N canales y 64 bandas de QMF. Además, el renderizador 800 binaural recibe un número de parámetros de entrada para controlar el procesamiento de la señal de audio. Los parámetros de entrada incluyen la binaural respuesta 804 al impulso de la habitación (BRIR) para los canales $2 \times N$ y 64 bandas de QMF, una indicación 806 $K_{\text{máx}}$ de la banda máxima que se utiliza para la convolución de la señal de entrada de audio con la parte de reflexión 804 temprana de las BRIR, y los parámetros del reverberador 808 y 810 se ha mencionado con anterioridad (RT60 y la energía de reverberación). El renderizador 800 binaural comprende un renderizador 81 de convolución rápida para el procesamiento de la señal de audio de entrada 802 con la parte temprana de las BRIR 804 recibidas. El procesador 812 genera en una salida de la señal 814 temprana procesada incluye dos canales y bandas $K_{\text{máx}}$ QMF. El renderizador 800 binaural comprende,

además de la rama tramitación anticipada que tiene el procesador de convolución 812 rápida, también una rama de reverberación que incluye dos reverberadores 816a y 816b cada reciben como parámetro de entrada la información 808 RT60 y la información 810 de energía de reverberación. La rama de reverberación incluye además un procesador 818 de mezclado descendente estéreo y un procesador 820 de análisis de correlación en donde ambos reciben la señal 802 de audio de entrada. Además, se proporcionan dos etapas 821a y 821b de ganancia entre el procesador 818 de mezclado descendente estéreo y los respectivos reverberadores 816a y 816b para controlar la ganancia de un mezclado descendente la señal 822 proporcionada por el procesador 818 de mezclado descendente estéreo. El procesador de mezclado 818 descendente estéreo proporciona sobre la base de la señal de entrada 802 la señal 822 de mezclado descendente que tiene dos bandas de 64 y bandas de QMF. La ganancia de las etapas de ganancia 821a y 821b está controlada por señales respectivas 824a y 824b de control proporcionadas por el procesador 820 de análisis de correlación. La ganancia controla la señal de mezclado descendente que se introduce en los reverberadores 816a y 816b respectivos para generar señales respectivas 826a, 826b de reverberado. La señal 814 temprana procesada y las señales 826a, 826b reverberadas son recibidas por un mezclador 828 que combina las señales recibidas en la señal 830 de audio de salida que tiene dos canales y 64 bandas de QMF. Además, de acuerdo con la presente invención, el procesador 812 de convolución rápida y los reverberadores 816a y 816b reciben un parámetro 832 de entrada adicional que indica la transición en la respuesta 804 al impulso de la habitación de la parte temprana a la reverberación tardía determinado de acuerdo con lo discutido con anterioridad.

El módulo 800 de renderizador binaural (por ejemplo, el renderizador 236 binaural de la Fig. 2 o la Fig. 4) tiene como entrada 802 la secuencia de datos descodificada. La señal es procesada por un banco de filtros de análisis de QMF de acuerdo con lo indicado en la norma ISO/IEC 14496-3:2009, 4.B.18.2 subcláusula con las modificaciones indicadas en la norma ISO/IEC 14496-3:2009, subcláusula 8.6.4.2. El módulo de procesador 800 también puede procesar datos de entrada de dominio de QMF; en este caso el de filtros de análisis se puede omitir. Las respuestas al impulso de la habitación binaurales (BRIR) 804 se representan como filtros de dominio de QMF complejos. La conversión del dominio del tiempo las respuestas al impulso de la habitación binaurales al complejo representación filtro QMF 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 temprana de reflexión 301, 302 (véase la Fig. 5) y la reverberación 304 difusa tardía no está incluida. El punto 832 de transición de las reflexiones tempranas de reverberación tardía se determina de acuerdo con lo descrito con anterioridad, por ejemplo, por un análisis de la BRIR 804 en un paso de procesamiento previo del procesamiento binaural. Las señales de audio de dominio de QMF 802 y las BRIR de dominio de QMF 804 son entonces procesadas por un Convolución rápida banda a banda 812 para realizar el procesamiento binaural. Un dominio de QMF reverberador 816a, 816b se utiliza para generar un dominio de reverberación tardía de 2 canales QMF 826a, 826b. El módulo 816a, 816b de reverberación utiliza un conjunto de reverberación dependiente de la frecuencia 808 y los valores 810 de energía para adaptar las características de la reverberación. La forma de onda de la reverberación está basada en un equipo 818 estéreo de mezclado descendente de la señal 802 de entrada de audio y se escala en forma 821a, 821b adaptativa en amplitud en función de un análisis 820 de correlación de la señal 802 de audio multi-canal. El resultado del dominio de 2 canales de QMF convolucional 814 y el dominio de 2 canales QMF 816a, 816b de reverberación se combinan luego 828 y finalmente, dos bancos de filtros de síntesis QMF calculan las señales 830 de salida de dominio de tiempo binaural de acuerdo con lo indicado en la norma ISO/IEC 14496-3:2009, subsección 4.6.18.4.2. El renderizador también puede producir datos de salida de dominio QMF; luego se omite el banco de filtros de síntesis.

Definiciones

Las señales 802 de audio que se introducen en el módulo 800 de renderizador binaural de aquí en adelante se conocen como señales de entrada. Las señales 830 de audio que son el resultado del procesamiento binaural se conocen como señales de salida. Las señales 802 de entrada del módulo 800 de renderizador binaural son señales de salida de audio del decodificador de núcleo (véase, por ejemplo las señales 228 en la Fig. 2). Se utilizan 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 mezcla descendente que contiene los coeficientes de mezclado descendente no negativos con valores reales (ganancias de mezclado descendente). M_{DMX} es de dimensión $N_{out} \times N_{in}$
L	Longitud del marco medida en muestras de audio de dominio de tiempo.
v	Índice de muestra del dominio de tiempo
n	Índice de segmentos de tiempo QMF (índice de muestra subbanda)
L_n	Longitud del marco medido en ranuras de tiempo QMF
F	Índice de marcos (número de cuadro)

K	Número de bandas de frecuencia de QMF, $K = 64$
k	Índice de banda QMF (1..64)
A, B, ch	Índices de canales (números de canal de configuraciones de canal)
L_{trans}	Longitud de la parte de reflexión temprana de la BRIR en muestras de dominio de tiempo
$L_{trans,n}$	Longitud de la parte de reflexión temprana de la BRIR en intervalos de tiempo de QMF
N_{BRIR}	Número de pares de BRIR en un conjunto de datos BRIR
L_{FFT}	Duración de la transformada de FFT
$\Re(\cdot)$	Parte real de una señal de valor complejo
$\Im(\cdot)$	Parte imaginaria de una señal de valor complejo
m_{conv}	Vector que las señales de que el canal de la señal de entrada que pertenece al par BRIR en el conjunto de datos BRIR
f_{max}	Frecuencia máxima utilizada para el procesamiento 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 utiliza para la convolución de la señal de entrada de audio con la parte de reflexión temprana de las BRIR
a	Coefficiente de matriz de mezclado descendente
$c_{eq,k}$	Factor de igualación de energía banda a banda
ε	Constante numérica, $\varepsilon = 10^{-20}$
d	El retraso en los intervalos de tiempo de dominio de QMF
$\tilde{y}_{ch}^{n',k}$	Representación de la señal de dominio Pseudo-FFT en la banda de frecuencia k
n'	Índice de frecuencia de Pseudo-FFT
$\tilde{h}^{n',k}$	Dominio de la representación Pseudo-FFT de BRIR en la banda de frecuencia k
$\tilde{z}_{ch,conv}^{n',k}$	Resultado Pseudo-FFT de convolución de dominio en la banda de frecuencia k
$\hat{z}_{ch,conv}^{n,k}$	Señal intermedia: resultado de convolución de 2 canales en el dominio de QMF
$\hat{z}_{ch,rev}^{n,k}$	Señal intermedia: reverberación de 2 canales en el dominio de QMF
K_{ana}	Número de bandas de frecuencia de análisis (utilizado para el reverberador)
$f_{c,ana}$	Frecuencias centrales de las bandas de frecuencia de análisis
$N_{DMX,act}$	Número de canales que se mezclan en un canal de la mezcla estéreo y son activos en el marco de señal real
c_{corr}	Coefficiente global de correlación para un marco de señal
$c_{corr}^{A,B}$	Coefficiente de correlación para la combinación de canales A, B
$\sigma_{\hat{y}_{ch,A}^n}$	Desviación estándar para el intervalo de tiempo n de la señal $\hat{y}_{ch,A}^n$
\mathbf{c}_{scale}	Vector de dos factores de escala
$\tilde{\mathbf{c}}_{scale}$	Vector de dos factores de escala, suavizado con el tiempo

Procesamiento

5 A continuación se describe el procesamiento de la señal de entrada. El módulo de renderizador binaural opera en cuadros contiguos, no solapados de longitud $L = 2048$ muestras en el dominio de tiempo de las señales de audio de entrada y salidas de un marco de L muestras por marco de entrada de la longitud del procesado L .

(1) Inicialización y procesamiento previo

ES 2 758 757 T3

La inicialización del bloque de procesamiento binaural se lleva a cabo antes de que se lleve a cabo el procesamiento de las muestras de audio suministradas por el decodificador del núcleo (véase, por ejemplo, el decodificador de 200 en la Fig. 2). La inicialización se compone de varios pasos de procesamiento.

5 (a) Lectura de valores de análisis

El módulo del reverberador 816a, 816b tiene un conjunto dependiente de la frecuencia de la reverberación 808 y valores de energía 810 como parámetros de entrada. Estos valores se leen desde una interfaz en la inicialización del módulo de procesamiento binaural 800. Además, el momento de transición 832 de las reflexiones tempranas y finales de la reverberación en muestras de dominio de tiempo que se lee. Los valores se pueden almacenar en un archivo binario escrito con 32 bits por muestra, valores de coma flotante, ordenamiento little-endian. Los valores de lectura que se necesitan para el procesamiento 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 $f_{c,ana}$ de bandas de frecuencia	K_{ana}	Flotante
Tiempos de reverberación RT60 en segundos	K_{ana}	Flotante
Valores de energía que representan la energía (amplitud de la potencia de dos) de la parte de reverberación tardía de una BRIR	K_{ana}	Flotante

15 (b) Lectura y procesamiento previo de BRIR

Las respuestas 804 al impulso de la habitación binaural se leen a partir de dos archivos dedicados que almacenan individualmente las BRIR de los oídos izquierdo y derecho. Las muestras en el dominio de tiempo de las BRIR se almacenan en archivos enteros de onda con una resolución de 24 bits por muestra y 32 canales. El orden de BRIR en el archivo está de acuerdo con lo indicado en la siguiente tabla:

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

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

- 5 Como un paso de procesamiento previo, el conjunto dado de respuestas al impulso de la habitación binaurales (BRIR) se transforma de filtros de dominio de tiempo al dominio de filtros QMF de valor complejo. La implementación de los filtros de dominio de tiempo dados en el dominio de QMF de valor complejo se realiza de acuerdo a la norma ISO/IEC FDIS 23003-1:2006, Anexo B. Los coeficientes del filtro prototipo para la conversión de filtro se utilizan de acuerdo con ISO/IEC FDIS 23003 -1:2006, Anexo B, Tabla B.1. La representación en el dominio de tiempo $\tilde{\mathbf{h}}_{\text{ch}}^v = [\tilde{h}_1^v \cdots \tilde{h}_{N_{\text{BRIR}}}^v]$
- 10 con $1 \leq v \leq L_{\text{trans}}$ se procesa para obtener una valorada filtro de dominio QMF complejo $\hat{\mathbf{h}}_{\text{ch}}^{n,k} = [\hat{h}_1^{n,k} \cdots \hat{h}_{N_{\text{BRIR}}}^{n,k}]$ con $1 \leq n \leq L_{\text{trans},n}$.

(2) *Procesamiento de la señal de audio*

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

El tratamiento tiene como entrada

- 20
- los datos 802 de audio decodificados desde el decodificador núcleo,
 - el complejo dominio de representación de QMF de la parte de la reflexión 804 temprana de la BRIR fijo, y
- 25
- el parámetro dependiente de la frecuencia 808, 810, 832 fija que se utiliza por el 816a reverberador de dominio de QMF, 816b para generar la reverberación 826a, 826b tardía.

(a) *Análisis de QMF de la señal de audio*

- 30 Como el primer paso de procesamiento, las transformadas del módulo renderizador binaural $L = 2048$ muestras en el dominio de tiempo de la N_{in} Canales de señal de entrada el dominio del tiempo (que viene del decodificador del núcleo) $[\tilde{\mathbf{y}}_{\text{ch},1}^v \cdots \tilde{\mathbf{y}}_{\text{ch},N_{\text{in}}}^v] = \tilde{\mathbf{y}}_{\text{ch}}^v$ a una N_{in} Representación de la señal de dominio de QMF-canal 802 de dimensión $L_n = 32$ intervalos de tiempo de QMF (Indicador de la ranura n) Y $K = 64$ bandas de frecuencia (índice de banda k).
- 35 Un análisis de QMF de acuerdo con lo indicado 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. se lleva a cabo en un fotograma de la señal de dominio de tiempo $\tilde{\mathbf{y}}_{\text{ch}}^v$ para obtener un marco de la señal de dominio de QMF $[\hat{\mathbf{y}}_{\text{ch},1}^{n,k} \cdots \hat{\mathbf{y}}_{\text{ch},N_{\text{in}}}^{n,k}] = \hat{\mathbf{y}}_{\text{ch}}^{n,k}$ con $1 \leq v \leq L$ y $1 \leq n \leq L_n$.

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

A continuación, una convolución 812 rápida banda a banda se lleva a cabo para el procesamiento de la señal 802 de audio de dominio de QMF y las BRIR de dominio de QMF 804. Se puede llevar a cabo un análisis FFT para cada banda de frecuencia de QMF para cada canal de la señal 802 de entrada y cada BRIR 804.

- 45 Debido a los valores complejos en el dominio de QMF, se lleva a cabo un análisis FFT en la parte real de la representación de la señal de dominio de QMF y un análisis FFT sobre las partes imaginarias de la representación de la

señal de dominio de QMF. Los resultados se combinan entonces para formar la señal de dominio final Banda a banda valor complejo pseudo-FFT

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

5 y los valores complejos BRIR-Banda a banda

$$\tilde{h}_1^{n',k} = \text{FFT}(\hat{h}_1^{n',k}) = \text{FFT}(\Re(\hat{h}_1^{n',k})) + j \cdot \text{FFT}(\Im(\hat{h}_1^{n',k}))$$

para el oído izquierdo

$$\tilde{h}_2^{n',k} = \text{FFT}(\hat{h}_2^{n',k}) = \text{FFT}(\Re(\hat{h}_2^{n',k})) + j \cdot \text{FFT}(\Im(\hat{h}_2^{n',k}))$$

para el oído derecho.

La longitud de transformada de FFT se determina de acuerdo a la longitud del valor complejo de filtros BRIR de dominio de QMF $L_{trans,n}$ y la longitud del marco en intervalos de tiempo de dominio de QMF L_n de manera tal que

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

Las señales de dominio pseudo-FFT de valores complejos se multiplican con los filtros BRIR dominio pseudo-FFT de valores complejos para formar los resultados de convolución rápida. Un vector \mathbf{m}_{conv} se utiliza para indicar qué canal de la señal de entrada corresponde a la que BRIR par en el conjunto de datos BRIR.

Esta multiplicación se realiza banda a banda para todas las bandas de frecuencia de QMF k con $1 \leq k \leq K_{max}$. La banda máxima K_{max} está determinado por la banda de QMF que representa una frecuencia de 18 kHz o ya sea la frecuencia de la señal máxima que está presente en la señal de audio del decodificador núcleo

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

Los resultados de la multiplicación de cada canal de entrada de audio con cada par BRIR se resumen en cada banda de frecuencia QMF k con $1 \leq k \leq K_{max}$ resultando en un intermedio K_{max} señal de 2 canales de dominio de banda de pseudo-FFT.

$$\tilde{z}_{ch,1,conv}^{n',k} = \sum_{ch=1}^{ch=N_{in}} \tilde{y}_{ch,ch}^{n',k} \cdot \tilde{h}_{1,\mathbf{m}_{conv}[ch]}^{n',k} \quad \text{y} \quad \tilde{z}_{ch,2,conv}^{n',k} = \sum_{ch=1}^{ch=N_{in}} \tilde{y}_{ch,ch}^{n',k} \cdot \tilde{h}_{2,\mathbf{m}_{conv}[ch]}^{n',k}$$

son el resultado de convolución de pseudo-FFT $\tilde{z}_{ch,conv}^{n',k} = [\tilde{z}_{ch,1,conv}^{n',k}, \tilde{z}_{ch,2,conv}^{n',k}]$ en la banda de frecuencia de dominio de QMF k .

A continuación, una síntesis de FFT banda a banda se lleva a cabo para transformar el resultado de convolución de nuevo al dominio de QMF que resulta en un intermedio K_{max} -Banda de 2 canales de la señal de dominio de QMF con L_{FFT} ranuras de tiempo $\hat{z}_{ch,conv}^{n,k} = [\hat{z}_{ch,1,conv}^{n,k}, \hat{z}_{ch,2,conv}^{n,k}]$ con $1 \leq n \leq L_{FFT}$ y $1 \leq k \leq K_{max}$.

Para cada marco de la señal de entrada con QMF dominio $L = 32$ intervalos de tiempo de un marco de la señal se devuelve un resultado de convolución con $L = 32$ ranuras de tiempo. Los restantes $L_{FFT} - 32$ intervalos de tiempo se almacenan y un procesamiento de solape-suma se lleva a cabo en los marcos siguientes.

(c) Generación de reverberación tardía

Como una segunda señal 826a, 826b intermedia una señal de reverberación llamada $\hat{z}_{ch,rev}^{n,k} = [\hat{z}_{ch,1,rev}^{n,k}, \hat{z}_{ch,2,rev}^{n,k}]$ es generada por un módulo de reverberador 816a, 816b de dominio de frecuencia. El reverberador 816a, 816b dominio de la frecuencia toma como entrada

- un dominio estéreo QMF de mezclado 822 descendente de un marco de la señal de entrada,
- un conjunto de parámetros que contiene la reverberación dependiente de la frecuencia 808 y los valores 810 de energía.

El reverberador 816a, 816b del dominio de la frecuencia devuelve una cola del dominio de 2 canales QMF de reverberación tardía.

El número de banda máximo utilizado del conjunto de parámetros dependientes de la frecuencia se calcula en función de la frecuencia máxima.

5 En primer lugar, un dominio 818 de QMF mezclado descendente estéreo de un marco de la señal de entrada $\hat{\mathbf{y}}_{ch}^{n,k}$ se lleva a cabo 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 mezclado descendente \mathbf{M}_{DMX} . Son de valor real y no negativo y la matriz de mezclado descendente es de dimensión $N_{out} \times N_{in}$. Contiene un valor distinto de cero, donde un canal de la señal de entrada se asigna a uno de los dos canales de salida.

10 Los canales que representan las posiciones de los altavoces en el hemisferio izquierdo se asignan al canal de salida a la izquierda y los canales que representan los altavoces situados en el hemisferio derecho se asignan al canal de salida correcto. Las señales de estos canales se ponderan con un coeficiente de 1. Los canales que representan los altavoces situados en el plano medio se asignan a los dos canales de salida de la señal binaural. Las señales de entrada de estos canales se ponderan con un coeficiente

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

Además, un paso de compensación de energía se lleva a cabo en el mezclado descendente. Se adapta la energía banda a banda de un canal de mezclado descendente para ser igual a la suma de la energía banda a banda de los canales de señal de entrada que están contenidos en este canal de mezclado descendente. Esta igualación de energía se lleva a cabo por una multiplicación banda a banda con un coeficiente de valor real

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

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

La Fig. 13 representa en forma esquemática el procesamiento en el reverberador 816a, 816b del dominio de frecuencia del renderizador 800 binaural de acuerdo con una realización de la presente invención.

30 En el reverberador del dominio de frecuencia de un mezclado descendente mono de la entrada estéreo se calcula por el uso de un mezclador 900 de entrada. Esto se hace aplicando incoherentemente un desplazamiento de fase de 90° en el segundo canal de entrada.

35 Esta señal mono se alimenta entonces a un bucle 902 de retardo de realimentación en cada banda de frecuencia k , lo que crea una secuencia de descomposición de los impulsos. Le siguen los decorrelacionadores FIR paralelos que distribuyen la energía de la señal de una manera decadente en los intervalos entre los impulsos y crean incoherencia entre los canales de salida. Una descomposición de densidad aparente de filtro se aplica para crear el decaimiento de la energía. Las operaciones de eliminación del grifo del filtro se limitan a cuatro opciones para implementar un correlador escasa y multiplicador libre.

40 Tras el cálculo de la reverberación de una coherencia (ICC) de corrección entre canales 904 se incluye en el módulo reverberador para cada banda de frecuencia de QMF. En la frecuencia dependen de las ganancias directas g_{direct} y ganancias de mezcla cruzada g_{cross} paso de corrección ICC sirven para adaptar la CPI.

45 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 un número de puntos de frecuencia que se asigna internamente a la $K = 64$ bandas de frecuencia de QMF.

Se utilizan dos instancias del reverberador de dominio de la frecuencia para calcular la señal intermedio final

50 $\hat{\mathbf{z}}_{ch,rev}^{n,k} = \left[\hat{z}_{ch,1,rev}^{n,k}, \hat{z}_{ch,2,rev}^{n,k} \right]$. 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 en el marco de señal de reverberación final que tiene la dimensión de 2 canales, 64 bandas y 32 ranuras de tiempo.

55 El mezclado 822 descendente estéreo es dos veces escalados 821a, b de acuerdo con una medida 820 de correlación del marco de señal de entrada para asegurar la escala derecha de la salida de reverberador. El factor de escala se

define como un valor en el intervalo de $\left[\sqrt{N_{\text{DMX,act}}}, N_{\text{DMX,act}} \right]$ linealmente en función de un coeficiente de correlación c_{corr} entre 0 y 1 con

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

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

5 donde $\sigma_{\hat{y}_{\text{ch},A}^n}$ significa la desviación estándar a través de una ranura de tiempo n de canal A , El operador $\{^*\}$

denota el conjugado complejo y \hat{y} es la versión de promedio cero de la señal de dominio de QMF \hat{y} en el marco de señal real.

10 c_{corr} se calcula dos veces: una vez para todos los canales A, B que están activos en el marco de la señal actual F y se incluyen en el canal izquierdo de la mezcla estéreo y una vez para todos los canales A, B que están activos en el marco de la señal actual F y que se incluyen en el canal derecho de la mezcla estéreo.

15 $N_{\text{DMX,act}}$ es el número de canales de entrada que reducen la mezcla a un canal de mezclado descendente A (Número de elemento de matriz en la A th fila de la matriz de mezclado descendente M_{DMX} que son diferente de cero) y que están activos en el marco actual.

Los factores de escala son, pues,

$$\mathbf{c}_{\text{scale}} = \left[c_{\text{scale},1}, c_{\text{scale},2} \right]$$

$$= \left[\sqrt{N_{\text{DMX,act},1}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},1} - \sqrt{N_{\text{DMX,act},1}}), \sqrt{N_{\text{DMX,act},2}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},2} - \sqrt{N_{\text{DMX,act},2}}) \right]$$

20 Los factores de escala se alisan por medio de marcos de señal de audio por medio de un filtro de paso bajo de 1^{er} orden resulta en factores de escala suavizadas $\tilde{\mathbf{c}}_{\text{scale}} = \left[\tilde{c}_{\text{scale},1}, \tilde{c}_{\text{scale},2} \right]$.

25 Los factores de escala se inicializan en el primer marco de datos de entrada de audio por un análisis de correlación de dominio de tiempo con los mismos medios.

La entrada de la primera instancia reverberador se escala con el factor de escala $\tilde{c}_{\text{scale},1}$ y la entrada de la segunda instancia reverberador se escala con el factor de escala $\tilde{c}_{\text{scale},2}$.

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

30 A continuación, el resultado convolucional 814, $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} = \left[\hat{z}_{\text{ch},1,\text{conv}}^{n,k}, \hat{z}_{\text{ch},2,\text{conv}}^{n,k} \right]$, y la salida 826a reverberador, 826b,

$\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k} = \left[\hat{z}_{\text{ch},1,\text{rev}}^{n,k}, \hat{z}_{\text{ch},2,\text{rev}}^{n,k} \right]$, para un dominio de marco de entrada de audio de QMF se combinan por un proceso de mezcla 828 que suma banda a banda las dos señales. Se debe tener en cuenta que las bandas superiores mayores que K_{max} son cero en $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k}$ porque la convolución sólo se lleva a cabo en las bandas hasta K_{max} .

35 La salida de la reverberación tardía se retrasa por un importe de $d = ((L_{\text{trans}} - 20 \cdot 64 + 1) / 64 + 0.5) + 1$ ranuras de tiempo en el proceso de mezcla.

40 El retraso d tiene en cuenta el momento de transición desde las reflexiones tempranas a las reflexiones finales de los años en las BRIR y un retraso inicial del reverberador de 20 ranuras de tiempo de QMF, así como un retraso de análisis de 0,5 intervalos de tiempo para el análisis de QMF de las BRIR para garantizar la inserción de reverberación tardía en un espacio de tiempo razonable. La señal combinada $\hat{\mathbf{z}}_{\text{ch}}^{n,k}$ en una ranura de tiempo n calculado por $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} + \hat{\mathbf{z}}_{\text{ch,rev}}^{n-d,k}$.

(e) Síntesis de QMF de la señal de dominio de QMF binaural

Un marco de 2 canales de 32 intervalos de tiempo de la señal de salida en el dominio de QMF $\hat{z}_{ch}^{n,k}$ se transforma en un marco de señal de dominio de tiempo de 2 canales con una longitud por la síntesis QMF de acuerdo con la norma ISO/IEC 14496-3:2009, subsección 4.6.18.4.2. lo que da la señal de salida en el dominio tiempo final 830,

$$\tilde{z}_{ch}^v = \left[\tilde{z}_{ch,1}^v \cdots \tilde{z}_{ch,2}^v \right].$$

Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a un paso de procedimiento o una función de un paso de procedimiento. En forma análoga, los aspectos descritos en el contexto de un paso método también representan una descripción de un bloque o elemento o función de un aparato correspondiente. Algunos o todos los pasos del método pueden ser ejecutados por (o por medio de) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, algunos uno o más de los pasos del método más importantes pueden ser ejecutados por un aparato de este tipo.

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

Algunas realizaciones de acuerdo con la invención comprenden un portador de datos que tiene señales de control legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de manera tal que se lleve a cabo uno de los métodos descritos en el presente documento.

En general, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, el código de programa siendo operativo para llevar a cabo uno de los métodos, cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede, por ejemplo, ser almacenado en un soporte legible por máquina.

Otras realizaciones comprenden el programa de ordenador para llevar a cabo uno de los métodos descritos en el presente documento, almacenado en un soporte legible por máquina.

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

Una realización adicional del método 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 de ordenador para realizar uno de los métodos descritos en el presente documento. El soporte de datos, el medio de almacenamiento digital o el medio de grabado en forma típica son tangibles y/o no transitorios.

Una realización adicional del método de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa de ordenador para llevar a cabo uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales se pueden, por ejemplo, configurar para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

Una realización comprende además un medio de procesamiento, por ejemplo, un ordenador o un dispositivo lógico programable, configurado para, o programado para, que lleve a cabo uno de los métodos descritos en el presente documento.

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

Una realización adicional de acuerdo con la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, por vía electrónica u ópticamente) un programa de ordenador para realizar uno de los métodos descritos en el presente documento a un receptor. El receptor puede, por ejemplo, ser un ordenador, un dispositivo móvil, un dispositivo de memoria o similares. El aparato o sistema puede, por ejemplo, comprender un servidor de archivos para transferir el programa de ordenador al receptor.

5 En algunas realizaciones, un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en campo) se puede utilizar para llevar a cabo todas o algunas de las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programable en campo puede cooperar con un microprocesador para llevar a cabo uno de los métodos descritos en el presente documento. En general, los métodos se llevan a cabo preferiblemente por cualquier aparato de hardware.

10 Las realizaciones descritas con anterioridad son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para aquéllos con experiencia en la técnica. Por lo tanto, la intención es limitarse sólo por el alcance de acuerdo con las reivindicaciones adjuntas y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de la presente.

Literatura

- 5 [1] T. Hidaka et al: "A new definition of boundary point between early reflections and late reverberation in room impulse responses". Forum Acusticum, 2005.
- [2] Jot et al: "Analysis and synthesis of room reverberation based on a statistical time frequency model".
- 10 [3] J.S. Abel, P. Huang: "A Simple, Robust Measure of Reverberation Echo Density". AES Convention, San Francisco, 2006.
- [4] R. Stewart, M. Sandler: "Statistical Measures of Early Reflections of Room Impulse Responses". DAFx, 2007.
- 15 [5] Reilly et al: "Using Auralisation for Creating Animated 3-D Sound Fields Across Multiple Speakers". AES Convention, Nueva York, 1995.
- [6] Usher, J.: "An improved method to determine the onset timings of reflections in an acoustic impulse response". Journal of the Acoustical Society of America, (2010, volumen 127) band 4, p. 172 a 177.
- 20 [7] Masri, P.: "Computer Modelling of Sound for Transformation and Synthesis of Musical Signals". PhD thesis, University of Bristol, 1996.

REIVINDICACIONES

5 1. Un método para el procesamiento de una señal (400, 802) de audio de acuerdo con una respuesta (300, 804) al impulso de la habitación, el método comprende:

10 procesar por separado (402, 404, 406, 422, 424, 812, 816a, 816b) de la señal (400, 802) de audio con una parte (301, 302) temprana de la respuesta (300, 804) al impulso de la habitación y una reverberación (304) tardía de la respuesta (300, 804) al impulso de la habitación o una reverberación sintética; y

15 combinar (418, 432, 828) la señal de audio procesada con la parte (414, 428, 814) temprana de la respuesta (300, 804) al impulso de la habitación y la señal (416, 430, 826a, 826b) de audio procesada con la reverberación (304) tardía de la respuesta (300, 804) al impulso de la habitación o con la reverberación sintética;

20 determinar una transición de la parte (301, 302) temprana a la reverberación (304) tardía en la respuesta al impulso de la habitación como un tiempo cuando una medida de correlación alcanza un umbral,

25 en el que la medida de correlación describe con respecto a la respuesta (300, 804) al impulso de la habitación la similitud de la decadencia en la energía acústica que incluye el estado inicial y de la decadencia en la energía acústica partiendo en cualquier punto en el tiempo, dicho punto en el tiempo sigue el estado inicial en un intervalo de frecuencia predefinida,

30 en el que el umbral se establece dependiendo de la medida de correlación para dicho punto en el tiempo, dicho punto en el tiempo siendo un tiempo de uno seleccionado de las reflexiones (302, 602) tempranas en la parte (301, 302) temprana de la respuesta (300, 804) al impulso de la habitación y en el que la seleccionada de las reflexiones tempranas es la primera reflexión (602).

2. El método de acuerdo con la reivindicación 1, en el que la determinación de la transición comprende:

35 determinar (500) una distribución de la energía acústica en base a la respuesta (300, 804) al impulso de la habitación; y;

determinar (502) una pluralidad de medidas de correlación que indican para una pluralidad de porciones de la distribución determinada una correlación entre la energía acústica en la porción respectiva de la distribución determinada y la energía acústica en un estado inicial.

3. El método de acuerdo con la reivindicación 2, en el que

40 la determinación de la distribución comprende la determinación de una distribución de tiempo-frecuencia de la energía acústica, y

una porción de la distribución comprende un bloque de tiempo de una longitud predefinida, el estado inicial se define por el primero de la pluralidad de bloques de tiempo de la distribución de tiempo-frecuencia.

4. El método de acuerdo con las reivindicaciones 2 o 3, en el que la determinación de la distribución comprende calcular el alivio del decaimiento de energía, *EDR*, de la respuesta (300, 804) al impulso de la habitación.

5. El método de acuerdo con la reivindicación 4, en el que el alivio del decaimiento de energía, *EDR*, se calcula de acuerdo con lo presentado a continuación:

$$E(t, \omega) = \left| \int_t^\infty h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

50 donde

$E(t, \omega)$ = alivio del decaimiento de energía,
 $h(\tau)$ = respuesta (300, 804) al impulso de la habitación,
 $\omega = 2\pi F$.

55 6. El método de acuerdo con una de las reivindicaciones 3 a 5, en el que la respuesta (300, 804) al impulso de la habitación tiene una longitud efectiva predefinida, y en el que la determinación de la distribución de tiempo-frecuencia comprende el cálculo del espectro de FFT de la respuesta (300, 804) al impulso de la habitación por el uso de una ventana que tiene una longitud que corresponde a la longitud efectiva de la respuesta al impulso de la habitación.

60 7. El método de acuerdo con la reivindicación 6, en el que

la energía acústica en el estado inicial se determina al tomar la longitud efectiva completa de la respuesta (300, 804) (300, 804) al impulso de la habitación al calcular el espectro de FFT y al tomar el cuadrado de los valores absolutos, y

la energía acústica de un bloque de tiempo se determina por medio del desplazamiento de la ventana por el tiempo asociado con el bloque de tiempo, el relleno con ceros de las muestras de ventana a la longitud efectiva, el cálculo del espectro de FFT y al tomar el cuadrado de los valores absolutos.

5

8. El método de la una de acuerdo con las reivindicaciones 1 a 7, en el que la medida de correlación se calcula de acuerdo con lo presentado a continuación:

$$\rho(t) = \frac{\sum_{\omega}(E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega}(E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega}(E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega}(E(t, \omega) - \bar{E}(t, \omega))^2}}$$

donde

- 10 $\rho(t)$ = medida de correlación,
 $E(1, \omega)$ = alivio del decaimiento de energía en la frecuencia f,
 $\bar{E}(1, \omega)$ = valor medio a lo largo de todas las frecuencias del alivio del decaimiento de energía inicial,
 $E(t, \omega)$ = alivio del decaimiento de energía en la frecuencia f a partir de un tiempo t,
 $\bar{E}(t, \omega)$ = valor medio a lo largo de todas las frecuencias del alivio del decaimiento de energía de
15 intervalo completo a partir del tiempo t,
 ω = $2\pi f$.

9. El método de acuerdo con una de las reivindicaciones 1 a 8, en el que el umbral se determina en base a un valor constante y la medida de correlación para la seleccionada de las reflexiones (302, 602) tempranas.

20

10. El método de acuerdo con la reivindicación 9, en el que la constante es 1/e, y en el que el umbral se define de acuerdo con lo presentado a continuación:

$$\rho(t) = c \cdot \rho(t_F)$$

donde

- 25 $\rho(t_F)$ = medida de correlación para la seleccionada de las reflexiones (302, 602),
 t_F = índice de tiempo donde la seleccionada de las reflexiones (302, 602) tempranas después de que el sonido directo (300) incide,
 c = el valor constante que está basado en $\frac{1}{e}$, siendo e el número de Euler.

30 11. Un producto de programa de computación no tangible que comprende un medio legible por ordenador que almacena instrucciones que, cuando se ejecuta en un ordenador, llevan a cabo el método de acuerdo con las reivindicaciones 1 a 10.

12. Una unidad de procesamiento de señales, que comprende

35 una entrada para recibir una señal (400, 802) de audio;

un procesador configurado para el procesamiento de la señal (400, 802) de audio recibida con una parte (301, 302) temprana de una respuesta (300, 804) al impulso de la habitación y con una reverberación (304) tardía del impulso de respuesta (300, 804) de la habitación o una reverberación sintética; y
40 una salida para combinar (418, 432, 828) la señal de audio procesada con la parte (414, 428, 814) temprana de la respuesta (300, 804) al impulso de la habitación y la señal (416, 430, 826a, 826b) de audio procesada con la reverberación (304) tardía de la respuesta (300, 804) al impulso de la habitación o con la reverberación sintética, en una señal de audio de salida; en la que el procesador está configurado para determinar una transición desde la parte (301, 302) temprana a la reverberación (304) tardía en la respuesta (300, 804) al impulso de la habitación como un tiempo en
45 el que una medida de correlación alcanza un umbral,

en la que la medida de correlación describe con respecto a la respuesta (300, 804) de impulso de la habitación una similitud de un decaimiento en energía acústica que incluye un estado inicial y del decaimiento en la energía acústica a partir de un punto en el tiempo, siguiendo dicho punto en el tiempo el estado inicial de un intervalo de frecuencia predefinida,
50

en el que el umbral se establece dependiendo de la medida de correlación para dicho punto en el tiempo, dicho punto en el tiempo seleccionándose de una de las reflexiones (302, 602) tempranas en la parte (301, 302) temprana de la respuesta (300, 804) al impulso de la habitación, y en la que la seleccionada de las reflexiones tempranas es la primera reflexión (602).
55

13. La unidad de procesamiento de señales de acuerdo con la reivindicación 12, que comprende:

60 un procesador (402, 404, 422, 812) de la parte temprana para el procesamiento de la señal (400, 802) de audio recibida de acuerdo con la parte (301, 302, 602) temprana de la respuesta (300, 804) al impulso de la habitación; y

un procesador 406, 424, 816a, 816b de reverberación tardía para el procesamiento de la señal (400, 802) de audio recibida de acuerdo con la reverberación (304) tardía de la respuesta (300, 804) al impulso de la habitación.

- 5 14. Un codificador de audio para codificar una señal de audio, en el que
- el codificador de audio está configurado para el procesamiento de una señal de audio a ser codificada de acuerdo con una respuesta (300, 804) al impulso de la habitación de acuerdo con el método de acuerdo con una de las reivindicaciones 1 a 10.
- 10 15. El codificador de audio de acuerdo con la reivindicación 14, en el que el codificador de audio comprende una unidad de procesamiento de señales de acuerdo con la reivindicación 12 o 13.
- 15 16. Un decodificador de audio para decodificar una señal de audio codificada, en el que el decodificador de audio está configurado para el procesamiento de una señal (400, 802) de audio decodificada de acuerdo con una respuesta (300, 804) al impulso de la habitación de acuerdo con un método de acuerdo con una de las reivindicaciones 1 a 10.
- 20 17. El decodificador de audio de acuerdo con la reivindicación 16, en el que el decodificador de audio comprende una unidad de procesamiento de señales de acuerdo con la reivindicación 12 o 13.
18. Un renderizador binaural, que comprende una unidad de procesamiento de señales de acuerdo con la reivindicación 12 o 13.

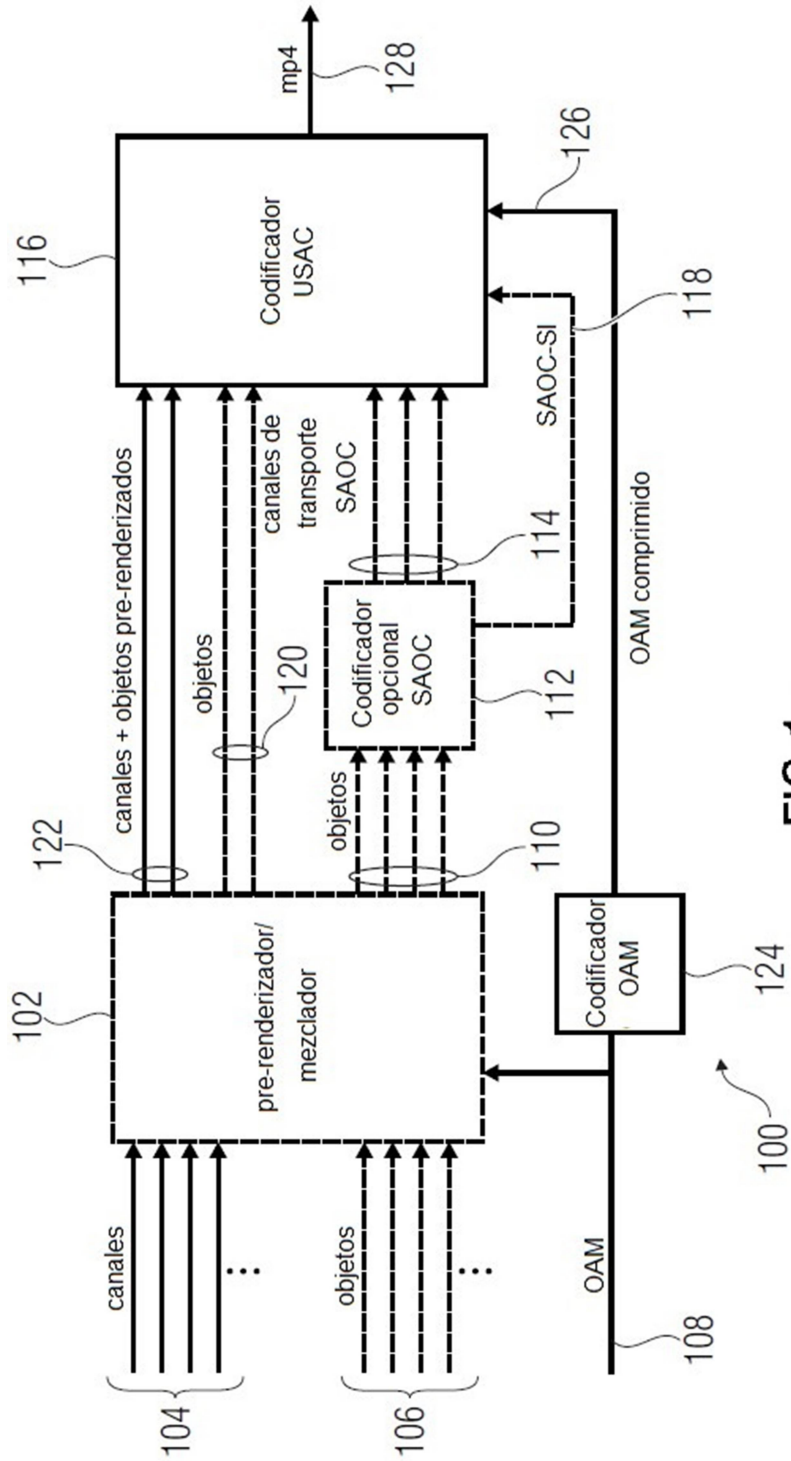


FIG 1

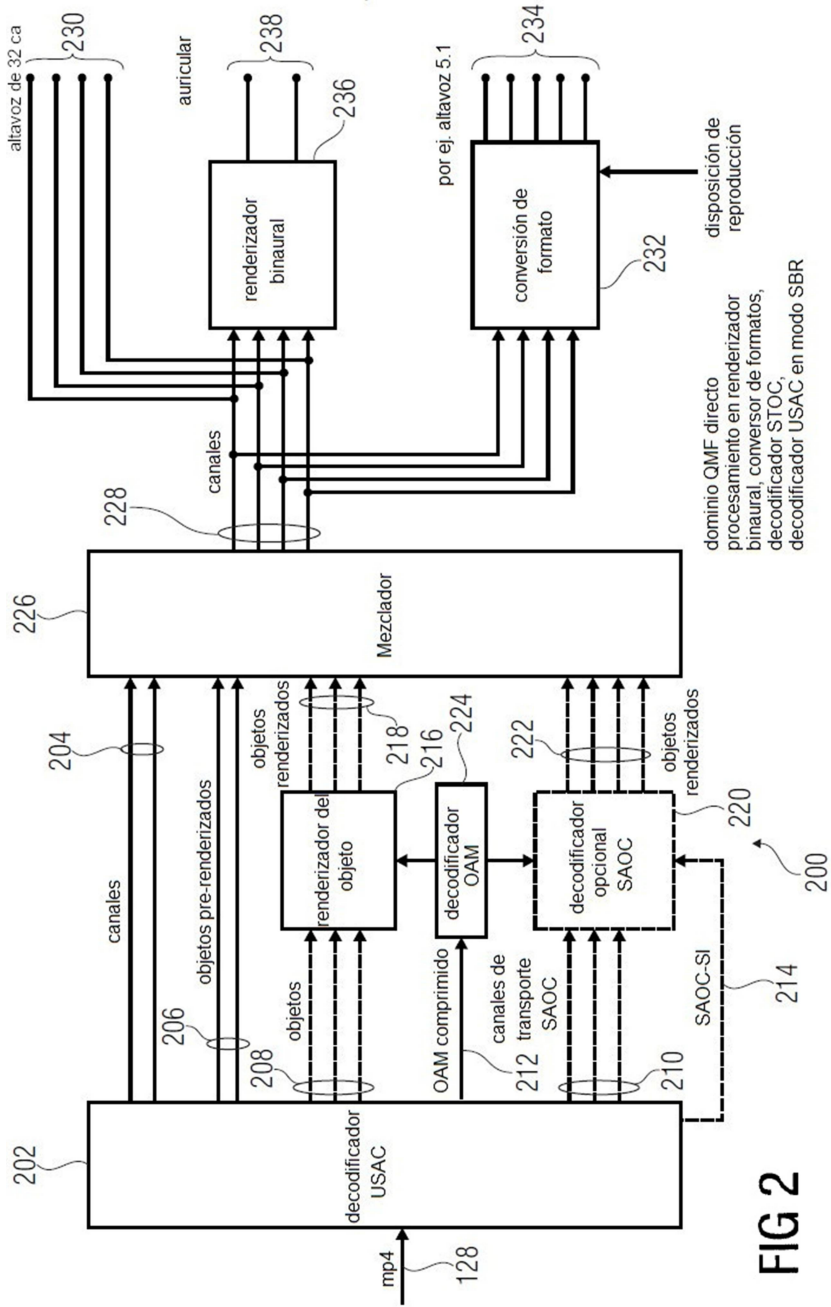


FIG 2

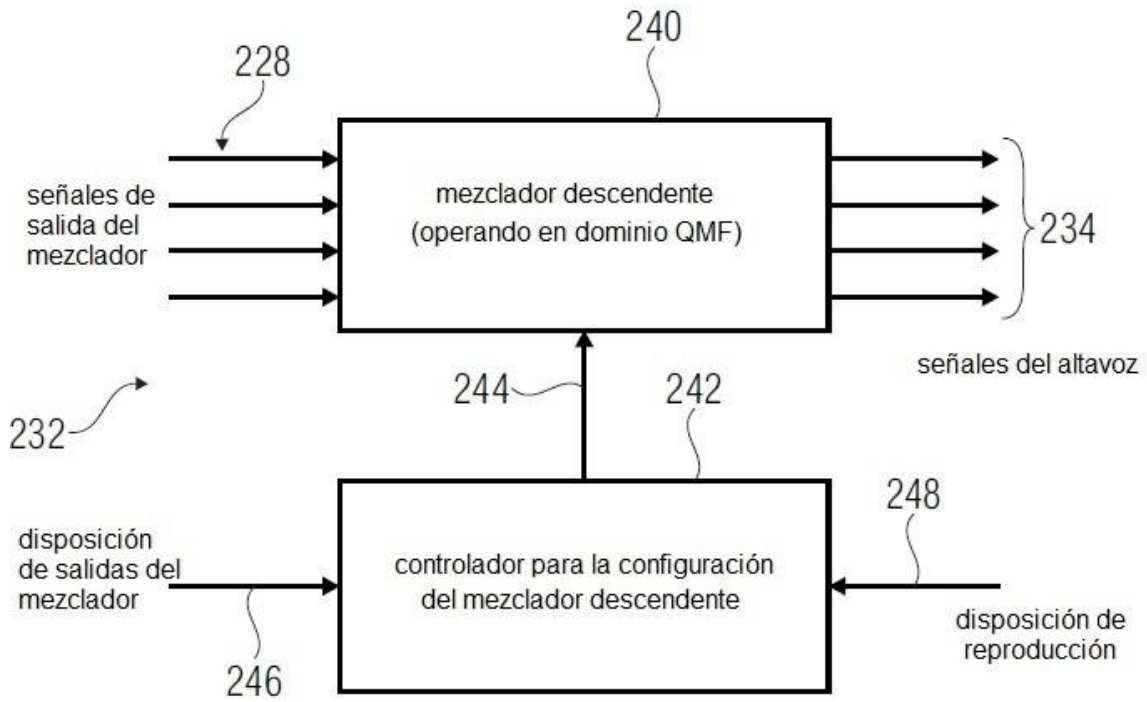


FIG 3

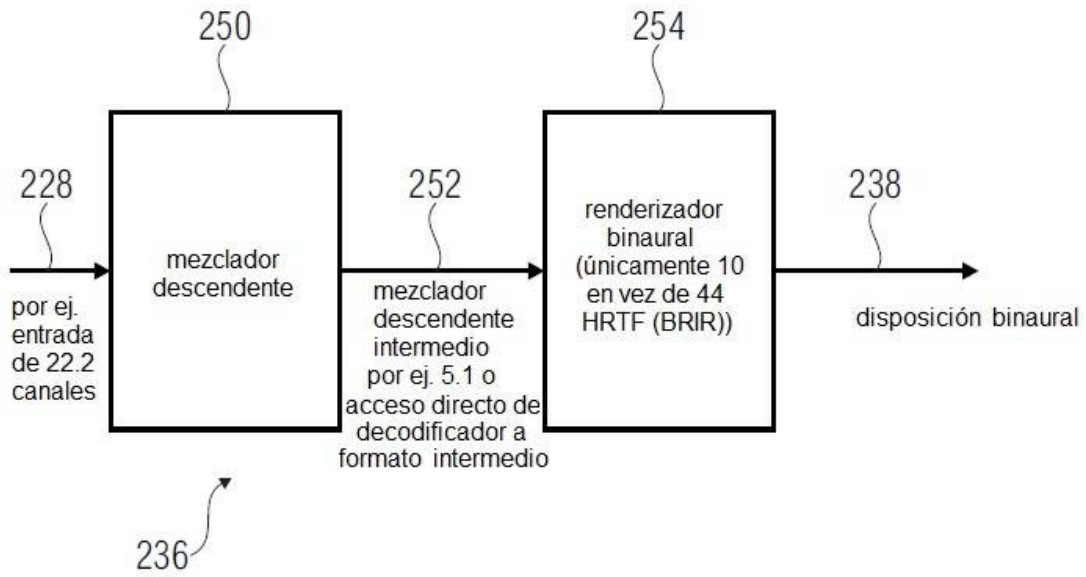


FIG 4

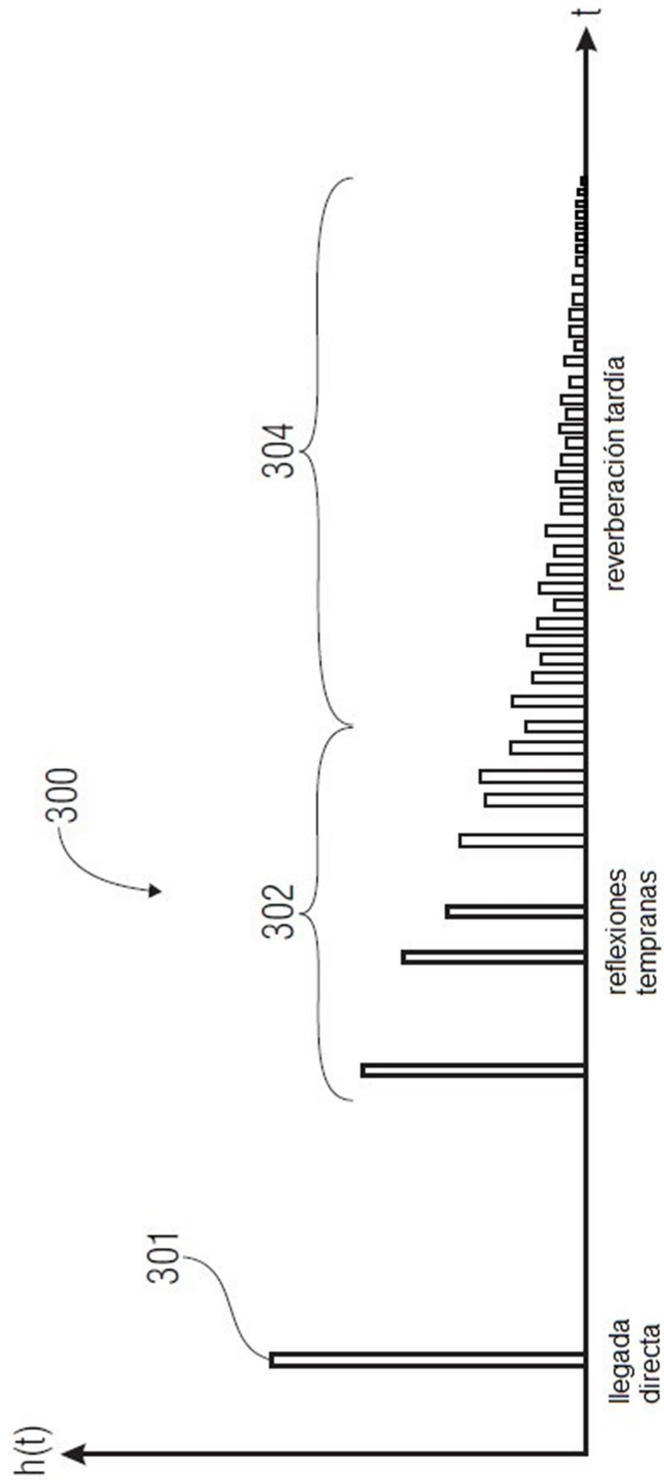


FIG 5

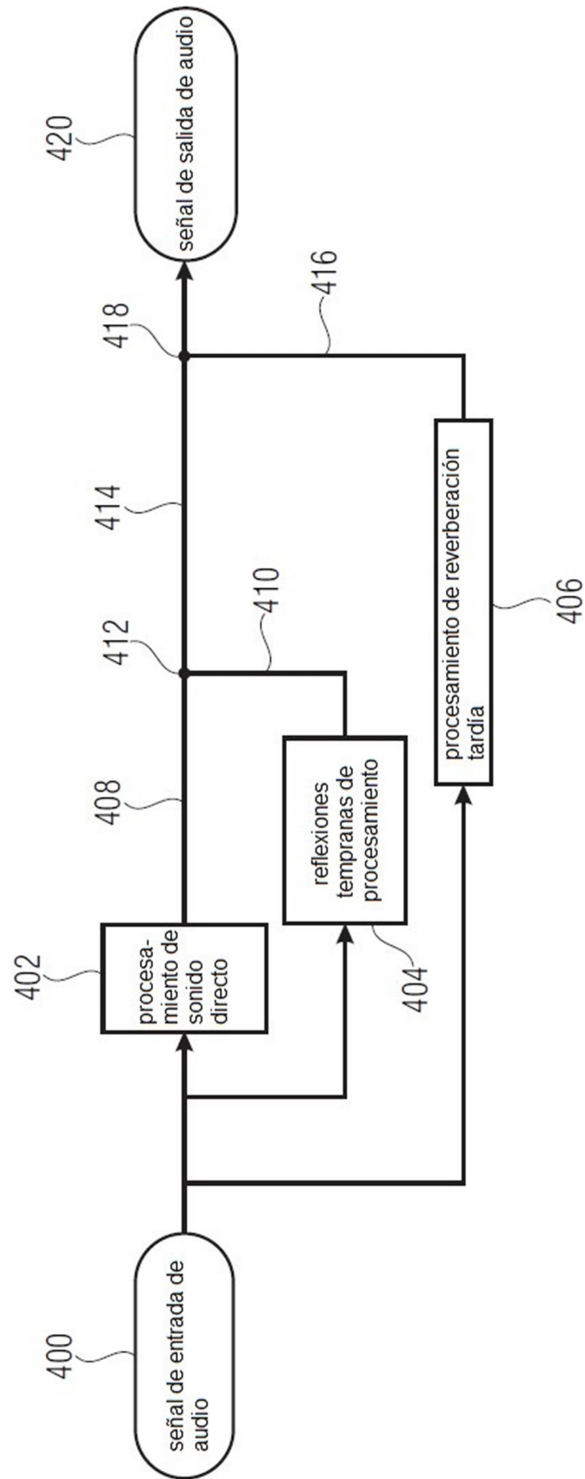


FIG 6A

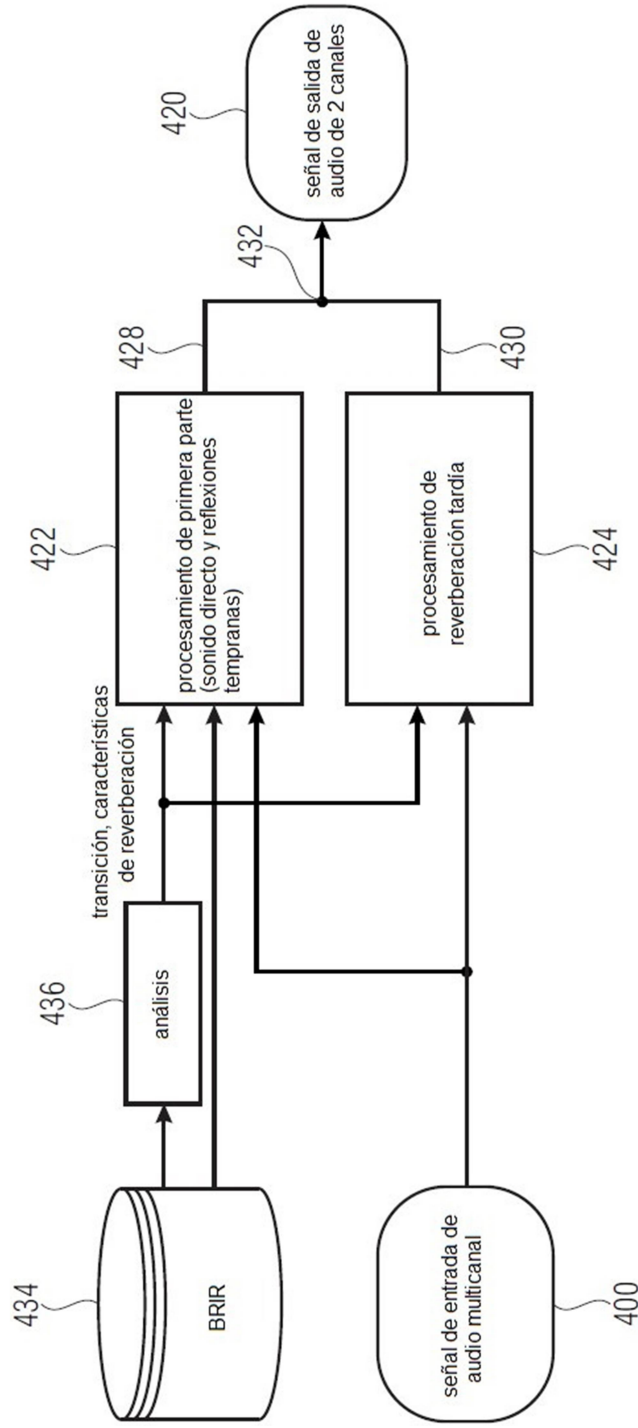


FIG 6B

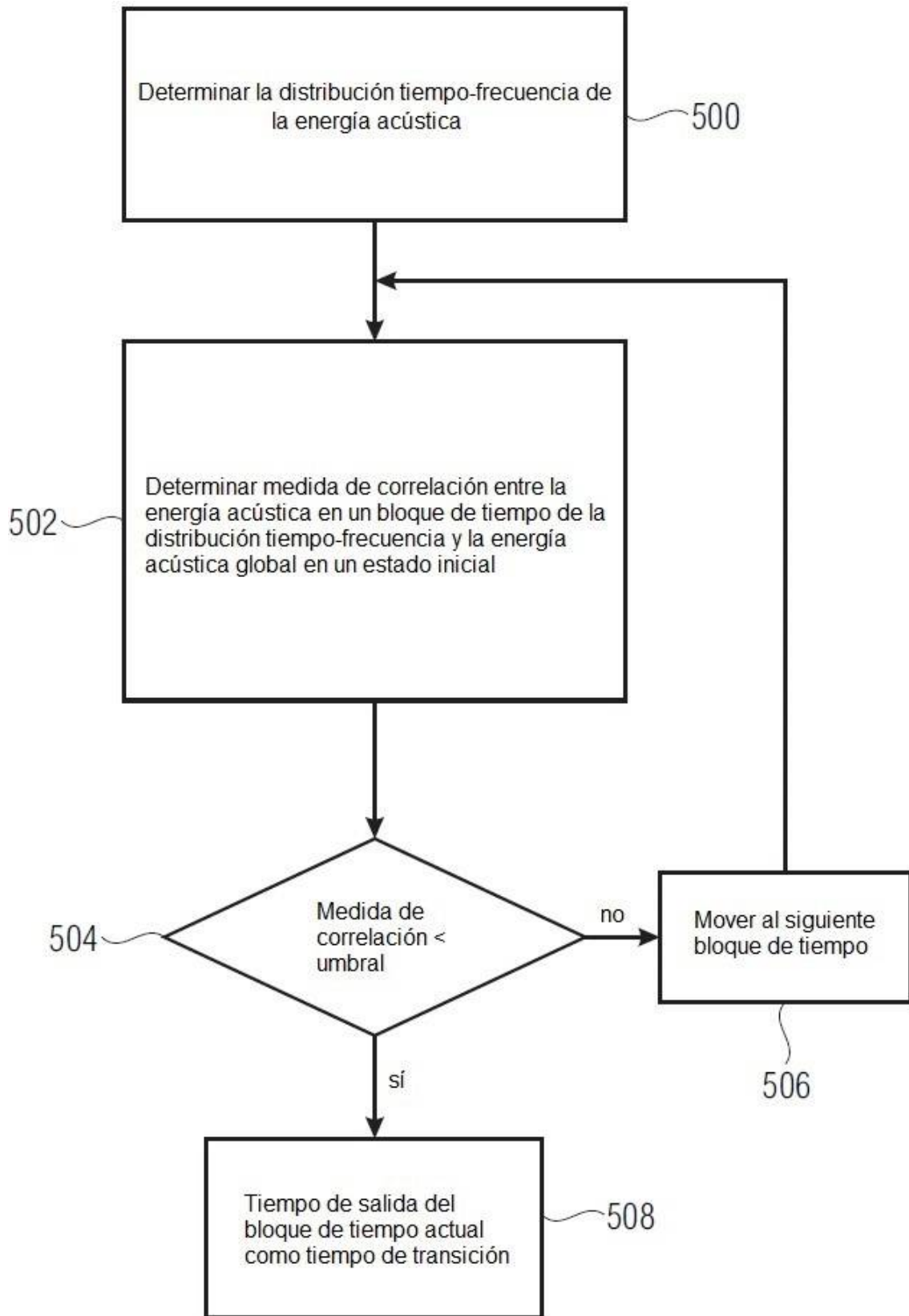


FIG 7

alivio de decaimiento de la energía

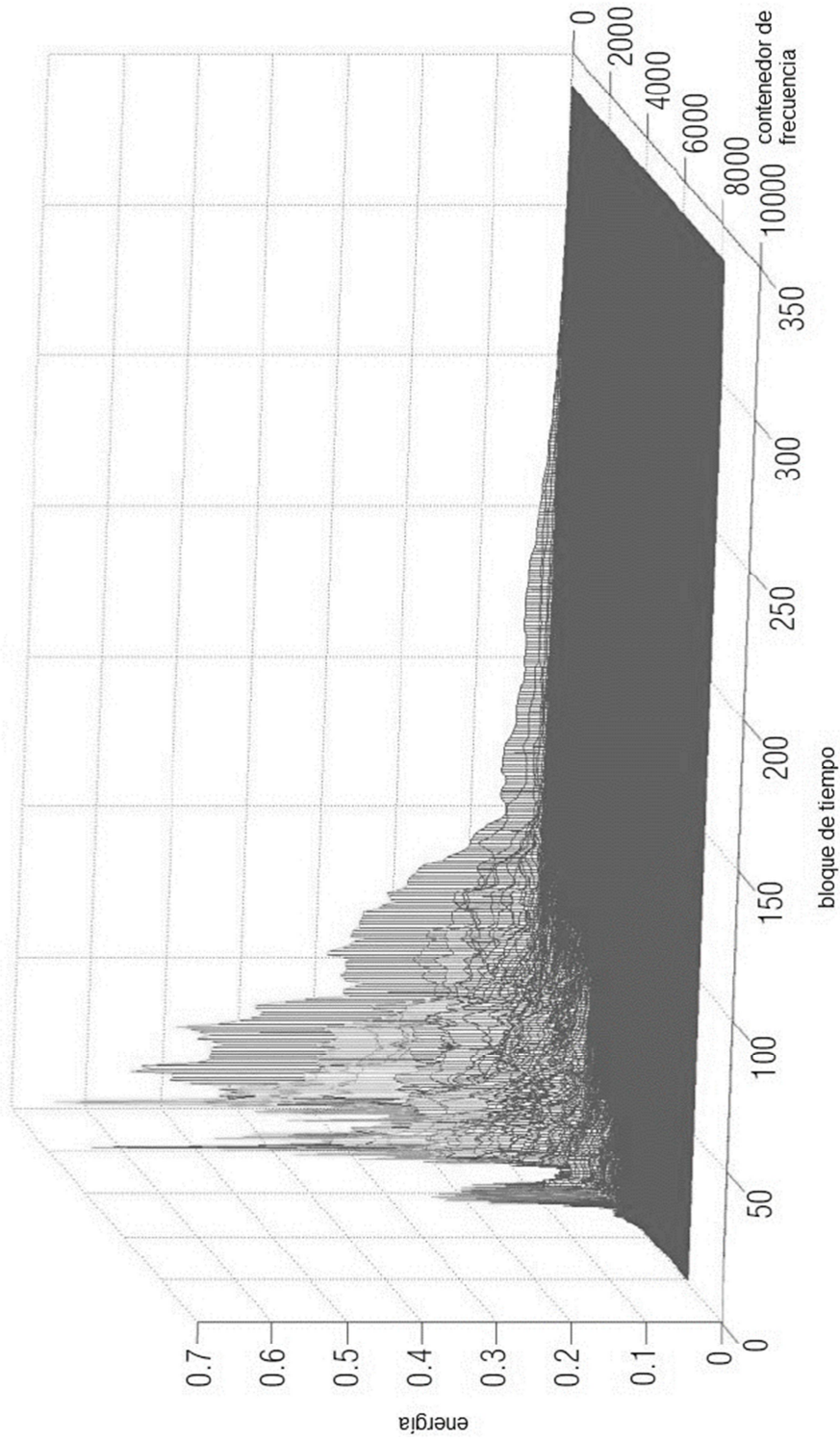


FIG 8

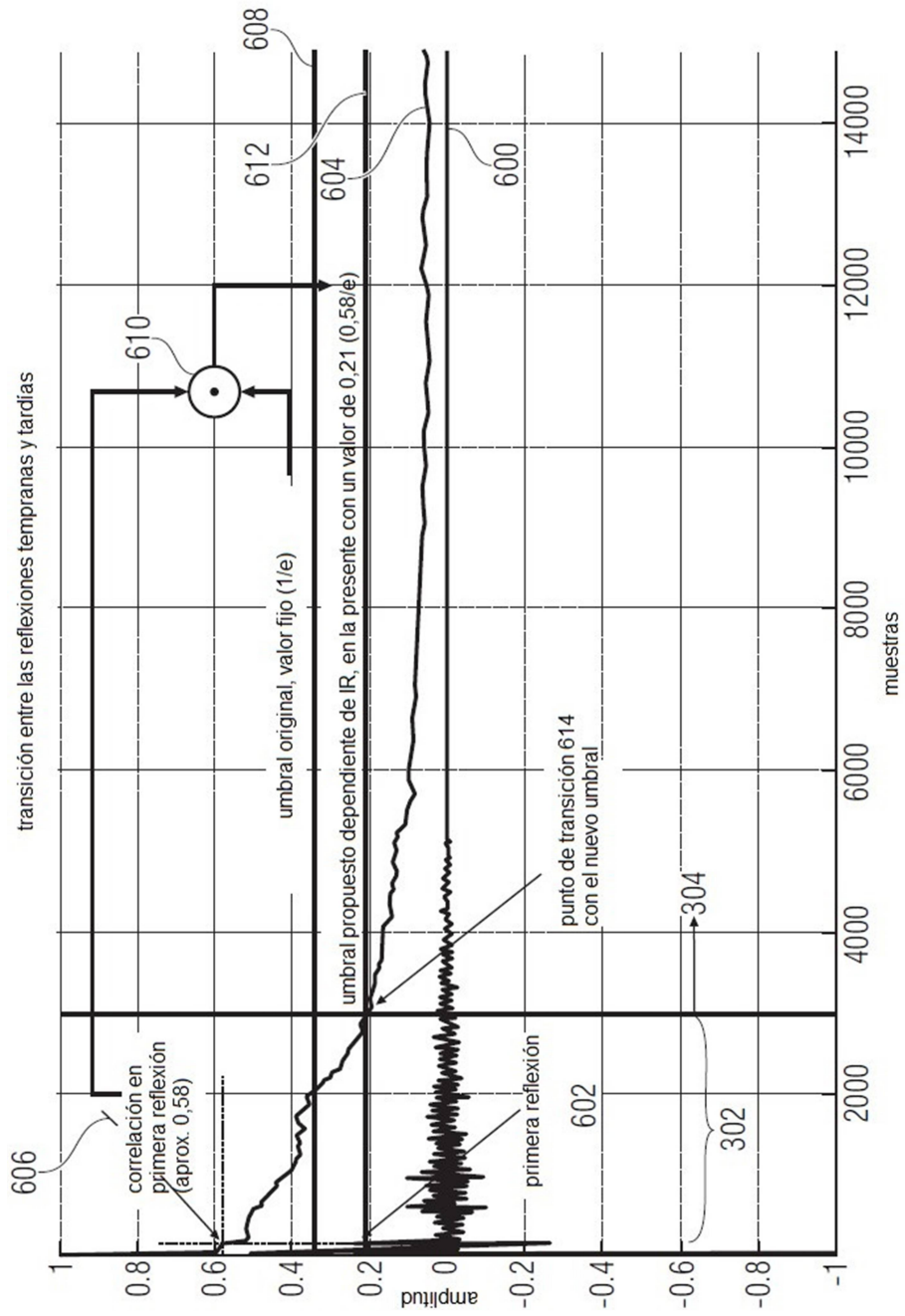


FIG 9

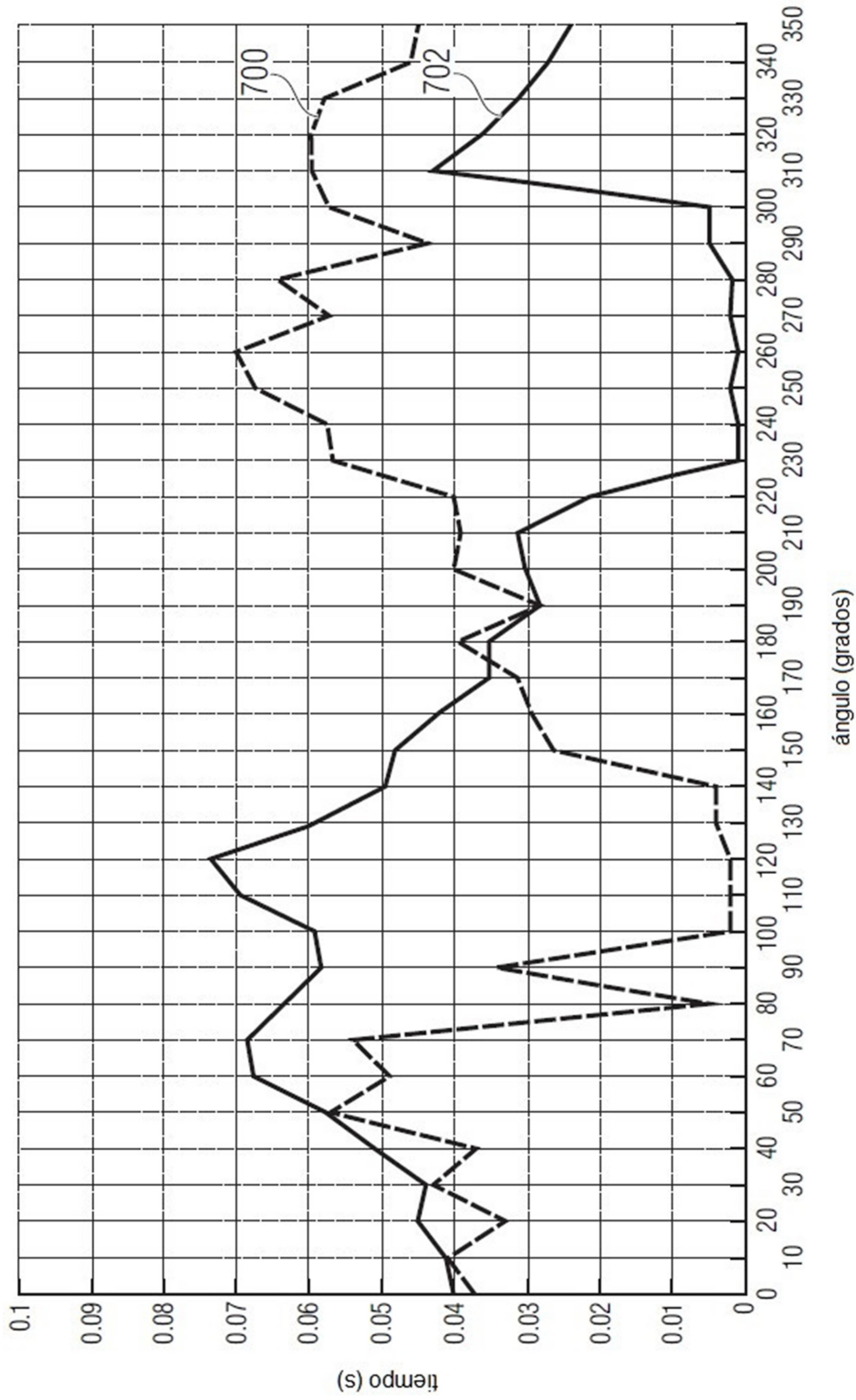


FIG 10

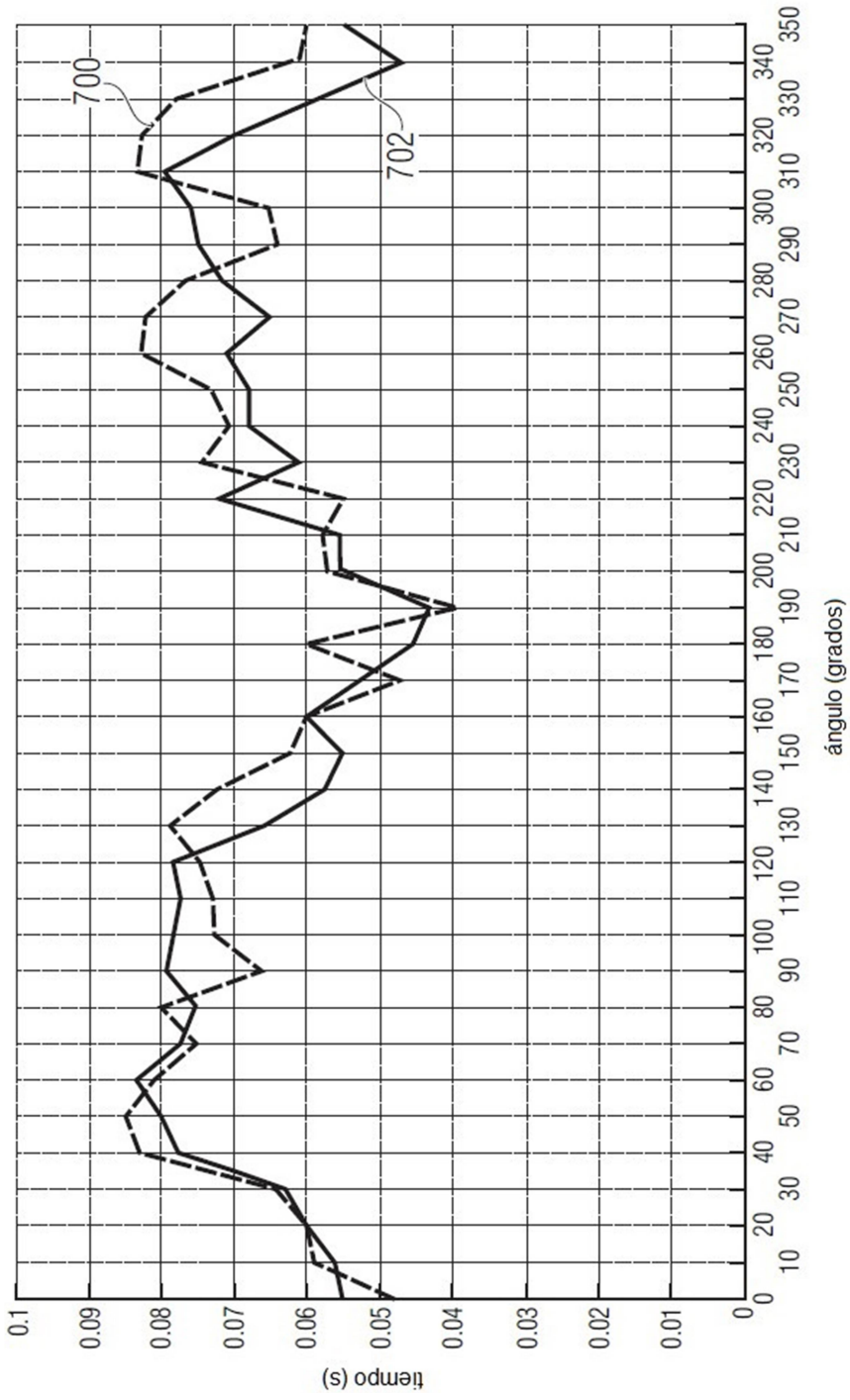


FIG 11

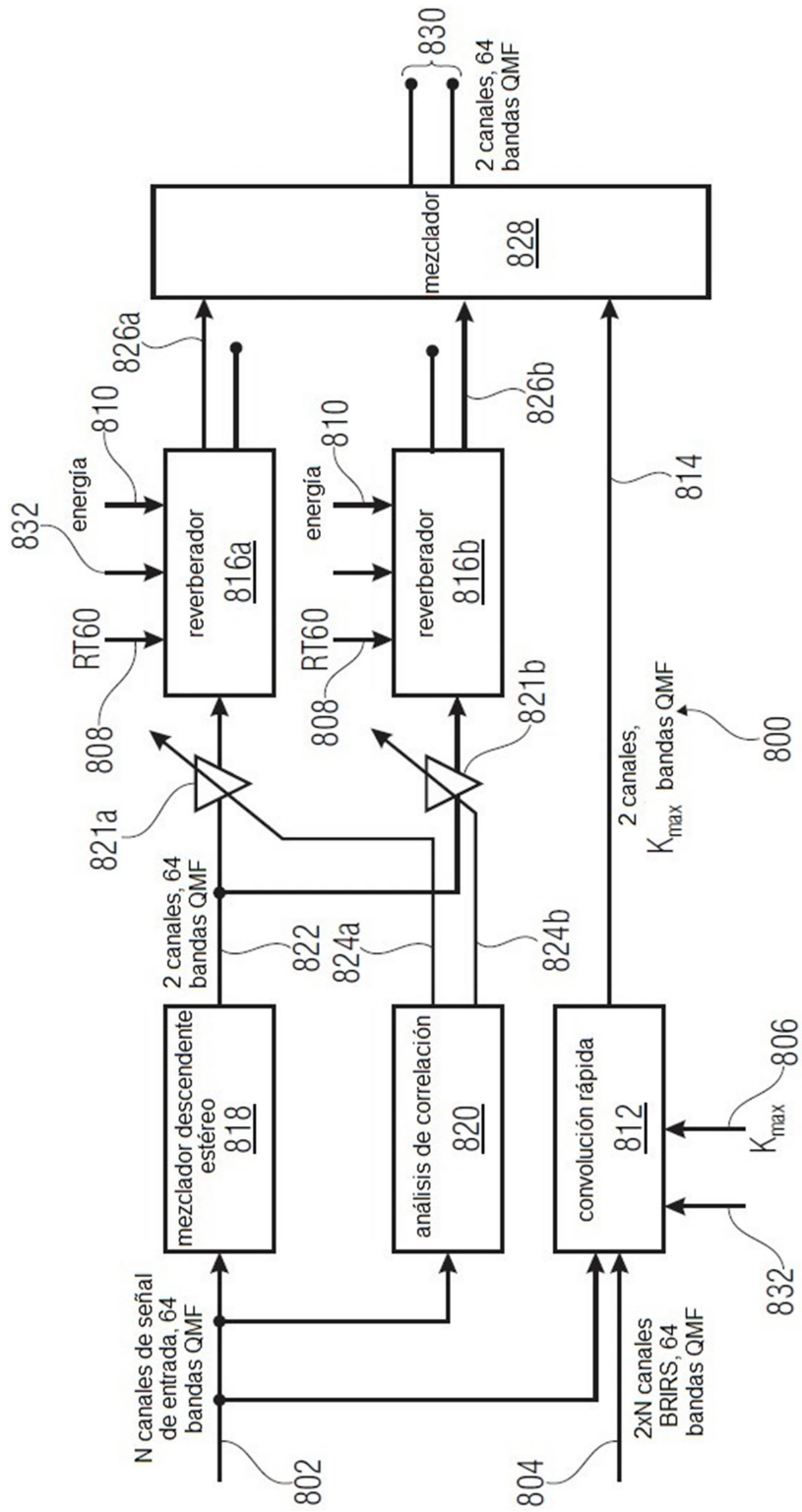


FIG 12

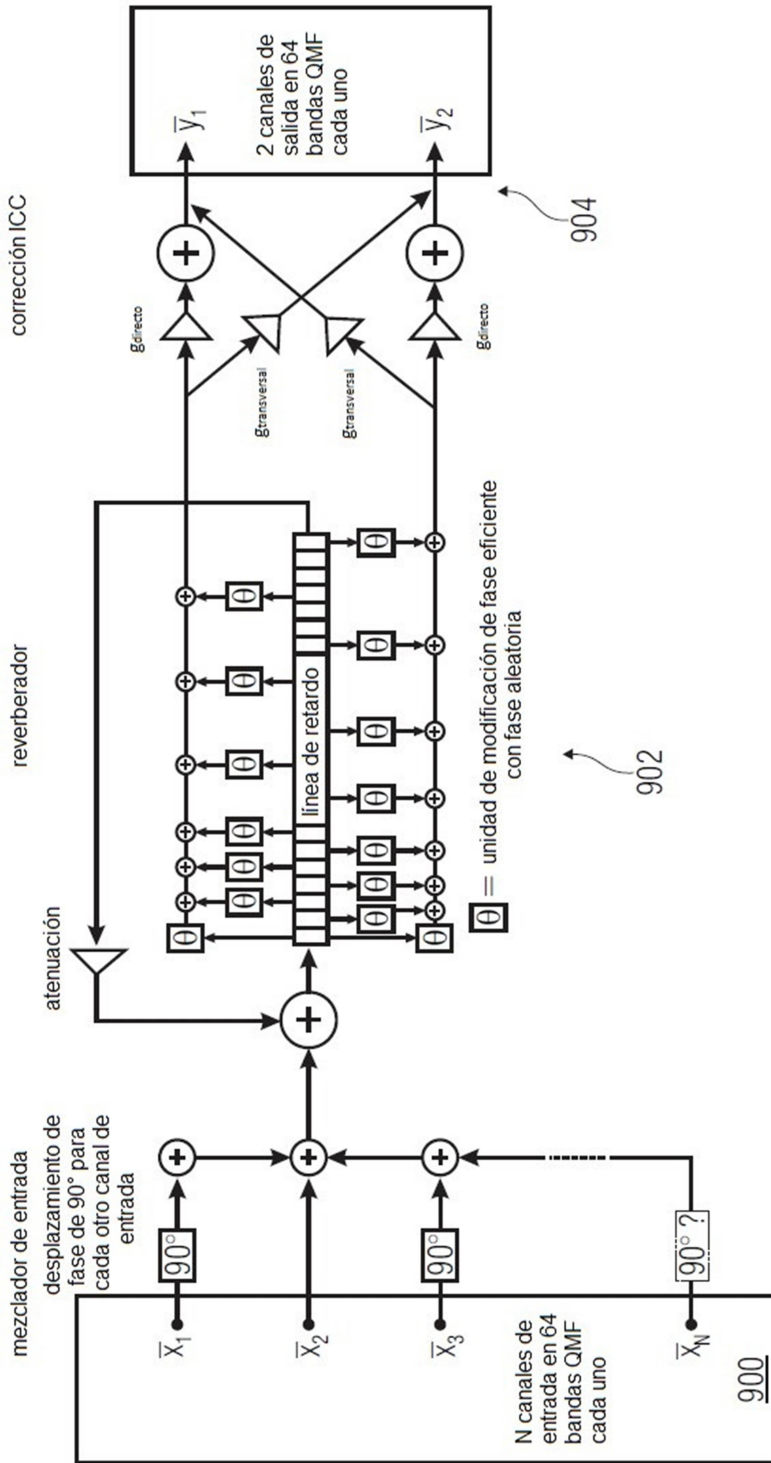


FIG 13