

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 573 802**

51 Int. Cl.:

H04R 3/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **29.11.2012 E 12794942 (8)**

97 Fecha y número de publicación de la concesión europea: **27.04.2016 EP 2786593**

54 Título: **Aparato y método para colocar micrófonos basándose en una densidad de potencia espacial**

30 Prioridad:

**02.12.2011 EP 11191828
12.04.2012 US 201213445560**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.06.2016

73 Titular/es:

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

72 Inventor/es:

**DEL GALDO, GIOVANNI;
THIERGART, OLIVER;
KÜCH, FABIAN;
HABETS, EMANUEL y
CRACIUN, ALEXANDRA**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 573 802 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).

Aparato y método para colocar micrófonos basándose en una densidad de potencia espacial

DESCRIPCIÓN

5 La presente invención se refiere al procesamiento de señales de audio y, en particular, a un aparato y a un método para la colocación automática de micrófonos.

10 El procesamiento de señales de audio es cada vez más importante. En particular, el registro de sonidos espaciales se emplea en una pluralidad de aplicaciones. El registro de sonidos espaciales está destinado a capturar un campo de sonido con la ayuda de múltiples micrófonos de tal manera que, al momento de la reproducción, un oyente perciba la imagen sonora tal como se desarrolló en el lugar del registro.

15 Los enfoques convencionales para el registro de sonidos espaciales normalmente implican micrófonos omnidireccionales, espaciados (por ejemplo, estereofonía AB) micrófonos direccionales coincidentes (por ejemplo, en estereofonía de intensidad), o micrófonos más sofisticados, tales como un micrófono de formato B, por ejemplo, en Ambisonics, véase, por ejemplo,

20 [1] Michael A. Gerzon. Ambisonics in multichannel broadcasting and video. J. Audio Eng. Soc, 33(11): 859-871, 1985.

Un micrófono espacial, por ejemplo, los micrófonos direccionales, conjuntos de micrófonos, etc., es capaz de registrar el sonido espacial. El término "micrófono espacial" hace referencia a cualquier aparato para la adquisición direccionalmente selectiva del sonido espacial (por ejemplo micrófonos direccionales, conjuntos de micrófonos, etc.).

25 Para la reproducción de sonidos, los enfoques no paramétricos existentes obtienen señales de reproducción de audio deseadas directamente de las señales registradas del micrófono. Una ventaja principal de estos enfoques es que la imagen espacial registrada siempre es relativa al micrófono espacial usado.

30 En muchas aplicaciones, no es posible ni factible colocar un micrófono espacial en la posición deseada que, por ejemplo, puede ser una posición cercana a una o más fuentes de sonido. En este caso, sería beneficioso colocar múltiples micrófonos espaciales más alejados de las fuentes de sonido activas y aún así poder capturar la escena sonora tal como se desea.

35 Algunas aplicaciones emplean dos o más micrófonos espaciales reales. Debería observarse que el término "micrófono espacial real" se refiere al tipo de micrófono o combinación de micrófonos deseados (por ejemplo un micrófono direccional, un par de micrófonos direccionales tal como se usan en los micrófonos estéreo comunes, pero también un conjunto de micrófonos), que existe físicamente.

40 Para cada micrófono espacial real, la Dirección de Llegada (DOA, del inglés *Direction Of Arrival*) se puede estimar en el dominio tiempo-frecuencia. Usando la información recogida por los micrófonos espaciales reales, junto con el conocimiento de su posición relativa, puede ser posible calcular las señales de salida de un micrófono espacial virtualmente colocado en una posición arbitraria (a voluntad) en el ambiente. Este micrófono espacial se denomina como "micrófono espacial virtual" de aquí en adelante.

45 En dichas aplicaciones, la posición y la orientación del uno o más micrófonos virtuales necesitan introducirse manualmente. No obstante, se apreciaría si una posición y/u orientación óptima de dichos uno o más micrófonos virtuales se determinara automáticamente. El documento US 2009/0129609 proporciona una solución a este problema.

50 Sería ventajoso, si un aparato y método pudieran estar disponibles para determinar dónde colocar un micrófono virtual, dónde colocar un micrófono físico o determinar una posición de escucha óptima. Asimismo, sería ventajoso cómo colocar un micrófono en una orientación óptima. Los términos "colocación de micrófonos" e "información sobre la colocación" se refieren a cómo determinar una posición adecuada de un micrófono o un oyente, así como también cómo determinar una orientación adecuada de un micrófono o un oyente.

55 El objetivo de la presente invención es proporcionar conceptos mejorados para la colocación de micrófonos. El objetivo de la presente invención se consigue por medio de un aparato de acuerdo con la reivindicación 1, por medio de un método de acuerdo con la reivindicación 10, y por medio de un programa informático de acuerdo con la reivindicación 11.

60 Se proporciona un aparato para determinar las posiciones óptimas de escucha o del micrófono. El aparato comprende un determinador de densidad de potencia espacial y un estimador de información espacial. El determinador de densidad de potencia espacial se adapta para determinar una densidad de potencia espacial que indica valores de potencia para una pluralidad de ubicaciones en un ambiente basándose en información de fuentes

de sonido que indica uno o más valores de potencia y uno o más valores de posición de una o más fuentes de sonido ubicadas en el ambiente. El estimador de información espacial se adapta para estimar información espacial acústica basándose en la densidad de potencia espacial.

5 En lo sucesivo, la expresión “micrófono virtual” hará referencia, en general, a cualquier tipo de micrófono. En particular, el término “micrófono virtual” se refiere tanto a los micrófonos virtuales espaciales como a los micrófonos no espaciales, así como también a los micrófonos espaciales físicamente existentes como a los micrófonos no espaciales para los cuales ha de determinarse la información sobre la colocación.

10 El estimador de información espacial se adapta para determinar una posición óptima de micrófono virtual o una orientación óptima de micrófono virtual en un ambiente basándose en la densidad de potencia espacial determinada por el determinador de densidad de potencia espacial. La densidad de potencia espacial se determina por el determinador de densidad de potencia espacial basándose en los valores de potencia de las fuentes de sonido y la correspondiente información de posición.

15 Se proporciona una forma automática de determinar una posición y/u orientación óptima de uno o más micrófonos para describir la escena sonora, por ejemplo, uno o más micrófonos virtuales.

20 En algunas formas de realización, el determinador de densidad de potencia espacial se puede adaptar para hacer uso de la información opcional proporcionada por una métrica de importancia que, por ejemplo, representa una medida de la fiabilidad para la estimación de las posiciones de la ESS.

25 Por ejemplo, en algunas formas de realización, la Psi de difusión del sonido se puede usar como la métrica de importancia. El término (1-Psi) se puede entonces multiplicar simplemente a los valores de potencia de la fuente mientras se calcula la densidad de potencia espacial, de tal manera que el sonido difuso contribuirá menos que el sonido directo en la determinación de la densidad de potencia espacial.

30 Una ventaja importante de los conceptos propuestos es que se pueden aplicar independientemente de la condición de la sala y no requieren de ninguna información a priori con respecto al número o la posición de los interlocutores y/o las fuentes físicas de sonido. Por consiguiente, el sistema es independiente y se puede adaptar a cualquier clase de escenario usando solamente análisis de sonido. De acuerdo con la técnica anterior, la información a priori debe estar disponible para determinar una posición y/u orientación óptima de uno o más micrófonos. Esto limita la aplicación, o bien debe realizarse una estimación, limitando la precisión. Al emplear las formas de realización descritas anteriormente, esto no es necesario. La posición del micrófono virtual (o la pluralidad de micrófonos virtuales) se calcula al realizar un análisis de escena semiciego y a cambiarlo a continuación de acuerdo con los requisitos de la aplicación buscada.

35 A diferencia de otros métodos para estimar una posición y/u orientación óptima de los micrófonos virtuales, el método propuesto no requiere ninguna información sobre la escena geométrica considerada. Por ejemplo, no es necesaria una información a priori sobre el número de fuentes activas de sonido (por ejemplo, el número de participantes de una conferencia), ni ninguna información sobre las posiciones relativas de las fuentes activas de sonido (por ejemplo, la disposición de los participantes en una sala de conferencias). La información sobre el sonido se obtiene solamente de las propiedades de las fuentes activas de sonido, que se denominan “fuentes efectivas de sonido” (ESS, del inglés *Effective Sound Sources*), que describen la escena sonora. Las ESS modelan una escena sonora espacial en que una o más ESS están activas en cierto instante temporal o en cierto segmento de tiempo-frecuencia. En lo sucesivo, el término “fuente física” se usa para describir una fuente real de la escena sonora, por ejemplo, un interlocutor, mientras que el término fuente efectiva de sonido (ESS), (también denominada “fuente de sonido”), se usa para describir un evento sonoro que está activo en un único instante o segmento de tiempo-frecuencia. Cada ESS está caracterizada por una posición y por una potencia. Esta información permite construir una densidad de potencia espacial, por ejemplo una densidad de potencia espacial, que permite determinar la posición u orientación óptima del micrófono virtual.

40 Los parámetros de la ESS pueden obtenerse, por ejemplo, empleando los conceptos que se explican más adelante para el aparato para generar una señal de salida de audio de un micrófono virtual en una posición virtual configurable. La estimación de posición de los eventos sonoros se explica más adelante para el aparato para generar una señal de salida de audio de un micrófono virtual, y en particular se explica con referencia a las Figs. 15 - 17. Los conceptos allí descritos se pueden emplear para determinar la posición de una fuente efectiva de sonido. La compensación de propagación se explica más adelante para el aparato para generar una señal de salida de audio de un micrófono virtual, y en particular se explica con referencia a las Figs. 17 - 20. Los conceptos allí descritos se pueden emplear para determinar la potencia de una fuente efectiva de sonido.

60 De acuerdo con una forma de realización, el estimador de información espacial puede comprender un estimador de centro de escena sonora para estimar una posición de un centro de una escena sonora en el ambiente. El estimador de información espacial puede comprender, además, un calculador de la posición del micrófono para calcular una

posición de un micrófono como la información espacial acústica basándose en la posición del centro de la escena sonora.

5 En otra forma de realización, el calculador de la posición del micrófono se puede adaptar para calcular la posición del micrófono, en la que el micrófono es un micrófono virtual.

Asimismo, de acuerdo con otra forma de realización, el estimador de centro de escena sonora se puede adaptar para calcular un centro de gravedad de la densidad de potencia espacial para estimar el centro de la escena sonora.

10 En otra forma de realización, el estimador de centro de escena sonora se puede configurar para determinar un perfil de retardo de potencia basándose en la densidad de potencia espacial y para determinar un retardo de acuerdo con un valor cuadrático medio basándose en el perfil de retardo de potencia para cada una de una pluralidad de ubicaciones en el ambiente. El estimador de centro de escena sonora se puede configurar para determinar la posición de la ubicación de la pluralidad de ubicaciones como el centro de la escena sonora, que tiene el retardo
15 mínimo de acuerdo con un valor cuadrático medio de los retardos de acuerdo con un valor cuadrático medio de la pluralidad de ubicaciones.

20 En otra forma de realización, el estimador de centro de escena sonora se puede adaptar para realizar la integración en círculo para estimar el centro de la escena sonora, en la que el estimador de centro de escena sonora se puede adaptar para realizar la integración en círculo al convolucionar la densidad de potencia espacial con un círculo aplicando, por ejemplo, la fórmula

$$g(x, y) = \Gamma(x, y) * C_{(r,o)}(x, y)$$

25 en la que $\Gamma(x, y)$ es la densidad de potencia espacial, y en la que $C_{(r,o)}(x, y)$ indica un círculo, para determinar un valor de integración en círculo para cada una de la pluralidad de ubicaciones del ambiente, cuando el ambiente es un ambiente bidimensional.

30 Como alternativa, el estimador de centro de escena sonora se puede adaptar para realizar la integración en círculo al convolucionar la densidad de potencia espacial con una esfera aplicando, por ejemplo, la fórmula

$$g(x, y, z) = \Gamma(x, y, z) * C_{(r,o)}(x, y, z)$$

35 en la que, $\Gamma(x, y, z)$ es la densidad de potencia espacial, y en la que $C_{(r,o)}(x, y, z)$ indica una esfera, para determinar un valor de integración en círculo para cada una de la pluralidad de ubicaciones del ambiente, cuando el ambiente es un ambiente tridimensional.

40 Asimismo, de acuerdo con una forma de realización, el estimador de centro de escena sonora se puede adaptar para determinar un máximo de los valores de integración en círculo de cada una de la pluralidad de ubicaciones del ambiente para estimar el centro de la escena sonora.

45 En otra forma de realización, el calculador de la posición del micrófono se puede adaptar para determinar una línea de ancho más amplio de una pluralidad de líneas a través del centro de la escena sonora en el ambiente. Cada una de la pluralidad de líneas a través del centro de la escena sonora puede tener un ancho de energía, en la que la línea de ancho más amplio puede ser la línea de la pluralidad de líneas a través del centro de la escena sonora que tiene el ancho de energía más grande.

50 De acuerdo con una forma de realización, el ancho de energía de una línea considerada de la pluralidad de líneas puede indicar una longitud más grande de un segmento en la línea considerada, de tal manera que el primer punto del segmento que limita el segmento, y de tal manera que un segundo punto diferente del segmento que limita el segmento tengan, ambos, un valor de potencia indicado por la densidad de potencia espacial, que puede ser mayor o igual a un valor de potencia predefinido. El calculador de la posición del micrófono se puede adaptar para determinar la posición del micrófono de tal manera que una segunda línea, que atraviesa el centro de la escena sonora y la posición del micrófono puedan ser ortogonales a la línea de ancho más amplio.

55 En una forma de realización, el calculador de la posición del micrófono puede estar configurado para aplicar una descomposición en valores singulares a una matriz que tiene una pluralidad de columnas. Las columnas de la matriz pueden indicar las posiciones de las ubicaciones en el ambiente relativas al centro de la escena sonora. Asimismo, las columnas de la matriz pueden solamente indicar las posiciones de las ubicaciones que tienen los valores de potencia indicados por la densidad de potencia espacial que son superiores a un valor umbral predefinido, o las columnas de la matriz pueden solamente indicar las posiciones de las ubicaciones que tienen los valores de potencia indicados por la densidad de potencia espacial que son mayores o iguales a un valor umbral predefinido.

60 De acuerdo con otra forma de realización, el estimador de información espacial puede comprender un determinador

de orientación para determinar una orientación del micrófono basándose en la densidad de potencia espacial. El determinador de orientación se puede adaptar para determinar la orientación del micrófono de tal manera que el micrófono se oriente hacia el centro de la escena sonora. El determinador de orientación se puede configurar para determinar un valor de integración $f(\varphi)$ para cada una de una pluralidad de direcciones φ aplicando la fórmula

$$f(\varphi) = \int_0^{r_{\max}} \Gamma(r \cos(\varphi), r \sin(\varphi)) \cdot r \, dr,$$

5 en el que r_{\max} define una distancia máxima desde el micrófono, y en el que el determinador de orientación está configurado para determinar la orientación del micrófono basándose en los valores de integración determinados $f(\varphi)$.

10 En otra forma de realización, el determinador de densidad de potencia espacial se puede adaptar para determinar la densidad de potencia espacial para la pluralidad de ubicaciones del ambiente para un segmento de tiempo-frecuencia (k, n) aplicando la fórmula

$$\Gamma(x, y, k, n) = \sum_{i=1}^N \text{potencia}_i(k, n) \cdot g(\gamma_i, x - x_{\text{ESS}i}, y - y_{\text{ESS}i}, k, n),$$

15 cuando el ambiente es un ambiente bidimensional, o aplicando la fórmula

$$\Gamma(x, y, z, k, n) = \sum_{i=1}^N \text{potencia}_i(k, n) \cdot g(\gamma_i, x - x_{\text{ESS}i}, y - y_{\text{ESS}i}, z - z_{\text{ESS}i}, k, n),$$

20 cuando el ambiente es un ambiente tridimensional, en donde k indica el índice de frecuencia y n indica el índice de tiempo, en el que x, y, z indican coordenadas de una de la pluralidad de ubicaciones, en el que $\text{potencia}_i(k, n)$ indica el valor de potencia a una i -ésima fuente de sonido para el segmento de tiempo-frecuencia (k, n) , en el que $x_{\text{ESS}i}, y_{\text{ESS}i}, z_{\text{ESS}i}$ indican coordenadas de la fuente de sonido i -ésima, en el que γ_i es un valor escalar que puede representar un indicador de cómo de fiables son las estimaciones de la posición de cada fuente efectiva de sonido y en el que g es una función que depende de $x, y, z, x_{\text{ESS}i}, y_{\text{ESS}i}, z_{\text{ESS}i}, k, n$ y γ_i .

Las formas de realización de la presente invención se explican con referencia a los dibujos adjuntos, en los cuales:

- 30 Fig. 1 ilustra un aparato para la colocación de micrófonos de acuerdo con una forma de realización,
 Fig. 2 representa un aparato para la colocación de micrófonos de acuerdo con otra forma de realización.
 35 Fig. 3 ilustra las entradas y salidas de un aparato para la colocación de micrófonos de acuerdo con una forma de realización,
 Fig. 4a-4c muestran una pluralidad de escenarios de aplicación para un aparato para la colocación de micrófonos,
 40 Fig. 5 representa un determinador de densidad de potencia espacial 21 de acuerdo con una forma de realización,
 Fig. 6a ilustra funciones delta para construir la función g ,
 45 Fig. 6b representa funciones de densidad para construir la función g ,
 Fig. 7 ilustra un estimador de información espacial de acuerdo con una forma de realización,
 Fig. 8 muestra un estimador de información espacial de acuerdo con otra forma de realización,
 50 Fig. 9 ilustra un calculador de la posición/orientación del micrófono 44 de acuerdo con otra forma de realización que representa más detalles,
 Fig. 10a-10c representan optimización basándose en el ancho de energía proyectada de acuerdo con una forma de realización,
 55 Fig. 11 ilustra un estimador de información espacial de acuerdo con otra forma de realización, en el que el estimador de información espacial además comprende un determinador de orientación,
 60 Fig. 12 ilustra un aparato para generar una señal de salida de audio de acuerdo con una forma de realización,

- Fig. 13 ilustra las entradas y salidas de un aparato y un método para generar una señal de salida de audio de acuerdo con una forma de realización,
- 5 Fig. 14 ilustra la estructura básica de un aparato para generar una señal de salida de audio de acuerdo con una forma de realización que comprende un estimador de posición de eventos sonoros y un módulo de cálculo de información,
- 10 Fig. 15 muestra un escenario de ejemplo en el cual se representan micrófonos espaciales reales según los Conjuntos Lineales Uniformes de 3 micrófonos cada uno,
- Fig. 16 representa dos micrófonos espaciales en 3D para estimar la dirección de llegada en espacio 3D,
- 15 Fig. 17 ilustra una geometría donde una fuente de sonido del estilo de punto isotrópico del segmento de tiempo-frecuencia actual (k, n) se ubica en una posición $p_{IPLS}(k, n)$,
- Fig. 18 representa el módulo de cálculo de información de acuerdo con una forma de realización,
- Fig. 19 representa el módulo de cálculo de información de acuerdo con otra forma de realización,
- 20 Fig. 20 muestra dos micrófonos espaciales reales, un evento sonoro localizado y una posición de un micrófono virtual espacial,
- Fig. 21 ilustra cómo obtener la dirección de llegada relativa a un micrófono virtual de acuerdo con una forma de realización,
- 25 Fig. 22 representa una posible forma de obtener la DOA del sonido desde el punto de vista del micrófono virtual de acuerdo con una forma de realización,
- 30 Fig. 23 ilustra un bloque de cálculo de información que comprende una unidad de cálculo de difusión de acuerdo con una forma de realización,
- Fig. 24 representa una unidad de cálculo de difusión de acuerdo con una forma de realización,
- 35 Fig. 25 ilustra un escenario, donde la estimación de la posición de los eventos sonoros no es posible,
- Fig. 26 muestra un aparato para generar una señal de salida virtual
- 40 Fig. 27a-27c ilustran escenarios donde dos conjuntos de micrófonos reciben sonido directo, sonido reflejado por una pared y sonido difuso.

La Fig. 1 ilustra un aparato para la colocación de micrófonos de acuerdo con una forma de realización. El aparato comprende un determinador de densidad de potencia espacial 10 y un estimador de información espacial 20. El determinador de densidad de potencia espacial 10 se adapta para determinar una densidad de potencia espacial spd que indica valores de potencia para una pluralidad de ubicaciones en un ambiente basándose en información de fuentes de sonido ssi que indica uno o más valores de potencia y uno o más valores de posición de una o más fuentes efectivas de sonido (EES) que se encuentran en el ambiente. El estimador de información espacial 20 se adapta para estimar la información espacial acústica $aspi$ basándose en la densidad de potencia espacial.

50 La Fig. 2 ilustra un aparato para la colocación de micrófonos de acuerdo con otra forma de realización. El aparato comprende un determinador de densidad de potencia espacial 21 para determinar una densidad de potencia espacial (SPD), que también se conoce como distribución de potencia espacial, que indica valores de potencia para una pluralidad de ubicaciones de un ambiente basándose en la información de fuentes efectivas de sonido que indica uno o más valores centrales y valores de posición de una o más fuentes efectivas de sonido ubicadas en el ambiente. El aparato además comprende un estimador de información espacial 22 para estimar una posición y/u orientación de un micrófono virtual (VM) basándose en la densidad de potencia espacial.

La Fig. 3 ilustra las entradas y salidas de un aparato para la colocación de micrófonos de acuerdo con una forma de realización. Las entradas 91, 92, ... 9N al aparato comprenden la potencia, por ejemplo, el valor absoluto de la presión del campo de sonido al cuadrado y la posición, por ejemplo, coordenadas Cartesianas 2D o 3D. Las fuentes efectivas de sonido (ESS) están describiendo la escena sonora (campo de sonido).

Las fuentes efectivas de sonido pueden, por ejemplo, ser iguales a las fuentes de sonido puntuales instantáneas (IPLS *Instantaneous Point-Like Sound Sources*) tal como se describen más adelante para el aparato para generar

una señal de salida de audio de un micrófono virtual en una posición virtual configurable.

En la salida, se devuelve la posición y ubicación de uno o más micrófonos virtuales. En lo sucesivo, el término “fuente física” se usa para describir una fuente real de la escena sonora, por ejemplo, un interlocutor, mientras que el

5 término “fuente efectiva de sonido” (ESS), (también denominada “fuente de sonido”), se usa para describir un evento sonoro que está activo en un único momento o segmento de tiempo-frecuencia, como también se usa para la IPLS descrita a continuación con respecto al aparato para generar una señal de salida de audio de un micrófono virtual en una posición virtual configurable.

10 Asimismo, debería observarse que el término “fuente de sonido” cubre tanto fuentes físicas así como también fuentes efectivas de sonido.

La entrada del aparato de acuerdo con la forma de realización de la Fig. 2, 91, 92, ..., 9N comprende información sobre la posición y correspondiente potencia de la pluralidad de N fuentes efectivas de sonido localizadas dentro de un lapso de tiempo o un segmento de tiempo-frecuencia tal como se describe más adelante para el aparato para

15 generar una señal de salida de audio de un micrófono virtual en una posición virtual configurable, y como se describe también en

[20] Giovanni Del Galdo, Oliver Thiergart, Tobias Weller, y E. A. P. Habets. Generating microphone virtual signals using geometrical information gathered by distributed arrays. In Third Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA '11), Edinburgo, Reino Unido, Mayo 2011.

20

Por ejemplo, esta información puede estar comprendida en la salida 106 de la Fig. 14 del módulo de cálculo de información del aparato para generar una señal de salida de audio de un micrófono virtual en una posición virtual configurable considerada más adelante, para 1, 2, ..., N diferentes segmentos de frecuencia cuando se aplica la transformada de Fourier de tiempo reducido (STFT, del inglés *Short-Time Fourier Transform*).

25

Con respecto al aparato para la colocación de micrófonos, se pueden activar diferentes modos de operación durante cierto intervalo de tiempo, cada uno de los cuales implica diversos escenarios para colocar y orientar dichos uno o más micrófonos virtuales. Un aparato para la colocación de micrófonos se puede emplear para una pluralidad de escenarios de aplicación:

30

En un primer escenario de aplicación, N micrófonos omnidireccionales virtuales se pueden colocar dentro de la escena sonora (ver Fig. 4a). Por tanto, en este escenario de aplicación, un número de micrófonos virtuales están cubriendo toda la escena sonora.

35

En un segundo escenario de aplicación, un único micrófono virtual está ubicado en un centro acústico de la escena sonora. Por ejemplo, micrófonos virtuales omnidireccionales, micrófonos virtuales cardioides, o un micrófono virtual espacial (tal como un micrófono de formato B) se colocan de manera tal que todos los participantes se capturan de manera óptima (Fig. 4b).

40

En un tercer escenario de aplicación, se coloca un micrófono espacial ‘fuera’ de la escena sonora. Por ejemplo, un micrófono estéreo virtual se coloca de tal manera que se obtenga una imagen espacial amplia, tal como se ilustra en la Fig. 4c.

45

En un cuarto escenario de aplicación, la orientación óptima del micrófono virtual se estima mientras el micrófono virtual se encuentra en una posición fija (posición predeterminada), por ejemplo la posición y la direccionalidad del micrófono virtual podrían predefinirse, y solamente la orientación se calcula en forma automática.

50

Debería apreciarse que todas las aplicaciones anteriores pueden incluir adaptabilidad temporal. Por ejemplo, la posición/orientación del foco (*spot*) del micrófono virtual sigue un interlocutor a medida que el interlocutor se mueve en la sala.

55

En las Fig. 2 y 3, se proporciona información opcional por una métrica significativa 13, que, por ejemplo, representa una medida de la fiabilidad para la estimación de las posiciones de la ESS. Por ejemplo, dicha métrica puede obtenerse de las varianzas de los estimadores de la dirección de llegada (cuando se usan dos o más conjuntos de micrófonos según lo explicado) tal como se explica a continuación para el aparato para generar una señal de salida de audio de un micrófono virtual en una posición virtual configurable, o a partir del parámetro de difusión calculado de acuerdo con lo descrito en

60

[21] Ville Pulkki. Spatial sound reproduction with directional audio coding. J. Audio Eng. Soc, 55(6): 503–516, junio de 2007.

La métrica puede expresarse ya sea con respecto a todas las entradas 91, ..., 9N, (por ejemplo, puede usarse un valor constante de la métrica para todas las entradas), o puede definirse de manera diferente para cada entrada 91,

..., 9N. Las salidas 15, 16 del aparato de la Fig. 2 pueden comprender la posición y/u orientación de dichos uno o más micrófonos virtuales. Según la aplicación, pueden generarse las salidas (posiciones y orientaciones) para una pluralidad de micrófonos virtuales, cada una de las cuales corresponde a un micrófono virtual específico.

5 La Fig. 5 ilustra un determinador de densidad de potencia espacial 21 de acuerdo con una forma de realización. El determinador de densidad de potencia espacial comprende una unidad de procesamiento principal de densidad de potencia espacial 31 y una unidad postprocesamiento de densidad de potencia espacial 32. El determinador de densidad de potencia espacial 21 se adapta para determinar (o bien calcular) una densidad de potencia espacial (SPD) modificada, indicada en lo sucesivo mediante $\Gamma(x, y, z, k, n)$, que expresa la potencia localizada en un cierto punto, por ejemplo, (x, y, z) en el espacio para cada segmento de tiempo-frecuencia (k, n) . La SPD es generada al integrar los valores de potencia en las posiciones de las fuentes efectivas de sonido 91, ..., 9N, que se introducen en el determinador de densidad de potencia espacial 21.

El cálculo de la SPD para un segmento de tiempo-frecuencia (k, n) se puede realizar de acuerdo con la fórmula

$$15 \quad \Gamma(x, y, z, k, n) = \sum_{i=1}^N \text{potencia}_i(k, n) \cdot g(\gamma_i, x - x_{ESS_i}, y - y_{ESS_i}, z - z_{ESS_i}, k, n), \quad (1)$$

en la que, (x, y, z) representan las coordenadas del sistema y $x_{ESS_i}, y_{ESS_i}, z_{ESS_i}$ son las coordenadas de la fuente efectiva de sonido i . La métrica de importancia γ_i representa un indicador de cómo de fiables son las estimaciones de la posición de cada fuente efectiva de sonido. Por defecto, la métrica de importancia puede ser igual a 1. Debería observarse aquí que la potencia, y las coordenadas x_{ESS_i}, y_{ESS_i} y z_{ESS_i} corresponden a la entrada 9i de la Fig. 3. Asimismo, debería observarse que, para simplificar su denominación, la extensión (k, n) no se escribirá en lo sucesivo. No obstante, las siguientes fórmulas aún dependen del segmento de tiempo-frecuencia (k, n) considerado particular.

25 La SPD generada por la espacial unidad de procesamiento principal de densidad de potencia 31 (por ejemplo en la Fig. 5), puede procesarse adicionalmente por la unidad de postprocesamiento de densidad de potencia espacial 32 (postprocesamiento de SPD y módulo de integración temporal) e integrarse en tiempo, por ejemplo, al emplear un filtro auto-regresivo. Para ser más robusto contra los valores atípicos de la escena sonora (es decir, causado por la estimación incorrecta de la posición), cualquier clase de filtro postprocesamiento puede aplicarse sobre la SPD. Dicho filtro postprocesamiento puede ser, por ejemplo, un filtro de paso bajo o un filtro morfológico (erosión, dilatación).

35 Cuando se calculan la posición y/u orientación de dichos uno o más micrófonos virtuales, se puede emplear un parámetro opcional que depende de la SPD. Este parámetro puede referirse, por ejemplo, a regiones prohibidas y/o preferidas de la sala donde se colocan los micrófonos virtuales (VM), o puede hacer referencia a la SPD, eligiendo rangos específicos de SPD, que satisfacen algunas reglas predeterminadas.

Tal como se puede observar en la fórmula (1), g es una función de la métrica de importancia γ (o bien γ_i) en el espacio, que tiene, por defecto, un valor igual a 1. Por el contrario, γ se puede usar para tener en cuenta diferentes contribuciones. Por ejemplo, si σ^2 es la varianza de la estimación de la posición, entonces, por ejemplo γ se puede fijar a $\gamma = \frac{1}{\sigma^2}$.

45 Como alternativa, es posible emplear la difusión ψ promedio calculada en los conjuntos de micrófonos, dando como resultado $\gamma = 1 - \psi$.

Por consiguiente, γ se puede elegir de tal manera que se reduzca para estimaciones no fiables y se incremente para estimaciones más fiables.

50 Existe una pluralidad de posibilidades para construir la función g . Dos ejemplos particularmente útiles en la práctica son:

$$g(\gamma, x, y, z) = \gamma \cdot \delta(x) \cdot \delta(y) \cdot \delta(z) \quad (2)$$

$$g(\gamma, \mathbf{s}) = \frac{1}{(2\pi)^{3/2} |\Sigma_\gamma|^{1/2}} e^{-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu})' \Sigma_\gamma^{-1} (\mathbf{s}-\boldsymbol{\mu})} \quad (3)$$

55 En la primera función, $\delta(x)$, $\delta(y)$ e $\delta(z)$ indican funciones delta (véase la Fig. 6a que ilustra las funciones delta). En

una segunda función, $\mathbf{s} = [x \ y \ z]^T$, $\boldsymbol{\mu} = [\mu_x \ \mu_y \ \mu_z]^T$ es el vector medio y Σ_γ es la matriz de covarianza de la función g de distribución de Gauss (véase la Fig. 6b que ilustra las funciones de distribución). La matriz de covarianza se calcula usando la siguiente fórmula:

$$\Sigma_\gamma = E[(\mathbf{s} - \boldsymbol{\mu})(\mathbf{s} - \boldsymbol{\mu})^T], \quad (4)$$

que depende de la elección de γ para el escenario donde $\gamma = \frac{1}{\sigma^2}$, teniendo en cuenta que, por ejemplo, para el caso 1D:

$$\sigma^2 = E[(x - \mu_x)^2]. \quad (5)$$

Tal como se puede ver en la fórmula (3), la función g se puede describir mediante una función de distribución en torno a las posiciones de las fuentes efectivas de sonido dadas por las entradas 91 ... 9N, donde por ejemplo, la métrica de importancia es el valor inverso de la varianza de una distribución de Gauss. Si la estimación de una posición de la fuente de sonido tiene alta fiabilidad, la distribución acorde será un tanto acotada, mientras que una estimación poco fiable correspondería a una alta variante y, por tanto, a una distribución amplia; véase, por ejemplo, la Fig. 6b que ilustra un ejemplo 1D.

La Fig. 7 ilustra un estimador de información espacial 22 de acuerdo con una forma de realización. El estimador de información espacial comprende un estimador de centro de escena sonora 41 para estimar una posición de un centro de una escena sonora en el ambiente. Asimismo, el estimador de información espacial comprende un calculador de la posición del micrófono 42 para calcular una posición de un micrófono como la información espacial acústica basándose en la posición del centro de la escena sonora.

La Fig. 8 ilustra un estimador de información espacial 22 de acuerdo con otra forma de realización. El estimador de información espacial comprende un calculador de la posición del micrófono virtual 44 que se adapta para calcular una posición de un micrófono virtual y además se adapta para determinar una orientación de un micrófono virtual. Por tanto, el calculador de la posición del micrófono virtual 44 también se denomina como calculador de la posición/orientación del micrófono 44.

El estimador de información espacial 22 de la Fig. 8 usa como entradas la SPD 23 previamente generada. Devuelve como salidas la posición 15 y la orientación 16 de uno o más micrófonos virtuales, según la aplicación buscada. El primer bloque de procesamiento, el estimador de centro de escena sonora 41, proporciona una estimación del centro de la escena sonora. La salida 43 del bloque 41, por ejemplo, la posición del centro de la escena sonora, se proporciona a continuación como entrada al segundo bloque de procesamiento, el calculador de la posición/orientación del micrófono virtual 44. El calculador de la posición/orientación del micrófono virtual 44 realiza la estimación real de la posición final 15 y la orientación 16 de uno o más micrófonos virtuales, según la aplicación buscada.

El estimador de centro de escena sonora 41 proporciona una estimación del centro de la escena sonora. La salida del estimador de centro de escena sonora 41 se proporciona a continuación como entrada al calculador de la posición/orientación del micrófono 44. El calculador de la posición/orientación del micrófono 44 realiza la estimación real de la posición final 15 y/o la orientación 16 de dichos uno o más micrófonos virtuales de acuerdo con el modo de operación que caracteriza la aplicación buscada.

Las formas de realización del estimador de centro de escena sonora se explican ahora en mayor detalle. Para obtener el centro de la escena sonora, existen varios conceptos posibles.

De acuerdo con un primer concepto de una primera forma de realización, el centro de la escena sonora se obtiene al calcular el centro de gravedad de la SPD $\Gamma(x,y,z)$. El valor de $\Gamma(x,y,z)$ puede interpretarse como la masa existente en el punto (x,y,z) en el espacio.

De acuerdo con un segundo concepto de una segunda forma de realización, la posición en el espacio con una dispersión de tiempo mínimo del canal deberá encontrarse. Esto se consigue al considerar la dispersión del retardo de acuerdo con un valor cuadrático medio (RMS, *Root Mean Squared*). En primer lugar, para cada punto en el espacio $p = (x_0, y_0)$, se calcula un perfil de retardo de potencia (PDP) $A_p(\tau)$ basándose en la SPD $\Gamma(x, y, z)$, por ejemplo usando

$$A_p(\tau) = \int \int \Gamma(x, y) \cdot \delta(t - \tau) dy dx$$

donde $\tau = \sqrt{(x-x_0)^2 + (y-y_0)^2} / c$

A partir de $A_p(\tau)$, se calcula entonces el retardo RMS usando la siguiente ecuación:

$$\tau_{\text{RMS},p} = \sqrt{\frac{\int_0^\infty (\tau - \bar{\tau}_s)^2 A_p(\tau) d\tau}{\int_0^\infty A_p(\tau) d\tau}},$$

5 donde $\bar{\tau}_s$ representa el retardo medio de $A_p(\tau)$. La posición para la cual el retardo medio $\tau_{\text{RMS},p}$ es mínimo representará el centro de la escena sonora.

De acuerdo con un tercer concepto de una tercera forma de realización, que se puede emplear como una alternativa a la estimación del centro de escena sonora, se propone una "integración en círculo". Por ejemplo, en el caso 2D, la SPD $\Gamma(x, y)$ se convoluciona con un círculo $C_{(r,o)}$, de acuerdo con la siguiente fórmula:

10

$$g(x, y) = \Gamma(x, y) * C_{(r,o)}(x, y),$$

en la que r es el radio del círculo, y en la que o define el centro del círculo. El radio r puede ser constante o puede variar según el valor de potencia en el punto (x,y). Por ejemplo, una potencia alta en el punto (x,y) puede corresponder a un radio grande, mientras que una potencia baja puede corresponder a un radio pequeño. También pueden existir otros factores dependientes de la potencia. Un ejemplo tal debería ser convolucionar el círculo con una función de Gauss bivariada antes de usarlo para construir la función g (x, y). De acuerdo con dicha forma de realización, la matriz de covarianza de la función de Gauss bivariada se hace dependiente de la potencia en la posición (x,y), es decir, la potencia alta corresponde a la varianza baja, mientras que la potencia baja corresponde a la varianza alta.

20

Una vez que se calcula g (x, y), el centro de la escena sonora se puede determinar acuerdo con la siguiente fórmula:

$$\hat{p}_{\text{centro}} = \arg \max_{x,y} g(x, y).$$

En otras formas de realización, este concepto se extiende a 3D al emplear una convolución 3D de $\Gamma(x, y, z)$ con una esfera, de manera análoga.

25

La Fig. 9 ilustra un calculador de la posición/orientación del micrófono 44 de acuerdo con otra forma de realización que ilustra más detalles. El centro de la escena sonora 43 se proporciona como entrada al calculador de la posición/orientación del micrófono 44 junto con la SPD 23. En el calculador de la posición/orientación del micrófono 44, la información sobre el centro de la escena sonora 43 se puede copiar, según la operación requerida por la aplicación buscada, a la salida, y usarse directamente como la posición de un micrófono virtual, por ejemplo, cuando el escenario de aplicación de la Fig. 4b es aplicable, en relación con el escenario con un micrófono virtual ubicado en el centro acústico de la escena sonora.

30

Como alternativa, la información sobre el centro de la escena sonora 43 se puede usar como un parámetro de modificación dentro del calculador de la posición/orientación del micrófono 44.

35

Es posible aplicar diferentes conceptos para calcular una posición del micrófono, por ejemplo:

40

- optimización basándose en el ancho de energía proyectada,
- optimización basándose en el análisis de componentes principales.

Puede suponerse, para fines ilustrativos, que la posición del micrófono se calcule de acuerdo con el escenario de aplicación de la Fig. 4c en relación con el escenario de un micrófono espacial fuera de la escena complementaria. No obstante, las explicaciones se aplican de igual modo a cualquier otro escenario de aplicación.

45

Los conceptos para estimar la posición de los micrófonos virtuales de acuerdo con las formas de realización, que se enumeraron previamente, se describirán en mayor detalle a continuación.

50

La optimización basándose en el ancho de energía proyectada define un conjunto de líneas equitativamente espaciadas M que atraviesan el centro de la escena sonora. Para cada una de estas líneas, en por ejemplo, un escenario 2D, la SPD $\Gamma(x,y)$ se proyecta ortogonalmente sobre ellas y se suma.

Las Figs. 10a - 10c ilustran la optimización basándose en el ancho de energía proyectada. En la Fig. 10a, la función

de potencia proyectada P_{proj} se calcula para cada una de las líneas $l_1, \dots, l_i, \dots, l_M$. Los correspondientes anchos de la función se calculan a continuación, véase la Fig. 10b. Por ejemplo, el ancho se puede definir como un ancho de -3 dB, que es equivalente a la distancia para la cual los puntos más a la izquierda y más a la derecha del segmento de distancia corresponden a un nivel de potencia predefinido, por ejemplo, un nivel de potencia superior a -3 dB. Posteriormente, se identifica la línea con el ancho más amplio y el micrófono virtual se coloca en dirección ortogonal a la misma. La orientación del micrófono virtual se puede establecer de tal manera que apunte al centro de la escena sonora, tal como se explica en la siguiente sección. Con este enfoque, se obtienen dos posibles posiciones del micrófono virtual (VM), ya que el VM se puede colocar ya sea en dirección ortogonal positiva o negativa.

La distancia a la cual el VM se coloca puede calcularse, por ejemplo, basándose en consideraciones geométricas junto con el ángulo de apertura del micrófono virtual. Esto se ilustra en la Fig. 10c. La distancia a la cual el VM se coloca varía según el modo de operación de la aplicación buscada. Esto implica construir un triángulo de tal manera que el ancho i de la Fig. 10c represente un lado del triángulo y el centro de gravedad COG es el punto medio del lado. Al llevar la línea ortogonal al COG y definirla como el bisector del ángulo de apertura del VM α , se encuentra el tercer vértice del triángulo. La longitud del bisector proporciona entonces la distancia entre la posición del VM y el centro de la escena sonora.

De acuerdo con otra forma de realización, el concepto de optimización descrito basándose en la energía proyectada se puede extender a 3D. En este caso, se definen planos equitativamente espaciados M^2 (en dirección acimutal y en elevación) en lugar de líneas M . El ancho corresponde entonces al diámetro del círculo que comprende la parte más grande de la energía proyectada. La posición final se obtiene al colocar el VM en la normal a la superficie del plano del diámetro del círculo más grande. De acuerdo con una forma de realización, la distancia desde el centro de la escena sonora hasta la posición del VM puede calcularse nuevamente, de manera similar al caso 2D, es decir, usando consideraciones geométricas y el ángulo de apertura especificado por el modo de operación.

De acuerdo con otra forma de realización, se emplea optimización basándose en un análisis de componentes principales. La optimización basándose en un procesamiento de tipo análisis de componentes principales usa directamente la información disponible a través de la SPD. Al principio, la SPD $\Gamma(x,y,z)$ se cuantifica y se aplica un filtro selectivo de umbral sobre el conjunto de datos cuantificados. De esta manera, se descartan todos los puntos que tienen niveles de energía inferiores a cierto umbral. Con posterioridad, los puntos restantes $h_i = [h_{x,i}, h_{y,i}, h_{z,i}]^T$ se centran en promedio (es decir, los puntos centrados en promedio representan las coordenadas de la fuente efectiva i -ésima menos las coordenadas del centro de la escena sonora), y a continuación se reorganizan en una matriz de datos H de la siguiente manera:

$$H = \begin{pmatrix} h_{x,0} & h_{x,1} & h_{x,2} & \dots & h_{x,N} \\ h_{y,0} & h_{y,1} & h_{y,2} & \dots & h_{y,N} \\ h_{z,0} & h_{z,1} & h_{z,2} & \dots & h_{z,N} \end{pmatrix},$$

donde N define el número de puntos después de establecer el umbral. A continuación, la descomposición en valores singulares (SVD, del inglés *Singular Value Decomposition*) se aplica a H , de tal manera que se factoriza en el siguiente producto:

$$H = U \cdot \Sigma \cdot V^T.$$

La primera columna de U representa el componente principal, que tiene la mayor variabilidad del conjunto de datos. La segunda columna de U es ortogonal a la primera y representa la dirección en la cual queremos colocar el VM. El ancho se proporciona implícitamente por el primer valor singular en la matriz E . Conociendo el ancho, así como también la dirección, podemos calcular la posición y la orientación del VM tal como se describe en el método de optimización basándose en el ancho de energía proyectada, según lo descrito con anterioridad en referencia a las Figs. 10a - 10c.

En otra forma de realización, estos métodos se aplican a un problema 2D, que es directo, ya que solamente es necesario ignorar/eliminar el componente de eje z de las ecuaciones y consideraciones.

Para otras aplicaciones, tales como el escenario de aplicación de la Fig. 4a (una pluralidad de micrófonos virtuales que cubren toda la escena sonora), se puede emplear un concepto diferente, tal como un esquema de optimización iterativa. En una primera etapa, se identifica la posición con el valor máximo de la SPD. En consecuencia, se designa la ubicación del primer VM del total de micrófonos virtuales N . Después de esto, toda la energía que rodea esta posición (es decir, hasta cierta distancia) se elimina de la SPD. Las etapas anteriores se repiten hasta que se encuentran todas las posiciones de los micrófonos virtuales N . En el caso de que N no se defina, la iteración se realiza hasta que el valor máximo en la SPD sea menor que un cierto umbral.

La Fig. 11 ilustra otra forma de realización, en la que un estimador de información espacial 22 comprende, además, un determinador de orientación 45. El determinador de orientación 45 se adapta para determinar una orientación 16 (adecuada) del micrófono basándose en la densidad de potencia espacial 23.

5 A continuación se describirá la estimación de la orientación. Los enfoques de optimización basándose en el ancho de energía proyectada así como también el análisis de componentes principales calculan la orientación del micrófono virtual 15 en forma implícita, ya que se asume que el micrófono virtual está orientado hacia el centro de la escena sonora.

10 Para algunos otros escenarios de aplicación, no obstante, puede ser adecuado calcular la orientación de manera explícita, por ejemplo, en un escenario de aplicación, en el que se estima la orientación óptima del micrófono virtual, en el que el micrófono virtual se ubica en una posición fija. En este caso, la orientación debería determinarse, de tal manera que el micrófono virtual capte la mayor parte de la energía en la escena sonora.

15 De acuerdo con una forma de realización, para determinar la orientación de un micrófono virtual, en primer lugar, las posibles direcciones φ se muestrean y se realiza la integración de acuerdo con la energía de cada una de estas direcciones. Se obtiene la siguiente función de φ :

$$20 \quad f(\varphi) = \int_0^{r_{\max}} \Gamma(r \cos(\varphi), r \sin(\varphi)) \cdot r \, dr,$$

donde r_{\max} se define como la distancia máxima desde el VM y controla el patrón de captación del VM. A continuación, la orientación final ϕ del VM se calcula como:

$$\hat{\phi} = \arg \max_{\phi} \int_{-\pi}^{\pi} f(\varphi) \cdot w_{\phi}(\varphi) \, d\varphi,$$

25 donde $w_{\phi}(\varphi)$ es una función de ponderación basándose en las características de entrada del VM. Por ejemplo, $w_{\phi}(\varphi)$ puede ser la función que define cómo se escala la energía que proviene de la dirección φ dada una cierta dirección de visualización ϕ y un patrón de captación específico del VM.

30 A continuación, se explica un aparato para generar una señal de salida de audio para simular un registro de un micrófono virtual en una posición virtual configurable en un ambiente. Puede emplearse un aparato para la colocación de micrófonos de acuerdo con una de las formas de realización descritas con anterioridad para determinar la posición virtual para el aparato para generar la señal de salida de audio.

35 La Fig. 12 ilustra un aparato para generar una señal de salida de audio para simular un registro de un micrófono virtual en una posición virtual configurable posVmic en un ambiente. El aparato comprende un estimador de la posición de eventos sonoros 110 y un módulo de cálculo de información 120. El estimador de la posición de eventos sonoros 110 recibe una primera información de dirección di1 desde un primer micrófono espacial real y una segunda información de dirección di2 desde un segundo micrófono espacial real. El estimador de la posición de eventos sonoros 110 se adapta para estimar una posición de fuente de sonido ssp que indica una posición de una fuente de sonido en el ambiente, emitiendo la fuente de sonido de onda de sonido, en el que el estimador de la posición de eventos sonoros 110 se adapta para estimar la posición de fuente de sonido ssp basándose en una primera información de dirección di1 proporcionada por un primer micrófono espacial real que está ubicado en una primera posición de micrófono real pos1mic en el ambiente, y basándose en una segunda información de dirección di2 proporcionada por un segundo micrófono espacial real que está ubicado en una segunda posición de micrófono real en el ambiente. El módulo de cálculo de información 120 se adapta para generar la señal de salida de audio basándose en una primera señal de entrada de audio registrada is1 que se registra por el primer micrófono espacial real, basándose en la primera posición de micrófono real pos1mic y basándose en la posición virtual posVmic del micrófono virtual. El módulo de cálculo de información 120 comprende un compensador de propagación que está adaptado para generar una primera señal de audio modificada al modificar la primera señal de entrada de audio registrada is1 compensando un primer retardo o decaimiento de la amplitud entre una llegada de la onda de sonido emitida por la fuente de sonido en el primer micrófono espacial real y una llegada de la onda de sonido en 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 registrada es 1, para obtener la señal de salida de audio.

55 La Fig. 13 ilustra las entradas y salidas de un aparato y un método de acuerdo con una forma de realización. La información de dos o más micrófonos espaciales reales 111, 112, ..., 11N se alimenta al aparato/se procesa por el método. Esta información comprende señales de audio captadas por los micrófonos espaciales reales así como

también información de dirección desde los micrófonos espaciales reales, por ejemplo estimaciones de dirección de llegada (DOA). Las señales de audio y la información de dirección, tal como las estimaciones de dirección de llegada se pueden expresar en un dominio de tiempo-frecuencia. Si, por ejemplo, se desea una reconstrucción geométrica 2D y se elige un dominio STFT (transformada de Fourier de tiempo reducido) tradicional para la representación de las señales, la DOA se puede expresar como ángulos acimutales que dependen de k y n , en concreto los índices de frecuencia y tiempo.

En las formas de realización, la ubicación del evento sonoro en el espacio, así como también la descripción de la posición del micrófono virtual se pueden realizar basándose en las posiciones y orientaciones de los micrófonos espaciales virtuales y reales en un sistema común de coordenadas. Esta información puede representarse por las entradas 121 ... 12N y la entrada 104 en la Fig. 13. La entrada 104 puede especificar adicionalmente la característica del micrófono virtual espacial, por ejemplo, su posición y patrón de captación, tal como se analizará a continuación. Si el micrófono virtual espacial comprende múltiples sensores virtuales, sus posiciones y los correspondientes patrones de captación diferentes pueden considerarse.

La salida del aparato o un correspondiente método puede ser, cuando se desea, una o más señales de sonido 105, que pueden haber sido captadas por un micrófono espacial definido y colocado tal como se especificó para 104. Asimismo, el aparato (o, en su defecto, el método) puede proporcionar como salida la correspondiente información complementaria espacial 106 que puede estimarse empleando el micrófono virtual espacial.

La Fig. 14 ilustra un aparato de acuerdo con una forma de realización, que comprende dos unidades de procesamiento principales, un estimador de la posición de eventos sonoros 201 y un módulo de cálculo de información 202. El estimador de la posición de eventos sonoros 201 puede llevar a cabo la reconstrucción geométrica basándose en las DOA comprendidas en las entradas 111 ... 11N y basándose en el conocimiento de la posición y la orientación de los micrófonos espaciales reales, cuando las DOA se han calculado. La salida del estimador de la posición de eventos sonoros 205 comprende las estimaciones de posición (ya sea en 2D o 3D) de las fuentes de sonido cuando los eventos sonoros ocurren para cada segmento de tiempo y frecuencia. El segundo bloque de procesamiento 202 es un módulo de cálculo de información. De acuerdo con la forma de realización de la Fig. 14, el segundo bloque de procesamiento 202 calcula una señal del micrófono virtual y la información complementaria espacial. Por tanto, también se denomina como bloque de señal del micrófono virtual y cálculo de información complementaria 202. El bloque de señal del micrófono virtual y cálculo de información complementaria 202 usa las posiciones de los eventos sonoros 205 para procesar las señales de audio comprendidas en 111...11N para emitir la señal de audio del micrófono virtual 105. El bloque 202, si se requiere, también puede calcular la información complementaria espacial 106 correspondiente al micrófono virtual espacial. Las formas de realización que siguen ilustran posibilidades de cómo los bloques 201 y 202 pueden operar.

A continuación, la estimación de la posición de un estimador de la posición de eventos sonoros de acuerdo con una forma de realización se describe en mayor detalle.

Según la dimensión del problema (2D o 3D) y el número micrófonos espaciales, varias soluciones para la estimación de la posición son posibles.

Si existen dos micrófonos espaciales en 2D, (el caso posible más simple) es posible una triangulación simple. La Fig. 15 muestra un escenario de ejemplo en el cual los micrófonos espaciales reales se ilustran como Conjuntos Lineales Uniformes (ULA) de 3 micrófonos cada uno. La DOA, expresada como los ángulos acimutales $\alpha_1(k, n)$ y $\alpha_2(k, n)$, se calcula para el segmento de tiempo-frecuencia (k, n) . Esto se consigue al emplear un estimador de la DOA apropiado, tal como ESPRIT,

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

[14] R. Schmidt, "Multiple emitter location and signal parameter estimation", IEEE Transactions on Antennas and Propagation, vol. 34, n.º 3, pág. 276–280, 1986

a las señales de presión transformadas en el dominio de tiempo-frecuencia.

En la Fig. 15, se ilustran dos micrófonos espaciales reales, aquí, dos conjuntos de micrófonos espaciales reales 410, 420. Las dos DOA estimadas $\alpha_1(k, n)$ y $\alpha_2(k, n)$ están representadas por dos líneas, una primera línea 430 que representa DOA $\alpha_1(k, n)$ y una segunda línea 440 que representa DOA $\alpha_2(k, n)$. La triangulación es posible mediante consideraciones geométricas simples conociendo la posición y orientación de cada conjunto.

La triangulación falla cuando las dos líneas 430, 440 están exactamente paralelas. En las aplicaciones reales, no

obstante, esto es muy improbable. No obstante, no todos los resultados de la triangulación corresponden a una posición física o factible para el evento sonoro en el espacio considerado. Por ejemplo, la posición estimada del evento sonoro podría estar demasiado lejos o incluso fuera del espacio asumido, lo que indica que probablemente las DOA no correspondan a ningún evento sonoro que pueda interpretarse físicamente con el modelo usado. Dichos resultados pueden producirse por el ruido del sensor o una reverberación de la sala demasiado fuerte. Por tanto, de acuerdo con una forma de realización, dichos resultados indeseados se señalizan de tal manera que el módulo de cálculo de información 202 los pueda tratar de forma apropiada.

La Fig. 16 representa un escenario donde la posición de un evento sonoro se estima en espacio 3D. Se emplean micrófonos espaciales apropiados, por ejemplo, un conjunto de micrófonos plano o 3D. En la Fig. 16, se ilustra un primer micrófono espacial 510, por ejemplo, un primer conjunto de micrófonos 3D, y un segundo micrófono espacial 520, por ejemplo, un primer conjunto de micrófonos 3D. La DOA en el espacio 3D puede expresarse, por ejemplo, como azimut y elevación. Los vectores unitarios 530, 540 se pueden emplear para expresar las DOA. Dos líneas 550, 560 se proyectan de acuerdo con las DOA. En 3D, incluso con estimaciones muy fiables, las dos líneas 550, 560 proyectadas de acuerdo con las DOA podrían no intersectarse. No obstante, la triangulación puede todavía llevarse a cabo, por ejemplo, al elegir el punto medio del segmento más pequeño que conecta las dos líneas.

De manera similar al caso 2D, la triangulación puede fallar o puede generar resultados no factibles para ciertas combinaciones de direcciones, que pueden entonces señalizarse, por ejemplo al módulo de cálculo de información 202 de la Fig. 14.

Si existen más de dos micrófonos espaciales, varias soluciones son posibles. Por ejemplo, la triangulación explicada con anterioridad podría llevarse a cabo para todos los pares de los micrófonos espaciales reales (si N = 3, 1 con 2, 1 con 3, y 2 con 3). Las posiciones resultantes pueden entonces promediarse (a lo largo de x e y, y, si se considera 3D, z).

Como alternativa, se pueden usar conceptos más complejos. Por ejemplo, los enfoques probabilísticos se pueden aplicar tal como se describe en

[15] J. Michael Steele, "Optimal Triangulation of Random Samples in the Plane", The Annals of Probability, Vol. 10, N.º 3 (agosto de 1982), págs. 548-553.

De acuerdo con una forma de realización, el campo de sonido se puede analizar en el dominio de tiempo-frecuencia, por ejemplo, obtenido a través de una transformada de Fourier de tiempo reducido (STFT), en donde k y n indican 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 para cierto k y n se modela como una única onda esférica emitida por una fuente del tipo de punto isotrópico de banda estrecha, por ejemplo al emplear la fórmula:

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

donde $P_{IPLS}(k, n)$ 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 ejemplo, introduce modificaciones apropiadas de fase y magnitud. Aquí se puede asumir que, en cada segmento de tiempo-frecuencia, solamente una IPLS está activa. No obstante, las múltiples IPLS de banda estrecha ubicadas en diferentes posiciones también pueden estar activas en un único instante de tiempo.

Cada IPLS modela el sonido directo o un reflejo de sala distinto. Su posición $p_{IPLS}(k, n)$ puede idealmente corresponder a una fuente de sonido real ubicada en el interior de la sala, o una fuente de sonido especular ubicada fuera, respectivamente. Por tanto, la posición $p_{IPLS}(k, n)$ también puede indicar la posición de un evento sonoro.

Ha de destacarse que la expresión "fuente de sonido real" indica la fuente de sonido real, que existe físicamente en el ambiente de registro, tal como los interlocutores o instrumentos musicales. Por el contrario, con las "fuentes de sonido" o los "eventos sonoros" o "IPLS", hacemos referencia a las fuentes efectivas de sonido, que están activas en cierto instante de tiempo o en ciertos segmentos de tiempo-frecuencia, en el que las fuentes de sonido pueden representar, por ejemplo, fuentes de sonido reales o fuentes especulares.

Las Figs. 27a-27b ilustran conjuntos de micrófonos que localizan fuentes de sonido. Las fuentes de sonido localizadas pueden tener diferentes interpretaciones físicas según su naturaleza. Cuando los conjuntos de micrófonos reciben el sonido directo, pueden localizar la posición de una fuente de sonido verdadera (por ejemplo interlocutores). Cuando los conjuntos de micrófonos reciben reflejos, pueden localizar la posición de una fuente especular. Las fuentes especulares también son fuentes de sonido.

La Fig. 27a ilustra un escenario, donde dos conjuntos de micrófonos 151 y 152 reciben sonido directo de una fuente

de sonido real (una fuente de sonido físicamente existente) 153.

La Fig. 27b ilustra un escenario, donde dos conjuntos de micrófono 161, 162 reciben el sonido reflejado, en el que el sonido se ha reflejado por una pared. Debido a la reflexión, los conjuntos de micrófonos 161, 162 localizan la posición, de donde parece provenir el sonido, a una posición de una fuente de imagen especular 165, que es diferente de la posición del altavoz 163.

La fuente de sonido real 153 de la Fig. 27a, así como la fuente de imagen especular 165 son fuentes de sonido.

La Fig. 27c ilustra un escenario, donde dos conjuntos de micrófono 171, 172 reciben sonido difuso y no pueden localizar una fuente de sonido.

Si bien este modelo de onda única es preciso solo para los ambientes moderadamente reverberantes dado que las señales fuente cumplen la condición de ortogonalidad disjunta W (WDO), es decir, la superposición de tiempo-frecuencia es suficientemente pequeña. Esto normalmente es válido para las señales del habla, véase, por ejemplo,

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

Sin embargo, el modelo también proporciona una buena estimación para otros ambientes y en consecuencia también es aplicable para estos ambientes.

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 un IPLS activo en un cierto segmento de tiempo-frecuencia, y en consecuencia la estimación de un evento de sonido en un segmento de tiempo-frecuencia, se estima por medio de la triangulación basándose en la dirección de llegada (DOA) del sonido medido en al menos dos puntos de observación diferentes.

La Fig. 17 ilustra una geometría, donde el IPLS del intervalo de frecuencia-tiempo actual (k, n) se ubica en la posición desconocida $p_{IPLS}(k, n)$. Para determinar la información de DOA requerida, se emplean dos micrófonos espaciales reales, aquí, dos conjuntos de micrófonos, que tiene una geometría, posición y orientación conocidas, que se colocan en las posiciones 610 y 620, respectivamente. Los vectores p_1 y p_2 señalan las posiciones 610, 620, respectivamente. Las orientaciones del conjunto están definidas por los vectores unitarios c_1 y c_2 . La DOA del sonido se determina en las posiciones 610 y 620 para cada (k, n) usando un algoritmo de estimación de DOA, por ejemplo como proporciona en el análisis DirAC (véase [2], [3]). Debido a esto, un primer vector unitario de punto de vista $e_1^{POV}(k, n)$ y un segundo vector unitario de punto de vista $e_2^{POV}(k, n)$ con respecto a un punto de vista de los conjuntos de micrófono (ambos no mostrados en la Fig. 17) se pueden proporcionar como resultado del análisis DirAC. Por ejemplo, cuando se opera en el 2D, el primer vector unitario de punto de vista produce:

$$e_1^{POV}(k, n) = \begin{bmatrix} \cos(\varphi_1(k, n)) \\ \text{sen}(\varphi_1(k, n)) \end{bmatrix}, \quad (2)$$

Aquí, $\varphi_1(k, n)$ representa el azimut de la DOA estimada en el primer conjunto de micrófono, como se ilustra en la Fig. 17. Los correspondientes vectores unitarios DOA $e_1(k, n)$ y $e_2(k, n)$, con respecto al sistema de coordenadas global en el origen, se puede calcular mediante la aplicación de las fórmulas:

$$\begin{aligned} e_1(k, n) &= R_1 \cdot e_1^{POV}(k, n), \\ e_2(k, n) &= R_2 \cdot e_2^{POV}(k, n), \end{aligned} \quad (3)$$

donde R son matrices de transformación de las coordenadas, por ejemplo,

$$R_1 = \begin{bmatrix} c_{1,x} & -c_{1,y} \\ c_{1,y} & c_{1,x} \end{bmatrix}, \quad (4)$$

cuando opera en 2D y $c_1 = [c_{1,x}, c_{1,y}]^T$. Para llevar a cabo la triangulación, los vectores de dirección $d_1(k, n)$ y $d_2(k, n)$ se pueden calcular como:

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

5 donde $d_1(k, n) = \|\mathbf{d}_1(k, n)\|$ y $d_2(k, n) = \|\mathbf{d}_2(k, n)\|$ son las distancias no conocidas entre IPLS y los dos conjuntos de micrófono. La siguiente ecuación

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

10 se puede resolver para $d_1(k, n)$. Finalmente, la posición $\mathbf{p}_{\text{IPLS}}(k, n)$ de la IPLS se da en

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

15 En otra forma de realización, la ecuación (6) se puede resolver con $d_2(k, n)$ y $\mathbf{p}_{\text{IPLS}}(k, n)$ se calcula de modo análogo empleando $d_2(k, n)$.

La ecuación (6) siempre proporciona una solución cuando se opera en 2D, a menos que $\mathbf{e}_1(k, n)$ y $\mathbf{e}_2(k, n)$ sean paralelos. Sin embargo, cuando se usan más de dos conjuntos de micrófono 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 puede calcular el punto que es más cercano a todos los vectores de dirección \mathbf{d} y el resultado se puede usar como la posición del IPLS.

25 En una forma de realización, todos los puntos de observación $\mathbf{p}_1, \mathbf{p}_2, \dots$ se deben ubicar de modo que el sonido emitido por la IPLS se incluye en el mismo bloque temporal n . Este requerimiento se puede cumplir simplemente cuando la distancia Δ entre cualquiera de dos de los puntos de observación es menor de

$$\Delta_{\max} = c \frac{n_{\text{FFT}}(1 - R)}{f_s}, \tag{8}$$

30 donde n_{FFT} es la longitud de la ventana STFT, $0 \leq R < 1$ especifica la superposición entre las tramas de tiempo sucesivas y f_s es la frecuencia de muestreo. Por ejemplo, para un STFT de 1024 puntos a 48 kHz con 50 % de superposición ($R = 0,5$), el espaciado máximo entre los conjuntos para cumplir el requisito anterior es $\Delta = 3,65$ m.

35 A continuación, se describe con más detalle un módulo de cálculo de información 202, por ejemplo, una señal de micrófono virtual y el módulo de cálculo de la información secundaria, de acuerdo con una forma de realización.

40 La Fig. 18 ilustra una vista general esquemática de un módulo de cálculo de la información 202 de acuerdo con una forma de realización. La unidad de cálculo de la información comprende un compensador de propagación 500, un combinador 510 y una unidad de ponderación espectral 520. El módulo de cálculo de información 202 recibe las estimaciones de la posición de la fuente de sonido ssp estimadas por un estimador de la posición de los eventos sonoros, una o más señales de entrada de audio se registran 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 posVm del micrófono virtual. Esto produce una señal de salida de audio os que representa una señal de audio del micrófono virtual.

45 La Fig. 19 ilustra un módulo de cálculo de información de acuerdo con otra forma de realización. El módulo de cálculo de información de la Fig. 19 comprende un compensador de propagación 500, un combinador 510 y una unidad de ponderación espectral 520. El compensador de propagación 500 comprende un módulo de cálculo de los parámetros de propagación 501 y un módulo de compensación de la propagación 504. El combinador 510 comprende un módulo de cálculo de los factores de combinación 502 y un módulo de combinación 505. La unidad de ponderación espectral 520 comprende una unidad de cálculo de ponderaciones espectrales 503, un módulo de aplicación de ponderaciones espectrales 506 y un módulo de cálculo de información secundaria espacial 507.

55 Para calcular la señal de audio del micrófono virtual, la información geométrica, por ejemplo, la posición y orientación de los micrófonos espaciales reales 121 ... 12N, la posición, orientación y características del micrófono espacial virtual 104, y las estimaciones de posición de los eventos sonoros 205 se alimentan en el módulo de cálculo de

información 202, en particular, en el módulo de cálculo de los parámetros de propagación 501 del compensador de propagación 500, en el módulo de cálculo de los factores de combinación 502 del combinador 510 y en la unidad de cálculo de ponderaciones espectrales 503 de la unidad de ponderación espectral 520. El módulo de cálculo de los parámetros de propagación 501, el módulo de cálculo de los factores de combinación 502 y la unidad de cálculo de ponderaciones espectrales 503 calculan los parámetros usados 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 ponderaciones espectrales 506.

En el módulo de cálculo de información 202, las señales de audio 111 ... 11N en primer lugar se pueden modificar para compensar los efectos dados por las diferentes longitudes de propagación entre las posiciones del evento sonoro y los micrófonos espaciales reales. A continuación las señales se pueden combinar para mejorar por ejemplo la relación señal a ruido (SNR). Finalmente, la señal resultante a continuación se puede ponderar espectralmente para tener en cuenta el patrón de captación direccional del micrófono virtual, así como cualquier función de ganancia dependiente de la distancia. Estas tres etapas se describen con más detalle a continuación.

La compensación de la propagación se explica a continuación con más detalle. En la porción superior de la Fig. 20, se ilustran dos micrófonos espaciales reales (un primer conjunto de micrófonos 910 y un segundo conjunto de micrófonos 920), la posición de un evento sonoro localizado 930 para el intervalo de tiempo-frecuencia (k, n), y la posición del micrófono espacial virtual 940.

La porción inferior de la Fig. 20 ilustra un eje temporal. Se supone que un evento sonoro se emite en el tiempo t0 y a continuación se propaga a los micrófonos espaciales reales y virtuales. Los retardos de tiempo de llegada así como las amplitudes cambian con la distancia, de modo a mayor longitud de propagación, amplitud más débil y mayor retardo del tiempo de llegada.

Las señales en los dos conjuntos reales son comparables solo si el retardo relativo Dt12 entre ellos es pequeño. Por otra parte, una de las dos señales necesita realinearse temporalmente para compensar el retardo relativo Dt12, y posiblemente, escalarse para compensar los diferentes decaimientos.

La compensación del retardo entre la llegada al micrófono virtual y llegada a los conjuntos de micrófono reales (en uno de los micrófonos espaciales reales) cambia el retardo independientemente de la ubicación del evento sonoro, lo que lo hace superfluo para la mayor parte de las aplicaciones.

Nuevamente con respecto a la Fig. 19, el módulo de cálculo de los parámetros de propagación 501 está adaptado para calcular los retardos para corregir para cada micrófono espacial real y para cada evento sonoro. Si se desea, también calcula los factores de ganancia que se deben considerar para compensar los diferentes decaimientos de amplitud.

Por consiguiente, el módulo de compensación de la propagación 504 se configura para usar esta información para modificar las señales de audio. Si las señales tienen que desplazarse en una pequeña cantidad de tiempo (en comparación con la ventana de tiempo del banco de filtro), entonces es suficiente una rotación de fase simple. Si los retardos son mayores, se necesitan implementaciones más complicadas.

La salida del módulo de compensación de la propagación 504 son las señales de audio modificadas expresadas en el dominio de tiempo-frecuencia original.

A continuación, una estimación particular de compensación de la propagación para un micrófono virtual de acuerdo con una forma de realización se describirá con referencia a la Fig. 17 que entre otros ilustra la posición 610 de un primer micrófono espacial real y la posición 620 de un segundo micrófono espacial real.

En la forma de realización que se explica a continuación, se supone que al menos una primera señal de entrada de audio registrada, por ejemplo, una señal de presión de al menos uno de los micrófonos espaciales reales (por ejemplo, los conjuntos de micrófono) está disponible, por ejemplo, la señal de presión de un primer micrófono espacial real. Los autores se referirán 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 no solo se puede realizar con respecto a solo una señal de presión, sino también con respecto a las señales de presión de una pluralidad o el total de los micrófonos espaciales reales.

La relación entre la señal de presión $P_{IPLS}(k, n)$ emitida por el IPLS y una señal de presión de referencia $P_{ref}(k, n)$ de un micrófono de referencia ubicado en p_{ref} puede expresarse por la fórmula (9):

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

En general, el factor complejo $\gamma(k, p_a, p_b)$ expresa la rotación de fases y decaimiento de amplitud introducido por la propagación de una onda esférica de su origen en p_a a p_b . Sin embargo, los ensayos prácticos indicados que consideran solo el decaimiento de la amplitud en γ conducen a impresiones plausibles de la señal del micrófono virtual con significativamente menos artefactos en comparación también con la consideración de la rotación de la fase.

La energía del sonido que se puede medir en un punto determinado en el espacio depende fuertemente de la distancia r desde la fuente de sonido, en la Fig 6 desde la posición p_{IPLS} de la fuente de sonido. En muchas situaciones, esta dependencia se puede modelar con suficiente precisión usando principios físicos bien conocidos, por ejemplo, decaimiento $1/r$ 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 de sonido, y cuando también se conoce la distancia del micrófono virtual desde la fuente de sonido, entonces, se puede estimar la energía del sonido en la posición del micrófono virtual desde la señal y la energía del micrófono de referencia, por ejemplo el primer micrófono espacial real. Esto significa, que la señal de salida del micrófono virtual se puede obtener por la aplicación de ganancias apropiadas a la señal de presión de referencia.

Suponiendo que el primer micrófono espacial real es el micrófono de referencia, entonces $p_{ref} = p_1$. En la Fig. 17, el micrófono virtual se ubica en p_v . Debido a que se conoce en detalle la geometría de la Fig. 17, la distancia $d_1(k, n) = ||d_1(k, n)||$ entre el micrófono de referencia (en la Fig. 17: el primer micrófono espacial real) y el IPLS se puede determinar fácilmente, así como la distancia $s(k, n) = ||s(k, n)||$ entre el micrófono virtual y el IPLS, en concreto

$$s(k, n) = ||s(k, n)|| = ||p_1 + d_1(k, n) - p_v||. \quad (10)$$

La presión sonora $P_v(k, n)$ en la posición del micrófono virtual se calcula por la combinación de las fórmulas (1) y (9), que lleva a

$$P_v(k, n) = \frac{\gamma(k, p_{IPLS}, p_v)}{\gamma(k, p_{IPLS}, p_{ref})} P_{ref}(k, n). \quad (11)$$

Como se mencionó antes, en algunas formas de realización, los factores γ solo pueden considerar el decaimiento de la amplitud debido a la propagación. Suponiendo por ejemplo que la presión sonora disminuye con $1/r$, entonces

$$P_v(k, n) = \frac{d_1(k, n)}{s(k, n)} P_{ref}(k, n). \quad (12)$$

Cuando se mantiene el modelo de la fórmula (1), por ejemplo, cuando solo está presente el sonido directo, entonces la fórmula (12) puede reconstruir con precisión la información de la magnitud. Sin embargo, en caso de campos del sonido difuso puros, por ejemplo, cuando no se cumplen las suposiciones del modelo, el modelo presentado produce una des-reverberación implícita de la señal cuando el micrófono virtual se mueve lejos de las posiciones de los conjuntos sensores. En efecto, como se describió antes, en los campos del sonido difuso, se espera que la mayor parte de IPLS se localicen cerca de los dos conjuntos sensores. En consecuencia, cuando el micrófono virtual se mueve lejos de estas posiciones, probablemente se aumenta la distancia $s = ||s||$ en la Fig. 17. En consecuencia, la magnitud de la presión de referencia disminuye cuando se aplica una ponderación de acuerdo con la fórmula (11). En forma correspondiente, cuando se mueve el micrófono virtual cerca de una fuente de sonido real, los segmentos de tiempo-frecuencia que corresponden al sonido directo se amplificarán de modo que se percibirá la señal de audio global menos difusa. Mediante el ajuste de la regla en la fórmula (12), se puede controlar la amplificación del sonido directo y supresión del sonido difuso a voluntad.

Mediante la realización de la compensación de la propagación sobre la señal de entrada de audio registrada (por ejemplo la señal de presión) del primer micrófono espacial real, se obtiene una primera señal de audio modificada.

En las formas de realización, una segunda señal de audio modificada se puede obtener por la realización de la compensación de la propagación sobre una segunda señal de entrada registrada (segunda señal de presión) del segundo micrófono espacial real.

En otras formas de realización, otras señales de audio se pueden obtener por la realización de la compensación de la propagación sobre otras señales de entrada de audio registradas (otras señales de presión) de otros micrófonos

espaciales reales.

A continuación, se explica con más detalle la combinación de los bloques 502 y 505 de la Fig. 19 de acuerdo con una forma de realización. Se supone que dos o más señales de audio de una pluralidad de micrófonos espaciales reales diferentes se han modificado para compensar las diferentes trayectorias de propagación para obtener dos o más señales de audio modificadas. Una vez que las señales de audio de los diferentes micrófonos espaciales reales se han modificado para compensar las diferentes trayectorias de propagación, se pueden combinar para mejorar la calidad del audio. Al hacerlo, por ejemplo, se puede aumentar la SNR o se puede reducir la reverberancia.

10 Las posibles soluciones para la combinación comprenden:

- Promediado ponderado, por ejemplo, considerar SNR, o la distancia al micrófono virtual, o el grado de difusión que se estimó mediante los micrófonos espaciales reales. Se pueden emplear las soluciones tradicionales, por ejemplo, relación de combinación máxima (MRC) o combinación de ganancia igual (EQC), o
- Combinación lineal de algunas o todas las señales de audio modificadas para obtener una señal de combinación. Las señales de audio modificadas se pueden ponderar en la combinación lineal para obtener la señal de la combinación, o
- Selección, por ejemplo, solo se usa una señal, por ejemplo, dependiente de SNR o distancia o grado de difusión.

La tarea del módulo 502, si es aplicable, es calcular los parámetros para la combinación, lo que se lleva a cabo en el módulo 505.

25 A continuación, se describe con más detalle la ponderación espectral de acuerdo con las formas de realización. Para esto, se hace referencia a los bloques 503 y 506 de la Fig. 19. En esta etapa final, la señal de audio resultante de la combinación o de la compensación de la propagación de la señales de audio de entrada se pondera en el dominio de tiempo-frecuencia de acuerdo con las características espaciales del micrófono espacial virtual como se especifica en la entrada 104 y/o de acuerdo con la geometría reconstruida (dada en 205).

30 Para cada segmento de tiempo-frecuencia, la reconstrucción geométrica nos permite obtener fácilmente la DOA relativa al micrófono virtual, como se muestra en la Fig. 21. Además, también se puede calcular fácilmente la distancia entre el micrófono virtual y la posición del evento sonoro.

35 La ponderación para el segmento de tiempo-frecuencia se calcula a continuación considerando el tipo de micrófono virtual deseado.

En caso de micrófonos direccionales, las ponderaciones espectrales se pueden calcular 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 $g(\theta)$,

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

45 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.

Otra posibilidad son las funciones de decaimiento artístico (no físicas). En ciertas aplicaciones, se puede desear suprimir eventos sonoros alejados del micrófono virtual con un factor más grande que el que caracteriza la propagación en campo libre. Para este fin, algunas formas de realización introducen una función de ponderación adicional que depende de la distancia entre el micrófono virtual y el evento sonoro. En una forma de realización, solo se deben captar los eventos sonoros dentro de una determinada distancia (por ejemplo en metros) del micrófono virtual.

55 Con respecto a la direccionalidad del micrófono virtual, se puede aplicar patrones de direccionalidad arbitraria para el micrófono virtual. De este modo, se puede separar por ejemplo una fuente de una escena sonora compleja.

Debido a que la DOA del sonido se puede calcular en la posición p_v del micrófono virtual, en concreto

$$\varphi_v(k, n) = \arccos \left(\frac{s \cdot c_v}{\|s\|} \right), \tag{13}$$

60 donde c_v es un vector unitario que describe la orientación del micrófono virtual, se pueden realizar las

direccionales arbitrarias para el micrófono virtual. Por ejemplo, suponiendo que $P_v(k,n)$ indica la señal de combinación o la señal de audio modificado compensado por propagación, entonces la fórmula:

$$\tilde{P}_v(k, n) = P_v(k, n) [1 + \cos(\varphi_v(k, n))] \quad (14)$$

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

En las formas de realización, uno o más micrófonos no espaciales reales, por ejemplo, un micrófono omnidireccional o un micrófono direccional tal como un cardioide, se colocan en la escena sonora además de los micrófonos espaciales reales para mejorar adicionalmente la calidad del sonido de las señales del micrófono virtual 105 en la Figura 8. Estos micrófonos no se usan para reunir ninguna información geométrica, sino más bien solo para proporcionar una señal de audio más limpia. Estos micrófonos se pueden colocar más cercanos a las fuentes de sonido que los micrófonos espaciales. En este caso, de acuerdo con una forma de realización, las señales de audio de los micrófonos no espaciales reales y sus posiciones se alimentan simplemente al módulo de compensación de la propagación 504 de la Fig. 19 para el procesamiento, en vez de las señales de audio de los micrófonos espaciales reales. La compensación de la propagación a continuación se realiza para la una o más señales de audio registradas de los micrófonos no espaciales con respecto a la posición de uno o más micrófonos no espaciales. Por esta razón, se realiza una forma de realización usando micrófonos no espaciales adicionales.

En una forma de realización adicional, se realiza el cálculo de la información secundaria espacial del micrófono virtual. Para calcular la información secundaria espacial 106 del micrófono, el módulo de cálculo de información 202 de la Fig. 19 comprende un módulo de cálculo de información secundaria espacial 507, que está adaptado para recibir como entrada las posiciones de la fuente de sonido 205 y la posición, orientación y características 104 del micrófono virtual. En ciertas formas de realización, de acuerdo con la información secundaria 106 que se necesita calcular, la señal de audio del micrófono virtual 105 también se puede tener en cuenta como entrada al módulo de cálculo de información secundaria espacial 507.

La salida del módulo de cálculo de información secundaria espacial 507 es la información secundaria del micrófono virtual 106. Esta información secundaria puede ser, por ejemplo, la DOA o el grado de difusión para cada segmento de tiempo-frecuencia (k, n) desde el punto de vista del micrófono virtual. Otra información secundaria posible puede ser, por ejemplo, el vector de intensidad del sonido activo $l_a(k, n)$ que puede haberse medido en la posición del micrófono virtual. A continuación se describirán cómo se pueden obtener estos parámetros.

De acuerdo con una forma de realización, se realiza la estimación de DOA para el micrófono espacial virtual. El módulo de cálculo de información 120 está adaptado para estimar la dirección de la llegada al micrófono virtual como información secundaria espacial, basándose en un vector de posición del micrófono virtual y basándose en un vector de posición del evento sonoro como se ilustra en la Fig. 22.

La Fig. 22 ilustra una manera posible de obtener la DOA del sonido desde el punto de vista del micrófono virtual. La posición del evento sonoro, proporcionada por el bloque 205 en la Fig. 19, se puede describir para cada componente de tiempo-frecuencia (k, n) con un vector de posición $r(k, n)$, el vector de posición del evento sonoro. De modo similar, la posición del micrófono virtual, proporcionada como la entrada 104 en la Fig. 19, se puede describir con un vector de posición $s(k,n)$, el vector de posición del micrófono virtual. La dirección de mira del micrófono virtual se puede describir con un vector $v(k, n)$. La DOA relativa al micrófono virtual se proporciona mediante $a(k,n)$. Esta representa el ángulo entre v y la trayectoria de propagación del sonido $h(k,n)$. $h(k, n)$ se puede calcular mediante el empleo de la fórmula:

$$h(k, n) = s(k,n) - r(k, n).$$

La DOA deseada $a(k, n)$ a continuación se puede calcular para cada (k, n) por ejemplo por medio de la definición del producto punto de $h(k, n)$ y $v(k,n)$, es decir

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

En otra forma de realización, el módulo de cálculo de información 120 se puede adaptar para estimar la intensidad del sonido activo en el micrófono virtual como información secundaria espacial, basándose en un vector de posición del micrófono virtual y basándose en un vector de posición del evento sonoro como se ilustra en la Fig. 22.

A partir de la DOA $a(k, n)$ definida anteriormente, se puede obtener la intensidad del sonido activo $l_a(k, n)$ en la posición del micrófono virtual. Para esto, se supone que la señal de audio del micrófono virtual 105 en la Fig. 19 corresponde a la salida de un micrófono omnidireccional, por ejemplo, se supone que el micrófono virtual es un

micrófono omnidireccional. Más aún, la dirección de mirada v en la Fig. 22 se supone que es paralela al eje x del sistema de coordenadas. Debido al vector de intensidad del sonido activo deseado $la(k, n)$ describe el flujo neto de energía a través de la posición del micrófono virtual, podemos calcular $la(k, n)$, por ejemplo de acuerdo con la fórmula:

$$5 \quad \mathbf{la}(k, n) = - (1/2 \text{ rho}) |P_v(k, n)|^2 * [\cos a(k, n), \text{sen } a(k, n)]^T,$$

donde $[\]^T$ indica un vector transpuesto, rho es la densidad del aire, y $P_v(k, n)$ es la presión sonora medida por el micrófono espacial virtual, por ejemplo, la salida 105 del bloque 506 en la Fig. 19.

10 Si el vector de intensidad activo se debiera calcular expresado en el sistema de coordenadas general pero aún en la posición del micrófono virtual, se puede aplicar la siguiente fórmula:

$$15 \quad \mathbf{la}(k, n) = (1/2 \text{ rho}) |P_v(k, n)|^2 \mathbf{h}(k, n) / \|\mathbf{h}(k, n)\|.$$

El grado de difusión del sonido expresa cómo se difunde el campo de sonido en un intervalo de tiempo-frecuencia determinado (véase, por ejemplo, [2]). El grado de difusión se expresa como un valor ψ , en el que $0 \leq \psi \leq 1$. Un grado de difusión de 1 indica que la energía total del campo de sonido de un campo sonoro es completamente difusa. Esta información es importante por ejemplo en la reproducción del sonido espacial. Tradicionalmente, el grado de difusión se calcula en el punto específico del espacio en que se coloca un conjunto de micrófonos.

De acuerdo con una forma de realización, el grado de difusión se puede calcular como un parámetro adicional a la información secundaria generada por el micrófono virtual (VM), que se puede colocar a voluntad en una posición arbitraria en la escena sonora. Por esta razón, un aparato que también calcula el grado de difusión además de la señal de audio en una posición virtual de un micrófono virtual se puede observar como un extremo delantero DirAC virtual, ya que es posible producir una corriente DirAC, en concreto una señal de audio, dirección de llegada y grado de difusión, para un punto arbitrario de la escena sonora. La corriente DirAC también se puede procesar, almacenar, transmitir, y reproducir en un ajuste multi-altavoz arbitrario. En este caso, el oyente experimenta la escena sonora como si él o ella estuvieran en la posición especificada por el micrófono virtual y estuvieran mirando en la dirección determinada por su orientación.

La Fig. 23 ilustra un bloque de cálculo de la información de acuerdo con una forma de realización que comprende una unidad de cálculo del grado de difusión 801 para calcular el grado de difusión en el micrófono virtual. El bloque de cálculo de la información 202 está adaptado para recibir las entradas 111 a 11N, que además de las entradas de la Fig. 14 también incluye el grado de difusión en los micrófonos espaciales reales. $\psi^{(SM1)}$ a $\psi^{(SMN)}$ indican estos valores. Estas entradas adicionales se alimentan al módulo de cálculo de información 202. La salida 103 de la unidad de cálculo del grado de difusión 801 es el parámetro del grado de difusión calculado en la posición del micrófono virtual.

Una unidad de cálculo del grado de difusión 801 de una forma de realización se ilustra en la Fig. 24 que representa 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. Entonces, usando la información sobre las posiciones del IPLS, y la información en las posiciones de los micrófonos espaciales y virtuales, se obtienen N estimaciones de estas energías en la posición del micrófono virtual. Finalmente, las estimaciones se pueden combinar para mejorar la precisión de la estimación y se puede calcular fácilmente el parámetro del grado de difusión en el micrófono virtual.

45 $E_{\text{dir}}^{(SM1)}$ a $E_{\text{dir}}^{(SMN)}$ y $E_{\text{diff}}^{(SM1)}$ a $E_{\text{diff}}^{(SMN)}$ indican las estimaciones de las energías del sonido directo y difuso para los N micrófonos espaciales calculados por la unidad de análisis de energía 810. Si P_i es la señal de presión compleja y ψ_i es el grado de difusión para el micrófono espacial i -ésimo, a continuación las energías se pueden calcular, por ejemplo, de acuerdo con las fórmulas:

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

$$E_{\text{diff}}^{(SMi)} = \Psi_i \cdot |P_i|^2$$

55 La energía del sonido difuso debe ser igual en todas las posiciones, en consecuencia, se puede calcular simplemente una estimación de la energía de sonido difuso $E_{\text{diff}}^{(VM)}$ en el micrófono virtual promediando $E_{\text{diff}}^{(SM1)}$ a $E_{\text{diff}}^{(SMN)}$, por ejemplo en una unidad de combinación de grado de difusión 820, por ejemplo, de acuerdo con la fórmula:

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

5 Una combinación más efectiva de las estimaciones $E_{\text{diff}}^{(\text{SM}1)}$ a $E_{\text{diff}}^{(\text{SM}N)}$ se puede realizar considerando la varianza de los estimadores, por ejemplo, considerando la SNR.

La energía del sonido directo depende de la distancia a la fuente debido a la propagación. En consecuencia, $E_{\text{dir}}^{(\text{SM}1)}$ a $E_{\text{dir}}^{(\text{SM}N)}$ se puede modificar teniendo esto en cuenta. Esto se puede llevar a cabo, por ejemplo, mediante una unidad de ajuste de la propagación del sonido directo 830. Por ejemplo, si se supone que la energía del campo de sonido directo decae con 1 respecto de la distancia al cuadrado, entonces la estimación para el sonido directo en el micrófono virtual para el micrófono espacial i-ésimo se puede calcular de acuerdo con la fórmula:

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

15 De modo similar a la unidad de combinación del grado de difusión 820, se pueden combinar las estimaciones de la energía del sonido directo obtenida en diferentes micrófonos espaciales, por ejemplo por una unidad de combinación de sonido directo 840. El resultado es $E_{\text{dir}}^{(\text{VM})}$, por ejemplo, la estimación para la energía del sonido directo en el micrófono virtual. Se puede calcular el grado de difusión del micrófono virtual $\Psi^{(\text{VM})}$, por ejemplo, por un sub-calculador del grado de difusión 850, por ejemplo de acuerdo con la fórmula:

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

20 Como se mencionó antes, en algunos casos, la estimación de la posición de los eventos sonoros realizada por un estimador de la posición de los eventos sonoros falla, por ejemplo, en caso de una dirección equivocada de la estimación de llegada. La Fig. 25 ilustra un escenario de este tipo. En estos casos, independientemente de los parámetros de grado de difusión estimados en el diferente micrófono espacial y recibido como entradas 111 a 11N, se puede ajustar el grado de difusión para el micrófono virtual 103 a 1 (es decir, completamente difuso), ya que no es posible la reproducción espacialmente coherente.

Adicionalmente, se puede considerar la fiabilidad de las estimaciones de DOA en los N micrófonos espaciales. Esto se puede expresar por ejemplo en términos de la varianza del estimador de DOA o SNR. Tal información se puede tener en cuenta mediante el sub-calculador del grado de difusión 850, de modo que el grado de difusión VM 103 se puede aumentar artificialmente en caso de que las estimaciones de DOA no sean fiables. En efecto, como consecuencia, las estimaciones de la posición 205 también serán no fiables.

35 La Fig. 26 ilustra un aparato 991 para generar una señal de salida virtual de acuerdo con una forma de realización. El aparato 991 para generar una señal de salida virtual comprende un aparato 992 para la colocación del micrófono de acuerdo con una de las formas de realización anteriormente descritas que comprende un calculador de la posición del micrófono 993. Además, el aparato para generar una señal de salida virtual comprende un aparato 994 para generar una señal de salida de audio de acuerdo con una de las formas de realización anteriormente descritas. La señal de salida generada por el aparato 994 para generar una señal de salida de audio es la señal de salida virtual vos. El calculador de la posición del micrófono 992 del aparato para la colocación de micrófonos 991 está configurado para calcular la posición de un micrófono como una posición de micrófono calculada cmp. El aparato 994 para generar una señal de salida de audio está configurado para simular un registro de un micrófono virtual en la posición de micrófono calculada por el aparato 992 para la colocación del micrófono. Por esto, el aparato 992 para colocación del micrófono calcula la posición virtual del micrófono virtual para el aparato 994 para generar una señal de salida de audio.

Si bien se han descrito algunos aspectos en el contexto de un aparato, es claro que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa del método o un rasgo de la etapa del método. De modo análogo, los aspectos descritos en el contexto de una etapa del método también representan una descripción de un correspondiente bloque o elemento o rasgo de un correspondiente aparato.

La señal descompuesta de la invención se puede almacenar en un medio de almacenamiento digital o se puede

transmitir en un medio de transmisión tal como un medio de transmisión inalámbrica o un medio de transmisión por cable tal como internet.

5 Dependiendo de ciertos requisitos de implementación, las formas de realización de la invención se pueden implementar en hardware o en software. La implementación se puede realizar usando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control legibles electrónicamente almacenadas en la misma, que cooperan (o que pueden cooperar) con un sistema informático programable de modo que se realiza el método respectivo.

10 Algunas formas de realización de acuerdo con la invención comprenden un portador de datos no transitorios que tienen señales de control legibles electrónicamente, que pueden cooperar con un sistema informático programable de modo que se realiza uno de los métodos descritos en la presente.

15 Generalmente, las formas de realización de la presente invención se pueden implementar como un producto de programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los métodos cuando el producto de programa informático funciona en un ordenador. El código de programa por ejemplo se puede almacenar en un portador legible por máquina.

20 Otras formas de realización comprenden el programa informático para realizar uno de los métodos descritos en la presente, almacenados en un portador legible por máquina.

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

Una forma de realización de los métodos de la invención es, en consecuencia, un transportador de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, registrado en el mismo, el programa informático para realizar uno de los métodos descritos en la presente.

30 Una forma de realización del método de invención adicional es, en consecuencia, una corriente de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en la presente. La corriente de datos o la secuencia de señales por ejemplo se pueden configurar para transferirse mediante una conexión de comunicación de datos, por ejemplo mediante internet.

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

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

45 En algunas formas de realización, un dispositivo lógico programable (por ejemplo un campo de matrices de puertas programables) se puede usar para realizar algunas o todas las funcionalidades de los métodos descritos en la presente. En algunas formas de realización, un campo de matrices de puertas programables puede cooperar con un microprocesador para realizar uno de los métodos descritos en la presente. Generalmente, los métodos se realizan con preferencia con algún aparato de hardware.

50 Las formas de realización anteriormente descritas son solo ilustrativas para los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en la presente serán evidentes para los expertos en la materia. Es la intención, en consecuencia, estar limitado solo por el alcance de las reivindicaciones de patente próximas y no por los detalles específicos presentados a modo de descripción y explicación de las formas de realización en la presente.

Referencias

55 [1] Michael A. Gerzon. Ambisonics in multichannel broadcasting and video. J. Audio Eng. Soc, 33(11): 859–871, 1985.

[2] V. Pulkki, "Directional audio coding in spatial sound reproduction and stereo upmixing", in Proceedings of the AES 28th International Conference, pág. 251-258, Piteå, Suecia, 30 de junio - 2 de julio de 2006.

60 [3] V. Pulkki, "Spatial sound reproduction with directional audio coding", J. Audio Eng. Soc., vol. 55, n.º 6, pág. 503-516, junio de 2007.

[4] C. Faller: "Microphone Front-Ends for Spatial Audio Coders", in Proceedings of the AES 125th International

Convention, San Francisco, octubre de 2008.

- 5 [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, Múnich, Alemania, mayo de 2009.
- [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 Reino Unido, mayo de 2010.
- 10 [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 Reino Unido, mayo de 2010.
- 15 [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.
- 20 [10] A. Walther y C. Faller, "Linear simulation of spaced microphone arrays using b-format recordings", in Audio Engineering Society Convention 128, Londres Reino Unido, mayo de 2010.
- [11] Documento US61/287.596: An Apparatus and a Method for Converting a First Parametric Spatial Audio Signal into a Second Parametric Spatial Audio Signal.
- 25 [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.
- [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, Estados Unidos, abril de 1986.
- 30 [14] R. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation*, vol. 34, n.º 3, pág. 276-280, 1986.
- 35 [15] J. Michael Steele, "Optimal Triangulation of Random Samples in the Plane", *The Annals of Probability*, Vol. 10, N.º 3 (agosto de 1982), pág. 548-553.
- 40 [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", in Audio Engineering Society Convention 124, Ámsterdam, Países Bajos, mayo de 2008.
- 45 [18] M. Kallinger, F. Küch, R. Schultz-Amling, G. Del Galdo, T. Ahonen y V. Pulkki, "Enhanced direction estimation using microphone arrays for directional audio coding," in *Hands-Free Speech Communication and Microphone Arrays*, 2008. HSCMA 2008, mayo de 2008, pág. 45-48.
- 50 [19] R. K. Furness, "Ambisonics - An overview", in AES 8th International Conference, abril de 1990, pág. 181-189.
- [20] Giovanni Del Galdo, Oliver Thiergart, TobiasWeller, y E. A. P. Habets. Generating virtual microphone signals using geometrical information gathered by distributed arrays. In *Third Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA '11)*, Edimburgo, Reino Unido, mayo de 2011.
- 55 [21] Ville Pulkki. Spatial sound reproduction with directional audio coding. *J. Audio Eng. Soc.*, 55(6): 503–516, junio de 2007.

REIVINDICACIONES

1. Un aparato para la ubicación de micrófonos, que comprende:

5 un determinador de densidad de potencia espacial (10; 21) para determinar una densidad de potencia espacial que indica los valores de potencia para una pluralidad de ubicaciones de un ambiente basándose en la información de la fuente de sonido que indica uno o más valores de potencia y uno o más valores de posición de una o más fuentes de sonido ubicadas en el ambiente, y
 10 un estimador de información espacial (20; 22) para estimar una posición de un micrófono basándose en la densidad de potencia espacial,
 en el que el estimador de información espacial (20; 22) comprende un estimador del centro de la escena sonora (41) para estimar una posición de un centro de una escena sonora en el ambiente,
 en el que el estimador de información espacial (20; 22) comprende además un calculador de la posición del micrófono (42; 44) para determinar la posición de un micrófono basándose en la posición del centro de la escena sonora,
 15 en el que el estimador de información espacial (20; 22) comprende un determinador de la orientación (45) para determinar una orientación del micrófono basándose en la densidad de potencia espacial,
 en el que el determinador de densidad de potencia espacial (10; 21) está adaptado para determinar la densidad de potencia espacial aplicando la fórmula

$$\Gamma(x,y,k,n) = \sum_{i=1}^N \text{potencia}_i(k,n) \cdot g(\gamma_i, x - x_{ESSi}, y - y_{ESSi}, k, n),$$

para calcular los valores de potencia $\Gamma(x,y,k,n)$ para la pluralidad de ubicaciones en el ambiente durante un segmento de tiempo-frecuencia (k, n), en el que k indica el índice de frecuencia y n indica el índice de tiempo, en el que N indica un número de las fuentes de sonido, en el que x, y, indican coordenadas de una de la pluralidad de ubicaciones, en el que la potencia $i(k,n)$ indica el valor de potencia a una fuente de sonido i-ésima para el segmento de tiempo-frecuencia (k, n), en el que x_{ESSi} , y_{ESSi} indican coordenadas de la fuente de sonido i-ésima, en el que γ_i es un valor escalar y en el que g es una función que depende de x, y, x_{ESSi} , y_{ESSi} , k, n y γ_i , cuando el ambiente es un ambiente bidimensional, o
 25 en el que el determinador de densidad de potencia espacial (10; 21) está adaptado para determinar la densidad de potencia espacial aplicando la fórmula

$$\Gamma(x,y,z,k,n) = \sum_{i=1}^N \text{potencia}_i(k,n) \cdot g(\gamma_i, x - x_{ESSi}, y - y_{ESSi}, z - z_{ESSi}, k, n),$$

35 para calcular los valores de potencia $\Gamma(x,y,z,k,n)$ para la pluralidad de ubicaciones del ambiente para un segmento de tiempo-frecuencia (k, n), en el que k indica el índice de frecuencia y n indica el índice de tiempo, en el que N indica un número de las fuentes de sonido, en el que x, y, z indican las coordenadas de una de la pluralidad de ubicaciones, en el que la potencia $i(k,n)$ indica el valor de potencia en una fuente de sonido i-ésima para el segmento de tiempo-frecuencia (k, n), en el que x_{ESSi} , y_{ESSi} , z_{ESSi} indican las coordenadas de la fuente de sonido i-ésima, en el que γ_i es un valor escalar y en el que g es una función que depende de x, y, z, x_{ESSi} , y_{ESSi} , z_{ESSi} , k, n y γ_i , cuando el ambiente es un ambiente tridimensional.

2. Un aparato de acuerdo con la reivindicación 1, en el que el determinador de la orientación (45) está adaptado para determinar la orientación del micrófono de manera que el micrófono está orientado hacia el centro de la escena sonora.

3. Un aparato de acuerdo con la reivindicación 1 o 2, en el que el calculador de la posición del micrófono (42; 44) está adaptado para calcular la posición del micrófono, en el que el micrófono es un micrófono espacial virtual.

4. Un aparato de acuerdo con una de las reivindicaciones anteriores, en el que el estimador del centro de la escena sonora (41) está adaptado para calcular un centro de gravedad de la densidad de potencia espacial para estimar el centro de la escena sonora.

5. Un aparato de acuerdo con una de las reivindicaciones 1 a 3,
 55 en el que el estimador de centro de escena sonora (41) está configurado para determinar un perfil de retardo de potencia basándose en la densidad de potencia espacial y para determinar un retardo de acuerdo con un valor cuadrático medio basándose en el perfil de retardo de potencia para cada una de una pluralidad de ubicaciones en el ambiente, y
 en el que el estimador de centro de escena sonora (41) está configurado para determinar la ubicación de la

pluralidad de ubicaciones como el centro de la escena sonora, que tiene el retardo de acuerdo con un valor cuadrático medio mínimo de los retardos de acuerdo con unos valores cuadráticos medios de la pluralidad de ubicaciones.

5 6. Un aparato de acuerdo con una de las reivindicaciones anteriores,
 en el que el calculador de la posición del micrófono (42; 44) está adaptado para determinar una línea de ancho más
 amplio de una pluralidad de líneas a través del centro de la escena sonora en el ambiente,
 en el que cada una de la pluralidad de líneas a través del centro de la escena sonora está asociada con un ancho de
 energía, y en el que la línea de ancho más amplio se define como la línea de la pluralidad de líneas a través del
 10 centro de la escena sonora que tiene el ancho de energía más grande,
 en el que el calculador de la posición del micrófono (42; 44) está adaptado para determinar la posición del micrófono
 de manera que una segunda línea, que atraviesa el centro de la escena sonora y la posición del micrófono es
 ortogonal a la línea de ancho más amplio.

15 7. Un aparato de acuerdo con la reivindicación 6, en el que el ancho de energía de una línea considerada de la
 pluralidad de líneas indica un ancho más grande de un segmento de la línea considerada, de manera que el primer
 punto del segmento que limita el segmento, y de manera que un segundo punto diferente del segmento que limita el
 segmento, tienen ambos un valor de potencia indicado por la densidad de potencia espacial, que es mayor que o
 igual a un valor de potencia predefinido.

20 8. Un aparato de acuerdo con una de las reivindicaciones 1 a 5, en el que el calculador de la posición del micrófono
 (42; 44) está configurado para aplicar una descomposición de valor singular a una matriz que tiene una pluralidad de
 columnas,
 en el que las columnas de la matriz indican posiciones de ubicaciones en el ambiente con relación al centro de la
 25 escena sonora, y
 en el que las columnas de la matriz indican únicamente las posiciones de ubicaciones que tienen valores de
 potencia indicados por la densidad de potencia espacial que son mayores que un valor umbral predefinido, o las
 columnas de la matriz únicamente indican las posiciones de las ubicaciones que tienen valores de potencia
 indicados por la densidad de potencia espacial que son mayores que o iguales a un valor umbral predefinido.

30 9. Un aparato (991) para generar una señal de salida virtual, que comprende:
 un aparato para colocación de micrófonos (992) de acuerdo con una de las reivindicaciones 1 a 8, en el que el
 calculador de la posición del micrófono (993) del aparato para la colocación de micrófonos está configurado para
 35 calcular la posición de un micrófono como una posición de micrófono calculada, y
 un aparato (994) para generar una señal de salida de audio como la señal de salida virtual para simular una
 grabación de un micrófono virtual en la posición de micrófono calculada, en el que el aparato para generar una
 señal de salida de audio comprende:

40 un estimador de posición de eventos sonoros (110) para estimar una posición de fuente de sonido que indica
 una posición de una fuente de sonido en el ambiente, emitiendo la fuente de sonido una onda de sonido, en
 el que el estimador de posición de eventos sonoros (110) está adaptado para estimar la posición de la fuente
 de sonido basándose en una primera información de dirección proporcionada mediante un primer micrófono
 espacial real que está ubicado en una primera posición de micrófono real en el ambiente,
 45 y basándose en una segunda información de dirección proporcionada mediante un segundo micrófono
 espacial real que está ubicado en una segunda posición de micrófono real en el ambiente; y
 un módulo de cálculo de información (120) para generar la señal de salida de audio basándose en una
 primera señal de entrada de audio grabada que se graba mediante el primer micrófono espacial real,
 basándose en la primera posición de micrófono real y basándose en la posición de micrófono calculada.

50 10. Un método para colocación de micrófonos, que comprende:
 determinar una densidad de potencia espacial que indica valores de potencia para una pluralidad de ubicaciones
 de un ambiente basándose en información de la fuente de sonido que indica uno o más valores de potencia y
 55 uno o más valores de posición de una o más fuentes de sonido ubicadas en el entorno, y
 estimar una posición de un micrófono basándose en la densidad de potencia espacial, y
 determinar una orientación del micrófono,
 en el que estimar la posición del micrófono basándose en la densidad de potencia espacial se realiza estimando
 una posición de un centro de una escena sonora en el ambiente, y determinando la posición del micrófono
 60 basándose en la posición del centro de la escena sonora,
 en el que el estimador de información espacial (20; 22) comprende un determinador de orientación (45) para
 determinar una orientación del micrófono, en el que el determinador de orientación (45) está adaptado para
 determinar la orientación del micrófono basándose en la densidad de potencia espacial,
 en el que el determinador de densidad de potencia espacial (10; 21) está adaptado para determinar la densidad

de potencia espacial aplicando la fórmula

$$\Gamma(x,y,k,n) = \sum_{i=1}^N \text{potencia}_i(k,n) \cdot g(\gamma_i, x - x_{ESSi}, y - y_{ESSi}, k, n),$$

- 5 para calcular los valores de potencia $\Gamma(x,y,k,n)$ para la pluralidad de ubicaciones en el ambiente durante un segmento de tiempo-frecuencia (k, n), en el que k indica el índice de frecuencia y n indica el índice de tiempo, en el que N indica un número de las fuentes de sonido, en el que x, y, indican coordenadas de una de la pluralidad de ubicaciones, en el que la potencia $i(k,n)$ indica el valor de potencia a una fuente de sonido i-ésima para el segmento de tiempo-frecuencia (k, n), en el que x_{ESSi}, y_{ESSi} indican coordenadas de la fuente de sonido i-ésima,
- 10 en el que γ_i es un valor escalar y en el que g es una función que depende de x, y, x_{ESSi}, y_{ESSi}, k, n y γ_i , cuando el ambiente es un ambiente bidimensional, o en el que el determinante de densidad de potencia espacial (10; 21) está adaptado para determinar la densidad de potencia espacial aplicando la fórmula

$$\Gamma(x,y,z,k,n) = \sum_{i=1}^N \text{potencia}_i(k,n) \cdot g(\gamma_i, x - x_{ESSi}, y - y_{ESSi}, z - z_{ESSi}, k, n),$$

- 15 para calcular los valores de potencia $\Gamma(x,y,z,k,n)$ para la pluralidad de ubicaciones del ambiente para un segmento de tiempo-frecuencia (k, n), en el que k indica el índice de frecuencia y n indica el índice de tiempo, en el que N indica un número de las fuentes de sonido, en el que x, y, z indican las coordenadas de una de la pluralidad de ubicaciones, en el que la potencia $i(k,n)$ indica el valor de potencia en una fuente de sonido i-ésima para el segmento de tiempo-frecuencia (k, n), en el que $x_{ESSi}, y_{ESSi}, z_{ESSi}$ indican las coordenadas de la fuente de sonido i-ésima, en el que γ_i es un valor escalar y en el que g es una función que depende de x, y, z, $x_{ESSi}, y_{ESSi}, z_{ESSi}, k, n$ y γ_i , cuando el ambiente es un ambiente tridimensional.
- 20
- 25 11. Producto de programa informático adecuado para implementar el método de acuerdo con la reivindicación 10 cuando el programa informático se ejecuta en un ordenador o procesador.

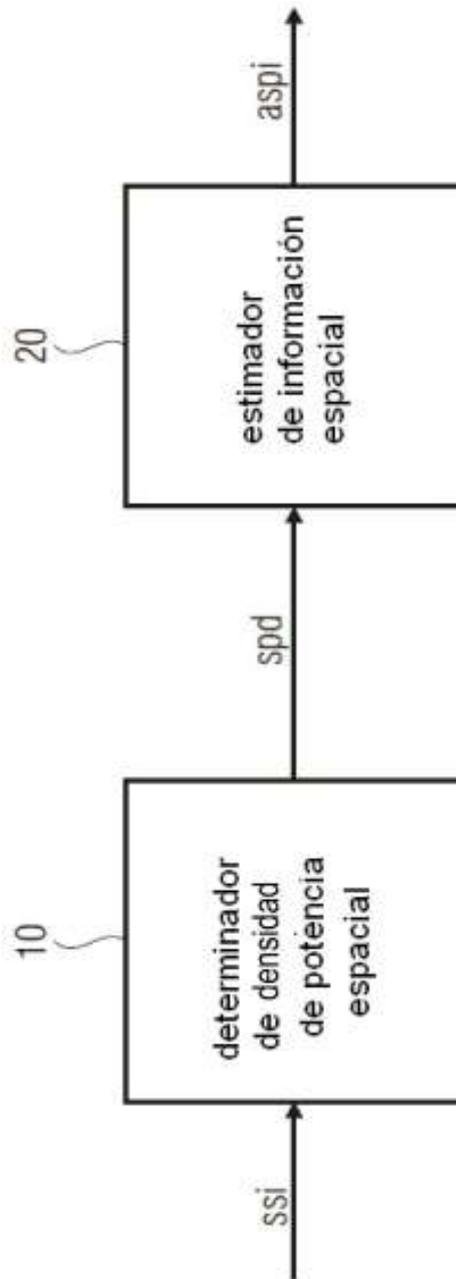


FIG 1

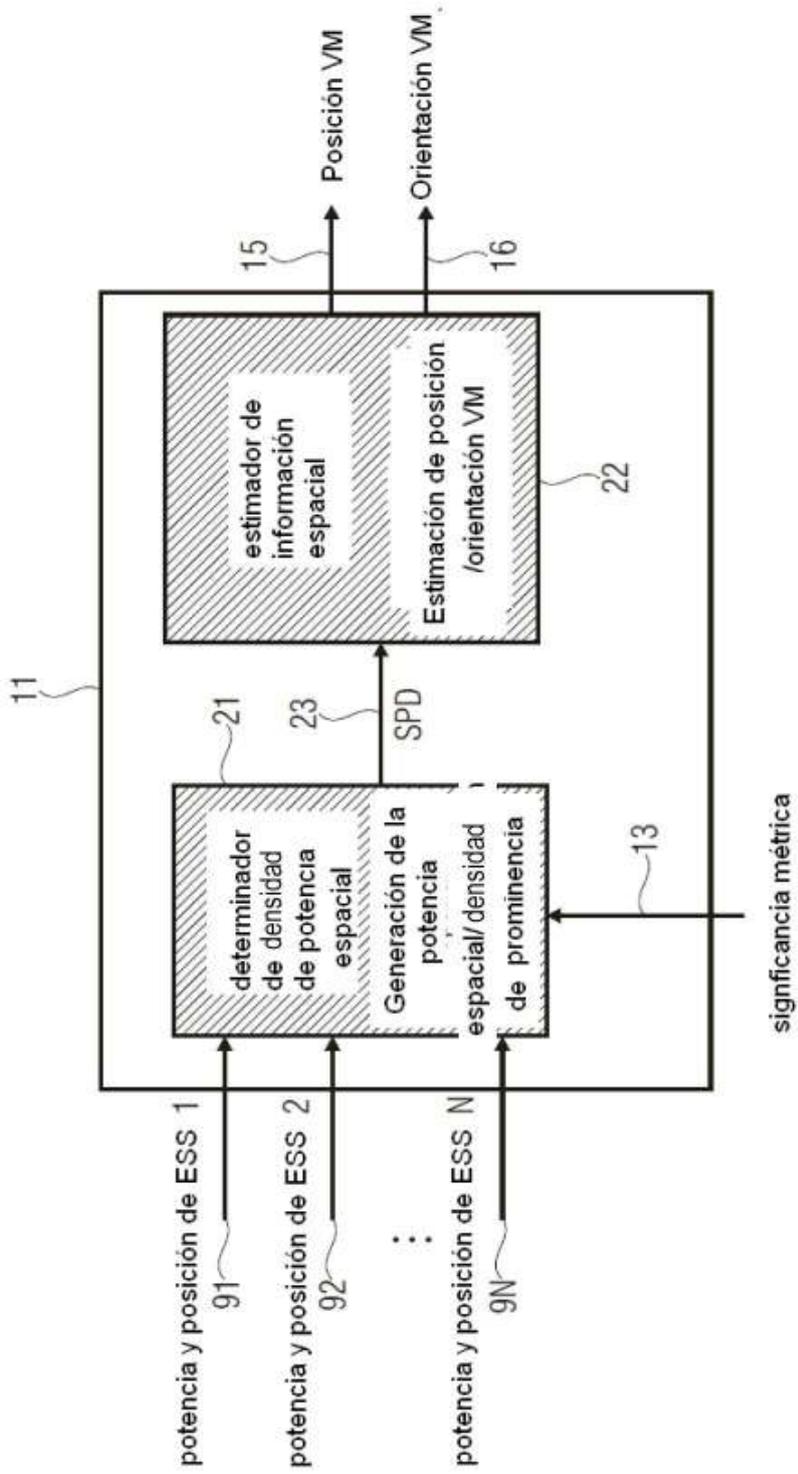


FIG 2

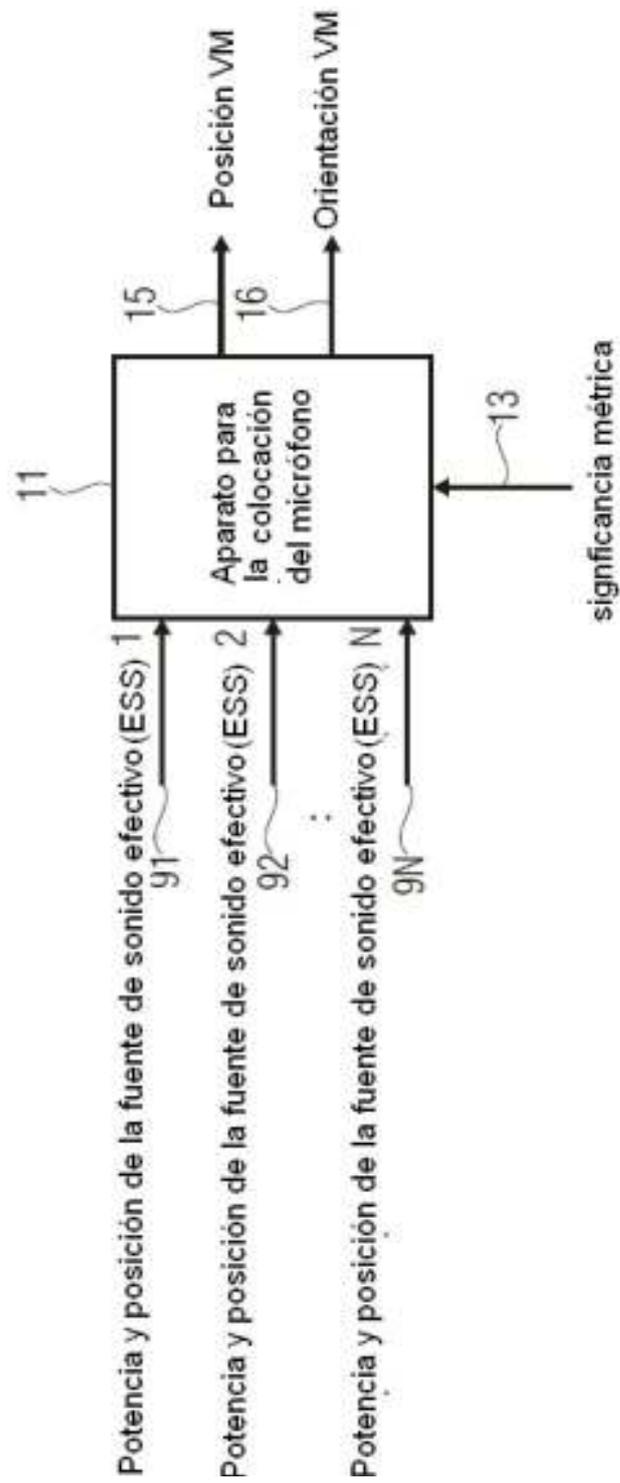


FIG 3

d (Micrófono virtual omnidireccional

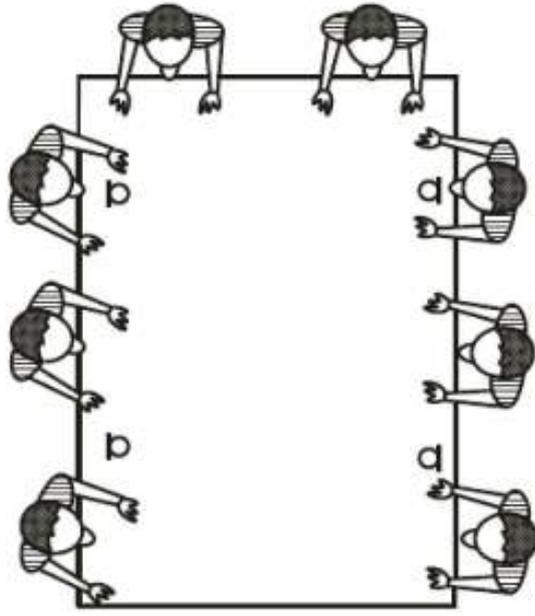


FIG 4A

d Micrófono virtual omnidireccional

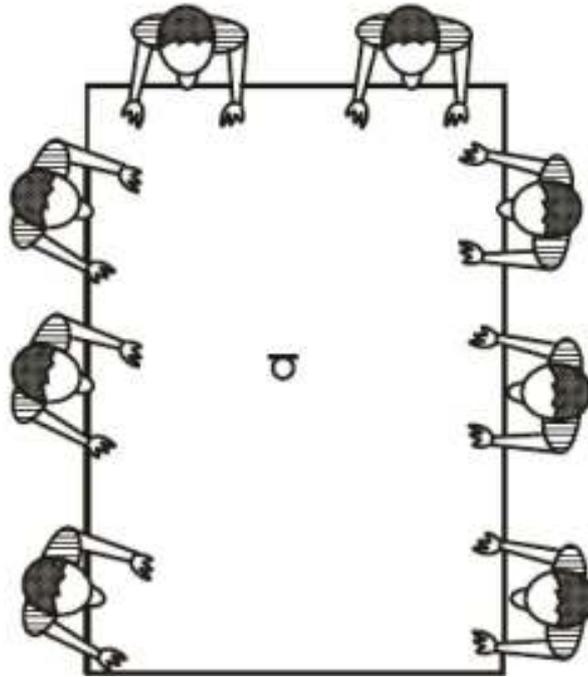


FIG 4B

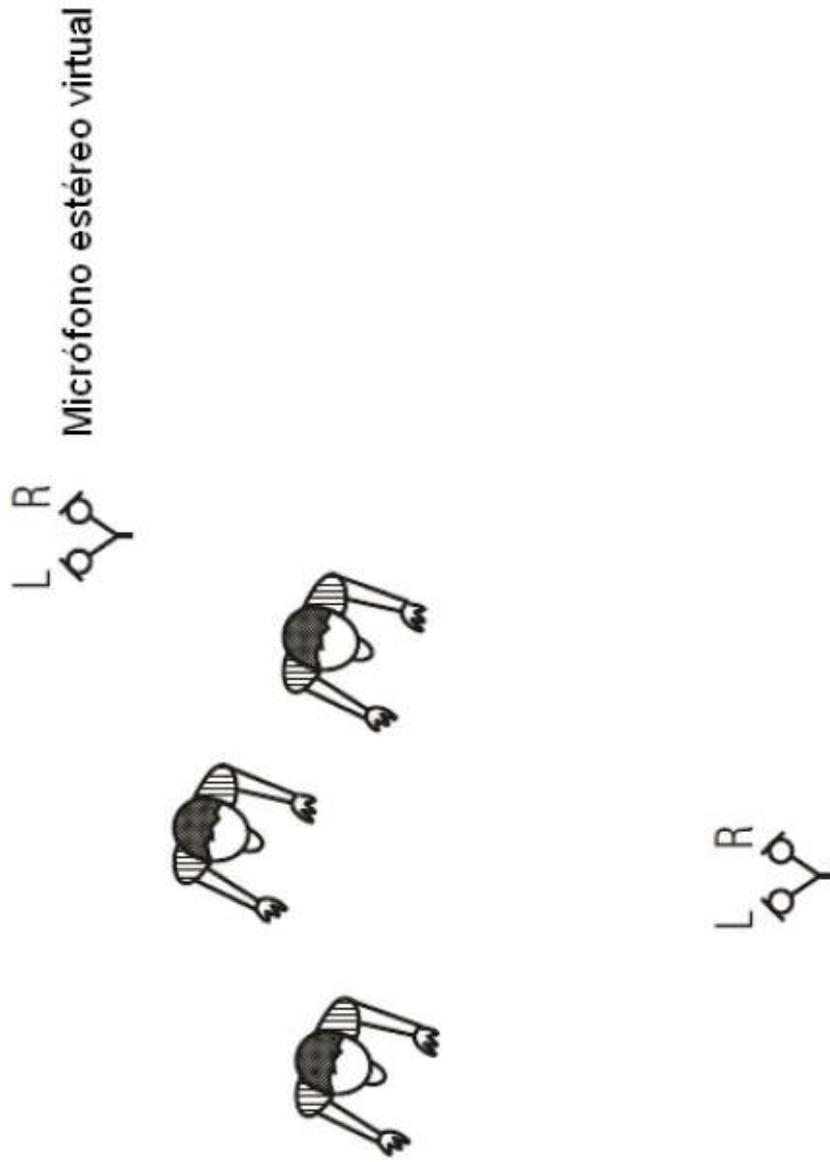


FIG 4C

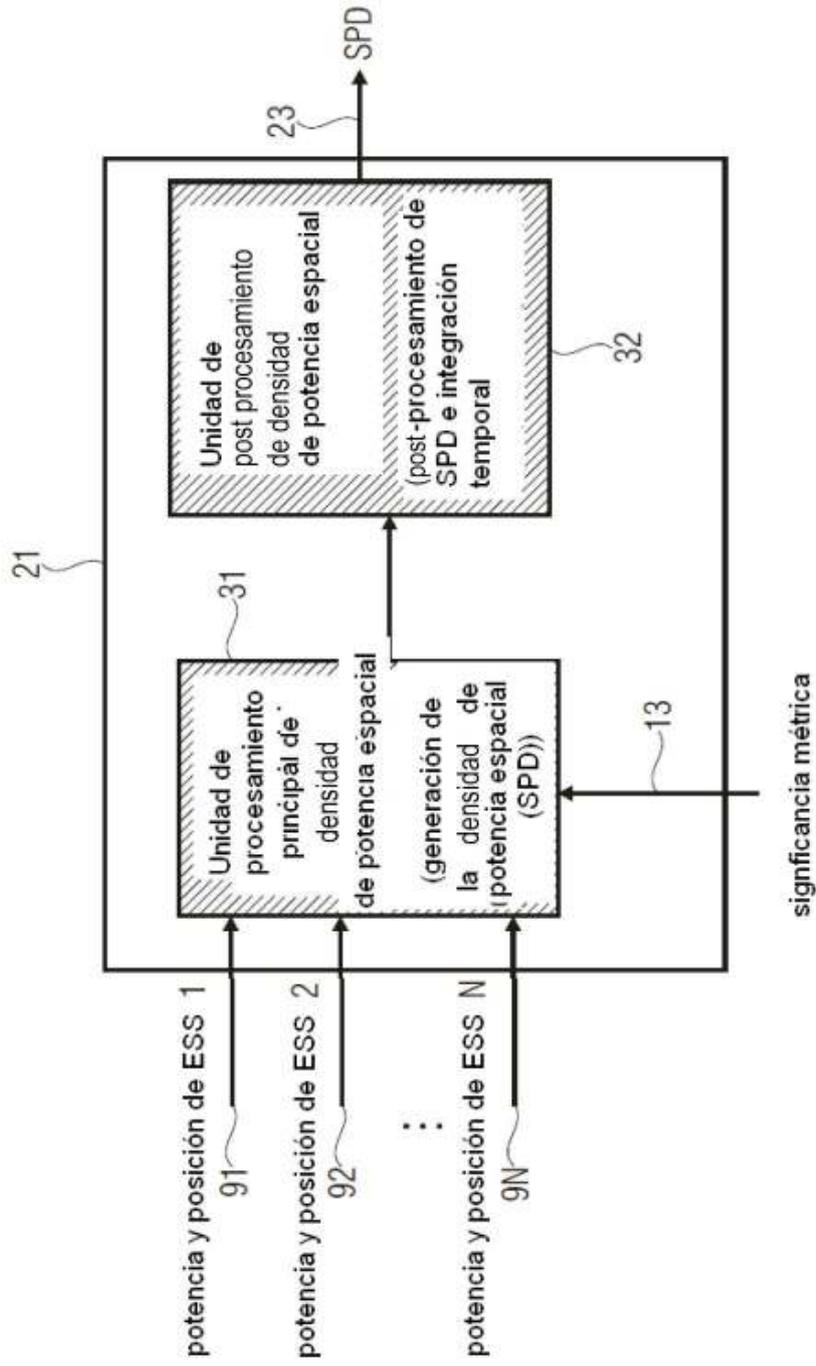


FIG 5

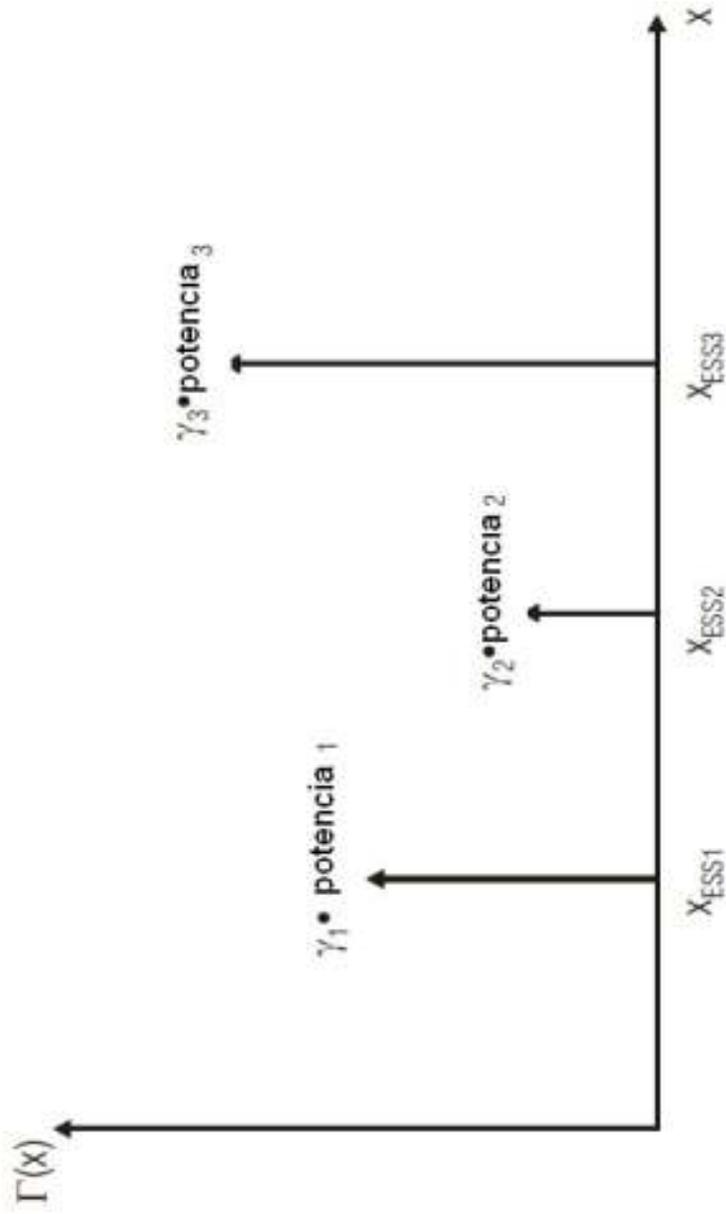


FIG 6A

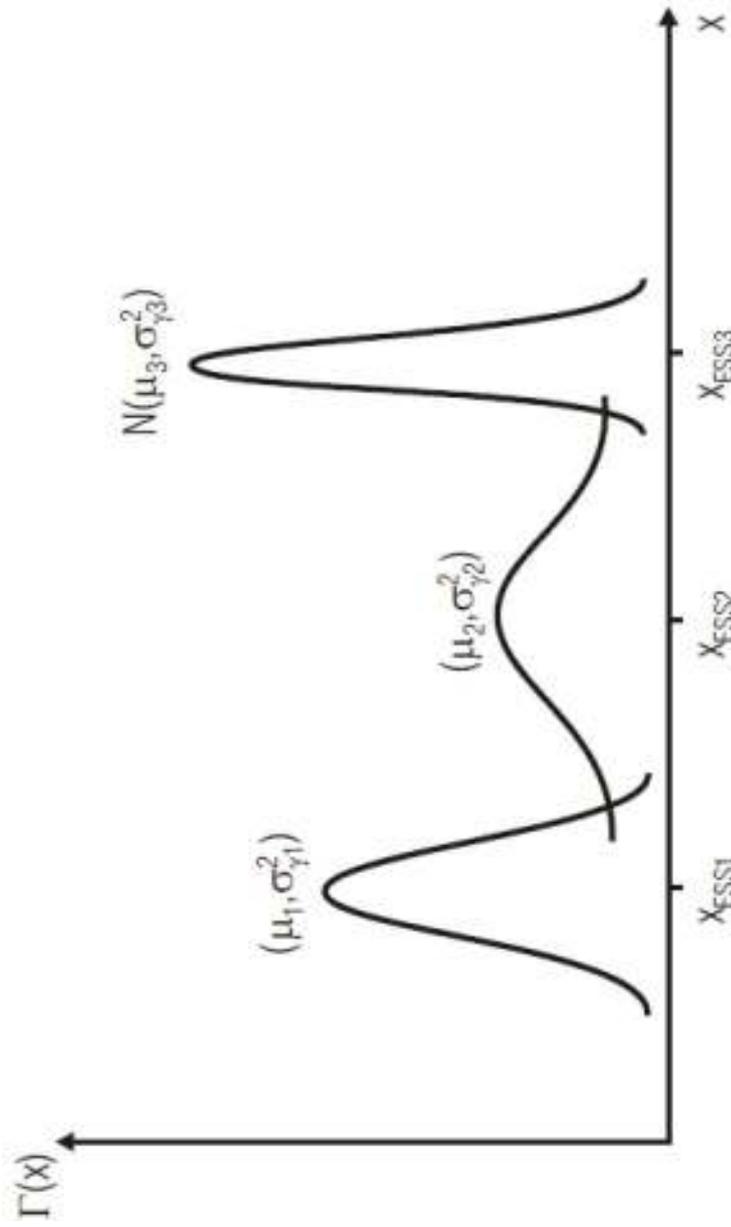


FIG 6B

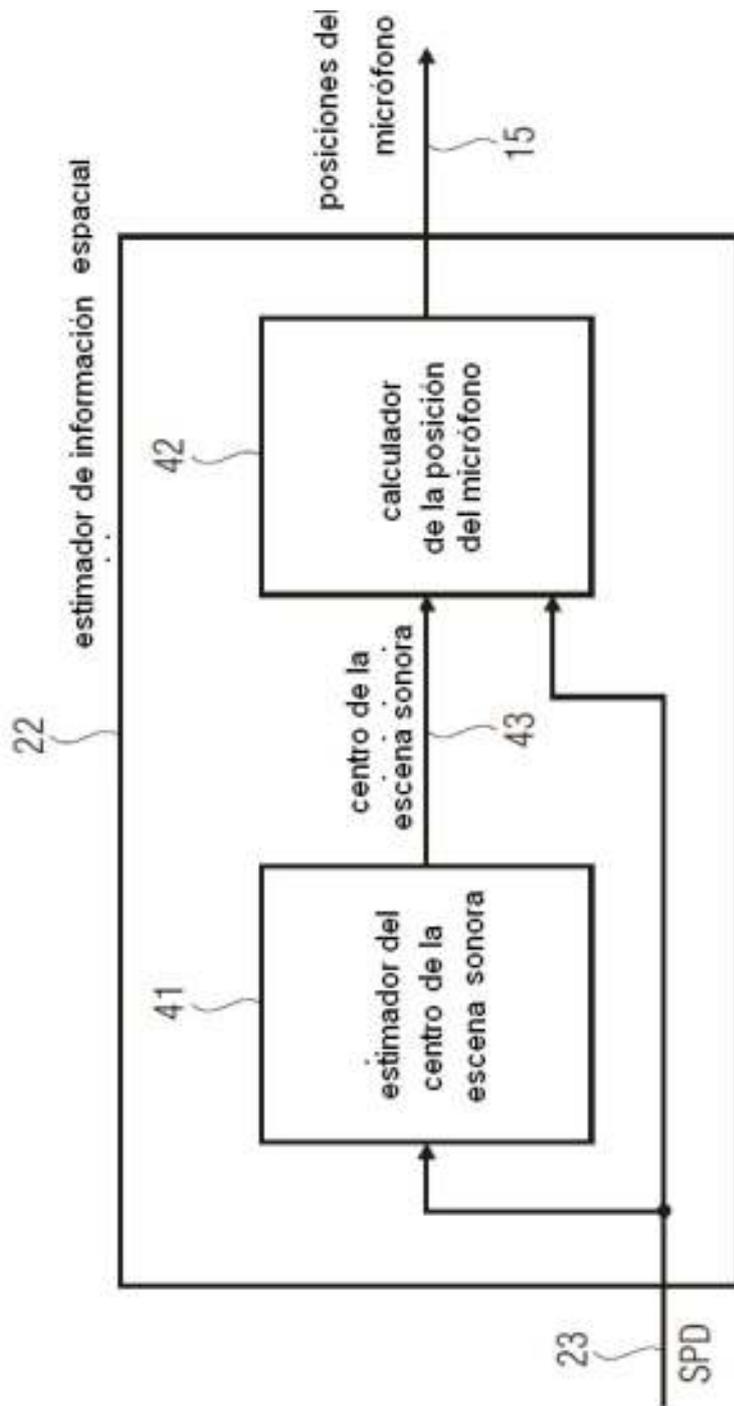


FIG 7

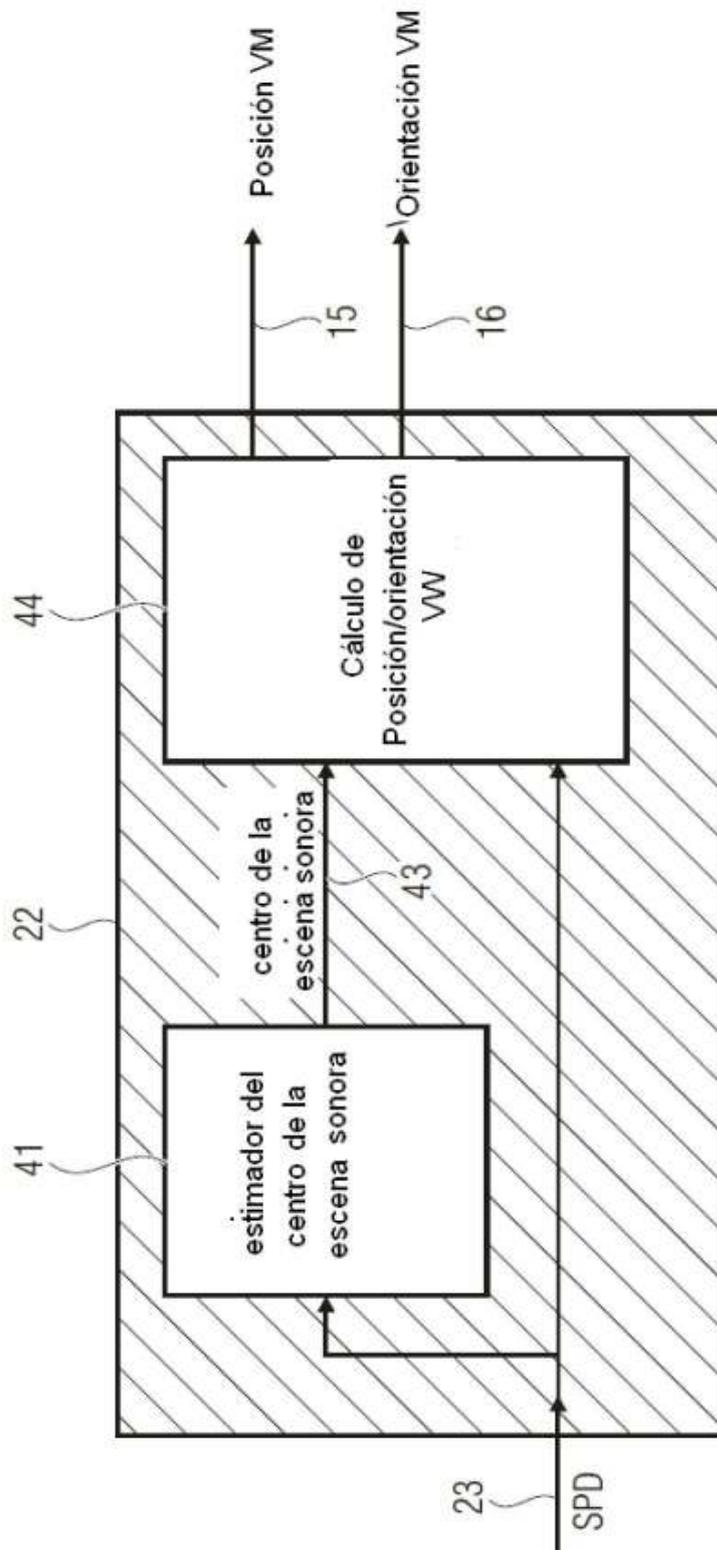


FIG 8

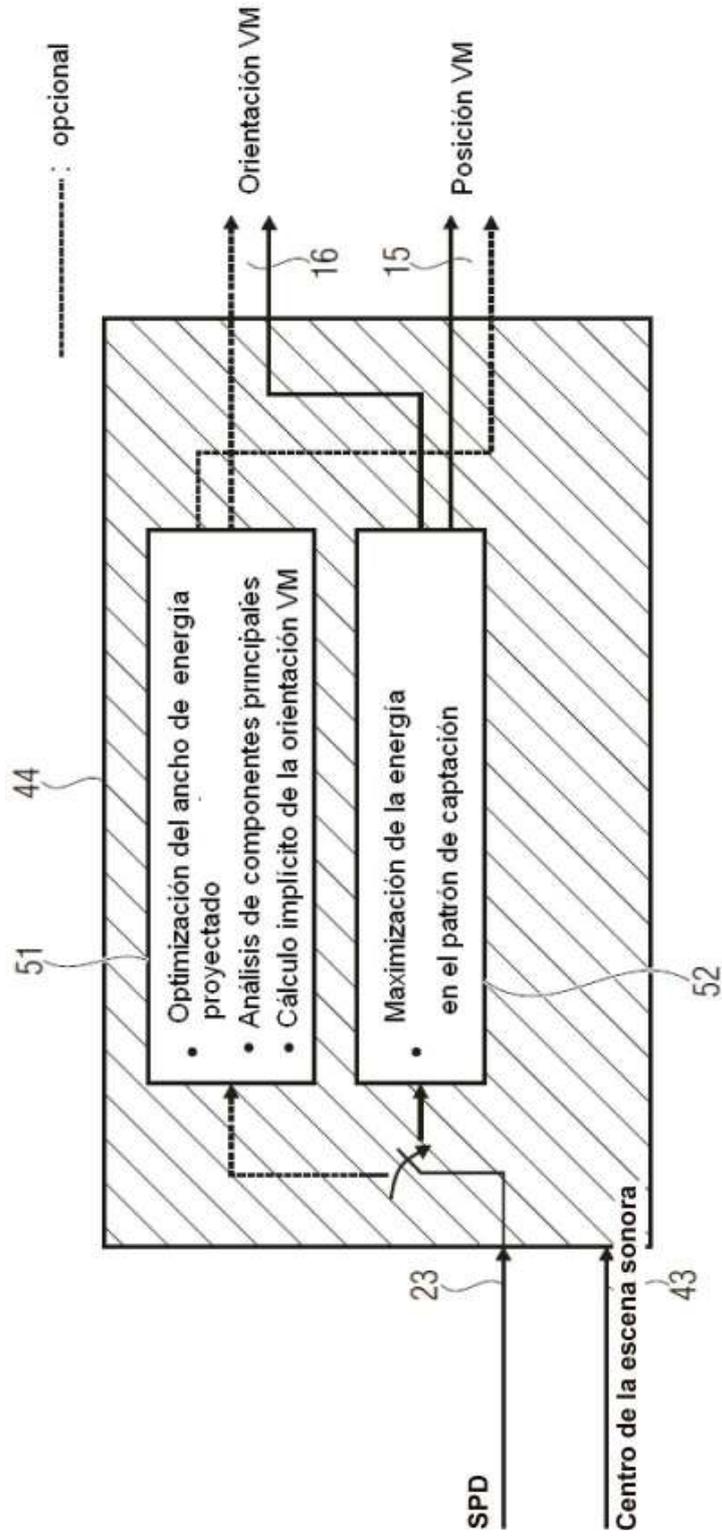


FIG 9

COG: centro de gravedad de la escena sonora

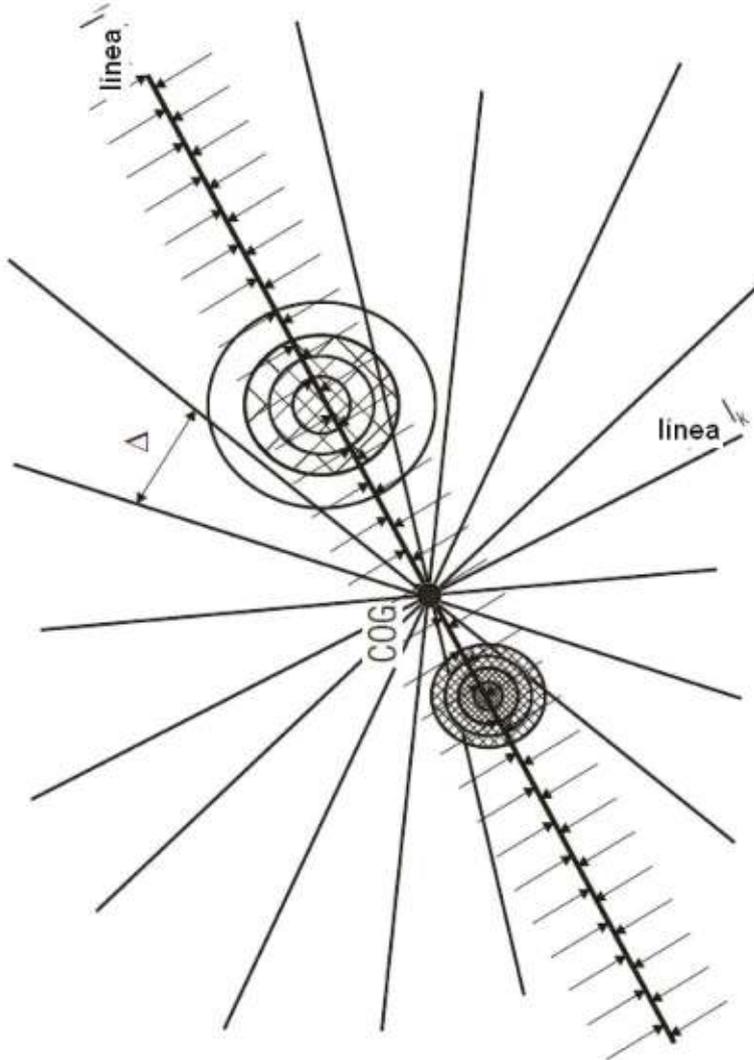


FIG 10A

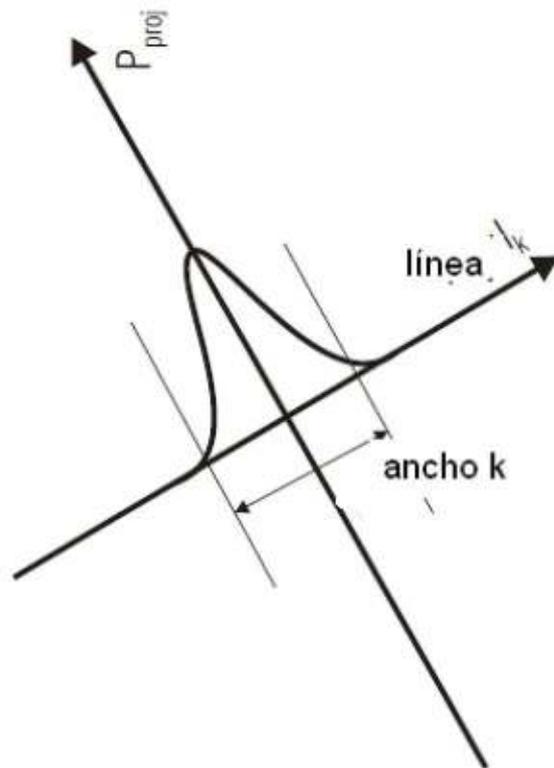
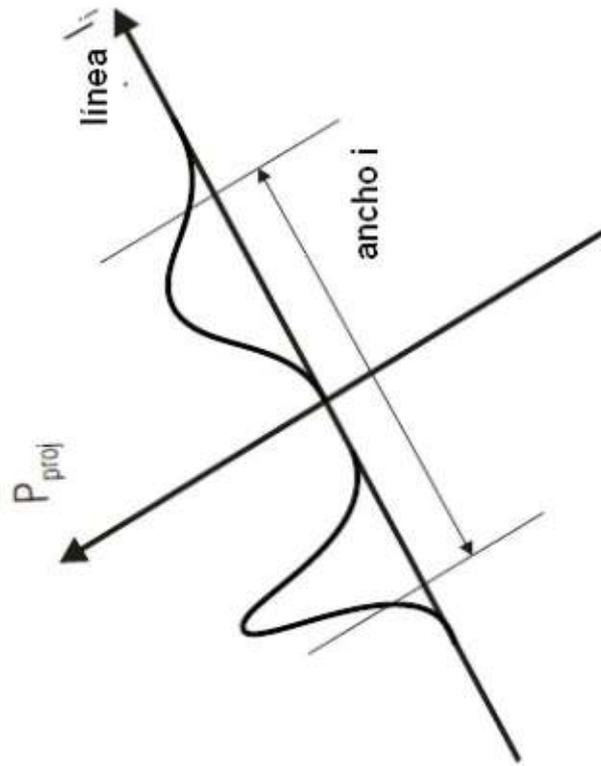


FIG 10B

α : Ángulo de apertura del micrófono virtual
 $d_{\text{COG-VM}}$: (Distancia entre COG y VM

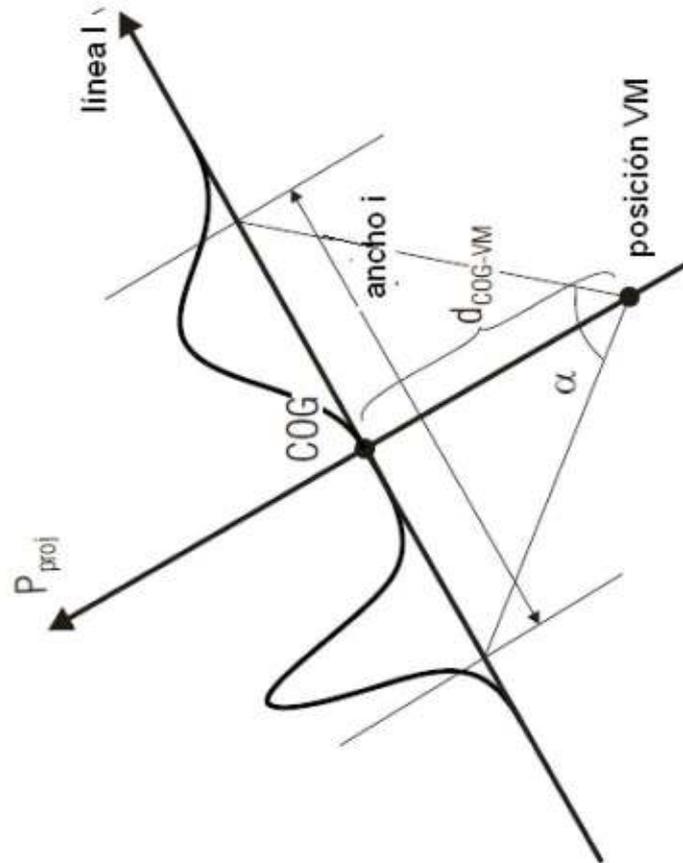


FIG 10C

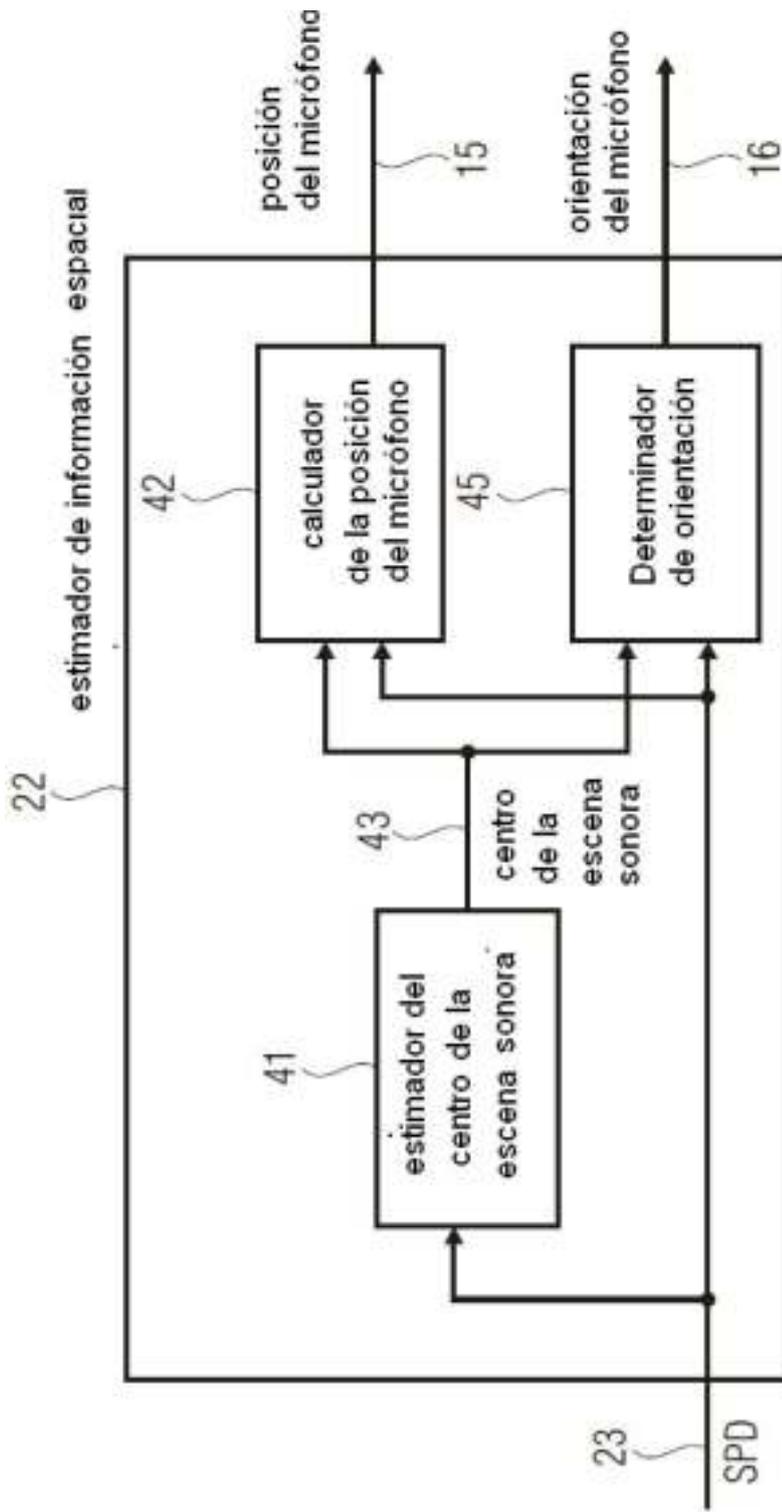


FIG 11

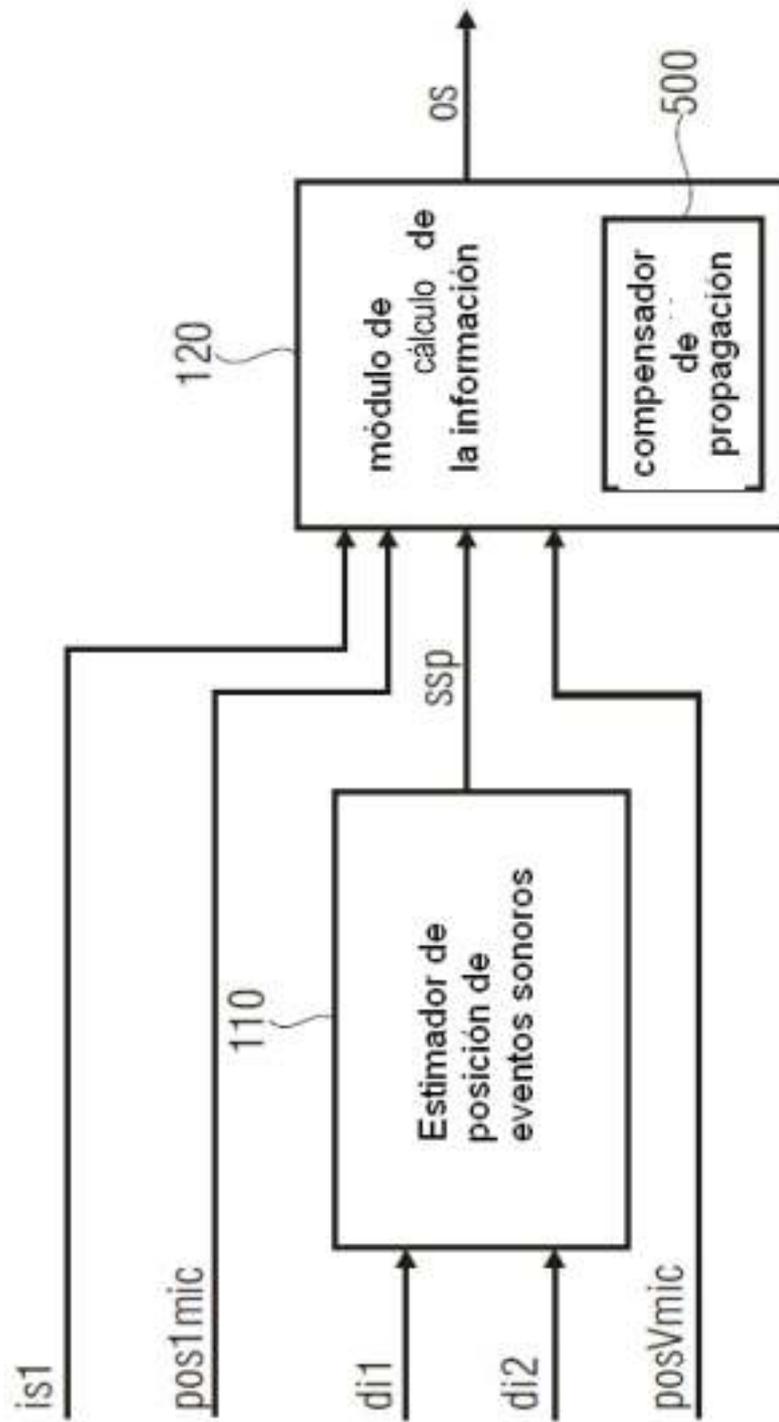


FIG 12

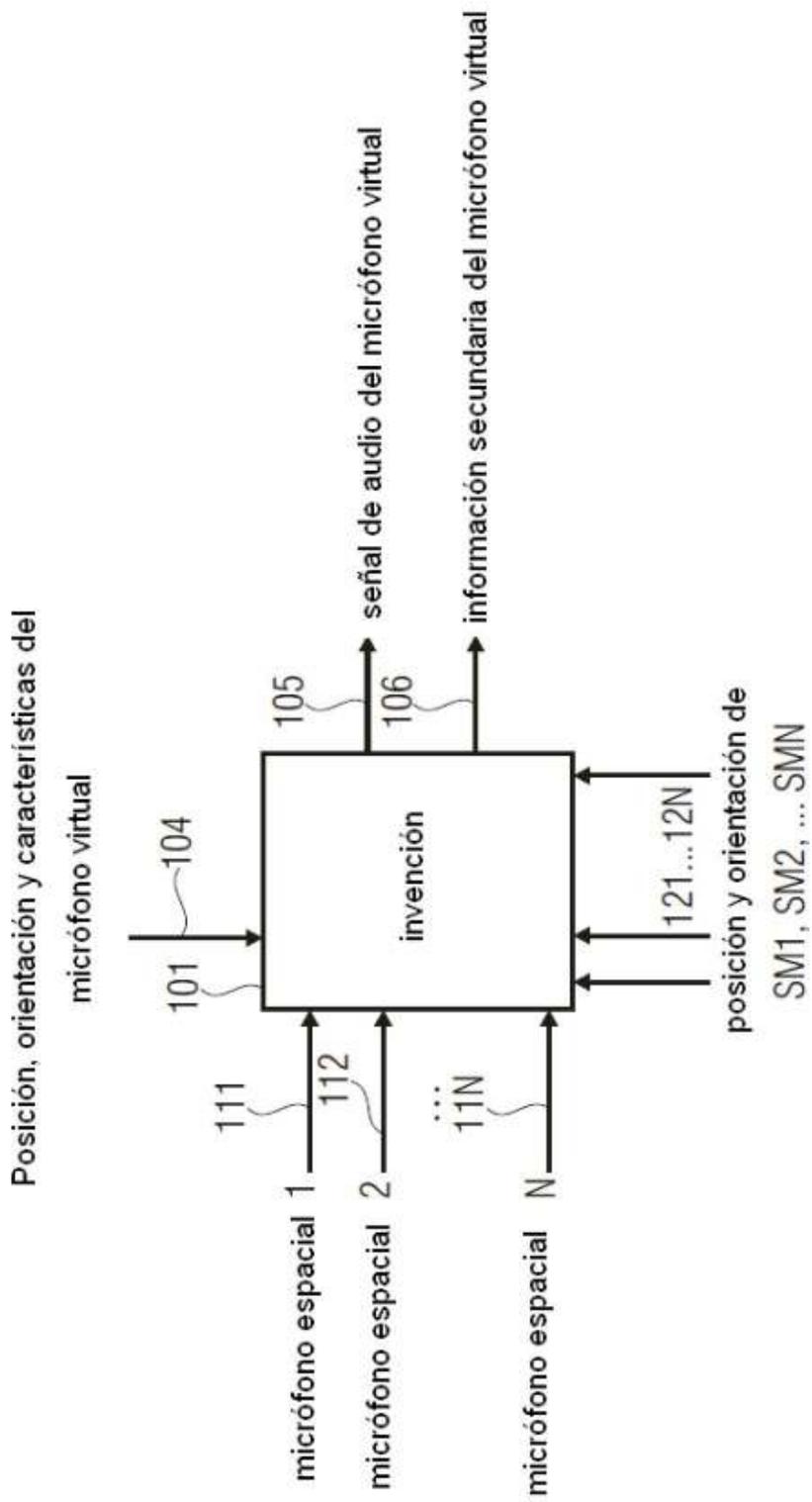


FIG 13

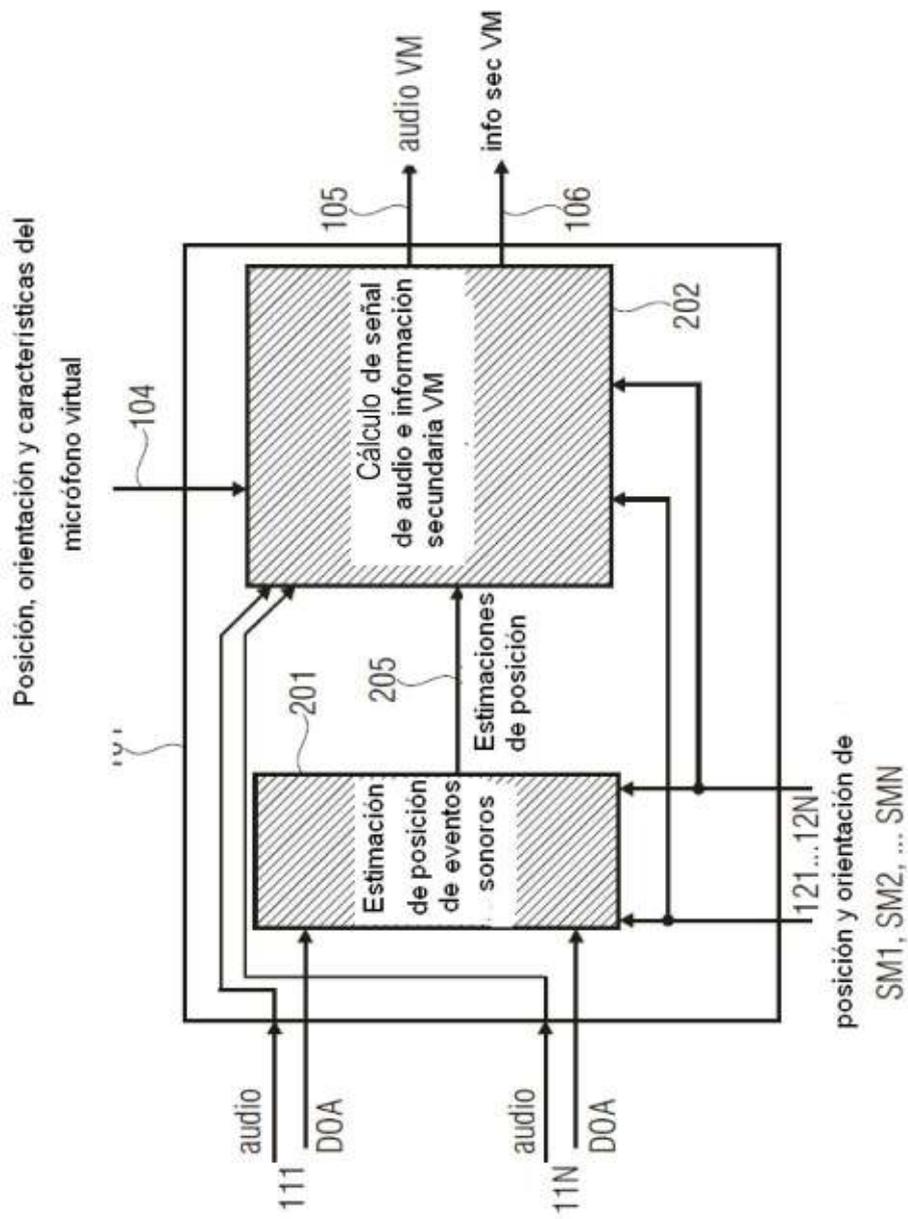


FIG 14

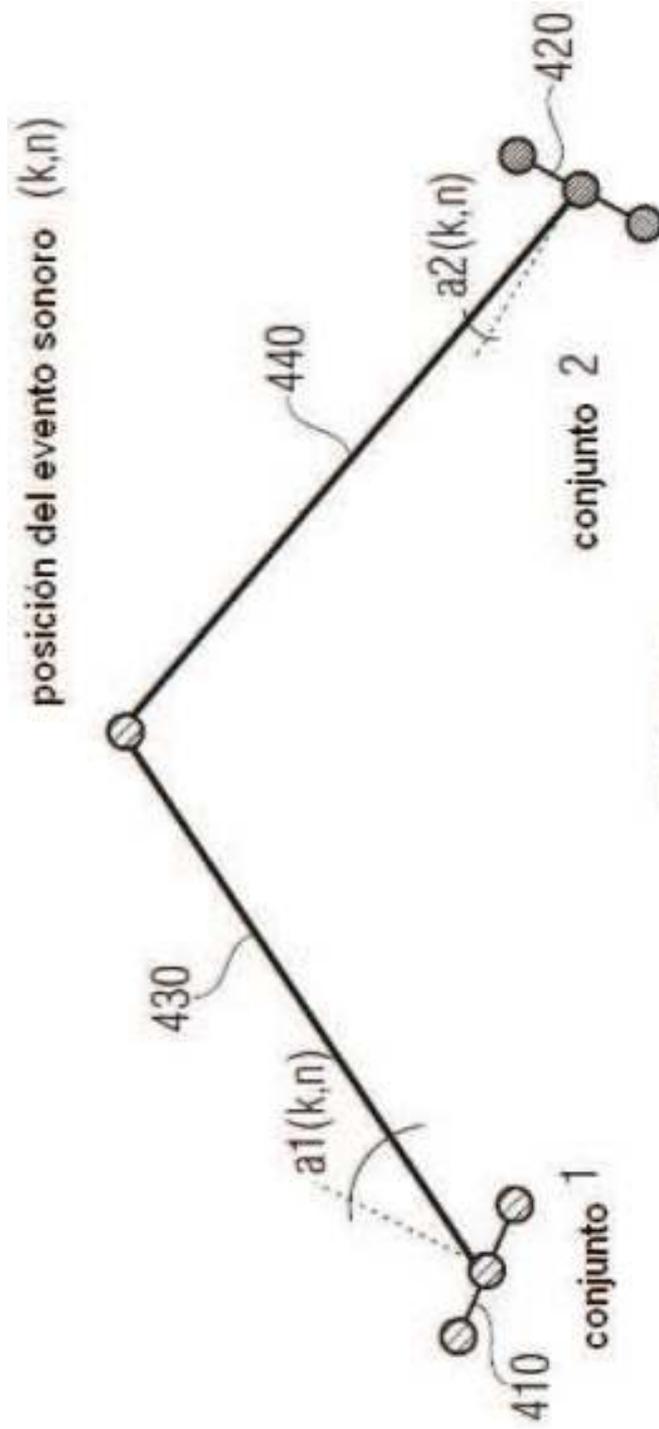


FIG 15

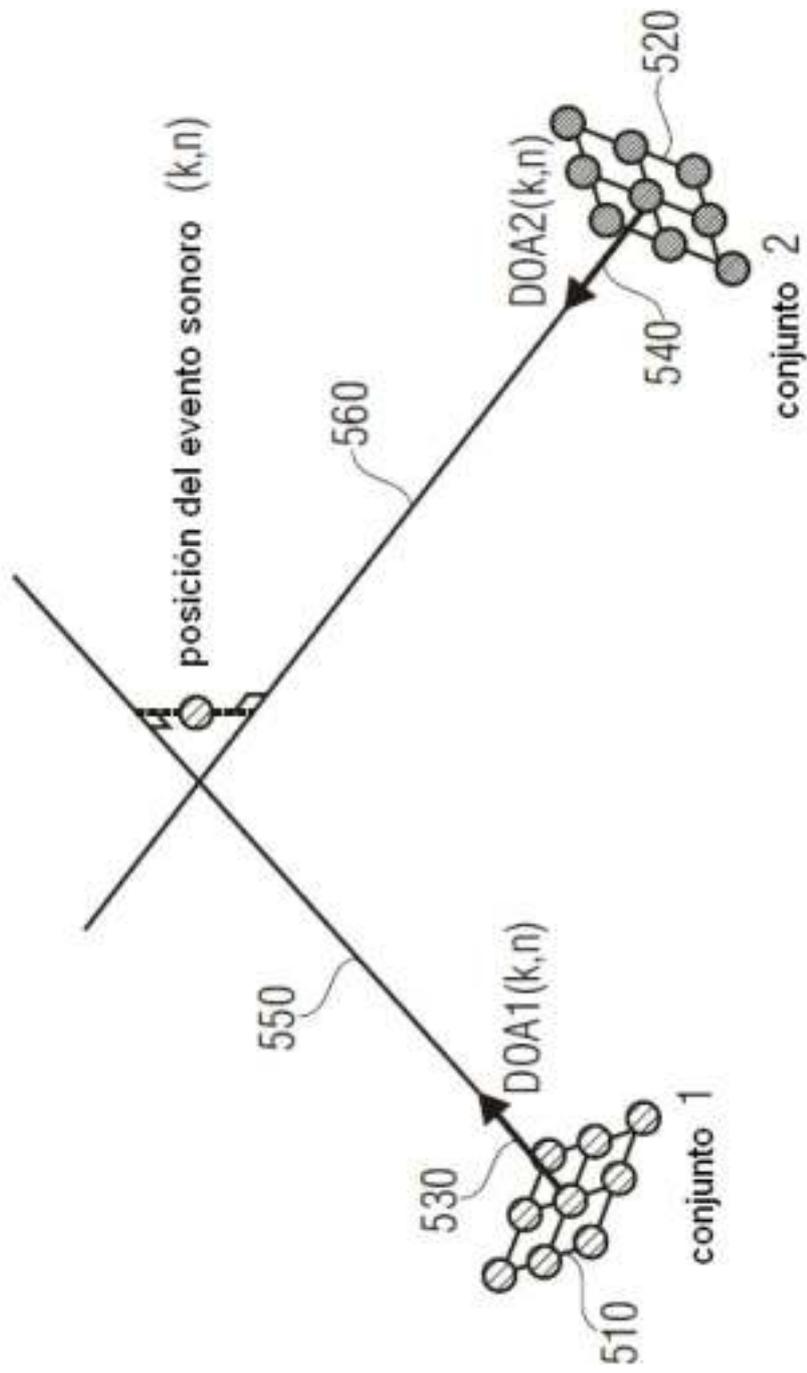


FIG 16

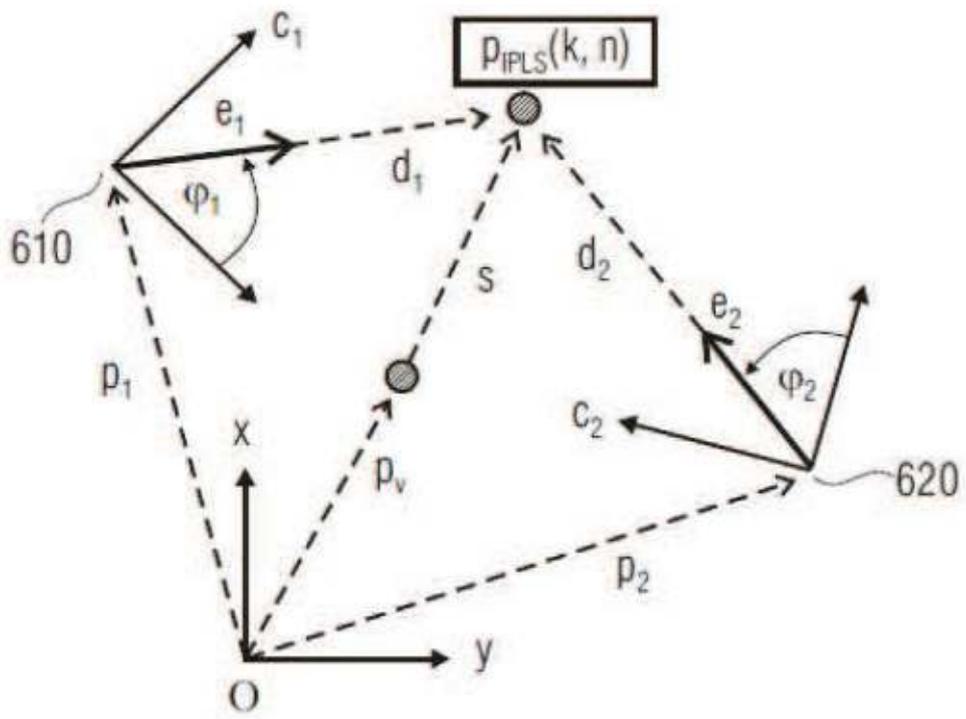


FIG 17

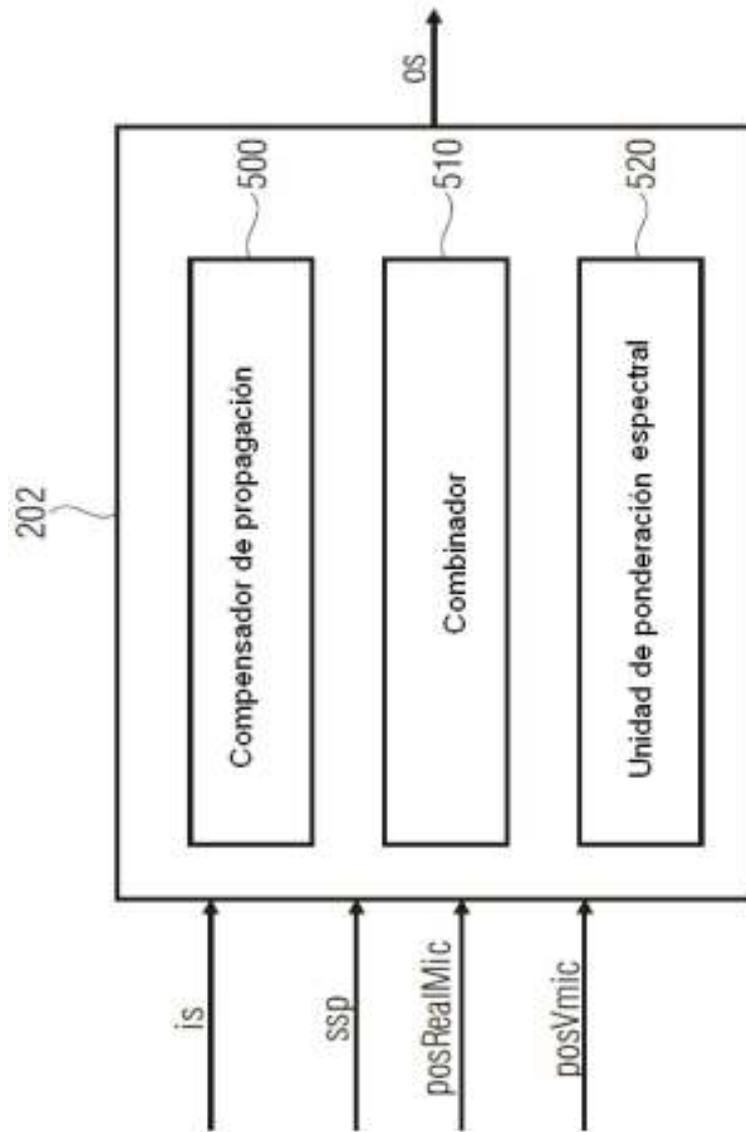


FIG 18

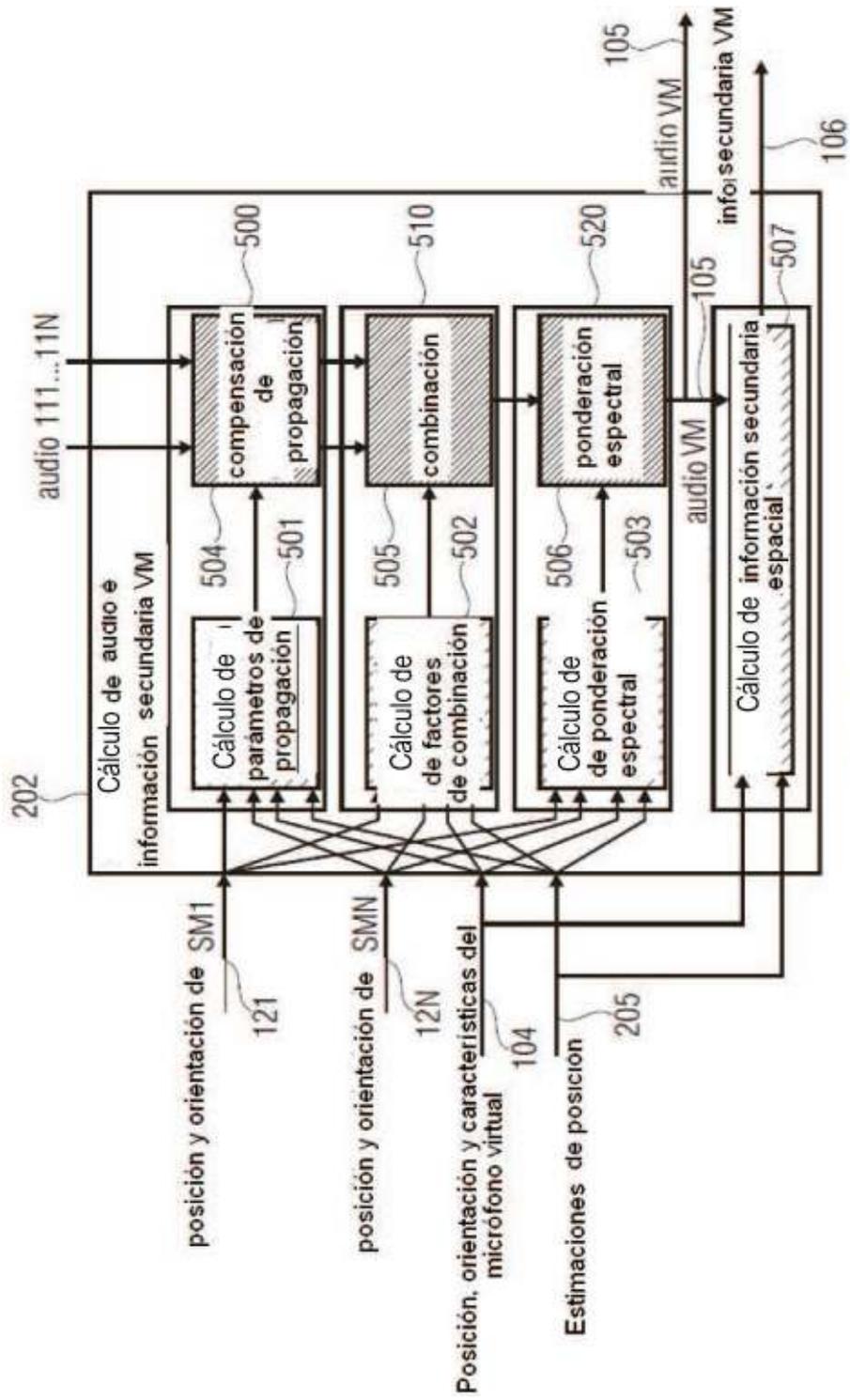


FIG 19

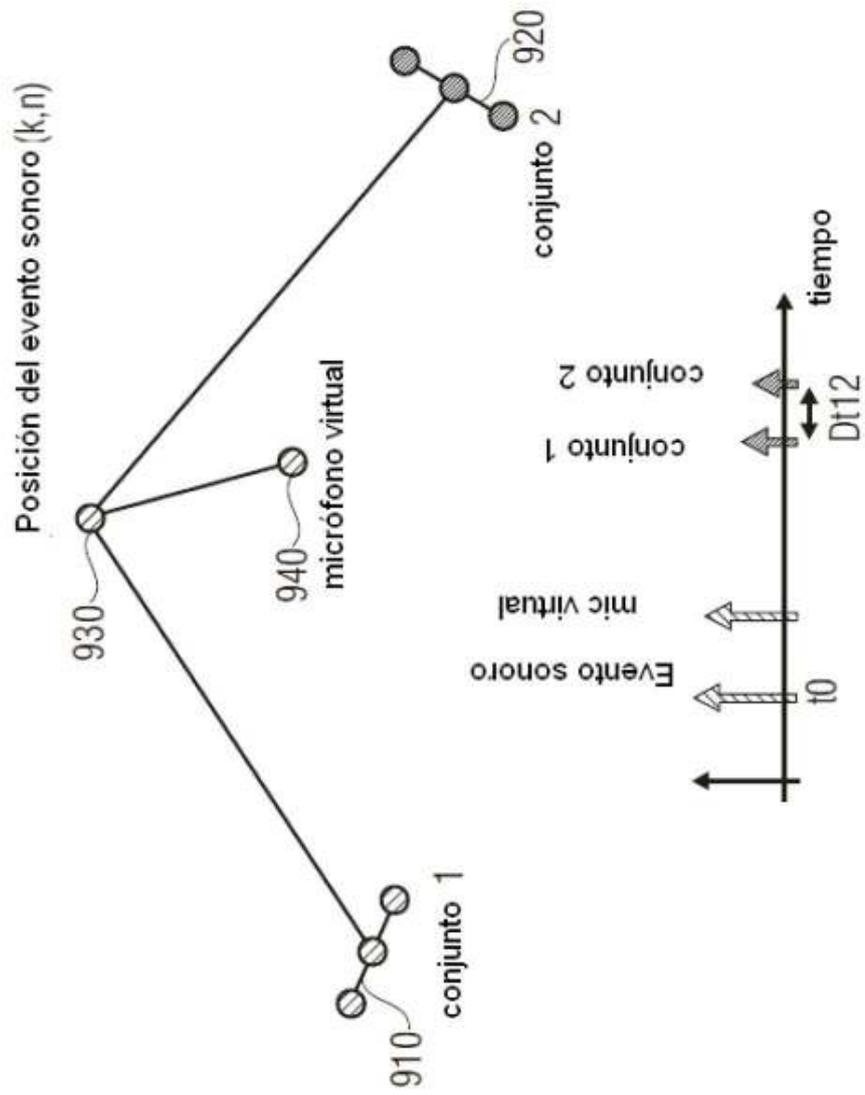


FIG 20

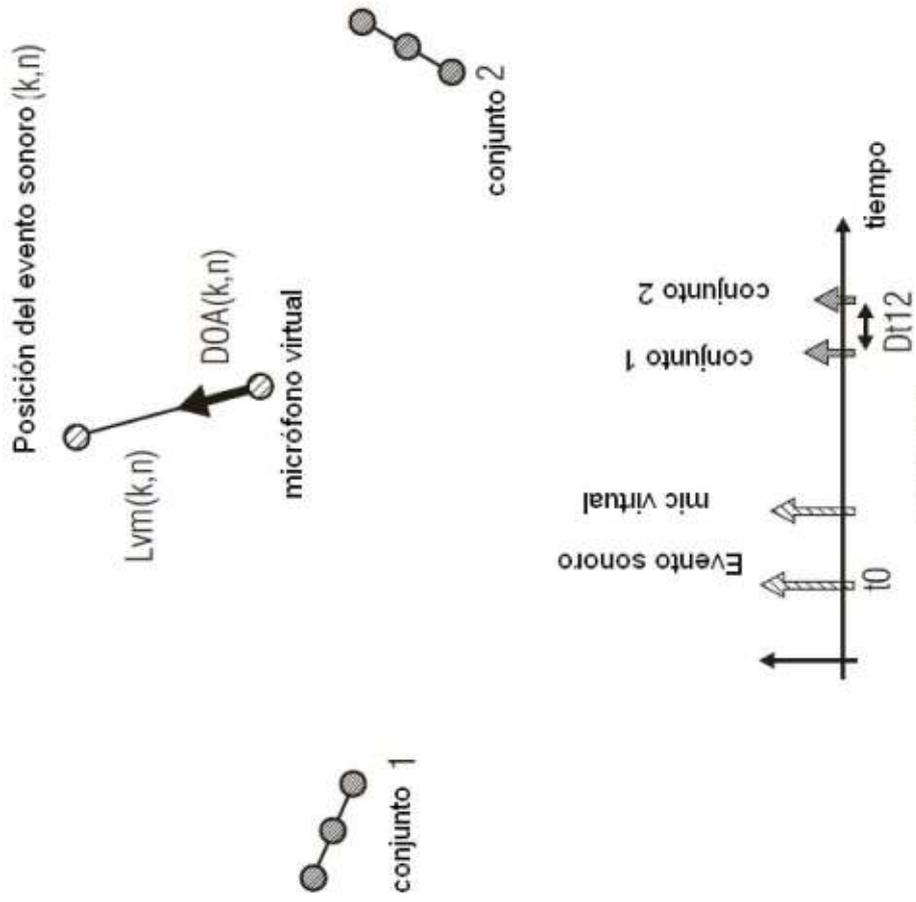


FIG 21

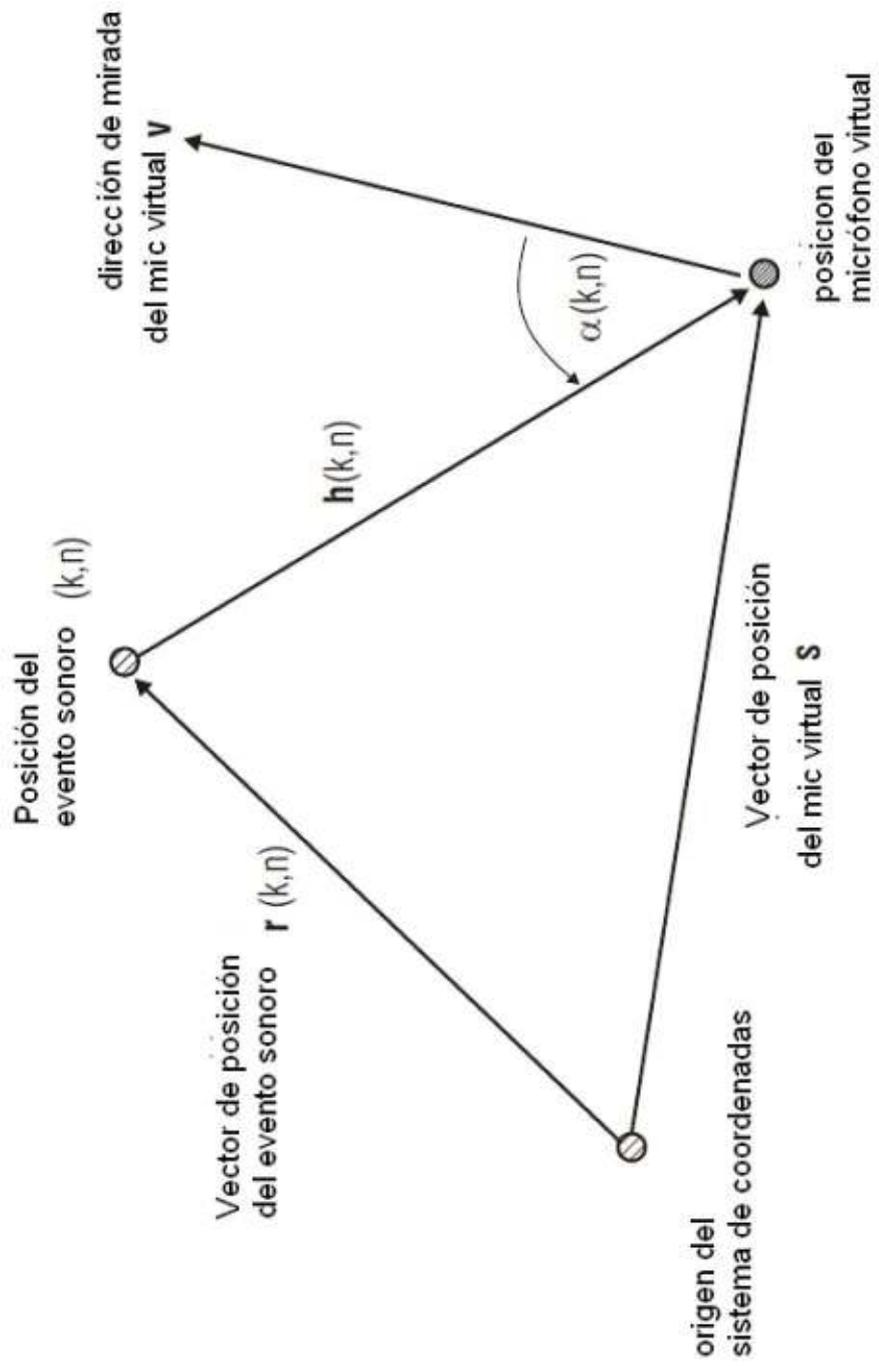


FIG 22

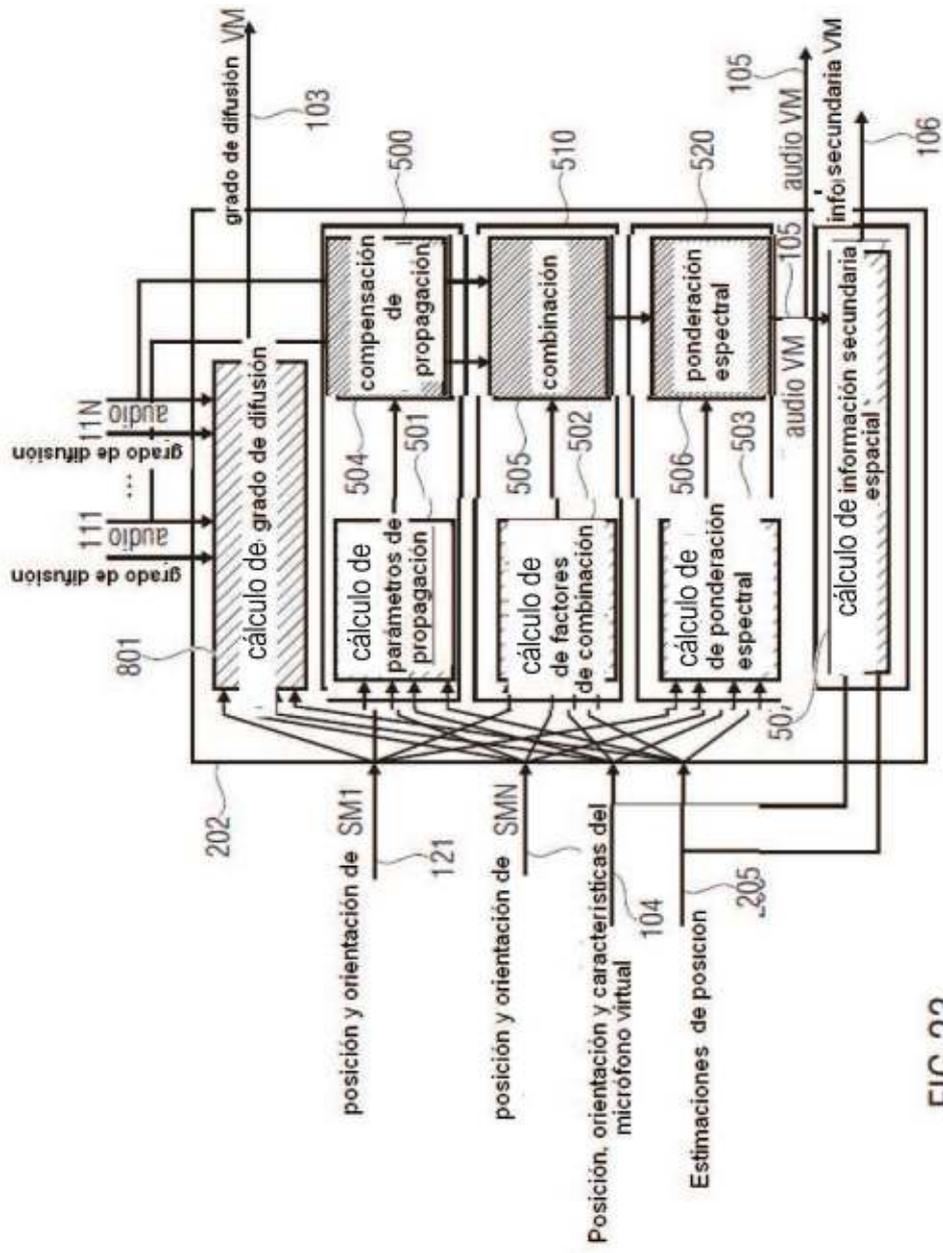


FIG 23

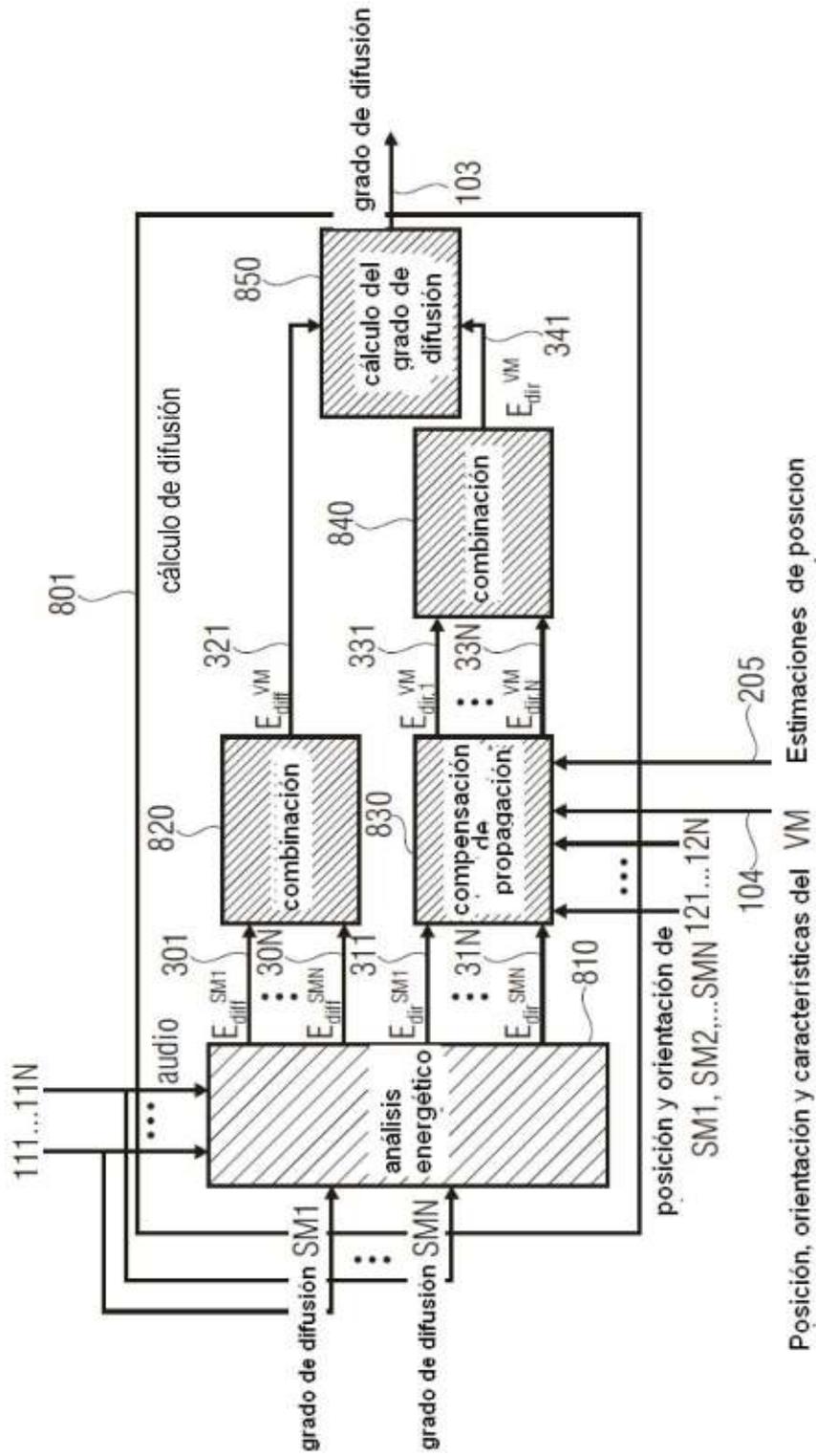


FIG 24

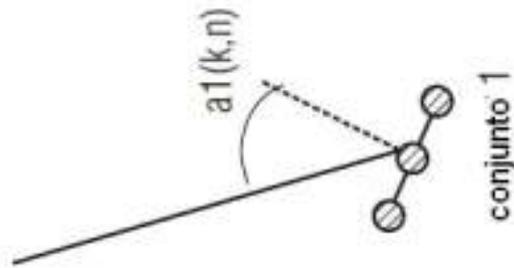
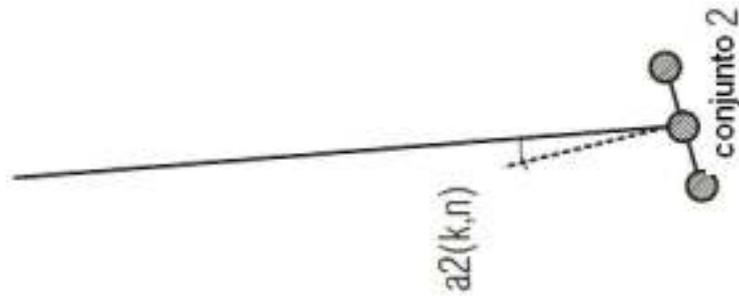


FIG 25

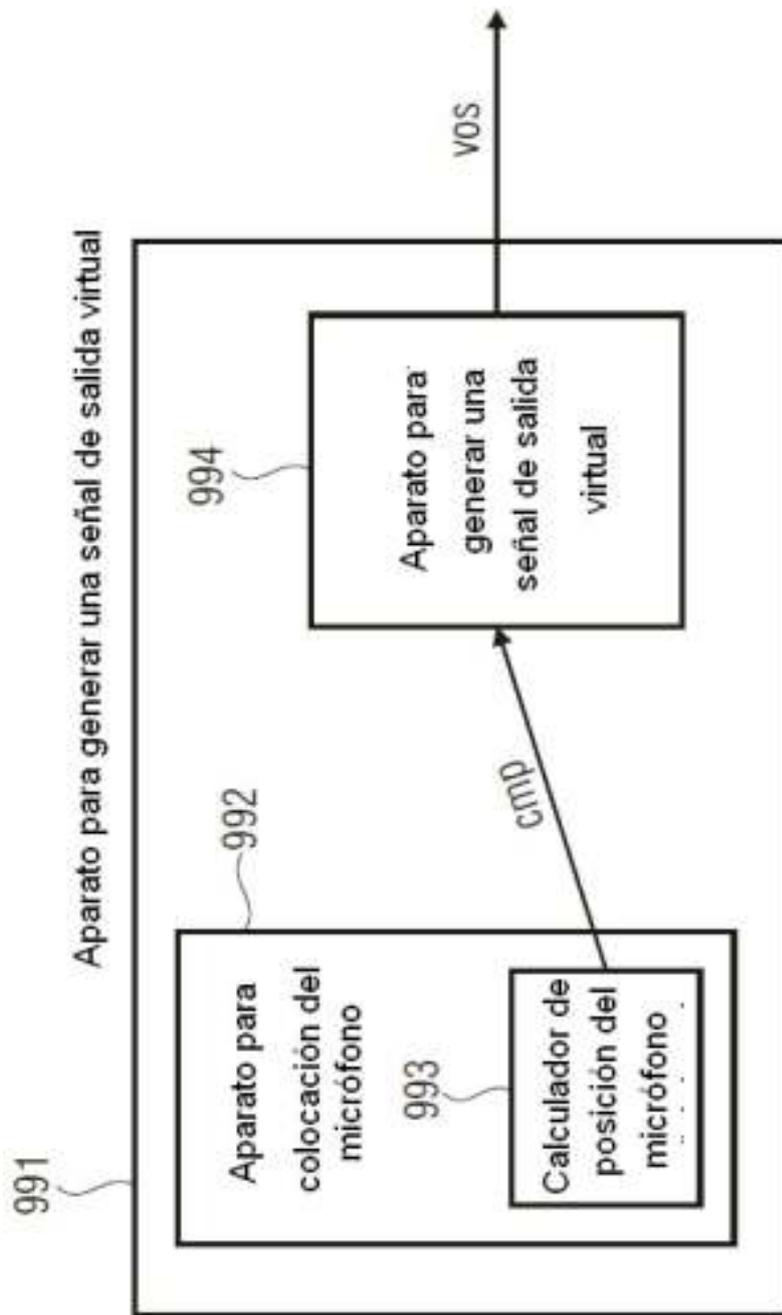


FIG 26

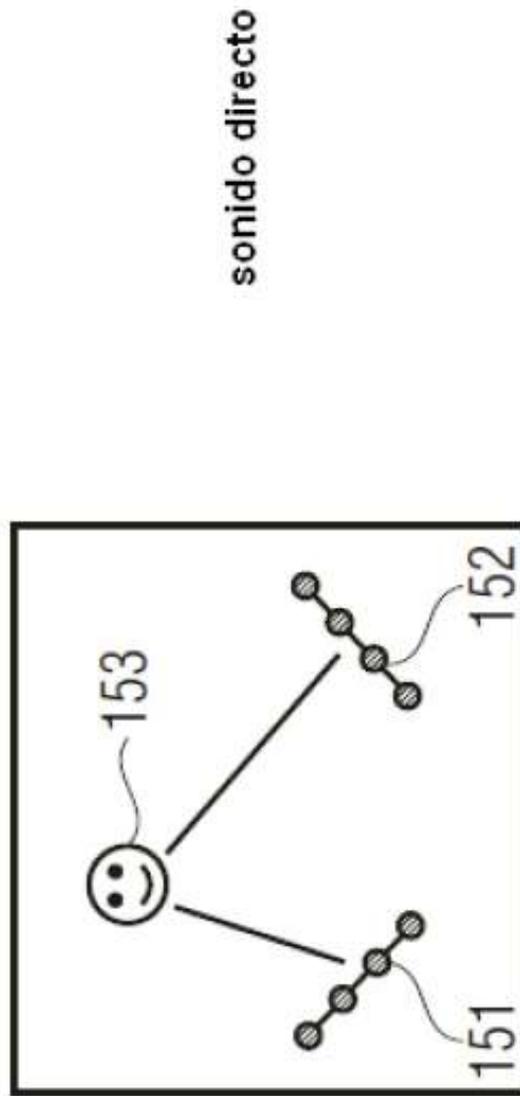
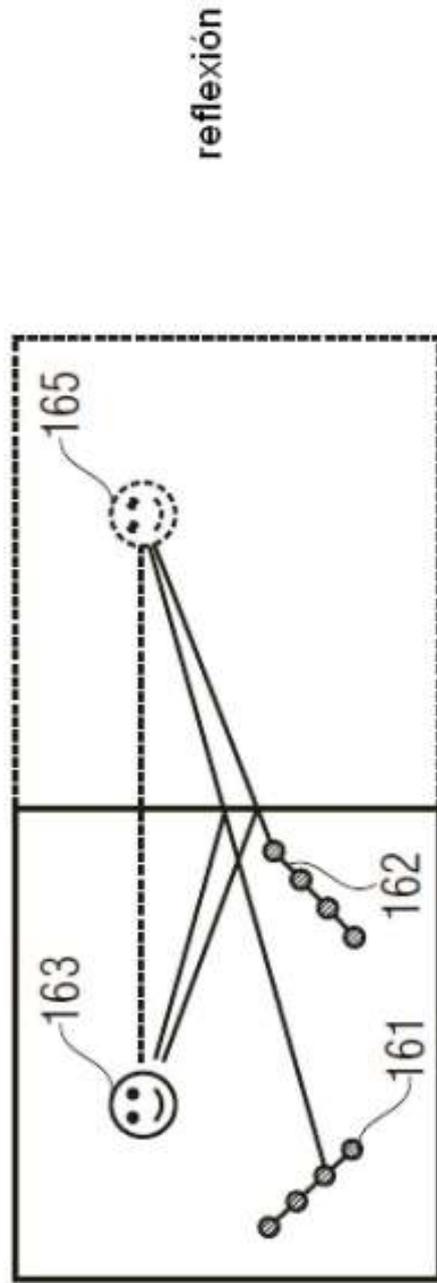
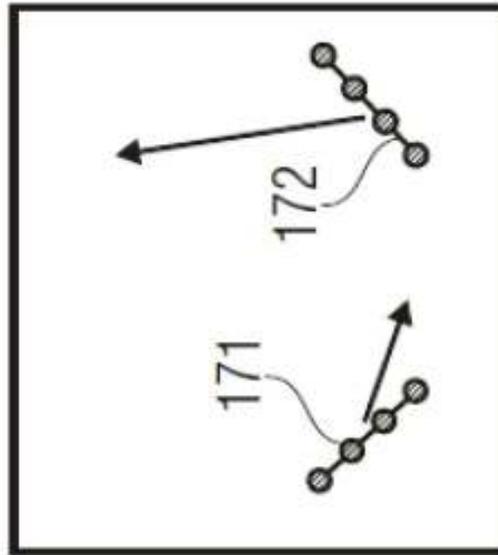


FIG 27A



reflexión

FIG 27B



sonido directo y ruido

FIG 27C