

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 525 839**

51 Int. Cl.:

G10L 19/02 (2013.01)

G10L 19/00 (2013.01)

H04R 1/32 (2006.01)

H04R 3/00 (2006.01)

G10L 19/20 (2013.01)

G10L 19/16 (2013.01)

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.12.2011 E 11801647 (6)**

97 Fecha y número de publicación de la concesión europea: **29.10.2014 EP 2647222**

54 Título: **Adquisición de sonido mediante la extracción de información geométrica de estimativos de dirección de llegada**

30 Prioridad:

03.12.2010 US 419623 P
06.12.2010 US 420099 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.12.2014

73 Titular/es:

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

72 Inventor/es:

HERRE, JÜRGEN;
KÜCH, FABIAN;
KALLINGER, MARKUS;
DEL GALDO, GIOVANNI;
THIERGART, OLIVER;
MAHNE, DIRK;
KUNTZ, ACHIM;
KRATSCHMER, MICHAEL y
CRACIUN, ALEXANDRA

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 525 839 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

Adquisición de sonido mediante la extracción de información geométrica de estimativos de dirección de llegada

- 5 **[0001]** La presente invención se relaciona con el procesamiento de audio y, en particular, con un aparato y procedimiento para la adquisición de sonido mediante la extracción de información geométrica de estimativos de dirección de llegada.
- 10 **[0002]** La grabación de sonido espacial tradicional apunta a capturar un campo sonoro con múltiples micrófonos de tal manera que, del lado de la reproducción, un oyente perciba la imagen sonora como si estuviera en el lugar de la grabación. Las estrategias normales para la grabación de sonido espacial por lo general utilizan micrófonos omnidireccionales espaciados, por ejemplo en al estereofonía AB, o micrófonos direccionales coincidentes, como por ejemplo, en Ambisonics; véase, por ejemplo,
- 15 [1] R. K. Furness, "Ambisonics – An overview," en la 8a Conferencia Internacional de AES, abril de 1990, pág. 181–189.
- [0003]** En cuanto a la reproducción del sonido, estas estrategias no paramétricas derivan las señales de reproducción de audio deseadas (por ej., las señales que han de ser enviadas a los parlantes) directamente de las
- 20 señales de micrófono grabadas.
- [0004]** Por otro lado, se pueden aplicar procedimientos basados en la representación paramétrica de campos sonoros, a los que se hace referencia como codificadores de audio espacial paramétricos. Estos procedimientos emplean con frecuencia matrices de micrófonos para determinar una o más señales de mezcla descendente junto
- 25 con información secundaria espacial que describe el sonido espacial. Son ejemplos la Codificación de Audio direccional (DirAC) o el denominado enfoque de micrófonos de audio espacial (SAM). Se pueden encontrar más detalles sobre la DirAC en
- [2] Pulkki, V., "Directional audio coding in spatial sound reproduction and stereo upmixing," en el Acta de la 28a Conferencia Internacional de AES, pág. 251–258, Piteå, Suecia, Junio 30 – Julio 2 de 2006,
- 30 [3] V. Pulkki, "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc., vol. 55, no. 6, pág. 503–516, Junio de 2007.
- Para más detalles sobre la estrategia de micrófonos de audio espacial, se hace referencia a
- [4] C. Faller: "Microphone Front–Ends for Spatial Audio Coders", en el Acta de la 125a Convención Internacional de AES, San Francisco, Oct. 2008.
- 35 **[0005]** En la DirAC, por ejemplo, el caso de la información de la pista espacial comprende la dirección de llegada (DOA) del sonido y la difusividad del campo sonoro computada en un dominio de dominio del tiempo–frecuencia. Para la reproducción del sonido, se pueden derivar las señales de reproducción de audio sobre la base de la descripción paramétrica. En algunas aplicaciones, la adquisición de sonido espacial apunta a capturar una escena
- 40 sonora completa. En otras aplicaciones, la adquisición de sonido espacial sólo tiene por fin capturar ciertos componentes deseados. Con frecuencia se usan micrófonos para hablar muy cerca para grabar fuentes de sonido individuales con alta relación señal a ruido (SNR) y baja reverberación, en tanto que las configuraciones más distantes como la estereofonía XY representan una manera de capturar la imagen espacial de una escena sonora total. Se puede obtener una mayor flexibilidad en términos de directividad con la formación de haces, donde se
- 45 utiliza una matriz de micrófonos para obtener patrones de captación direccionables. Se obtiene una flexibilidad aun mayor con los procedimientos antes mencionados, como por ejemplo la codificación direccional de audio (DirAC) (ver [2], [3]), en la cual es posible incluir filtros espaciales con patrones arbitrarios de captación, como se describe en
- [5] M. Kallinger, H. Ochsenfeld, G. Del Galdo, F. Küch, D. Mahne, R. Schultz–Amling, y O. Thierngart, "A spatial filtering approach for directional audio coding," en Audio Engineering Society Convention 126, Munich, Alemania, mayo de 2009,
- 50 como así también otras manipulaciones del procesamiento de señales de le la escena sonora; ver, por ejemplo,
- [6] R. Schultz–Amling, F. Küch, O. Thierngart y M. Kallinger, "Acoustical zooming based on a parametric sound field representation," en Audio Engineering Society Convention 128, Londres, R.U., mayo de 2010,
- [7] J. Herre, C. Falch, D. Mahne, G. Del Galdo, M. Kallinger y O. Thierngart, "Interactive teleconferencing combining spatial audio object coding and DirAC technology," en Audio Engineering Society Convention 128, Londres R.U., mayo de 2010.
- 55 **[0006]** Lo que tienen en común todos los conceptos antes mencionados es que los micrófonos están dispuestos en una geometría fija conocida. El espaciamiento entre los micrófonos es el menor posible para la microfónica coincidente, en tanto que normalmente es de sólo unos pocos centímetros en los demás procedimientos. A continuación nos referimos a cualquier aparato para la grabación del sonido espacial con capacidad para rescatar la dirección de llegada del sonido (por ej. una combinación de micrófonos direccionales o una matriz de micrófonos, etc.) como un micrófono espacial.
- 60 **[0007]** Más aun, lo que tienen en común todos los procedimientos precedentemente mencionados es que se limitan a una representación del campo sonoro con respecto a un solo punto, es decir la ubicación de medición. Por consiguiente, los micrófonos necesarios deben estar situados en puntos muy específicos, cuidadosamente
- 65

seleccionados, por ej. cerca de las fuentes o de tal manera que se pueda captar de manera óptima la imagen espacial.

5 **[0008]** Sin embargo, en numerosas aplicaciones, esto no es factible y, por lo tanto, sería ventajoso colocar varios micrófonos a mayor distancia de las fuentes de sonido y de todas maneras poder captar el sonido pretendido.

10 **[0009]** Existen varios procedimientos de reconstrucción de campo para estimar el campo sonoro en un punto del espacio aparte de aquél en el cual se lo midiera. Uno de estos procedimientos es la holografía acústica, descrita en [8] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, 1999.

15 **[0010]** La holografía acústica permite computar el campo sonoro en cualquier punto con un volumen arbitrario, siempre que se conozca la presión sonora y la velocidad de partícula en toda su superficie. Por lo tanto, cuando el volumen es grande, se requiere un número de sensores no práctico por lo grande. Por añadidura, el procedimiento presupone que no hay fuentes sonoras presentes dentro del volumen, lo que hace que el algoritmo sea inviable para nuestras necesidades. La extrapolación de campos de ondas relacionada (remitirse también a [8]) tiene por fin extrapolar el campo sonoro conocidos sobre la superficie de un volumen a regiones externas. No obstante, la precisión de la extrapolación se degrada rápidamente en el caso de distancias de extrapolación más grandes, como así también en el caso de extrapolaciones hacia direcciones ortogonales a la dirección de propagación del sonido; véase

20 [9] A. Kuntz y R. Rabenstein, "Limitations in the extrapolation of wave fields from circular measurements," en 15th European Signal Processing Conference (EUSIPCO 2007), 2007.

25 [10] A. Walther y C. Faller, "Linear simulation of spaced matrices de micrófonos using b-format recordings," en Audio Engineering Society Convention 128, Londres R.U., mayo de 2010, describe un modelo de onda plana, en el cual la extrapolación de campo sólo es posible en puntos lejanos a las fuentes sonoras reales, por ej., cerca del punto de medición.

30 **[0011]** Una desventaja importante de los enfoques tradicionales es que la imagen espacial registrada siempre es relativa al micrófono espacial utilizado. En muchas aplicaciones, no es posible o factible colocar un micrófono espacial en la posición indicada, por ej., cerca de las fuentes del sonido. En este caso, sería más ventajoso colocar múltiples micrófonos espaciales más alejados de la escena sonora y de todas maneras poder capturar el sonido como se pretende.

35 [11] US61/287,596: An apparatus and Method for Converting a First Parametric Spatial Audio Signal into a Second Parametric Spatial Audio Signal, propone un procedimiento para correr la posición real de grabación a otra posición cuando se reproducen por parlantes o auriculares. Sin embargo, esta estrategia se limita a una sencilla escena sonora en la cual se presume que todos los objetos sonoros están a una distancia igual con respecto al micrófono espacial utilizado para la grabación. Más aun, el procedimiento sólo puede aprovechar un micrófono espacial.

40 **[0012]** Otro ejemplo de procedimiento para generar una señal de salida de audio para simular una grabación de la señal de salida de audio mediante un micrófono virtual en una posición virtual configurable se describe en US 2005/0281410 A1.

45 **[0013]** Es un objeto de la presente invención proporcionar conceptos mejorados para la adquisición de sonido mediante la extracción de información geométrica. El objetivo de la presente invención se alcanza por medio de un aparato de acuerdo con la reivindicación 1, un procedimiento de acuerdo con la reivindicación 17 y un programa de computación de acuerdo con la reivindicación 18.

50 **[0014]** De acuerdo con una realización, se presenta un aparato para generar una señal de salida de audio a fin de simular una grabación de un micrófono virtual en una posición virtual configurable de un ambiente. El aparato comprende un estimador de posición de eventos sonoros y un módulo de cómputo de información. El estimador de posición de eventos sonoros está adaptado para estimar la posición de una fuente sonora que indica una posición de una fuente de sonido en el ambiente, donde el estimador de posición de eventos sonoros está adaptado para estimar la posición de la fuente sonora sobre la base de una primera información de dirección provista por un primer micrófono espacial real que está situado en una primera posición del micrófono real en el ambiente, y sobre la base de una segunda información de dirección provista por un segundo micrófono espacial real que está situado en una segunda posición de micrófono real en el ambiente.

55 **[0015]** El módulo de cómputo de información está adaptado para generar la señal de salida de audio sobre la base de una primera señal de entrada de audio grabada que es registrada por el primer micrófono espacial real, sobre la base de la primera posición del micrófono real, sobre la base de la posición virtual del micrófono virtual, y sobre la base de la posición de la fuente sonora.

60 **[0016]** En una forma de realización, el módulo de cómputo de información comprende un compensador de la propagación, donde el compensador de la propagación está adaptado para generar una primera señal de audio modificada mediante la modificación de la primera señal de entrada de audio grabada, sobre la base de la degradación de una primera amplitud entre la fuente del sonido y el primer micrófono espacial real y sobre la base de una segunda degradación de la amplitud entre la fuente del sonido y el micrófono virtual, mediante el ajuste de un

valor de amplitud, un valor de magnitud o un valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio. En una forma de realización, la primera degradación de la amplitud puede ser la degradación de la amplitud de una onda sonora emitida por una fuente de sonido y la segunda degradación de la amplitud puede ser la degradación de la amplitud de la onda sonora emitida por la fuente del sonido.

5
10
[0017] De acuerdo con otra forma de realización, el módulo de cómputo de información comprende un compensador de la propagación que está adaptado para generar una primera señal de audio modificada mediante la modificación de la primera señal de entrada de audio grabada compensando un primer retardo entre la llegada de una onda sonora emitida por la fuente del sonido en el lugar del primer micrófono espacial real y la llegada de la onda sonora en el lugar del micrófono virtual mediante el ajuste de un valor de amplitud, un valor de magnitud o un valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio.

15
[0018] De acuerdo con una forma de realización, se presume el uso de de dos o más micrófonos espaciales, a los que se hace referencia en lo sucesivo como micrófonos espaciales reales. Por cada micrófono espacial real, se puede estimar la DOA del sonido en el dominio del tiempo–frecuencia. De la información reunida por los micrófonos espaciales reales, junto con el conocimiento de su posición relativa, es posible constituir la señal de salida de un micrófono espacial arbitrario ubicado virtualmente a voluntad en el ambiente. En lo sucesivo se hace referencia a este micrófono espacial como micrófono espacial virtual.

20
[0019] Nótese que la Dirección de Llegada (DOA) se puede expresar como ángulo azimutal en el caso del espacio 2D, o por un par de ángulos azimutal y de elevación en 3D. De modo similar, se puede utilizar un vector normal unitario que apunte a la DOA.

25
[0020] En las realizaciones, se presentan medios para capturar sonido de manera espacialmente selectiva, por ej., se puede captar el sonido que se origina en una ubicación específica, como si se hubiera instalado un “micrófono de punto” cercano en esta ubicación. En lugar de instalar, en realidad, este micrófono de punto, se puede simular su señal de salida utilizando dos o más micrófonos espaciales colocados en otras posiciones distantes.

30
[0021] El término “micrófono espacial” se refiere a cualquier aparato para la adquisición de sonido espacial con capacidad para recuperar la dirección de llegada del sonido (por ej. una combinación de micrófonos direccionales, matrices de micrófonos, etc.) .

35
[0022] El término “micrófono no espacial” se refiere a cualquier aparato que no esté adaptado para recuperar la dirección de llegada del sonido, como por ejemplo un micrófono omnidireccional o directivo único.

[0023] Se debe tener en cuenta que el término “micrófono espacial real” se refiere a un micrófono espacial de acuerdo con lo definido anteriormente, que existe físicamente.

40
[0024] En lo que respecta al micrófono espacial virtual, cabe señalar que el micrófono espacial virtual puede representar cualquier tipo de micrófono o combinación de micrófonos conveniente, por ej. puede representar, por ejemplo, un micrófono omnidireccional único, un micrófono direccional, un par de micrófonos direccionales como se utiliza en los micrófonos estéreo comunes, aunque también una matriz de micrófonos.

45
50
[0025] La presente invención se basa en el hallazgo de que cuando se utilizan dos o más micrófonos espaciales reales, es posible estimar la posición en el espacio 2D o 3D de los eventos sonoros, y por consiguiente se puede obtener la localización de la posición. Utilizando las posiciones determinadas de los eventos sonoros, se puede computar la señal de sonido que habría sido registrada por un micrófono espacial virtual situado y orientado arbitrariamente en el espacio, como así también la correspondiente información complementaria espacial, tal como la Dirección de Llegada desde el punto de vista del micrófono espacial virtual.

55
60
65
[0026] Para este fin, se puede presumir que cada evento sonoro representa una fuente de sonido puntual, por ej. una fuente de sonido de punto isotrópico. En lo sucesivo la expresión “fuente sonora real” se refiere a una fuente de sonido real con existencia física en el ambiente de grabación, como por ejemplo personas que hablan o instrumentos musicales, etc.. Por el contrario, en lo sucesivo utilizamos la expresión “fuente sonora” o “evento sonoro” para referirnos a una fuente de sonido efectivo, que está activa en un determinado instante de tiempo o en una determinada caja de tiempo–frecuencia, donde las fuentes del sonido pueden representar, por ejemplo, fuentes sonoras reales o fuentes de imágenes espejadas. De acuerdo con una forma de realización, se supone implícitamente que se puede modelar la escena sonora como multitud de dichos eventos sonoros o como fuentes sonoras puntuales. Más aun, se puede suponer que cada fuente está activa sólo dentro de una ranura específica de tiempo y frecuencia en una representación de tiempo–frecuencia predefinida. La distancia entre los micrófonos espaciales reales puede ser tal que la diferencia temporal consiguiente de los tiempos de propagación sea más corta que la resolución temporal de la representación de tiempo–frecuencia. Esta última presunción garantiza que un determinado evento sonoro sea captado por todos los micrófonos espaciales dentro de la misma ranura de tiempo. Esto implica que las DOAs estimadas en diferentes micrófonos espaciales correspondientes a la misma ranura de tiempo–frecuencia corresponden al mismo evento sonoro. Esta presunción no es difícil de cumplir con micrófonos espaciales reales colocados a pocos metros entre sí en recintos grandes (tales como salas de estar o salones de conferencias) con una resolución temporal de incluso unos pocos ms.

5 **[0027]** Se pueden utilizar matrices de micrófonos para localizar fuentes sonoras. Las fuentes sonoras localizadas pueden tener diferentes interpretaciones físicas dependiendo de su naturaleza. Cuando las matrices de micrófonos reciben sonido directo, pueden tener la capacidad de localizar la posición de una fuente sonora verdadera (por ej. voces humanas). Al recibir reflexiones las matrices de micrófonos, éstas pueden localizar la posición de una fuente de imagen espejada. Las fuentes de imágenes espejadas también son fuentes de sonido.

10 **[0028]** Se da a conocer un procedimiento paramétrico que puede estimar la señal sonora de un micrófono virtual situado en una ubicación arbitraria. A diferencia de los procedimientos antes descritos, el procedimiento propuesto no apunta directamente a la reconstrucción del campo sonoro, sino que tiene por objeto producir un sonido perceptualmente similar al que sería captado por un micrófono situado físicamente en ese punto. Esto se puede obtener empleando un modelo paramétrico del campo sonoro sobre la base de fuentes sonoras puntuales, por ej. de punto isotrópico (IPLS). Se puede obtener la información geométrica requerida, es decir la posición instantánea de todas las IPLS, llevando a cabo la triangulación de las direcciones de llegada estimadas con dos o más matrices de micrófonos distribuidas. Esto se puede lograr obteniendo el conocimiento de la posición y orientación relativas de las matrices. Pese a ello, no es necesario conocer de antemano el número y la posición de las fuentes sonoras reales (por ej., las voces humanas). Dada la naturaleza paramétrica de los conceptos propuestos, por ejemplo el aparato o procedimiento propuesto, el micrófono virtual puede poseer un patrón de directividad arbitrario, como así también comportamientos físicos o no físicos arbitrarios, por ej., con respecto a la degradación de la presión con la distancia. La estrategia presentada ha sido verificada estudiando la precisión de la estimación de parámetros basada en mediciones del entorno reverberante.

25 **[0029]** Mientras que las técnicas de grabación convencionales para el audio espacial son limitadas en la medida en que la imagen espacial obtenida siempre es relativa a la posición en la cual se han colocado físicamente los micrófonos, las realizaciones de la presente invención toman en cuenta que, en numerosas aplicaciones, es ventajoso colocar los micrófonos fuera de la escena sonora y, de todas maneras, pueden capturar el sonido desde una perspectiva arbitraria. De acuerdo con las realizaciones, se presentan conceptos que colocan virtualmente un micrófono virtual en un punto arbitrario del espacio, computando una señal perceptualmente similar a la que se habría captado si se hubiera colocado físicamente el micrófono en la escena sonora. Las realizaciones pueden aplicar conceptos que pueden emplear un modelo paramétrico del campo sonoro sobre la base de las fuentes sonoras puntuales, por ej. fuentes sonoras puntuales isotrópicas. La información geométrica necesaria puede ser colectada por dos o más matrices de micrófonos distribuidas.

35 **[0030]** De acuerdo con una forma de realización, el estimador de posición de eventos sonoros puede estar adaptado para estimar la posición de la fuente sonora sobre la base de una primera dirección de llegada de la onda sonora emitida por la fuente del sonido en la primera posición del micrófono real como primera información de dirección y sobre la base de una segunda dirección de llegada de la onda sonora en la segunda posición de micrófono real como segunda información de dirección.

40 **[0031]** En otra forma de realización, el módulo de cómputo de información puede comprender un módulo de cómputo de información complementaria espacial para computar la información complementaria espacial. El módulo de cómputo de información puede estar adaptado para estimar la dirección de llegada o una intensidad sonora activa en el lugar del micrófono virtual como información complementaria espacial, sobre la base de un vector de posición del micrófono virtual y sobre la base de un vector de posición del evento sonoro.

45 **[0032]** De acuerdo con una forma de realización adicional, el compensador de la propagación puede estar adaptado para generar la primera señal de audio modificada en un dominio del tiempo–frecuencia, compensando el primer retardo o degradación de la amplitud entre la llegada de la onda sonora emitida por la fuente del sonido en el lugar del primer micrófono espacial real y la llegada de la onda sonora en el lugar del micrófono virtual mediante el ajuste de dicho valor de magnitud de la primera señal de entrada de audio grabada que está representada en un dominio del tiempo–frecuencia.

55 **[0033]** En una forma de realización, el compensador de la propagación puede estar adaptado para llevar a cabo la compensación de la propagación mediante la generación de un valor de magnitud modificado de la primera señal de audio modificada aplicando la fórmula:

$$P_v(k, n) = \frac{d_1(k, n)}{s(k, n)} P_{\text{ref}}(k, n)$$

60 donde $d_1(k, n)$ es la distancia entre la posición del primer micrófono espacial real y la posición del evento sonoro, donde $s(k, n)$ es la distancia entre la posición virtual del micrófono virtual y la posición de la fuente sonora del evento sonoro, donde $P_{\text{ref}}(k, n)$ es un valor de magnitud de la primera señal de entrada de audio grabada que está representada en un dominio del tiempo–frecuencia y donde $P_v(k, n)$ es el valor de magnitud modificado.

65 **[0034]** En otra forma de realización, el módulo de cómputo de información puede comprender además un combinador, donde el compensador de la propagación puede estar adaptado además para modificar una segunda

señal de entrada de audio grabada, que es registrada por el segundo micrófono espacial real, mediante la compensación de un segundo retardo o degradación de la amplitud entre la llegada de la onda sonora emitida por la fuente del sonido en el lugar del segundo micrófono espacial real y la llegada de la onda sonora en el lugar del micrófono virtual, mediante el ajuste de un valor de amplitud, un valor de magnitud o un valor de fase de la segunda señal de entrada de audio grabada para obtener una segunda señal de audio modificada, y donde el combinador puede estar adaptado para generar una señal combinatoria combinando la primera señal de audio modificada y la segunda señal de audio modificada, para obtener la señal de salida de audio.

[0035] De acuerdo con otra forma de realización, el compensador de la propagación puede estar adaptado asimismo para modificar una o más señales de entrada de audio grabadas adicionales, que es registrada por dichos uno o más micrófonos espaciales reales adicionales, compensando los retardos entre una llegada de la onda sonora en el lugar del micrófono virtual y la llegada de la onda sonora emitida por la fuente del sonido en cada uno de los micrófonos espaciales reales adicionales. cada uno de los retardos o degradaciones de la amplitud puede ser compensado mediante el ajuste de un valor de amplitud, un valor de magnitud o un valor de fase de cada una de las señales de entrada de audio grabadas adicionales para obtener una pluralidad de terceras señales de audio modificadas. El combinador puede estar adaptado para generar una señal combinatoria combinando la primera señal de audio modificada y la segunda señal de audio modificada y la pluralidad de terceras señales de audio modificadas, para obtener la señal de salida de audio.

[0036] En otra forma de realización, el módulo de cómputo de información puede comprender una unidad de ponderación espectral para generar una señal de audio ponderada mediante la modificación de la primera señal de audio modificada dependiendo de la dirección de llegada de la onda sonora en la posición virtual del micrófono virtual y dependiendo de la orientación virtual del micrófono virtual para obtener la señal de salida de audio, donde la primera señal de audio modificada puede ser modificada en un dominio del tiempo–frecuencia.

[0037] Más aun, el módulo de cómputo de información puede comprender una unidad de ponderación espectral para generar una señal de audio ponderada modificando la señal combinatoria dependiendo de la dirección de llegada o la onda sonora en la posición virtual del micrófono virtual y una orientación virtual del micrófono virtual para obtener la señal de salida de audio, donde la señal combinatoria puede ser modificada en un dominio del tiempo–frecuencia.

[0038] De acuerdo con otra forma de realización, la unidad de ponderación espectral puede estar adaptada para aplicar el factor de ponderación

$\alpha + (1-\alpha)\cos(\varphi_v(k, n))$, o el factor de ponderación

$$0,5 + 0,5 \cos(\varphi_v(k, n))$$

a la señal de audio ponderada,

donde $\varphi_v(k, n)$ indica un vector de dirección de llegada de la onda sonora emitida por la fuente del sonido en la posición virtual del micrófono virtual.

[0039] En una forma de realización, el compensador de la propagación está adaptado además para generar una tercera señal de audio modificada modificando una tercera señal de audio de entrada grabada registrada por un micrófono omnidireccional mediante la compensación de un tercer retardo o degradación de la amplitud entre una llegada de la onda sonora emitida por la fuente del sonido en el micrófono omnidireccional y la llegada de la onda sonora en el lugar del micrófono virtual mediante el ajuste de un valor de amplitud, un valor de magnitud o un valor de fase de la tercera señal de audio de entrada grabada, para obtener la señal de salida de audio.

[0040] En otra forma de realización, el estimador de posición de eventos sonoros puede estar adaptado para estimar la posición de una fuente sonora en un ambiente tridimensional.

[0041] Más aun, de acuerdo con otra forma de realización, el módulo de cómputo de información puede comprender asimismo una unidad de cómputo de la difusividad que está adaptada para estimar una intensidad sonora difusa en el lugar del micrófono virtual o una intensidad sonora directa en el lugar del micrófono virtual.

[0042] La unidad de cómputo de la difusividad puede estar adaptada, de acuerdo con una forma de realización adicional, para estimar la intensidad sonora difusa $E_{dif}^{(VM)}$ en el lugar del micrófono virtual by applying the formula:

$$E_{dif}^{(VM)} = \frac{1}{N} \sum_{i=1}^N E_{dif}^{(SMi)}$$

en la cual N es el número de una pluralidad de micrófonos espaciales reales que comprende el primero y segundo micrófonos espaciales reales, y donde $E_{\text{dir}}^{(\text{SM}i)}$ es la intensidad sonora difusa en el i-ésimo micrófono espacial real.

5 **[0043]** En otra forma de realización, la unidad de cómputo de la difusividad puede estar adaptada para estimar la intensidad sonora directa aplicando la fórmula:

$$E_{\text{dir},i}^{(\text{VM})} = \left(\frac{\text{distancia SMi - IPLS}}{\text{distancia VM - IPLS}} \right)^2 E_{\text{dir}}^{(\text{SM}i)}$$

10 en la cual “distancia SMi – IPLS” es la distancia entre una posición del i-ésimo micrófono real y la posición de la fuente sonora, donde “distancia VM – IPLS” es la distancia entre la posición virtual y la posición de la fuente sonora, y donde $E_{\text{dir}}^{(\text{SM}i)}$ es la intensidad directa en el i-ésimo micrófono espacial real.

15 **[0044]** Más aun, de acuerdo con otra forma de realización, la unidad de cómputo de la difusividad puede estar adaptada además para estimar la difusividad en el lugar del micrófono virtual mediante el cálculo de la intensidad sonora difusa en el lugar del micrófono virtual y la intensidad sonora directa en el lugar del micrófono virtual y aplicando la fórmula:

$$\Psi^{(\text{VM})} = \frac{E_{\text{dif}}^{(\text{VM})}}{E_{\text{dif}}^{(\text{VM})} + E_{\text{dir}}^{(\text{VM})}}$$

20 en la cual $\Psi^{(\text{VM})}$ indica la difusividad en el lugar del micrófono virtual que se está estimando, donde $E_{\text{dif}}^{(\text{VM})}$ indica la intensidad sonora difusa que se estima y donde $E_{\text{dir}}^{(\text{VM})}$ indica la intensidad sonora directa que se está estimando.

[0045] A continuación se describen las realizaciones preferidas de la presente invención, en las cuales:

25 La Fig. 1 ilustra un aparato para generar una señal de salida de audio de acuerdo con una forma de realización,

La Fig. 2 ilustra las entradas y salidas de un aparato y un procedimiento para generar una señal de salida de audio de acuerdo con una forma de realización,

30 La Fig. 3 ilustra la estructura básica de un aparato de acuerdo con una forma de realización que comprende un estimador de posición de eventos sonoros y un módulo de cómputo de información,

La Fig. 4 expone un ejemplo de escanerio en el que inos micrófonos espaciales reales se representan como tiras lineales uniformes de 3 micrófonos cada una,

35 La Fig. 5 ilustra dos micrófonos espaciales en 3D para estimar la dirección de llegada en el espacio 3D,

La Fig. 6 ilustra una geometría en la que una Fuente de sonido de tipo puntual isotrópica del contenedor de tiempo-frecuencia (k, n) actual está localizado en una posición $p_{\text{IPLS}}(k, n)$,

40 La Fig. 7 ilustra el módulo de cómputo de información de acuerdo con una forma de realización,

La Fig. 8 ilustra el módulo de cómputo de información de acuerdo con otra forma de realización,

45 La Fig. 9 expone dos micrófonos espaciales reales, un evento sonoro localizado y una posición de un micrófono espacial virtual, junto con los correspondientes retardos y degradaciones de la amplitud,

La Fig. 10 ilustra, cómo obtener la dirección de llegada con respecto a un micrófono virtual de acuerdo con una forma de realización,

50 La Fig. 11 ilustra una manera posible de derivar la DOA del sonido desde el punto de vista del micrófono virtual de acuerdo con una forma de realización,

55 La Fig. 12 ilustra un bloque de cálculo de información que comprende además una unidad de cómputo de la difusividad de acuerdo con una forma de realización,

La Fig. 13 ilustra una unidad de cómputo de la difusividad de acuerdo con una forma de realización,

La Fig. 14 ilustra una situación en que no es posible la estimación de la posición de los eventos sonoros y

Las Figs. 15a–15c ilustran situaciones en que dos matrices de micrófonos reciben sonido directo, sonido reflejado por una pared y sonido difuso.

[0046] La Fig. 1 ilustra un aparato para generar una señal de salida de audio para simular una grabación de un micrófono virtual en una posición virtual configurable $posVmic$ en un ambiente. El aparato comprende un estimador de posición de eventos sonoros 110 y un módulo de cómputo de información 120. El estimador de posición de eventos sonoros 110 recibe una primera información de la dirección $di1$ de un primer micrófono espacial real y una segunda información de la dirección $di2$ desde un segundo micrófono espacial real. El estimador de posición de eventos sonoros 110 está adaptado para estimar la posición de una fuente sonora ssp que indica una posición de una fuente de sonido en el ambiente, donde la fuente del sonido emite una onda sonora, donde el estimador de posición de eventos sonoros 110 está adaptado para estimar la posición de la fuente sonora ssp sobre la base de una primera información de dirección $di1$ provista por un primer micrófono espacial real que está situado en una primera posición del micrófono real $pos1mic$ en el ambiente, y sobre la base de una segunda información de dirección $di2$ provista por un segundo micrófono espacial real que está situado en una segunda posición de micrófono real en el ambiente. El módulo de cómputo de información 120 está adaptado para generar la señal de salida de audio sobre la base de una primera señal de entrada de audio grabada $is1$ que es registrada por el primer micrófono espacial real, sobre la base de la primera posición del micrófono real $pos1mic$ y sobre la base de la posición virtual $posVmic$ del micrófono virtual. El módulo de cómputo de información 120 comprende un compensador de la propagación que está adaptado para generar una primera señal de audio modificada mediante la modificación de la primera señal de entrada de audio grabada $is1$ compensando un primer retardo o degradación de la amplitud entre una llegada de la onda sonora emitida por la fuente del sonido en el lugar del primer micrófono espacial real y la llegada de la onda sonora en el lugar del micrófono virtual mediante el ajuste de un valor de amplitud, un valor de magnitud o un valor de fase de la primera señal de entrada de audio grabada $is1$, para obtener la señal de salida de audio.

[0047] La Fig. 2 ilustra las entradas y salidas de un aparato y un procedimiento de acuerdo con una forma de realización. Se alimenta información de dos o más micrófonos espaciales reales 111, 112, ..., 11N al aparato/es procesada por el procedimiento. Esta información comprende las señales de audio captadas por los micrófonos espaciales reales, como así también información de la dirección de los micrófonos espaciales reales, por ej. estimativos de la dirección de llegada (DOA). Las señales de audio y la información de dirección, como por ejemplo los estimativos de dirección de llegada, pueden ser expresadas en un dominio del tiempo–frecuencia. Si, por ejemplo, es conveniente una reconstrucción geométrica 2D y se opta por un dominio de STFT tradicional (transformación de fourier de corto tiempo) para la representación de las sañes, la DOA se puede expresar en términos de ángulos azimutales dependientes de k y n , es decir los índices de frecuencia y tiempo.

[0048] En las realizaciones, la localización del evento sonoro en el espacio, como así también la descripción de la posición del micrófono virtual se puede llevar a cabo basándose en las posiciones y orientaciones de los micrófonos espaciales reales y virtuales en un sistema de coordenadas común. Esta información puede estar representada por las entradas 121 ... 12N y la entrada 104 de la Fig. 2. La entrada 104 puede especificar además la característica del micrófono espacial virtual, por ej., su posición y patrón de captación, como se describe más adelante. Si el micrófono espacial virtual comprende múltiples sensores virtuales, se pueden considerar sus posiciones y los correspondientes patrones de captación diferentes.

[0049] La salida del aparato o un procedimiento correspondiente puede consistir, en caso de ser conveniente, una o más señales sonoras 105, que pueden haber sido captadas por un micrófono espacial definido y colocado de acuerdo con lo especificado en 104. Más aun, el aparato (o, más precisamente el procedimiento) puede producir como salida la información complementaria espacial correspondiente 106 que puede ser estimada empleando el micrófono espacial virtual.

[0050] La Fig. 3 ilustra un aparato de acuerdo con una forma de realización, que comprende dos unidades principales de procesamiento, un estimador de posición de eventos sonoros 201 y un módulo de cómputo de información 202. El estimador de posición de eventos sonoros 201 puede llevar a cabo la reconstrucción geométrica sobre la base de las DOAs comprendidas en las entradas 111 ... 11N y basándose en el conocimiento de la posición y orientación de los micrófonos espaciales reales, donde se han computado las DOAs. La salida del estimador de posición de eventos sonoros 205 comprende los estimativos de posición (ya sea en 2D o 3D) de las fuentes de sonido en que tienen lugar los eventos sonoros por cada caja de tiempo y frecuencia. El segundo bloque de procesamiento 202 es un módulo de cómputo de información. De acuerdo con la forma de realización de la Fig. 3, el segundo bloque de procesamiento 202 computa una señal de micrófono virtual y la información complementaria espacial. Por lo tanto, también se lo denomina bloque de cómputo de señal de micrófono virtual e información complementaria 202. El bloque de cómputo de señal de micrófono virtual e información complementaria 202 utiliza las posiciones de los eventos sonoros 205 para procesar las señales de audio comprendidas en 111...11N para dar salida a la señal de audio del micrófono virtual 105. El bloque 202, en caso de ser necesario, puede computar asimismo la información complementaria espacial 106 correspondiente al micrófono espacial virtual. Las siguientes realizaciones ilustran posibilidades de cómo pueden funcionar los bloques 201 y 202.

[0051] A continuación se describe en forma más detallada la estimación de posiciones por un estimador de posición de eventos sonoros de acuerdo con una forma de realización.

5 [0052] Dependiendo de la dimensionalidad del problema (2D o 3D) y del número de micrófonos espaciales, hay varias soluciones posibles para la estimación de posición.

10 [0053] En caso de existir dos micrófonos espaciales en 2D, (el caso más sencillo posible) es posible una triangulación simple. La Fig. 4 expone una situación ilustrativa en la cual se muestran los micrófonos espaciales reales en forma de Matrices Uniformes Lineales (ULAs) de 3 micrófonos cada una. Se computa la DOA, que se expresa en términos de ángulos azimutales $a_1(k, n)$ y $a_2(k, n)$, correspondiente a la caja de tiempo–frecuencia (k, n) . Esto se obtiene empleando un estimador de DOA apropiado tal como ESPRIT,

15 [13] R. Roy, A. Paulraj, and T. Kailath, "Direction-of-arrival estimation by subspace rotation methods – ESPRIT," en IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Stanford, CA, USA, abril de 1986, o (raíz) MÚSICA, ver

[14] R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pág. 276–280, 1986 a las señales de presión transformadas al dominio del tiempo–frecuencia.

20 [0054] En la Fig. 4, se ilustran dos micrófonos espaciales reales, en este caso dos matrices de micrófonos espaciales reales 410, 420. Las dos DOAs estimadas $a_1(k, n)$ y $a_2(k, n)$ están representadas por dos líneas, una primera línea 430 que representa la DOA $a_1(k, n)$ y una segunda línea 440 que representa la DOA $a_2(k, n)$. La triangulación es posible mediante sencillas consideraciones geométricas conociendo la posición y orientación de cada matriz.

25 [0055] La triangulación falla cuando las dos líneas 430, 440 son exactamente paralelas. Sin embargo, en aplicaciones reales, esto es muy poco probable. No obstante, no todos los resultados de la triangulación corresponden a una posición física o factible para el evento sonoro en el espacio en cuestión. Por ejemplo, la posición estimada del evento sonoro podría estar demasiado lejos o incluso fuera del espacio presunto, indicado que probablemente las DOAs no corresponden a ningún evento sonoro que pueda ser físicamente interpretado con el modelo utilizado. Esos resultados pueden ser causados por ruido de los sensores o por una reverberación demasiado fuerte del recinto. Por lo tanto, de acuerdo con una forma de realización, esos resultados no deseados son identificados con banderas de tal manera que el módulo de cómputo de información 202 los pueda tratar correctamente.

30 [0056] La Fig. 5 ilustra una situación en la cual se estima la posición de un evento sonoro en el espacio 3D. Se emplean micrófonos espaciales apropiados, por ejemplo, una matriz de micrófonos planos o 3D. En la Fig. 5, se ilustra un primer micrófono espacial 510, por ejemplo, una primera matriz de micrófonos 3D y un segundo micrófono espacial 520, por ej. , una primera matriz de micrófonos 3D. La DOA del espacio 3D se puede expresar, por ejemplo, en términos de azimut y elevación. Se pueden emplear vectores unitarios 530, 540 para expresar las DOAs. Se proyectan dos líneas 550, 560 de acuerdo con las DOAs. En 3D, incluso con cálculos estimativos muy confiables, las dos líneas 550, 560 proyectadas de acuerdo con las DOAs podrían no intersectarse. Sin embargo, de todas maneras se puede llevar a cabo la triangulación, por ejemplo, eligiendo el punto medio del segmento más pequeño que conecta las dos líneas.

45 [0057] De manera similar al caso 2D, la triangulación puede fallar o puede producir resultados no factibles en el caso de ciertas combinaciones de direcciones, que también pueden ser indicadas con banderas, por ej. con el módulo de cómputo de información 202 de la Fig. 3.

50 [0058] Si existen más de dos micrófonos espaciales, hay varias soluciones posibles. Por ejemplo, se podría llevar a cabo la triangulación antes explicada sobre todos los pares de micrófonos espaciales reales (si $N = 3$, 1 con 2, 1 con 3, y 2 con 3). A continuación se pueden promediar las posiciones así obtenidas (a lo largo de x e y , en caso de considerarse 3D, z).

55 [0059] Por otro lado, se pueden utilizar conceptos más complejos. Por ejemplo, se pueden aplicar los enfoques probabilísticos descritos en [15] J. Michael Steele, "Optimal Triangulation of Random Samples in the Plane", The Annals of Probability, Vol. 10, No.3 (Ago., 1982), pág. 548–553.

60 [0060] De acuerdo con una forma de realización, el campo sonoro puede ser analizado en el dominio del tiempo–frecuencia, por ejemplo, obtenido mediante una transformada de Fourier de corto tiempo (STFT), en la cual k y n denotan el índice de frecuencia k y el índice de tiempo n , respectivamente. La presión compleja $P_v(k, n)$ en una posición arbitraria p_v correspondiente a un k y n determinados se modela en forma de onda esférica única emitida por una fuente puntual isotrópica de banda angosta, por ej. empleando la fórmula:

65

$$P_v(k, n) = P_{IPLS}(k, n) \cdot \gamma(k, p_{IPLS}(k, n), p_v), \quad (1)$$

en la cual $P_{IPLS}(k, n)$ es la señal emitida por la IPLS en su posición $p_{IPLS}(k, n)$. El factor complejo $\gamma(k, p_{IPLS}, p_v)$ expresa la propagación de $p_{IPLS}(k, n)$ a p_v , por ej., introduce modificaciones apropiadas de fase y magnitud. En este caso, se puede aplicar la presunción de que, en cada caja de tiempo–frecuencia, sólo una IPLS está activa. Pese a esto, también pueden estar activas múltiples IPLSs de banda angosta en diferentes posiciones en un solo instante de tiempo.

[0061] Cada IPLS modela el sonido directo, o bien una reflexión diferenciada del recinto. Su posición $p_{IPLS}(k, n)$ puede corresponder idealmente a una fuente de sonido real situada dentro del recinto o a una fuente de sonido de imagen espejada situada en el exterior, respectivamente. Por lo tanto, la posición $p_{IPLS}(k, n)$ puede indicar asimismo la posición de un evento sonoro.

[0062] Sírvase notar que el término “fuentes sonoras reales” denota las fuentes de sonido real físicamente existentes en el ambiente de grabación, tales como voces humanas o instrumentos musicales. Por el contrario, con la expresión “fuentes de sonido” o “eventos sonoros” o “IPLS” nos referimos a fuentes sonoras efectivas, que son activas en determinados instantes de tiempo o en ciertas cajas de tiempo–frecuencia, donde las fuentes del sonidos pueden representar, por ejemplo, fuentes sonoras reales o fuentes de imágenes espejadas.

[0063] Las Fig. 15a–15b ilustran matrices de micrófonos que localizan fuentes sonoras. Las fuentes sonoras localizadas pueden tener diferentes interpretaciones físicas según su naturaleza. Cuando las matrices de micrófonos reciben sonido directo, pueden ser aptas para localizar la posición de una verdadera fuente de sonido (por ej. voces humanas). Cuando las matrices de micrófonos reciben reflexiones, pueden localizar la posición de una fuente de imágenes espejadas. Las fuentes de imágenes espejadas también son fuentes sonoras.

[0064] La Fig. 15a ilustra una situación en que dos matrices de micrófonos 151 y 152 reciben sonido directo de una fuente sonora real (una fuente sonora de existencia física) 153.

[0065] La Fig. 15b ilustra una situación en la cual dos matrices de micrófonos 161, 162 reciben sonido reflejado, donde el sonido ha sido reflejado por una pared. Debido a la reflexión, las matrices de micrófonos 161, 162 localizan la posición de la cual parece que el sonido proviene, en una posición de una fuente de imágenes espejadas 165, que es diferente de la posición de la persona que habla 163.

[0066] Tanto la fuente sonora real 153 de la Fig. 15a, como la fuente de imágenes espejadas 165 son fuentes sonoras.

[0067] La Fig. 15c ilustra una situación en la cual dos matrices de micrófonos 171, 172 reciben sonido difuso y no pueden localizar una fuente de sonido.

[0068] Si bien este modelo de onda única es preciso sólo en ambientes con poca reverberación, dado que las señales de la fuente cumplen la condición de ortogonalidad W disjunta (WDO), es decir que el traslape de tiempo–frecuencia es suficientemente pequeño. Esto es así normalmente en el caso de las señales de voz; véase, por ejemplo,

[12] S. Rickard and Z. Yilmaz, "On the approximate W -disjoint orthogonality of speech," in *Acoustics, Speech and Signal Processing*, 2002. ICASSP 2002. IEEE International Conference on, abril de 2002, vol. 1.

[0069] Sin embargo, el modelo también ofrece un cálculo estimativo satisfactorio con respecto a otros ambientes y, por lo tanto, también es aplicable a esos ambientes.

[0070] A continuación se explica la estimación de las posiciones $p_{IPLS}(k, n)$ de acuerdo con una forma de realización. La posición $p_{IPLS}(k, n)$ de una IPLS activa de una cierta caja de tiempo–frecuencia, y por consiguiente la estimación de un evento sonoro en una caja de tiempo–frecuencia, se realiza por medio de la triangulación sobre la base de la dirección de llegada (DOA) del sonido medida en por lo menos dos puntos de observación diferentes.

[0071] La Fig. 6 ilustra una geometría en la cual la IPLS de la ranura de tiempo–frecuencia (k, n) actual está situada en una posición desconocida $p_{IPLS}(k, n)$. Para determinar la información de DOA requerida, se emplean dos micrófonos espaciales reales, en este caso dos matrices de micrófonos, con una geometría, posición y orientación conocidas, que se colocan en las posiciones 610 y 620, respectivamente. Los vectores p_1 y p_2 apuntan a las posiciones 610, 620, respectivamente. Las orientaciones de la matriz están definidas por los vectores unitarios c_1 y c_2 . La DOA del sonido se determina en las posiciones 610 y 620 correspondientes a cada (k, n) empleando un algoritmo de cálculo de DOA, por ejemplo el provisto por el análisis de DirAC (ver [2], [3]). Al hacerlo de esta manera, se puede producir un vector unitario de un primer punto de vista $e_1^{POV}(k, n)$ y un vector unitario del segundo punto de vista $e_2^{POV}(k, n)$ con respecto a un punto de vista de las matrices de micrófonos (ninguna de las cuales está ilustrada en la Fig. 6) como salida del análisis de DirAC. Por ejemplo, cuando se opera en 2D, el primer vector unitario de punto de vista da como resultado:

$$\mathbf{e}_1^{\text{POV}}(k, n) = \begin{bmatrix} \cos(\varphi_1(k, n)) \\ \sin(\varphi_1(k, n)) \end{bmatrix}, \quad (2)$$

5 **[0072]** En este caso, $\varphi_1(k, n)$ representa el azimut de la DOA estimada en la primera matriz de micrófonos ilustrada en la Fig. 6. Los correspondientes vectores unitarios de DOA $\mathbf{e}_1(k, n)$ y $\mathbf{e}_2(k, n)$, con respecto al sistema global de coordenadas en el origen, pueden ser computados aplicando las fórmulas:

$$\begin{aligned} \mathbf{e}_1(k, n) &= \mathbf{R}_1 \cdot \mathbf{e}_1^{\text{POV}}(k, n), \\ \mathbf{e}_2(k, n) &= \mathbf{R}_2 \cdot \mathbf{e}_2^{\text{POV}}(k, n), \end{aligned} \quad (3)$$

donde \mathbf{R} son matrices de transformación de coordenadas, por ej.,

$$\mathbf{R}_1 = \begin{bmatrix} c_{1,x} & -c_{1,y} \\ c_{1,y} & c_{1,x} \end{bmatrix}, \quad (4)$$

20 al operar en 2D y $\mathbf{c}_1 = [c_{1,x}, c_{1,y}]^T$. Para llevar a cabo la triangulación, los vectores de dirección $\mathbf{d}_1(k, n)$ y $\mathbf{d}_2(k, n)$ pueden ser calculados de la siguiente manera:

$$\begin{aligned} \mathbf{d}_1(k, n) &= d_1(k, n) \mathbf{e}_1(k, n), \\ \mathbf{d}_2(k, n) &= d_2(k, n) \mathbf{e}_2(k, n), \end{aligned} \quad (5)$$

25 donde $d_1(k, n) = \|\mathbf{d}_1(k, n)\|$ y $d_2(k, n) = \|\mathbf{d}_2(k, n)\|$ son las distancias desconocidas entre las IPLS y las dos matrices de micrófonos. La siguiente ecuación

$$\mathbf{p}_1 + \mathbf{d}_1(k, n) = \mathbf{p}_2 + \mathbf{d}_2(k, n) \quad (6)$$

puede ser resuelta con respecto a $\mathbf{d}_1(k, n)$. Por último, la posición $\mathbf{p}_{\text{IPLS}}(k, n)$ de la IPLS está dada por

$$\mathbf{p}_{\text{IPLS}}(k, n) = d_1(k, n) \mathbf{e}_1(k, n) + \mathbf{p}_1. \quad (7)$$

40 **[0073]** En otra forma de realización, se puede resolver la ecuación (6) correspondiente a $\mathbf{d}_2(k, n)$ y $\mathbf{p}_{\text{IPLS}}(k, n)$ se computa de manera análoga empleando $\mathbf{d}_2(k, n)$.

45 **[0074]** La ecuación (6) siempre ofrece una solución cuando se opera en 2D, a menos que $\mathbf{e}_1(k, n)$ y $\mathbf{e}_2(k, n)$ son paralelos. Sin embargo, cuando se utilizan más de dos matrices de micrófonos o cuando se opera en 3D, no se puede obtener una solución cuando los vectores de dirección \mathbf{d} no se intersectan. De acuerdo con una forma de realización, en este caso, se debe computar el punto más cercano a todos los vectores de dirección \mathbf{d} y se puede utilizar el resultado como posición de la IPLS.

50 **[0075]** En una forma de realización, todos los puntos de observación $\mathbf{p}_1, \mathbf{p}_2, \dots$ deben estar situados de tal manera que el sonido emitido por la IPLS caiga dentro del mismo bloque temporal n . Este requisito se puede cumplir simplemente cuando la distancia Δ entre dos cualesquiera de los puntos de observación es menor que

$$\Delta_{\max} = c \frac{n_{\text{FFT}}(1 - R)}{f_s},$$

(8)

5 donde n_{FFT} es la longitud de ventana de STFT, $0 \leq R < 1$ especifica el traslape entre los sucesivos cuadros de tiempo y f_s es la frecuencia de muestreo. Por ejemplo, en el caso de una STFT de 1024 puntos a 48 kHz con 50 % de traslape ($R = 0,5$), el espaciamiento máximo entre las matrices para cumplir el requisito mencionado es $\Delta = 3,65$ m.

10 **[0076]** A continuación se describe en forma más detallada un módulo de cómputo de información 202, por ej. un módulo de cómputo de la señal de micrófono virtual e información complementaria de acuerdo con una forma de realización.

15 **[0077]** La Fig. 7 ilustra una reseña esquemática de un módulo de cómputo de información 202 de acuerdo con una forma de realización. La unidad de cómputo de información comprende un compensador de la propagación 500, un combinador 510 y una unidad de ponderación espectral 520. El módulo de cómputo de información 202 recibe los estimativos de posición de la fuente sonora ssp estimados por un estimador de posición de eventos sonoros, una o más señales de entrada de audio es registrada por uno o más de los micrófonos espaciales reales, las posiciones posRealMic de uno o más de los micrófonos espaciales reales y la posición virtual posVmic del micrófono virtual. Produce como salida una señal de salida de audio os que representa una señal de audio del micrófono virtual.

20 **[0078]** La Fig. 8 ilustra un módulo de cómputo de información de acuerdo con otra forma de realización. El módulo de cómputo de información de la Fig. 8 comprende un compensador de la propagación 500, un combinador 510 y una unidad de ponderación espectral 520. El compensador de la propagación 500 comprende un módulo de cómputo de parámetros de propagación 501 y un módulo de compensación de la propagación 504. El combinador 510 comprende un módulo de cómputo de factores de combinación 502 y un módulo de combinación 505. La unidad de ponderación espectral 520 comprende una unidad una unidad de cómputo de ponderaciones espectrales 503, un módulo de aplicación de ponderación espectral 506 y un módulo de cómputo de información complementaria espacial 507.

30 **[0079]** Para computar la señal de audio del micrófono virtual, se alimenta la información geométrica, por ej. la posición y orientación de los micrófonos espaciales reales 121 ... 12N, la posición, orientación y características del micrófono espacial virtual 104, y los estimativos de posición de los eventos sonoros 205 al módulo de cómputo de información 202, en particular, al módulo de cómputo de parámetros de propagación 501 del compensador de la propagación 500, al módulo de cómputo de factores de combinación 502 del combinador 510 y a la unidad de cómputo de ponderaciones espectrales 503 de la unidad de ponderación espectral 520. El módulo de cómputo de parámetros de propagación 501, el módulo de cómputo de factores de combinación 502 y la unidad de cómputo de ponderaciones espectrales 503 computan los parámetros empleados en la modificación de las señales de audio 111 ... 11N en el módulo de compensación de la propagación 504, el módulo de combinación 505 y el módulo de aplicación de ponderación espectral 506.

40 **[0080]** En el módulo de cómputo de información 202, las señales de audio 111 ... 11N pueden ser modificadas en un principio para compensar los efectos producidos por las longitudes de propagación diferentes entre las posiciones de los eventos sonoros y los micrófonos espaciales reales. A continuación se pueden combinar las señales para mejorar, por ejemplo, la relación señal a ruido (SNR). Por último, se puede realizar la ponderación espectral de la señal obtenida para tomar en cuenta el patrón de captación direccional del micrófono virtual, como así también cualquier función de ganancia dependiente de la distancia. Más adelante se describen en forma más detallada estos tres pasos.

50 **[0081]** Se explica ahora la compensación de la propagación en forma más detallada. En la porción superior de la Fig. 9, se ilustran dos micrófonos espaciales reales (una primera matriz de micrófonos 910 y una segunda matriz de micrófonos 920), la posición de un evento sonoro localizado 930 correspondiente a la caja de tiempo–frecuencia (k, n) y la posición of the micrófono espacial virtual 940.

55 **[0082]** La porción inferior de la Fig. 9 ilustra un eje temporal. Se presume que un evento sonoro es emitido en el momento t_0 y luego se propaga a los micrófonos espaciales reales y virtuales. Los retardos de tiempo de llegada, como así también las amplitudes, cambian con la distancia, por lo que cuanto mayor sea la longitud de propagación, más débil es la amplitud y más prolongado el retardo de tiempo de llegada.

60 **[0083]** Las señales presentes en las dos matrices reales son comparables sólo si el retardo relativo Dt_{12} entre ellos es pequeño. De lo contrario, una de las dos señales debe ser realineada temporalmente para compensar el retardo relativo Dt_{12} , y posiblemente para ser escalada a fin de compensar las degradaciones diferentes.

65 **[0084]** La compensación del retardo entre la llegada al lugar del micrófono virtual y la llegada a las matrices de micrófonos reales (en uno de los micrófonos espaciales reales) cambia el retardo independientemente de la localización del evento sonoro, lo que hace que sea superflua en la mayoría de las aplicaciones.

[0085] Volviendo a la Fig. 8, el módulo de cómputo de parámetros de propagación 501 está adaptado para computar los retardos que se han de corregir por cada micrófono espacial real y por cada evento sonoro. Si resulta ventajoso, también computa los factores de ganancia a considerar para compensar las diferentes degradaciones de la amplitud.

[0086] El módulo de compensación de la propagación 504 está configurado para usar esta información para modificar las señales de audio de modo correspondiente. Si las señales se deben desplazar en una pequeña cantidad de tiempo (en comparación con la ventana temporal del banco de filtros), luego basta con una simple rotación de fases. Si los retardos son más grandes, son necesarias implementaciones más complicadas.

[0087] La salida del módulo de compensación de la propagación 504 está constituida por las señales de audio modificadas expresadas en el dominio del tiempo–frecuencia original.

[0088] A continuación se describe una estimación específica de la compensación de la propagación correspondiente a un micrófono virtual de acuerdo con una forma de realización, con referencia a la Fig. 6 que ilustra, entre otras cosas, la posición 610 de un primer micrófono espacial real y la posición 620 de un segundo micrófono espacial real.

[0089] En la forma de realización que se pasa a explicar ahora, se presume que se dispone de por lo menos una primera señal de entrada de audio grabada, por ej. una señal de presión de por lo menos uno de los micrófonos espaciales reales (por ej. las matrices de micrófonos), por ejemplo, la señal de presión de un primer micrófono espacial real. Nos referimos al micrófono considerado como micrófono de referencia, a su posición como posición de referencia p_{ref} y a su señal de presión como señal de presión de referencia $P_{ref}(k, n)$. Sin embargo, la compensación de la propagación puede no sólo llevarse a cabo con respecto a una sola señal de presión, sino también con respecto a las señales de presión de una pluralidad o todos los micrófonos espaciales reales.

[0090] La relación entre la señal de presión $P_{IPLS}(k, n)$ emitida por la IPLS y una señal de presión de referencia $P_{ref}(k, n)$ de un micrófono de referencia situado en p_{ref} puede estar expresada por la fórmula (9):

$$P_{ref}(k, n) = P_{IPLS}(k, n) \cdot \gamma(k, \mathbf{p}_{IPLS}, \mathbf{p}_{ref}), \quad (9)$$

[0091] En general, el factor complejo $\gamma(k, p_a, p_b)$ expresa la rotación de fases y la degradación de la amplitud introducida por la propagación de una onda esférica desde su origen en p_a a p_b . Sin embargo, las pruebas prácticas indicaron que considerando sólo la degradación de la amplitud en γ se llega a impresiones plausibles de la señal del micrófono virtual con significativamente menos distorsiones en comparación con la consideración asimismo de la rotación de fases.

[0092] La intensidad sonora que se puede medir además en un cierto punto del espacio depende fuertemente de la distancia r de la fuente del sonido, en la Fig 6 desde la posición p_{IPLS} de la fuente del sonido. En muchas situaciones, esta dependencia puede ser modelada con suficiente precisión utilizando principios físicos muy conocidos, por ejemplo, la degradación $1/r$ de la presión sonora en el campo lejano de una fuente puntual. Cuando se conoce la distancia de un micrófono de referencia, por ejemplo, el primer micrófono real, de la fuente del sonido, y cuando también se conoce la distancia del micrófono virtual de la fuente del sonido, luego se puede estimar la intensidad sonora en la posición del micrófono virtual de la señal y la intensidad del micrófono de referencia, por ej. el primer micrófono espacial real. Esto significa que se puede obtener la señal de salida del micrófono virtual aplicando las ganancias adecuadas a la señal de presión de referencia.

[0093] Suponiendo que el primer micrófono espacial real es el micrófono de referencia, luego $p_{ref} = p_1$. En la Fig. 6, el micrófono virtual está situado en p_v . Dado que en la Fig. 6 se conoce en detalle la geometría, se puede determinar fácilmente la distancia $d_1(k, n) = \|d_1(k, n)\|$ entre el micrófono de referencia (en la Fig. 6: el primer micrófono espacial real) y la IPLS, como así también la distancia $s(k, n) = \|s(k, n)\|$ entre el micrófono virtual y la IPLS, es decir

$$s(k, n) = \|s(k, n)\| = \|\mathbf{p}_1 + \mathbf{d}_1(k, n) - \mathbf{p}_v\|. \quad (10)$$

[0094] La presión del sonido $P_v(k, n)$ en la posición del micrófono virtual se computa combinando las fórmulas (1) y (9), par dar

$$P_v(k, n) = \frac{\gamma(k, \mathbf{p}_{IPLS}, \mathbf{p}_v)}{\gamma(k, \mathbf{p}_{IPLS}, \mathbf{p}_{ref})} P_{ref}(k, n). \quad (11)$$

[0095] Como se mencionara anteriormente, en algunas de las realizaciones, los factores γ pueden considerarse sólo la degradación de la amplitud debido a la propagación. Suponiendo, por ejemplo, que la presión sonora se reduce con $1/r$, luego

$$P_v(k, n) = \frac{d_1(k, n)}{s(k, n)} P_{\text{ref}}(k, n).$$

(12)

5
10
15
20

[0096] Cuando prima el modelo de la fórmula (1), por ej., cuando sólo está presente el sonido directo, luego la fórmula (12) puede reconstruir con precisión la información de magnitud. Sin embargo, en el caso de los campos sonoros difusos, por ej., en que no se cumplen las presunciones del modelo, el procedimiento presentado da una desreverberación implícita de la señal al alejar el micrófono virtual de las posiciones de las matrices de sensores. De hecho, como se explicara anteriormente, en los campos sonoros difusos, esperamos que la mayoría de las IPLS estén situadas cerca de las dos matrices de sensores. De esa manera, al alejar el micrófono virtual de estas posiciones, es probable que se incremente la distancia $s = \|s\|$ de la Fig. 6. Por lo tanto, se reduce la magnitud de la presión de referencia al aplicar una ponderación de acuerdo con la fórmula (11). De manera correspondiente, al acercar el micrófono virtual a una fuente sonora real, las cajas de tiempo–frecuencia que corresponden al sonido directo se amplifican de tal manera que la señal de audio general se percibe en forma menos difusa. Ajustando la regla de la fórmula (12), se puede controlar la amplificación del sonido directo y la supresión del sonido difuso a voluntad.

[0097] Efectuando la compensación de la propagación sobre la señal de audio de entrada grabada (por ej. la señal de presión) del primer micrófono espacial real, se obtiene una primera señal de audio modificada.

25 **[0098]** En ciertas realizaciones, se puede obtener una segunda señal de audio modificada llevando a cabo la compensación de la propagación en una segunda señal de entrada de audio grabada (segunda señal de presión) del segundo micrófono espacial real.

30 **[0099]** En otras realizaciones, se pueden obtener otras señales de audio llevando a cabo la compensación de la propagación en otras señales de entrada de audio grabadas (otras señales de presión) de otros micrófonos espaciales reales.

35 **[0100]** Se explica ahora, en forma más detallada, la combinación efectuada en los bloques 502 y 505 de la Fig. 8 de acuerdo con una forma de realización. Se presume que dos o más señales de audio de una pluralidad de micrófonos espaciales reales diferentes han sido modificadas para compensar los diferentes trayectos de propagación a fin de obtener dos o más señales de audio modificadas. Una vez que las señales de audio procedentes de los diferentes micrófonos espaciales reales han sido modificadas para compensar los diferentes trayectos de propagación, se las puede combinar para mejorar la calidad del audio. Al así hacerlo, por ejemplo, se puede incrementar la SNR o se puede reducir la reverberación.

40 **[0101]** Las soluciones posibles para la combinación comprenden:

45 - Promedio ponderado, por ej., considerando la SNR, o la distancia hasta el micrófono virtual, o la difusividad estimada por los micrófonos espaciales reales. Se pueden emplear soluciones tradicionales, por ejemplo la Combinación de Relación Máxima (MRC) o la Combinación de Ganancias Iguales (o

50 - La combinación lineal de algunas o todas las señales de audio modificadas para obtener una señal combinatoria. Las señales de audio modificadas pueden ser ponderadas en la combinación lineal para obtener la señal combinatoria o

- La selección, por ej., el uso de una sola señal, por ejemplo dependiendo de la SNR o la distancia o difusividad, por ejemplo, dependiente de la SNR o la distancia o la difusividad.

55 **[0102]** La tarea del módulo 502 consiste, en caso de ser aplicable, en computar los parámetros para la combinación, que se lleva a cabo en el módulo 505.

60 **[0103]** Se describe ahora la ponderación espectral de acuerdo con las realizaciones en forma más detallada. Para ello, se hace referencia a los bloques 503 y 506 de la Fig. 8. En este paso final, se pondera la señal de audio que resulta de la combinación o de la compensación de la propagación de las señales de audio de entrada en el dominio del tiempo–frecuencia de acuerdo con las características espaciales del micrófono espacial virtual según lo especificado por la entrada 104 y/o de acuerdo con la geometría reconstruida (dada en 205).

65 **[0104]** Por cada caja de tiempo–frecuencia la reconstrucción geométrica nos permite obtener fácilmente la DOA con respecto al micrófono virtual, como se ilustra en la Fig. 10. Más aun, la distancia entre el micrófono virtual y la posición del evento sonoro también puede ser fácilmente calculada.

[0105] A continuación se computa la ponderación correspondiente a la caja de tiempo–frecuencia considerando el tipo de micrófono virtual deseado.

5 **[0106]** En el caso de los micrófonos direccionales, las ponderaciones espectrales pueden ser computadas de acuerdo con un patrón de captación predefinido. Por ejemplo, de acuerdo con una forma de realización, un micrófono cardioide puede tener un patrón de captación definido por la función $g(\theta)$,

$$g(\theta) = 0,5 + 0,5 \cos(\theta),$$

10 donde θ es el ángulo entre la dirección de mirada del micrófono espacial virtual y la DOA del sonido desde el punto de vista del micrófono virtual.

15 **[0107]** Otra posibilidad es la de funciones de degradación artística (no física). En ciertas aplicaciones, puede ser ventajoso suprimir los eventos sonoros alejados del micrófono virtual en un factor superior al que caracteriza la propagación libre en el campo. Por esta razón, algunas realizaciones introducen una función de ponderación adicional que depende de la distancia entre el micrófono virtual y el evento sonoro. En una forma de realización, sólo se deben captar los eventos sonoros dentro de una determinada distancia (por ej. en metros) del micrófono virtual.

20 **[0108]** Con respecto a la directividad del micrófono virtual, se pueden aplicar patrones de directividad arbitrarios al micrófono virtual. Al hacerlo, se puede separar, por ejemplo, una fuente de una escena de sonido compleja.

[0109] Dado que se puede computar la DOA del sonido en la posición p_v del micrófono virtual, es decir

$$\varphi_v(k, n) = \arccos \left(\frac{\mathbf{s} \cdot \mathbf{c}_v}{\|\mathbf{s}\|} \right),$$

(13)

30 donde \mathbf{c}_v es un vector unitario que describe la orientación del micrófono virtual, se pueden obtener directividades arbitrarias para el micrófono virtual. Por ejemplo, suponiendo que $P_v(k, n)$ indica la señal combinatoria o la señal de audio modificada con compensación de la propagación, luego la fórmula:

$$\tilde{P}_v(k, n) = P_v(k, n) [1 + \cos(\varphi_v(k, n))]$$

(14)

calcula la salida de un micrófono virtual con directividad cardioide. Los patrones direccionales, que potencialmente pueden generarse de esta manera, dependen de la precisión de la estimación de posición.

40 **[0110]** En las realizaciones, se coloca uno o más micrófonos reales no espaciales, por ejemplo, un micrófono omnidireccional o un micrófono direccional tal como un cardioide, en la escena sonora además de los micrófonos espaciales reales para mejorar aun más la calidad del sonido de las señales del micrófono virtual 105 de la Figura 8. Estos micrófonos no se utilizan para coleccionar información geométrica alguna, sino que sólo para obtener una señal de audio más nítida. Estos micrófonos pueden estar ubicados más cerca de las fuentes del sonido que los micrófonos espaciales. En este caso, de acuerdo con una forma de realización, las señales de audio de los micrófonos reales no espaciales y sus posiciones son alimentadas al módulo de compensación de la propagación 504 de la Fig. 8 para su procesamiento, en lugar de las señales de audio de los micrófonos espaciales reales. A continuación se ejecuta la compensación de la propagación correspondiente a las unas o más señales de audio registradas de los micrófonos no espaciales con respecto a la posición de uno o más micrófonos no espaciales. De esta manera, se obtiene una forma de realización utilizando micrófonos no espaciales adicionales.

55 **[0111]** En otra forma de realización, se realiza el cómputo de la información complementaria espacial del micrófono virtual. Para computar la información complementaria espacial 106 del micrófono, el módulo de cómputo de información 202 de la Fig. 8 comprende un módulo de cómputo de información complementaria espacial 507, que está adaptado para recibir, como entrada, las posiciones de las fuentes del sonido 205 y la posición, orientación y características 104 del micrófono virtual. En ciertas realizaciones, de acuerdo con la información complementaria 106 que se debe computar, también se puede tomar en cuenta la señal de audio del micrófono virtual 105 como entrada al módulo de cómputo de información complementaria espacial 507.

60 **[0112]** La salida del módulo de cómputo de información complementaria espacial 507 es la información complementaria del micrófono virtual 106. Esta información complementaria puede ser, por ejemplo, la DOA o la difusividad del sonido correspondiente a cada caja de tiempo–frecuencia (k, n) desde el punto de vista del micrófono virtual. Otra información complementaria posible podría ser, por ejemplo, el vector de intensidad sonora $I_a(k, n)$ que

se habría medido en la posición del micrófono virtual. Se pasa a describir ahora cómo se pueden derivar estos parámetros.

5 **[0113]** De acuerdo con una forma de realización, se ejecuta la estimación de la DOA correspondiente al micrófono espacial virtual. El módulo de cómputo de información 120 está adaptado para estimar la dirección de llegada en el lugar del micrófono virtual como información complementaria espacial, sobre la base de un vector de posición del micrófono virtual y sobre la base de un vector de posición del evento sonoro, como se ilustra en la Fig. 11.

10 **[0114]** La Fig. 11 ilustra una manera posible de derivar la DOA del sonido desde el punto de vista del micrófono virtual. Se puede describir la posición del evento sonoro, provista por el bloque 205 de la Fig. 8, por cada caja de tiempo–frecuencia (k, n) con un vector de posición $\mathbf{r}(k, n)$, la posición vector del evento sonoro. De modo similar, la posición del micrófono virtual, provista como entrada 104 en la Fig. 8, puede ser descrita con un vector de posición $\mathbf{s}(k, n)$, la posición vector del micrófono virtual. La dirección de mirada del micrófono virtual puede ser descrita por un vector $\mathbf{v}(k, n)$. La DOA con respecto al micrófono virtual está dada en $a(k, n)$. Representa el ángulo entre \mathbf{v} y $\mathbf{h}(k, n)$. Se puede computar el trayecto de propagación del sonido $\mathbf{h}(k, n)$ empleando la fórmula:

$$\mathbf{h}(k, n) = \mathbf{s}(k, n) - \mathbf{r}(k, n).$$

20 **[0115]** Ahora se puede computar la DOA deseada $a(k, n)$ por cada (k, n), por ejemplo por medio de la definición del producto puntual de $\mathbf{h}(k, n)$ y $\mathbf{v}(k, n)$, es decir

$$a(k, n) = \arcs (\mathbf{h}(k, n) \cdot \mathbf{v}(k, n) / (\|\mathbf{h}(k, n)\| \|\mathbf{v}(k, n)\|)).$$

25 **[0116]** En otra forma de realización, el módulo de cómputo de información 120 puede estar adaptado para estimar la intensidad del sonido activo en el lugar del micrófono virtual como información complementaria espacial, sobre la base de un vector de posición del micrófono virtual y sobre la base de un vector de posición del evento sonoro como se ilustra en la Fig. 11.

30 **[0117]** De la DOA $a(k, n)$ antes definida podemos derivar la intensidad del sonido activo $\mathbf{Ia}(k, n)$ en la posición del micrófono virtual. Para ello, se presume que la señal de audio del micrófono virtual 105 de la Fig. 8 corresponda a la salida de un micrófono omnidireccional, por ej., suponemos que el micrófono virtual es un micrófono omnidireccional. Más aun, se supone que la dirección de mirada \mathbf{v} de la Fig. 11 es paralela al eje x del sistema de coordenadas. Dado que el vector de intensidad del sonido activo $\mathbf{Ia}(k, n)$ describe el flujo de energía neto a través de la posición del micrófono virtual, podemos computar $\mathbf{Ia}(k, n)$, por ej. de acuerdo con la fórmula:

$$\mathbf{Ia}(k, n) = - (1/2 \rho) |P_v(k, n)|^2 * [\cos a(k, n), \text{sen } a(k, n)]^T,$$

40 donde $[\]^T$ denota un vector transpuesto, ρ es la densidad del aire y $P_v(k, n)$ es la presión sonora medida por el micrófono espacial virtual, por ej., la salida 105 del bloque 506 de la Fig. 8.

[0118] Si se computa el vector de intensidad activa expresado en el sistema general de coordenadas, pero aún en la posición del micrófono virtual, se puede aplicar la siguiente fórmula:

$$\mathbf{Ia}(k, n) = (1/2 \rho) |P_v(k, n)|^2 \mathbf{h}(k, n) / \|\mathbf{h}(k, n)\|.$$

45 **[0119]** La difusividad del sonido expresa cuán difuso es el campo sonoro en una ranura de tiempo–frecuencia dada (ver, por ejemplo, [2]). La difusividad está expresada por un valor ψ , donde $0 \leq \psi \leq 1$. Una difusividad de 1 indica que la energía del campo sonoro total es completamente difusa. Esta información es importante, por ej. en la reproducción del sonido espacial. Tradicionalmente, la difusividad se computa en el punto específico del espacio en el cual se coloca una matriz de micrófonos.

50 **[0120]** De acuerdo con una forma de realización, la difusividad puede ser computada como parámetro adicional a la información complementaria generada con respecto al micrófono virtual (VM), que puede ser colocado a voluntad en una posición arbitraria de la escena de sonido. Al hacerlo, se puede considerar que un aparato que también calcula la difusividad aparte de la señal de audio en una posición virtual de un micrófono virtual es un DirAC frontal, ya que es posible producir un flujo de DirAC, es decir una señal de audio, dirección de llegada y difusividad, con respecto a un punto arbitrario de la escena sonora. El flujo de DirAC puede ser posteriormente procesado, almacenado, transmitido y reproducido en una configuración arbitraria de múltiples parlantes. En este caso, el oyente experimenta la escena sonora como si él o ella estuviera en la posición especificada por el micrófono virtual y estuviera mirando en la dirección determinada por su orientación.

55 **[0121]** La Fig. 12 ilustra un bloque de cómputo de la información de acuerdo con una forma de realización que comprende una unidad de cómputo de la difusividad 801 para computar la difusividad en el lugar del micrófono virtual. El bloque de cómputo de la información 202 está adaptado para recibir entradas 111 a 11N, que además de las entradas de la Fig. 3, incluyen también difusividad en los micrófonos espaciales reales. Digamos que $\psi^{(SM1)}$ a $\psi^{(SMN)}$ denotan estos valores. Estas entradas adicionales son alimentadas al módulo de cómputo de información 202.

La salida 103 de la unidad de cómputo de la difusividad 801 es el parámetro de difusividad computado en la posición del micrófono virtual.

5 [0122] Una unidad de cómputo de la difusividad 801 de acuerdo con una forma de realización es la ilustrada en la Fig. 13 que expone más detalles. De acuerdo con una forma de realización, se estima la energía del sonido directo y difuso en cada uno de los N micrófonos espaciales. A continuación, utilizando la información sobre las posiciones de la IPLS y la información sobre las posiciones de los micrófonos espaciales y virtuales, se obtienen N cálculos estimativos de estas energías en la posición del micrófono virtual. Por último, se pueden combinar los estimativos para mejorar la precisión de estimación y se puede computar fácilmente el parámetro de difusividad en el lugar del micrófono virtual.

10 [0123] Digamos que $E_{\text{dir}}^{(\text{SM}1)}$ a $E_{\text{dir}}^{(\text{SM}N)}$ y $E_{\text{dif}}^{(\text{SM}1)}$ a $E_{\text{dif}}^{(\text{SM}N)}$ denotan los estimativos de las energías del sonido directo y difuso correspondientes a los N micrófonos espaciales computados por la unidad de análisis de energía 810. Si P_i es la señal de presión compleja y ψ_i es la difusividad correspondiente al i -ésimo micrófono espacial, luego se pueden computar las energías, por ejemplo, de acuerdo con las fórmulas:

$$E_{\text{dir}}^{(\text{SM}i)} = (1 - \Psi_i) \cdot |P_i|^2$$

$$E_{\text{dif}}^{(\text{SM}i)} = \Psi_i \cdot |P_i|^2$$

20 [0124] La energía del sonido difuso debe ser igual en todas las posiciones; por lo tanto, se puede computar un estimativo de la intensidad sonora difusa $E_{\text{dif}}^{(\text{VM})}$ en el lugar del micrófono virtual simplemente promediando $E_{\text{dif}}^{(\text{SM}1)}$ a $E_{\text{dif}}^{(\text{SM}N)}$, por ej. en una unidad de combinación de difusividades 820, por ejemplo, de acuerdo con la fórmula:

$$E_{\text{dif}}^{(\text{VM})} = \frac{1}{N} \sum_{i=1}^N E_{\text{dif}}^{(\text{SM}i)}$$

25 [0125] Se podría llevar a cabo una combinación más eficaz de los estimativos $E_{\text{dir}}^{(\text{SM}1)}$ a $E_{\text{dir}}^{(\text{SM}N)}$ considerando la varianza de los estimadores, por ejemplo considerando la SNR.

30 [0126] La energía del sonido directo depende de la distancia hasta la fuente debido a la propagación. Por lo tanto, $E_{\text{dir}}^{(\text{SM}1)}$ a $E_{\text{dir}}^{(\text{SM}N)}$ pueden ser modificadas para tomar esto en cuenta. Esto se puede llevar a cabo, por ej., mediante una unidad de ajuste de la propagación del sonido directo 830. Por ejemplo, si se presume que la energía del campo de sonido directo se degrada en 1 sobre la distancia al cuadrado, luego se puede calcular el estimativo correspondiente al sonido directo en el lugar del micrófono virtual correspondiente al i -ésimo micrófono espacial de acuerdo con la fórmula:

$$E_{\text{dir},i}^{(\text{VM})} = \left(\frac{\text{distancia SM}i - \text{IPLS}}{\text{distancia VM} - \text{IPLS}} \right)^2 E_{\text{dir}}^{(\text{SM}i)}$$

40 [0127] De manera similar a la unidad de combinación de difusividades 820, se pueden combinar los estimativos de la intensidad sonora directa obtenida en diferentes micrófonos espaciales, por ej. por medio de una unidad de combinación de sonido directo 840. El resultado es $E_{\text{dir}}^{(\text{VM})}$, por ej., el estimativo correspondiente a la intensidad sonora directa en el lugar del micrófono virtual. La difusividad en el lugar del micrófono virtual $\psi^{(\text{VM})}$ puede ser computada, por ejemplo, por una subcalculadora de difusividad 850, por ej. de acuerdo con la fórmula:

$$\Psi^{(\text{VM})} = \frac{E_{\text{dif}}^{(\text{VM})}}{E_{\text{dif}}^{(\text{VM})} + E_{\text{dir}}^{(\text{VM})}}$$

[0128] Como ya se mencionara, en algunos casos, falla la estimación de posición de los eventos sonoros llevada a cabo por un estimador de posición de eventos sonoros, por ej., en el caso de una estimación incorrecta de la

dirección de llegada. La Fig. 14 ilustra esa situación. En estos casos, independientemente de los parámetros de difusividad estimados en los diferentes micrófonos espaciales y recibidos como entradas 111 a 11N, se puede ajustar a 1 la difusividad correspondiente al micrófono virtual 103 (es decir, completamente difuso), ya que no es posible una reproducción espacialmente coherente.

5 [0129] Además, se puede considerar la confiabilidad de las estimaciones de DOA en los N micrófonos espaciales. Esto se puede expresar, por ej., en términos de la varianza del estimador de DOA o SNR. Esa información puede ser tomada en cuenta por la subcalculadora de difusividad 850, de manera que la difusividad del VM 103 pueda ser artificialmente incrementada en caso de que los estimativos de DOA no sean confiables. De hecho, tampoco son
10 confiables, como consecuencia, los estimativos de posición 205.

[0130] Si bien se han descrito algunos aspectos en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del procedimiento correspondiente, en el cual un bloque o dispositivo corresponde a un paso del procedimiento o a una característica de un paso del procedimiento. De manera análoga,
15 los aspectos descritos en el contexto de un paso del procedimiento también representan una descripción de un bloque o ítem correspondiente o de una característica de un aparato correspondiente.

[0131] La señal descompuesta de la presente invención puede ser almacenada en un medio de almacenamiento digital o se la puede transmitir por un medio de transmisión tal como un medio de transmisión inalámbrico o un
20 medio de transmisión conectado por cables tal como la Internet

[0132] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo un disco blando, un DVD, un CD, una ROM, una PROM, una EPROM, una
25 EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema de computación programable de tal manera que se ejecute el procedimiento respectivo.

[0133] Algunas realizaciones comprenden un transportador de datos no transitorio que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema de computación programable de tal
30 manera que se ejecute uno de los procedimientos descritos en la presente.

[0134] En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto programa de computación con un código de programa, donde el código de programa cumple la función de ejecutar uno de los procedimientos al ejecutarse el programa de computación en una computadora. El código de programa
35 puede ser almacenado, por ejemplo, en un portador legible por una máquina.

[0135] Otras realizaciones comprenden el programa de computación para ejecutar uno de los procedimientos aquí descritos, almacenado en un portador legible por una máquina.
40

[0136] En otras palabras, una realización del procedimiento consiste, por lo tanto, en un programa de computación que consta de un código de programa para realizar uno de los procedimientos aquí descritos al ejecutarse el programa de computación en una computadora.
45

[0137] Una forma de realización adicional de los procedimientos consiste, por lo tanto, en un portador de datos (o medio de almacenamiento digital, o medio legible por computadora) que comprende, grabado en el mismo, el programa de computación para ejecutar uno de los procedimientos aquí descritos.
50

[0138] Otra realización del procedimiento es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa de computación para ejecutar uno de los procedimientos aquí descritos. El flujo de datos o la secuencia de señales pueden estar configurada, por ejemplo, para ser transferida a través de una conexión de comunicación de datos, por ejemplo por la Internet.
55

[0139] Otra de las realizaciones comprende un medio de procesamiento, por ejemplo una computadora, un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los procedimientos aquí descritos.
60

[0140] Una forma de realización adicional comprende una computadora en la que se ha instalado el programa de computación para ejecutar uno de los procedimientos aquí descritos.
65

[0141] En algunas formas de realización, se puede utilizar un dispositivo lógico programable (por ejemplo una matriz de puertas programables en el campo) para ejecutar algunas o todas las funcionalidades de los procedimientos aquí descritos. En algunas realizaciones, una matriz de puertas programables en el campo puede cooperar con un microprocesador para ejecutar uno de los procedimientos aquí descritos. Por lo general, los procedimientos son ejecutados preferentemente por cualquier aparato de hardware.
70

[0142] Las realizaciones precedentemente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles aquí descritos han de

ser evidentes para las personas con capacitación en la técnica. Por lo tanto, sólo es intención limitarse al alcance de las siguientes reivindicaciones de patente y no a los detalles específicos presentados a manera de descripción y explicación de las realizaciones aquí presentadas.

5

Literatura:

[0143]

- 10 [1] R. K. Furness, "Ambisonics – An overview," in AES 8th International Conference, Abril de 1990, pág. 181–189.
- [2] V. Pulkki, "Directional audio coding in spatial sound reproduction and stereo upmixing," en el Acta de la 28a Conferencia Internacional de AES, pág. 251–258, Piteå, Suecia, Junio 30 – Julio 2, 2006.
- 15 [3] V. Pulkki, "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc., vol. 55, no. 6, pág. 503–516, Junio 2007.
- [4] C. Faller: "Microphone Front–Ends for Spatial Audio Coders", en el Acta de la 125a Convención Internacional de AES, San Francisco, Oct. 2008.
- 20 [5] M. Kallinger, H. Ochsenfeld, G. Del Galdo, F. Küch, D. Mahne, R. Schultz–Amling, y O. Thiergart, "A spatial filtering approach for directional audio coding," in Audio Engineering Society Convention 126, Munich, Alemania, Mayo de 2009.
- 25 [6] R. Schultz–Amling, F. Küch, O. Thiergart y M. Kallinger, "Acoustical zooming based on a parametric sound field representation," in Audio Engineering Society Convention 128, Londres R.U., Mayo de 2010.
- [7] J. Herre, C. Falch, D. Mahne, G. Del Galdo, M. Kallinger y O. Thiergart, "Interactive teleconferencing combining spatial audio object coding and DirAC technology," in Audio Engineering Society Convention 128, Londres R.U., mayo de 2010.
- 30 [8] E. G. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography, Academic Press, 1999.
- [9] A. Kuntz y R. Rabenstein, "Limitations in the extrapolation of wave fields from circular measurements," in 15th European Signal Processing Conference (EUSIPCO 2007), 2007.
- [10] A. Walther y C. Faller, "Linear simulation of spaced matrices de micrófonos using b–format recordings," in Audio Engineering Society Convention 128, Londres R.U., mayo de 2010.
- 40 [11] US61/287,596: An apparatus and a method for Converting a First Parametric Spatial Audio Signal into a Second Parametric Spatial Audio Signal.
- [12] S. Rickard y Z. Yilmaz, "On the approximate W–disjoint orthogonality of speech," in Acoustics, Speech and Signal Processing, 2002. ICASSP 2002. IEEE International Conference on, abril de 2002, vol. 1.
- 45 [13] R. Roy, A. Paulraj y T. Kailath, "Direction–of–arrival estimation by subspace rotation methods – ESPRIT," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Stanford, CA, EE.UU., abril de 1986.
- 50 [14] R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pág. 276–280, 1986.
- [15] J. Michael Steele, "Optimal Triangulation of Random Samples in the Plane", The Annals of Probability, Vol. 10, No.3 (Aug., 1982), pág. 548–553.
- 55 [16] F. J. Fahy, Sound Intensity, Essex: Elsevier Science Publishers Ltd., 1989.
- [17] R. Schultz–Amling, F. Küch, M. Kallinger, G. Del Galdo, T. Ahonen y V. Pulkki, "Planar microphone array processing for the analysis and reproduction of spatial audio using directional audio coding," en Audio Engineering Society Convention 124, Amsterdam, Países Bajos, mayo de 2008.
- 60 [18] M. Kallinger, F. Küch, R. Schultz–Amling, G. Del Galdo, T. Ahonen y V. Pulkki, "Enhanced direction estimation using matrices de micrófonos for directional audio coding," in Hands–Free Speech Communication and microphone arrays, 2008. HSCMA 2008, mayo de 2008, pág. 45–48.
- 65

REIVINDICACIONES

1. Aparato para generar una señal de salida de audio para simular una grabación de la señal de salida de audio de un micrófono virtual en una posición virtual configurable en un entorno, que comprende:

5 un estimador de posición de acontecimientos sonoros (110) destinado a estimar una posición de un acontecimiento sonoro que indica una posición de un acontecimiento sonoro en el entorno, en el que el acontecimiento sonoro está activo en un determinado momento o en un determinado contenedor de tiempo-frecuencia, en el que el acontecimiento sonoro es una fuente sonora real o una fuente en imagen-espejo, en el que el estimador de posición de acontecimientos sonoros (110) está configurado para estimar la posición del acontecimiento sonoro que indica una posición de una fuente en imagen-espejo en el entorno cuando el acontecimiento sonoro es una fuente en imagen-espejo, y en el que el estimador de posición de acontecimientos sonoros (110) está adaptado para estimar la posición del acontecimiento sonoro a partir de una primera información de dirección suministrada por un primer micrófono espacial real situado en una posición del primer micrófono real en el entorno, y a partir de una segunda información de dirección suministrada por un segundo micrófono espacial real situado en una posición del segundo micrófono real en el entorno, donde el primer micrófono espacial real y el segundo micrófono espacial real son unos micrófonos espaciales que existen físicamente; y donde el primer micrófono espacial real y el segundo micrófono espacial real son unos aparatos destinados a la adquisición de sonido espacial capaz de determinar la dirección de llegada del sonido, y

10 un módulo de cálculo de informaciones (120) destinado a generar la señal de salida de audio a partir de una primera señal de entrada de audio grabada, a partir de la posición del primer micrófono real, a partir de la posición virtual del micrófono virtual, y a partir de la posición del acontecimiento sonoro, en el cual el primer micrófono espacial real está configurado para grabar la primera señal de entrada de audio grabada, o en el cual un tercer micrófono está configurado para grabar la primera señal de entrada de audio grabada,

15 en el cual el estimador de posición de acontecimientos sonoros (110) está adaptado para estimar la posición del acontecimiento sonoro a partir de una primera dirección de llegada de la onda sonora emitida por el acontecimiento sonoro en la posición del primer micrófono real como primera información de dirección y a partir de una segunda dirección de llegada de la onda sonora en la posición del segundo micrófono real como segunda información de dirección, y

20 en el cual el módulo de cálculo de informaciones (120) comprende un compensador de propagación (500), en el cual el compensador de propagación (500) está adaptado para generar una primera señal de audio modificada modificando la primera señal de entrada de audio grabada a partir de una primera disminución de amplitud entre el acontecimiento sonoro y el primer micrófono espacial real y a partir de una segunda disminución de amplitud entre el acontecimiento sonoro y el micrófono virtual, ajustando un valor de amplitud, un valor de magnitud o un valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio; o en el cual el compensador de propagación (500) está adaptado para generar una primera señal de audio modificada compensando un primer retardo entre una llegada de una onda sonora emitida por el acontecimiento sonoro en el primer micrófono espacial real y una llegada de la onda sonora al micrófono virtual ajustando un valor de amplitud,

25 un valor de magnitud o un valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio.

2. Aparato según la reivindicación 1, en el cual el módulo de cálculo de informaciones (120) comprende un módulo de cálculo de informaciones laterales espaciales (507) destinado a calcular las informaciones laterales espaciales, en el cual el módulo de cálculo de informaciones (120) está adaptado para estimar la dirección de llegada o una intensidad de sonido activo al micrófono virtual como informaciones laterales espaciales a partir de un vector de posición del micrófono virtual y a partir de un vector de posición del acontecimiento sonoro.

3. Aparato según la reivindicación 1, en el cual el compensador de propagación (500) está adaptado para generar la primera señal de audio modificada modificando la primera señal de entrada de audio grabada a partir de la primera disminución de amplitud entre el acontecimiento sonoro y el primer micrófono espacial real y a partir de la segunda disminución de amplitud entre el acontecimiento sonoro y el micrófono virtual ajustando el valor de amplitud, el valor de magnitud o el valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio, en el cual el compensador de propagación (500) está adaptado para generar la primera señal de audio modificada en un ámbito de tiempo-frecuencia a partir de la primera disminución de amplitud entre el acontecimiento sonoro y el primer micrófono espacial real y a partir de la segunda disminución de amplitud entre el acontecimiento sonoro y el micrófono virtual ajustando dicho valor de amplitud de la primera señal de entrada de audio grabada representada en un ámbito de tiempo-frecuencia.

4. Aparato según la reivindicación 1, en el cual el compensador de propagación (500) está adaptado para generar la primera señal de audio modificada compensando el primer retardo entre la llegada de una onda sonora emitida por el acontecimiento sonoro en el primer micrófono espacial real y la llegada de la onda sonora al micrófono virtual ajustando el valor de amplitud, el valor de magnitud o el valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio, en el cual el compensador de propagación (500) está adaptado para generar la primera señal de

audio modificada en el ámbito de tiempo-frecuencia compensando el primer retardo entre la llegada de la onda sonora emitida por el acontecimiento sonoro en el primer micrófono espacial real y la llegada de la onda sonora al micrófono virtual ajustando dicho valor de amplitud de la primera señal de entrada de audio grabada representada en un ámbito de tiempo-frecuencia.

5 **5.** Aparato según cualquiera de las reivindicaciones anteriores, en el cual el compensador de propagación (500) está adaptado para realizar la compensación de propagación generando un valor de magnitud modificada de la primera señal de audio modificada aplicando la fórmula:

$$P_v(k, n) = \frac{d_1(k, n)}{s(k, n)} P_{ref}(k, n)$$

10 en la que $d_1(k, n)$ es la distancia entre la posición del primer micrófono espacial real y la posición del acontecimiento sonoro, en la que $s(k, n)$ es la distancia entre la posición virtual del micrófono virtual y la posición del acontecimiento sonoro, en la que $P_{ref}(k, n)$ es un valor de magnitud de la primera señal de entrada de audio grabada representada en un ámbito de tiempo-frecuencia, y en la que $P_v(k, n)$ es el valor de magnitud modificada correspondiente a la señal del micrófono virtual, en la que k designa un índice de frecuencia y en el que n designa un índice de tiempo.

15 **6.** Dispositivo según cualquiera de las reivindicaciones anteriores, en el cual el módulo de cálculo de información (120) comprende por otro lado un combinador (510), en el cual el compensador de propagación (500) está adaptado por otro lado para modificar una segunda señal de entrada de audio grabada que se graba mediante el segundo micrófono espacial real compensando un segundo retardo o una segunda disminución de amplitud entre una llegada de la onda sonora emitida por el acontecimiento sonoro al segundo micrófono espacial real y una llegada de la onda sonora al micrófono virtual ajustando un valor de amplitud, un valor de magnitud o un valor de fase de la segunda señal de entrada de audio grabada, para obtener una segunda señal de audio modificada, y
20 en el cual el combinador (510) está adaptado para generar una señal de combinación combinando la primera señal de audio modificada y la segunda señal de audio modificada, para obtener la señal de salida de audio.

7. Aparato según la reivindicación 6, en el cual el compensador de propagación (500) está adaptado por otro lado para modificar una o varias otras señales de entrada de audio grabadas que se graban mediante uno o varios otros micrófonos espaciales reales compensando los retardos o las disminuciones de amplitud entre una llegada de la onda sonora al micrófono virtual y una llegada de la onda sonora emitida por el acontecimiento sonoro a cada uno de los otros micrófonos espaciales reales, en el cual el compensador de propagación (500) está adaptado para compensar cada uno de los retardos o disminuciones de amplitud ajustando un valor de amplitud, un valor de magnitud o un valor de fase de cada una de las otras señales de entrada de audio grabadas, para obtener una pluralidad de terceras señales de audio modificadas, y
30 en el cual el combinador (510) está adaptado para generar una señal de combinación combinando la primera señal de audio modificada y la segunda señal de audio modificada y la pluralidad de terceras señales de audio modificadas, para obtener la señal de salida de audio.

40 **8.** Dispositivo según cualquiera de las reivindicaciones 1 a 5, en el cual el módulo de cálculo de informaciones (120) comprende una unidad de ponderación espectral (520) destinada a generar una señal de audio ponderada modificando la primera señal de audio modificada en función de una dirección de llegada de la onda sonora en la posición virtual del micrófono virtual y en función de un vector de unidad que describe la orientación del micrófono virtual, para obtener la señal de salida de audio, en el cual la primera señal de audio modificada se modifica en un ámbito de tiempo-frecuencia.

9. Aparato según la reivindicación 6 o la 7, en el cual el módulo de cálculo de informaciones (120) comprende una unidad de ponderación espectral (520) destinada a generar una señal de audio ponderada modificando la señal de combinación en función de una dirección de llegada o de la onda sonora en la posición virtual del micrófono virtual y en función de un vector de unidad que describe la orientación del micrófono virtual, para obtener la señal de salida de audio, en el cual la señal de combinación se modifica en un ámbito de tiempo-frecuencia.

50 **10.** Aparato según la reivindicación 8 o la 9, en el cual la unidad de ponderación espectral (520) está adaptada para aplicar el factor de ponderación
55 $a + (1 - \alpha)\cos(\varphi_v(k, n))$, o el factor de ponderación

$$0.5 + 0.5 \cos(\varphi_v(k, n))$$

a la señal de audio ponderada,

en la que $\Pi_v(k, n)$ indica un ángulo que especifica una dirección de llegada de la onda sonora emitida por el acontecimiento sonoro en la posición virtual del micrófono virtual, en la que k designa un índice de frecuencia y en la que n designa un índice de tiempo.

5 11. Aparato según cualquiera de las reivindicaciones 1 a 6, en el cual el compensador de propagación (500) está adaptado por otro lado para generar una tercera señal de audio modificada modificando una tercera señal de entrada de audio grabada que se graba mediante un cuarto micrófono compensando un tercer retardo o una tercera
10 disminución de amplitud entre una llegada de la onda sonora emitida por el acontecimiento sonoro al cuarto micrófono y una llegada de la onda sonora al micrófono virtual ajustando un valor de amplitud, un valor de magnitud o un valor de fase del tercera señal de entrada de audio grabada, para obtener la señal de salida de audio.

12. Aparato según cualquiera de las reivindicaciones anteriores, en el cual el estimador de posición de acontecimientos sonoros (110) está adaptado para estimar una posición del acontecimiento sonoro en un entorno tridimensional.

13. Aparato según cualquiera de las reivindicaciones anteriores, en el cual el módulo de cálculo de informaciones (120) comprende por otro lado una unidad de cálculo de difusión (801) adaptada para estimar una energía de sonido difuso en el micrófono virtual o una energía de sonido directo en el micrófono virtual, en el cual la unidad de cálculo de difusión (801) está adaptado para estimar la energía de sonido difuso en el micrófono virtual a partir de las energías de sonido difuso al primer y al segundo micrófono espacial real.

14. Aparato según la reivindicación 13, en el cual la unidad de cálculo de difusión (801) está adaptada para estimar la energía de sonido difuso $E_{diff}^{(VM)}$ en el micrófono virtual aplicando la fórmula:

$$E_{diff}^{(VM)} = \frac{1}{N} \sum_{i=1}^N E_{diff}^{(SMi)}$$

25 en la que N es el número de una pluralidad de micrófonos espaciales reales que comprende el primer y el segundo micrófono espacial real, y en la que $E_{diff}^{(SMi)}$ es la energía del sonido difuso en el i-ésimo micrófono espacial real.

15. Aparato según la reivindicación 13 o la 14, en el cual la unidad de cálculo de difusión (801) está adaptado para estimar la energía de sonido directo aplicando la fórmula:

$$E_{dir,i}^{(VM)} = \left(\frac{\text{distance SMi} - \text{IPLS}}{\text{distance VM} - \text{IPLS}} \right)^2 E_{dir}^{(SMi)}$$

30 en la que la "distancia SMi - IPLS" es la distancia entre una posición de el i-ésimo micrófono espacial real y la posición del acontecimiento sonoro, en la que la "distancia VM - IPLS" es la distancia entre la posición virtual y la posición del acontecimiento sonoro, y en la que $E_{dir}^{(SMi)}$ es la energía directa en el i-ésimo micrófono espacial real.

35 16. Aparato según cualquiera de las reivindicaciones 13 a 15, en el cual la unidad de cálculo de difusión (801) está adaptada para estimar la difusión en el micrófono virtual estimando la energía de sonido difuso en el micrófono virtual y la energía de sonido directo en el micrófono virtual y aplicando la fórmula:

$$\Psi^{(VM)} = \frac{E_{diff}^{(VM)}}{E_{diff}^{(VM)} + E_{dir}^{(VM)}}$$

40 en la que $\Psi^{(VM)}$ indica la difusión en el micrófono virtual estimada, en la que $E_{diff}^{(VM)}$ indica la energía de sonido difuso estimada y en la que $E_{dir}^{(VM)}$ indica la energía de sonido directo estimada.

17. Procedimiento de generación de una señal de salida de audio para simular una grabación de la señal de salida de audio por un micrófono virtual en una posición virtual configurable en un entorno, que comprende:

5 estimar una posición del acontecimiento sonoro que indica una posición de un acontecimiento sonoro en el entorno, en el que el acontecimiento sonoro está activo en un determinado momento o en un determinado contenedor de tiempo-frecuencia, en el que el acontecimiento sonoro es una fuente sonora real o una fuente en imagen-espejo, en el que la etapa de estimación de la posición del acontecimiento sonoro comprende estimar la posición del acontecimiento sonoro indicando una posición de una fuente en imagen-espejo en el entorno cuando el acontecimiento sonoro es una fuente en imagen-espejo, y en el que la etapa de estimación de la posición del acontecimiento sonoro se basa en una primera información de dirección suministrada por un primer micrófono espacial real situado en una posición del primer micrófono real en el entorno, y basada en una segunda información de dirección suministrada por un segundo micrófono espacial real situado en una posición del segundo micrófono real en el entorno, donde el primer micrófono espacial real y el segundo micrófono espacial real son unos micrófonos espaciales que existen físicamente; y

10 donde el primer micrófono espacial real y el segundo micrófono espacial real son unos aparatos destinados a la adquisición de sonido espacial capaz de determinar la dirección de llegada del sonido, y generar la señal de salida de audio a partir de una primera señal de entrada de audio grabada a partir de la posición del primer micrófono real, a partir de la posición virtual del micrófono virtual, y a partir de la posición del acontecimiento sonoro, en el cual el primer micrófono espacial real está configurado para grabar la primera señal de entrada de audio grabada, o en el cual un tercer micrófono está configurado para grabar la primera señal de entrada de audio grabada, en el cual la estimación de la posición del acontecimiento sonoro se realiza a partir de una primera dirección de llegada de la onda sonora emitida por el acontecimiento sonoro en la posición del primer micrófono real como primera información de dirección y a partir de una segunda dirección de llegada de la onda sonora en la posición del segundo micrófono real como segunda información de dirección, en el cual la etapa de generación de la señal de audio de salida comprende generar una primera señal de audio modificada modificando la primera señal de entrada de audio grabada a partir de una primera disminución de amplitud entre el acontecimiento sonoro y el primer micrófono espacial real y a partir de una segunda disminución de amplitud entre el acontecimiento sonoro y el micrófono virtual ajustando un valor de amplitud, un valor de magnitud o un valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio; o

20 en el cual la etapa de generación de la señal de salida de audio comprende generar una primera señal de audio modificada compensando un primer retardo entre una llegada de una onda sonora emitida por el acontecimiento sonoro del primer micrófono espacial real y una llegada de la onda sonora al micrófono virtual ajustando un valor de amplitud, un valor de magnitud o un valor de fase de la primera señal de entrada de audio grabada, para obtener la señal de salida de audio.

25

30

35

18. Programa de ordenador para la realización del procedimiento según la reivindicación 17 cuando se ejecuta en un ordenador o un procesador de señal.

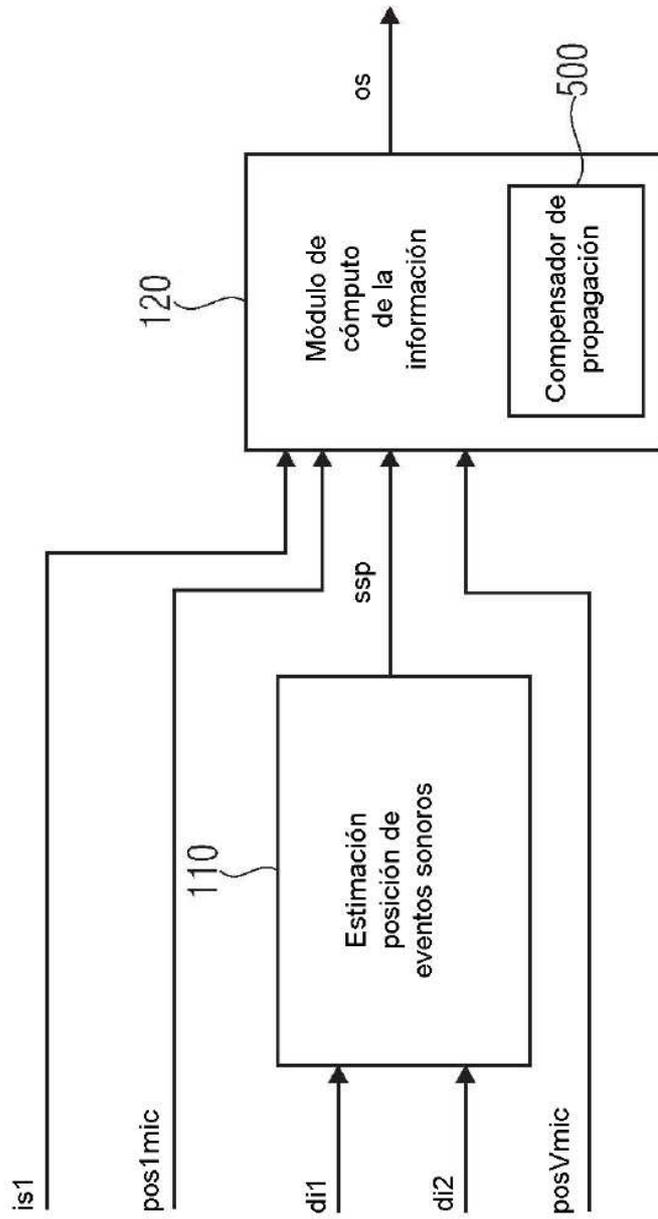


FIG 1

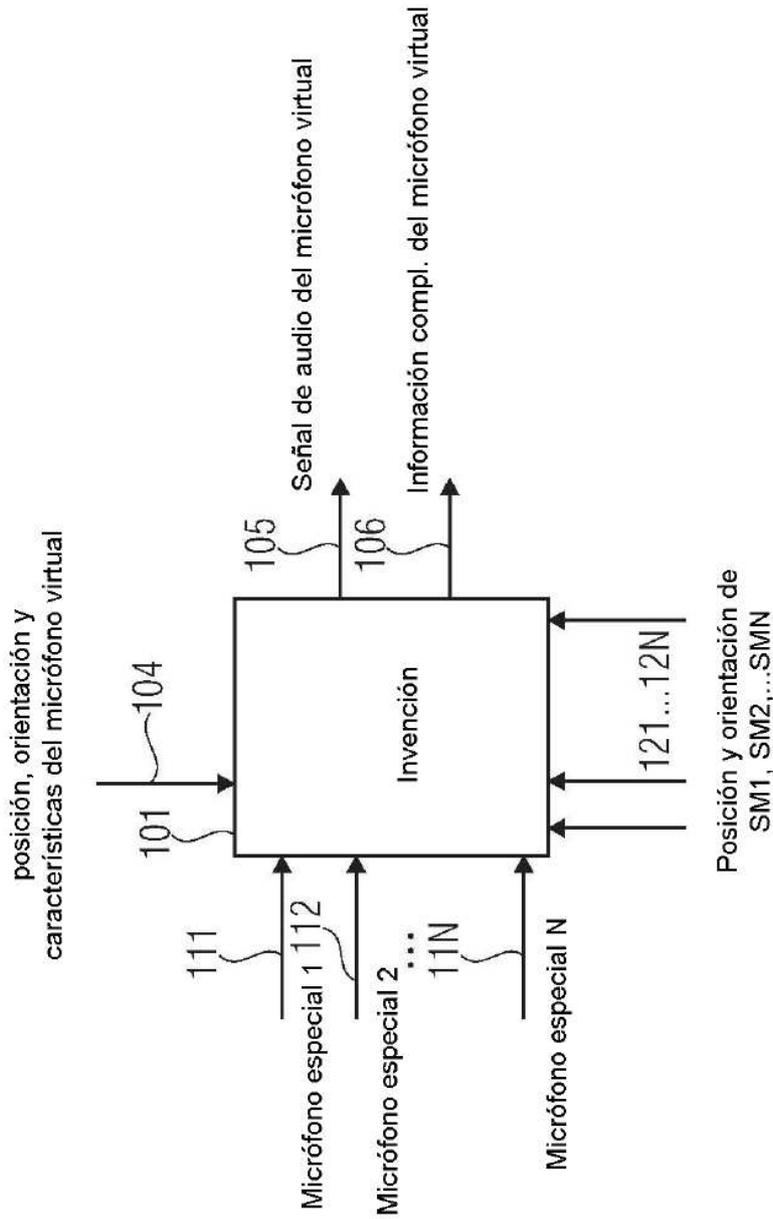


FIG 2

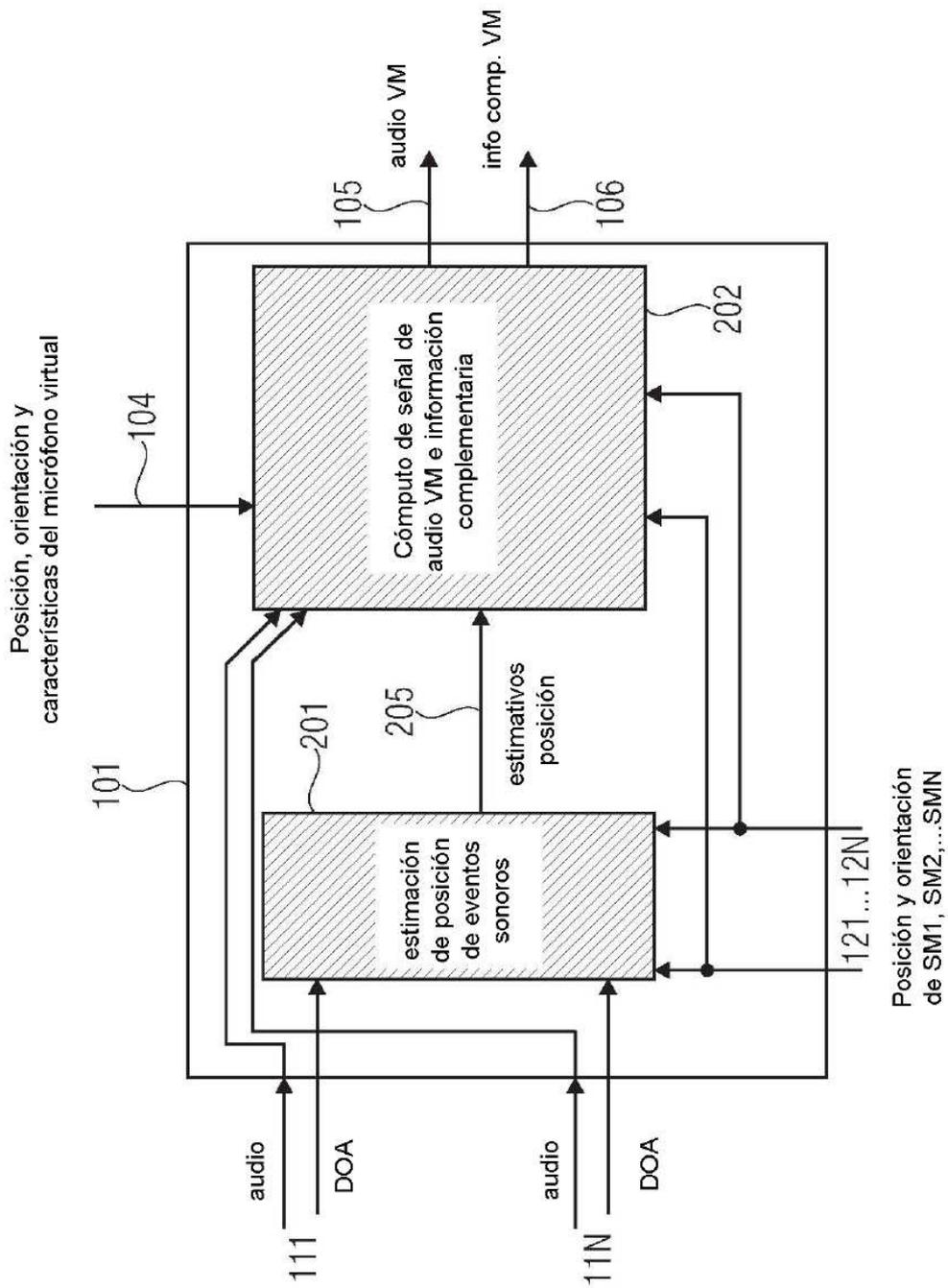


FIG 3

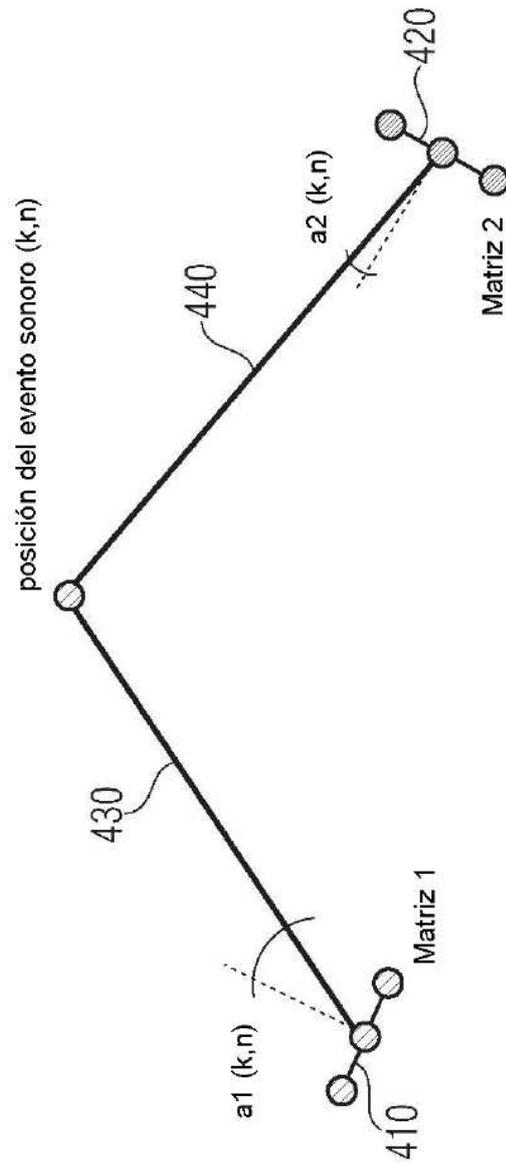


FIG 4

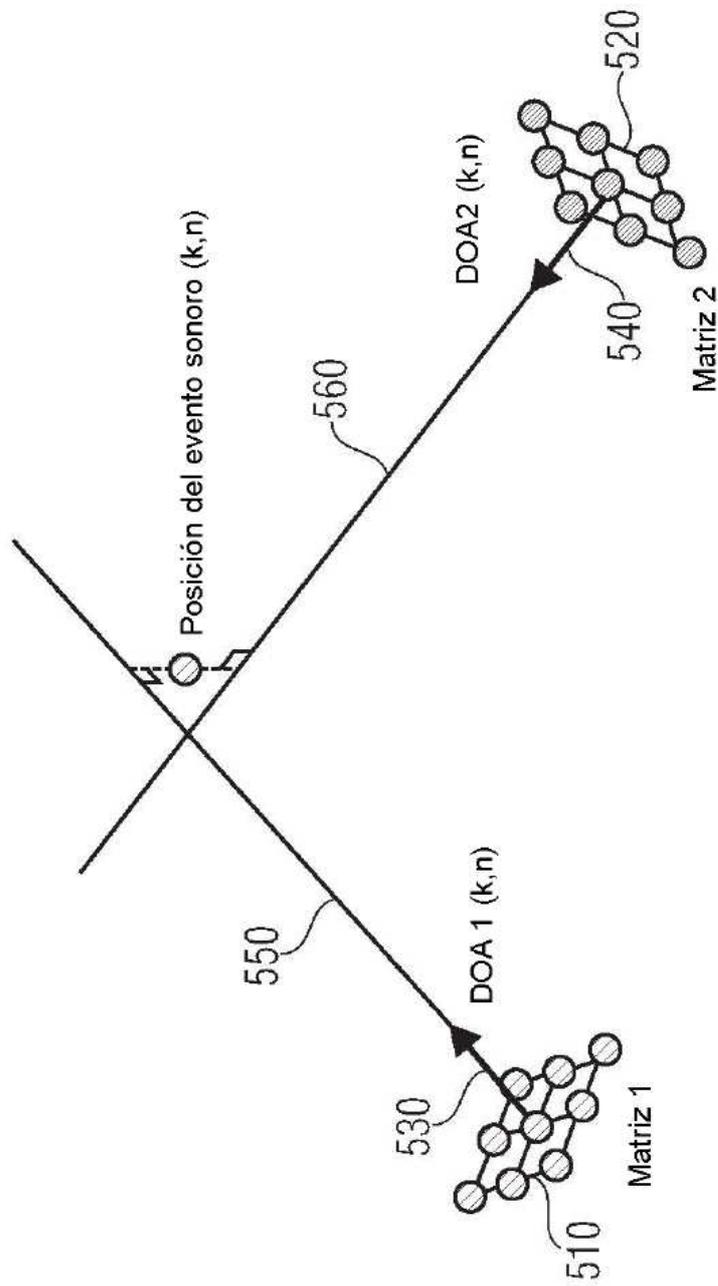


FIG 5

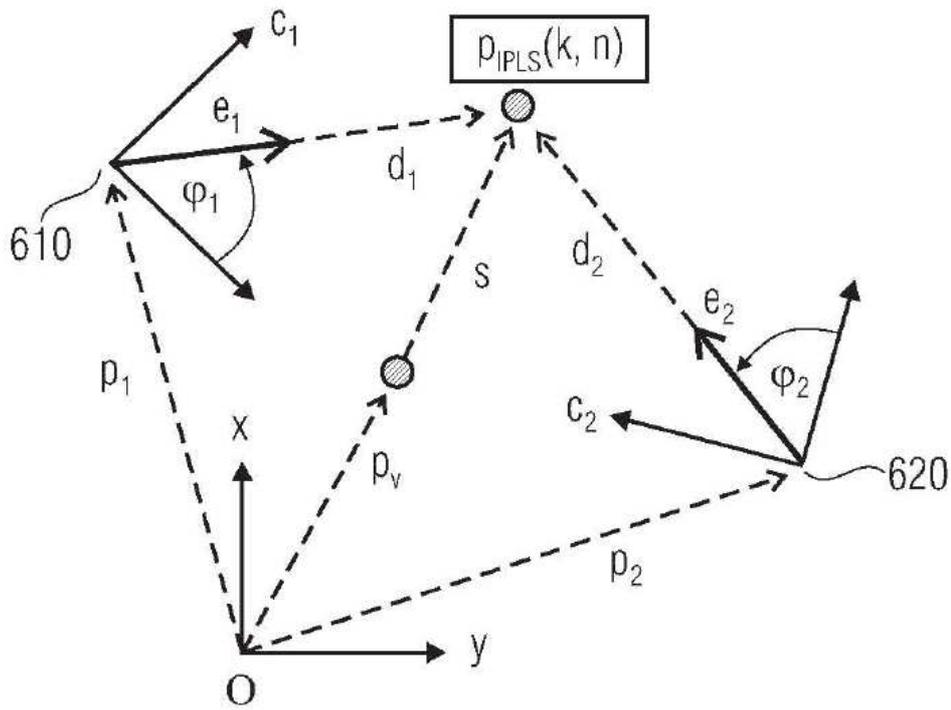


FIG 6

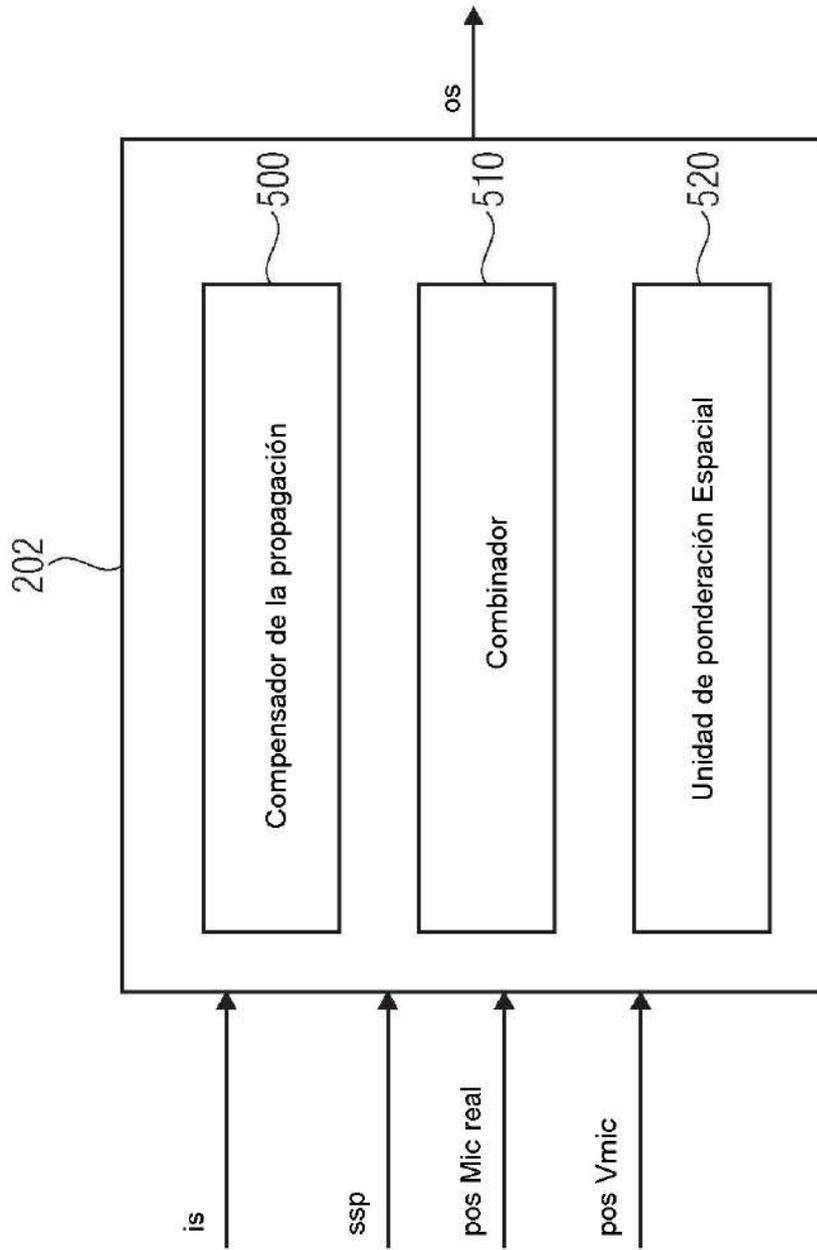


FIG 7

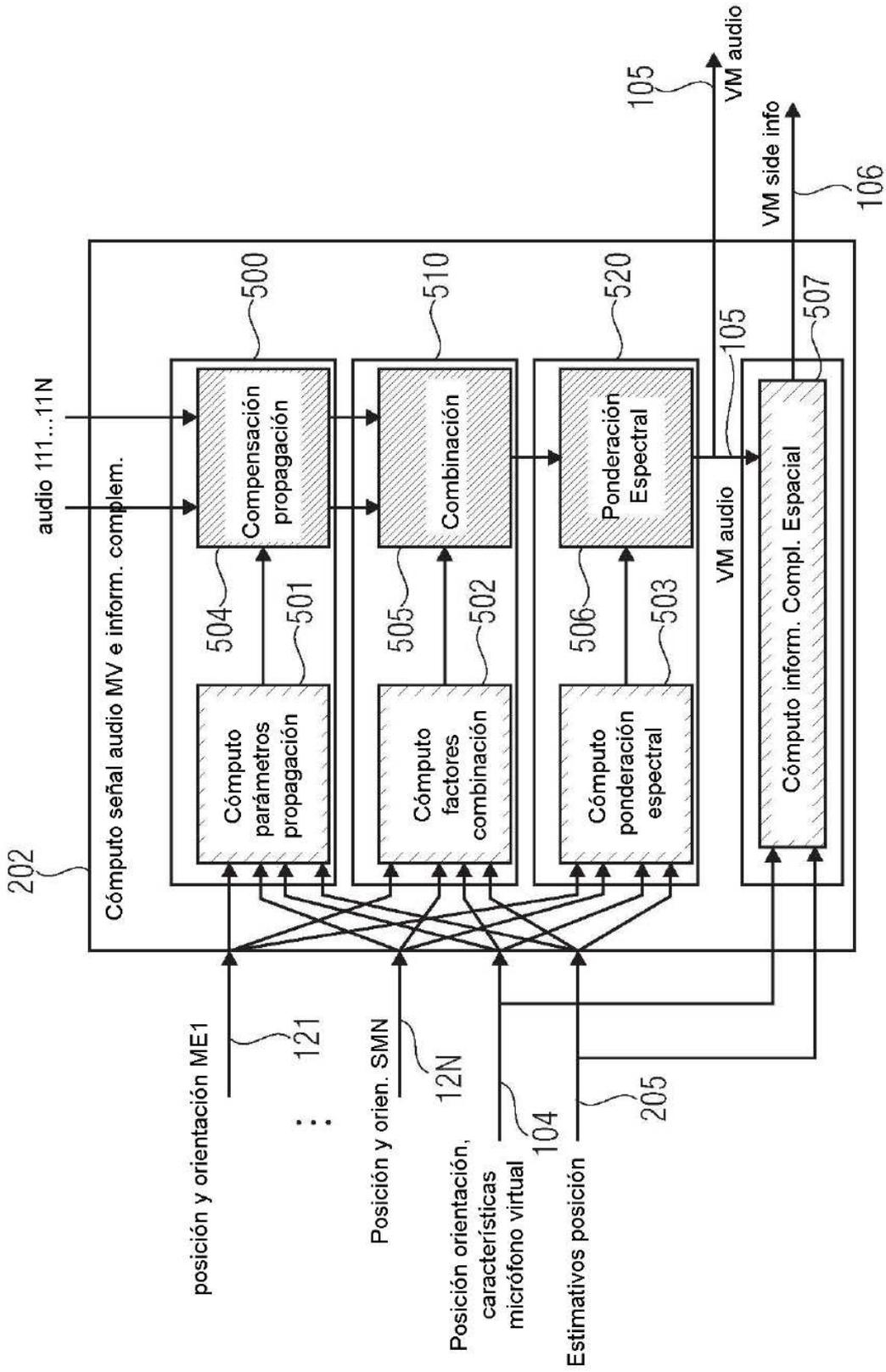


FIG 8

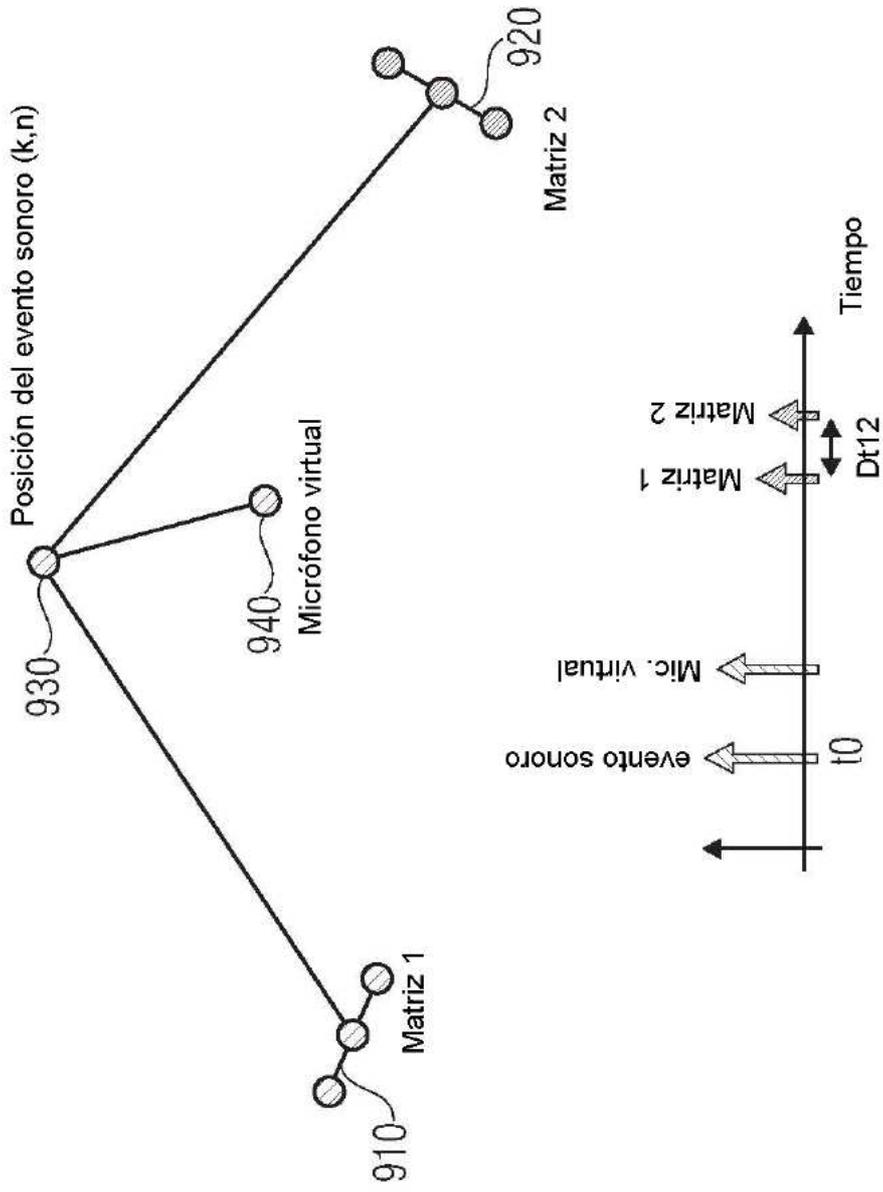


FIG 9

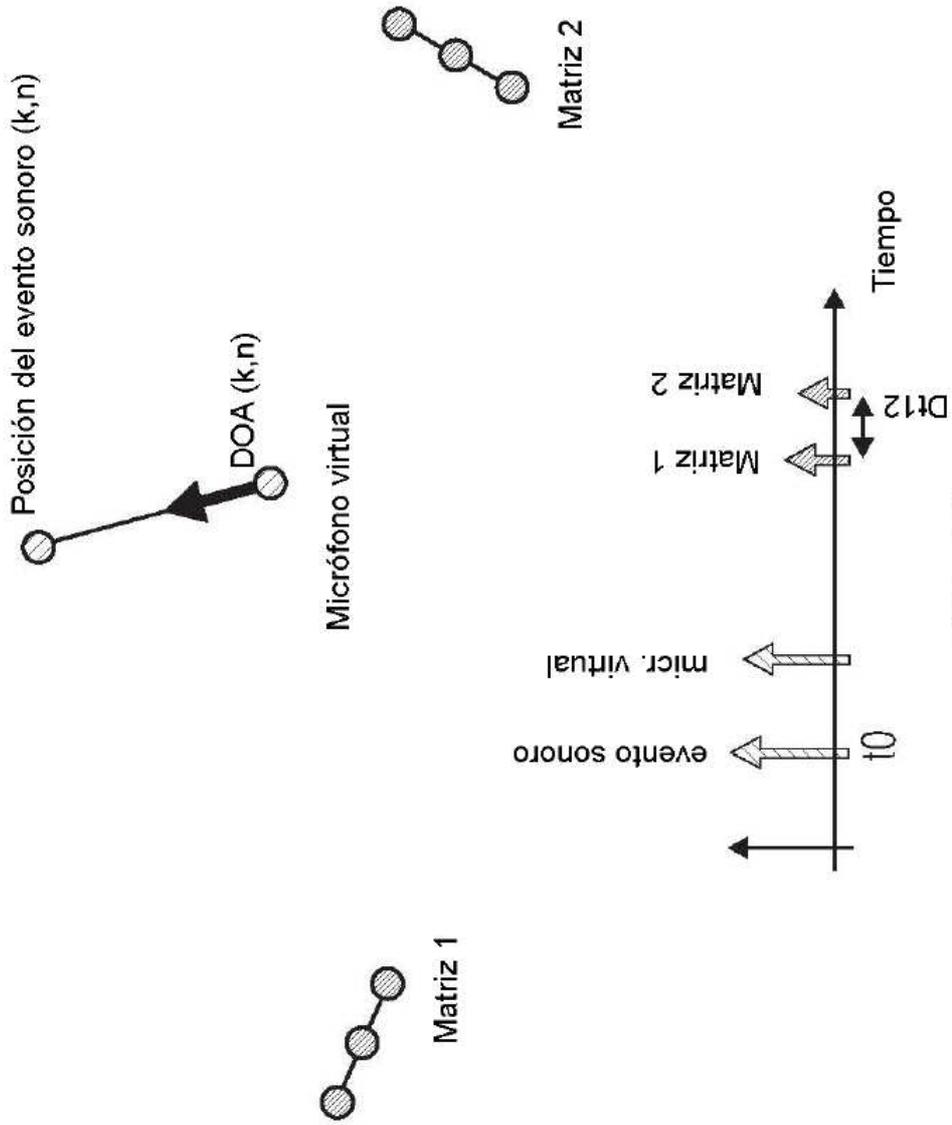


FIG 10

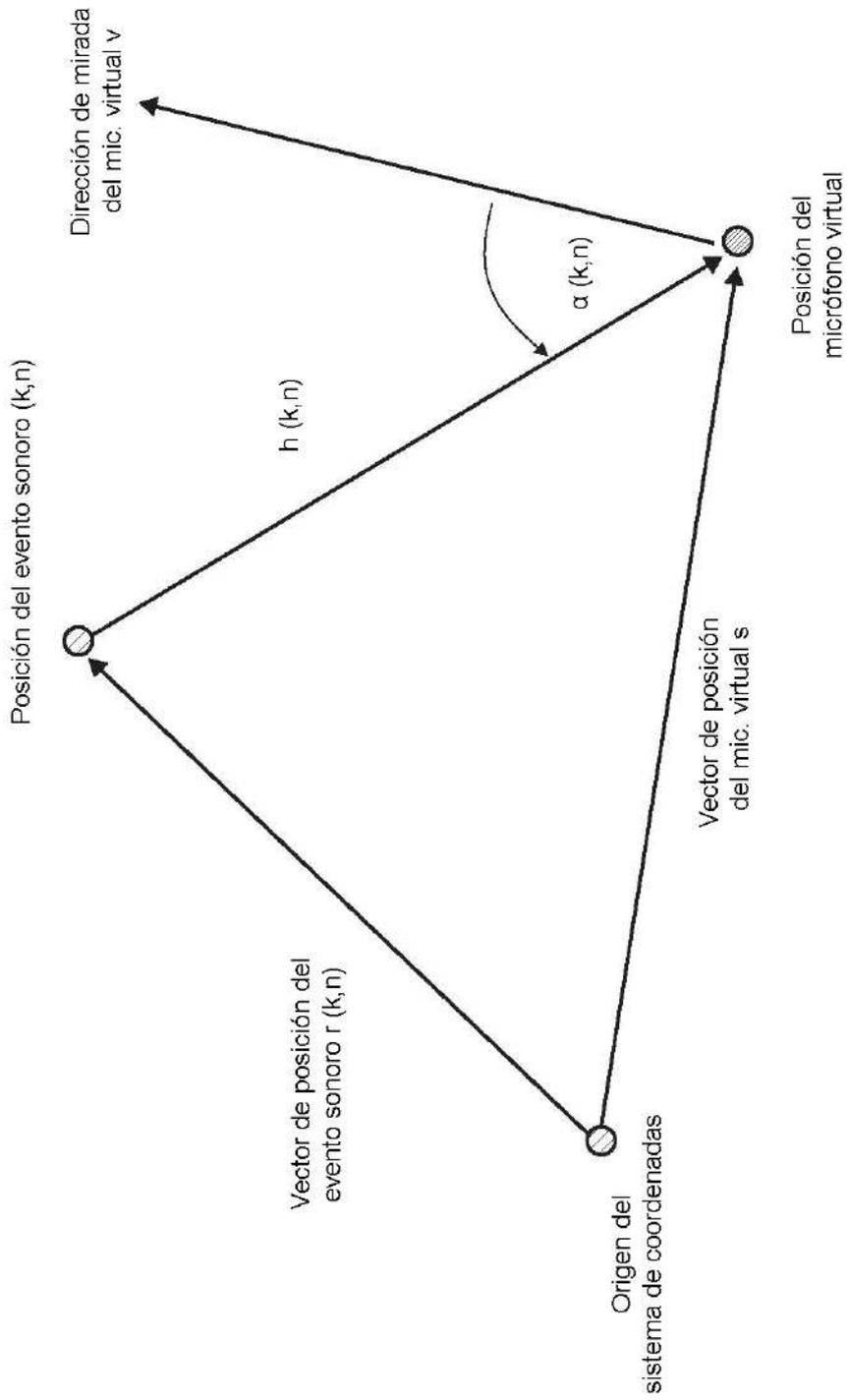


FIG 11

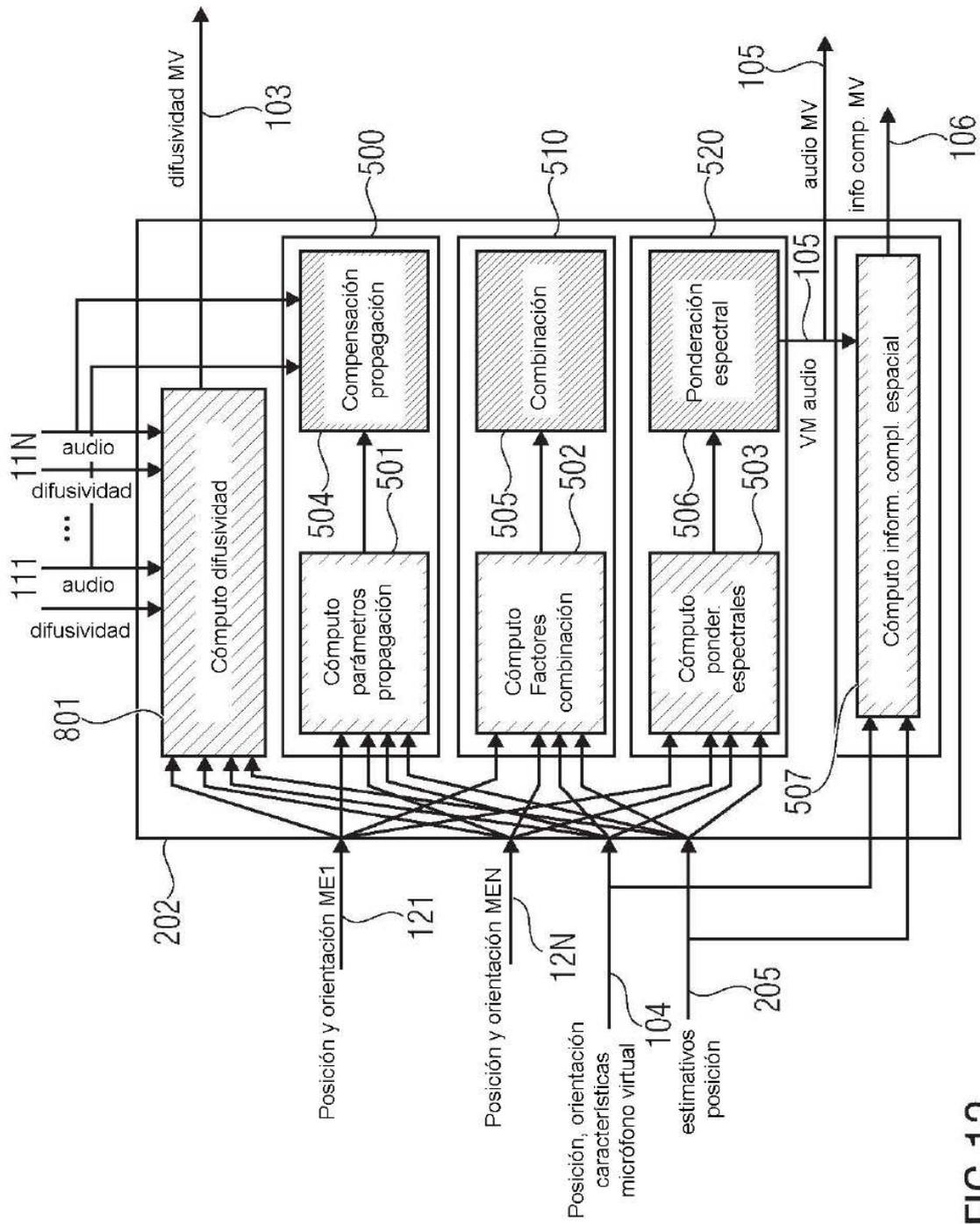


FIG 12

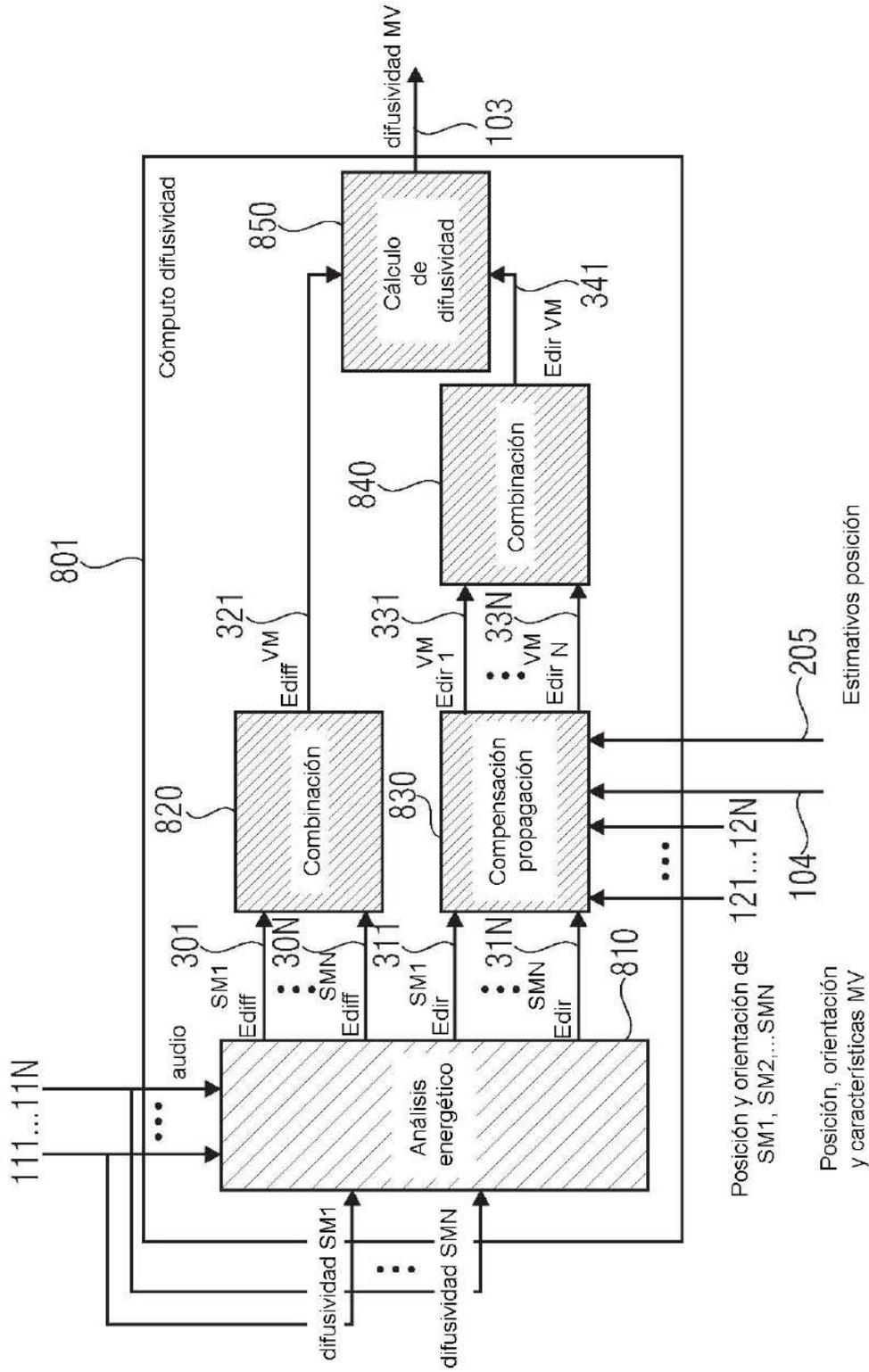


FIG 13

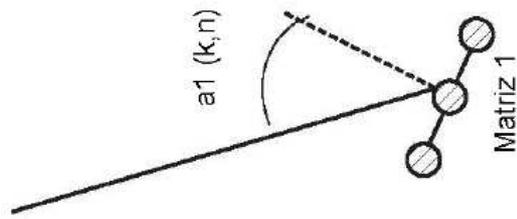
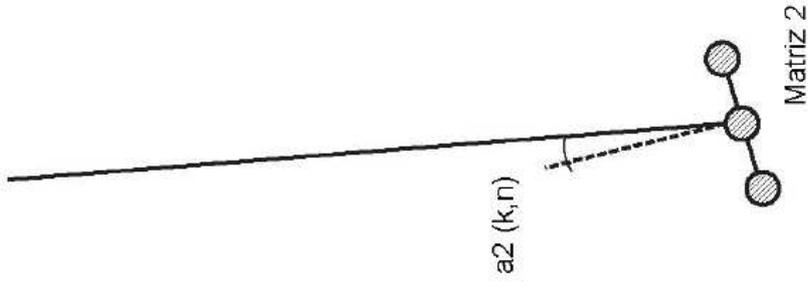


FIG 14

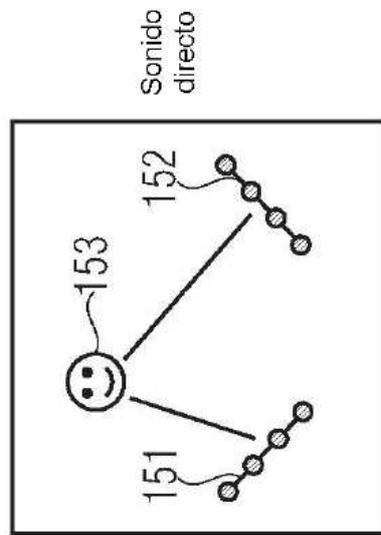


FIG 15A

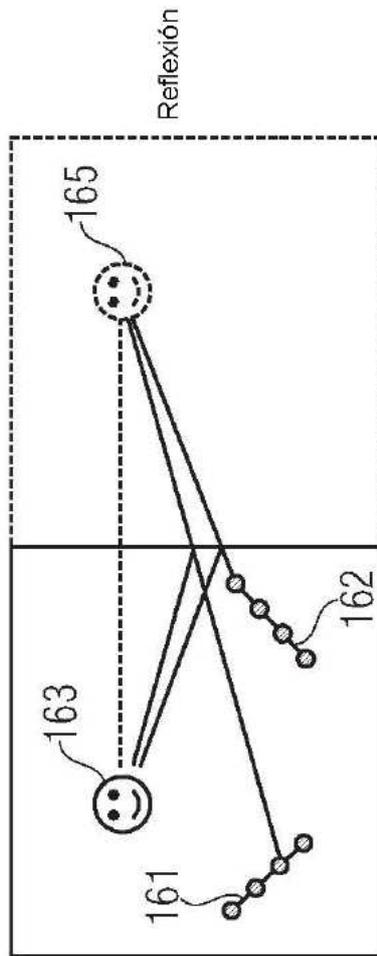


FIG 15B

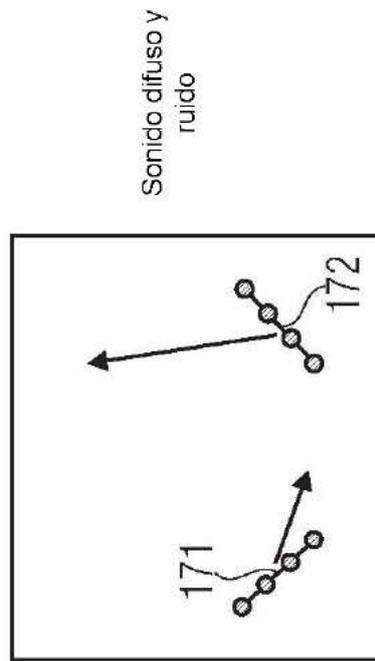


FIG 15C