

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 754 245**

51 Int. Cl.:

H04S 1/00 (2006.01)

H04S 7/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.07.2014 PCT/FR2014/051728**

87 Fecha y número de publicación internacional: **29.01.2015 WO15011359**

96 Fecha de presentación y número de la solicitud europea: **04.07.2014 E 14748239 (2)**

97 Fecha y número de publicación de la concesión europea: **11.09.2019 EP 3025514**

54 Título: **Espacialización sonora con efecto de sala**

30 Prioridad:

24.07.2013 FR 1357299

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

16.04.2020

73 Titular/es:

**ORANGE (100.0%)
78, rue Olivier de Serres
75015 Paris, FR**

72 Inventor/es:

**PALLONE, GRÉGORY y
EMERIT, MARC**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 754 245 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

Espacialización sonora con efecto de sala

5 La invención se refiere al tratamiento de datos sonoros y, de manera más particular, a la espacialización (denominada "renderizado 3D") de señales audio.

10 Una operación de este tipo se ejecuta, por ejemplo, durante la decodificación de una señal audio 3D codificada, representada en un número de canales determinado, hacia un número de canales diferentes, dos, por ejemplo, para permitir la restitución de los efectos 3D audio en un casco de auriculares.

15 La invención se refiere igualmente a la transmisión y a la restitución de señales audio multicanales y a su conversión hacia un dispositivo de restitución, transductor, impuesto por el equipo de un usuario. Este es, por ejemplo, el caso para la restitución de una escena sonora 5.1 mediante un casco de auriculares audio o mediante un par de altavoces.

La invención se refiere igualmente a la restitución, en el marco de un juego o grabación en video, por ejemplo, de una o varias muestras sonoras almacenadas en ficheros, con vistas a su espacialización.

20 En el caso de una fuente monofónica estática, la binauralización se basa en el filtrado de la señal monofónica por la función de transferencia entre la posición deseada de la fuente y cada uno de los dos oídos. La señal binaural (dos canales) obtenida puede alimentar entonces un casco audio y proporcionar al oyente la sensación de una fuente en la posición simulada. De este modo, el término "binaural" tiene por objeto la restitución de una señal sonora con efectos de espacialización.

25 Cada una de las funciones de transferencia que simulan diferentes posiciones pueden medirse en cámara sorda, lo que resulta de este modo en un conjunto de HRTF (por sus siglas de "Head Related Transfer Functions" o "Funciones de transferencias relativas a la cabeza") en las que ningún efecto de sala está presente.

30 Estas funciones de transferencia pueden medirse igualmente en una sala "clásica", lo que resulta de este modo en un conjunto de BRIR (por sus siglas de "Binaural Room Impulse Response" o "Respuesta de impulso binaural de sala") en las que el efecto de sala, o reverberación, está presente. El conjunto de las BRIR corresponde por tanto a un conjunto de funciones de transferencia entre una posición dada y los oídos de un oyente (real o cabeza artificial) situado en una sala.

35 La técnica habitual de medición de BRIR consiste en enviar sucesivamente a cada uno de los altavoces reales, posicionados alrededor de una cabeza (real o artificial) equipada con micrófonos en los oídos, una señal de prueba (por ejemplo, una señal de barrido "sweep", una secuencia binaria pseudo-aleatoria o un ruido blanco). Esta señal de prueba permite, durante un tratamiento fuera de tiempo real, la reconstitución (generalmente por deconvolución) de la respuesta de impulso entre la posición del altavoz y cada uno de los dos oídos.

40 La diferencia entre un conjunto de HRTF y de BRIR se sitúa principalmente en la longitud de la respuesta de impulso, del orden del milisegundo para las HRTF, al orden del segundo para las BRIR.

45 Al basarse el filtrado en la convolución entre la señal monofónica y la respuesta de impulso, la complejidad para efectuar una binauralización con las BRIR (que contiene un efecto de sala) es significativamente más elevada que con las HRTF.

50 Mediante esta técnica, es posible simular en el casco o en un número limitado de altavoces una escucha de un contenido multicanal (L canales) generado por L altavoces en una sala. En efecto, basta con considerar cada uno de los altavoces L como una fuente virtual posicionada idóneamente con respecto al oyente, medir en la sala a simular las funciones de transferencia (para los oídos izquierdo y derecho) de cada uno de estos L altavoces, después aplicar a cada una de las L señales audio (que supuestamente alimentan los L altavoces reales) los filtros BRIR correspondientes a los altavoces. Las señales que alimentan cada uno de los oídos se suman para proporcionar una señal binaural que alimenta un casco audio.

55 Se denomina $I(1)$ (siendo $1=[1, L]$) a la señal de entrada que supuestamente alimenta los L altavoces. Se denomina $BRIR^{i/d}(1)$, a las BRIR de cada uno de los altavoces para cada uno de los dos oídos y se denomina $O^{i/d}$ a la señal binaural de salida. La binauralización de la señal multicanal se escribe por tanto:

60
$$O^i = \sum_{l=1}^L I(l) * BRIR^i(l)$$

$$O^d = \sum_{l=1}^L I(l) * BRIR^d(l)$$

Dónde * representa al operador de convolución.

65 A continuación, el índice 1 como $l \in [1, L]$ hace referencia a uno de los L altavoces. Se tiene una BRIR para una señal 1.

De este modo, con referencia a la figura 1, dos convoluciones (una para cada oído) están presentes para cada altavoz (etapas S11 a S1L).

5 Para L altavoces, la binauralización necesita por tanto 2.L convoluciones. Se puede calcular la complejidad C_{conv} en el caso de una implementación rápida por bloque. Una implementación rápida por bloque se da, por ejemplo, mediante una transformada de Fourier rápida (FFT de "Fast Fourier Transform"). El documento "Submission and Evaluation Procedures for 3D Audio" (MPEG 3D Audio) precisa una formula posible para el cálculo de C_{conv} :

10
$$C_{conv} = (L + 2) \cdot (nBloques) \cdot (6 \cdot \log_2(2Fs/nBloques))$$

15 En esta ecuación, L representa el número de FFT para transformar en frecuencia las señales de entrada (1FFT por señal de entrada), el 2 representa el número de FFT inversa para obtener la señal binaural temporal (2 FFT inversa para los dos canales del binaural), el 6 indica un coeficiente de complejidad por FFT, el segundo 2 indica un relleno de ceros necesario para evitar los problemas debidos a la convolución circular, Fs indica el tamaño de cada una de las BRIR, y nBloques representa el hecho de usar un tratamiento por bloque, más realista en una perspectiva en la que la latencia no debe ser excesivamente elevada, y \cdot representa la multiplicación.

20 De este modo, para un uso típico con nBloques=10, Fs=48000, L=22, la complejidad por muestra de señal multicanal para una convolución directa basada en una FFT es de $C_{conv} = 19049$ multiplicaciones-adiciones.

25 Esta complejidad es demasiada elevada para una implementación realista hoy en día en procesadores corrientes (móviles por ejemplo), por lo tanto es necesario disminuir esta complejidad sin degradar demasiado el renderizado de la binauralización.

Para que la espacialización sea de buena calidad, se debe aplicar la totalidad de la señal temporal de las BRIR.

30 El documento US 2006/045294 A1 divulga un sistema de espacialización sonora de señales multicanales para suministrar al menos dos canales de salida, y que incluye el particionamiento de las respuestas de impulso en dos partes, conteniendo la primera la llegada directa y las primeras reflexiones, y conteniendo la segunda la reverberación, sumándose y ponderándose las segundas partes de al menos dos respuestas de impulso para obtener una función de transferencia única que representa la reverberación que hay que aplicar a las señales de entrada, con el fin de reducir la complejidad de las operaciones de convolución. El documento US 2006/045294 A1 también divulga un filtrado por bloques de FFT.

35 El documento STEWARD REBECCA ET AL: "Generating a Spatial Average Reverberation Tail Across Multiple Impulse Responses", CONFERENCE: 35TH INTERNATIONAL CONFERENCE: AUDIO FOR GAMES; FEBRUARY 2009, AES, NEW YORK, EE. UU., del 1 de febrero de 2009 (2009-02-01) divulga un particionamiento similar de respuestas de impulso, así como la promediación de las partes de respuestas de impulso que representan la reverberación, para reducir la redundancia en las bases de datos de respuestas de impulso.

40 El documento GARDNER W G: "EFFICIENT CONVOLUTION WITHOUT INPUT-OUTPUT DELAY", JOURNAL OF THE AUDIO ENGINEERING SOCIETY, AUDIO ENGINEERING SOCIETY, NEW YORK, NY, EE. UU., VOL. 43, N.º 3, del 1 de marzo de 1995 (1995-03-01), páginas 127-135 describe una implementación eficaz para la convolución de respuestas de impulso, en concreto en el contexto de las simulaciones de reverberación, en forma de bloques FFT. Este artículo sugiere reutilizar las transformadas de las señales de entrada lo antes posible, para reducir los costes de cálculo, y muestra que un particionamiento de una respuesta de impulso en sub-bloques, cuya longitud es múltiplo los unos de los otros, permite reducir la complejidad de cálculo.

50 La presente invención viene a mejorar la situación.

Tiene por objeto disminuir en gran medida la complejidad de la binauralización de una señal multicanal con efecto de sala mientras se conserva lo mejor posible la calidad audio.

55 La presente invención propone para tal efecto un procedimiento de espacialización sonora como el que se define en las reivindicaciones 1 a 10.

La invención tiene también por objeto un programa informático que incluye unas instrucciones para la implementación del procedimiento.

60 La invención puede implementarse mediante un dispositivo de espacialización sonora como el que se define en la reivindicación 12.

La invención puede implementarse igualmente en un módulo de descodificación de señales sonoras, como señales de entrada, que incluye el dispositivo de espacialización anterior.

5 Otras ventajas y características de la invención se pondrán de manifiesto tras la lectura de la descripción detallada a continuación, de unos ejemplos de realización de la invención y del examen de los dibujos en los que:

- la figura 1 ilustra un procedimiento de espacialización de la técnica anterior,
- la figura 2 ilustra esquemáticamente las etapas de un procedimiento en el sentido de la invención, en un ejemplo de realización,
- 10 - la figura 3 representa una respuesta de impulso binaural de sala BRIR,
- la figura 4 ilustra esquemáticamente las etapas de un procedimiento en el sentido de la invención, en un ejemplo de realización,
- la figura 5 ilustra esquemáticamente las etapas de un procedimiento en el sentido de la invención, en un ejemplo de realización,
- 15 - la figura 6 representa esquemáticamente un dispositivo que incluye unos medios de implementación del procedimiento en el sentido de la invención.

Se hace referencia a la figura 6 para ilustrar en primer lugar un posible contexto de implementación de la presente invención en un dispositivo de tipo terminal conectado TER (por ejemplo, un teléfono, smartphone u otro, o incluso una tableta conectada, un ordenador conectado u otros). Un dispositivo de este tipo TER incluye unos medios de recepción (normalmente, una antena) de señales audio Xc codificadas en compresión, un dispositivo de descodificación DESCOD que suministra señales descodificadas X listas para ser tratadas por un dispositivo de espacialización antes de la restitución de las señales audio (por ejemplo, en binaural en un casco con auriculares CAS). Por supuesto, en ciertos casos, puede ser ventajoso guardar las señales parcialmente descodificadas (por ejemplo, en el ámbito de las sub-bandas) si el tratamiento de espacialización se efectúa en el mismo ámbito (tratamiento frecuencial en el ámbito de las sub-bandas, por ejemplo).

También con referencia a la figura 6, el dispositivo de espacialización se presenta con una combinación de elementos:

- 30 - hardware que normalmente incluye uno o varios circuitos CIR que cooperan con una memoria de trabajo MEM y un procesador PROC,
- y software, respecto al cual, las figuras 2 y 4 son unos ejemplos de organigramas que ilustran el algoritmo general.

35 En este caso, la cooperación entre los elementos de hardware y software produce un efecto técnico que conlleva, en concreto, un ahorro de complejidad de espacialización para sustancialmente un mismo renderizado audio (misma sensación para un oyente), como se verá más adelante.

40 Se hace referencia ahora a la figura 2 para describir un tratamiento en el sentido de la invención, implementado de este modo mediante unos medios informáticos.

En una primera etapa S21, se efectúa una preparación de los datos. Esta preparación es opcional, las señales pueden tratarse según las etapas S22 y siguientes sin este pretratamiento.

45 En particular, esta preparación consiste en efectuar un truncamiento de cada BRIR para ignorar las muestras inaudibles al inicio y al final de la respuesta de impulso.

Esta preparación, para el truncamiento al inicio de la respuesta de impulso TRUNCADO S, en la etapa S211, consiste en determinar un instante de inicio de ondas sonoras directas y puede implementarse mediante las siguientes etapas:

- 50 - Se calcula una suma acumulada de las energías de cada uno de los filtros BRIR(1). Normalmente, esta energía se calcula mediante una suma al cuadrado de las amplitudes de las muestras 1 a j, estando j comprendida en [1 ; J] siendo J el número de muestras de un filtro BRIR.
- 55 - Se calcula el valor de energía del filtro de energía máximo valMáx (entre los filtros relativos al oído izquierdo y al oído derecho).
- Para cada uno de los altavoces 1, se calcula el índice para el que la energía de cada uno de los filtros BRIR (1) supera un determinado umbral en dB calculado con respecto a valMáx (por ejemplo, valMáx-50dB).
- 60 - El índice de truncamiento iT seleccionado para todas las BRIR es el índice mínimo de entre todos los índices de las BRIR y se considera como el instante de inicio de las ondas sonoras directas.

65 El índice iT obtenido corresponde por tanto al número de muestras que hay que ignorar para cada una de las BRIR. Un truncamiento brutal al inicio de la respuesta de impulso con una ventana rectangular puede llevar a artefactos

audibles si se aplica en una parte demasiada energética. Por lo tanto, puede ser preferible aplicar una ventana de fundido de entrada adaptada, sin embargo, si se han tomado unas precauciones en el umbral seleccionado, esta ventana es inútil por ser inaudible (solo se corta señal inaudible).

5 El sincronismo entre BRIR permite la aplicación de un plazo constante para todas las BRIR en aras de una mayor simplicidad de implementación, incluso si una optimización de la complejidad es posible.

10 El truncamiento de cada BRIR para ignorar las muestras inaudibles al final de la respuesta de impulso TRUNCADO E, en la etapa S212, puede realizarse a partir de etapas similares a las que se han descrito anteriormente, adaptadas para convenir al final de la respuesta de impulso. Un truncamiento brutal al final de la respuesta de impulso con una ventana rectangular puede llevar a artefactos audibles en señales de impulso en las que la cola de reverberación puede revelarse audible. De este modo, en un modo de realización, se aplica una ventana de fundido de salida adaptada.

15 En la etapa 22, se efectúa un aislamiento sincronizado AISL A/B. Este aislamiento sincronizado consiste en separar, para cada BRIR, la parte de "sonido directo" y "primeras reflexiones" (o Directo, denotado A) y de la parte de "sonido difuso", denotado B). En efecto, el tratamiento que hay que efectuar en la parte de "sonido difuso" puede ser ventajosamente diferente del que hay que efectuar en la parte de "sonido directo", en la medida en que es preferible tener una mejor calidad de tratamiento en la parte de "sonido directo" que en la parte de "sonido difuso". Esto hace posible una optimización de la relación calidad/complejidad.

20 En particular, para realizar el aislamiento sincronizado, se determina un índice "iDD" de muestra única y común para todas las BRIR (de ahí el término "sincronizado") a partir del cual se considera que el resto de la respuesta de impulso corresponde a un campo difuso. Se particionan por tanto las respuestas de impulso BRIR(1) en dos: A(1) y B(1), en las que la concatenación de las dos corresponde a BRIR(1).

La figura 3 muestra el índice de particionamiento iDD en la muestra 2000. La parte a la izquierda de este índice iDD corresponde a la parte A. La parte a la derecha de este índice iDD corresponde a la parte B.

30 En un modo de realización, estas dos partes están aisladas, sin ventana, con el fin de someterse a tratamientos diferentes. En una variante, se aplica una ventana entre las partes A(1) y B(1).

35 El índice iDD puede ser específico de la sala para la que se han determinado las BRIR. El cálculo de este índice puede depender por tanto de la envolvente espectral, de la correlación de las BRIR o incluso del ecograma de estas BRIR. Por ejemplo, se puede determinar el iDD con una fórmula de tipo $iDD = \sqrt{V_{sala}}$ siendo V_{sala} el volumen de la sala de medición.

40 En un modo de realización, iDD es un valor fijo, normalmente 2000. En una variante, iDD varía, ventajosamente de manera dinámica, en función del entorno a partir del cual se capturan las señales de entrada.

La señal de salida para los oídos izquierdo (i) y derecho (d), representada por $O^{i/d}$, se escribe, por tanto:

$$O^{i/d} = \sum_{l=1}^L I(l) * BRIR^{i/d}(l) = O_A^{i/d} + z^{-iDD} * O_B^{i/d}$$

$$\sum_{l=1}^L I(l) * A^{i/d}(l) + z^{-iDD} * \sum_{l=1}^L I(l) * B^{i/d}(l)$$

45 donde z^{-iDD} corresponde al plazo de iDD muestras.

La aplicación de este plazo a las señales se efectúa almacenando los valores calculados para $\sum_{l=1}^L I(l) * B^{i/d}(l)$ en una memoria temporal (por ejemplo, en un búfer) y restituyéndolos en el momento deseado.

50 En un modo de realización, los índices de muestras seleccionados para A y B pueden tomar en cuenta igualmente las longitudes de tramas en el caso de integración en un codificador audio. En efecto, los tamaños típicos de tramas de 1024 muestras pueden llevar a una selección tal que A haga 1024 y B haga 2048, asegurándose de que B sea en efecto una zona de campo difuso para todas las BRIR.

55 En particular, puede ser interesante que el tamaño de B sea un múltiplo del tamaño de A porque si el filtrado se implementa por bloques de FFT, entonces el cálculo de una FFT puede reutilizarse para B.

60 Un campo difuso se caracteriza por el hecho de que es estadísticamente idéntico en todos los puntos de la sala. De este modo, su respuesta en frecuencia varía poco en función del altavoz que hay que simular. La presente invención aprovecha esta característica con el fin de sustituir todos los filtros Difusos D(l) de todas las BRIR por un solo y único filtro "medio" B_{medio} con el fin de disminuir en gran medida la complejidad debida a las múltiples convoluciones. Para ello, se puede modificar la parte de campo difuso B en la etapa S23B, siempre con referencia a la figura 2.

En la etapa S23B1, se calcula el valor del filtro medio B_{medio} . Primero, es extremadamente poco frecuente que el sistema completo esté calibrado idóneamente, se puede aplicar por tanto una ganancia de ponderación que se aplicará en la señal de entrada con el fin de efectuar una única convolución por oído para la parte de campo difuso. Se descomponen por tanto las BRIR en filtros normalizados en energía, y se aplica la ganancia de normalización $\sqrt{E_{B^{i/d}(l)}}$ en la señal de entrada:

$$O_B^{i/d} = \sum_{l=1}^L [I(l) * B^{i/d}(l)] = \sum_{l=1}^L [I(l) * \sqrt{E_{B^{i/d}(l)}} \cdot B_{norm^{i/d}}(l)]$$

$$= \sum_{l=1}^L [(\frac{I(l)}{\sqrt{E_{B^{i/d}(l)}}}) * B_{norm^{i/d}}(l)]$$

siendo $B_{norm^{i/d}}(l) = \frac{B^{i/d}(l)}{\sqrt{E_{B^{i/d}(l)}}}$ donde $E_{B^{i/d}(l)}$ representa la energía de $B^{i/d}(l)$.

A continuación, se aproxima $B_{norm^{i/d}(l)}$ mediante un solo y único filtro medio $B_{medio^{i/d}}$ que ya no es función del altavoz 1, pero que es posible normalizar igualmente en energía:

$$O_B^{i/d} \approx \hat{O}_B^{i/d} = \sum_{l=1}^L [(\frac{I(l)}{\sqrt{E_{B^{i/d}(l)}}}) * (\frac{B_{medio^{i/d}}}{\sqrt{E_{B_{medio^{i/d}}}}})]$$

siendo $B_{medio^{i/d}} = \frac{1}{L} \sum_{l=1}^L [B_{norm^{i/d}}(l)]$

En un modo de realización, este filtro medio puede obtenerse promediando las muestras temporales. En una variante, se puede obtener mediante cualquier tipo de promediación como, por ejemplo, una promediación de las densidades espectrales de potencia.

En un modo de realización, la energía del filtro medio $E_{B_{medio^{i/d}}}$ puede medirse directamente a partir del filtro construido $B_{medio^{i/d}}$. En una variante, se puede estimar igualmente tomando en cuenta la hipótesis de que los filtros $B_{norm^{i/d}}(l)$ están descorrelacionados. En efecto, en ese caso, como se suman señales de energía unitaria, se tiene:

$$E_{B_{medio^{i/d}}} = \sum_{l=1}^L [B_{norm^{i/d}}(l)]^2 = \frac{1}{L^2} \cdot (L \cdot E_{B_{norm^{i/d}}}) = \frac{1}{L}$$

La energía puede calcularse sobre el conjunto de las muestras correspondientes a la parte de campo difuso.

En la etapa S23B2, se calcula el valor del factor de ponderación $W^{i/d}(l)$. Se calcula un único factor de ponderación que hay que aplicar a la señal de entrada, tomando en cuenta las normalizaciones de los filtros Difusos y del filtro medio:

$$\hat{O}_B^{i/d} = \sum_{l=1}^L [(\frac{I(l)}{\sqrt{E_{B^{i/d}(l)}}}) * B_{medio^{i/d}}] = \sum_{l=1}^L [(\frac{1}{W^{i/d}(l)} I(l)) * B_{medio^{i/d}}]$$

$$\text{siendo } W^{i/d}(l) = \frac{\sqrt{E_{B_{medio^{i/d}}}}}{\sqrt{E_{B^{i/d}(l)}}}$$

Al ser el filtro medio constante, puede salir de la suma:

$$\hat{O}_B^{i/d} = \sum_{l=1}^L [(\frac{1}{W^{i/d}(l)} I(l))] * B_{medio^{i/d}}$$

De este modo, las L convoluciones con la parte de campo difuso se sustituyen por una sola convolución con un filtro medio, por medio de una suma ponderada de la señal de entrada.

En la etapa S23B3, se puede calcular opcionalmente una ganancia G que corrige la ganancia del filtro medio $B_{medio^{i/d}}$. En efecto, en el caso de convolución entre las señales de entrada y los filtros no aproximados, sean cuales sean los valores de correlación entre las señales de entrada, el filtrado por filtros descorrelacionados que son los $B^{i/d}(l)$ lleva a señales que hay que sumar que entonces están a su vez también descorrelacionadas. A la inversa, en el caso de convolución entre las señales de entrada y el filtro medio aproximado, la energía de la señal procedente de la suma de las señales filtradas va a depender del valor de correlación que existe entre las señales de entrada.

Por ejemplo,

* Si todas las señales de entrada $I(l)$ son idénticas y de energía unitaria, y los filtros $B(l)$ están todos descorrelacionados (ya que se trata de campo difuso) y de energía unitaria, se tiene:

$$E_{O_B}^{i/d} = \text{energía} (\sum_{l=1}^L [I(l) * B_{norm}^{i/d}(l)]) = L$$

5 * si todas las señales de entrada I(1) están descorrelacionadas y son de energía unitaria, y los filtros B(l) son todos de energía unitaria pero sustituidos por filtros idénticos $\frac{B_{medio}^{i/d}}{\sqrt{E_B^{medio}^{i/d}}}$, se tiene:

$$\begin{aligned} E_{\hat{O}_B}^{i/d} &= \text{energía} (\sum_{l=1}^L [I(l) * (\frac{B_{medio}^{i/d}}{\sqrt{E_B^{medio}^{i/d}}})]) \\ &= \text{energía} (\frac{1}{\sqrt{E_B^{medio}^{i/d}}} \cdot \sum_{l=1}^L [I(l) * B_{medio}^{i/d}]) = (\frac{1}{\sqrt{E_B^{medio}^{i/d}}})^2 \cdot (L \cdot \frac{1}{L}) = L \end{aligned}$$

10 Porque se añaden las energías de las señales descorrelacionadas.

Este caso es equivalente al anterior en el sentido en que las señales procedentes del filtrado están todas descorrelacionadas, gracias a las señales de entrada en el primer caso, y gracias a los filtros en el segundo caso.

15 * si todas las señales de entrada I(1) son idénticas y de energía unitaria, y los filtros B(l) son todos de energía unitaria, pero sustituidos por filtros idénticos $\frac{B_{medio}^{i/d}}{\sqrt{E_B^{medio}^{i/d}}}$, se tiene:

$$\begin{aligned} E_{\hat{O}_B}^{i/d} &= \text{energía} (\sum_{l=1}^L [I(l) * (\frac{B_{medio}^{i/d}}{\sqrt{E_B^{medio}^{i/d}}})]) \\ &= \text{energía} (\frac{1}{\sqrt{E_B^{medio}^{i/d}}} \cdot \sum_{l=1}^L [I(l) * B_{medio}^{i/d}]) = (\frac{1}{\sqrt{E_B^{medio}^{i/d}}})^2 \cdot (L^2 \cdot \frac{1}{L}) = L^2 \end{aligned}$$

20 Porque las energías de las señales idénticas se añaden en cuadratura (porque sus amplitudes se añaden).

De este modo,

- 25 - si dos altavoces son activos simultáneamente, alimentados por señales descorrelacionadas, entonces no se aporta ninguna ganancia aplicando las etapas S23B1 y S23B2 con respecto al método clásico.
- si tres altavoces son activos simultáneamente, alimentados por señales idénticas, entonces se aporta una ganancia de $10 \cdot \log_{10}(L^2/L) = 10 \cdot \log_{10}(2^2/2) = 3,01 \text{ dB}$ aplicando las etapas S23B1 y S23B2 con respecto al método clásico.
- 30 - si tres altavoces son activos simultáneamente, alimentados por señales idénticas, entonces se aporta una ganancia de $10 \cdot \log_{10}(L^2/L) = 10 \cdot \log_{10}(3^2/3) = 4,77 \text{ dB}$ aplicando las etapas S23B1 y S23B2 con respecto al método clásico.

35 Los casos evocados anteriormente corresponden a casos extremos de señales idénticas o descorrelacionadas. Estos casos son sin embargo realistas: una fuente posicionada en el medio de dos altavoces, virtuales o reales, proporcionará una señal idéntica a estos dos altavoces (por ejemplo, con una técnica de tipo VBAP, por sus siglas de "Vector base amplitude panning", en español, amplitud de panoramización de base de vector). En el caso de un posicionamiento en un sistema 3D, los 3 altavoces pueden recibir la misma señal al mismo nivel.

40 De este modo, se puede aplicar una compensación con el fin de respetar la energía de las señales binauralizadas.

Idóneamente, esta ganancia de compensación G se determinará en función de la señal de entrada (es decir G(I(1))) y se aplicará a la suma de las señales de entrada ponderadas:

$$45 \quad \hat{O}_B^{i/d} = G \cdot \sum_{l=1}^L [(\frac{1}{W^{i/d}(l)} \cdot I(l)) * B_{medio}^{i/d}]$$

La ganancia G(I(l)) puede estimarse mediante un cálculo de correlación entre cada una de las señales. Puede estimarse igualmente comparando las energías de las señales antes y después de las sumas. En ese caso, la ganancia G puede variar dinámicamente a lo largo del tiempo, por ejemplo, en función de las correlaciones entre las señales de entrada, que varían ellas mismas a lo largo del tiempo.

50 En un modo de realización simplificado, es posible establecer una ganancia constante, por ejemplo $G = -3 \text{ dB} = 10^{-3/20}$, lo que evitará tener que efectuar una estimación de correlación que puede ser costosa. La ganancia constante G puede aplicarse entonces fuera de línea a los factores de ponderación (obteniéndose así $\frac{G}{W^{i/d}(l)}$), o al filtro $B_{medio}^{i/d}$, lo que evitará la aplicación de una ganancia suplementaria al vuelo.

Una vez calculadas las funciones de transferencia A y B aisladas y los filtros $B_{medio}^{i/d}$ (opcionalmente los pesos $W^{i/d}(l)$ y G), se aplican estas funciones de transferencia y estos filtros a las señales de entrada.

5 En un primer modo de realización, descrito con referencia a la figura 4, el tratamiento de la señal multicanal mediante la aplicación de los filtros Directo (A) y Difuso (B) para cada uno de los oídos se efectúa de la siguiente manera:

- Se aplica (etapas S4A1 a S4AL) a la señal multicanal de entrada un filtrado eficaz (por ejemplo, convolución directa basada-FFT) mediante los filtros Directo (A), como se describe en el estado de la técnica. Se obtiene una señal $\hat{O}_A^{i/d}$
- 10 - En función de las relaciones entre las señales de entrada, concretamente, en función de su correlación, se puede corregir opcionalmente en la etapa S4B11 la ganancia del filtro medio $B_{medio}^{i/d}$ mediante la aplicación de la ganancia G a las señales de salida después de la suma de las señales de entrada previamente ponderadas (etapas M4B1 a M4BL).
- 15 - Se aplica a la señal multicanal B en la etapa S4B1 un filtrado eficaz mediante el filtro Difuso medio B_{medio} . Esta etapa tiene lugar, en efecto, después de la suma de las señales de entrada previamente ponderadas (etapas M4B1 a M4BL). Se obtiene la señal $\hat{O}_B^{i/d}$.
- 20 - Se aplica a la señal $\hat{O}_B^{i/d}$ un plazo iDD con el fin de compensar el retraso introducido durante la etapa de aislamiento de la señal B en la etapa S4B2.
- Se suman las señales $\hat{O}_A^{i/d}$ y $\hat{O}_B^{i/d}$.
- 25 - Si se ha realizado un truncamiento que elimina las muestras inaudibles en el inicio de las respuestas de impulso, entonces, en la etapa S41 se aplica a la señal de entrada un plazo iT correspondiente a las muestras inaudibles suprimidas.

30 En una variante, con referencia a la figura 5, las señales no solo se calculan para los oídos izquierdo y derecho (índices i y d, en lo sucesivo) sino para k dispositivos de restitución (normalmente altavoces).

En un segundo modo de realización, se aplica la ganancia G antes de la suma de las señales de entrada, es decir durante las etapas de ponderación (etapas M4B1 a M4BL).

35 En un tercer modo de realización, se aplica una descorrelación a las señales de entrada. De este modo, las señales están descorrelacionadas después de la convolución mediante el filtro B_{medio} sean cuales sean las correlaciones de origen entre señales de entrada. Se puede usar una implementación eficaz de descorrelación (por ejemplo, usando una red de retrasos en bucle) con el fin de evitar el uso de filtros de descorrelación costosos.

40 De este modo, al suponer de manera realista que unas BRIR con una longitud de 48000 muestras pueden estar:

- truncadas entre la muestra 150 y la muestra 3222 mediante la técnica descrita en la etapa S21,
- 45 - descompuestas en dos partes: campo directo A de 1024 muestras, y campo difuso B de 2048 muestras, mediante la técnica descrita en la etapa S22,

entonces, la complejidad de binauralización puede venir dada aproximadamente por:

$$50 \quad C_{inv} = C_{invA} + C_{invB} = (L+2) \cdot (6 \cdot \log_2(2 \cdot NA)) + (L+2) \cdot (6 \cdot \log_2(2 \cdot NB))$$

Siendo NA y NB los tamaños en muestras de A y B

De este modo, para nBloques=10, Fs=48000, L=22, NA=1024 y NB=2048, la complejidad por muestra de señal multicanal para una convolución basada-FFT es de $C_{conv} = 3312$ multiplicaciones-adiciones. Este resultado, sin embargo, ha de compararse lógicamente con una solución simple que implementa solo el truncamiento, es decir para nBloques=10, Fs=3072, L=22:

$$60 \quad C_{truncado} = (L+2) \cdot (nBloques) \cdot (6 \cdot \log_2(2 \cdot Fs/nBloques)) = 13339$$

60 Existe por tanto un factor $19049/3312 = 5,75$ de complejidad entre el estado de la técnica y la presente invención, e incluso un factor $13339/3312=4$ de complejidad entre el estado de la técnica que se beneficia del truncamiento y la presente invención.

65 Si el tamaño de B es un múltiplo del tamaño de A, entonces si el filtrado se implementa por bloques de FFT, el cálculo de una FFT para A puede reutilizarse para B. Se necesita por tanto L FFT sobre NA puntos, que servirán a la vez para

el filtrado por A y por B, dos FFT inversas sobre NA puntos para obtener la señal binaural temporal, y la multiplicación de los espectros de frecuencia.

5 En este caso, se puede aproximar la complejidad (se ignoran las adiciones, (L+1) corresponde a la multiplicación de los espectros, L para A y 1 para B) por :

$$C_{inv2} = (L+2).(6.\log_2(2.NA)) + (L+1) = 1607$$

10 Con este enfoque, se gana también un factor 2, y por tanto un factor 12 y 8 con respecto al estado de la técnica no truncada y truncada.

La invención puede encontrar una aplicación directa en la norma MPEG-H 3D Audio.

15 Por supuesto, la presente invención no se limita a la forma de realización descrita anteriormente; se extiende a otras variantes quedando al mismo tiempo dentro del alcance de la protección definida en las reivindicaciones adjuntas.

20 De este modo, anteriormente se ha descrito, por ejemplo, un modo de realización en el que la señal Directa A no está aproximada por un filtro medio. Por supuesto, se puede usar un filtro medio de A para hacer las convoluciones (etapas S4A1 a S4AL) con las señales procedentes de los altavoces.

Se ha descrito anteriormente un modo de realización basado en el tratamiento de un contenido multicanal generado para L altavoces. Por supuesto, el contenido multicanal puede ser generado por cualquier tipo de fuente audio como, por ejemplo, la voz, un instrumento de música, cualquier ruido, etc.

25 Se ha descrito anteriormente un modo de realización basado en valores de BRIR determinadas en una sala. Por supuesto, se puede implementar la presente invención para cualquier tipo de entorno exterior (por ejemplo, una sala de concierto, al air libre, etc.).

30 Se ha descrito anteriormente un modo de realización basado en la aplicación de dos funciones de transferencia. Por supuesto, se puede implementar la presente invención con más de dos funciones de transferencia. Por ejemplo, se puede aislar de manera sincronizada una parte relativa a los sonidos emitidos directamente, una parte relativa a las primeras reflexiones y una parte relativa a los sonidos difusos.

REIVINDICACIONES

1. Procedimiento de espacialización sonora en el que se aplica al menos un filtrado en al menos dos señales de entrada (I(1), I(2), ..., I(L)) para suministrar al menos dos señales de salida (O(1), O(2), ..., O(K)), incluyendo el filtrado:

- ponderar (M4B1, M4B2, ..., M4BL) dichas al menos dos señales de entrada con pesos de ponderación ($W^k(1)$, ..., $W^k(L)$) respectivos, siendo cada peso de ponderación específico para cada una de las señales de entrada;
- para cada respuesta de impulso que incorpora un efecto de sala entre una pluralidad de respuestas de impulso que incorporan un efecto de sala, estando dicha respuesta de impulso que incorpora un efecto de sala respectivamente asociada a una señal de entrada entre dichas al menos dos señales de entrada (I(1), I(2), ..., I(L)) y a una señal de salida entre dichas al menos dos señales de salida (O(1), O(2), ..., O(K)):

- o particionar (S22), en un dominio temporal, dicha respuesta de impulso en una primera parte (A) y en una segunda parte (B), efectuándose dicho particionamiento de modo que:

- dicha primera parte representa unas propagaciones sonoras directas y unas primeras reflexiones sonoras de dichas propagaciones y se extiende por un primer número de muestras, y
- dicha segunda parte representa un campo sonoro difuso presente después de dichas primeras reflexiones y se extiende por un segundo número de muestras, siendo dicho segundo número de muestras un múltiplo de dicho primer número de muestras;

- o determinar una primera función de transferencia ($A^k(1)$, $A^k(2)$, ..., $A^k(L)$) a partir de dicha primera parte;
- o determinar una segunda función de transferencia a partir de dicha segunda parte;

- para cada señal de salida (O(1), O(2), ... O(K)) entre dichas al menos dos señales de salida (O(1), O(2), ..., O(K)):

- o determinar (S23B1) una tercera función de transferencia (B_{medio}^k) a partir de una media de dichas segundas funciones de transferencia correspondiente a la señal de salida (O(1), O(2), ..., O(K));
- o aplicar (S4A1, S4A2, ..., S4AL) a cada señal de entrada (I(1), I(2), ..., I(L)) la primera función de transferencia ($A^k(1)$, $A^k(2)$, ..., $A^k(L)$) correspondiente a la señal de entrada (I(1), I(2), ..., I(L)) y a la señal de salida (O(1), O(2), ..., O(K));
- o aplicar (S4B1) a cada señal de entrada la tercera función de transferencia (B_{medio}^k) correspondiente a la señal de salida (O(1), O(2), ... O(K));
- en donde la aplicación de las primeras y terceras funciones de transferencia se efectúa por bloques de FFT;

- sumar las señales resultantes de dichas aplicaciones de las primeras y terceras funciones de transferencia para obtener dichas al menos dos señales de salida (O(1), O(2), ..., O(K)).

2. Procedimiento según la reivindicación 1, caracterizado por que una ganancia de compensación de energía (G) se aplica (S4B11) a los pesos de ponderación ($W^k(1)$, ..., $W^k(L)$).

3. Procedimiento según una de las reivindicaciones anteriores, caracterizado por que dicho particionamiento de dicha respuesta de impulso incluye las operaciones:

- de determinación (S211) de un instante de inicio de presencia de ondas sonoras directas,
- de determinación de un instante de inicio de presencia de dicho campo sonoro difuso después de las primeras reflexiones, y
- de selección (S22), en dicha respuesta de impulso, de una parte de la respuesta que se extiende temporalmente entre dicho instante de inicio de presencia de ondas sonoras directas hasta dicho instante de inicio de presencia de campo difuso, correspondiendo dicha parte de respuesta seleccionada a dicha primera función de transferencia.

4. Procedimiento según la reivindicación 3, caracterizado por que dicho filtrado incluye la aplicación de al menos un plazo de compensación (S3B2) correspondiente a una diferencia temporal entre dicho instante de inicio de ondas sonoras directas y dicho instante de inicio de presencia de campo difuso.

5. Procedimiento según la reivindicación 4, caracterizado por que dichas primera y tercera funciones de transferencia se aplican en paralelo a dichas señales de entrada y por que dicho al menos un plazo de compensación se aplica a las señales de entrada filtradas por dichas terceras funciones de transferencia.

6. Procedimiento según una de las reivindicaciones anteriores, en el que dicha tercera función de transferencia viene dada por:

$$B_{medio}^k = \frac{1}{L} \sum_{l=1}^L [B_{norm}^k(l)]$$

siendo:

k un índice relativo a una señal de salida,

$l \in [1;L]$ un índice relativo a una señal de entrada entre dichas señales de entrada,
 L un número de señales de entrada,
 $B_{norm}^k(l)$ es una función de transferencia normalizada obtenida a partir de una segunda función de transferencia entre dichas segundas funciones de transferencia.

5
 7. Procedimiento según la reivindicación 6, caracterizado por que al menos una señal de salida O^k de dicho procedimiento viene dada por:

$$O^k = \sum_{l=1}^L (I(l) * A^k(l)) + z^{-iDD} \cdot \sum_{l=1}^L \left(\frac{1}{W^k(l)} \cdot I(l) \right) * B_{medio}^k$$

10 siendo:

$I(l)$ una señal de entrada entre dichas señales de entrada,
 $A^k(l)$ una primera función de transferencia entre dichas primeras funciones de transferencia,
 $W^k(l)$ un peso de ponderación entre dichos pesos de ponderación,
 z^{-iDD} corresponde a la aplicación de dicho plazo de compensación,
 donde \cdot es la multiplicación, y
 donde $*$ es el operador de convolución.

15
 8. Procedimiento según la reivindicación 6, caracterizado por que comprende una etapa de descorrelación de las señales de entrada, previamente a la aplicación de las terceras funciones de transferencia y por que al menos una señal de salida O^k de dicho procedimiento viene dada por:

$$O^k = \sum_{l=1}^L (I(l) * A^k(l)) + z^{-iDD} \cdot \sum_{l=1}^L \left(\frac{1}{W^k(l)} \cdot I_d(l) \right) * B_{medio}^k$$

25 siendo:

$I(l)$ una señal de entrada entre dichas señales de entrada,
 $I_d(l)$ una señal de entrada entre dichas señales de entrada que se han sometido a dicha etapa de descorrelación,
 $A^k(l)$ una primera función de transferencia entre dichas primeras funciones de transferencia,
 $W^k(l)$ un peso de ponderación entre dichos pesos de ponderación,
 z^{-iDD} corresponde a la aplicación de dicho plazo de compensación,
 donde \cdot es la multiplicación, y
 donde $*$ es el operador de convolución.

30
 9. Procedimiento según la reivindicación 6, caracterizado por que comprende una etapa de determinación de una ganancia de compensación de energía en función de las señales de entrada y por que al menos una señal de salida viene dada por:

$$O^k = \sum_{l=1}^L (I(l) * A^k(l)) + z^{-iDD} \cdot \sum_{l=1}^L \left(G(l) \frac{1}{W^k(l)} \cdot I(l) \right) * B_{medio}^k$$

40 siendo:

$I(l)$ una señal de entrada entre dichas señales de entrada,
 $G(l)$ dicha ganancia de compensación de energía determinada,
 $A^k(l)$ una primera función de transferencia entre dichas primeras funciones de transferencia,
 $W^k(l)$ un peso de ponderación entre dichos pesos de ponderación,
 z^{-iDD} corresponde a la aplicación de dicho plazo de compensación,
 donde \cdot es la multiplicación, y
 donde $*$ es el operador de convolución.

50 10. Procedimiento según una de las reivindicaciones 6 a 9, caracterizado por que dicho peso viene dado por:

$$W^k(l) = \frac{\sqrt{E_{B_{medio}^k}}}{\sqrt{E_{B^k(l)}}}$$

siendo k el índice relativo a una señal de salida,
 $l \in [1; L]$ el índice relativo a una señal de entrada entre dichas señales de entrada,
 L el número de señales de entrada,

55 siendo $E_{B_{medio}^k}$ una energía relativa a B_{medio}^k ,

$E_{B^k(l)}$ una energía relativa a una segunda función de transferencia entre las segundas funciones de transferencia.

11. Programa informático que incluye unas instrucciones para la implementación del procedimiento según una de las reivindicaciones 1 a 10, cuando estas instrucciones son ejecutadas por un procesador.

60

12. Dispositivo de espacialización sonora, que incluye al menos un filtro aplicado en al menos dos señales de entrada ($I(1), I(2), \dots, I(L)$), siendo el dispositivo apto para suministrar al menos dos señales de salida ($O(1), O(2), \dots, O(K)$), comprendiendo el dispositivo unos módulos de ponderación ($M4B1, M4B2, \dots, M4BL$) para ponderar dichas al menos dos señales de entrada con unos pesos de ponderación ($W^k(1), \dots, (W^k(L))$ respectivos, siendo cada peso de ponderación específico para cada una de las señales de entrada; estando además el dispositivo configurado para:

- para cada respuesta de impulso que incorpora un efecto de sala entre una pluralidad de respuestas de impulso que incorporan un efecto de sala, estando dicha respuesta de impulso que incorpora un efecto de sala respectivamente asociada a una señal de entrada entre dichas al menos dos señales de entrada ($I(1), I(2), \dots, I(L)$) y a una señal de salida entre dichas al menos dos señales de salida ($O(1), O(2), \dots, O(K)$):

o particionar (S22), en un dominio temporal, dicha respuesta de impulso en una primera parte (A) y en una segunda parte (B), efectuándose dicho particionamiento de modo que:

dicha primera parte representa unas propagaciones sonoras directas y unas primeras reflexiones sonoras de dichas propagaciones y se extiende por un primer número de muestras, y dicha segunda parte representa un campo sonoro difuso presente después de dichas primeras reflexiones y se extiende por un segundo número de muestras, siendo dicho segundo número de muestras un múltiplo de dicho primer número de muestras;

o determinar una primera función de transferencia ($A^k(1), A^k(2), \dots, A^k(L)$) a partir de dicha primera parte;

o determinar una segunda función de transferencia a partir de dicha segunda parte, comprendiendo el filtro:

- para cada señal de salida ($O(1), O(2), \dots, O(K)$) entre las dichas al menos dos señales de salida ($O(1), O(2), \dots, O(K)$):

o determinar (S23B1) una tercera función de transferencia (B_{medio}^k) a partir de una media de dichas segundas funciones de transferencia correspondiente a la señal de salida ($O(1), O(2), \dots, O(K)$);

o aplicar (S4A1, S4A2, ..., S4AL) a cada señal de entrada de las primeras funciones de transferencia correspondientes a la señal de salida ($O(1), O(2), \dots, O(K)$);

o aplicar (S4B1) a cada señal de entrada la tercera función de transferencia (B_{medio}^k) correspondiente a la señal de salida ($O(1), O(2), \dots, O(K)$);

en el que las señales resultantes de dichas aplicaciones de las primeras y terceras funciones de transferencia se suman para obtener dichas al menos dos señales de salida ($O(1), O(2), \dots, O(K)$).

13. Módulo de descodificación de señales sonoras, que incluye un dispositivo de espacialización según la reivindicación 12, de dichas señales sonoras como señales de entrada.

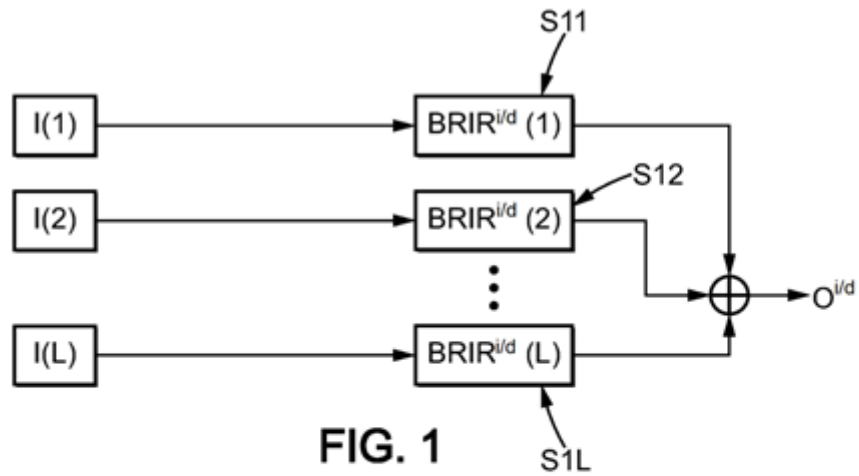


FIG. 1

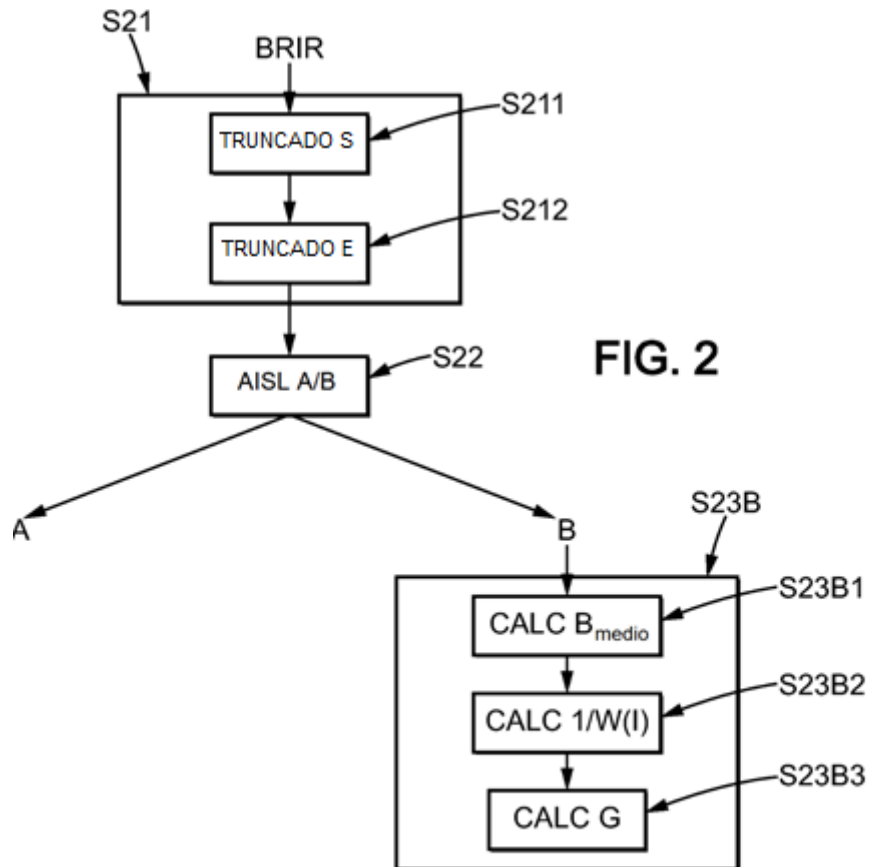


FIG. 2

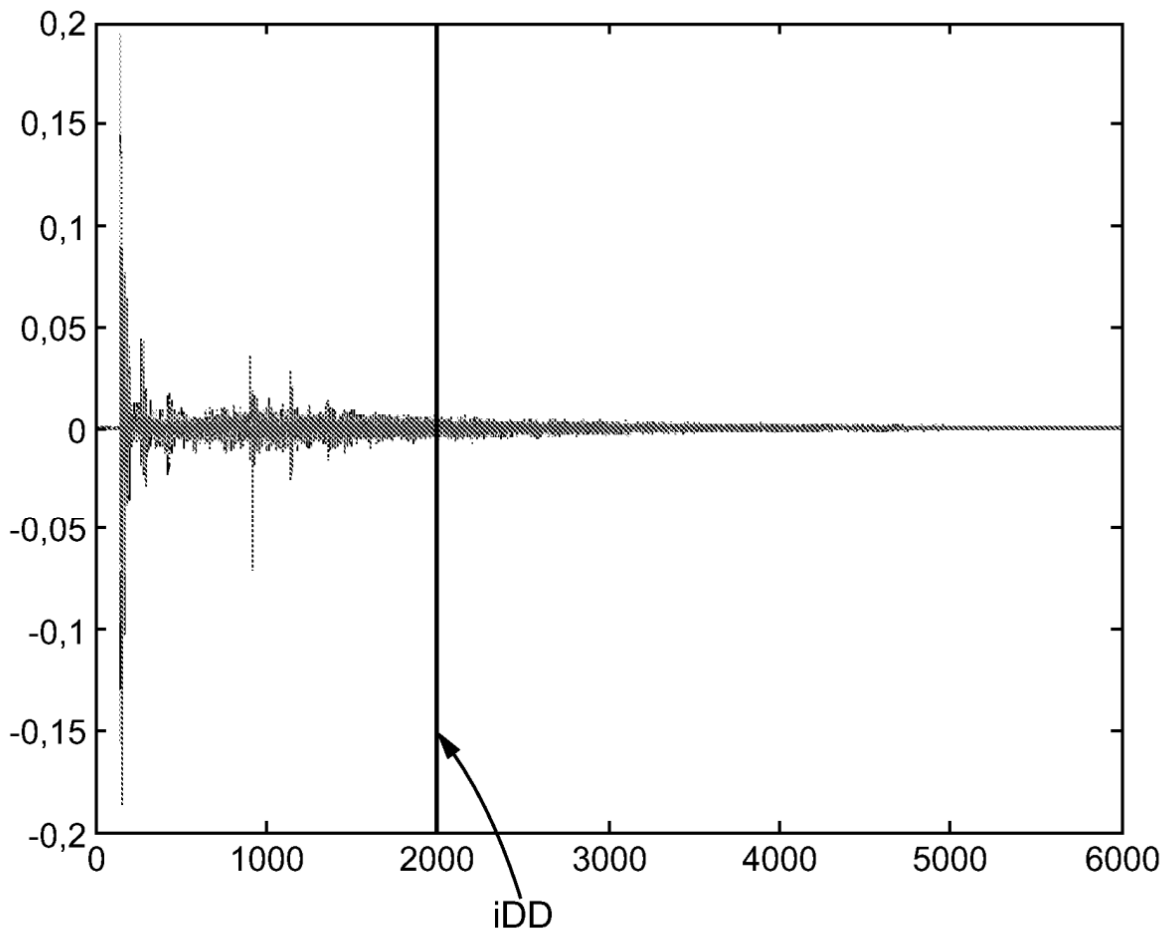
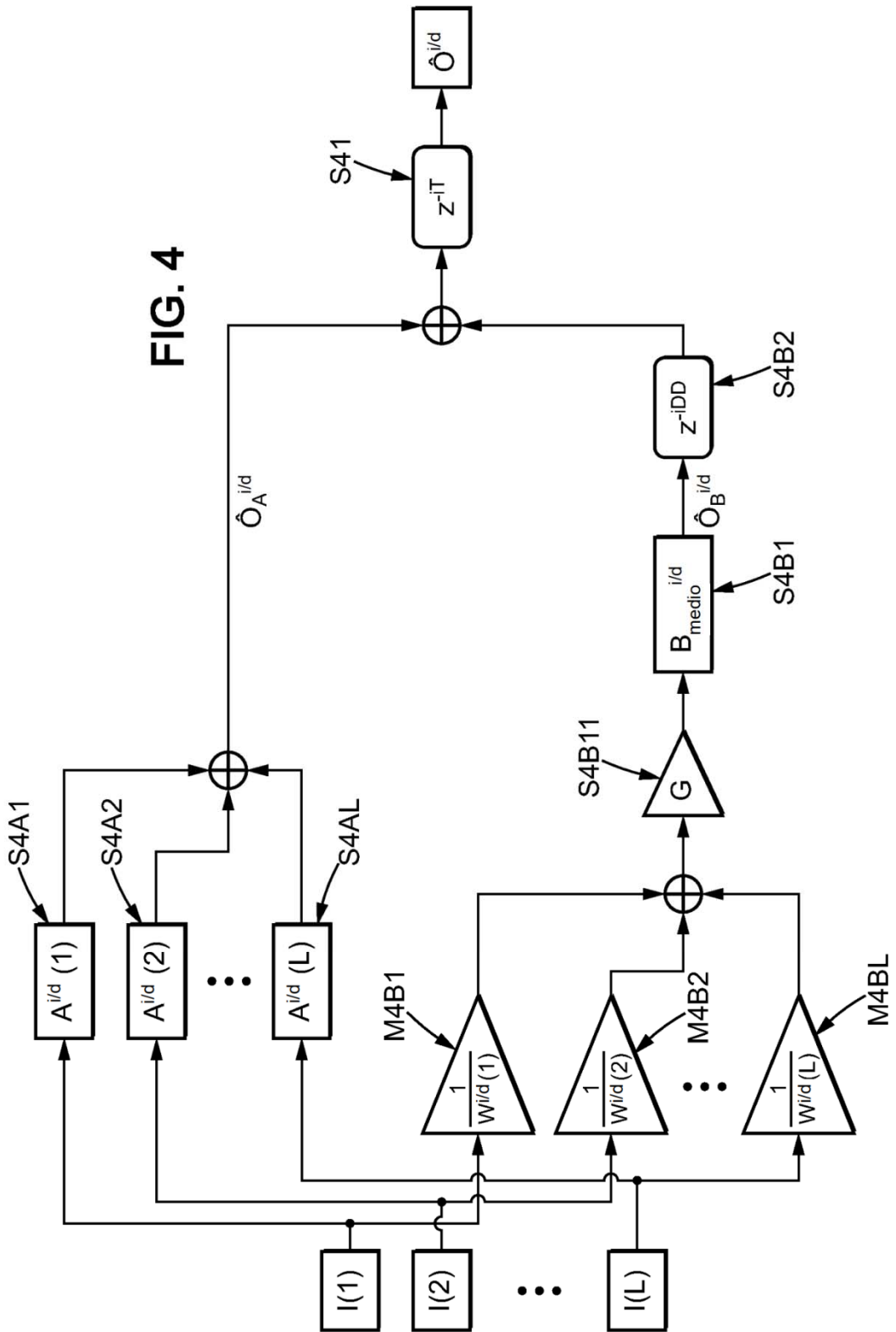
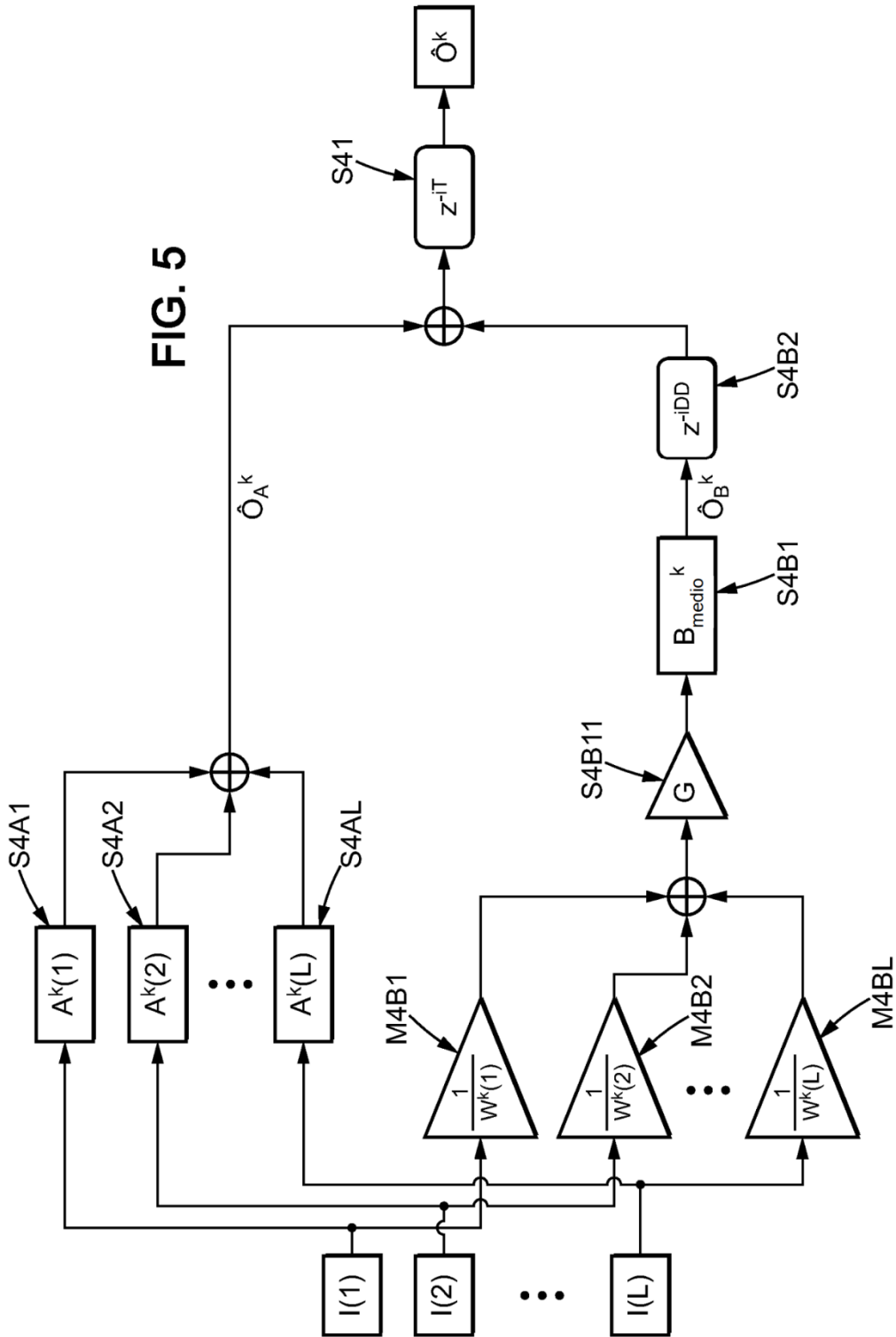


FIG. 3





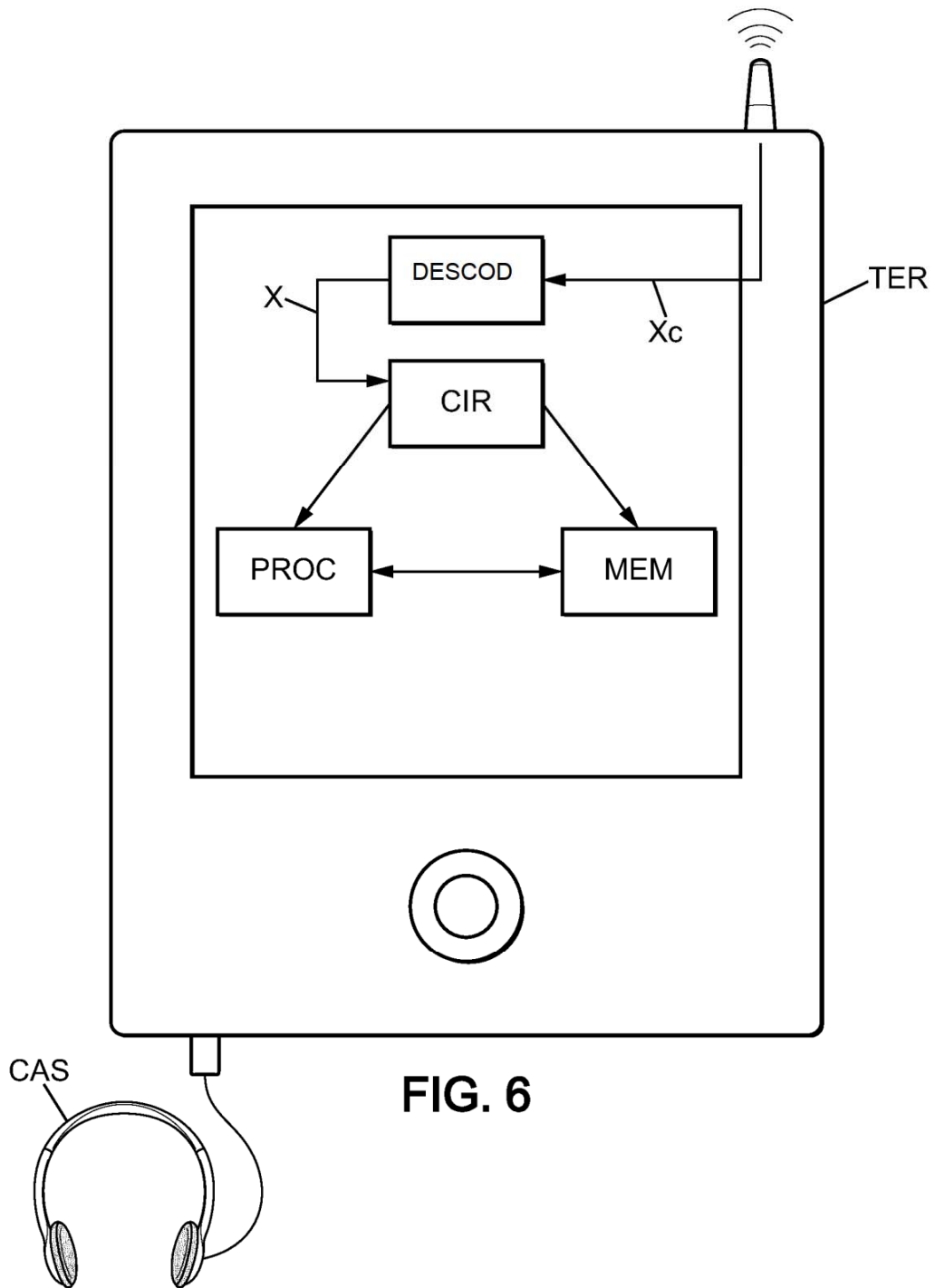


FIG. 6