

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 984 773**

51 Int. Cl.:

**G10L 19/005** (2013.01)

**G10L 19/008** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.06.2020 PCT/EP2020/065631**

87 Fecha y número de publicación internacional: **17.12.2020 WO20249480**

96 Fecha de presentación y número de la solicitud europea: **05.06.2020 E 20729787 (0)**

97 Fecha y número de publicación de la concesión europea: **24.04.2024 EP 3984027**

54 Título: **Ocultación de la pérdida de paquetes para la codificación de audio espacial basada en DirAC**

30 Prioridad:

**12.06.2019 EP 19179750**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**31.10.2024**

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR  
FÖRDERUNGDER ANGEWANDTEN FORSCHUNG  
E.V. (100.0%)  
Hansastr. 27c  
80686 München, DE**

72 Inventor/es:

**FUCHS, GUILLAUME;  
MULTRUS, MARKUS;  
DÖHLA, STEFAN y  
EICHENSEER, ANDREA**

74 Agente/Representante:

**ARIZTI ACHA, Monica**

**ES 2 984 773 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

Ocultación de la pérdida de paquetes para la codificación de audio espacial basada en DirAC

5 **Campo técnico**

Las realizaciones de la presente invención se refieren a un método para la ocultación de pérdida de parámetros espaciales de audio, un método para decodificar una escena de audio codificada con DirAC y a los programas informáticos correspondientes. Las realizaciones adicionales se refieren a un aparato de ocultación de pérdida para la ocultación de pérdida de parámetros espaciales de audio y a un decodificador que comprende un aparato de ocultación de pérdida de paquetes. Las realizaciones preferidas describen un concepto/método para compensar las degradaciones de calidad debido tramas o paquetes perdidos y corruptos que se presentan durante la transmisión de una escena de audio para la cual se codificó paramétricamente la imagen espacial por medio del paradigma de codificación de audio direccional (DirAC, Directional Audio Coding).

En la técnica anterior se abordan algunos enfoques para tratar una pérdida de daño de la información. Por ejemplo, el documento WO 2015/003027 A1 describe un concepto para la ocultación de la pérdida de paquetes. Además, la publicación titulada "Directional Audio Coding-Perception Based Reproduction and Spatial Sound" y el documento EP 2423702 forman parte del estado de la técnica.

20 **Introducción**

La comunicación de voz y audio puede estar sujeta a diferentes problemas de calidad debido a la pérdida de paquetes durante la transmisión. De hecho, las malas condiciones en la red, tales como errores de bits y fluctuaciones, pueden llevar a la pérdida de algunos paquetes. Estas pérdidas resultan en artefactos severos, como clics, plops o silencios indeseables que degradan en gran medida la calidad percibida de la voz o señal de audio reconstruida en el lado del receptor. Para combatir el impacto adverso de la pérdida de paquetes, se han propuesto algoritmos de ocultación de pérdida de paquetes (PLC, Packet Loss Concealment) en esquemas de codificación de voz y audio convencionales. Tales algoritmos operan normalmente en el lado del receptor generando una señal de audio sintética para ocultar los datos perdidos en el flujo de bits recibido.

DirAC es una técnica de procesamiento de audio espacial motivada por la percepción que representa de manera compacta y eficiente el campo de sonido por medio de un conjunto de parámetros espaciales y una señal de mezcla de audio. La señal de mezcla de audio puede ser una monofónica, estereofónica, o señales multicanal en un formato de audio tal como formato A o formato B, también conocidos como Ambisonics de primer orden (FAO, First Order Ambisonics). La señal de mezcla de audio se complementa por parámetros espaciales de DirAC que describen la escena de audio en términos de dirección de llegada (DoA, Direction-Of-Arrival) y difusividad por unidad de tiempo/frecuencia. En aplicaciones de almacenamiento, streaming o comunicación, la señal de mezcla de audio se codifica por medio de un codificador central convencional (por ejemplo, EVS o una extensión estéreo/multicanal de EVS o cualquier otro códecs mono/estéreo/multicanal), que tenga como objetivo conservar la forma de onda de audio de cada canal. El codificador central se puede construir en torno a un esquema de codificación basado en transformada o esquema de codificación de voz que opera en el dominio del tiempo, tal como CELP. Entonces, el codificador central puede integrar herramientas de resistencia a errores ya existentes, tales como algoritmos de ocultación de pérdida de paquetes (PLC).

Por otro lado, no hay solución existente para proteger los parámetros espaciales de DirAC.

Por lo tanto, hay una necesidad de un planteamiento mejorado.

50 **Breve descripción de las realizaciones**

Es un objetivo de la presente invención proporcionar un concepto para la ocultación de pérdidas en el contexto de DirAC.

55 Este objetivo fue resuelto por el objeto de las reivindicaciones independientes (reivindicación 1 del método, reivindicación 12 del medio de almacenamiento digital legible por ordenador y reivindicación 13 del aparato de ocultación de pérdidas). En la técnica anterior se abordan algunos enfoques para tratar una pérdida de daño de la información.

60 **Breve descripción de los dibujos**

Las realizaciones de la presente invención se comentarán posteriormente haciendo referencia a las figuras adjuntas, en el que

la Figura 1 muestra diagramas de bloques esquemáticos que ilustran un análisis y síntesis de DirAC;

la Figura 2 muestra un diagrama de bloques esquemático detallado de un análisis y síntesis de DirAC en el codificador de audio 3D de tasa de bits inferior;

5 la Figura 3a muestra un diagrama de flujo esquemático de un método para la ocultación de pérdidas de acuerdo con una realización básica;

la Figura 3b muestra un aparato de ocultación de pérdidas esquemático de acuerdo con una realización básica;

10 las Figuras 4a, 4b muestran diagramas esquemáticos de funciones de difusividad medida de DDR (Figura 4a, tamaño de ventana  $W = 16$ , Figura 4b, tamaño de ventana  $W = 512$ ) con el fin de ilustrar las realizaciones;

15 la Figura 5 muestra un diagrama esquemático de dirección medida (azimut y elevación) en la función de difusividad con el fin de ilustrar las realizaciones;

la Figura 6a muestra un diagrama de flujo esquemático de un método para decodificar una escena de audio codificada con DirAC de acuerdo con las realizaciones; y

20 la Figura 6b muestra un diagrama de bloques esquemático de un decodificador para una escena de audio codificada con DirAC de acuerdo con una realización.

A continuación, se describirán las realizaciones de la presente invención haciendo referencia a las figuras adjuntas, en las que se proporcionan números de referencia idénticos a los objetos/elementos que tienen una función idéntica o similar, de modo que la descripción de los mismos sea mutuamente aplicable e intercambiable. Antes de tratar en detalle las realizaciones de la presente invención, se ofrece una introducción al DirAC.

25

### **Descripción detallada de realizaciones**

30 Introducción a DirAC: DirAC es una reproducción de sonido espacial motivada perceptualmente. Se supone que en un instante en el tiempo y para una banda crítica, la resolución espacial del sistema auditivo está limitada a decodificar una señal para dirección y otra para coherencia interaural. Con base en estas suposiciones, DirAC representa el sonido espacial en una banda de frecuencia por medio de atenuación cruzada de dos flujos: un flujo difuso no direccional y un flujo no difuso direccional. El procesamiento de DirAC se lleva a cabo en dos fases:

35 La primera fase es el análisis como se ilustra por medio de la Figura 1a y la segunda fase es la síntesis como se ilustra por medio de la Figura 1b.

40 La Figura 1a muestra la etapa de análisis 10 que comprende uno o más filtros de paso de banda 12a-n que reciben las señales de micrófono W, X, Y y Z, una etapa de análisis 14e para la energía y 14i para la intensidad. Al utilizar el acomodo temporal, se puede determinar la difusividad  $\Psi$  (ver el número de referencia 16d). La difusividad  $\Psi$  se determina con base en el análisis de energía 14c y la intensidad 14i. Con base en el análisis de intensidad 14i, se puede determinar una dirección 16e. El resultado de la determinación de la dirección es el azimut y el ángulo de elevación.  $\Psi$ , azi y ele son salida como metadatos. Estos metadatos se utilizan por medio de la entidad de síntesis 20 mostrada en la Figura 1b.

45 La entidad de síntesis 20 como se muestra en la Figura 1b comprende un primer flujo 22a y un segundo flujo 22b. El primer flujo comprende una pluralidad de filtros de paso de banda 12a-n y una entidad de cálculo para micrófonos virtuales 24. El segundo flujo 22b comprende medios para procesar los metadatos, es decir 26 para el parámetro de difusividad y 27 para el parámetro de dirección. Además, se utiliza un decorrelador 28 en la etapa de síntesis 20, en donde esta entidad de decorrelación 28 recibe los datos de los dos flujos 22a, 22b. La salida del decorrelador 28 se puede proporcionar a los altavoces 29.

50 En la etapa de análisis de DirAC, un micrófono coincidente de primer orden en formato B se considera como entrada y se analiza la difusividad y dirección de llegada del sonido en el dominio de la frecuencia.

55 En la etapa de síntesis de DirAC, el sonido se divide en dos flujos, el flujo no difuso y el flujo difuso. El flujo no difuso se reproduce como fuentes puntuales utilizando panoramización de amplitud, que se puede hacer utilizando panoramización en amplitud a base de vectores (VBAP, Vector Base Amplitude Panning) [2]. El flujo difuso es responsable de la sensación de desarrollo y se produce al transmitir a los altavoces señales mutuamente decorrelacionadas.

60

Los parámetros de DirAC, también llamados metadatos espaciales o metadatos de DirAC en lo siguiente, constante de mensajes de datos de difusividad y dirección. La dirección se puede representar en coordenadas esféricas por

medio de dos ángulos, el azimut y la elevación, mientras que la difusividad es factor escalar entre 0 y 1.

Posteriormente, se comentará un sistema de una codificación de audio espacial de DirAC con respecto a la Figura 2. La Figura 2 muestra un análisis de DirAC 10' de dos etapas y una síntesis de DirAC 20'. Aquí, el análisis de DirAC comprende el análisis de banco de filtros 12, el estimador de dirección 16i y el estimador de difusividad 16d. Ambos, 16i y 16d proporcionan los datos de difusividad/dirección como metadatos espaciales. Estos datos se pueden codificar utilizando el codificador 17. El análisis directo 20' comprende el decodificador de metadatos espaciales 21, una síntesis de salida 23, una síntesis de banco de filtros 12 que permiten proporcionar una señal a los altavoces FOA/HOA.

En paralelo a la etapa de análisis directo 10' comentada y la etapa de síntesis directa 20', que procesan los metadatos espaciales, se utiliza un codificador/decodificador de EVS. En el lado del análisis, se lleva a cabo una conformación de haz/selección de señal con base en el formato B de señal de entrada (ver la entidad de conformación de haz/selección de señal 15). Después, la señal se codifica por EVS (ver el número de referencia 17). Después de la señal se codifica por EVS. En el lado de la síntesis (ver número de referencia 20'), se utiliza un decodificador de EVS 25. Este decodificador de EVS proporciona una señal a un análisis de banco de filtros 12, que proporciona su señal a la síntesis de salida 23.

Ya que ahora se ha comentado la estructura del análisis directo/síntesis directa 10'/20', se comentará a detalle su funcionalidad.

El codificador analiza 10' generalmente la escena de audio espacial en formato B. Alternativamente, el análisis de DirAC se puede ajustar para analizar formatos de audio diferentes como objetos de audio o señales multicanal o la combinación de cualquier formato de audio espacial. El análisis de DirAC extrae una representación paramétrica de la escena de audio de entrada. Una dirección de llegada (DoA) y una difusividad medida por unidad de tiempo-frecuencia forman los parámetros. El análisis de DirAC es seguido por un codificador de metadatos espaciales, que cuantifica y codifica los parámetros de DirAC para obtener una representación paramétrica de baja tasa de bits.

Junto con los parámetros, se deriva una señal de mezcla de audio a partir de las diferentes fuentes o se codifican señales de entrada de audio para transmisión por medio de un codificador central de audio convencional. En la realización preferida, se prefiere un codificador de audio de EVS para codificar la señal de mezcla de audio, pero la invención no se limita a este codificador central y se puede aplicar a cualquier codificador central. La señal de mezcla de audio consta de diferentes canales, llamados canales de transporte: la señal puede ser, por ejemplo, las señales de cuatro coeficientes que componen una señal de formato B, un par estéreo o una mezcla de audio monofónica dependiendo de la tasa de bits objetivo. Los parámetros espaciales codificados y el flujo de bits de audio codificado se multiplexan antes de transmitirse a través del canal de comunicación.

En el decodificador, los canales de transporte se decodifican por medio del decodificador central, mientras que los metadatos de DirAC se decodifican primero antes de transmitirse con los canales de transporte decodificados a la síntesis de DirAC. La síntesis de DirAC utiliza los metadatos decodificados para controlar la reproducción del flujo de sonido directo y su mezcla con el flujo de sonido difuso. El campo de sonido reproducido se puede reproducir en un diseño de altavoces arbitrario o se puede generar en formato Ambisonics (HOA/FOA) con un orden arbitrario.

Estimación de parámetros de DirAC: En cada banda de frecuencia, se estima la dirección de llegada de sonido junto con la difusividad del sonido. A partir del análisis de tiempo-frecuencia de los componentes de formato B de entrada  $w^i(n), x^i(n), y^i(n), z^i(n)$ , se pueden determinar los vectores de presión y velocidad como:

$$P^i(n, k) = W^i(n, k)$$

$$U^i(n, k) = X^i(n, k)e_x + Y^i(n, k)e_y + Z^i(n, k)e_z$$

donde  $i$  es el índice de la entrada y,  $k$  y  $n$ , índices de tiempo y frecuencia del mosaico de tiempo-frecuencia, y  $e_x, e_y, e_z$  representa los valores de unidad cartesiana.  $P(n, k)$  y  $U(n, k)$  se utilizan para calcular los parámetros de DirAC, es decir DoA y difusividad mediante el cómputo del vector de intensidad:

$$I(k, n) = \frac{1}{2} \Re\{P(k, n) \overline{U(n, k)}\},$$

donde  $\overline{(\cdot)}$  denota conjugación compleja. La difusividad del campo de sonido combinado está dada por:

$$\psi(k, n) = 1 - \frac{\|E\{I(k, n)\}\|}{cE\{E(k, n)\}}$$

donde  $E\{\}$  denota el operador de promediado temporal,  $c$  a velocidad del sonido y  $E(k, n)$  la energía del campo de sonido dada por:

$$E(n, k) = \frac{\rho_0}{4} \|U(n, k)\|^2 + \frac{1}{\rho_0 c^2} |P(n, k)|^2$$

La difusividad del campo de sonido se define como la relación entre la intensidad del sonido y la densidad de la energía que tienen valores entre 0 y 1.

La dirección de llegada (DoA) se expresa por medio del vector unitario *dirección*  $(n, k)$ , definido como

$$\text{dirección}(n, k) = \frac{I(n, k)}{\|I(n, k)\|}$$

La dirección de llegada se determina por medio de un análisis energético de la entrada de formato B y se puede definir como contraria a la dirección del vector de intensidad. La dirección está definida en coordenadas cartesianas pero se puede transformar fácilmente a coordenadas esféricas definidas por un radio unitario, el ángulo de azimut y ángulo de elevación.

En el caso de transmisión, los parámetros son necesarios para transmitirse al lado del receptor a través de un flujo de bits. Para una transmisión robusta a través de una red con capacidad limitada, se prefiere un flujo de bits de baja tasa de bits que se puede lograr diseñando un esquema de codificación eficiente para los parámetros de DirAC. Puede emplear, por ejemplo, técnicas tales como agrupamiento de banda de frecuencia promediando los parámetros a través de diferentes bandas de frecuencia y/o unidades de tiempo, predicción, cuantificación y codificación entrópica. En el decodificador, los parámetros transmitidos se pueden decodificar para cada unidad de tiempo/frecuencia  $(k, n)$  en caso de que no ocurra ningún error en la red.

Sin embargo, si las condiciones de la red no son lo suficientemente buenas para asegurar una transmisión de paquetes apropiada, se puede perder un paquete durante la transmisión. La presente invención tiene como propósito proporcionar una solución en el último caso.

Originalmente, la DirAC se pretendía para procesar señales de grabación en formato B, también conocidas como señales Ambisonics de primer orden. Sin embargo, el análisis se puede extender fácilmente a cualquier sistema de micrófonos que combine micrófonos omnidireccionales o direccionales. En este caso, la presente invención es todavía relevante ya que la esencia de los parámetros de DirAC no tiene cambios.

Adicionalmente, los parámetros de DirAC, también conocidos como metadatos, se pueden calcular directamente durante el procesamiento de señal de micrófono antes de transmitirse al codificador de audio espacial. El sistema de codificación espacial basado en DirAC después se proporciona por medio de parámetros espaciales de audio equivalentes o similares a los parámetros de DirAC en la forma de metadatos y una forma de onda de audio de una señal de mezcla de audio. La DoA y difusividad se pueden derivar fácilmente por banda de parámetros a partir de los meta datos introducidos. Tal formato de entrada se denomina en ocasiones como formato de audio espacial asistido por metadatos (MASA, Metadata-assisted spatial audio). MASA permite al sistema ignorar la especificidad de los sistemas de micrófonos y sus factores de forma necesarios para calcular los parámetros espaciales. Estos se derivarán fuera del sistema de codificación de audio espacial utilizando un procesamiento específico al dispositivo que incorpore los micrófonos.

Las realizaciones de la presente invención pueden utilizar un sistema de codificación espacial como se ilustra por medio de la Figura 2, donde se representan el codificador y decodificador de audio espacial basados en DirAC. Las realizaciones se comentarán con respecto a las Figuras 3a y 3b, en donde se comentarán antes las extensiones al modelo de DirAC.

De acuerdo con las realizaciones, el modelo de DirAC también se puede extender permitiendo diferentes componentes direccionales con el mismo mosaico de Tiempo/Frecuencia. Se puede extender en dos formas principales:

La primera extensión consiste en enviar dos o más DoAs por mosaico de T/F. Cada DoA se debe asociar entonces

con una energía, o una relación de energía. Por ejemplo, la  $l$ -ésima DoA se puede asociar con una relación de energía  $\eta_l$  entre la energía del componente direccional y la energía de la escena de audio general:

$$\eta_l(k, n) = \frac{\|E\{I_l(k, n)\}\|}{cE\{E(k, n)\}}$$

5 donde  $I_l(k, n)$  es el vector de intensidad asociado con la  $l$ -ésima dirección. Si se transmiten  $L$  DoAs junto con sus  $L$  porciones de energía, se pueden deducir las difusividades a partir de las  $L$  relaciones de energía como:

$$\psi(k, n) = 1 - \sum_{l=1}^L \eta_l(k, n)$$

10 Los parámetros espaciales transmitidos en el flujo de bits pueden ser las  $L$  direcciones junto con las  $L$  relaciones de energía o estos últimos parámetros también se pueden convertir a  $L-1$  relaciones de energía + un parámetro de difusividad.

$$\psi = 1 - \sum_{l=1}^L \eta_l$$

15 La segunda extensión consiste en dividir el espacio 2D o 3D en sectores no superpuestos y transmitir para cada uno de los sectores un conjunto de parámetros de DirAC (DoA+difusividad por sector). Hablamos entonces sobre DirAC de orden superior tal como se introduce en [5].

20 Ambas extensiones se pueden combinar realmente, y la presente invención es relevante para ambas extensiones. Las Figuras 3a y 3b ilustran realizaciones de la presente invención, en donde la Figura 3a muestra el planteamiento con enfoque en el concepto básico/método utilizado 100, en donde el aparato utilizado 50 se muestra por la Figura 3b.

25 La Figura 3a ilustra el método 100 que comprende las etapas básicas 110, 120 y 130.

30 Las primeras etapas 110 y 120 son comparables entre sí, es decir se refieren a la recepción de conjuntos de parámetros espaciales de audio. En la primera etapa 110, se recibe el primer conjunto, en donde en la segunda etapa 120, se recibe el segundo conjunto. Adicionalmente, pueden estar presentes etapas de recepción adicionales (no mostrados). Se debe observar que el primer conjunto puede referirse al primer punto en el tiempo/primer trama, el segundo conjunto puede referirse a un segundo punto en el tiempo (posterior)/segunda trama (posterior), etc. Como se comentó anteriormente, el primer conjunto así como el segundo conjunto pueden comprender una información de difusividad ( $\psi$ ) y/o una información de dirección (azimut y elevación). Esta información se puede codificar utilizando un codificador de metadatos espaciales. Ahora se hace la suposición de que el segundo conjunto de información se pierde o daña durante la transmisión. En este caso, el segundo conjunto se reemplaza por un primer conjunto. Esto permite una ocultación de pérdida de paquetes para parámetros espaciales de audio como los parámetros de DirAC.

40 En el caso de pérdida de paquetes, los parámetros de DirAC borrados de las tramas perdidas necesitan ser restituidos para limitar el impacto en la calidad. Esto se puede lograr generando sintéticamente los parámetros faltantes considerando los parámetros recibidos pasados. Una imagen espacial inestable se puede percibir como desagradable y como un artefacto, aunque una imagen espacial estrictamente constante se puede percibir como innatural.

45 El planteamiento 100, como se comenta con respecto a la Figura 3a se puede llevar a cabo por medio de la entidad 50 como se muestra en la Figura 3b. El aparato para la ocultación de pérdidas 50 comprende una interfaz 52 y un procesador 54. Por medio de la interfaz, se pueden recibir los conjuntos de parámetros espaciales de audio,  $\psi_1$ ,  $\text{azi}_1$ ,  $\text{ele}_1$ ,  $\psi_2$ ,  $\text{azi}_2$ ,  $\text{ele}_2$ ,  $\psi_n$ ,  $\text{azi}_n$ ,  $\text{ele}_n$ . El procesador 54 analiza los conjuntos recibidos y, en caso de un conjunto perdido o dañado, reemplaza el conjunto perdido o dañado, por ejemplo, por un conjunto recibido previamente o un conjunto comparable. Se pueden utilizar estas diferentes estrategias, lo cual se comentará a continuación.

50 Estrategia de retención: Generalmente es seguro considerar que la imagen espacial debe ser relativamente estable en el tiempo, lo cual se puede traducir para los parámetros de DirAC, es decir, la dirección de llegada y difusión que no cambian mucho entre tramas. Por esta razón, un planteamiento simple, pero efectivo, es mantener los parámetros de la última trama bien recibida para las tramas pérdidas durante la transmisión.

55 Extrapolación de la dirección: Alternativamente, se puede prever estimar la trayectoria de los eventos de sonido en la

escena de audio y después intentar extrapolar la trayectoria estimada. Es especialmente relevante si el evento de sonido está bien localizado en el espacio como una fuente puntual, que se refleja en el modelo de DirAC por una baja difusividad. La trayectoria estimada se puede calcular a partir de observaciones de direcciones pasadas y ajustando una curva entre estos puntos, lo cual puede involucrar cualquiera de interpolación o suavizado. También se puede emplear un análisis de regresión. Después se lleva a cabo la extrapolación evaluando la curva ajustada más allá del intervalo de datos observados.

En DirAC, las direcciones a menudo se expresan, cuantifican y codifican en coordenadas polares.

Sin embargo, generalmente es más conveniente procesar las direcciones y después la trayectoria en coordenadas cartesianas para evitar manejar operaciones de módulo  $2\pi$ .

Tramado de la dirección: Cuando el evento de sonido es más difuso, las direcciones son menos significativas y se pueden considerar como la realización de un proceso estocástico. El tramado puede entonces ayudar a que el campo de sonido representado sea más natural y agradable al inyectar ruido aleatorio en las direcciones previas antes de utilizarlo para las tramas perdidas. El ruido inyectado y su varianza pueden ser función de la difusividad.

Utilizando un análisis de escena de audio de DirAC estándar, podemos estudiar la influencia de la difusividad en la precisión y significación de la dirección del modelo. Utilizando una señal de formato B artificial para la cual está dada la relación de energía directa a difusa (DDR, Direct-to-Diffuse energy Ratio) entre un componente de onda plana y componente de campo difuso, podemos analizar los parámetros de DirAC resultantes y su precisión.

La difusividad teórica  $\Psi$  es función de la relación de energía directa a difusa (DDR),  $\Gamma$ , y se expresa como:

$$\Psi = \frac{P_{diff}}{P_{diff} + P_{pw}} = \frac{1}{1 + \frac{P_{pw}}{P_{diff}}} = \frac{1}{1 + 10^{\Gamma/20}}$$

donde  $P_{pw}$  y  $P_{diff}$  son la onda plana y las potencias de difusividad, respectivamente, y  $\Gamma$  es la DDR expresada en escala de dB.

Desde luego, es posible que se utilice una o una combinación de las tres estrategias comentadas. La estrategia utilizada se selecciona por el procesador 54 dependiendo de los conjuntos de parámetros espaciales de audio recibidos. Para esto, los parámetros de audio se pueden analizar, de acuerdo con las realizaciones, para permitir la aplicación de diferentes estrategias de acuerdo con las características de la escena de audio y más particularmente de acuerdo con la difusividad.

Esto significa que, de acuerdo con las realizaciones, el procesador 54 está configurado para proporcionar ocultación de pérdida de paquetes para audio paramétrico espacial utilizando información direccional bien recibida previamente y tramado. De acuerdo con una realización adicional, el tramado es una función de la difusividad estimada o relación de energía entre componentes direccionales y no direccionales de la escena de audio. De acuerdo con las realizaciones, el tramado es una función de la tonalidad medida de la señal de mezcla de audio transmitida. Por lo tanto, el analizador lleva a cabo su análisis con base en la difusividad estimada, la relación de energía y/o una tonalidad.

En las Figuras 3a y 3b, la difusividad medida está dada en función de DDR simulando el campo difuso con  $N = 466$  ruidos rosas no correlacionados posicionados uniformemente en una esfera y la onda plana por un ruido rosa independiente colocado en un azimut de 0 grados y elevación de 0 grados. Se confirma que la difusividad medida en análisis de DirAC, es un buen estimado de la difusividad teórica si la longitud de la ventana de observación  $W$  es lo suficientemente larga. Esto implica que la difusividad tenga características a largo plazo, lo cual confirma que el parámetro se puede predecir bien en caso de pérdida de paquete al simplemente mantener el valor bien recibido previamente.

Por otro lado, la estimación de parámetros de dirección también se puede evaluar en función de difusividad verdadera, lo cual se reporta en la Figura 4. Se puede demostrar que la elevación y azimut estimados de la posición de onda plana se desvían de la posición real del suelo (0 grados de azimut y 0 grados de elevación) con una desviación estándar que aumenta con la difusividad. Para una difusividad de 1, la desviación estándar es de aproximadamente 90 grados para el ángulo de azimut definido entre 0 y 360 grados, que corresponden a un ángulo completamente aleatorio para una distribución uniforme. En otras palabras, el ángulo de azimut es entonces sin sentido. La misma observación se puede hacer para la elevación. En general, la precisión de la dirección estimada y su significación disminuye con la difusividad. Por lo tanto se espera que la dirección en DirAC fluctuará con el tiempo y se desviará de

su valor esperado con una función de varianza de la difusividad. Esta dispersión natural es parte del modelo de DirAC, que es esencial para una reproducción fiel de la escena de audio. De hecho, representar en una dirección constante el componente direccional de DirAC aún si la difusividad es alta, generará una fuente puntual que en realidad se debería percibir más amplia.

5 Por las razones expuestas anteriormente, proponemos aplicar un tramado en la dirección en lugar de la estrategia de retención. La amplitud del tramado se hace función de la difusividad y puede seguir, por ejemplo los modelos evocados en la Figura 4. Se pueden derivar dos modelos para la elevación y los ángulos medidos de elevación para lo cual se expresa la desviación estándar como:

10

$$\sigma_{azi} = 654^{3.8} + \sigma_{ele}$$

$$\sigma_{ele} = 33.254 + 1.25$$

15 El pseudocódigo de ocultación de parámetros de DirAC puede ser entonces:

```

para k in frame_start:frame_end
{
  si (bad_frame_indicador[k])
  {
    para band in band_start:band_end
    {
      diff_index = diffuseness_index[k-1][band];
      diffuseness[k][band] = unquantize_diffuseness(diff_index);

      azimuth_index[k][b] = azimuth_index[k-1][b];
      azimuth[k][b] = unquantize_azimuth(azimuth_index[k][b])
      azimuth[k][b] = azimuth[k][b] + random() * dithering_azi_scale[diff_index]

      elevation_index[k][b] = elevation_index[k-1][b];
      elevation[k][b] = unquantize_elevation(elevation_index[k][b])
      elevation[k][b] = elevation[k][b] + random() * dithering_ele_scale[diff_index]
    }
  }
  de lo contrario
  {
    para band in band_start:band_end
    {
      diffuseness_index[k][b] = read_diffuseness_index()
      azimuth_index[k][b] = read_azimuth_index()
      elevation_index[k][b] = read_elevation_index()

      diffuseness[k][b] = unquantize_diffuseness(diffuseness_index[k][b])
      azimuth[k][b] = unquantize_azimuth(azimuth_index[k][b])
      elevation[k][b] = unquantize_elevation(elevation_index[k][b])
    }

    output_frame[k] = Dirac_synthesis(diffuseness[k][b], azimuth[k][b],
    elevation[k][b])
  }
}

```

20 donde bad\_frame\_indicador[k] es un marcador que indica si la trama en el índice k se recibió bien o no. En el caso de una buena trama, se leen los parámetros de DirAC, se decodifican y descuantifican para cada banda de parámetro que corresponde a una gama de frecuencias dada. En caso de una mala trama, la difusividad se retiene directamente desde la última trama bien recibida en la misma banda de parámetros, mientras que el azimut y la elevación se derivan a partir de la descuantificación de los últimos índices bien recibidos con inyección de un valor aleatorio escalado por medio de una función de factor del índice de difusividad. La función random() genera un valor aleatorio de acuerdo con una distribución dada. El proceso aleatorio puede seguir, por ejemplo, una distribución normal estándar con media 25 cero y varianza unitaria. Alternativamente, puede seguir una distribución uniforme entre -1 y 1 o seguir una densidad de probabilidad de triángulo utilizando, por ejemplo, el siguiente pseudocódigo:

30

```

    random()
    {
        rand_val = uniform_random();

        if ( rand_val <= 0.5f )
        {
            return 0.5f * sqrt(rand_val + 1.0f) - 0.5f;
        }
        // lo contrario
        {
            return 0.5f - 0.5f * sqrt(1.0f - rand_val);
        }
    }
}

```

5 Las escalas de tramado son funciones del índice de difusividad heredado de la última trama bien recibida en la misma banda de parámetros y se puede derivar a partir de los modelos deducidos de la Figura 4. Por ejemplo, en el caso de que la difusividad se codifique en 8 índices, pueden corresponder a las siguientes tablas: tramado\_azi\_scale[8] = { 6.716062e-01f, 1.011837e+00f, 1.799065e+00f, 2.824915e+00f, 4.800879e+00f, 9.206031e+00f, 1.469832e+01f, 2.566224e+01f }; tramado\_ele\_scale[8] = { 6.716062e-01f, 1.011804e+00f, 1.796875e+00f, 2.804382e+00f, 4.623130e+00f, 7.802667e+00f, 1.045446e+01f, 1.379538e+01f };

10 Adicionalmente, también se puede dirigir la intensidad de tramado dependiendo de la naturaleza de la señal de mezcla de audio. De hecho, una señal muy tonal tiende a percibirse como una fuente más localizada que las señales no tonales. Por lo tanto, el tramado se puede ajustar en función de la tonalidad de la mezcla de audio transmitida, por medio de una disminución del efecto de tramado para elementos tonales. La tonalidad se puede medir, por ejemplo, en el dominio del tiempo calculando una ganancia de predicción a largo plazo o en el dominio de la frecuencia midiendo la planitud espectral.

15 Con respecto a las Figuras 6a y 6b, se comentarán realizaciones adicionales con referencia a un método para decodificar una escena de audio codificada por DirAC (ver la Figura 6a, método 200) y un decodificador 17 para una escena de audio codificada por DirAC (ver la Figura 6b).

20 La Figura 6a ilustra el nuevo método 200 que comprende las etapas 110, 120 y 130 del método 100 y una etapa adicional de decodificación 210. La etapa de decodificación permite la decodificación de una escena de audio codificada por DirAC que comprende una mezcla de audio (no mostrada) mediante el uso del primer conjunto de parámetros espaciales de audio y un segundo conjunto de parámetros espaciales de audio, en donde aquí, se utiliza el segundo conjunto reemplazado, proporcionado por la etapa 130. Este concepto se utiliza por medio del aparato 17, mostrado en la Figura 6b. La Figura 6b muestra un decodificador 70 que comprende el procesador para la ocultación de pérdidas de parámetros espaciales de audio 15 y un decodificador DirAC 72. El decodificador DirAC 72 o, a mayor detalle el procesador del decodificador DirAC 72, recibe una señal de mezcla de audio y los conjuntos de parámetros espaciales de audio, por ejemplo, directamente de la interfaz 52 y/o procesados por medio del procesador 52 de acuerdo con el planteamiento comentado anteriormente.

30 Aunque algunos aspectos se han descrito en el contexto de un aparato, es claro que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa de método o una característica de una etapa de método. Análogamente, los aspectos descritos en el contexto de una etapa de método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente. Algunos o todas las etapas de método se pueden ejecutar por (o utilizando) un aparato de hardware, como por ejemplo, un microprocesador, una ordenador programable o un circuito electrónico. En algunas realizaciones, alguna o más de las etapas de método más importantes pueden ejecutarse por un aparato de este tipo.

40 La señal de audio codificado inventiva se puede almacenar en un medio de almacenamiento digital o se puede transmitir en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión alámbrico tal como Internet.

45 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 utilizando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria flash, que tenga señales de control legibles electrónicamente almacenadas en el mismo, que coopere (o sea capaz de cooperar) con un sistema de cómputo programable de tal manera que se lleve a cabo el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

50 Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, que pueden actuar conjuntamente con un sistema informático programable, de tal manera que se realiza uno de los métodos descritos en el presente documento.

Generalmente, pueden implementarse realizaciones de la presente invención como un producto de programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los métodos cuando se ejecuta el producto de programa informático en un ordenador. El código de programa se puede almacenar, por ejemplo, en un portador legible por máquina.

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

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

15 Una realización adicional de los métodos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El portador de datos, el medio de almacenamiento digital o el medio grabado son generalmente tangibles y/o no transitorios.

20 Una realización adicional del método de la invención es, por consiguiente, una corriente de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales se puede configurar, por ejemplo, para transferirse por medio de una conexión de comunicación de datos, por ejemplo, a través de Internet.

25 Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado o adaptado para realizar uno de los métodos descritos en el presente documento.

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

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

35 En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en el campo) para realizar algunas o todas de las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programable en el campo puede actuar conjuntamente con un microprocesador con el fin de realizar uno de los métodos descritos en el presente documento. Generalmente, los métodos se llevan a cabo preferiblemente por medio de cualquier aparato de hardware.

40 Las realizaciones descritas anteriormente son simplemente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de los acomodos y los detalles descritos en este documento serán aparentes para otros experimentados en la técnica. La intención es, por tanto, quedar limitado solo por el alcance de las reivindicaciones de patente inminente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

## **Referencias**

50 [1] V. Pulkki, M-V. Laitinen, J. Vilkkamo, J. Ahonen, T. Lokki, y T. Pihlajamäki, "Directional audio coding - perception-based reproduction of spatial sound", (Codificación de audio direccional - reproducción basada en percepción de sonido espacial), International Workshop on the Principles and Application on Spatial Hearing, noviembre de 2009, Zao; Miyagi, Japón.

55 [2] V. Pulkki, "Virtual source positioning using vector base amplitude panning", (Posicionamiento de fuente virtual utilizando panoramización en amplitud a base de vectores), J. Audio Eng. Soc., 45(6):456-466, junio de 1997.

[3] J. Ahonen y V. Pulkki, "Diffuseness estimation using temporal variation of intensity vectors", (Estimación de difusividad utilizando variación temporal de vectores de intensidad), en Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA, Mohonk Mountain House, New Paltz, 2009.

60 [4] T. Hirvonen, J. Ahonen, y V. Pulkki, "Perceptual compression methods for metadata in Directional Audio Coding applied to audiovisual teleconference", (Métodos de compresión perceptual para metadatos en codificación de audio direccional aplicada a teleconferencia audiovisual), AES 126th Convention 2009, 7 a 10 de mayo, Múnich, Alemania.

[5] A. Politis, J. Vilkamo y V. Pulkki, "Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain", (Reproducción de campo de sonido paramétrico basada en sector en el dominio armónico esférico), en *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852-866, agosto de 2015.

## REIVINDICACIONES

1. Un método (100) para la ocultación de pérdida de parámetros espaciales de audio, los parámetros espaciales de audio comprenden al menos una información de dirección de llegada, el método comprende las siguientes etapas:
- 5 recibir (110) un primer conjunto de parámetros espaciales de audio que comprende al menos una primera información de dirección de llegada (azi1, ele1);
- 10 recibir (120) un segundo conjunto de parámetros espaciales de audio que comprende al menos una segunda información de dirección de llegada (azi2, ele2); y
- 15 reemplazar la segunda información de dirección de llegada (azi2, ele2) de un segundo conjunto por una información de dirección de llegada de reemplazo derivada a partir de la primera información de dirección de llegada (azi1, ele1) si se pierde o daña al menos la segunda información de dirección de llegada (azi2, ele2) o una porción de la segunda información de dirección de llegada (azi2, ele2);
- 20 el método se caracteriza por que la etapa de sustitución comprende la etapa de tramado inyectando ruido aleatorio a la primera dirección (azi1, ele1) de la información de llegada para obtener la dirección de sustitución de la información de llegada y en la que la etapa de inyección se realiza, si la primera o segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ) indica una alta difusividad; y/o
- 25 si la primera o segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ) está por encima de un umbral predeterminado para la información de difusividad,
- en el que el primer (1<sup>er</sup> conjunto) y el segundo (2<sup>o</sup> conjunto) conjunto de parámetros espaciales de audio comprenden una primera y una segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ), respectivamente.
- 30 2. Método (100) según la reivindicación 1, en el que la primera o una segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ) se deriva de al menos una relación de energía relacionada con al menos una información de dirección de llegada.
3. Método (100) de acuerdo con la reivindicación 1 o 2, en el que el método comprende además sustituir una segunda información de difusividad ( $\psi_2$ ) de un segundo conjunto (2<sup>o</sup> conjunto) por una información de difusividad de sustitución derivada de la primera información de difusividad ( $\psi_1$ ).
- 35 4. El método (100) de acuerdo con una de las reivindicaciones anteriores, en el que la dirección de sustitución de la información de llegada coincide con la primera dirección (azi1, ele1) de la información de llegada.
- 40 5. El método (100) de acuerdo con una de las reivindicaciones 1 a 4, en el que la información de difusividad comprende o se basa en una relación entre componentes direccionales y no direccionales de una escena de audio descrita por el primer (1<sup>er</sup> conjunto) y/o el segundo conjunto de (2<sup>o</sup> conjunto) parámetros espaciales de audio.
- 45 6. El método (100) de acuerdo con una de las reivindicaciones 1 a 5, en el que el ruido aleatorio a inyectar depende de la primera y/o segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ); y/o
- 50 en el que el ruido aleatorio que debe inyectarse se escala por un factor que depende de la primera y/o segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ).
7. El método (100) de acuerdo con una de las reivindicaciones 1 a 6, que comprende además la etapa de analizar la tonalidad de una escena de audio descrita por el primer (1<sup>er</sup> conjunto) y/o segundo conjunto (2<sup>o</sup> conjunto) de parámetros de audio espacial o de analizar la tonalidad de una mezcla descendente transmitida perteneciente al primer (1<sup>er</sup> conjunto) y/o segundo conjunto (2<sup>o</sup> conjunto) de parámetros de audio espacial para obtener un valor de tonalidad que describa la tonalidad; y
- 55 en el que el ruido aleatorio a inyectar depende del valor de tonalidad.
8. El método (100) de acuerdo con la reivindicación 7, en el que el ruido aleatorio se reduce en un factor decreciente junto con el inverso del valor de tonalidad o si la tonalidad aumenta.
- 60 9. El método (100) de acuerdo con una de las reivindicaciones anteriores, en el que el primer conjunto (1<sup>er</sup> conjunto) de parámetros espaciales de audio pertenecen a un primer punto en el tiempo y/o a un primer fotograma y en el que el segundo conjunto (2<sup>o</sup> conjunto) de parámetros espaciales de audio pertenecen a un

segundo punto en el tiempo y/o a un segundo fotograma; o bien

en el que el primer conjunto (1<sup>er</sup> conjunto) de parámetros espaciales de audio pertenece a un primer punto en el tiempo y en el que el segundo punto en el tiempo es posterior al primer punto en el tiempo o en el que el segundo fotograma es posterior al primer fotograma.

- 5
10. El método (100) de acuerdo con una de las reivindicaciones anteriores, en el que el primer conjunto (1<sup>er</sup> conjunto) de parámetros de audio espacial comprende un primer subconjunto de parámetros de audio espacial para una primera banda de frecuencias y un segundo subconjunto de parámetros de audio espacial para una segunda banda de frecuencias; y/o
- 10

en el que el segundo conjunto (2<sup>o</sup> conjunto) de parámetros de audio espacial comprende otro primer subconjunto de parámetros de audio espacial para la primera banda de frecuencias y otro segundo subconjunto de parámetros de audio espacial para la segunda banda de frecuencias.

- 15
11. Un método (200) para decodificar una escena de audio codificada con DirAC, que comprende las siguientes etapas:

20

decodificación de la escena de audio codificada con DirAC que comprende una mezcla descendente, un primer conjunto de parámetros de audio espacial y un segundo conjunto de parámetros de audio espacial;

realización del método (100) para la ocultación de pérdidas de parámetros espaciales de audio como se define en una de las reivindicaciones 1-11.

- 25
12. Medio de almacenamiento digital legible por ordenador que tiene almacenado en el mismo un programa informático que tiene un código de programa para realizar, cuando se ejecuta en un ordenador, un método (100, 200) de acuerdo con una de las reivindicaciones anteriores.

- 30
13. Aparato de ocultación de pérdidas (50) para la ocultación de pérdida de parámetros espaciales de audio, los parámetros espaciales de audio comprenden al menos una información de dirección de llegada, el aparato comprende:

35

un receptor (52) para recibir (110) un primer conjunto de parámetros de audio espacial que comprende una primera dirección (azi1, ele1) de información de llegada y para recibir (120) un segundo conjunto de parámetros de audio espacial que comprende una segunda dirección (azi2, ele2) de información de llegada;

40

un procesador (54) configurado para sustituir la segunda dirección de llegada (azi2, ele2) del segundo conjunto por una información de dirección de llegada de reemplazo derivada a partir de la primera información de dirección de llegada (azi1, ele1) si se pierde o daña al menos la segunda información de dirección de llegada (azi2, ele2) o una porción de la segunda información de dirección de llegada (azi2, ele2);

45

en donde la sustitución comprende la etapa de tramado inyectando ruido aleatorio a la primera dirección (azi1, ele1) de la información de llegada para obtener la dirección de sustitución de la información de llegada y en la que la etapa de inyección se realiza, si la primera o segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ) indica una alta difusividad; y/o si la primera o segunda información de difusividad ( $\psi_1$ ,  $\psi_2$ ) está por encima de un umbral predeterminado para la información de difusividad;

50

en el que el primer (1<sup>er</sup> conjunto) y el segundo (2<sup>o</sup> conjunto) conjunto de parámetros espaciales de audio comprenden una primera y una segunda información de difusividad ( $\Psi_1$ ,  $\Psi_2$ ), respectivamente.

14. Un decodificador (70) para una escena de audio codificada en DirAC que comprende el aparato de ocultación de pérdidas de acuerdo con la reivindicación 13.

10

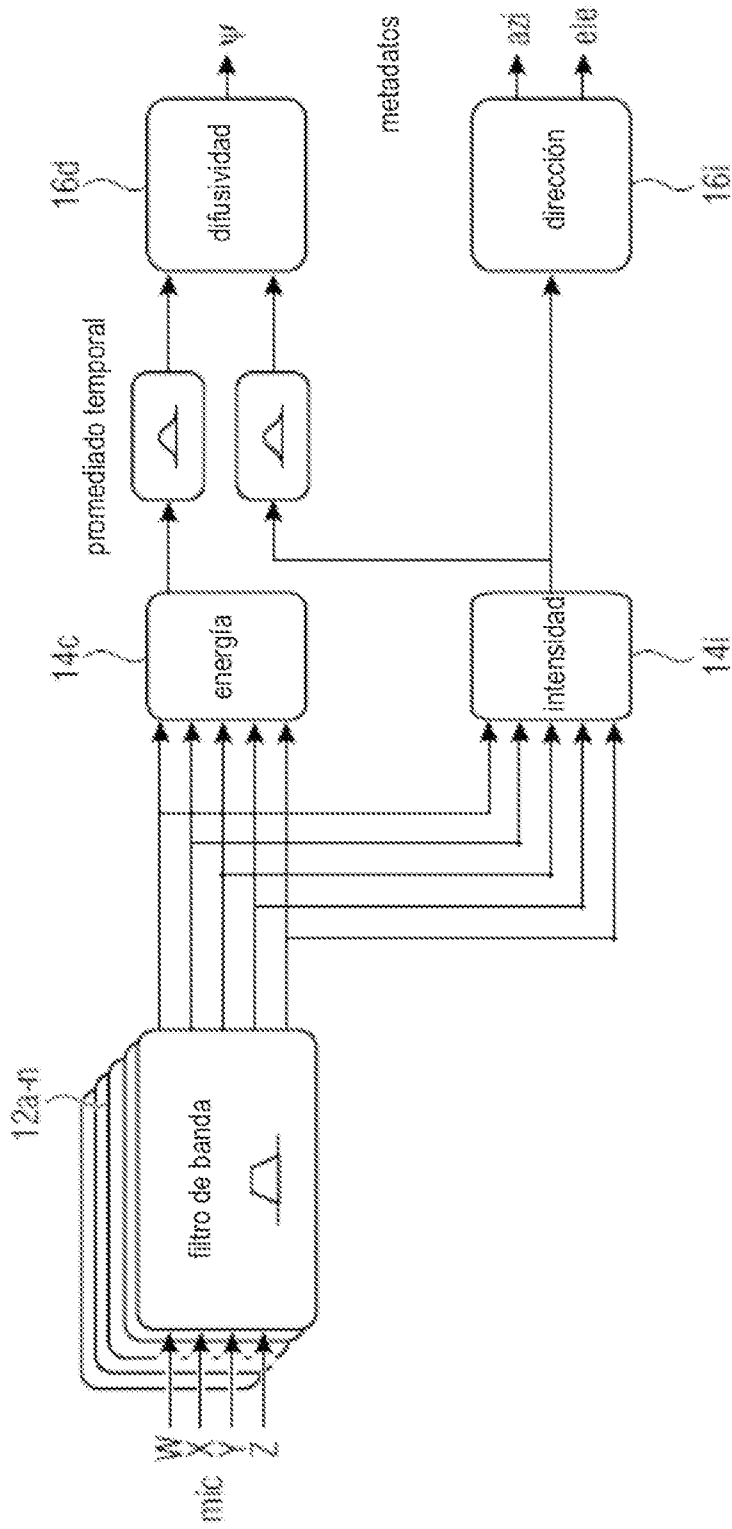


Fig. 1a

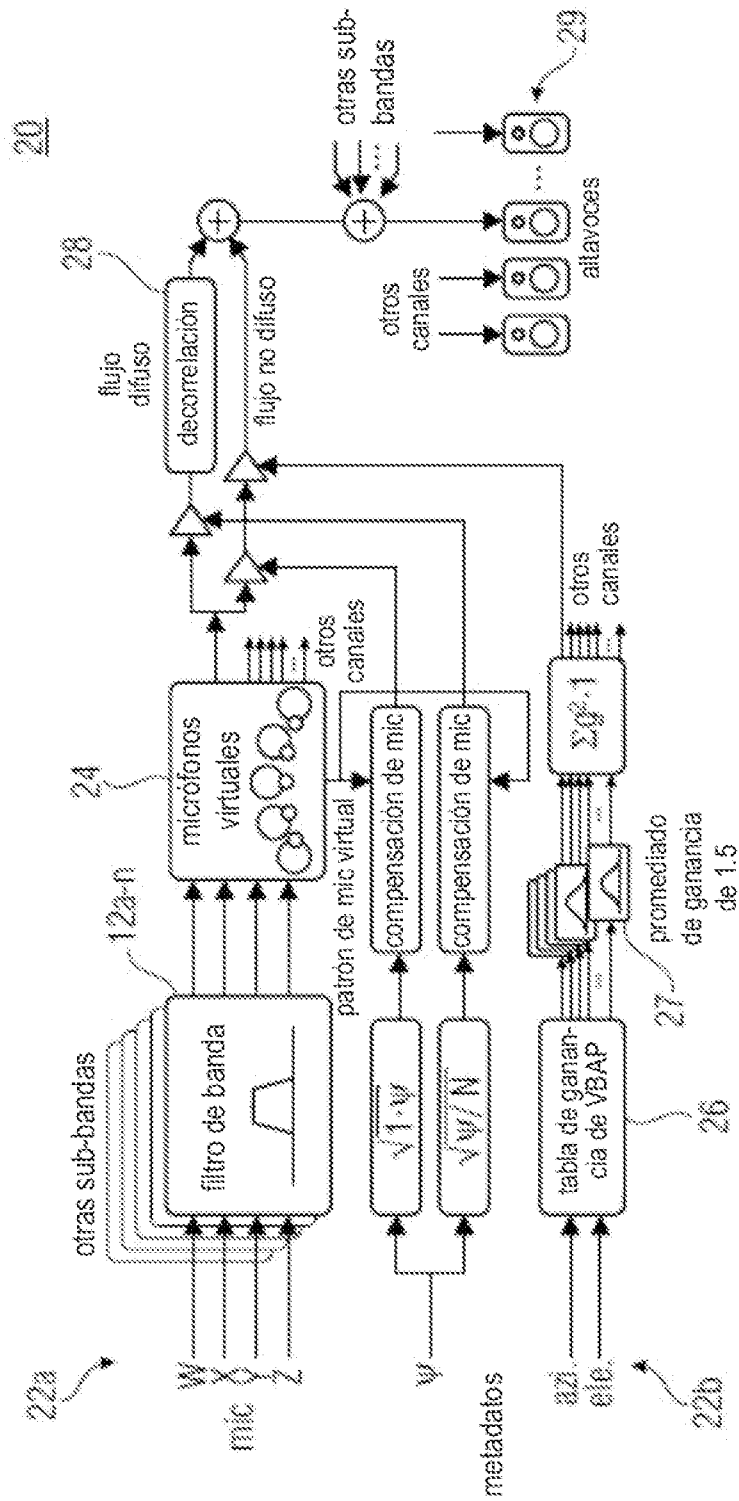


Fig. 1b

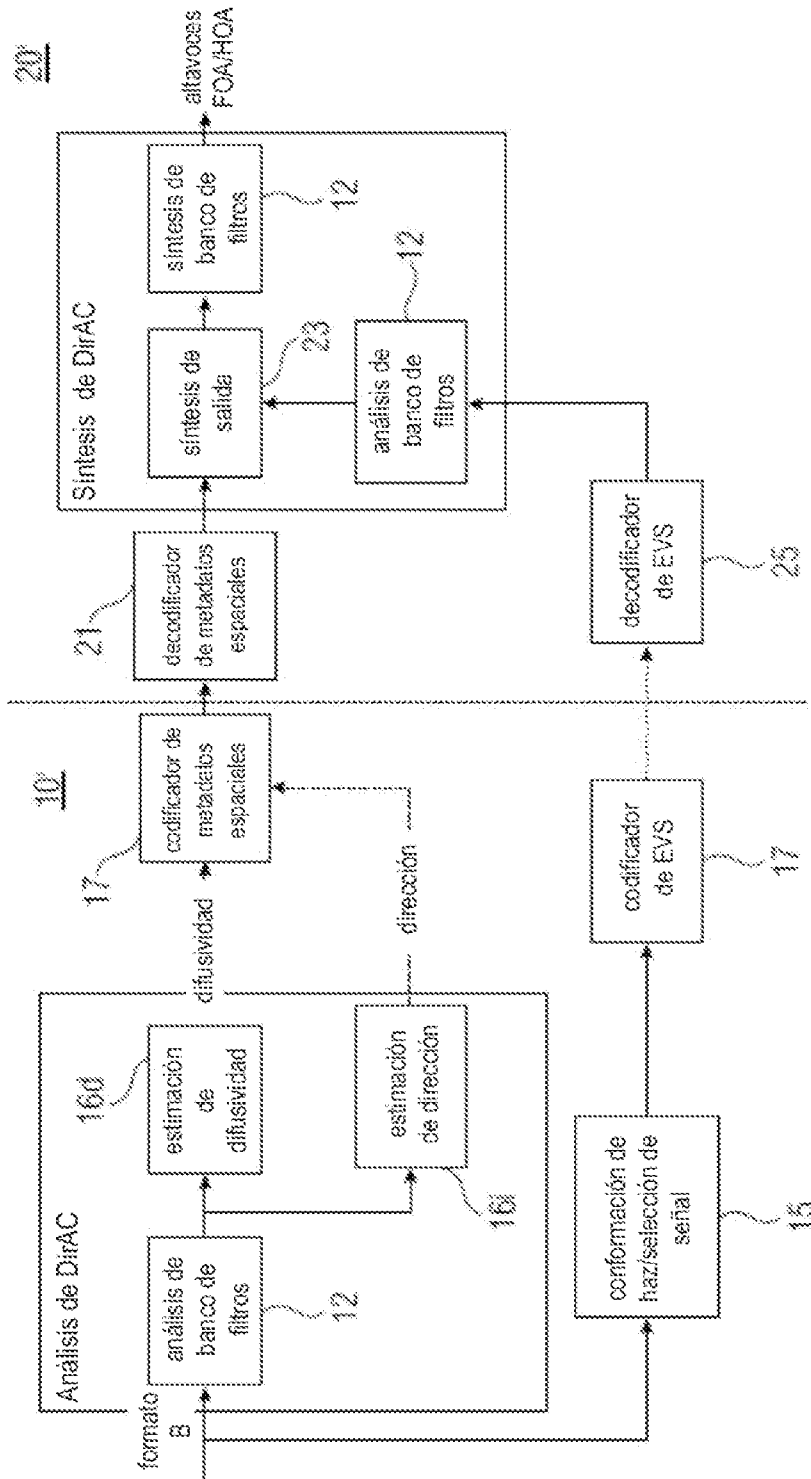


Fig. 2

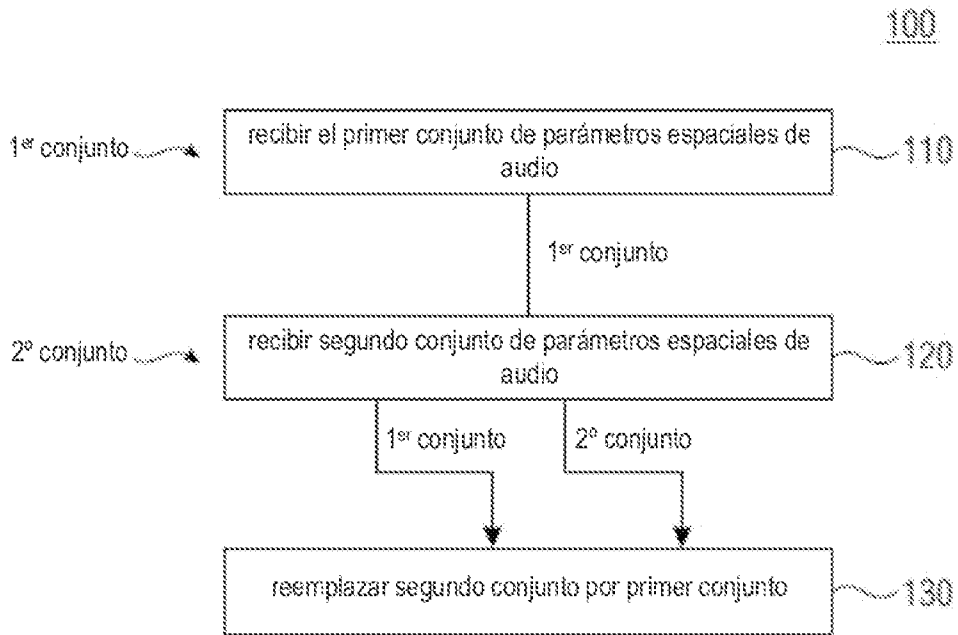


Fig. 3a

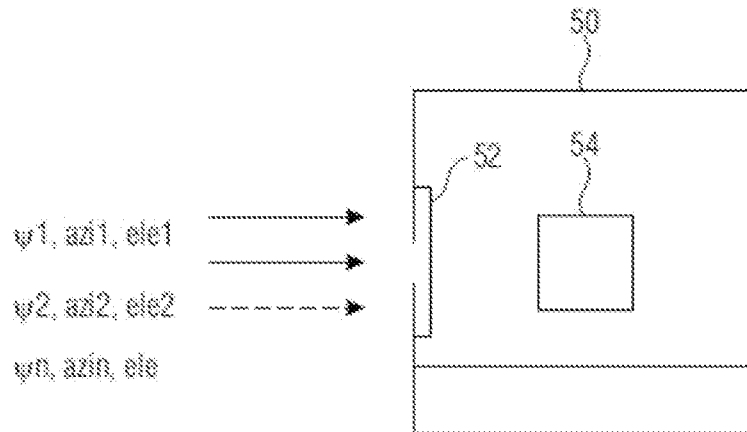


Fig. 3b

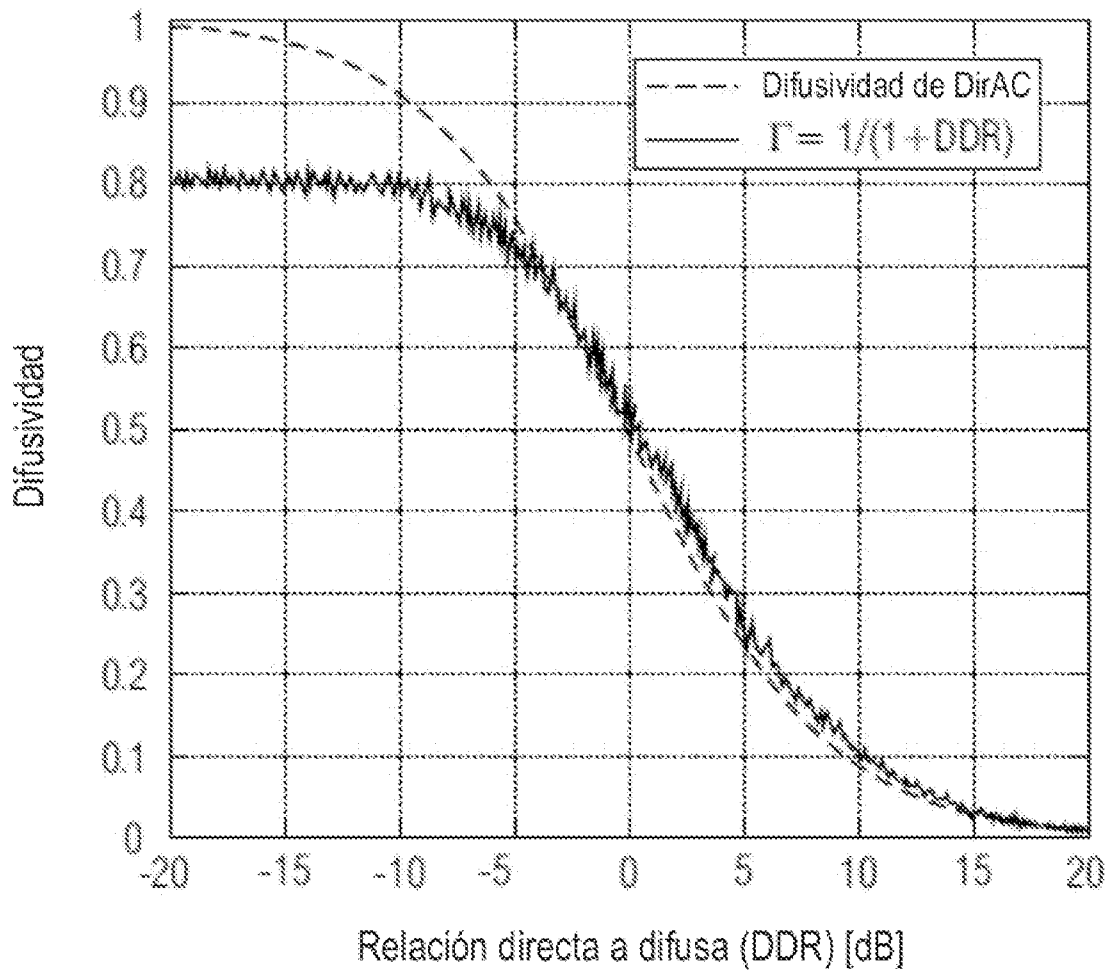


Fig. 4a

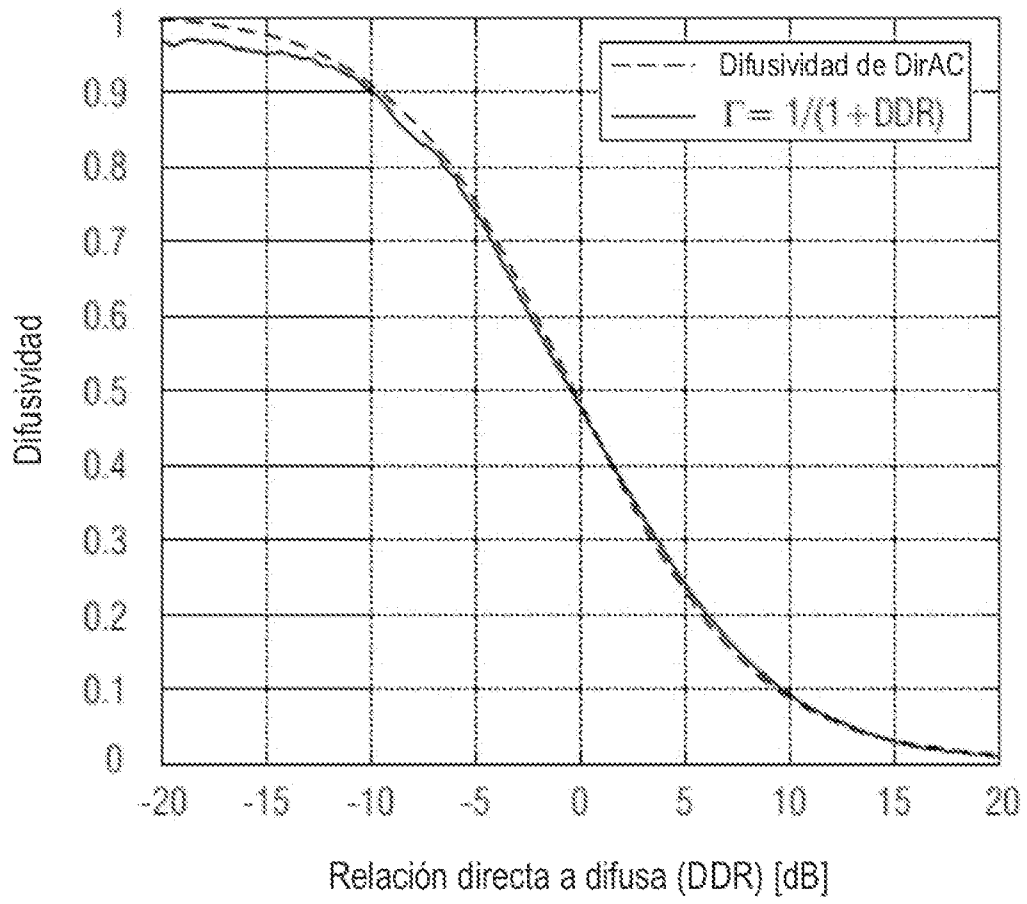


Fig. 4b

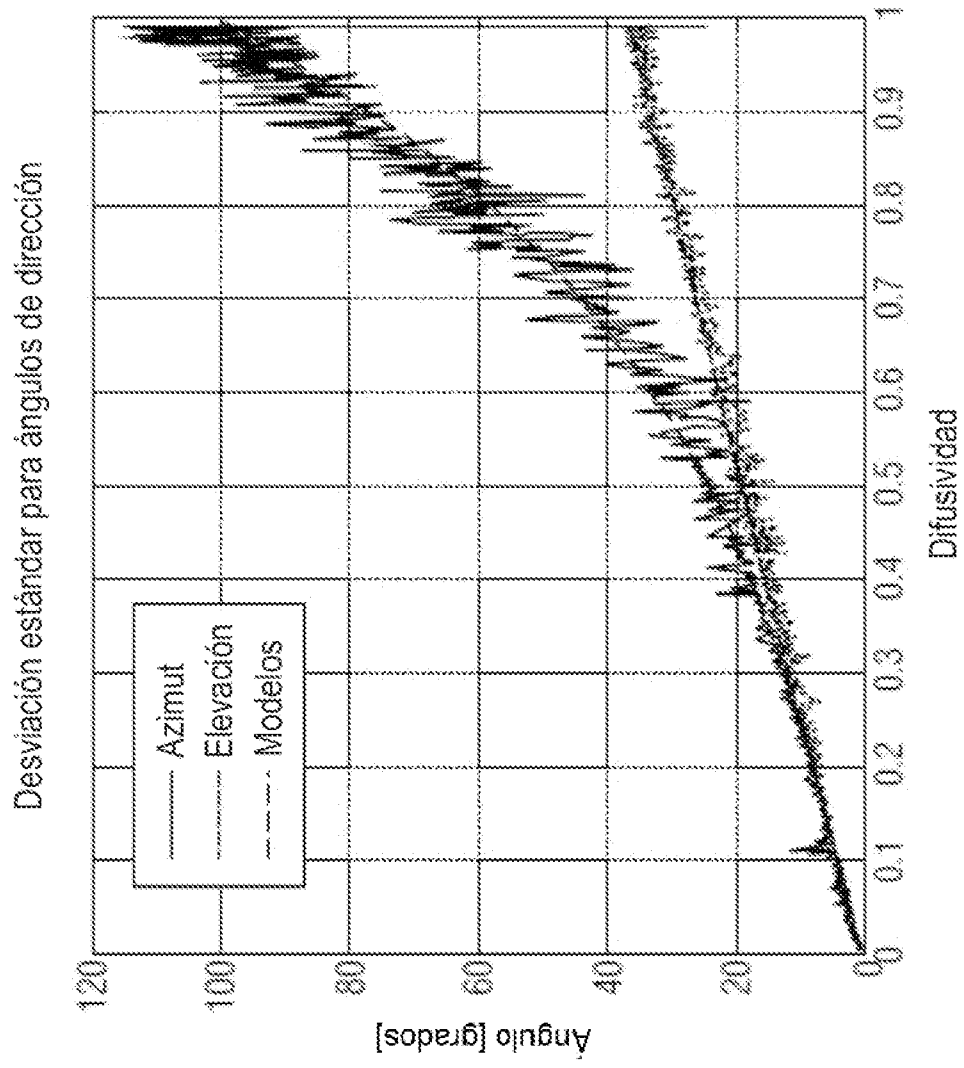


Fig. 5

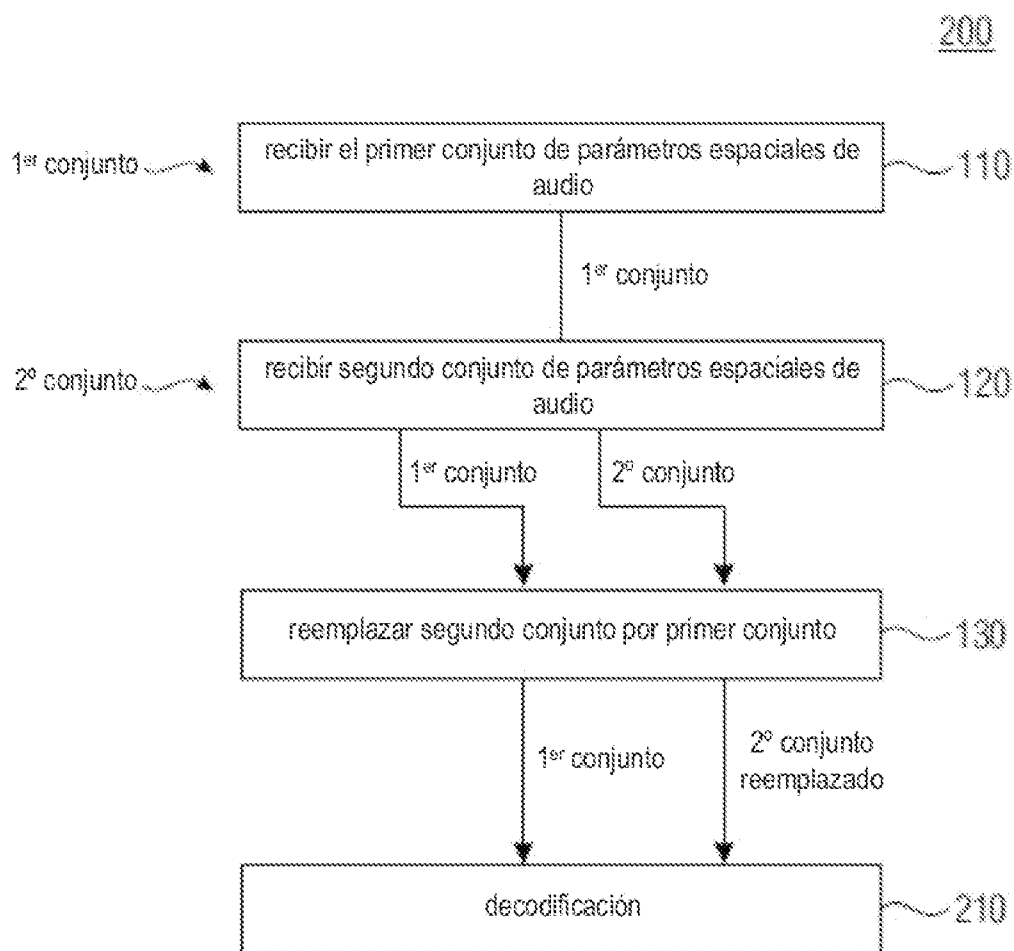


Fig. 6a

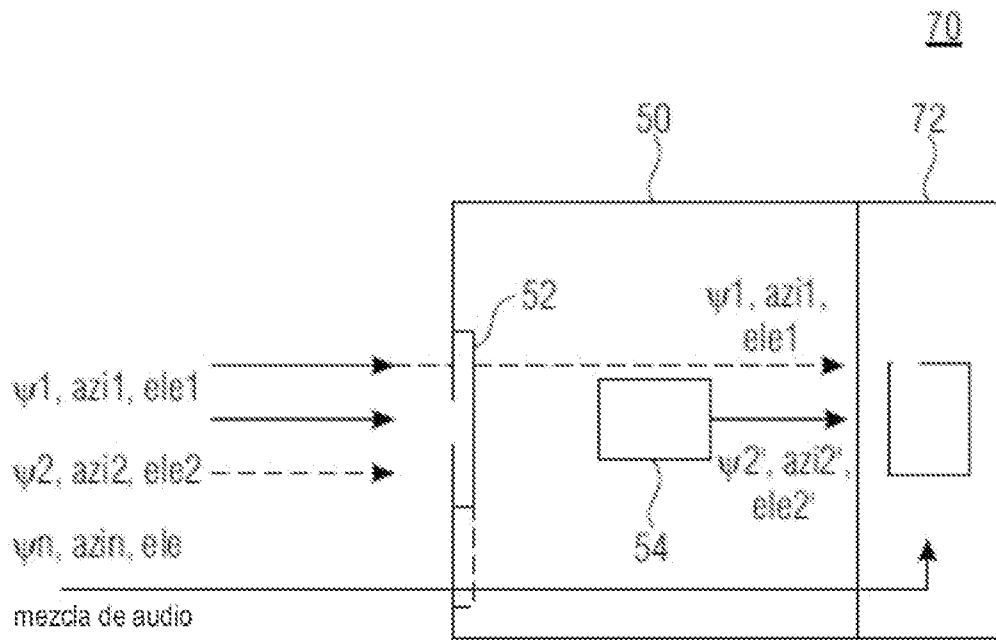


Fig. 6b