

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 523 793**

51 Int. Cl.:

G10H 1/00 (2006.01)

H04S 3/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **12.08.2009 E 09806394 (4)**

97 Fecha y número de publicación de la concesión europea: **30.07.2014 EP 2311026**

54 Título: **Aparato para determinar una señal de audio espacial convertida**

30 Prioridad:

13.08.2008 US 88513 P

25.08.2008 US 91682 P

02.02.2009 EP 09001398

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

01.12.2014

73 Titular/es:

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

72 Inventor/es:

**DEL GALDO, GIOVANNI;
KUECH, FABIAN;
KALLINGER, MARKUS;
PULKKI, VILLE;
LAITINEN, MIKKO-VILLE y
SCHULTZ-AMLING, RICHARD**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 523 793 T3

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

DESCRIPCIÓN

Aparato para determinar una señal de audio espacial convertida.

5 [0001] La presente invención es concerniente con el campo de procesamiento de audio, especialmente procesamiento de audio espacial y conversión de diferentes formatos de audio espacial.

10 [0002] La codificación de audio de DirAC (DirAC = Codificación de Audio Direccional) es un procesamiento para la reproducción y procesamiento de audio espacial. Los sistemas convencionales aplican DirAC en la reproducción de alta calidad bidimensional y tridimensional de sonido grabado, aplicaciones de teleconferencia, micrófonos direccionales y mezcla ascendente de estéreo a surround, por ejemplo V. Pulkki y C. Faller, Directional audio coding: Filterbank and STFT-based design, in 120th AES Convention, May 20-23, 2006, Paris, France May 2006, V. Pulkki and C. Faller, Directional audio coding in spatial sound reproduction and stereo upmixing, in AES 28th International Conference, Pitea, Sweden, June 2006, V. Pulkki, Spatial sound reproduction with directional audio coding, Journal of the Audio Engineering Society, 55(6):503-516, June 2007, Jukka Ahonen, V. Pulkki and Tapio Lokki, Teleconference application and B-format microphone array for directional audio coding, in 30th AES International Conference.

20 [0003] Otras aplicaciones convencionales que usan DirAC son, por ejemplo, el formato de codificación universal y cancelación de ruido. En DirAC, algunas propiedades direccionales del sonido son analizadas en bandas de frecuencia dependientes del tiempo. Los datos de análisis son transmitidos junto con los datos de audio y sintetizados para diferentes propósitos. El análisis se hace comúnmente utilizando señales de formato B, aunque teóricamente DirAC no está limitado a este formato. El formato B, ver Michael Gerzon, Surround sound psychoacoustics, in Wireless World, volume 80, pages 483-486, December 1974, fue desarrollado en el trabajo de Ambisonics, un sistema desarrollado por investigadores británicos en los 70's para traer el sonido surround a salas de concierto a salas de estar. El formato B consiste de cuatro señales, es decir $w(t), x(t), y(t)$ y $z(t)$. Las primeras tres corresponden a la presión medida por un micrófono omnidireccional, mientras que las últimas tres son lecturas de presión de micrófonos que tienen patrones de captación de cifra de ocho dirigidos hacia los tres ejes de un sistema de coordenadas Cartesiano. Las señales $x(t), y(t)$ y $z(t)$ son proporcionales a los componentes del vector de velocidad de partículas dirigidos hacia x, y y z respectivamente.

35 [0004] La corriente de DirAC consiste de 1-4 canales de audio con metadatos direccionales. En teleconferencia y en algunos otros casos, la corriente consiste de solamente un solo canal de audio con metadatos, llamado una corriente de DirAC mono. Esta es una manera muy compacta para describir audio espacial, ya que solamente un solo canal de audio necesita ser transmitido junto con información lateral, que por ejemplo, da buena separación espacial entre los parlantes. Sin embargo, en tales casos algunos tipos de sonido, tales como escenarios de sonido reverberado o sonido ambiente pueden ser reproducidos con calidad limitada. Para producir mejor calidad en estos casos, canales de audio adicionales necesitan ser transmitidos.

40 [0005] La conversión del formato B a DirAC es descrita en V. Pulkki, A method for reproducing natural or modified spatial impression in multichannel listening, Patent WO 2004/077884 A1, September 2004. La Codificación de Audio Direccional es un procedimiento eficiente para el análisis y reproducción de sonido espacial. DirAC utiliza una representación paramétrica de campos de sonido basados en los elementos que son relevantes para la percepción de sonido espacial, es decir la DOA (DOA = dirección de llegada) y difusividad del campo de sonido en sub-bandas de frecuencia. En efecto, DirAC supone que las diferencias de tiempo interaurales (ITD) y diferencias de nivel interaural (ILD) son percibidas correctamente cuando la DOA de un campo de sonido es reproducida correctamente, mientras que la coherencia interaural (IC) es percibida correctamente, si la difusividad es reproducida exactamente. Estos parámetros, es decir DOA y difusividad, representan información lateral que acompaña una señal mono en lo que es denominada como una corriente de DirAC mono.

55 [0006] La Figura 7 muestra el codificador de DirAC, del cual señales de micrófono apropiadas calcula un canal de audio mono e información lateral, es decir difusividad $\psi(k, n)$ y dirección de llegada $e_{DOA}(k, n)$. La Figura 7 muestra un codificador de DirAC 200, que es apto para calcular un canal de audio mono e información lateral de señales de micrófono apropiadas. En otras palabras, la Figura 7 ilustra un codificador de DirAC 200 para determinar la difusividad y dirección de llegada de señales de micrófono. La Figura 7 muestra un codificador de DirAC 200 que comprende una unidad de estimación P/U 210, en donde $P(k, n)$ representa una señal de presión y $U(k, n)$ representa un vector de velocidad de partícula. La unidad de estimación de P/U recibe las señales de micrófono como información de entrada, en la cual la estimación de P/U está basada. Una etapa de análisis energético 220 permite la estimación de la dirección de llegada y el parámetro de a difusividad de la corriente de DirAC mono.

60 [0007] Los parámetros de DirAC, como por ejemplo una representación de audio mono $W(k, n)$, un parámetro de difusividad $\psi(k, n)$ y una dirección de llegada (DOA) $e_{DOA}(k, n)$, pueden ser obtenidos a partir de una representación de frecuencia-tiempo de las señales de micrófono. Por consiguiente, los parámetros son dependientes del tiempo y de la frecuencia. En el lado de reproducción, esta información permite la presentación espacial exacta. Para recrear el sonido espacial en una posición de escucha deseada, se requiere un montaje de múltiples altavoces. Sin

embargo, su geometría puede ser arbitraria. En efecto, los canales de los altavoces pueden ser determinados como función de los parámetros de DirAC.

5 **[0008]** Hay diferencias sustanciales entre DirAC y codificación de audio de multicanal paramétrica, tal como MPEG Surround, véase Lars Villemoos, Juergen Herre, Jeroen Breebaart, Gerard Hotho, Sascha Disch, Heiko Purnhagen, and Kristofer Kjrling, MPEG surround: The forthcoming ISO standard for spatial audio coding, in AES 28th International Conference, Pitea, Sweden, June 2006, aunque comparten estructuras de procesamiento similares. Mientras que MPEG Surround está basado en un análisis de tiempo/frecuencia de los diferentes canales de altavoces, DirAC toma como entrada los canales de micrófonos coincidentes, que describen efectivamente el campo de sonido en un punto. Así, DirAC también representa una técnica de grabación eficiente para audio espacial.

10 **[0009]** Otro sistema que trata con audio espacial es SAOC (SAOC = Codificación de Objeto de Audio Espacial), véase Jonas Engdegard, Barbara Resch, Cornelia Falch, Oliver Hellmuth, Johannes Hilpert, Andreas Hoelzer, Leonid Terentiev, Jeroen Breebaart, Jeroen Koppens, Erik Schuijers, and Werner Oomen, Spatial audio object (SAOC) the upcoming MPEG standard on parametric object based audio coding, in 12th AES Convention, May 17-20, 2008, Amsterdam, The Netherlands, 2008, actualmente bajo normalización de ISO/MPEG. Se integra sobre el motor de presentación de MPEG Surround y trata diferentes fuentes de sonido como objetos. Esta codificación de audio ofrece eficiencia muy alta en términos de proporción de bits y da libertad sin precedentes de interacción en el lado de reproducción. Este procedimiento promete nuevos elementos componentes y funcionalidad en sistemas heredados, también como varias otras nuevas aplicaciones.

15 **[0010]** US 2006/045275 A1 da a conocer un procedimiento para procesar datos de audio y un dispositivo de adquisición de sonido que lleva a cabo este procedimiento. El procedimiento comprende la codificación de señales que representan un sonido que se propaga en el espacio tridimensional utilizando componentes expresados en una base armónica esférica, y la aplicación de una compensación de un efecto de campo próximo a estos componentes.

20 **[0011]** US 6.259.759 B1 describe un procedimiento y aparato para procesar audio especializado, donde se aplica la función de transferencia relacionada con al menos una cabeza a cada componente espacial para producir una serie de señales de transmisión. Las señales de transmisión se transmiten a múltiples usuarios, donde se determina una orientación actual de un usuario actual.

25 **[0012]** La publicación técnica " Spatial Sound Reproduction with Directional Audio Coding ", V. Pulkki, J. Audio Ing. Soc, Vol. 55, No. 6, junio de 2007 da a conocer detalles sobre la codificación de audio direccional (Dir Ac).

30 **[0013]** La publicación técnica " A distributed system for the creation and delivery of ambisonic surround sound audio " R. Foss et al., AES, 16a Conferencia Internacional, 1 de enero de 1999 describe un sistema para la producción de composiciones de sonido envolvente ambisónico usando una arquitectura cliente-surfer.

35 **[0014]** La publicación técnica " Realtime Room Acoustics Using Ambisonics " J. Papa et al., AES, 16a Conferencia Internacional, marzo de 1999, da a conocer una técnica de dos etapas para la simulación acústica de la sala. La primera etapa se basa en un modelo exacto de la respuesta de impulso de la sala y una etapa en tiempo real realizada posteriormente.

40 **[0015]** Es el objeto de la presente invención proveer un concepto mejorado para el procesamiento espacial.

45 **[0016]** El objetivo es obtenido por un aparato para determinar una señal de audio espacial convertida de acuerdo con la reivindicación 1 y un procesamiento correspondiente de acuerdo con la reivindicación 13.

50 **[0017]** La presente invención está basada en el descubrimiento de que el procesamiento espacial mejorado puede ser obtenido, por ejemplo cuando se convierte una señal de audio espacial codificada como una corriente de DirAC mono a una señal de formato B. En modalidades, la señal de formato B convertida puede ser procesada o presentada antes de ser agregada a algunas otras señales de audio y codificada de regreso a una corriente de DirAC. Algunas modalidades pueden tener diferentes aplicaciones, por ejemplo, mezcla de tipos diferentes de corriente de DirAC y corrientes de formato B, basados en DirAC, etc. Algunas modalidades pueden introducir una operación inversa a WO 2004/077884 A1, es decir la conversión de una corriente de DirAC mono a formato B.

55 **[0018]** La presente invención está basada en el descubrimiento de que el procesamiento mejorado puede ser obtenido, si las señales de audio son convertidas a componentes direccionales. En otras palabras, es el descubrimiento de la presente invención de que el procesamiento espacial mejorado puede ser obtenido, cuando el formato de una señal de audio espacial corresponde a componentes direccionales tal como son grabados, por ejemplo, mediante un micrófono direccional de formato B. Además, es un descubrimiento de la presente invención que los compuestos direccionales u omnidireccionales de fuentes diferentes pueden ser procesados conjuntamente y con los mismos con una eficiencia incrementada. En otras palabras, especialmente cuando se procesan señales de audio espacial a partir de múltiples fuentes de audio, el procesamiento se puede llevar a cabo más eficientemente, si las señales de las múltiples fuentes de audio están disponibles en el formato de sus componentes omnidireccionales y direccionales, ya que estos pueden ser procesados conjuntamente. En algunas modalidades,

por consiguiente, los generadores de efecto de audio o procesadores de audio pueden ser usados más eficientemente mediante el procesamiento de componentes combinados de múltiples fuentes.

5 **[0019]** En algunas modalidades, las señales de audio espacial pueden ser representadas como una corriente de DirAC mono que denota una técnica de flujo de DirAC en donde los datos de los medios están acompañados por solamente un canal de audio en la transmisión. Este formato puede ser convertido, por ejemplo, a una corriente de formato B, que tiene múltiples componentes direccionales. Algunas modalidades pueden habilitar el procesamiento espacial mejorado al convertir las señales de audio espacial a componentes direccionales.

10 **[0020]** Algunas modalidades pueden proveer una ventaja con respecto a la descodificación de DirAC, en donde solamente un canal de audio es usado para crear todas las señales de altavoz, en que el procesamiento espacial adicional es habilitado en base a los componentes de audio direccionales, que son determinados antes de crear las señales de altavoz. Algunas modalidades pueden proveer la ventaja de que los problemas en la creación de sonidos reverberantes son reducidos.

15 **[0021]** En modalidades, por ejemplo, una corriente de DirAC puede usar una señal de audio estéreo en una señal de audio mono, en donde los canales estéreo son L (L = canal estéreo izquierdo) y R (R = canal de estéreo derecho) y son transmitidos para ser usados en la descodificación de DirAC. Algunas modalidades pueden obtener una mejor calidad para sonido reverberante y proveer una compatibilidad directa con los sistemas de altavoz estéreo, por ejemplo.

20 **[0022]** Algunas modalidades pueden proveer la ventaja de que la descodificación de DirAC de micrófono virtual puede ser habilitada. Detalles en cuanto a descodificación de DirAC de micrófono virtual se pueden encontrar en V. Pulkki, Spatial sound reproduction with directional audio coding, Journal of the Audio Engineering Society, 55 (6): 503-516, June 2007. Estas modalidades obtienen las señales de audio para los altavoces colocando micrófonos virtuales orientados hacia la posición de los altavoces y que tienen fuentes de sonido semejantes a punto, cuya posición es determinada por los parámetros de DirAC. Algunas modalidades pueden proveer la ventaja de que, mediante la conversión, se puede habilitar la combinación lineal conveniente de las señales de audio.

30 **[0023]** Las modalidades de la presente invención serán detalladas usando las figuras adjuntas, en las cuales:

La Figura 1a muestra una modalidad de un aparato para determinar una señal de audio espacial convertida;

35 La Figura 1b muestra los componentes de presión de un vector de velocidad de partículas en un plano Gaussiano para una onda plana;

La Figura 2 muestra otra modalidad para convertir una corriente de DirAC mono a una señal de formato B;

40 La Figura 3 muestra una modalidad para combinar múltiples señales de audio espacial convertidas;

Las Figuras 4a-4d muestran modalidades para combinar múltiples señales de audio espacial basadas en DirAC aplicando diferentes efectos de audio;

45 La Figura 5 ilustra una modalidad de un generador de efecto de audio;

La Figura 6 muestra una modalidad de un generador de efecto de audio que aplica múltiples efectos de audio sobre componentes direccionales; y

50 La Figura 7 muestra un codificador de DirAC del estado del arte.

La Figura 1a muestra un aparato 100 para determinar una señal de audio espacial convertida, la señal de audio espacial convertida tiene un componente omnidireccional y por lo menos un componente direccional (X;Y;Z), a partir de una señal de audio espacial de entrada, la señal de audio espacial de entrada tiene una representación de audio de entrada (W) y una dirección de llegada de entrada (ϕ).

55 **[0024]** El aparato 100 comprende un estimador 110 para estimar una representación de ola que comprende una medida de campo de onda y una medida de dirección de llegada de onda basada en la representación de audio de entrada (W) y la dirección de llegada de entrada (ϕ). Además, el aparato 100 comprende un procesador 120 para procesar la medida de campo de onda y la medida de dirección de llegada de onda para obtener el componente omnidireccional y el por lo menos un componente direccional. El estimador 110 puede ser apto para estimar la representación de onda como una representación de onda plana.

60 **[0025]** En algunas modalidades, el procesador puede ser apto para proveer la representación de audio de entrada (W) como el componente de audio omnidireccional (W). En otras palabras, el componente de audio omnidireccional W' puede ser igual a la representación de audio de entrada W. Por consiguiente, de acuerdo con las líneas de puntos de la Figura 1a, la representación de audio de entrada puede omitir el estimador 110, el procesador 120 o

ambos. En otras modalidades, el componente de audio omnidireccional W puede estar basado en la intensidad de onda y la dirección de llegada de onda siendo procesada por el procesador 120 junto con la representación de audio de entrada W. En algunas modalidades, múltiples componentes de audio direccionales (X; Y; Z) pueden ser procesados, por ejemplo un primero (X), un segundo (Y) y/o un tercero (Z) componente de audio direccional correspondiente a diferentes direcciones espaciales. En algunas modalidades, por ejemplo tres componentes de audio direccional diferentes (X; Y; Z) pueden ser derivados de acuerdo con las diferentes direcciones de un sistema de coordenadas Cartesiano.

[0026] El estimador 110 puede ser apto para estimar la medida de campo de onda en términos de una amplitud de campo de onda y una fase de campo de onda. En otras palabras, en algunas modalidades, la medida de campo de onda puede ser estimada como una cantidad de valor complejo. La amplitud de campo de onda puede corresponder a una magnitud de presión de sonido y la fase de campo de onda puede corresponder a una fase de presión de sonido en algunas modalidades.

[0027] En algunas modalidades, la medida de dirección de llegada de onda puede corresponder a cualquier cantidad direccional, expresada por ejemplo por un vector, uno o más ángulos, etc. y puede ser derivada de cualquier medida direccional que representa un componente de audio como por ejemplo, un vector de intensidad, un vector de velocidad de partícula, etc. La medida de campo de onda puede corresponder a cualquier cantidad física que describe un componente de audio, que puede ser de valor real o valor complejo, corresponder a una señal de presión, una amplitud de magnitud de velocidad de partícula, volumen, etc. Además, las medidas pueden ser consideradas en el dominio de tiempo y/o dominio de frecuencia.

[0028] Algunas modalidades pueden estar basadas en la estimación de una representación de onda plana para cada una de las corrientes de entrada, que se puede llevar a cabo por el estimador 110 en la Figura 1a. En otras palabras, la medida de campo de onda puede ser modelada utilizando una representación de onda plana. En general, existen varias descripciones exhaustivas (esto es, completas) equivalentes de una onda u ondas planas en general. En lo siguiente, se presentará una descripción matemática para calcular los parámetros de difusividad y direcciones de llegada o medidas de dirección para diferentes componentes. Aunque solamente unas pocas descripciones son concernientes directamente con cantidades físicas, como por ejemplo presión, velocidad de partícula, etc., potencialmente existen un número infinito de diferentes maneras para describir representaciones de onda, de las cuales una será presentado como ejemplo subsecuentemente, sin embargo, no se propone ser limitante de ninguna manera a las modalidades de la presente invención. Cualquier combinación puede corresponder a la medida de campo de onda y la medida de dirección de llegada de onda.

[0029] Con el fin de detallar adicionalmente diferentes descripciones potenciales se consideran dos números reales a y b. La información contenida en a y b puede ser transferida al enviar c y d, cuando

$$\begin{bmatrix} c \\ d \end{bmatrix} = \Omega \begin{bmatrix} a \\ b \end{bmatrix},$$

en donde Ω es una matriz de 2x2 conocida. El ejemplo considera solamente combinaciones lineales, en general cualquier combinación, esto es, también una combinación no lineal es concebible.

[0030] En lo siguiente, los escalares son representados por letras minúsculas *a,b,c*, mientras que los vectores de columna son representados por letras minúsculas en negritas ***a,b,c***. El superíndice $(\cdot)^T$ denota las transpuesta, respectivamente, mientras que $(\cdot)^*$ y $(\cdot)^{\dagger}$ denotan conjugación completa. La notación de fasor complejo es distinguida de la temporal. Por ejemplo, la presión $p(t)$, que es un número real y del cual una medida de campo de onda posible puede ser derivada, puede ser expresada por medio del fasor P , que es un número complejo y del cual otra medida de campo de onda posible puede ser derivada por

$$p(t) = \text{Re}\{P e^{j\omega t}\},$$

en donde $\text{Re}\{\cdot\}$ denota la parte real y $\omega = 2\pi f$ es la frecuencia angular. Además, las letras mayúsculas usadas para cantidades físicas representan fasores en lo siguiente. Para la siguiente notación ejemplar introductoria y para evitar confusión, por favor nótese que todas las cantidades con el subíndice "PW" se refieren a ondas planas.

[0031] Para una onda plana monocromática ideal, el vector de velocidad de partícula U_{PW} puede ser indicado como

$$\mathbf{U}_{PW} = \frac{P_{PW}}{\rho_0 c} \mathbf{e}_d = \begin{bmatrix} U_x \\ U_y \\ U_z \end{bmatrix},$$

en donde el vector unitario \mathbf{e}_d apunta hacia la dirección de propagación de la onda, por ejemplo, correspondiente a

una medida de dirección. Se puede probar que

$$\begin{aligned} I_a &= \frac{1}{2\rho_0 c} |P_{PW}|^2 \mathbf{e}_d \\ E &= \frac{1}{2\rho_0 c^2} |P_{PW}|^2, \quad (a) \\ \Psi &= 0 \end{aligned}$$

en donde I_a denota la intensidad activa, ρ_0 denota a densidad de del aire, c denota la velocidad del sonido, E denota la energía del campo de sonido y Ψ denota la difusividad.

[0032] Es interesante notar que, puesto que todos los componentes de \mathbf{e}_d son números reales, los componentes de \mathbf{U}_{PW} están todos en fase con P_{PW} . La Figura 1b ilustra un \mathbf{U}_{PW} y P_{PW} en el plano Gaussiano. Como se mencionó, todos los componentes de \mathbf{U}_{PW} comparten la misma fase como P_{PW} , es decir θ . Sus magnitudes, por otra parte, están limitadas a

$$\frac{|P_{PW}|}{c} = \sqrt{|U_x|^2 + |U_y|^2 + |U_z|^2} = \|\mathbf{U}_{PW}\|.$$

[0033] Algunas modalidades de la presente invención pueden proveer un procesamiento para convertir una corriente de DirAC mono a una señal de formato B. Una corriente de DirAC mono puede ser representada por una señal de presión capturada, por ejemplo por un micrófono omnidireccional y mediante información lateral. La información lateral puede comprender medidas dependientes del tiempo-frecuencia de difusividad y dirección de llegada del sonido.

[0034] En algunas modalidades, la señal de audio espacial de entrada puede comprender adicionalmente un parámetro de difusividad Ψ y el estimador 110 puede ser apto para estimar la medida de campo de onda basado adicionalmente en el parámetro de difusividad Ψ .

[0035] La dirección de llegada de entrada y la medida de la dirección de llegada de onda se pueden referir a un punto de referencia correspondiente a un sitio de grabación de la señal de audio espacial de entrada, esto es, en otras palabras todas las direcciones se pueden referir al mismo punto de referencia. El punto de referencia puede ser el sitio en donde un micrófono está colocado o múltiples micrófonos direccionales son colocados con el fin de grabar un campo de sonido.

[0036] En algunas modalidades, la señal de audio espacial convertida puede comprender un primero (X), un segundo (Y) y un tercero (Z) componente direccional. El procesador 120 puede ser apto para el procesamiento adicional de la medida de campo de onda y la medida de dirección de llegada de onda para obtener el primero (X) y/o el segundo (Y) y/o el tercero (Z) componentes direccionales y/o los componentes de audio omnidireccionales.

[0037] En las siguientes notaciones un modelo de datos será introducido.

[0038] Sea $p(t)$ y $\mathbf{u}(t)=[u_x(t), u_y(t), u_z(t)]^T$ el vector depresión y velocidad de partícula, respectivamente, para un punto específico en el espacio, en donde $[\cdot]^T$ denota la transpuesta. $p(t)$ puede corresponder a una representación de audio y $\mathbf{u}(t)=[u_x(t), u_y(t), u_z(t)]^T$ puede corresponder a componentes direccionales. Estas señales pueden ser transformadas a un dominio de tiempo-frecuencia por medio de un banco de filtros apropiado o una STFT (STFT = Transformada de Fourier en el Tiempo Corta) como se sugiere por ejemplo por V. Pulkki and C. Faller, Directional audio coding: Filterbank and STFT-based design, in 120th AES Convention, May 20-23, 2006, Paris, France, May 2006.

[0039] Sea $P(k,n)$ y $\mathbf{U}(k,n)=[U_x(k,n), U_y(k,n), U_z(k,n)]^T$ que denoten las señales transformadas, en donde k y n son índices para la frecuencia (o banda de frecuencia) y tiempo, respectivamente. El vector de intensidad activa $\mathbf{I}_a(k,n)$ puede ser definido como

$$\mathbf{I}_a(k,n) = \frac{1}{2} \text{Re}\{P(k,n) \cdot \mathbf{U}^*(k,n)\}, \quad (1)$$

en donde $(\cdot)^*$ denota la conjugación compleja $\text{Re}\{\cdot\}$ extrae la parte real. El vector de intensidad activa puede expresar el flujo neto de energía que caracteriza el campo de sonido, véase F.J. Fahy, Sound Intensity, Essex: Elsevier Science Publishers Ltd., 1989.

[0040] Sea c que denote la velocidad de sonido en el medio considerado y E la energía de campo de sonido definida por F.J. Fahy

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

en donde $\|\cdot\|$ calcula la 2-norma. En lo siguiente, el contenido de una corriente de DirAC mono será detallado.

[0041] La corriente de DirAC mono puede consistir de la señal mono $p(t)$ o representación de audio y la información lateral, por ejemplo una medida de dirección de llegada. Esta información lateral puede comprender la dirección de llegada dependiente del tiempo-frecuencia y una medida de difusividad dependiente del tiempo-frecuencia. La primera puede ser denotada por $e_{DOA}(k, n)$, que es un vector unitario que apunta hacia la dirección de la cual el sonido llega, esto es, puede ser el modelado de la dirección de llegada. La última, difusividad, puede ser denotada por

$$\Psi(k, n).$$

[0042] En algunas modalidades, el estimador 110 y/o el procesador 120 pueden ser aptos para estimar/procesar la DOA de entrada y/o la medida de DOA de onda en términos de un vector unitario $e_{DOA}(k, n)$. La dirección de llegada puede ser obtenida como

$$e_{DOA}(k, n) = -e_I(k, n),$$

en donde el vector unitario $e(k, n)$ indica la dirección hacia la cual la intensidad activa apunta, es decir

$$I_a(k, n) = \|I_a(k, n)\| \cdot e_I(k, n),$$

$$e_I(k, n) = I_a(k, n) / \|I_a(k, n)\|, \quad (3)$$

respectivamente. Como alternativa, en algunas modalidades, la DOA o medida de DOA puede ser expresada en términos de azimut y ángulo de elevación en un sistema de coordenadas esféricas. Por ejemplo, si $\varphi(k, n)$ y $\vartheta(k, n)$ son azimut y ángulos de elevación, respectivamente, entonces

$$e_{DOA}(k, n) = [\cos(\varphi(k, n)) \cdot \cos(\vartheta(k, n)), \sin(\varphi(k, n)) \cdot \cos(\vartheta(k, n)), \sin(\vartheta(k, n))]^T$$

$$= [e_{DOA,x}(k, n), e_{DOA,y}(k, n), e_{DOA,z}(k, n)]$$

(4)

en donde $e_{DOA,x}(k, n)$ es un componente del vector unitario $e_{DOA}(k, n)$ de la dirección de llegada de entrada a lo largo de un eje x de un sistema de coordenadas Cartesiano, $e_{DOA,y}(k, n)$ es un componente de $e_{DOA}(k, n)$ a lo largo del eje y y $e_{DOA,z}(k, n)$ es un componente de $e_{DOA}(k, n)$ a lo largo del eje z.

[0043] en algunas modalidades, el estimador 110 puede ser apto para estimar la medida de campo de onda basado adicionalmente en el parámetro de difusividad Ψ , opcionalmente también expresado por $\Psi(k, n)$ de manera dependiente del tiempo-frecuencia. El estimador 110 puede ser apto para estimar en base al parámetro de difusividad en términos de

$$\Psi(k, n) = 1 - \frac{\| \langle I_a(k, n) \rangle_t \|}{c \langle E(k, n) \rangle_t}, \quad (5)$$

en donde $\langle \cdot \rangle_t$, indica un promedio temporal.

[0044] Existen diferentes estrategias para obtener $P(k, n)$ y $U(k, n)$ en la práctica. Una posibilidad es usar un micrófono de formato B, que alimenta 4 señales, es decir $w(t)$, $x(t)$, $y(t)$ y $z(t)$. La primera $w(t)$, puede corresponder a la lectura de presión de un micrófono omnidireccional. Las últimas tres pueden corresponder a lecturas de presión de micrófonos que tienen patrones de captación de cifra de ocho dirigidos hacia los tres ejes de un sistema de coordenadas Cartesiano. Estas señales también son proporcionales a la velocidad de partícula. Por consiguiente, en algunas modalidades:

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

$$U(k, n) = -\frac{1}{\sqrt{2}\rho_0 c} [X(k, n), Y(k, n), Z(k, n)]^T \quad (6)$$

en donde $W(k,n)$, $X(k,n)$, $Y(k,n)$ y $Z(k,n)$ son las señales de formato B transformadas correspondientes al componente omnidireccional $W(k,n)$ y los tres componentes direccionales $X(k,n)$, $Y(k,n)$, $Z(k,n)$. Nótese que el factor $\sqrt{2}$ en (6) viene de la convención usada en la definición de señales de formato B, véase Michael Gerzon, Surround sound psychoacoustics, in Wireless World, volume 80, pages 483-486, December 1974.

[0045] Alternativamente, $P(k,n)$ y $\mathbf{U}(k,n)$ pueden ser estimados por medio de un arreglo de micrófonos omnidireccionales, como se sugiere en J. Merimaa, Applications of a 3-D microphone array, in 112th AES Convention, Paper 5501, Munich, May 2002. Las etapas de procesamiento descritas anteriormente también son ilustradas en la Figura 7.

[0046] La Figura 7 muestra un codificador de DirAC 200, que es apto para calcular un canal de audio mono e información lateral a partir de señales de micrófono apropiadas. En otras palabras, la Figura 7 ilustra un codificador de DirAC 200 para determinar la difusividad $\Psi(k,n)$ y dirección de llegada $\mathbf{e}_{DOA}(k,n)$ a partir de señales de micrófono apropiadas. La Figura 7 muestra un codificador de DirAC 200 que comprende una unidad de estimación de P/\mathbf{U} 210. La unidad de estimación de P/\mathbf{U} recibe las señales de micrófono como información de entrada, a las cuales está basada la estimación de P/\mathbf{U} . Puesto que toda la información está disponible, la estimación de P/\mathbf{U} es directa de acuerdo con las ecuaciones anteriores. Una etapa de análisis energético 220 permite la estimación de la dirección de llegada y el parámetro de difusividad de la corriente combinada.

[0047] En algunas modalidades, el estimador 110 puede ser apto para determinar la medida o amplitud del campo de onda en base a una fracción $\beta(k,n)$ de la representación de audio de entrada $P(k,n)$. La Figura 2 muestra las etapas de procesamiento de una modalidad para calcular las señales de formato B a partir de una corriente de DirAC mono. Todas las cantidades dependen de los índices de tiempo y frecuencia (k,n) y son parcialmente omitidas en lo siguiente por simplicidad.

[0048] En otras palabras, la Figura 2 ilustra otra modalidad. De acuerdo con la Ecuación (6), $W(k,n)$ es igual a la presión $P(k,n)$. Por consiguiente, el problema de sintetizar el formato B a partir de una corriente de DirAC mono se reduce a la estimación de del vector de velocidad de partícula $\mathbf{U}(k,n)$, ya que sus componentes son proporcionales a $X(k,n)$, $Y(k,n)$, y $Z(k,n)$.

[0049] Algunas modalidades pueden aproximar la estimación basada en la suposición de que el campo consiste de una onda plana sumada a un campo difuso. Por consiguiente, la presión y velocidad de partícula pueden ser expresados como:

$$P(k,n) = P_{PW}(k,n) + P_{diff}(k,n) \quad (7)$$

$$\mathbf{U}(k,n) = \mathbf{U}_{PW}(k,n) + \mathbf{U}_{diff}(k,n) \quad (8)$$

en donde los subíndices "PW" y "diff" denotan la onda plana y el campo difuso, respectivamente.

[0050] Los parámetros de DirAC portan información solamente con respecto a la intensidad activa. Por consiguiente, el vector de velocidad de partícula $\mathbf{U}(k,n)$ es estimado con $\hat{\mathbf{U}}_{PW}(k,n)$, que es el estimador para la velocidad de partícula de la onda plana solamente. Puede ser definido como:

$$\hat{\mathbf{U}}_{PW}(k,n) = -\frac{1}{\rho_0 c} \beta(k,n) \cdot P(k,n) \cdot \mathbf{e}_{DOA}(k,n), \quad (9)$$

en donde el número real $\beta(k,n)$ es un factor de ponderación apropiado, que en general es dependiente de la frecuencia y puede exhibir una proporcionalidad inversa a la difusividad $\Psi(k,n)$. En efecto, para baja difusividad, esto es $\Psi(k,n)$ cercano a cero, se puede suponer que el campo está compuesto de una sola onda plana, de tal manera que:

$$\hat{\mathbf{U}}_{PW}(k,n) \approx -\frac{1}{\rho_0 c} P(k,n) \cdot \mathbf{e}_{DOA}(k,n) = \hat{\mathbf{U}}_{PW}(k,n) \Big|_{\beta(k,n)=1}, \quad (10)$$

lo cual implica que $\beta(k,n) = 1$.

[0051] En otras palabras, el estimador 110 puede ser apto para estimar la medida de campo de onda con una alta amplitud para un parámetro de baja difusividad Ψ y para estimar la medida de campo de onda con una baja amplitud para un parámetro de alta difusividad Ψ . En algunas modalidades, el parámetro de difusividad $\Psi = [0..1]$. El

parámetro de difusividad puede indicar una relación entre la energía en un componente direccional y la energía en un componente omnidireccional. En algunas modalidades, el parámetro de difusividad Ψ puede ser una medida para un ancho espacial de un componente direccional.

5 **[0052]** Considerando la ecuación anterior y la Ecuación (6), los componentes omnidireccional y/o el primero y/o segundo y/o tercero componente direccionales puede ser expresados como:

$$\begin{aligned}
 W(k, n) &= P(k, n) \\
 X(k, n) &= \sqrt{2}\beta(k, n) \cdot P(k, n) \cdot e_{DOA,x}(k, n) \\
 Y(k, n) &= \sqrt{2}\beta(k, n) \cdot P(k, n) \cdot e_{DOA,y}(k, n) \\
 Z(k, n) &= \sqrt{2}\beta(k, n) \cdot P(k, n) \cdot e_{DOA,z}(k, n)
 \end{aligned}
 \tag{11}$$

15 en donde $e_{DOA,x}(k, n)$ es el componente del vector unitario $e_{DOA}(k, n)$ de la dirección de llegada de entrada a lo largo del eje x de un sistema de coordenadas Cartesiano, $e_{DOA,y}(k, n)$ es el componente de $e_{DOA}(k, n)$ a lo largo del eje y y $e_{DOA,z}(k, n)$ es el componente $e_{DOA}(k, n)$ a lo largo del eje z. En la modalidad mostrada en la Figura 2, la medida de dirección de llegada de onda estimada por el estimador 110 corresponde a $e_{DOA,x}(k, n)$, $e_{DOA,y}(k, n)$ y $e_{DOA,z}(k, n)$ y la medida de campo de onda corresponde a $\beta(k, n)P(k, n)$. El primer componente direccional tal como es emitido por el procesador 120 puede corresponder a cualquiera de $X(k, n)$, $Y(k, n)$ o $Z(k, n)$ y el segundo componente direccional de acuerdo con cualquiera de $X(k, n)$, $Y(k, n)$ o $Z(k, n)$.

25 **[0053]** En lo siguiente, dos modalidades prácticas serán presentadas en como determinar el factor de $\beta(k, n)$.

[0054] La primera modalidad tiene como objetivo estimar la presión de una onda plana primero, es decir $P_{PW}(k, n)$, y luego, de la misma, derivar el vector de velocidad de partícula.

30 **[0055]** Ajustando la densidad del aire ρ_0 igual a 1, y abandonando la dependencia funcional (k, n) por simplicidad, se puede escribir:

$$\psi = 1 - \frac{\langle |P_{PW}|^2 \rangle_t}{\langle |P_{PW}|^2 \rangle_t + 2c^2 \langle E_{diff} \rangle_t} \tag{12}$$

35 **[0056]** Dadas las propiedades estadísticas de los campos difusos, se puede introducir una aproximación por:

$$\langle |P_{PW}|^2 \rangle_t + 2c^2 \langle E_{diff} \rangle_t \approx \langle |P|^2 \rangle_t \tag{13}$$

en donde E_{diff} es la energía del campo difuso. El estimador puede así ser obtenido mediante:

$$\langle |P_{PW}| \rangle_t \approx \langle |\hat{P}_{PW}| \rangle_t = \sqrt{1 - \Psi} \langle |P| \rangle_t \tag{14}$$

45 **[0057]** Para calcular valores estimativos instantáneos, esto es, para cada mosaico de frecuencia en el tiempo, los operadores de esperanza pueden ser removidos, obteniendo:

$$\hat{P}_{PW}(k, n) = \sqrt{1 - \Psi(k, n)} P(k, n) \tag{15}$$

50 **[0058]** Al aprovechar la suposición de onda plana, el valor estimativo para la velocidad de partícula puede ser derivado directamente:

$$\hat{U}_{PW}(k, n) = \frac{1}{\rho_0 c} \hat{P}_{PW}(k, n) \cdot e_t(k, n) \tag{16}$$

de lo cual se sigue que:

$$\beta(k, n) = \sqrt{1 - \Psi(k, n)} \tag{17}$$

60 **[0059]** En otras palabras, el estimador 110 puede ser apto para estimar la fracción $\beta(k, n)$ en base al parámetro de difusividad $\Psi(k, n)$, de acuerdo con:

$$\beta(k, n) = \sqrt{1 - \Psi(k, n)}$$

y la medida de campo de onda de acuerdo con:

$$\beta(k, n)P(k, n) ,$$

en donde el procesador 120 puede ser apto para obtener la magnitud del primer componente direccional $X(k, n)$ y/o el segundo componente direccional $Y(k, n)$ y/o el tercer componente direccional $Z(k, n)$ y/o el componente de audio omnidireccional $W(k, n)$ mediante

$$\begin{aligned} W(k, n) &= P(k, n) \\ X(k, n) &= \sqrt{2}\beta(k, n) \cdot P(k, n) \cdot e_{DOA,x}(k, n) \\ Y(k, n) &= \sqrt{2}\beta(k, n) \cdot P(k, n) \cdot e_{DOA,y}(k, n) \\ Z(k, n) &= \sqrt{2}\beta(k, n) \cdot P(k, n) \cdot e_{DOA,z}(k, n) \end{aligned}$$

en donde la medida de dirección de llegada de onda es representada por el vector unitario $[e_{DOA,x}(k, n), e_{DOA,y}(k, n), e_{DOA,z}(k, n)]^T$, en donde x, y y z indican las direcciones en un sistema de coordenadas Cartesiano.

[0060] Una solución alternativa en algunas modalidades puede ser derivada al obtener el factor $\beta(k, n)$ directamente de la expresión de la difusividad $\Psi(k, n)$. Como ya se mencionó, la velocidad de partícula $U(k, n)$ puede ser modelada como

$$U(k, n) = \beta(k, n) \cdot \frac{P(k, n)}{\rho_0 c} \cdot e_i(k, n) . \quad (18)$$

[0061] La Ecuación (18) puede ser sustituida en (5) conduciendo a:

$$\Psi(k, n) = 1 - \frac{\frac{1}{\rho_0 c} \left\| \langle |\beta(k, n) \cdot P(k, n)|^2 \cdot e_i(k, n) \rangle \right\|}{c < \frac{1}{2\rho_0 c^2} |P(k, n)|^2 \cdot (\beta^2(k, n) + 1) > , \quad (19)$$

[0062] Para obtener los valores instantáneos, los operadores de esperanza pueden ser removidos y resolviendo para $\beta(k, n)$ produce:

$$\beta(k, n) = \frac{1 - \sqrt{1 - (1 - \Psi(k, n))^2}}{1 - \Psi(k, n)} . \quad (20)$$

[0063] En otras palabras, en algunas modalidades, el estimador 110 puede ser apto para estimar la fracción $\beta(k, n)$ en base a $\Psi(k, n)$, de acuerdo con:

$$\beta(k, n) = \frac{1 - \sqrt{1 - (1 - \Psi(k, n))^2}}{1 - \Psi(k, n)} .$$

[0064] En algunas modalidades, la señal de audio espacial de entrada puede corresponder a una señal de DirAC mono. Algunas modalidades pueden ser extendidas para el procesamiento de otras corrientes. En caso de que la corriente o la señal de audio espacial de entrada no lleva un canal omnidireccional, algunas modalidades pueden combinar los canales disponibles para aproximar un patrón de captación omnidireccional. Por ejemplo, en el caso de una corriente de DirAC estéreo como señal de audio espacial de entrada, la señal de presión P en la Figura 2 puede ser aproximada al sumar los canales L y R .

[0065] En lo siguiente, una modalidad con $\Psi = 1$ será ilustrada. La Figura 2 ilustra que si la difusividad es igual a uno para ambas modalidades, el sonido es enrutado exclusivamente al canal W ya que β es igual a cero, de tal manera que las señales X , Y y Z , esto es, los componentes direccionales, pueden ser también cero. Si $\Psi = 1$

constantemente en el tiempo, el canal de audio mono puede así ser enrutado al canal W sin ningún cálculo adicional. La interpretación física de esto es que la señal de audio es presentada al escucha como un campo reactivo puro, ya que el vector de velocidad de partícula tiene magnitud cero.

5 **[0066]** Otro caso cuando $\Psi = 1$ ocurre considerando una situación en donde una señal de audio está presente solamente en una o cualquier sub-conjunto de señales de dipolo y no en la señal W . En el análisis de difusividad de DirAC, este escenario es analizado para tener $\Psi = 1$ con la Ecuación 5, puesto que el vector de intensidad tiene constantemente la duración de cero, ya que la presión P es cero en la Ecuación (1). La interpretación física de esto es también que la señal de audio es presentada al escucha siendo reactiva, ya que esta señal de presión en el
10 tiempo es constantemente cero, mientras que el vector de velocidad de partícula no es cero.

[0067] Debido al hecho de que el formato B es inherentemente una representación independiente de montaje de los altavoces, algunas modalidades pueden usar el formato B como un lenguaje común hablado por diferentes dispositivos de audio, lo que significa que la conversión de uno a otro se puede hacer posible por modalidades vía una conversión intermedia a formato B. Por ejemplo, algunas modalidades pueden unir corrientes de DirAC de diferentes medios ambientes acústicos grabados con diferentes medios ambientes de sonido sintetizados en formato B. La unión de corriente de DirAC mono a corriente de formato B puede también ser habilitada por algunas modalidades.

20 **[0068]** Algunas modalidades pueden permitir la unión de señales de audio de multicanal en cualquier formato surround con una corriente de DirAC mono. Además, algunas modalidades pueden permitir la unión de una corriente de DirAC mono con cualquier corriente de formato B. Además, algunas modalidades pueden permitir la unión de una corriente de DirAC mono con una corriente de formato B.

25 **[0069]** Estas modalidades pueden proveer una ventaja, por ejemplo, en la creación de reverberación o introducir efectos de audio, como será detallado subsecuentemente. En la producción de música, los reverberadores pueden ser usados como dispositivos de efecto que colocan perceptualmente el audio procesado a un espacio virtual. En realidad virtual, la síntesis de reverberación puede ser necesaria cuando las fuentes virtuales son auralizadas al interior de un espacio cerrado, por ejemplo en salas o salones de conciertos.

30 **[0070]** Cuando una señal para reverberación está disponible, tal auralización puede ser efectuada por algunas modalidades mediante aplicación de sonido seco y sonido reverberado a diferentes corrientes de DirAC. Algunas modalidades pueden usar diferentes procedimientos en cuanto a como procesar la señal reverberada en el contexto de DirAC, en donde algunas modalidades pueden producir el sonido reverberado que es máximamente difuso
35 alrededor del escucha.

[0071] La Figura 3 ilustra una modalidad de un aparato 300 para determinar una señal de audio espacial convertida combinada, la señal de audio espacial convertida combinada tiene por lo menos un primer componente combinado y un segundo componente combinado, en donde la señal de audio espacial convertida combinada es determinada a partir de una primera y una segunda señal de audio espacial de entrada que tiene una primera y una segunda representación de audio de entrada y una primera y una segunda dirección de llegada.

40 **[0072]** El aparato 300 comprende una primera modalidad del aparato 101 para determinar una señal de audio espacial convertida de acuerdo con la descripción anterior, para proveer una primera señal convertida que tiene un primer componente omnidireccional y por lo menos un componente direccional del primer aparato 101. Además, el aparato 300 comprende otra modalidad de un aparato 102 para determinar una señal de audio espacial convertida de acuerdo con la descripción anterior para proveer una segunda señal convertida, que tiene un segundo componente omnidireccional y por lo menos un componente direccional del segundo aparato 102.

50 **[0073]** En general, algunas modalidades no están limitados a comprender solamente dos de los aparatos 100, en general, una pluralidad de los aparatos descritos anteriormente pueden estar comprendidos en el aparato 300, por ejemplo, el aparato 300 puede ser apto para combinar una pluralidad de señales de DirAC.

55 **[0074]** De acuerdo con la Figura 3, el aparato 300 comprende además un generador de efectos de audio 301 para presentar el primer compuesto omnidireccional y el primer componente de audio direccional del primer aparato 101 para obtener un primer componente presentado.

60 **[0075]** Además, el aparato 300 comprende un primer combinador 311 para combinar el primer componente presentado con los primeros y segundos componentes omnidireccionales, o para combinar el primer componente presentado con componentes direccionales del primer aparato 101 y el segundo aparato 102 para obtener el primer componente combinado. El aparato 300 comprende además un segundo combinador 312 para combinar los primeros y segundos componentes omnidireccionales o los componentes direccionales a partir del primero o segundo aparato 101 y 102 para obtener el segundo componente combinado.

65 **[0076]** En otras palabras, el generador de efectos de audio 301 puede presentar el primer componente omnidireccional, de tal manera que el primer combinador 311 puede luego combinar el primer componente

omnidireccional presentado, el primer componente omnidireccional y el segundo componente omnidireccional para obtener el primer componente combinado. El primer componente combinado puede luego corresponder, por ejemplo, a un componente omnidireccional combinado. En esta modalidad, el segundo combinador 312 puede combinar el componente direccional del primer aparato 101 y el componente direccional del segundo aparato para obtener el segundo componente combinado, por ejemplo, correspondiente a un primer componente direccional combinado.

[0077] En algunas modalidades, el generador de efectos de audio 301 puede presentar los componentes direccionales. En estas modalidades el combinador 311 puede combinar el componente direccional del primer aparato 101, el componente direccional del segundo aparato 102 y el primer componente presentado para obtener el primer componente combinado, en este caso correspondiente a un componente direccional combinado. En esta modalidad, el segundo combinador 312 puede combinar los primeros y segundos componentes omnidireccionales del primer aparato 101 y el segundo aparato 102 para obtener el segundo componente combinado, esto es, un componente omnidireccional combinado.

[0078] En otras palabras, la Figura 3 muestra una modalidad de un aparato 300 apto para determinar una señal de audio espacial convertida combinada, la señal de audio espacial convertida combinada tiene por lo menos un primer componente combinado y un segundo componente combinado, a partir de una primera y una segunda señal de audio espacial de entrada, la primera señal de audio espacial de entrada tiene una primera representación de audio de entrada y una primera dirección de llegada, la segunda señal de entrada espacial tiene una segunda representación de audio de entrada y una segunda dirección de llegada.

[0079] El aparato 300 comprende un primer aparato 101 que comprende un aparato 100 apto para determinar una señal de audio espacial convertida, la señal de audio espacial convertida tiene un componente de audio omnidireccional W' y por lo menos un componente de audio direccional $X; Y; Z$, a partir de una señal de audio espacial de entrada, la señal de audio espacial de entrada tiene una representación de audio de entrada y una dirección de llegada de entrada. El aparato 100 comprende un estimador 110 apto para estimar una representación de onda, la representación de onda comprende una medida de campo de onda y una medida de dirección de llegada de onda, basada en la representación de audio de entrada y la dirección de llegada de entrada.

[0080] Además, el aparato 100 comprende un procesador 120 apto para procesar la medida de campo de onda y la medida de d_i de llegada de onda para obtener el componente omnidireccional (W') y el por lo menos un componente direccional ($X;Y;Z$). El primer aparato 101 es apto para proveer una primera señal convertida basada en la primera señal de audio espacial de entrada, que tiene un primer componente omnidireccional y por lo menos un componente direccional del primer aparato 101.

[0081] Además, el aparato 300 comprende un segundo aparato 102 que comprende otro aparato 100 apto para proveer una segunda señal convertida basada en la segunda señal de audio espacial de entrada, que tiene un segundo componente omnidireccional y por lo menos un componente direccional del segundo aparato 102. Además, el aparato 300 comprende un generador de efectos de audio 301 apto para presentar el primer componente omnidireccional para obtener un primer componente presentado o para presentar el componente direccional del primer aparato 101 para obtener el primer componente presentado.

[0082] Además, el aparato 300 comprende un primer combinador 311 apto para combinar el primer componente presentado, el primer componente omnidireccional y el segundo componente omnidireccional, o para combinar el primer componente presentado, el componente direccional del primer aparato 101, y el componente direccional del segundo aparato 102 para obtener el primer componente combinado. El aparato 300 comprende un segundo combinador 312 apto para combinar el componente direccional del primer aparato 101 y el componente direccional del segundo aparato 102 o para combinar el primer componente omnidireccional y el segundo componente omnidireccional para obtener el segundo componente combinado.

[0083] En otras palabras, la Figura 3 muestra una modalidad de un aparato 300 apto para determinar una señal de audio espacial convertida combinada, la señal de audio espacial convertida combinada tiene por lo menos un primer componente combinado y un segundo componente combinado, a partir de una primera y una segunda señal de audio espacial de entrada, la primera señal de audio espacial de entrada tiene una primera representación de audio de entrada y una primera dirección de llegada, la segunda señal de entrada espacial tiene una segunda representación de audio de entrada y una segunda dirección de llegada. El aparato 300 comprende primeros medios 101 aptos para determinar una primera señal convertida, la primera señal convertida tiene un primer componente omnidireccional y por lo menos un primer componente direccional ($X;Y;Z$), a partir de la primera señal de audio espacial de entrada. Los primeros medios 101 pueden comprender una modalidad del aparato descrito anteriormente 100.

[0084] Los primeros medios 101 comprenden un estimador apto para estimar una primera representación de onda, la primera representación de onda comprende una primera medida de campo y una primera medida de dirección de llegada de onda, basada en la primera representación de audio de entrada y la primera dirección de llegada de entrada. El estimador puede corresponder a una modalidad del estimador 110 descrito anteriormente.

- 5 [0085] Los primeros medios 101 comprenden además un procesador aptos para procesar la primera medida de campo de onda y la primera medida de dirección de llegada de onda para obtener el primer componente omnidireccional y el por lo menos un primer componente direccional. El procesador puede corresponder a una modalidad del procesador 120 descrito anteriormente.
- 10 [0086] Los primeros medios 101 pueden ser aptos además para proveer la primera señal convertida que tiene el primer componente omnidireccional y el por lo menos un primer componente direccional.
- 15 [0087] Además, el aparato 300 comprende segundos medios 102 aptos para proveer una segunda señal convertida basada en la segunda señal de audio espacial de entrada, que tiene un segundo componente omnidireccional y por lo menos un segundo componente direccional. Los segundos medios pueden comprender una modalidad del aparato 100 descrito anteriormente.
- 20 [0088] Los segundos medios 102 comprenden además otro estimador apto para estimar una segunda representación de onda, la segunda representación de onda comprende una segunda medida de campo de onda y una segunda medida de dirección de llegada de onda, basada en la segunda representación de audio de entrada y la segunda dirección de llegada de entrada. El otro estimador puede corresponder a una modalidad del estimador 110 descrito anteriormente.
- 25 [0089] Los segundos medios 102 comprenden además otro procesador apto para procesar la segunda medida de campo de onda y la segunda medida de dirección de llegada de onda para obtener el segundo componente omnidireccional y el por lo menos un segundo componente direccional. El otro procesador puede corresponder a una modalidad del procesador 120 descrito anteriormente.
- 30 [0090] Además, los segundos medios 101 son aptos para proveer la segunda señal convertida que tiene el segundo componente omnidireccional y por lo menos un segundo componente direccional.
- 35 [0091] Además, el aparato 300 comprende un generador de efectos de audio 301 apto para presentar el primer componente omnidireccional para obtener un primer componente presentado o para presentar el primer componente direccional para obtener el primer componente presentado. El aparato 300 comprende un primer combinador 311 apto para combinar el primer componente presentado, el primer componente omnidireccional y el segundo componente omnidireccional, o para combinar el primer componente presentado, el primer componente direccional, y el segundo componente direccional para obtener el primer componente combinado.
- 40 [0092] Además, el aparato 300 comprende un segundo combinador 312 apto para combinar el primer componente direccional y el segundo componente direccional, o para combinar el primer componente omnidireccional y el segundo componente omnidireccional para obtener el segundo componente combinado.
- 45 [0093] En modalidades, se puede efectuar un procesamiento para determinar una señal de audio espacial convertida combinada, la señal de audio espacial convertida combinada tiene por lo menos un primer componente combinado y un segundo componente combinado, a partir de una primera y una segunda señal de audio espacial de entrada, la primera señal de audio espacial de entrada tiene una primera representación de audio de entrada y una primera dirección de llegada, la segunda señal de entrada espacial tiene una segunda representación de audio de entrada y una segunda dirección de llegada.
- 50 [0094] El procesamiento puede comprender las etapas de determinar una primera señal de audio espacial convertida, la primera señal de audio espacial convertida tiene un primer componente omnidireccional (W') y por lo menos un primer componente direccional ($X;Y;Z$), a partir de la primera señal de audio espacial de entrada, al usar las sub-etapas de estimar una primera representación de onda, la primera representación de onda comprende una primera medida de campo de onda y una primera medida de dirección de llegada de onda, basada en la primera representación de audio de entrada y la primera dirección de llegada de entrada; y procesar la primera medida de campo de onda y la primera medida de dirección de llegada de onda para obtener el primer componente omnidireccional (W') y el por lo menos un primer componente direccional ($X;Y;Z$).
- 55 [0095] El procesamiento puede comprender además la etapa de proveer la primera señal convertida que tiene el primer componente omnidireccional y el por lo menos un primer componente direccional.
- 60 [0096] Además, el procesamiento puede comprender determinar una segunda señal de audio espacial convertida, la segunda señal de audio espacial convertida tiene un segundo componente omnidireccional (W') y por lo menos un segundo componente direccional ($X;Y;Z$), a partir de la segunda señal de audio espacial de entrada, al usar las sub-etapas de estimar una segunda representación de onda, la segunda representación de onda comprende una segunda medida de campo de onda y una segunda medida de dirección de llegada de onda, basada en la segunda representación de audio de entrada y la segunda dirección de llegada de entrada; y procesar la segunda medida de campo de onda y la segunda medida de dirección de llegada de onda para obtener el segundo componente omnidireccional (W') y el por lo menos un segundo componente direccional ($X;Y;Z$).
- 65

[0097] Además el procesamiento puede comprender proveer la segunda señal convertida que tiene el segundo componente omnidireccional y el por lo menos un segundo componente direccional.

5 **[0098]** El procesamiento puede comprender además presentar el primer componente omnidireccional para obtener un primer componente presentado o presentar el primer componente direccional para obtener el primer componente presentado; y combinar el primer componente presentado, el primer componente omnidireccional y el segundo componente omnidireccional, o combinar el primer componente presentado, el primer componente direccional, y el segundo componente direccional para obtener el primer componente combinado.

10 **[0099]** Además, el procesamiento puede comprender combinar el primer componente direccional y el segundo componente direccional, o combinar el primer componente omnidireccional y el segundo componente omnidireccional para obtener el segundo componente combinado.

15 **[0100]** De acuerdo con las modalidades descritas anteriormente, cada uno de los aparatos pueden producir múltiples componentes direccionales, por ejemplo un componente X, un componente Y y un componente Z. En algunas modalidades, se pueden usar múltiples generadores de efecto de audio, lo que se indica en la Figura 3 por los bloques de líneas discontinuas 302, 303 y 304. Estos generadores de efecto de audio opcionales pueden generar componentes presentados correspondientes, basados en señales de entrada omnidireccionales y/o direccionales. En una modalidad, un generador de efecto de audio puede presentar un componente direccional en base a un componente omnidireccional. Además, el aparato 300 puede comprender múltiples combinadores, esto es combinadores 311, 312, 313 y 314 con el fin de combinar un componente combinado omnidireccional y múltiples componentes direccionales combinados, por ejemplo, para las tres dimensiones espaciales.

20 **[0101]** Una de las ventajas de la estructura del aparato 300 es que un máximo de cuatro generadores de efecto de audio son necesarios para presentar en general un número ilimitado de fuentes de audio.

25 **[0102]** Como se indica por los combinadores de líneas discontinuas 331, 332, 333 y 334 en la Figura 3, un generador de efecto de audio puede ser apto para presentar una combinación de componentes direccionales u omnidireccionales a partir de los aparatos 101 y 102. En una modalidad, el generador de efecto de audio 301 puede ser apto para presentar una combinación de los componentes omnidireccionales del primer aparato 101 y el segundo aparato 102 o para presentar una combinación de los componentes direccionales del primer aparato 101 y el segundo aparato 102 para obtener el primer componente presentado. Como se indica por las trayectorias discontinuas en la Figura 3, se pueden proveer combinaciones de múltiples componentes a los diferentes generadores de efecto de audio.

30 **[0103]** En una modalidad, todos los componentes omnidireccionales de todas las fuentes de sonido, en la Figura 3, representadas por el primer aparato 101 y el segundo aparato 102, pueden ser combinados con el fin de generar múltiples componentes presentados. En cada una de las cuatro trayectorias mostradas en la Figura 3, cada generador de efecto de audio puede generar un componente presentado a ser agregado a los componentes direccionales u omnidireccionales correspondientes de las fuentes de sonido.

35 **[0104]** Además, como se muestra en la Figura 3, se pueden usar múltiples etapas de retardo y escalamiento 321 y 322. En otras palabras, cada aparato 101 ó 102 puede tener en su trayectoria de salida una etapa de retardo y escalamiento 321 ó 322, con el fin de retardar uno o más de sus componentes de salida. En algunas modalidades, las etapas de retardo y escalamiento pueden retardar y escalar los respectivos componentes omnidireccionales solamente. En general, las etapas de retardo y escalamiento pueden ser usadas para componentes omnidireccionales y direccionales.

40 **[0105]** En algunas modalidades, el aparato 300 puede comprender una pluralidad de aparatos 100 que representan fuentes de audio y correspondientemente una pluralidad de generadores de efecto de audio, en donde el número de generadores de efecto de audio es menor que el número de aparatos correspondientes a las fuentes de sonido. Como ya se mencionó anteriormente, en una modalidad, pueden haber hasta cuatro generadores de efecto de audio, con un número básicamente ilimitado de fuentes de sonido. En algunas modalidades, un generador de efecto de audio puede corresponder a un reverberador.

45 **[0106]** La Figura 4a muestra otra modalidad de un aparato 300 en más detalle. La Figura 4a muestra dos aparatos 101 y 102 cada uno emitiendo un componente de audio omnidireccional W, y tres componentes direccionales X, Y, Z. De acuerdo con la modalidad mostrada en la Figura 4a, los componentes omnidireccionales de cada uno de los aparatos 101 y 102 son provistos a dos etapas de retardo y escalamiento 321 y 322, que emiten tres componentes retardados y escalados, que son luego agregados por los combinadores 331, 332, 333 y 334. Cada una de las señales combinadas es luego presentada separadamente por uno de los cuatro generadores de efecto de audio 301, 302, 303 y 304, que son implementados como reverberadores en la Figura 4a. Como se indica en la Figura 4a, cada uno de los generadores de efecto de audio emite un componente, correspondiente a un componente omnidireccional y tres componentes direccionales en total. Los combinadores 311, 312, 313 y 314 son luego usados para combinar los respectivos componentes presentados con los componentes originales emitidos por los aparatos

101 y 102, en donde en la Figura 4a en general, pueden haber una multiplicidad de aparatos 100.

5 **[0107]** En otras palabras, en el combinador 311 una versión presentada de las señales de salida omnidireccionales combinadas de todos los aparatos puede ser combinada con los componentes de salida omnidireccionales originales o sin presentar. Combinaciones similares se pueden llevar a cabo por otros combinadores con respecto a los componentes direccionales. En la modalidad mostrada en la Figura 4a, los componentes direccionales presentados son creados en base a versiones retardadas y escaladas de los componentes omnidireccionales.

10 **[0108]** En general, algunas modalidades pueden aplicar un efecto de audio como por ejemplo una reverberación eficientemente a una o más corrientes de DirAC. Por ejemplo, por lo menos dos corrientes de DirAC son introducidas a la modalidad del aparato 300, como se muestra en la Figura 4a. En algunas modalidades, estas corrientes pueden ser corrientes de DirAC reales o corrientes sintetizadas, por ejemplo al tomar una señal mono y agregar información lateral como dirección y difusividad. De acuerdo con la discusión anterior, los aparatos 101, 102 pueden generar hasta cuatro señales para cada corriente, es decir W, X, Y y Z. En general, las modalidades de los aparatos 101 ó 15 102 pueden proveer menos de tres componentes direccionales, por ejemplo solamente X o X y Y, o cualquier otra combinación de los mismos.

20 **[0109]** En algunas modalidades, los componentes omnidireccionales W pueden ser provistos a generadores de efecto de audio, como por ejemplo reverberadores con el fin de crear los componentes presentados. En algunas modalidades, para cada una de las corrientes de DirAC de entrada, las señales pueden ser copiadas a las cuatro ramas mostradas en la Figura 4a, que pueden ser retardadas independientemente, esto es, individualmente por aparato 101 ó 102 cuatro retardadas independientemente, por ejemplo por retardos $\tau_W, \tau_X, \tau_Y, \tau_Z$ y escaladas, por ejemplo por factores de escalamiento $\gamma_W, \gamma_X, \gamma_Y, \gamma_Z$, versiones pueden ser combinadas antes de ser provistas a un generador de efecto de audio.

25 **[0110]** De acuerdo con las Figuras 3 y 4a, las ramas de las diferentes corrientes, esto es, las salidas de los aparatos 101 y 102, pueden ser combinadas para obtener cuatro señales combinadas. Las señales combinadas pueden luego ser presentadas independientemente por los generadores de audio, por ejemplo reverberadores mono convencionales. Las señales presentadas resultantes pueden luego ser sumadas a las señales W, X, Y y Z emitidas originalmente de los diferentes aparatos 101 y 102.

30 **[0111]** En algunas modalidades, señales de formato B generales pueden ser obtenidas, que pueden luego, por ejemplo, ser reproducidas con un descodificador de formato B y se lleva a cabo por ejemplo en Ambisonics. En otras modalidades, las señales de formato B pueden ser codificadas como por ejemplo con el codificador de DirAC como se muestra en la Figura 7, de tal manera que la corriente de DirAC resultante puede luego ser transmitida, procesada o descodificada adicionalmente con un descodificador de DirAC mono convencional. La etapa de descodificación puede corresponder a calcular señales de altavoz para reproducción.

35 **[0112]** La Figura 4b muestra otra modalidad de un aparato 300. La Figura 4b muestra los dos aparatos 101 y 102 con los cuatro componentes de salida correspondientes. En la modalidad mostrada en la Figura 4b, solamente los componentes omnidireccionales W son usados para ser primero retardados individualmente y escalados en las etapas de retardo y escalamiento 321 y 322 antes de ser combinados por el combinador 331. La señal combinada es luego provista al generador de efecto de audio 301, que es otra vez implementado como un reverberador en la Figura 4b. La salida presentada del reverberador 301 es luego combinada con los componentes omnidireccionales originales de los aparatos 101 y 102 por el combinador 311. Los otros combinadores 312, 313 y 314 son usados para combinar los componentes direccionales X, Y y Z de los aparatos 101 y 102 con el fin de obtener componentes direccionales combinados correspondientes.

40 **[0113]** En relación con la modalidad ilustrada en la Figura 4a, la modalidad ilustrada en la Figura 4b corresponde a ajustar los factores de escalamiento para las ramas X, Y y Z a 0. En esta modalidad, solamente un generador de efecto de audio o reverberador 301 es usado. En una modalidad, el generador de efecto de audio 301 puede ser apto para reverberar el primer componente omnidireccional solamente para obtener el primer componente presentado, esto es, solamente W puede ser reverberado.

45 **[0114]** En general, ya que los aparatos 101, 102 y potencialmente N aparatos correspondientes a N fuentes de sonido, las potencialmente N etapas de retardo y escalamiento 321, que son opcionales, pueden simular las distancias de las fuentes de sonido, un retardo más corto puede corresponder a la percepción de una fuente de sonido virtual más cercana a la escucha. En general, la etapa de retardo y escalamiento 321, puede ser usada para presentar una relación espacial entre diferentes fuentes de sonido representadas por la señal convertida, señales de audio espacial convertidas respectivamente. La impresión espacial de un medio ambiente de los alrededores puede luego ser creado por los generadores de efecto de audio correspondiente 301 o reverberadores. En otras palabras, en algunas modalidades, las etapas de retardo y escalamiento 321 pueden ser usadas para introducir retardos específicos de fuente y escalamiento en relación con las otras fuentes de sonido. Una combinación de las señales relacionadas apropiadamente, esto es, retardadas y escaladas, convertidas, puede luego ser apta a un medio ambiente espacial por el generador de efecto de audio 301.

[0115] La etapa de retardo y escalamiento 321 puede ser vista como una clase de reverberador también. En algunas modalidades, el retardo introducido por la etapa de retardo y escalamiento 321 puede ser más corta que un retardo introducido por el generador de efecto de audio 301. En algunas modalidades, una base de tiempo común, por ejemplo como es provista por un generador de reloj, puede ser usada para la etapa de retardo y escalamiento 321 y el generador de efecto de audio 301. Un retardo puede luego ser expresado en términos de un número de períodos de muestra y el retardo introducido por la etapa de retardo y escalamiento 321 puede corresponder a un número más bajo de períodos de muestra que un retardo introducido por el generador de efecto de audio 301.

[0116] Las modalidades como se ilustran en las Figuras 3, 4a y 4b pueden ser utilizadas para casos cuando se usa decodificación de DirAC mono para N fuentes de sonido que son luego reverberadas conjuntamente. Ya que se puede suponer que la salida de un reverberador tiene una salida que es totalmente difusa, esto es, puede ser interpretada como una señal omnidireccional W también. Esta señal puede ser combinada con otras señales de formato sintetizadas, tales como las señales de formato B originadas de N fuentes de audio por sí mismas, que representan así la trayectoria directa al escucha. Cuando la señal de formato B resultante es codificada y decodificada por DirAC adicionalmente, el sonido reverberado se puede hacer disponible por algunas modalidades.

[0117] En la Figura 4c, se muestra otra modalidad del aparato 300. En la modalidad mostrada en la Figura 4c, en base a las señales omnidireccionales de salida de los aparatos 101 y 102, se crean componentes presentados reverberados direccionales. Por consiguiente, en base a la salida omnidireccional, las etapas de retardo y escalamiento 321 y 322 crean componentes retardados y escalados individualmente, que son combinados por los combinadores 331, 332 y 333. A cada una de las señales combinadas, diferentes reverberadores 301, 302 y 303 son aplicados, que en general corresponden a diferentes generadores de efecto de audio. De acuerdo con la descripción anterior, los componentes omnidireccionales, direccionales y presentados correspondientes son combinados por los combinadores 311, 312, 313 y 314, con el fin de proveer un componente omnidireccional combinado y componentes direccionales combinados.

[0118] En otras palabras, las señales W o señales omnidireccionales para cada corriente son alimentadas a tres generadores de efecto de audio, como por ejemplo reverberadores, como se muestra en las figuras. En general, también pueden haber solamente dos ramas dependiendo de si una señal de sonido bidimensional o tridimensional va a ser generada. Una vez que se obtiene las señales de formato B, las corrientes pueden ser decodificadas vía un decodificador de DirAC de micrófono virtual. El último es descrito en detalle en V. Pulkki, Spatial Sound Reproduction With Directional Audio Coding, Journal of the Audio Engineering Society, 55 (6): 503-516.

[0119] Con este decodificador, las señales de altavoz $D_p(k,n)$ pueden ser obtenidas como una combinación lineal de las señales W, X, Y y Z, por ejemplo de acuerdo con:

$$D_p(k,n) = G(k,n) \left[W(k,n)\sqrt{2} + X(k,n)\cos(\alpha_p)\cos(\beta_p) + Y(k,n)\sin(\alpha_p)\cos(\beta_p) + Z(k,n)\sin(\beta_p) \right]$$

en donde α_p y β_p son el azimut y la elevación del p-ésimo altavoz. El término $G(k,n)$ es una ganancia de toma panorámica dependiente de la dirección de llegada y de la configuración del altavoz.

[0120] En otras palabras, la modalidad mostrada en la Figura 4c puede proveer las señales de audio para los altavoces correspondiente a señales de audio obtenibles al colocar micrófonos virtuales orientados hacia la posición de los altavoces y que tienen fuentes de sonido semejantes a punto, cuya posición es determinada por los parámetros de DirAC. Los micrófonos virtuales pueden tener patrones de captación formados como caridioides, como dipolos o como cualquier patrón direccional de primer orden.

[0121] Los sonidos reverberados pueden por ejemplo ser usados eficientemente como X y Y en la suma de formato B. Tales modalidades pueden ser aplicadas a disposiciones físicas de altavoz horizontales que tienen cualquier número de altavoces, sin crear la necesidad de más reverberadores.

[0122] Como se discute anteriormente, la decodificación de DirAC mono tiene limitaciones en calidad de reverberación, en donde en modalidades la calidad puede ser mejorada con la decodificación de DirAC de micrófono virtual, que toma ventaja también de las señales de dipolo en una corriente de formato B.

[0123] La creación apropiada de señales de formato B para reverberar una señal de audio para la decodificación de DirAC de micrófono virtual se puede llevar a cabo en algunas modalidades. Un concepto simple y efectivo que puede ser usado por algunas modalidades es enrutar diferentes canales de audio diferentes señales de dipolo, por ejemplo, a X y Y canales. Algunas modalidades puede implementar esto por dos reverberadores que producen canales de audio mono incoherentes a partir del mismo canal de entrada, tratando sus salidas como canales de audio dipolo de formato B X y Y, respectivamente, como se muestra en la Figura 4c para los componentes direccionales. Ya que las señales no son aplicadas a W, serán analizadas para ser totalmente difusas en la codificación de DirAC subsecuente. También, la calidad incrementada para reverberación puede ser obtenida en la decodificación de DirAC de micrófono virtual, ya que los canales de dipolo contienen sonido reverberado diferentemente. Algunas modalidades pueden generar con las mismas una percepción "más amplia" y "más

envolvente” de reverberación que con la descodificación de DirAC mono. Algunas modalidades pueden por consiguiente usar un máximo de dos reverberadores en disposiciones físicas de altavoz horizontales y tres disposiciones físicas de altavoces 3-D en la reverberación basada en DirAC descrita.

5 **[0124]** Algunas modalidades pueden no estar limitadas a reverberación de señales, sino que pueden aplicar cualesquier otros efectos de audio que tienen por objetivo, por ejemplo en una percepción totalmente difusa del sonido. Similar a las modalidades descritas anteriormente, la señal de formato B reverberada puede ser sumada a otras señales de formato B sintetizadas en algunas modalidades, tales como las que se originan de las N fuentes de audio por sí mismas, representando así una trayectoria directa al escucha.

10 **[0125]** Todavía otra modalidad es mostrada en la Figura 4d. La Figura 4d muestra una modalidad similar como la Figura 4a, sin embargo, ninguna etapa de retardo o escalamiento 321 ó 322 está presente, esto es, los señalados individuales en las ramas son solamente reverberadas, en algunas modalidades solo los componentes omnidireccionales W son reverberados. La modalidad ilustrada en la Figura 4d puede también ser vista como similar a la modalidad ilustrada en la Figura 4a con los retardos y escalamientos o ganancias antes de que los reverberadores sean ajustados a 0 y 1, respectivamente, sin embargo, en esta modalidad, se supone que los reverberadores 301, 302, 303 y 304 no son arbitrarios e independientes. En la modalidad ilustrada en la Figura 4d, los cuatro generadores de efecto de audio son sumados para ser dependiente entre sí teniendo una estructura específica.

15 **[0126]** Cada uno de los generadores de efecto de audio o reverberadores pueden ser implementados como una línea de retardo derivada como será detallado subsecuentemente con la ayuda de la Figura 5. Los retardos y ganancias o escalamientos pueden ser escogidos apropiadamente de tal manera que cada una de las derivaciones modela un eco distinto cuya dirección, retardo y potencia pueden ser ajustados a voluntad.

20 **[0127]** En tal modalidad, el i-ésimo eco puede ser caracterizado por un factor de ponderación, por ejemplo en referencia a un sonido de DirAC ρ_i , un retardo τ_i y una dirección de llegada θ_i y ϕ_i correspondiente a la elevación y azimut respectivamente.

25 **[0128]** Los parámetros de los reverberadores pueden ser ajustados como sigue:

$$\tau_W = \tau_X = \tau_Y = \tau_Z = \tau_i$$

30 $\gamma_W = \rho_i$, para el W reverberador,

35 $\gamma_X = \rho_i \cos(\phi_i) \cos(\theta_i)$, para el X reverberador

$\gamma_Y = \rho_i \sin(\phi_i) \cos(\theta_i)$, para el Y reverberador

40 $\gamma_Z = \rho_i \sin(\theta_i)$, para el reverberador Z

[0129] En algunas modalidades, los parámetros físicos de cada eco pueden ser extraídos de procesos aleatorios o tomados de una respuesta de impulso de espacial de sala. La última podría por ejemplo ser medida o simulada con una herramienta de rastreo de rayos.

45 **[0130]** En general, algunas modalidades pueden proveer con las mismas la ventaja de que el número de generadores de efecto de audio es independiente del número de fuentes.

50 **[0131]** La Figura 5 ilustra una modalidad que utiliza un esquema conceptual de un efecto de audio mono como por ejemplo usado dentro de un generador de efecto de audio, que es extendido dentro del contexto de DirAC. Por ejemplo, un reverberador puede ser realizado de acuerdo con este esquema. La Figura 5 muestra una modalidad de un reverberador 500. La Figura 5 muestra en principio una estructura de filtro de FIR (FIR = Respuesta de Impulso Finito). Otras modalidades pueden utilizar filtros de IIR (IIR = Respuesta de Impulso Infinito) también. Una entrada señal es retardada por las K etapas de retardo marcadas por 511 a 51K. Las K copias retardadas, para las cuales los retardos son denotados por τ_1 a τ_K de la señal, son luego amplificadas por los amplificadores 521 a 52K con factores de amplificación γ_1 a γ_K antes de que sean sumadas en la etapa de suma 530.

55 **[0132]** La Figura 6 muestra otra modalidad con una extensión de la cadena de procesamiento de la Figura 5 dentro del contexto de DirAC. La salida del bloque de procesamiento puede ser una señal de formato B. La Figura 6 muestra una modalidad en donde múltiples etapas de suma 560, 562 y 564 son utilizadas dando como resultado las tres señales de salida W, X y Y. Con el fin de establecer diferentes combinaciones, las copias de señal retardadas pueden ser escaladas diferentemente antes de que sean agregadas en las tres etapas de adición diferentes 560, 562 y 564. Esto se lleva a cabo por los amplificadores adicionales 531 a 53K y 541 a 54K. En otras palabras, la modalidad 600 mostrada en la Figura 6 lleva a cabo reverberación para diferentes componentes de una señal de formato B basada en una corriente de DirAC mono. Tres diferentes copias reverberadas de la señal son generadas

utilizando tres filtros de FIR diferentes siendo establecidos por medio de diferentes coeficientes de filtro ρ_i a ρ_K y η_i a η_K .

5 **[0133]** La siguiente modalidad se puede aplicar a un reverberador o efecto de audio que puede ser modelado como la Figura 5. Una señal de entrada corre a través de una línea de retardo derivada simple, en donde múltiples copias de la misma son sumadas conjuntamente. La i -ésima de K ramas es retardada y atenuada por τ_i y γ_i respectivamente.

10 **[0134]** Los factores γ y τ pueden ser obtenidos dependiendo del efecto de audio deseado. En caso de un reverberador, estos factores imitan la respuesta de impulso de la sala que va a ser simulada. De cualquier manera, su determinación no es iluminada y así se supone que es dada.

15 **[0135]** Una modalidad es ilustrada en la Figura 6. El esquema en la Figura 5 es extendido de tal manera que dos capas más son obtenidas. En modalidades, a cada rama y ángulo de llegada θ puede ser asignados obtenido de un proceso estocástico. Por ejemplo, θ puede ser la realización de una distribución uniforme en el intervalo $[-\pi, \pi]$. La i -ésima rama es multiplicada con los factores η_i y ρ_i que pueden ser definidos como

$$\eta_i = \text{sen}(\theta_i) \quad (21)$$

20
$$\rho_i = \text{cos}(\theta_i) . \quad (22)$$

25 **[0136]** En algunas modalidades, el i -ésimo eco puede ser percibido como procedente de θ_i . La extensión a 3D es directa. En este caso, más capas necesitan ser agregadas y un ángulo de elevación necesita ser considerado. Una vez que la señal de formato B ha sido generada, es decir W , X , Y , y posiblemente Z , la combinación de la misma con otras señales de formato B se pueden llevar a cabo. Luego, puede ser enviada directamente a un descodificador de DirAC de micrófono virtual, o después de la codificación de DirAC la corriente de DirAC mono puede ser enviada a un descodificador de DirAC mono.

30 **[0137]** Algunas modalidades pueden comprender un procesamiento para determinar una señal de audio espacial convertida, la señal de audio espacial convertida tiene un primer componente de audio direccional y un segundo componente de audio direccional, a partir de una señal de audio espacial de entrada, la señal de audio espacial de entrada tiene una representación de audio de entrada y una dirección de llegada de entrada. El procesamiento comprende la etapa de estimar la representación de onda que comprende una medida de campo de onda y una medida de dirección de llegada de onda basada en la representación de audio de entrada y la dirección de llegada de entrada. Además, el procesamiento comprende la etapa de procesamiento de la medida de campo de onda y la medida de dirección de llegada de onda para obtener el primer componente direccional y el segundo componente direccional.

40 **[0138]** En algunas modalidades, el procesamiento para determinar una señal de audio espacial convertida puede estar comprendido por una obtener una corriente de DirAC mono que va a ser convertida a formato B. Opcionalmente W puede ser obtenido de P , cuando está disponible. Si no, se puede efectuar una etapa de aproximar W como una combinación lineal de las señales de audio disponibles. Subsecuentemente, se puede llevar a cabo una etapa de calcular el factor β como factor de ponderación dependiente de la potencia-tiempo inversamente proporcional a la difusividad, por ejemplo de acuerdo con

45
$$\beta(k, n) = \sqrt{1 - \Psi(k, n)} \quad \circ \quad \beta(k, n) = \frac{1 - \sqrt{1 - (1 - \Psi(k, n))^2}}{1 - \Psi(k, n)} .$$

50 **[0139]** El procesamiento puede comprender además la etapa de calcular las señales X , Y y Z a partir de e_{DOA} .

55 **[0140]** Para casos en los cuales $\Psi = 1$, la etapa de obtener W a partir de P puede ser reemplazada al obtener W de P con X , Y y Z siendo cero, obteniendo por lo menos una señal dipolo X , Y o Z a partir de P ; W es cero, respectivamente. Algunas modalidades de la presente invención pueden llevar a cabo el procesamiento de señales en el dominio de formato B, produciendo la ventaja de que el procesamiento de señal avanzada se puede llevar a cabo antes de que las señales de altavoz sean generadas.

60 **[0141]** Dependiendo de ciertos requerimientos de implementación de los procesamientos de la invención, los procesamientos de la invención pueden ser implementados en elementos físicos o elementos de programación. La implementación puede ser efectuada utilizando un medio de almacenamiento digital, y particularmente una memoria instantánea, un disco, un DVD o un CD que tiene señales de control que se pueden leer electrónicamente almacenadas en el mismo, que cooperan con un sistema de computadora programable, de tal manera que los procesamientos de la invención son efectuados. En general, la presente invención es, por consiguiente un código de programa de computadora con un código de programa almacenado en un portador que se puede leer por la máquina, el código de programa es operativo para efectuar los procesamientos de la invención cuando el programa de computadora se ejecuta en una computadora o procesador. En otras palabras, los procesamientos de la

invención son por consiguiente un programa de computadora que tiene un código de programa para efectuar por lo menos uno de los procesamientos de la invención, cuando el programa de computadora se ejecuta en una computadora.

REIVINDICACIONES

1. Aparato (300) adaptado para determinar una señal de audio espacial convertida combinada, presentando la señal de audio espacial convertida combinada al menos una primera componente combinada y una segunda componente combinada, a partir de unas señales de audio espacial de entrada primera y segunda, presentando la primera señal de audio espacial de entrada una primera representación de audio de entrada (P) y una primera dirección de llegada de entrada ($eDOA$), presentando la segunda señal espacial de entrada una segunda representación de audio de entrada y una segunda dirección de llegada de entrada ($eDOA$), que comprende:
- un primer medio (101) adaptado para determinar una primera señal convertida, presentando la primera señal convertida una primera componente omnidireccional (W) y al menos una componente direccional (X, Y, Z), a partir de la primera señal de audio espacial de entrada, comprendiendo el primer medio (101) un estimador (110) adaptado para estimar una primera representación de onda, comprendiendo la primera representación de onda una primera medida de campo de onda ($\beta(k,n)P(k,n)$) y una primera medida de dirección de llegada de onda ($eDOA_x, eDOA_y, eDOA_z$), donde el estimador está adaptado para estimar la primera representación de onda a partir de la primera representación de audio de entrada (P) y la primera dirección de llegada de entrada ($eDOA$); y
- un procesador (120) adaptado para tratar la primera medida de campo de onda ($\beta(k,n)P(k,n)$) y la primera medida de dirección de llegada de onda ($eDOA_x, eDOA_y, eDOA_z$), para obtener la al menos una componente direccional (X, Y, Z), donde la primera componente omnidireccional (W) corresponde a la primera representación de audio de entrada;
- en el cual el primer medio (101) está adaptado para suministrar la primera señal convertida que presenta la primera componente omnidireccional (W) y la al menos una componente direccional (X, Y, Z);
- un segundo medio (102) adaptado para suministrar una segunda señal convertida a partir de la segunda señal de audio espacial de entrada que presenta una segunda componente omnidireccional y al menos otra componente direccional, comprendiendo el segundo medio (102) otro estimador adaptado para estimar una segunda representación de onda, comprendiendo la segunda representación de onda una segunda medida de campo de onda y una segunda medida de dirección de llegada de onda, donde el otro estimador está adaptado para estimar la segunda representación de onda a partir de la segunda representación de audio de entrada y de la segunda dirección de llegada de entrada; y
- otro procesador adaptado para tratar la segunda medida de campo de onda y la segunda medida de dirección de llegada de onda, para obtener la al menos otra componente direccional, donde la segunda componente omnidireccional corresponde a la segunda representación de audio de entrada;
- en el cual el segundo medio (101) está adaptado para suministrar la segunda señal convertida que presenta la segunda componente omnidireccional y la al menos otra componente direccional;
- un generador de efecto de audio (301, 302, 303) adaptado para suministrar la primera componente omnidireccional, para obtener una primera componente representada, o para suministrar la al menos una componente direccional, para obtener la primera componente representada, donde el generador de efecto de audio (301, 302, 303) está adaptado para reverberar la primera componente omnidireccional o la al menos una componente direccional, para obtener la primera componente representada;
- un primer combinador (311) adaptado para combinar la primera componente representada, la primera componente omnidireccional y la segunda componente omnidireccional, o para combinar la primera componente representada, la al menos una componente direccional y la al menos otra componente direccional, para obtener la primera componente combinada; y
- un segundo combinador (312, 313) adaptado para combinar la al menos una componente direccional y la al menos otra componente direccional, o para combinar la primera componente omnidireccional y la segunda componente omnidireccional, para obtener la segunda componente combinada.
2. Aparato (300) según la reivindicación 1, en el cual el estimador está adaptado para estimar la primera medida de campo de onda en términos de una amplitud de campo de onda y de una fase de campo de onda, o en el cual el otro estimador está adaptado para estimar la segunda medida de campo de onda en términos de una amplitud de campo de onda y de una fase de campo de onda.
3. Aparato (300) según cualquiera de las reivindicaciones 1 ó 2, en el cual la primera señal de audio espacial de entrada comprende por otro lado un primer parámetro de difusión (Ψ) y en el cual el estimador está adaptado para estimar la primera medida de campo de onda por otro lado a partir del primer parámetro de difusión (Ψ), o en el cual la segunda señal de audio espacial de entrada comprende por otro lado un segundo parámetro de difusión, y en el cual el otro estimador está adaptado para estimar la segunda medida de campo de onda por otro lado a partir del segundo parámetro de difusión.
4. Aparato (300) según la reivindicación 3, en el cual la al menos una componente direccional comprende unas componentes direccionales primera (X), segunda (Y) y tercera (Z), y en el cual el procesador está adaptado para tratar aún más la primera medida de campo de onda y la primera medida de dirección de llegada de onda, para obtener las componentes direccionales primera (X), segunda (Y) y tercera (Z) para la primera señal convertida, o en el cual la al menos otra componente direccional comprende unas componentes direccionales adicionales primera, segunda y tercera, y en el cual el otro procesador está adaptado para tratar aún más la segunda medida de campo

de onda y la segunda medida de dirección de llegada de onda, para obtener les componentes direccionales adicionales primera, segunda y tercera para la segunda señal convertida.

5 5. Aparato (300) según la reivindicación 4, en el cual el estimador está adaptado para determinar la primera medida de campo de onda a partir de una primera fracción determinada como $\beta_1(k,n)$, de la primera representación de audio de entrada determinada como $P_1(k,n)$, donde k designa un índice de tiempo y n designa un índice de frecuencia, o en el cual el otro estimador está adaptado para determinar la segunda medida de campo de onda a partir de una segunda fracción determinada como $\beta_2(k,n)$, de la segunda representación de audio de entrada determinada como $P_2(k,n)$, donde k designa un índice de tiempo y n designa un índice de frecuencia.

10 6. Aparato (300) según la reivindicación 5, en el cual el procesador está adaptado para obtener una medida compleja de la primera componente direccional como $X_1(k,n)$ y/o de la segunda componente direccional como $Y_1(k,n)$ y/o de la tercera componente direccional como $Z_1(k,n)$ y/o de la primera componente omnidireccional como $W_1(k,n)$ para la primera señal convertida por

$$W_1(k,n) = P_1(k,n)$$

$$X_1(k,n) = \sqrt{2}\beta_1(k,n) \cdot P_1(k,n) \cdot e_{DOA,x,1}(k,n)$$

15

$$Y_1(k,n) = \sqrt{2}\beta_1(k,n) \cdot P_1(k,n) \cdot e_{DOA,y,1}(k,n)$$

$$Z_1(k,n) = \sqrt{2}\beta_1(k,n) \cdot P_1(k,n) \cdot e_{DOA,z,1}(k,n)$$

donde $e_{DOA,x,1}(k,n)$ es una componente de un vector unitario $e_{DOA,1}(k,n)$, que es la primera dirección de llegada de entrada en el eje x de un sistema de coordenadas cartesianas, $e_{DOA,y,1}(k,n)$ es una componente de $e_{DOA,1}(k,n)$ en el eje y y $e_{DOA,z,1}(k,n)$ es una componente de $e_{DOA,1}(k,n)$ en el eje z, o

20 en el cual el otro procesador está adaptado para obtener una medida compleja de la primera otra componente direccional como $X_2(k,n)$ y/o la segunda otra componente direccional como $Y_2(k,n)$ y/o la tercera otra componente direccional como $Z_2(k,n)$ y/o la segunda componente omnidireccional como $W_2(k,n)$ para la segunda señal convertida por

$$W_2(k,n) = P_2(k,n)$$

$$X_2(k,n) = \sqrt{2}\beta_2(k,n) \cdot P_2(k,n) \cdot e_{DOA,x,2}(k,n)$$

$$Y_2(k,n) = \sqrt{2}\beta_2(k,n) \cdot P_2(k,n) \cdot e_{DOA,y,2}(k,n)$$

$$Z_2(k,n) = \sqrt{2}\beta_2(k,n) \cdot P_2(k,n) \cdot e_{DOA,z,2}(k,n)$$

25 donde $e_{DOA,x,2}(k,n)$ es una componente de un vector unitario $e_{DOA,2}(k,n)$ que es la segunda dirección de llegada de entrada en el eje x de un sistema de coordenadas cartesianas, $e_{DOA,y,2}(k,n)$ es una componente de $e_{DOA,2}(k,n)$ en el eje y y $e_{DOA,z,2}(k,n)$ es una componente de $e_{DOA,2}(k,n)$ en el eje z.

30 7. Aparato (300) según cualquiera de las reivindicaciones 5 o 6, en el cual el estimador está adaptado para estimar la primera fracción $\beta_1(k,n)$ a partir del primer parámetro de difusión dado como $\Psi_1(k,n)$, según

$$\beta_1(k, n) = \sqrt{1 - \Psi_1(k, n)}$$

o en el cual el otro estimador está adaptado para estimar la segunda fracción $\beta_2(k, n)$ a partir del segundo parámetro de difusión dado como $\Psi_2(k, n)$, según

$$\beta_2(k, n) = \sqrt{1 - \Psi_2(k, n)}.$$

- 5 **8.** Aparato (300) según cualquiera de las reivindicaciones 5 o 6, en el cual el estimador está adaptado para estimar la primera fracción $\beta_1(k, n)$ a partir del primer parámetro de difusión dado como $\Psi_1(k, n)$, según

$$\beta_1(k, n) = \sqrt{\frac{1 - (1 - \Psi_1(k, n))^2}{1 - \Psi_1(k, n)}},$$

o en el cual el otro estimador está adaptado para estimar la segunda fracción $\beta_2(k, n)$ a partir del segundo parámetro de difusión dado como $\Psi_2(k, n)$, según

$$\beta_2(k, n) = \sqrt{\frac{1 - (1 - \Psi_2(k, n))^2}{1 - \Psi_2(k, n)}}.$$

- 10 **9.** Dispositivo (300) según cualquiera de las reivindicaciones 1 a 8, en el cual la primera señal de audio espacial de entrada corresponde a una señal audio codificada DirAc y en el cual el procesador está adaptado para obtener la primera componente omnidireccional (W) y la al menos una componente direccional (X; Y; Z) en términos de una señal de formato B, o en el cual la segunda señal de audio espacial de entrada corresponde a una señal audio codificada DirAc y en el cual el otro procesador está adaptado para obtener la segunda componente omnidireccional y la al menos otra componente direccional en términos de una señal de formato B.

- 15 **10.** Aparato (300) según cualquiera de las reivindicaciones 1 a 9, en el cual el generador de efecto de audio (301) está adaptado para suministrar una combinación de la primera componente omnidireccional y de la segunda componente omnidireccional, o para suministrar una combinación de la al menos una componente direccional y de la al menos otra componente direccional, para obtener la primera componente representada.

- 20 **11.** Aparato (300) según cualquiera de las reivindicaciones 1 a 10, que comprende por otro lado una primera fase de retardo y de puesta a escala (321), para retardar y/o poner a escala la primera componente omnidireccional y/o la al menos una componente direccional, y/o una segunda fase de retardo y de puesta a escala (322), para retardar y/o poner a escala la segunda componente omnidireccional y/o la al menos otra componente direccional.

- 25 **12.** Aparato (300) según cualquiera de las reivindicaciones 1 a 11, que comprende una pluralidad de medios (100) para convertir una pluralidad de señales audio espaciales de entrada, comprendiendo la pluralidad de medios (100) para convertir una pluralidad de señales audio espaciales de entrada el primer medio (101) y el segundo medio (102), comprendiendo el aparato (300) por otro lado una pluralidad de generadores de efecto de audio, en el cual el número total de generadores de efecto de audio es inferior al número total de medios (100).

- 30 **13.** Procedimiento para determinar una señal de audio espacial convertida combinada, presentando la señal de audio espacial convertida combinada al menos una primera componente combinada y una segunda componente combinada, a partir de unas señales de audio espacial de entrada primera y segunda, presentando la primera señal de audio espacial de entrada una primera representación de audio de entrada y una primera dirección de llegada de entrada, presentando la segunda señal de audio espacial de entrada una segunda representación de audio de entrada y una segunda dirección de llegada de entrada, que comprende las etapas consistentes en
 35 determinar una primera señal convertida, presentando la primera señal convertida una primera componente omnidireccional (W) y al menos una componente direccional (X; Y; Z), a partir de la primera señal de audio espacial de entrada, con ayuda de las sub-etapas consistentes en
 40

estimar una primera representación de onda, comprendiendo la primera representación de onda una primera medida de campo de onda y una primera medida de dirección de llegada de onda, donde la primera representación de onda es estimada a partir de la primera representación de audio de entrada y de la primera dirección de llegada de entrada; y

5 tratar la primera medida de campo de onda y la primera medida de dirección de llegada de onda, para obtener la al menos una componente direccional (X; Y; Z), donde la primera componente omnidireccional (W) corresponde a la primera representación de audio de entrada;

suministrar la primera señal convertida que presenta la primera componente omnidireccional y la al menos una componente direccional;

10 determinar una segunda señal convertida, presentando la segunda señal convertida una segunda componente omnidireccional y al menos otra componente direccional, a partir de la segunda señal de audio espacial de entrada, con ayuda de las sub-etapas consistentes en

estimar una segunda representación de onda, comprendiendo la segunda representación de onda una segunda medida de campo de onda y una segunda medida de dirección de llegada de onda, donde la segunda representación de onda es estimada a partir de la segunda representación de audio de entrada y de la segunda dirección de llegada de entrada; y

15 tratar la segunda medida de campo de onda y la segunda medida de dirección de llegada de onda, para obtener la al menos otra componente direccional, donde la segunda componente omnidireccional corresponde a la segunda representación de audio de entrada;

20 suministrar la segunda señal convertida que presenta la segunda componente omnidireccional y la al menos otra componente direccional; suministrar la primera componente omnidireccional, para obtener una primera componente representada, o suministrar la al menos una componente direccional, para obtener la primera componente representada, donde la etapa de representación comprende el hecho de reverberar la primera componente omnidireccional o la al menos una componente direccional, para obtener la primera componente representada;

25 combinar la primera componente representada, la primera componente omnidireccional y la segunda componente omnidireccional, o combinar la primera componente representada, la al menos una componente direccional y la al menos otra componente direccional, para obtener la primera componente combinada; y

30 combinar la al menos una componente direccional y la al menos otra componente direccional, o combinar la primera componente omnidireccional y la segunda componente omnidireccional, para obtener la segunda componente combinada.

35 **14.** Programa de ordenador que tiene un código de programa para realizar el procedimiento según la reivindicación 13 cuando el código de programa se ejecuta en un procesador de ordenador.

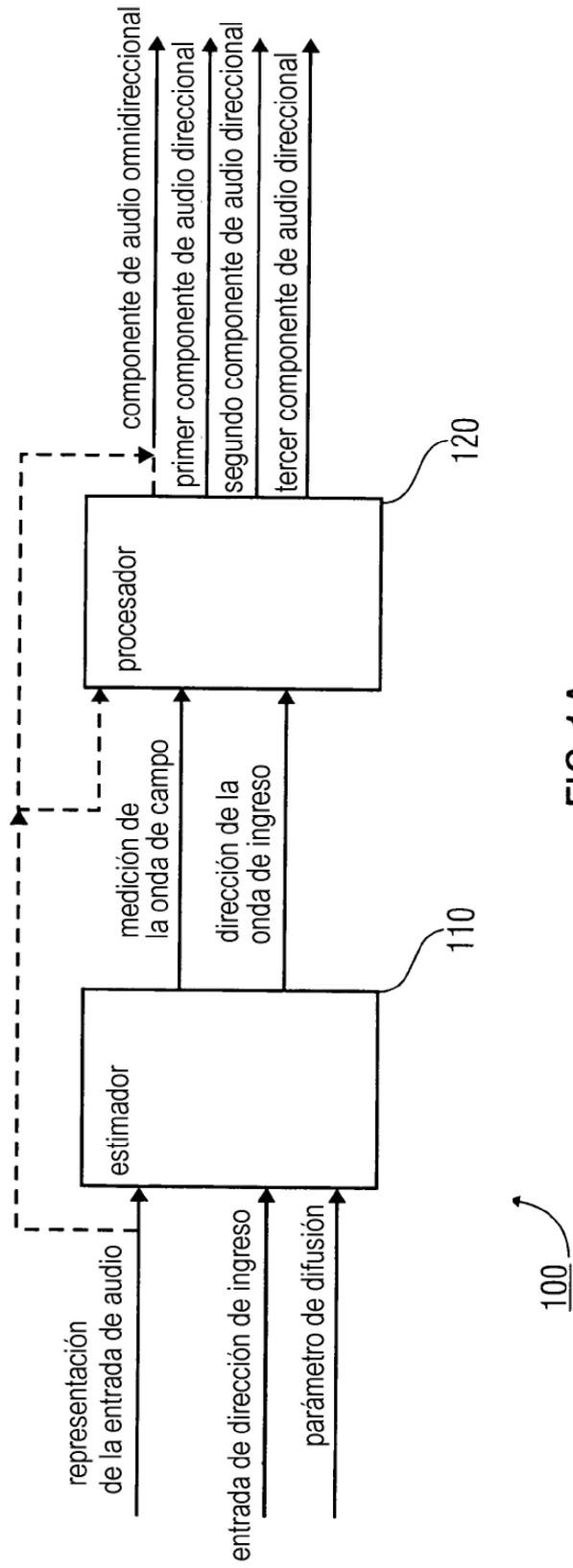


FIG 1A

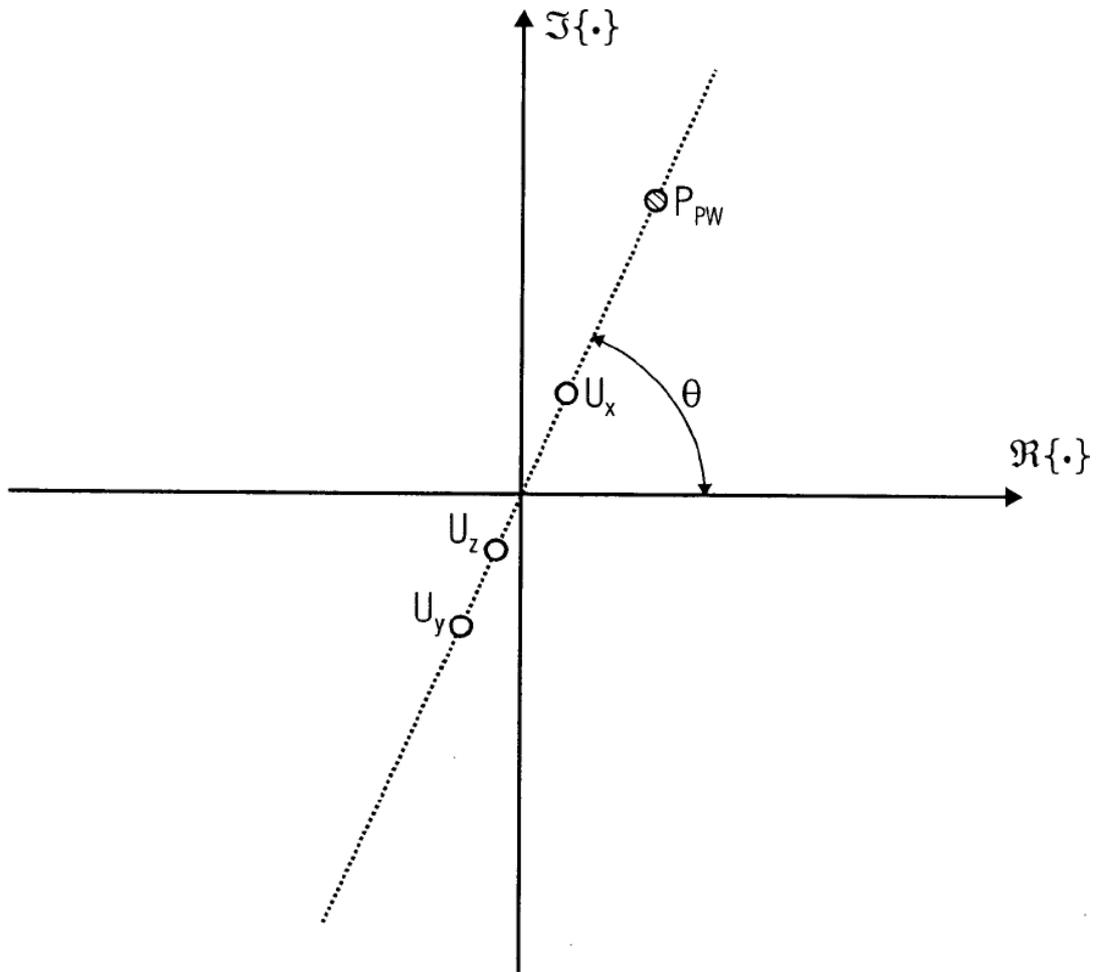


FIG 1B

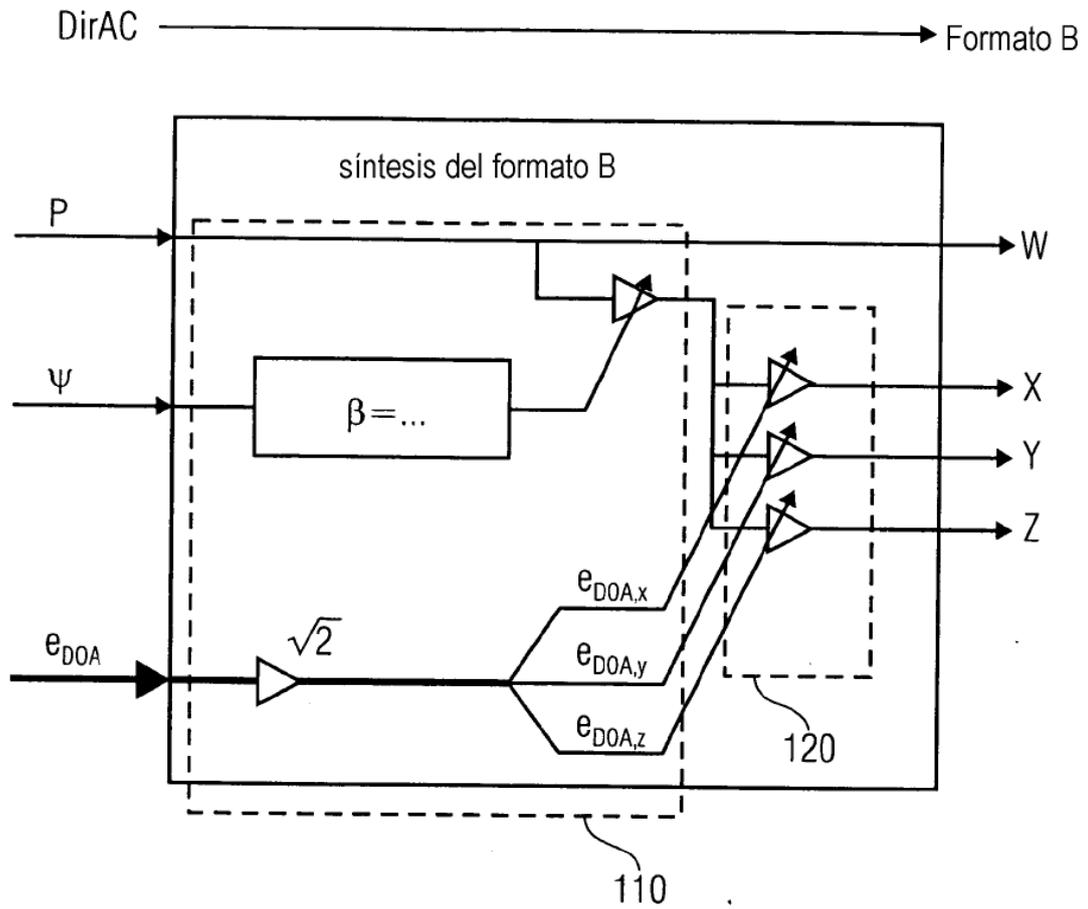


FIG 2

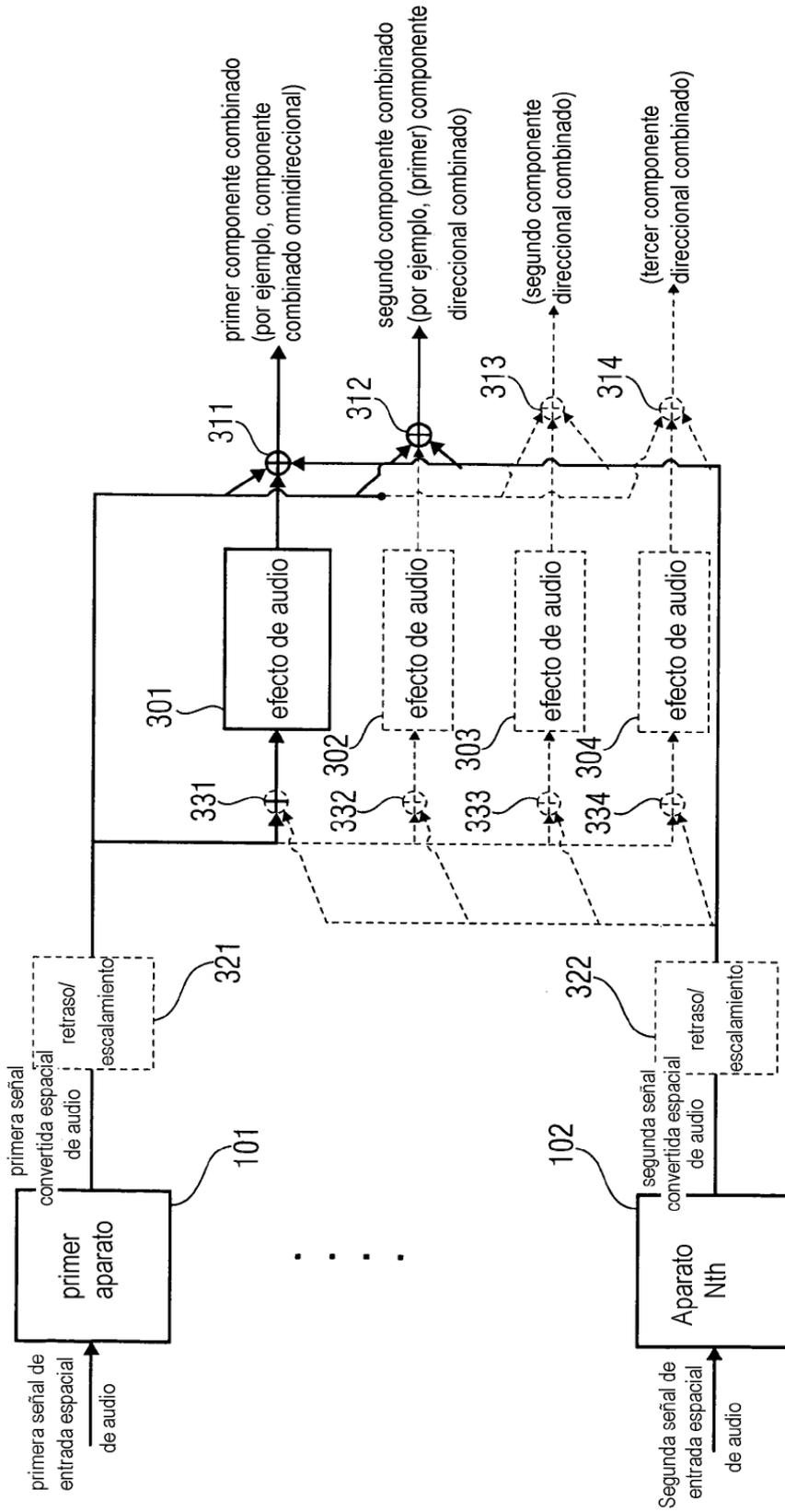


FIG 3

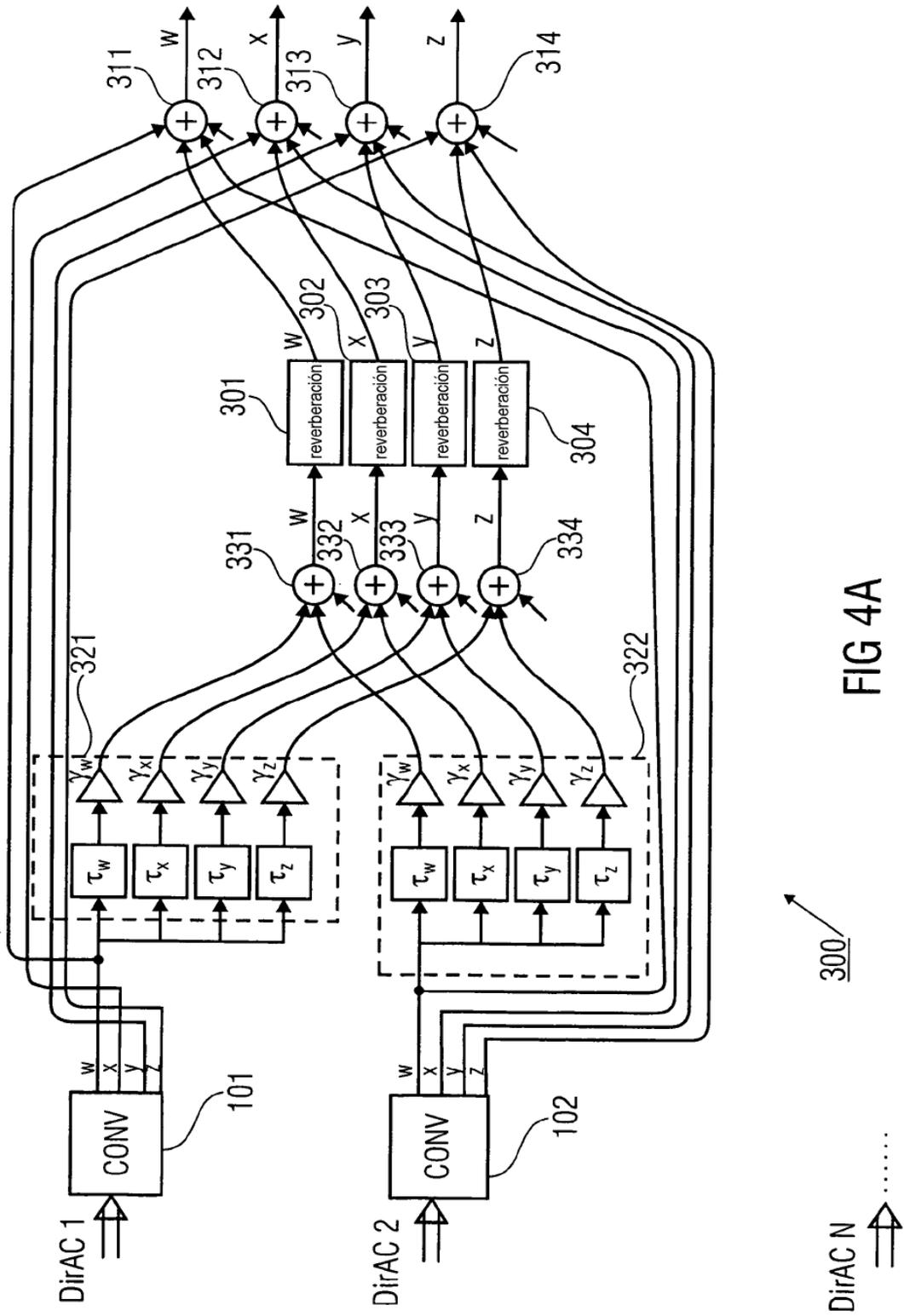


FIG 4A

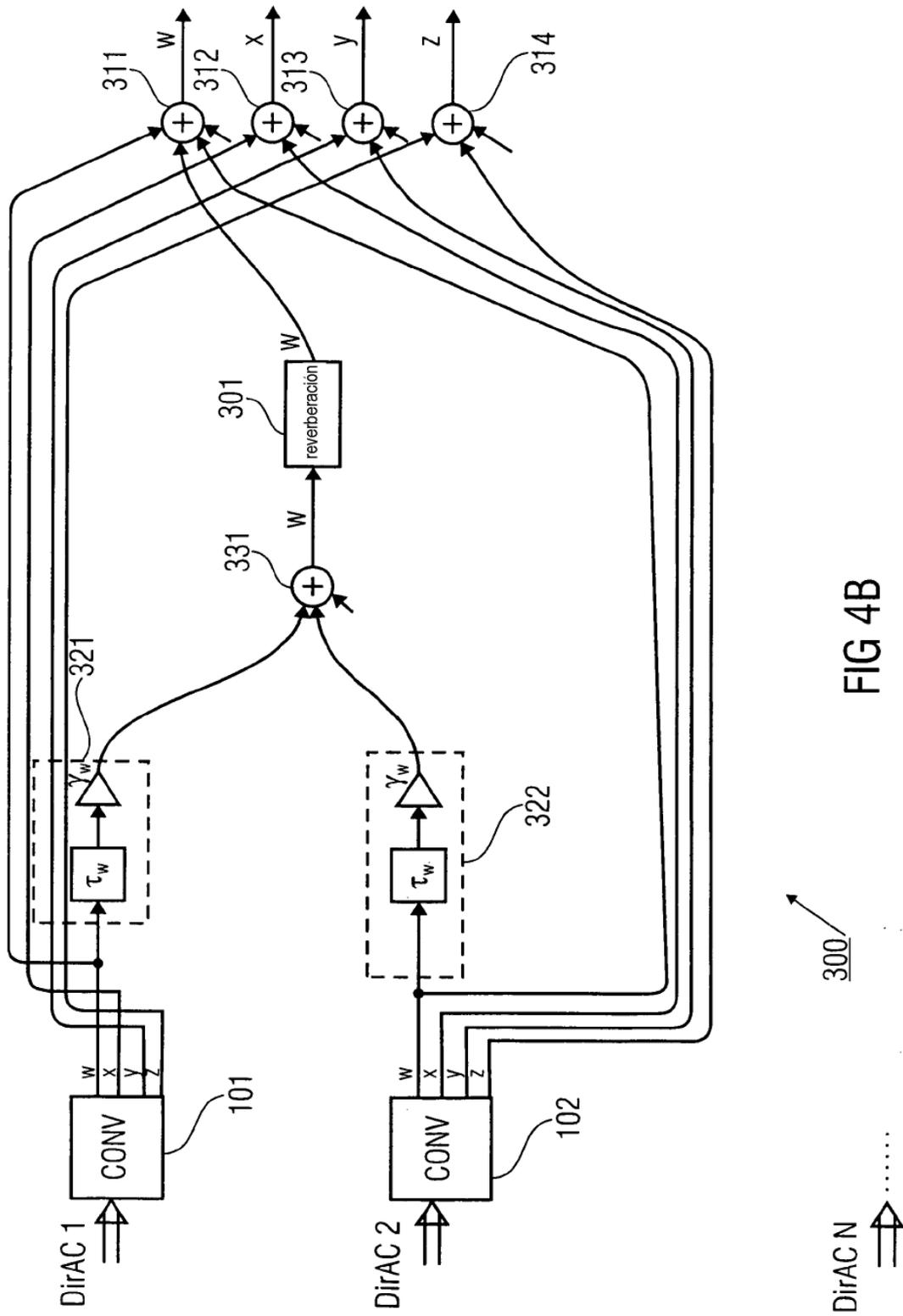


FIG 4B

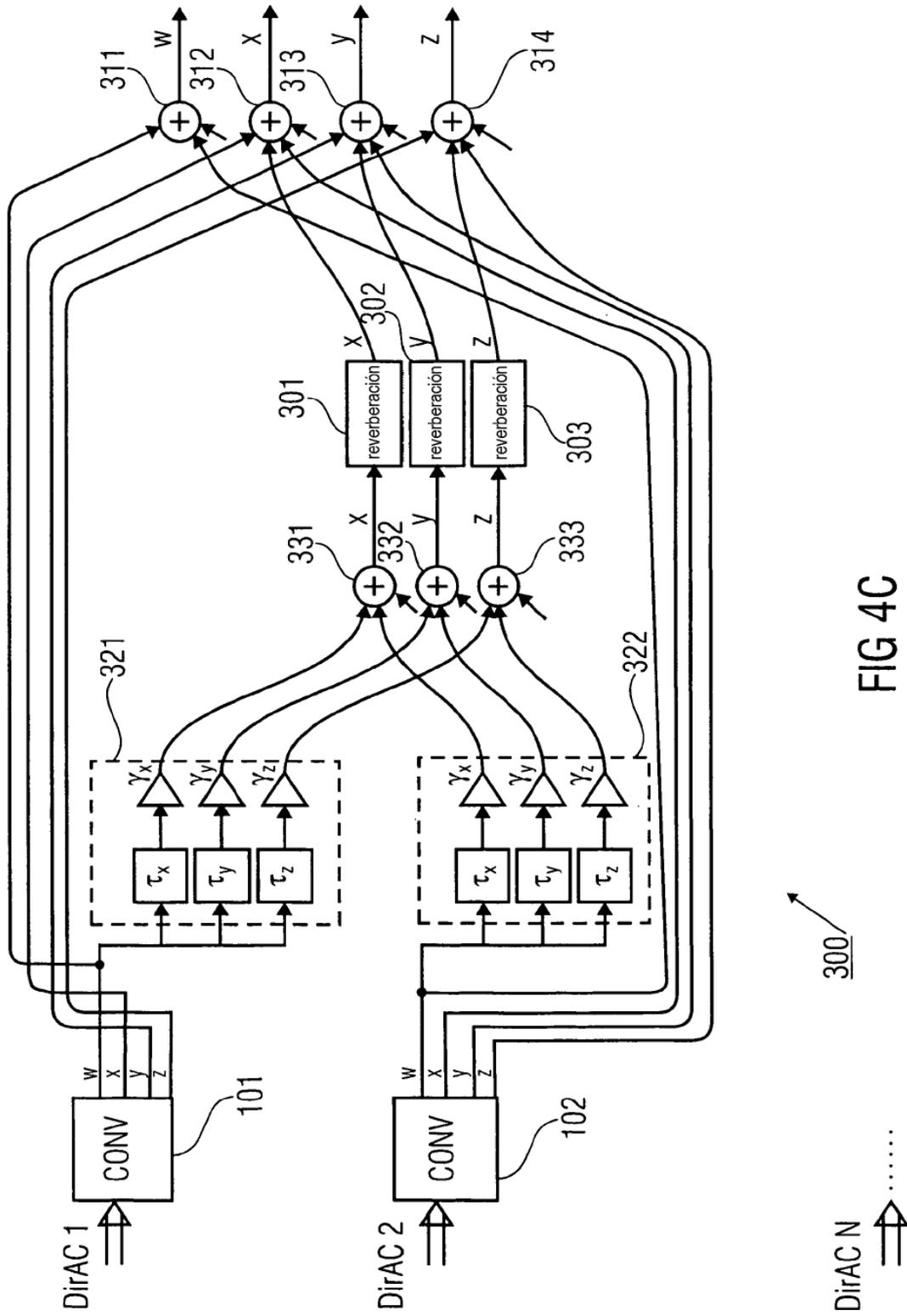


FIG 4C

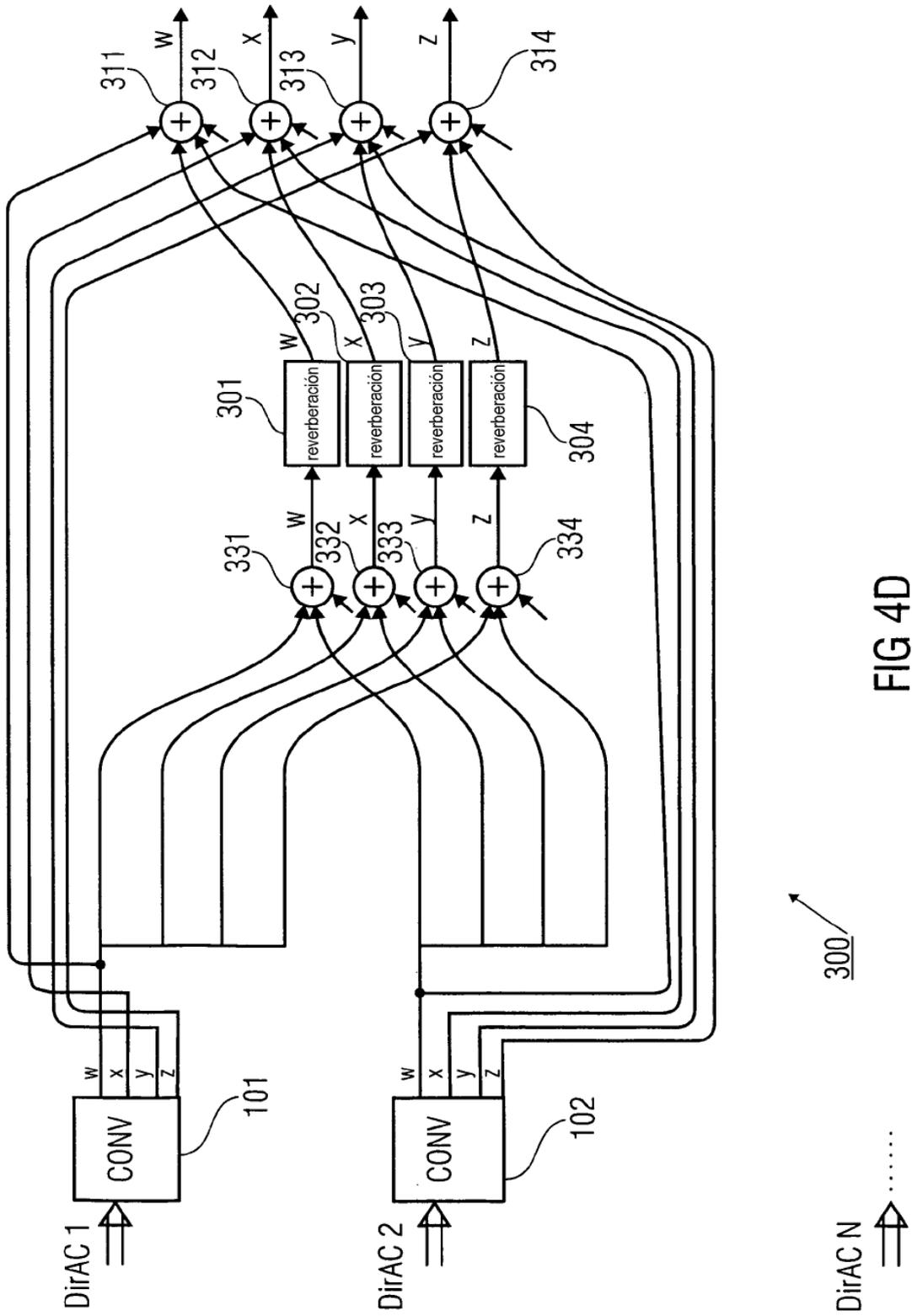


FIG 4D

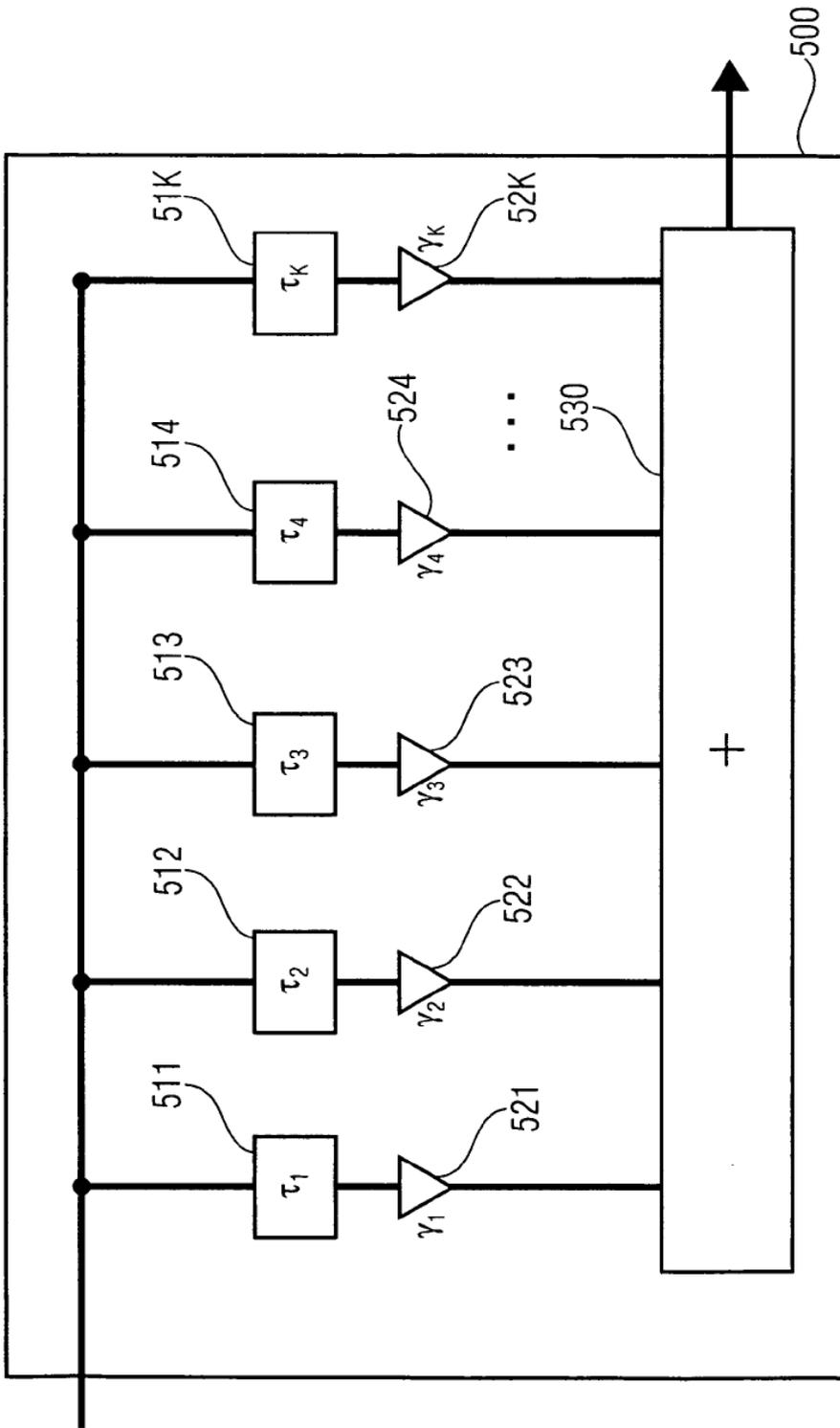


FIG 5

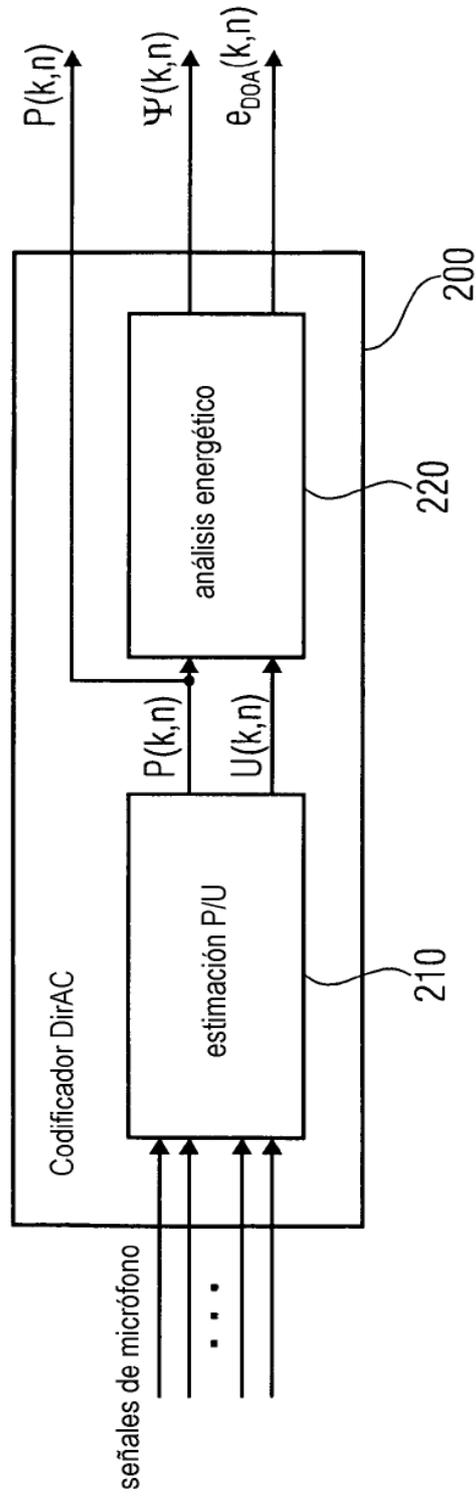


FIG 7