

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 999 614**

51 Int. Cl.:

H04R 3/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.03.2017 E 19187901 (4)**

97 Fecha y número de publicación de la concesión europea: **27.11.2024 EP 3579577**

54 Título: **Aparato, procedimiento o programa informático para generar una descripción de campo de sonido**

30 Prioridad:

15.03.2016 EP 16160504

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

26.02.2025

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.00%)
Hansastr. 27c
80686 München, DE**

72 Inventor/es:

**HABETS, EMANUEL;
THIERGART, OLIVER;
KÜCH, FABIAN;
NIEDERLEITNER, ALEXANDER;
KHAN, AFFAN-HASAN y
MAHNE, DIRK**

74 Agente/Representante:

PONTI & PARTNERS, S.L.P.

ES 2 999 614 T3

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

DESCRIPCIÓN

Aparato, procedimiento o programa informático para generar una descripción de campo de sonido

- 5 **[0001]** La presente invención se refiere a un aparato, un procedimiento o un programa informático para generar una descripción de campo de sonido y también a una síntesis de señales Ambisonics (orden superior) en el dominio del tiempo-frecuencia mediante el uso de información de dirección de sonido.
- 10 **[0002]** La presente invención está en el campo de la grabación y reproducción de sonido espacial. La grabación de sonido espacial tiene como propósito capturar un campo de sonido con múltiples micrófonos de tal manera que, en el lado de la reproducción, un oyente perciba la imagen de sonido como si estuviera en la ubicación de la grabación. Las estrategias estándar para grabación de sonido espacial generalmente utilizan micrófonos omnidireccionales espaciados (por ejemplo, en estereofonía AB), o micrófonos direccionales coincidentes (por ejemplo, en estereofonía de intensidad). Las señales grabadas se pueden reproducir desde una configuración estándar de altavoces estéreo 15 para lograr una imagen de sonido estéreo. Para reproducción de sonido envolvente, por ejemplo, utilizando una configuración de altavoces de 5.1, se pueden utilizar técnicas de grabación similares, por ejemplo, cinco micrófonos cardioides dirigidos hacia las posiciones de los altavoces [ArrayDesign]. Recientemente, han surgido sistemas de reproducción de sonido 3D, tal como la configuración de altavoces de 7.1+4, donde se utilizan 4 altavoces de altura para reproducir sonidos elevados. Las señales para tal configuración de altavoces pueden ser grabadas, por ejemplo, 20 con configuraciones de micrófono 3D espaciadas muy específicas [MicSetup3D]. Todas estas técnicas de grabación tienen en común que están diseñadas para una configuración de altavoces específica, lo cual limita la aplicabilidad práctica, por ejemplo, cuando el sonido grabado se debe reproducir en diferentes configuraciones de altavoces.
- 25 **[0003]** Se logra más flexibilidad cuando no se graban directamente las señales para una configuración de altavoces específica, sino que se graban las señales de un formato intermedio, desde el cual se pueden generar las señales de una configuración de altavoces arbitraria en el lado de la reproducción. Tal formato intermedio, que está bien establecido en la práctica, es representado por Ambisonics (de orden superior) [Ambisonics]. A partir de una señal Ambisonics, se pueden generar las señales de cada configuración de altavoces deseada incluyendo las señales binaurales para reproducción de auriculares. Esto requiere un renderizador específico que se aplica a la señal 30 Ambisonics, tal como un renderizador clásico de Ambisonics [Ambisonics], codificación de audio direccional (DirAC, Directional Audio Coding) [DirAC], o HARPEX [HARPEX].
- 35 **[0004]** Una señal Ambisonics representa una señal multicanal donde cada canal (denominado como componente Ambisonics) es equivalente al coeficiente de una llamada función de base espacial. Con una suma ponderada de estas funciones de base espacial (con los pesos correspondientes a los coeficientes) se puede recrear el campo de sonido original en la ubicación de la grabación [FourierAcoust]. Por lo tanto, los coeficientes de función de base espacial (es decir, los componentes Ambisonics) representan una descripción compacta del campo de sonido en la ubicación de grabación. Existen diferentes tipos de funciones de base espacial, por ejemplo, armónica esférica (SHs, Spherical Harmonics) [FourierAcoust] o armónica cilíndrica (CHs, Cylindrical Harmonics) [FourierAcoust]. CHs 40 se puede utilizar cuando se describe el campo de sonido en el espacio 2D (por ejemplo, para reproducción de sonido 2D) mientras que SHs se puede utilizar para describir el campo de sonido en el espacio 2D y 3D (por ejemplo, para reproducción de sonido 2D y 3D).
- 45 **[0005]** Las funciones de base espacial existen para diferentes órdenes l , y modos m en el caso de funciones de base espacial 3D (tal como SHs). En el último caso, existen $m = 2l + 1$ modos para cada orden l , donde m y l son enteros en el intervalo de $l \geq 0$ y $-l \leq m \leq l$. Un ejemplo correspondiente de las funciones de base espacial se muestra en la fig. 1a, la cual muestra funciones armónicas esféricas para diferentes órdenes l y modos m . Cabe observar que el orden l se denomina en ocasiones como niveles, y que los modos m también se pueden denominar como grados. Como se puede observar en la fig. 1a la armónica esférica del orden de ceros (nivel cero) $l = 0$ representa la presión 50 de sonido omnidireccional en la ubicación de grabación, mientras que la armónica esférica del primer orden (primer nivel) $l = 1$ representa componentes de dipolo a lo largo de las tres dimensiones del sistema de coordenadas cartesianas. Esto significa, que una función de base espacial de un orden (nivel) específico describe la directividad de un micrófono de orden l . En otras palabras, el coeficiente de una función de base espacial corresponde a la señal de un micrófono del orden (nivel) l y modo m . Cabe observar que las funciones de base espacial de diferentes órdenes y 55 modos son mutuamente ortogonales. Esto significa, por ejemplo, que, en un campo de sonido puramente difuso, los coeficientes de todas las funciones de base espacial son mutuamente no correlacionados.
- 60 **[0006]** Como se ha explicado anteriormente, cada componente Ambisonics de una señal Ambisonics corresponde a un coeficiente de función de base espacial de un nivel (y modo) específico. Un ejemplo, si el campo de sonido se describe hasta el nivel $l = 1$ utilizando SHs como función de base espacial, entonces la señal Ambisonics comprendería cuatro componentes Ambisonics (ya que tenemos un modo para el orden $l = 0$ más tres modos para el orden $l = 1$). Las señales Ambisonics de un orden máximo $l = 1$ se denominan como Ambisonics de primer orden (FOA, First-Order Ambisonics) en lo sucesivo, mientras que las señales Ambisonics de un orden máximo $l > 1$ se denominan como Ambisonics de orden superior (HOA, Higher-Order Ambisonics). Cuando se utilizan órdenes superiores l para 65 describir el campo de sonido, la resolución espacial se hace más alta, es decir, se puede describir o crear el campo

de sonido con mayor precisión.

[0007] Por lo tanto, se puede describir un campo de sonido con solamente menos órdenes que llevan a una menor precisión (pero menos datos) o se pueden utilizar mayores órdenes que llevan a mayor precisión (y más datos).

5

[0008] Existen definiciones matemáticas diferentes, pero estrechamente relacionadas, para las diferentes funciones de base espacial. Por ejemplo, se pueden calcular armónicas esféricas de valores complejos, así como armónicas esféricas de valores reales. Por otra parte, las armónicas esféricas se pueden calcular con diferentes términos de normalización tales como normalización SN3D, N3D, o N2D. Las diferentes definiciones se pueden encontrar por ejemplo en [Ambix]. Algunos ejemplos específicos se mostrarán posteriormente junto con la descripción de la invención y las realizaciones.

10

[0009] La señal de Ambisonics deseada se puede determinar a partir de las grabaciones con múltiples micrófonos. La forma directa de obtener señales Ambisonics es el cálculo directo de los componentes Ambisonics (coeficientes de función de base espacial) a partir de las señales de micrófono. Esta estrategia requiere medir la presión de sonido en muchas posiciones específicas, por ejemplo, en un círculo o en una superficie de una esfera. Después de eso, los coeficientes de función de base espacial se pueden calcular mediante la integración a través de las presiones de sonido medidas, como se describe por ejemplo en [FourierAcoust, p. 218]. Esta estrategia directa requiere una configuración de micrófonos específica, por ejemplo, una disposición circular o una disposición esférica de micrófonos omnidireccionales. Dos ejemplos específicos de configuraciones de micrófonos comercialmente disponibles son el micrófono SoundField ST350 o el micrófono EigenMike® [EigenMike]. Desafortunadamente, el requerimiento de una geometría de micrófono específica limita fuertemente la aplicabilidad práctica, por ejemplo, cuando los micrófonos necesitan estar integrados en un dispositivo pequeño o si la disposición de micrófonos necesita combinarse con una cámara de vídeo. Por otra parte, determinar los coeficientes espaciales de órdenes superiores con esta estrategia directa requiere un número relativamente alto de micrófonos para asegurar una robustez suficiente contra el ruido. Por lo tanto, la estrategia directa para obtener una señal Ambisonics es a menudo muy costosa. En el documento FR 2 858 512 se muestra un procedimiento para determinar una señal Ambisonics a partir de grabaciones con múltiples micrófonos.

15

20

25

30

[0010] Es un objetivo de la presente invención proporcionar un concepto mejorado para generar una descripción de campo de sonido que tenga una representación de componentes de campo de sonido.

[0011] Este objetivo se logra por medio de un aparato de conformidad con la reivindicación 1, un procedimiento de conformidad con la reivindicación 24 o un programa informático de conformidad con la reivindicación 25.

35

[0012] La presente invención se refiere a un aparato o un procedimiento o un programa informático para generar una descripción de campo de sonido que tiene una representación de componentes de campo de sonido. En un determinador de dirección, se determina una o más direcciones de sonido para cada cuadro de tiempo-frecuencia de una pluralidad de cuadros de tiempo-frecuencia de una pluralidad de señales de micrófono. Un evaluador de función de base espacial evalúa, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, una o más funciones de base espacial utilizando dichas una o más direcciones de sonido. Además, un calculador de componente de campo de sonido calcula para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de campo de sonido que corresponden a dichas una o más funciones de base espacial evaluadas utilizando dichas una o más direcciones de sonido y utilizando una señal de referencia para un cuadro de tiempo-frecuencia correspondiente, donde la señal de referencia se deriva de dichas una o más señales de micrófono de la pluralidad de señales de micrófono.

40

45

[0013] La presente invención se basa en el hallazgo de que la descripción de campo de sonido que describe un campo de sonido complejo arbitrario se puede derivar de una manera eficiente a partir de una pluralidad de señales de micrófono dentro de una representación de tiempo-frecuencia que consiste en cuadros de tiempo-frecuencia. Estos cuadros de tiempo-frecuencia, por una parte, se refieren a la pluralidad de señales de micrófono y, por otra parte, se utilizan para determinar las direcciones de sonido, por lo tanto, la determinación de dirección de sonido ocurre dentro del dominio espectral utilizando los cuadros de tiempo-frecuencia de la representación de tiempo-frecuencia. Después, la mayor parte del procesamiento subsecuente se lleva a cabo preferentemente dentro de la misma representación de tiempo-frecuencia. Para este fin, se lleva a cabo una evaluación de las funciones de base espacial utilizando dichas una o más direcciones de sonido determinadas para cada cuadro de tiempo-frecuencia. Las funciones de base espacial dependen de las direcciones de sonido, pero son independientes de la frecuencia. Por lo tanto, se aplica una evaluación de las funciones de base espacial con señales de dominio de la frecuencia, es decir, señales en los cuadros de tiempo-frecuencia. Dentro de la misma representación de tiempo-frecuencia, se calculan uno o más componentes de campo de sonido, que corresponden a dichas una o más funciones de base espacial que han sido evaluadas utilizando dichas una o más direcciones, junto con una señal de referencia que también existe dentro de la misma representación de tiempo-frecuencia.

50

55

60

[0014] Estos uno o más componentes de campo de sonido para cada bloque y cada contenedor de frecuencia de una señal, es decir, para cada cuadro de tiempo-frecuencia puede ser el resultado final o, alternativamente, se

65

puede llevar a cabo una conversión de vuelta al dominio del tiempo con el fin de obtener uno o más componentes de campo de sonido en el dominio del tiempo que corresponden a dichas una o más funciones de base espacial. Dependiendo de la implementación, los uno o más componentes de campo de sonido pueden ser componentes de campo de sonido directo determinados dentro de la representación de tiempo-frecuencia utilizando cuadros de tiempo-frecuencia o pueden ser componentes de campo de sonido difuso generalmente para ser determinados además de los componentes de campo de sonido directo. Los componentes de campo de sonido finales que tienen una parte directa y la parte difusa se pueden obtener entonces por medio de la combinación de los componentes de campo de sonido directo y los componentes de campo de sonido difuso, donde esta combinación se puede llevar a cabo ya sea en el dominio del tiempo o en el dominio de la frecuencia dependiendo de la implementación real.

10

[0015] Se pueden llevar a cabo varios procedimientos con el fin de derivar la señal de referencia a partir de dichas una o más señales de micrófono. Tales procedimientos pueden comprender la selección directa de una cierta señal de micrófono a partir de la pluralidad de señales de micrófono o una selección avanzada que se basa en dichas una o más direcciones de sonido. La determinación de señal de referencia avanzada selecciona una señal de micrófono específica de la pluralidad de señales de micrófono que es de un micrófono ubicado más cercano a la dirección de sonido entre los micrófonos de los cuales han sido derivadas las señales de micrófono. Una alternativa adicional es aplicar un filtro multicanal a dichas dos o más señales de micrófono con el fin de filtrar conjuntamente esas señales de micrófono de tal manera que se obtenga una señal de referencia común para todos los cuadros de frecuencia de un bloque de tiempo. Alternativamente, se pueden derivar diferentes señales de referencia para diferentes cuadros de frecuencia dentro de un bloque de tiempo. Naturalmente, se pueden generar también diferentes señales de referencia para diferentes bloques de tiempo, pero para las mismas frecuencias dentro de los diferentes bloques de tiempo. Por lo tanto, dependiendo de la implementación, la señal de referencia para un cuadro de tiempo-frecuencia se puede seleccionar o derivar libremente a partir de la pluralidad de señales de micrófono.

25 **[0016]**

En este contexto, se debe hacer énfasis en que los micrófonos se pueden ubicar en ubicaciones arbitrarias. Los micrófonos pueden tener diferentes características direccionales, también. Además, la pluralidad de señales de micrófono no necesariamente tiene que ser señales que hayan sido grabadas por micrófonos físicos reales. Más bien, las señales de micrófono pueden ser señales de micrófono que han sido creadas artificialmente a partir de un cierto campo de sonido utilizando ciertas operaciones de procesamiento de datos que imitan a los micrófonos físicos reales.

30

[0017] Para el propósito de determinar componentes de campo de sonido difuso en ciertas realizaciones, son posibles diferentes procedimientos y son útiles para ciertas implementaciones. Generalmente, una porción difusa se deriva a partir de la pluralidad de señales de micrófono como la señal de referencia y esta señal de referencia (difusa) se procesa después junto con una respuesta promedio de la función de base espacial de un cierto orden (o un nivel y/o un modo) con el fin de obtener el componente de sonido difuso para este orden o nivel o modo. Por lo tanto, un componente de sonido directo se calcula utilizando la evaluación de una cierta función de base espacial con una cierta dirección de llegada y un componente de sonido difuso, naturalmente, no se calcula utilizando una cierta dirección de llegada, sino que se calcula mediante el uso de la señal de referencia difusa combinando la señal de referencia difusa y la respuesta promedio de una función de base espacial de un cierto orden o nivel o modo por medio de una cierta función. Esta combinación funcional puede ser, por ejemplo, una multiplicación que también se puede llevar a cabo en el cálculo del componente de sonido directo o esta combinación puede ser una multiplicación ponderada o una adición o una sustracción, por ejemplo, cuando se llevan a cabo los cálculos en el dominio logarítmico. Se llevan a cabo otras combinaciones diferentes de una multiplicación o adición/sustracción utilizando una función no lineal o lineal adicional, donde las funciones no lineales son preferidas. Después de la generación del componente de campo de sonido directo y el componente de campo de sonido difuso de un cierto orden, se puede llevar a cabo una combinación al combinar el componente de campo de sonido directo y el componente de campo de sonido difuso dentro del dominio espectral para cada cuadro de tiempo/frecuencia individual. Alternativamente, los componentes de campo de sonido difuso y los componentes de campo de sonido directo para un cierto orden se pueden transformar del dominio de la frecuencia al dominio del tiempo y después se puede llevar a cabo también una combinación del dominio del tiempo de un componente de dominio de tiempo directo y un componente de dominio de tiempo difuso de un cierto orden.

35

40

45

50

[0018] Dependiendo de la situación, se pueden utilizar decorreladores adicionales para decorrelacionar los componentes de campo de sonido difuso. Alternativamente, los componentes de campo de sonido difuso decorrelacionados se pueden generar mediante el uso de diferentes señales de micrófono o diferentes cuadros de tiempo/frecuencia para diferentes componentes de campo de sonido difuso de diferentes órdenes o utilizando una señal de micrófono diferente para el cálculo del componente de campo de sonido directo y una señal de micrófono diferente adicional para el cálculo del componente de campo de sonido difuso.

60

[0019] En una realización preferida, las funciones de base espacial son funciones de base espacial asociadas con ciertos niveles (órdenes) y modos de la descripción del bien conocido campo de sonido Ambisonics. Un componente de campo de sonido de un cierto orden y un cierto modo corresponderían a un componente de campo de sonido Ambisonics asociado con un cierto nivel y un cierto modo. Generalmente, el primer componente de campo de sonido sería el componente de campo de sonido asociado con la función de base espacial omnidireccional como se

65

indica en la fig. 1a para el orden $l = 0$ y modo $m = 0$.

- [0020]** El segundo componente de campo de sonido podría estar asociado, por ejemplo, con una función de base espacial que tiene una directividad máxima dentro de la dirección X que corresponde al orden $l = 1$ y modo $m = -1$ con respecto a la fig. 1a. El tercer componente de campo de sonido podría ser, por ejemplo, una función de base espacial siendo direccional en la dirección Y que podría corresponder con el modo $m = 0$ y el orden $l = 1$ de la fig. 1a y un cuarto componente de campo de sonido podría ser, por ejemplo, una función de base espacial siendo direccional en la dirección Z que corresponde al modo $m = 1$ y el orden $l = 1$ de la fig. 1a.
- 10 **[0021]** Sin embargo, otras descripciones de campo de sonido además de Ambisonics son, desde luego, bien conocidas por aquellos experimentados en la materia y dichos otros componentes de campo de sonido que se basan en diferentes funciones de base espacial de las funciones de base espacial de Ambisonics también se podrían calcular convenientemente dentro de la representación del dominio tiempo-frecuencia como se discutió anteriormente.
- 15 **[0022]** Las realizaciones de la siguiente invención describen una forma fácil de obtener señales Ambisonics. En contraste a las estrategias del estado de la técnica mencionadas anteriormente, la presente estrategia se puede aplicar a configuraciones de micrófonos arbitrarias que poseen dos o más micrófonos. Sin embargo, los componentes de Ambisonics de órdenes superiores se pueden calcular utilizando relativamente pocos micrófonos solamente. Por lo tanto, la presente estrategia es comparativamente económica y práctica. En la realización que se propone, los componentes de Ambisonics no se calculan directamente a partir de la información de presión de sonido a lo largo de una superficie específica, como en las estrategias del estado de la técnica explicadas anteriormente, sino que se sintetizan con base en una estrategia paramétrica. Para este propósito, se asume un modelo de campo de sonido muy simple, similar al que se utiliza por ejemplo en DirAC [DirAC]. Más precisamente, se supone que el campo de sonido en la ubicación de grabación consiste en uno o unos pocos sonidos directos que llegan desde direcciones de sonido específicas más sonido difuso que llega de todas direcciones. Con base en este modelo, y utilizando información paramétrica sobre el campo de sonido tal como la dirección de sonido de los sonidos directos, es posible sintetizar los componentes de Ambisonics o cualquier otro componente de campo de sonido a partir de solamente unas pocas mediciones de la presión de sonido. La presente estrategia se explica en detalle en las siguientes secciones.
- 20
25
- 30 **[0023]** Las realizaciones preferidas de la presente invención se explican subsecuentemente con respecto a los dibujos adjuntos, en los cuales:
- La fig. 1a muestra funciones armónicas esféricas para diferentes órdenes y modos;
 - La fig. 1b muestra un ejemplo de cómo seleccionar el micrófono de referencia con base en la información de dirección-de-llegada;
 - La fig. 1c muestra una implementación preferida de un aparato o procedimiento para generar una descripción de campo de sonido;
 - La fig. 1d ilustra la conversión de tiempo-frecuencia de una señal de micrófono ejemplar donde se identifican específicamente los cuadros de tiempo-frecuencia (10, 1) específicos para un contenedor de frecuencia 10 y bloque de tiempo 1 por una parte y (5, 2) para un contenedor de frecuencia 5 y bloque de tiempo 2;
 - La fig. 1e ilustra la evaluación de cuatro funciones de base espacial que utilizan las direcciones de sonido para los contenedores de frecuencia (10, 1) y (5, 2) identificados;
 - La fig. 1f ilustra el cálculo de los componentes de campo de sonido para los dos contenedores (10, 1) y (5, 2) y la conversión de frecuencia-tiempo subsecuente y procesamiento de transición gradual/superposición-adición;
 - La fig. 1g ilustra una representación del dominio del tiempo de cuatro componentes de campo de sonido ejemplares b_1 a b_4 como se obtienen por medio del procesamiento de la fig. 1f;
 - La fig. 2a muestra un esquema de bloques general de la presente invención;
 - La fig. 2b muestra un esquema de bloques general de la presente invención donde se aplica la transformada de tiempo-frecuencia inversa antes del combinador;
 - La fig. 3a muestra una realización de la invención donde se calcula un componente Ambisonics de un nivel y modo deseados a partir de una señal de micrófono de referencia e información de dirección de sonido;
 - La fig. 3b muestra una realización de la invención donde el micrófono de referencia se selecciona con base en información de dirección-de-llegada;
 - La fig. 4 muestra una realización de la invención donde se calcula un componente Ambisonics de sonido directo y un componente Ambisonics de sonido difuso;
 - La fig. 5 muestra una realización de la invención donde el componente Ambisonics de sonido difuso está decorrelacionado;
 - La fig. 6 muestra una realización de la invención donde el sonido directo y el sonido difuso se extraen desde múltiples micrófonos e información de dirección de sonido;
 - La fig. 7 muestra una realización de la invención donde el sonido difuso se extrae desde múltiples micrófonos y donde el componente Ambisonics de sonido difuso está decorrelacionado; y
 - La fig. 8 muestra una realización de la invención donde se aplica un suavizado de ganancia a la respuesta de función de base espacial.
- 35
40
45
50
55
60

5 [0024] Una realización preferida se ilustra en la fig. 1c. La fig. 1c ilustra una realización de un aparato o procedimiento para generar una descripción de campo de sonido 130 que tiene una representación de componentes de campo de sonido tal como una representación del dominio del tiempo de componentes de campo de sonido o una representación del dominio de la frecuencia de componentes de campo de sonido, una representación codificada o decodificada o una representación intermedia.

10 [0025] Para este fin, un determinador de dirección 102 determina una o más direcciones de sonido 131 para cada cuadro de tiempo-frecuencia de una pluralidad de cuadros de tiempo-frecuencia de una pluralidad de señales de micrófono.

15 [0026] Por lo tanto, el determinador de dirección recibe, en su entrada 132, al menos dos señales de micrófono diferentes y, para cada una de esas dos señales de micrófono diferentes, está disponible una representación de tiempo-frecuencia que consiste generalmente en bloques subsecuentes de contenedores espectrales, donde un bloque de contenedores espectrales tiene un cierto índice de tiempo n asociado con el mismo, donde el índice de frecuencia es k . Un bloque de contenedores de frecuencia para un índice de tiempo representa un espectro de la señal del dominio del tiempo para un bloque de muestras del dominio del tiempo generadas por una cierta operación de formación de ventanas.

20 [0027] Las direcciones de sonido 131 son utilizadas por un evaluador de función de base espacial 103 para evaluar, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, una o más funciones de base espacial. Por lo tanto, el resultado del procesamiento en el bloque 103 es una o más funciones de base espacial evaluadas para cada cuadro de tiempo-frecuencia. Preferentemente, se utilizan dos o incluso más funciones de base espacial diferentes tal como cuatro funciones de base espacial como se discute con respecto a las figuras 1e y 1f. Por lo tanto, en la salida 133 del bloque 103, las funciones de base espacial evaluadas de diferentes órdenes y modos para los tres cuadros de tiempo-frecuencia diferentes de la representación de tiempo-espectro están disponibles y se introducen en el calculador de componente de campo de sonido 201. El calculador de componente de campo de sonido 201 utiliza adicionalmente una señal de referencia 134 generada por un calculador de señal de referencia (no mostrado en la fig. 1c). La señal de referencia 134 se deriva de una o más señales de micrófono de la pluralidad de señales de micrófono y se utiliza por el calculador de componente de campo de sonido dentro de la misma representación de tiempo/frecuencia.

35 [0028] Por lo tanto, el calculador de componente de campo de sonido 201 está configurado para calcular, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de campo de sonido que corresponden a dichas una o más funciones de base espacial evaluadas utilizando dichas una o más direcciones de sonido con la ayuda de una o más señales de referencia para el cuadro de tiempo-frecuencia correspondiente.

40 [0029] Dependiendo de la implementación, el evaluador de función de base espacial 103 está configurado para utilizar, para una función de base espacial, una representación parametrizada, donde un parámetro de la representación parametrizada es una dirección de sonido, siendo la dirección de sonido unidimensional en una situación bidimensional o bidimensional en una situación tridimensional, y para insertar un parámetro que corresponde a la dirección de sonido en la representación parametrizada para obtener un resultado de evaluación para cada función de base espacial.

45 [0030] Alternativamente, el evaluador de función de base espacial está configurado para utilizar una tabla de búsqueda para cada función de base espacial que tiene, como una entrada, una identificación de función de base espacial y la dirección de sonido y que tiene, como una salida, un resultado de evaluación. En esta situación, el evaluador de función de base espacial está configurado para determinar, para dichas una o más direcciones de sonido determinadas por el determinador de dirección 102, una dirección de sonido correspondiente de la entrada de la tabla de búsqueda. Generalmente, las diferentes entradas de dirección son cuantificadas en una forma que, por ejemplo, existe un cierto número de entradas de tabla tal como diez direcciones de sonido diferentes.

55 [0031] El evaluador de función de base espacial 103 está configurado para determinar, para una cierta dirección de sonido específica que no coincide inmediatamente con una entrada de dirección de sonido para la tabla de búsqueda, la entrada de tabla de búsqueda correspondiente. Esto se puede llevar a cabo, por ejemplo, mediante el uso, para una cierta dirección de sonido determinada, la siguiente dirección de sonido superior o inferior introducida en la tabla de búsqueda. Alternativamente, la tabla se utiliza de tal forma que se calcula una media ponderada entre las dos entradas vecinas de la tabla de búsqueda. Por lo tanto, el procedimiento sería que se determina la salida de la tabla para la siguiente entrada de dirección inferior. Además, la salida de la tabla de búsqueda para la siguiente entrada superior se determina y después se calcula un promedio entre esos valores.

65 [0032] Este promedio puede ser un promedio simple obtenido al agregar las dos salidas y al dividir los resultados entre dos o puede ser un promedio ponderado dependiendo de la posición de la dirección de sonido determinada con respecto a la siguiente salida de la tabla superior e inferior. Por lo tanto, de manera ejemplar, un factor de ponderación dependería de la diferencia entre la dirección de sonido determinada y la siguiente entrada

superior/inferior correspondiente en la tabla de búsqueda. Por ejemplo, cuando la dirección medida está cerca de la siguiente entrada inferior, entonces el resultado de la tabla de búsqueda para la siguiente entrada inferior se multiplica por un factor de ponderación superior en comparación con el factor de ponderación, por lo cual se pondera la salida de la tabla de búsqueda para la siguiente entrada superior. Por lo tanto, una pequeña diferencia entre la dirección
 5 determinada y la siguiente entrada inferior, la salida de la tabla de búsqueda para la siguiente entrada inferior se ponderaría con un factor de ponderación superior en comparación con un factor de ponderación utilizado para ponderar una salida de la tabla de búsqueda correspondiente a la siguiente entrada de tabla de búsqueda superior para la dirección del sonido.

10 **[0033]** Posteriormente, las figuras 1d a 1g se discuten para mostrar ejemplos para el cálculo específico de los diferentes bloques en mayor detalle.

[0034] La ilustración superior en la fig. 1d muestra una señal de micrófono esquemática. Sin embargo, la amplitud real de la señal de micrófono no se ilustra. En su lugar, se ilustran ventanas y, particularmente, las ventanas
 15 151 y 152. La ventana 151 define un primer bloque 1 y la ventana 152 identifica y determina un segundo bloque 2. Por lo tanto, se procesa una señal de micrófono con bloques preferentemente superpuestos donde la superposición es igual al 50 %. Sin embargo, se podría utilizar también una superposición superior o inferior, e incluso sería factible sin superposición del todo. Sin embargo, se lleva a cabo un procesamiento de superposición con el fin de evitar artefactos de bloqueo.

20 **[0035]** Cada bloque de valores de muestreo de la señal de micrófono se convierte en una representación espectral. La representación espectral o espectro para el bloque con el índice de tiempo $n = 1$, es decir, para el bloque 151, se ilustra en la representación intermedia en la fig. 1d, y la representación espectral del segundo bloque 2 que corresponde al número de referencia 152 se ilustra en la imagen inferior en la fig. 1d. Además, por razones ejemplares,
 25 se muestra que cada espectro tiene 10 contenedores de frecuencia, es decir, el índice de frecuencia k se extiende entre 1 y 10, por ejemplo.

[0036] Por lo tanto, el cuadro de tiempo-frecuencia (k, n) es el cuadro de tiempo-frecuencia (10, 1) en 153 y, un ejemplo adicional muestra otro cuadro de tiempo-frecuencia (5, 2) en 154. El procesamiento adicional llevado a cabo
 30 por el aparato para generar una descripción de campo de sonido se ilustra, por ejemplo, en la fig. 1d, ilustrado ejemplarmente utilizando estos cuadros de tiempo-frecuencia indicados por los números de referencia 153 y 154.

[0037] Se supone, además, que el determinador de dirección 102 determina una dirección de sonido o dirección de llegada (DOA, Direction Of Arrival) indicada ejemplarmente por medio del vector normal unitario n . Indicaciones de
 35 dirección alternativas comprenden un ángulo de azimut, un ángulo de elevación o ambos ángulos conjuntamente. Para este fin, todas las señales de micrófono de la pluralidad de señales de micrófono, donde cada señal de micrófono es representada por bloques subsiguientes de contenedores de frecuencia como se ilustra en la fig. 1d, son utilizadas por el determinador de dirección 102, y el determinador de dirección 102 de la fig. 1c determina entonces la dirección de sonido o DOA, por ejemplo. Por lo tanto, ejemplarmente, el cuadro de tiempo-frecuencia (10, 1) tiene la dirección
 40 de sonido $n(10, 1)$ y el cuadro de tiempo-frecuencia (5, 2) tiene la dirección de sonido $n(5, 2)$ como se ilustra en la porción superior de la fig. 1e. En el caso tridimensional, la dirección de sonido es un vector tridimensional que tiene un componente X, Y, o Z. Naturalmente, también se pueden utilizar otros sistemas de coordenadas tales como las coordenadas esféricas los cuales se basan en dos ángulos y un radio. Alternativamente, los ángulos pueden ser, por
 45 ejemplo, azimut y elevación. Entonces, no se requiere el radio. De manera similar, hay dos componentes de la dirección de sonido en un caso bidimensional tal como las coordenadas cartesianas, es decir, una dirección X e Y, pero, alternativamente, también se pueden utilizar coordenadas circulares que tienen un radio y un ángulo o azimut y ángulos de elevación.

[0038] Este procedimiento no solamente se lleva a cabo para los cuadros de tiempo-frecuencia (10, 1) y (5, 2),
 50 sino también para todos los cuadros de tiempo-frecuencia, por los cuales se representan las señales de micrófono.

[0039] Entonces, se determinan dichas una o más funciones de base espacial requeridas. Particularmente, se determina el número de componentes de campo de sonido o, generalmente, la representación de los componentes de campo de sonido que se debe generar. El número de funciones de base espacial que son utilizadas ahora por el
 55 evaluador de función de base espacial 103 de la fig. 1c determina finalmente el número de componentes de campo de sonido para cada cuadro de tiempo-frecuencia en una representación espectral o el número de componentes de campo de sonido en el dominio del tiempo.

[0040] Para la realización adicional, se supone que se va a determinar un número de cuatro componentes de
 60 campo de sonido donde, ejemplarmente, estos cuatro componentes de campo de sonido pueden ser un componente de campo de sonido omnidireccional (que corresponde al orden igual a 0) y tres componentes de campo de sonido direccional que son direccionales en las direcciones coordenadas correspondientes del sistema de coordenadas cartesianas.

65 **[0041]** La ilustración inferior en la fig. 1e ilustra las funciones de base espacial evaluadas G_i para diferentes

cuadros de tiempo-frecuencia. Por lo tanto, se hace más claro que, en este ejemplo, se determinan cuatro funciones de base espacial evaluadas para cada cuadro de tiempo-frecuencia. Cuando se supone ejemplarmente que cada bloque tiene diez cuadros de frecuencia, entonces se determina un número de 40 funciones de base espacial evaluadas G_i para cada bloque tal como para el bloque $n = 1$ y para el bloque $n = 2$ como se ilustra en la fig. 1e. Por lo tanto, todos juntos, cuando solamente se consideran dos bloques y cada bloque tiene diez contenedores de frecuencia, entonces el procedimiento resulta en 80 funciones de base espacial evaluadas, ya que hay veinte cuadros de tiempo-frecuencia en los dos bloques y cada cuadro de tiempo-frecuencia tiene cuatro funciones de base espacial evaluadas.

10 **[0042]** La fig. 1f ilustra implementaciones preferidas del calculador de componente de campo de sonido 201 de la Fig. 1c. La fig. 1f ilustra en las dos ilustraciones superiores dos bloques de contenedores de frecuencia para la entrada de señal de referencia determinada en el bloque 201 en la fig. 1c por medio de la línea 134. Particularmente, una señal de referencia que puede ser una señal de micrófono específica o una combinación de las señales de micrófono diferentes se ha procesado en la misma manera como se había discutido con respecto a la fig. 1d. Por lo tanto, ejemplarmente, la señal de referencia se representa por medio de un espectro de referencia para un bloque $n = 1$ y un espectro de señal de referencia para el bloque $n = 2$. Por lo tanto, la señal de referencia se descompone en el mismo patrón de tiempo-frecuencia como se ha utilizado para el cálculo de las funciones de base espacial evaluadas para los cuadros de tiempo-frecuencia proporcionados por medio de la línea 133 del bloque 103 al bloque 201.

20 **[0043]** Entonces, el cálculo real de los componentes de campo de sonido se lleva a cabo por medio de una combinación funcional entre el cuadro de tiempo-frecuencia correspondiente para la señal de referencia P y la función de base espacial evaluada asociada G , como se indica en 155. Preferentemente, una combinación funcional representada por $f(\dots)$ es una multiplicación ilustrada en 115 en las figuras 3a, 3b que se discuten subsecuentemente. Sin embargo, también se pueden utilizar otras combinaciones funcionales, como se discutió antes. Por medio de la combinación funcional en el bloque 155, dichos uno o más componentes de campo de sonido B_i se calculan para cada cuadro de tiempo-frecuencia con el fin de obtener la representación en el dominio de la frecuencia (espectral) de los componentes de campo de sonido B_i como se ilustra en 156 para el bloque $n = 1$ y en 157 para el bloque $n = 2$.

30 **[0044]** Por lo tanto, ejemplarmente, la representación en el dominio de la frecuencia de los componentes de campo de sonido B_i se ilustra para el cuadro de tiempo-frecuencia (10, 1) por una parte y también para el cuadro de tiempo-frecuencia (5, 2) para el segundo bloque por otra parte. Sin embargo, está claro de nuevo que el número de componentes de campo de sonido B_i que se ilustran en la fig. 1f en 156 y 157 es el mismo que el número de funciones de base espacial evaluadas que se ilustran en la porción inferior de la fig. 1e.

35 **[0045]** Cuando solamente se requieren los componentes de campo de sonido en el dominio de la frecuencia, el cálculo se completa con la salida de los bloques 156 y 157. Sin embargo, en otras realizaciones, se requiere una representación en el dominio del tiempo de los componentes de campo de sonido con el fin de obtener una representación en el dominio del tiempo para el primer componente de campo de sonido B_1 , una representación en el dominio del tiempo adicional para el segundo componente de campo de sonido B_2 y así sucesivamente.

40 **[0046]** Para este fin, los componentes de campo de sonido B_1 del contenedor de frecuencia 1 al contenedor de frecuencia 10 en el primer bloque 156 se insertan en un bloque de transferencia de frecuencia-tiempo 159 con el fin de obtener una representación en el dominio del tiempo para el primer bloque y el primer componente.

45 **[0047]** Análogamente, con el fin de determinar y calcular el primer componente en el dominio del tiempo, es decir, $b_1(t)$, los componentes de campo de sonido espectrales B_1 para el segundo bloque que va desde el contenedor de frecuencia 1 al contenedor de frecuencia 10 se convierten en una representación en el dominio del tiempo por medio de una transformada de frecuencia-tiempo adicional 160.

50 **[0048]** Debido al hecho de que se utilizaron ventanas superpuestas como se ilustra en la porción superior de la fig. 1d, se puede utilizar una operación de transición gradual o superposición-adición 161 que se ilustra en la parte inferior de la fig. 1f con el fin de calcular las muestras en el dominio del tiempo de salida de la primera representación espectral $b_1(d)$ en el intervalo de superposición entre el bloque 1 y el bloque 2 que se ilustra en 162 en la fig. 1g.

55 **[0049]** Se lleva a cabo el mismo procedimiento con el fin de calcular el segundo componente de campo de sonido en el dominio del tiempo $b_2(t)$ dentro de un intervalo de superposición 163 entre el primer bloque y el segundo bloque. Además, con el fin de calcular el tercer componente de campo de sonido en el mismo dominio del tiempo $b_3(t)$ y, particularmente, con el fin de calcular las muestras en el intervalo de superposición 164, los componentes D_3 del primer bloque y los componentes D_3 del segundo bloque se convierten correspondientemente en una representación en el dominio del tiempo por medio de los procedimientos 159, 160 y los valores resultantes se someten entonces a transición gradual o superposición-adición en el bloque 161.

60 **[0050]** Finalmente, se lleva a cabo el mismo procedimiento para los cuatro componentes B_4 para el primer bloque y B_4 para el segundo bloque con el fin de obtener las muestras finales del cuarto componente de campo de sonido de representación en el dominio del tiempo $b_4(t)$ en el intervalo de superposición 165 como se ilustra en la fig.

1g.

[0051] Se debe observar que no se requiere ninguna transición gradual/superposición-adición como se ilustra en el bloque 161, cuando el procesamiento, con el fin de obtener cuadros de tiempo-frecuencia, no se lleva a cabo con bloques superpuestos, pero se lleva a cabo con bloques no superpuestos.

[0052] Además, en el caso de una superposición superior donde más de dos bloques se superponen entre sí, se requiere un número correspondientemente superior de bloques 159, 160 y la transición gradual/superposición-adición del bloque 161 se calcula no solamente con dos entradas sino incluso con tres entradas para obtener finalmente muestras de las representaciones en el dominio del tiempo ilustradas en la fig. 1g.

[0053] Además, se debe observar que las muestras para las representaciones en el dominio del tiempo, por ejemplo, para el intervalo de superposición OL_{23} se obtienen aplicando los procedimientos en el bloque 159, 160 al segundo bloque y el tercer bloque. Correspondientemente, las muestras para el intervalo de superposición $OL_{0,1}$ se calculan llevando a cabo los procedimientos 159, 160 en los componentes de campo de sonido espectrales B_i para dicho cierto número i para el bloque 0 y el bloque 1.

[0054] Además, como ya se ha descrito, la representación de los componentes de campo de sonido puede ser una representación en el dominio de la frecuencia como se ilustra en la fig. 1f para 156 y 157. Alternativamente, la representación de los componentes de campo de sonido puede ser una representación en el dominio del tiempo como se ilustra en la fig. 1g, donde los cuatro componentes de campo de sonido representan señales de sonido directas que tienen una secuencia de muestras asociadas con una cierta tasa de muestreo. Además, se puede codificar cualquiera de la representación en el dominio de la frecuencia o la representación en el dominio del tiempo de los componentes de campo de sonido. Esta codificación se puede llevar a cabo por separado de tal manera que cada componente de campo de sonido se codifique como una señal mono o la codificación se puede llevar a cabo conjuntamente, de modo que, por ejemplo, los cuatro componentes de campo de sonido B_1 a B_4 se consideren como una señal multicanal que tiene cuatro canales. Por lo tanto, cualquiera de una representación codificada en el dominio de la frecuencia o una representación en el dominio del tiempo que es codificada con cualquier algoritmo de codificación útil también es una representación de los componentes de campo de sonido.

[0055] Además, incluso una representación en el dominio del tiempo antes de la transición gradual/superposición-adición llevada a cabo por el bloque 161 puede ser una representación útil de componentes de campo de sonido para una cierta implementación. Además, también se puede llevar a cabo un tipo de cuantificación de vector sobre los bloques n para un cierto componente tal como el componente 1 con el fin de comprimir la representación en el dominio de la frecuencia del componente de campo de sonido para transmisión o almacenamiento u otras tareas de procesamiento.

Realizaciones preferidas

[0056] La fig. 2a muestra la presente estrategia innovadora, dada por el Bloque (10), que permite sintetizar un componente Ambisonics de un orden (nivel) y modo deseados de las señales de múltiples (dos o más) micrófonos. A diferencia de las estrategias del estado de la técnica relacionadas, no se hace ninguna restricción para la configuración del micrófono. Esto significa que los múltiples micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.

[0057] Para obtener el componente Ambisonics deseado, las múltiples señales de micrófono primero se transforman en una representación de tiempo-frecuencia utilizando el bloque (101). Para este propósito, se puede utilizar, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT, Short-Time Fourier Transform). La salida del bloque (101) son las múltiples señales de micrófono en el dominio del tiempo-frecuencia. Cabe observar que el siguiente procesamiento es llevado a cabo de forma independiente para los cuadros de tiempo-frecuencia.

[0058] Después de transformar las múltiples señales de micrófono en el dominio del tiempo-frecuencia, determinamos una o más direcciones de sonido (para un cuadro de tiempo-frecuencia) en el Bloque (102) de dos o más señales de micrófono. Una dirección de sonido describe a partir de qué dirección un sonido prominente para un cuadro de tiempo-frecuencia está llegando en la disposición de micrófono. Esta dirección generalmente se denomina como la dirección de llegada (DOA) del sonido. Alternativamente a la DOA, se podría considerar también la dirección de propagación del sonido, que es la operación opuesta de la DOA, o cualquier otra medida que describa la dirección del sonido. Dichas una o múltiples direcciones de sonido o DOAs se estiman en el Bloque (102) utilizando, por ejemplo, estimadores de DOA de banda estrecha del estado de la técnica, los cuales están disponibles para casi cualquier configuración de micrófono. Estimadores de DOA ejemplares adecuados se enumeran en la Realización 1. El número de direcciones de sonido o DOAs (una o más) las cuales se calculan en el bloque (102), depende, por ejemplo, de la complejidad computacional tolerable, pero también de las capacidades del estimador de DOA utilizado o la geometría

del micrófono. Una dirección de sonido se puede estimar como, por ejemplo, en el espacio 2D (representado, por ejemplo, en la forma de un ángulo de azimut) o en el espacio 3D (representado, por ejemplo, en la forma de un ángulo de azimut y un ángulo de elevación). En lo sucesivo, la mayoría de las descripciones se basan en el caso más general de 3D, aunque es muy sencillo aplicar todas las etapas de procesamiento al caso 2D también. En muchos casos, el usuario especifica cómo se estiman muchas direcciones de sonido o DOAs (por ejemplo, 1, 2 o 3) por cuadro de tiempo-frecuencia estimado. Alternativamente, el número de sonidos prominentes se puede estimar utilizando estrategias del estado de la técnica, por ejemplo, las estrategias explicadas en [SourceNum].

[0059] Dichas una o más direcciones de sonido, que se estimaron en el Bloque (102) para un cuadro de tiempo-frecuencia, se utilizan en el bloque (103) para calcular, para el cuadro de tiempo-frecuencia, una o más respuestas de una función de base espacial del orden (nivel) y modo deseados. Se calcula una respuesta para cada dirección de sonido estimada. Como se explicó en la sección previa, una función de base espacial puede representar, por ejemplo, una armónica esférica (por ejemplo, si el procesamiento se lleva a cabo en el espacio 3D) o una armónica cilíndrica (por ejemplo, si el procesamiento se lleva a cabo en el espacio 2D). La respuesta de una función de base espacial es la función de base espacial evaluada en la dirección de sonido estimada correspondiente, como se explica con mayor detalle en la primera realización.

[0060] Dichas una o más direcciones de sonido, las cuales se estiman para un cuadro de tiempo-frecuencia, se utilizan adicionalmente en el Bloque (201), es decir para calcular, para el cuadro de tiempo-frecuencia, uno o más componentes Ambisonics del orden (nivel) y modo deseados. Tal componente Ambisonics sintetiza un componente Ambisonics para un sonido direccional que llega de la dirección de sonido estimada. Entrada adicional al Bloque (201) son dichas una o más respuestas de la función de base espacial que fueron calculadas para el cuadro de tiempo-frecuencia en el Bloque (103), así como una o más señales de micrófono para el cuadro de tiempo-frecuencia dado. En el Bloque (201) se calcula un componente Ambisonics del orden (nivel) y modo deseados para cada dirección de sonido estimada y respuesta correspondiente de la función de base espacial. Las etapas de procesamiento del Bloque (201) se discuten adicionalmente en las siguientes realizaciones.

[0061] La presente invención (10) contiene un Bloque (301) opcional que puede calcular, para un cuadro de tiempo-frecuencia, un componente Ambisonics de sonido difuso del orden (nivel) y modo deseados. Este componente sintetiza un componente Ambisonics, por ejemplo, para un campo de sonido puramente difuso o para sonido ambiente. La entrada al Bloque (301) son dichas una o más direcciones de sonido, las cuales fueron estimadas en el Bloque (102), así como una o más señales de micrófono. Las etapas de procesamiento del Bloque (301) se discuten adicionalmente en las realizaciones posteriores.

[0062] Los componentes Ambisonics de sonido difuso, los cuales se calculan en el Bloque (301) opcional, se pueden decorrelacionar adicionalmente en el Bloque (107) opcional. Para este propósito, se pueden utilizar decorreladores del estado de la técnica. Algunos ejemplos se enumeran en la Realización 4. Generalmente, se aplicarían diferentes decorreladores o diferentes realizaciones de un decorrelador para diferentes órdenes (niveles) y modos. Al hacerlo, los componentes Ambisonics de sonido difuso decorrelacionados de diferentes órdenes (niveles) y modos serán mutuamente decorrelacionados. Esto imita el comportamiento físico esperado, es decir que los componentes Ambisonics de diferentes órdenes (niveles) y modos son mutuamente decorrelacionados para sonidos difusos o sonidos ambientales, como se explica, por ejemplo, en [SpCoherence].

[0063] Dichos uno o más componentes Ambisonics (de sonido directo) del orden (nivel) y modo deseados, que se calcularon para un cuadro de tiempo-frecuencia en el Bloque (201), y el componente Ambisonics de sonido difuso correspondiente que fue calculado en el Bloque (301), se combinan en el Bloque (401). Como se discute en las Realizaciones posteriores, la combinación se puede realizar, por ejemplo, como una suma (ponderada). La salida del Bloque (401) es el componente Ambisonics sintetizado final del orden (nivel) y modo deseados para un cuadro de tiempo-frecuencia dado. Claramente, si solamente se calculó un solo componente Ambisonics (de sonido directo) del orden (nivel) y modo deseados en el Bloque (201) para un cuadro de tiempo-frecuencia (y ningún componente Ambisonics de sonido difuso), entonces el combinador (401) es superfluo.

[0064] Después de calcular el componente Ambisonics final del orden (nivel) y modo deseados para todos los cuadros de tiempo-frecuencia, el componente Ambisonics se puede transformar de vuelta al dominio del tiempo con la transformada de tiempo-frecuencia inversa (20), la cual se puede realizar, por ejemplo, como un banco de filtros inverso o una STFT inversa. Cabe observar que la transformada de tiempo-frecuencia inversa no es requerida en cada aplicación, y por lo tanto, no es parte de la presente invención. En la práctica, se calcularían los componentes Ambisonics para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado.

[0065] La fig. 2b muestra una realización ligeramente modificada de la misma presente invención. En esta figura, se aplica la transformada de tiempo-frecuencia inversa (20) antes del combinador (401). Esto es posible ya que la transformada de tiempo-frecuencia inversa es generalmente una transformación lineal. Al aplicar la transformada de tiempo-frecuencia inversa antes del combinador (401), es posible, por ejemplo, llevar a cabo la decorrelación en el dominio del tiempo (en lugar del dominio del tiempo-frecuencia como en la fig. 2a). Esto puede tener ventajas prácticas

para algunas aplicaciones cuando se implementa la invención.

[0066] Se debe observar que el banco de filtros inverso también puede estar en alguna otra parte. Generalmente, el combinador y el decorrelador se deben (y generalmente el último) aplicar en el dominio del tiempo.
5 Pero, también se pueden aplicar ambos o solamente un bloque en el dominio de la frecuencia.

[0067] Realizaciones preferidas comprenden, por lo tanto, un calculador de componente difuso 301 para calcular, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de sonido difuso. Además, tales realizaciones comprenden un combinador 401 para combinar
10 información de sonido difuso e información de campo de sonido directo para obtener una representación en el dominio de la frecuencia o una representación en el dominio del tiempo de los componentes de campo de sonido. Además, dependiendo de la implementación, el calculador de componente difuso además comprende un decorrelador 107 para decorrelacionar la información de sonido difuso, donde el decorrelador se puede implementar dentro del dominio de la frecuencia de tal manera que la correlación se lleve a cabo con la representación del cuadro de tiempo-frecuencia
15 del componente de sonido difuso. Alternativamente, el decorrelador está configurado para operar dentro del dominio del tiempo como se ilustra en la fig. 2b de tal manera que se lleve a cabo una decorrelación dentro del dominio del tiempo de la representación del tiempo de un cierto componente de sonido difuso de un cierto orden.

[0068] Realizaciones adicionales relacionadas con la presente invención comprenden un convertidor de tiempo-frecuencia tal como el convertidor de tiempo-frecuencia 101 para convertir cada una de la pluralidad de señales de micrófono en el dominio del tiempo en una representación de frecuencia que tiene la pluralidad de cuadros de tiempo-frecuencia. Realizaciones adicionales comprenden convertidores de frecuencia-tiempo tal como el Bloque 20 de la fig. 2a o la fig. 2b para convertir dichos uno o más componentes de campo de sonido o una combinación de dichos uno o más componentes de campo de sonido, es decir, los componentes de campo de sonido directo y los
25 componentes de sonido difuso en una representación en el dominio del tiempo del componente de campo de sonido.

[0069] En particular, el convertidor de frecuencia-tiempo 20 está configurado para procesar dichos uno o más componentes de campo de sonido para obtener una pluralidad de componentes de campo de sonido en el dominio del tiempo donde estos componentes de campo de sonido en el dominio del tiempo son los componentes de campo de sonido directo. Además, el convertidor de frecuencia-tiempo 20 está configurado para procesar los componentes (de campo) de sonido difuso para obtener una pluralidad de componentes (de campo de sonido) difusos en el dominio del tiempo y el combinador está configurado para llevar a cabo la combinación de los componentes de campo de sonido (directo) en el dominio del tiempo y los (componentes de campo de sonido) difusos en el dominio del tiempo como se ilustra, por ejemplo, en la fig. 2b. Alternativamente, el combinador 401 está configurado para combinar dichos
30 uno o más componentes de campo de sonido (directo) para un cuadro de tiempo-frecuencia y los componentes (de campo) de sonido difuso para el cuadro de tiempo-frecuencia correspondiente dentro del dominio de la frecuencia, y el convertidor de frecuencia-tiempo 20 está configurado entonces para procesar un resultado del combinador 401 para obtener los componentes de campo de sonido en el dominio del tiempo, es decir, la representación de los componentes de campo de sonido en el dominio del tiempo como se ilustra, por ejemplo, en la fig. 2a.
40

[0070] Las siguientes realizaciones describen con mayor detalle varias realizaciones de la presente invención. Cabe observar que las Realizaciones 1-7 consideran una dirección de sonido por cuadro de tiempo-frecuencia (y por lo tanto, solamente una respuesta de una función de base espacial y solamente un componente Ambisonics de sonido directo por nivel y modo y tiempo y frecuencia). La Realización 8 describe un ejemplo donde se considera más de una
45 dirección de sonido por cuadro de tiempo-frecuencia. El concepto de esta realización se puede aplicar de una manera directa a todas las demás realizaciones.

Realización 1

[0071] La figura 3a muestra una realización de la invención que permite sintetizar un componente Ambisonics de un orden (nivel) l y modo m deseados de las señales de múltiples (dos o más) micrófonos.
50

[0072] La entrada para la invención son las señales de múltiples (dos o más) micrófonos. Los micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.
55

[0073] Las múltiples señales de micrófono se transforman en el dominio del tiempo-frecuencia en el Bloque (101) utilizando, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT). La salida de la transformada de tiempo-frecuencia (101) son las múltiples señales de micrófono en el dominio del tiempo-frecuencia, que son denotadas por $P1...M(k,n)$, donde k es el índice de frecuencia, n es el índice de tiempo, y M es el número de micrófonos. Cabe observar que el siguiente procesamiento se lleva a cabo por separado para los cuadros de tiempo-frecuencia (k,n) .
60

[0074] Después de transformar las señales de micrófono en el dominio del tipo frecuencia, se lleva a cabo una
65

estimación de dirección de sonido en el Bloque (102) por tiempo y frecuencia utilizando dos o más de las señales de micrófono P1...M(k,n). En esta realización, se determina una sola dirección de sonido por tiempo y frecuencia. Para la estimación de dirección de sonido en (102) se pueden utilizar estimadores de dirección de llegada (DOA) de banda estrecha del estado de la técnica, los cuales están disponibles en la bibliografía para diferentes geometrías de disposición de micrófono. Por ejemplo, se puede utilizar el algoritmo MUSIC [MUSIC] el cual es aplicable para configuraciones de micrófono arbitrarias. En el caso de disposiciones lineales uniformes, disposiciones lineales no uniformes con puntos de retícula equidistantes, o disposiciones circulares de micrófonos omnidireccionales, se puede aplicar el algoritmo Root MUSIC [RootMUSIC1,RootMUSIC2,RootMUSIC3] que es computacionalmente más eficiente que MUSIC. Otro estimador de DOA de banda estrecha bien conocido, que se puede aplicar a disposiciones lineales o disposiciones planas con estructura de sub-disposición rotacionalmente invariante es ESPRIT [ESPRIT].

[0075] En esta realización, la salida del estimador de dirección de sonido (102) es una dirección de sonido para un caso de tiempo n e índice de frecuencia k . La dirección de sonido se puede expresar por ejemplo, en términos de un vector normal unitario $n(k,n)$ o en términos de un ángulo de azimut $\varphi(k,n)$ y/o ángulo de elevación $\vartheta(k,n)$, los cuales se relacionan, por ejemplo, como:

$$n(k, n) = \begin{bmatrix} \cos\varphi(k, n)\cos\vartheta(k, n) \\ \sin\varphi(k, n)\cos\vartheta(k, n) \\ \sin\vartheta(k, n) \end{bmatrix}.$$

[0076] Si no se estima en un ángulo de elevación (k,n) , podemos suponer elevación cero, es decir, $\vartheta(k,n) = 0$, en las siguientes etapas. En este caso, el vector normal unitario $n(k,n)$ se puede escribir como:

$$n(k, n) = \begin{bmatrix} \cos\varphi(k, n) \\ \sin\varphi(k, n) \end{bmatrix}.$$

[0077] Después de la estimación de la dirección de sonido en el Bloque (102), se determina una respuesta de una función de base espacial del orden (nivel) l y modo m deseados en el Bloque (103) individualmente por tiempo y frecuencia utilizando la información de dirección de sonido estimada. La respuesta de una función de base espacial del orden (nivel) l y modo m se denota por $G_l^m(k, n)$ y se calcula como:

$$G_l^m(k, n) = Y_l^m(\varphi, \vartheta).$$

[0078] Aquí, $Y_l^m(\varphi, \vartheta)$ es una función de base espacial del orden (nivel) l y modo m que depende de la dirección indicada por el vector $n(k,n)$ o el ángulo de azimut $\varphi(k,n)$ y/o el ángulo de elevación $\vartheta(k,n)$. Por lo tanto, la respuesta $G_l^m(k, n)$ describe la respuesta de una función de base espacial $Y_l^m(\varphi, \vartheta)$ para un sonido que llega de la dirección indicada por el vector $n(k,n)$ o el ángulo de azimut $\varphi(k,n)$ y/o el ángulo de elevación $\vartheta(k,n)$. Por ejemplo, cuando se consideran armónicas esféricas de valor real con normalización N3D como función de base espacial, $Y_l^m(\varphi, \vartheta)$ se puede calcular como [SphHarm,Ambix,FourierAcoust]

$$Y_l^m(\varphi, \vartheta) = \begin{cases} \sqrt{2}K_l^m \cos(m\varphi)L_l^m(\cos\vartheta)\sin > 0 \\ K_l^m L_l^m(\cos\vartheta)\sin = 0 \\ \sqrt{2}K_l^m \sin(-m\varphi)L_l^{-m}(\cos\vartheta)\sin < 0 \end{cases}$$

40 donde

$$K_l^m = \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}}$$

son las constantes de normalización N3D y $L_l^m(\cos\vartheta)$ es el polinomio de Legendre asociado del orden (nivel) l y modo m dependiendo del ángulo de elevación, el cual se define, por ejemplo, en [FourierAcoust]. Cabe observar que la respuesta de la función de base espacial Y_l^m del orden (nivel) l y modo m deseados también se pueden pre-calcular para cada ángulo de azimut y/o elevación y almacenar en una tabla de búsqueda y después seleccionar dependiendo de la dirección de sonido estimada.

[0079] En esta realización, sin pérdida de generalidad, la primera señal de micrófono se denomina como la señal de micrófono de referencia $P_{ref}(k,n)$, es decir,

$$P_{ref}(k, n) = P_1(k, n).$$

[0080] En esta realización, la señal de micrófono de referencia $P_{ref}(k, n)$ se combina tal como multiplicada 115 por el cuadro de tiempo-frecuencia (k, n) con la respuesta $G_l^m(k, n)$ de la función de base espacial determinada en el Bloque (103), es decir,

$$B_l^m(k, n) = P_{ref}(k, n)G_l^m(k, n),$$

que resulta en el componente Ambisonics $B_l^m(k, n)$ deseado del orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k, n) . Los componentes Ambisonics $B_l^m(k, n)$ resultantes eventualmente se pueden transformar de nuevo al dominio del tiempo utilizando un banco de filtros inverso o una STFT inversa, almacenar, transmitir, o utilizar, por ejemplo, para aplicaciones de reproducción de sonido espacial. En la práctica, se calcularían los componentes Ambisonics $B_l^m(k, n)$ para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado.

15

Realización 2

[0081] La figura 3b muestra otra realización de la invención que permite sintetizar un componente Ambisonics de un orden (nivel) l y modo m deseados a partir de las señales de múltiples (dos o más) micrófonos. La realización es similar a la Realización 1, pero adicionalmente contiene un Bloque (104) para determinar la señal de micrófono de referencia de la pluralidad de señales de micrófono.

[0082] Como en la realización 1, la entrada a la invención son las señales de múltiples (dos o más) micrófonos. Los micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.

[0083] Como en la Realización 1, las múltiples señales de micrófono se transforman en el dominio del tiempo-frecuencia en el Bloque (101) utilizando, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT). La salida de la transformada de tiempo-frecuencia (101) son las señales de micrófono en el dominio del tiempo-frecuencia, que son denotadas por $P_1 \dots M(k, n)$. El siguiente procesamiento se lleva a cabo por separado para los cuadros de tiempo-frecuencia (k, n) .

[0084] Como en la Realización 1, se lleva a cabo una estimación de dirección de sonido en el Bloque (102) por tiempo y frecuencia utilizando dos o más de las señales de micrófono $P_1 \dots M(k, n)$. Los estimadores correspondientes se discuten en la Realización 1. La salida del estimador de dirección de sonido (102) es una dirección de sonido para un caso de tiempo n e índice de frecuencia k . La dirección de sonido se puede expresar por ejemplo, en términos de un vector normal unitario $n(k, n)$ o en términos de un ángulo de azimut $\varphi(k, n)$ y/o ángulo de elevación $\vartheta(k, n)$, los cuales se relacionan como se explicó en la Realización 1.

40

[0085] Como en la Realización 1, la respuesta de una función de base espacial del orden (nivel) l y modo m deseados se determina en el Bloque (103) por tiempo y frecuencia utilizando la información de dirección de sonido estimada. La respuesta de la función de base espacial se denota por $G_l^m(k, n)$. Por ejemplo, podemos considerar armónicas esféricas de valor real con normalización N3D como función de base espacial y se puede determinar cómo se explicó en la Realización 1.

45

[0086] En esta Realización, una señal de micrófono de referencia $P_{ref}(k, n)$ se determina a partir de las múltiples señales de micrófono $P_1 \dots M(k, n)$ en el Bloque (104). Para este propósito, el Bloque (104) utiliza la información de dirección de sonido que se estimó en el Bloque (102). Diferentes señales de micrófonos de referencia se pueden determinar para diferentes cuadros de tiempo-frecuencia. Existen diferentes posibilidades para determinar la señal de micrófono de referencia $P_{ref}(k, n)$ a partir de las múltiples señales de micrófono $P_1 \dots M(k, n)$ con base en la información de dirección de sonido. Por ejemplo, se puede seleccionar por tiempo y frecuencia el micrófono de los múltiples micrófonos que está más cercano a la dirección de sonido estimada. Esta estrategia se visualiza en la fig. 1b. Por ejemplo, suponiendo que las posiciones de micrófono están dadas por los vectores de posición $d_1 \dots M$, el índice $i(k, n)$ del micrófono más cercano se puede encontrar resolviendo el problema

55

$$i(k, n) = \arg \min_{j \in [1, M]} \|d_j - n(k, n)\|$$

de tal manera que la señal de micrófono de referencia para el tiempo y frecuencia considerados está dada por

60

$$P_{ref}(k, n) = P_{i(k, n)}(k, n).$$

[0087] En el ejemplo en la fig. 1b, el micrófono de referencia para el cuadro de tiempo-frecuencia (k,n) sería el micrófono número 3, es decir, $i(k,n) = 3$, ya que d_3 está más cercana a $n(k,n)$. Una estrategia alternativa para determinar la señal de micrófono de referencia $P_{ref}(k,n)$ es aplicar un filtro multicanal a las señales de micrófono, es decir,

$$P_{ref}(k,n) = w^H(n)p(k,n),$$

donde $w(n)$ es el filtro multicanal que depende de la dirección de sonido estimada y el vector $p(k,n) = [P_1(k,n), \dots, P_M(k,n)]^T$ contiene las múltiples señales de micrófono. Existen muchos filtros multicanal $w(n)$ óptimos diferentes en la bibliografía que se pueden utilizar para calcular $P_{ref}(k,n)$, por ejemplo el filtro delay&sum o el filtro LCMV, los cuales se derivan, por ejemplo, en [OptArrayPr]. Utilizar los filtros multicanal proporciona diferentes ventajas y desventajas que se explican en [OptArrayPr], por ejemplo, nos permiten reducir el ruido propio del micrófono.

[0088] Como en la Realización 1, la señal de micrófono de referencia $P_{ref}(k,n)$ finalmente se combina tal como multiplicada 115 por el tiempo y la frecuencia con la respuesta $G_i^m(k,n)$ de la función de base espacial determinada en el Bloque (103), es decir, que resulta en el componente Ambisonics $B_i^m(k,n)$ deseado del orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k,n) . Los componentes Ambisonics $B_i^m(k,n)$ resultantes eventualmente se pueden transformar de nuevo al dominio del tiempo utilizando un banco de filtros inverso o una STFT inversa, almacenar, transmitir, o utilizar, por ejemplo, para reproducción de sonido espacial. En la práctica, se calcularían los componentes Ambisonics para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado.

Realización 3

[0089] La figura 4 muestra otra realización de la invención que permite sintetizar un componente Ambisonics de un orden (nivel) l y modo m deseados a partir de las señales de múltiples (dos o más) micrófonos. La realización es similar a la Realización 1, pero calcula los componentes Ambisonics para una señal de sonido directo y una señal de sonido difuso.

[0090] Como en la realización 1, la entrada a la invención son las señales de múltiples (dos o más) micrófonos. Los micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.

[0091] Como en la realización 1, las múltiples señales de micrófono se transforman en el dominio del tiempo-frecuencia en el Bloque (101) utilizando, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT). La salida de la transformada de tiempo-frecuencia (101) son las señales de micrófono en el dominio del tiempo-frecuencia, que son denotadas por $P_1 \dots M(k,n)$. El siguiente procesamiento se lleva a cabo por separado para los cuadros de tiempo-frecuencia (k,n) .

[0092] Como en la realización 1, se lleva a cabo una estimación de dirección de sonido en el Bloque (102) por tiempo y frecuencia utilizando dos o más de las señales de micrófono $P_1 \dots M(k,n)$. Los estimadores correspondientes se discuten en la Realización 1. La salida del estimador de dirección de sonido (102) es una dirección de sonido para un caso de tiempo n e índice de frecuencia k . La dirección de sonido se puede expresar por ejemplo, en términos de un vector normal unitario $n(k,n)$ o en términos de un ángulo de azimut $\varphi(k,n)$ y/o ángulo de elevación $\vartheta(k,n)$, los cuales se relacionan como se explicó en la Realización 1.

[0093] Como en la Realización 1, la respuesta de una función de base espacial del orden (nivel) l y modo m deseados se determina en el Bloque (103) por tiempo y frecuencia utilizando la información de dirección de sonido estimada. La respuesta de la función de base espacial se denota por $G_i^m(k,n)$. Por ejemplo, podemos considerar armónicas esféricas de valor real con normalización N3D como función de base espacial y $G_i^m(k,n)$ se puede determinar cómo se explicó en la Realización 1.

[0094] En esta Realización, una respuesta promedio de una función de base espacial del orden (nivel) l y modo m deseados, que es independiente del índice de tiempo n , se obtiene a partir del Bloque (106). Esta respuesta promedio se denota por $D_i^m(k)$ y describe la respuesta de una función de base espacial para los sonidos que llegan de todas las direcciones posibles (tal como sonidos difusos o sonidos ambientales). Un ejemplo para definir la respuesta promedio $D_i^m(k)$ es considerar la integral de la magnitud cuadrada de la función de base espacial $Y_i^m(\varphi, \vartheta)$ sobre todos los ángulos φ y/o ϑ posibles. Por ejemplo, cuando se integra a través de todos los ángulos en una esfera, obtenemos

$$D_l^m(k) = \int_0^{2\pi} \int_0^\pi |Y_l^m(\varphi, \vartheta)|^2 \sin\vartheta d\vartheta d\varphi.$$

[0095] Tal definición de la respuesta promedio $D_l^m(k)$ se puede interpretar del siguiente modo: como se explicó en la Realización 1, la función de base espacial $Y_l^m(\varphi, \vartheta)$ se puede interpretar como la directividad de un micrófono de orden l . Para órdenes mayores, tal micrófono se haría cada vez más directivo, y por lo tanto, se capturaría menos energía de sonido difuso o energía de sonido ambiental en un campo de sonido práctico en comparación con un micrófono omnidireccional (micrófono de orden $l = 0$). Con la definición de $D_l^m(k)$ dada anteriormente, la respuesta promedio $D_l^m(k)$ resultaría en un factor de valor real que describe por cuánto se atenúa la energía de sonido difuso o energía de sonido ambiental en la señal de un micrófono de orden l en comparación con un micrófono omnidireccional. Claramente, además de integrar la magnitud cuadrada de la función de base espacial $Y_l^m(\varphi, \vartheta)$ a través de las direcciones de una esfera, existen diferentes alternativas para definir la respuesta promedio $D_l^m(k)$, por ejemplo: integrar la magnitud cuadrada de $Y_l^m(\varphi, \vartheta)$ a través de las direcciones en un círculo, integrar la magnitud cuadrada de $Y_l^m(\varphi, \vartheta)$ a través de cualquier conjunto de direcciones (φ, ϑ) deseadas, promediar la magnitud cuadrada de $Y_l^m(\varphi, \vartheta)$ a través de cualquier conjunto de direcciones (φ, ϑ) deseadas, integrar o promediar la magnitud de $Y_l^m(\varphi, \vartheta)$ en lugar de la magnitud cuadrada, considerando una suma ponderada de $Y_l^m(\varphi, \vartheta)$ a través de cualquier conjunto de direcciones (φ, ϑ) deseadas, o especificar cualquier número de valor real deseado para $D_l^m(k)$ que corresponda a la sensibilidad deseada del micrófono imaginado antes mencionado de orden l con respecto a los sonidos difusos o sonidos ambientales.

[0096] La respuesta de función de base espacial promedio también se puede pre-calcular y almacenar en una tabla de búsqueda y la determinación de los valores de respuesta se lleva a cabo accediendo a la tabla de búsqueda y recuperando el valor correspondiente.

[0097] Como en la Realización 1, sin pérdida de generalidad, la primera señal de micrófono se denomina como la señal de micrófono de referencia $P_{ref}(k, n) = P_1(k, n)$.

[0098] En esta realización, la señal de micrófono de referencia $P_{ref}(k, n)$ se utiliza en el Bloque (105) para calcular una señal de sonido directo denotada por $P_{dir}(k, n)$ y una señal de sonido difuso denotada por $P_{diff}(k, n)$. En el Bloque (105), la señal de sonido directo $P_{dir}(k, n)$ se puede calcular, por ejemplo, aplicando un filtro de un solo canal $W_{dir}(k, n)$ a la señal de micrófono de referencia, es decir,

$$P_{dir}(k, n) = W_{dir}(k, n)P_{ref}(k, n).$$

[0099] Existen diferentes posibilidades en la bibliografía para calcular un filtro de un solo canal $W_{dir}(k, n)$ óptimo. Por ejemplo, se puede utilizar el filtro de Wiener de raíz cuadrada bien conocido, el cual se define en, por ejemplo, [VirtualMic] como

$$W_{dir}(k, n) = \sqrt{\frac{SDR(k, n)}{SDR(k, n) + 1}}$$

donde $SDR(k, n)$ es la relación señal-a-difuso (SDR, Signal-to-Diffuse Ratio) en un caso de tiempo n e índice de frecuencia k que describe la relación de potencia entre el sonido directo y el sonido difuso como se discute en [VirtualMic]. La SDR se puede estimar utilizando cualquier par de micrófonos de las múltiples señales de micrófono $P_1 \dots P_M(k, n)$ con un estimador de SDR del estado de la técnica disponible en la bibliografía, por ejemplo, los estimadores propuestos en [SDRstim] los cuales se basan en la coherencia espacial entre dos señales de micrófono arbitrarias. En el Bloque (105), se puede calcular la señal de sonido difuso $P_{diff}(k, n)$, por ejemplo, aplicando un filtro de un solo canal $W_{diff}(k, n)$ a la señal de micrófono de referencia, es decir,

$$P_{diff}(k, n) = W_{diff}(k, n)P_{ref}(k, n).$$

[0100] Existen diferentes posibilidades en la bibliografía para calcular un filtro de un solo canal $W_{diff}(k, n)$ óptimo. Por ejemplo, se puede utilizar el filtro de Wiener de raíz cuadrada bien conocido, el cual se define en, por ejemplo, [VirtualMic] como

$$W_{diff}(k, n) = \sqrt{\frac{1}{SDR(k, n) + 1}}$$

donde $SDR(k, n)$ es la SDR que se puede estimar como se discutió anteriormente.

- 5 **[0101]** En esta realización, la señal de sonido directo $P_{dir}(k, n)$ determinada en el Bloque (105) se combina tal como multiplicada 115a por tiempo y frecuencia con la respuesta $G_l^m(k, n)$ de la función de base espacial determinada en el Bloque (103), es decir,

$$B_{dir,l}^m(k, n) = P_{dir}(k, n)G_l^m(k, n),$$

10

que resulta en un componente Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ de orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k, n) . Por otra parte, la señal de sonido difuso $P_{diff}(k, n)$ determinada en el Bloque (105) se combina tal como multiplicada 115b por tiempo y frecuencia con la respuesta promedio $D_l^m(k)$ de la función de base espacial determinada en el Bloque (106), es decir,

15

$$B_{diff,l}^m(k, n) = P_{diff}(k, n)D_l^m(k),$$

que resulta en un componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ de orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k, n) .

20

- [0102]** Finalmente, el componente Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ se combinan, por ejemplo, por medio de la operación de suma (109), para obtener el componente Ambisonics final $B_l^m(k, n)$ del orden (nivel) l y modo m deseados para el cuadro de tiempo-frecuencia (k, n) , es decir,

25

$$B_l^m(k, n) = B_{dir,l}^m(k, n) + B_{diff,l}^m(k, n).$$

- [0103]** Los componentes Ambisonics $B_l^m(k, n)$ resultantes eventualmente se pueden transformar de nuevo al dominio del tiempo utilizando un banco de filtros inverso o una STFT inversa, almacenar, transmitir, o utilizar, por ejemplo, para reproducción de sonido espacial. En la práctica, se calcularían los componentes Ambisonics para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado.

30

- [0104]** Es importante enfatizar que la transformación de vuelta al dominio del tiempo utilizando, por ejemplo, un banco de filtros inverso o una STFT inversa se puede llevar a cabo antes del cálculo de $B_l^m(k, n)$, antes de la operación (109). Esto significa, que primero podemos transformar $B_{dir,l}^m(k, n)$ y $B_{diff,l}^m(k, n)$ de vuelta al dominio del tiempo y después sumar ambos componentes con la operación (109) para obtener el componente Ambisonics final B_l^m . Esto es posible ya que el banco de filtros inverso o la STFT inversa son en general operaciones lineales.

35

- [0105]** Cabe observar que el algoritmo en esta realización se puede configurar de tal manera que los componentes Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ se calculan para diferentes modos (órdenes) l . Por ejemplo, $B_{dir,l}^m(k, n)$ se puede calcular hasta el orden $l = 4$, mientras que $B_{diff,l}^m(k, n)$ se puede calcular solamente hasta el orden $l = 1$ (en este caso, $B_{diff,l}^m(k, n)$ sería cero para órdenes mayores que $l = 1$). Esto tiene ventajas específicas que se explican en la Realización 4. Si se desea, por ejemplo, calcular solamente $B_{dir,l}^m(k, n)$ pero no $B_{diff,l}^m(k, n)$ para un orden (nivel) l o modo m específicos, entonces, por ejemplo, el Bloque (105) se puede configurar de tal manera que la señal de sonido difuso $P_{diff}(k, n)$ se vuelve igual que cero. Esto se puede lograr, por ejemplo, ajustando el filtro $W_{diff}(k, n)$ en las ecuaciones previas a 0 y el filtro $W_{dir}(k, n)$ a 1. Alternativamente, se podría establecer anualmente la SDR en las ecuaciones previas en un valor muy alto.

45

50 Realización 4

[0106] La figura 5 muestra otra realización de la invención que permite sintetizar un componente Ambisonics de un orden (nivel) l y modo m deseados a partir de las señales de múltiples (dos o más) micrófonos. La realización es similar a la Realización 3, pero adicionalmente contiene decorreladores para los componentes Ambisonics difusos.

5 **[0107]** Como en la realización 3, la entrada a la invención son las señales de múltiples (dos o más) micrófonos. Los micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.

10 **[0108]** Como en la Realización 3, las múltiples señales de micrófono se transforman en el dominio del tiempo-frecuencia en el Bloque (101) utilizando, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT). La salida de la transformada de tiempo-frecuencia (101) son las señales de micrófono en el dominio del tiempo-frecuencia, que son denotadas por $P_1 \dots M(k, n)$. El siguiente procesamiento es llevado a cabo por separado para los cuadros de tiempo-frecuencia (k, n) .

15 **[0109]** Como en la Realización 3, se lleva a cabo una estimación de dirección de sonido en el Bloque (102) por tiempo y frecuencia utilizando dos o más de las señales de micrófono $P_1 \dots M(k, n)$. Los estimadores correspondientes se discuten en la Realización 1. La salida del estimador de dirección de sonido (102) es una dirección de sonido para un caso de tiempo n e índice de frecuencia k . La dirección de sonido se puede expresar por ejemplo, en términos de un vector normal unitario $\hat{n}(k, n)$ o en términos de un ángulo de azimut $\varphi(k, n)$ y/o ángulo de elevación $\vartheta(k, n)$, los cuales se relacionan como se explicó en la Realización 1.

25 **[0110]** Como en la Realización 3, la respuesta de una función de base espacial del orden (nivel) l y modo m deseados se determina en el Bloque (103) por tiempo y frecuencia utilizando la información de dirección de sonido estimada. La respuesta de la función de base espacial se denota por $G_l^m(k, n)$. Por ejemplo, podemos considerar armónicas esféricas de valor real con normalización N3D como función de base espacial y $G_l^m(k, n)$ se puede determinar como se explicó en la Realización 1.

30 **[0111]** Como en la Realización 3, una respuesta promedio de una función de base espacial del orden (nivel) l y modo m deseados, que es independiente del índice de tiempo n , se obtiene a partir del Bloque (106). Esta respuesta promedio se denota por $D_l^m(k)$ y describe la respuesta de una función de base espacial para los sonidos que llegan de todas las direcciones posibles (tal como sonidos difusos o sonidos ambientales). La respuesta promedio $D_l^m(k)$ se puede obtener como se describe en la Realización 3.

35 **[0112]** Como en la Realización 3, sin pérdida de generalidad, la primera señal de micrófono se denomina como la señal de micrófono de referencia $P_{ref}(k, n) = P_1(k, n)$.

40 **[0113]** Como en la Realización 3, la señal de micrófono de referencia $P_{ref}(k, n)$ se utiliza en el Bloque (105) para calcular una señal de sonido directo denotada por $P_{dir}(k, n)$ y una señal de sonido difuso denotada por $P_{diff}(k, n)$. El cálculo de $P_{dir}(k, n)$ y $P_{diff}(k, n)$ se explica en la Realización 3.

[0114] Como en la Realización 3, la señal de sonido directo $P_{dir}(k, n)$ determinada en el Bloque (105) se combina tal como multiplicada 115a por tiempo y frecuencia con la respuesta $G_l^m(k, n)$ de la función de base espacial determinada en el Bloque (103) que resulta en un componente Ambisonics de sonido directo $B_{dir, l}^m(k, n)$ de orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k, n) . Por otra parte, la señal de sonido difuso $P_{diff}(k, n)$ determinada en el Bloque (105) se combina tal como multiplicada 115b por tiempo y frecuencia con la respuesta promedio $D_l^m(k)$ de la función de base espacial determinada en el Bloque (106) que resulta en un componente Ambisonics de sonido difuso $B_{diff, l}^m(k, n)$ de orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k, n) .

50 **[0115]** En esta realización, el componente Ambisonics de sonido difuso $B_{diff, l}^m(k, n)$ calculado se decorrelaciona en el Bloque (107) utilizando un decorrelador que resulta en un componente Ambisonics de sonido difuso decorrelacionado, denotado por $\mathcal{B}_{diff, l}^m(k, n)$. Para la decorrelación se pueden utilizar técnicas de decorrelación del estado de la técnica. Generalmente se aplican diferentes decorreladores o realizaciones del decorrelador al componente Ambisonics de sonido difuso $B_{diff, l}^m(k, n)$ de orden (nivel) l y modo m diferentes de tal manera que los componentes Ambisonics de sonido difuso $\mathcal{B}_{diff, l}^m(k, n)$ decorrelacionados resultantes de diferente nivel y modo son mutuamente no correlacionados. Al hacerlo, los componentes Ambisonics de sonido difuso $\mathcal{B}_{diff, l}^m(k, n)$ poseen el comportamiento físico esperado, es decir que los componentes Ambisonics de diferentes órdenes y modos son mutuamente no correlacionados si el campo de sonido es ambiental o difuso [SpCoherence]. Cabe observar que el componente Ambisonics de sonido difuso $B_{diff, l}^m(k, n)$ se puede transformar de nuevo al dominio del tiempo utilizando, por ejemplo, un filtro de bancos inverso o una STFT inversa antes de aplicar el

decorrelador (107).

[0116] Finalmente, el componente Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ decorrelacionado se combinan, por ejemplo, por medio de la suma (109), para obtener el componente Ambisonics final $B_l^m(k, n)$ del orden (nivel) l y modo m deseados para el cuadro de tiempo-frecuencia (k, n) , es decir,

$$B_l^m(k, n) = B_{dir,l}^m(k, n) + B_{diff,l}^m(k, n).$$

[0117] Los componentes Ambisonics $B_l^m(k, n)$ resultantes eventualmente se pueden transformar de nuevo al dominio del tiempo utilizando un banco de filtros inverso o una STFT inversa, almacenar, transmitir, o utilizar, por ejemplo, para reproducción de sonido espacial. En la práctica, se calcularían los componentes Ambisonics para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado.

[0118] Es importante enfatizar que la transformación de vuelta al dominio del tiempo utilizando, por ejemplo, un banco de filtros inverso o una STFT inversa se puede llevar a cabo antes del cálculo de $B_l^m(k, n)$, es decir, antes de la operación (109). Esto significa, que primero podemos transformar $B_{dir,l}^m(k, n)$ y $B_{diff,l}^m(k, n)$ de vuelta al dominio del tiempo y después sumar ambos componentes con la operación (109) para obtener el componente Ambisonics final B_l^m . Esto es posible ya que el banco de filtros inverso o la STFT inversa son en general operaciones lineales. En la misma forma, el decorrelador (107) se puede aplicar al componente Ambisonics de sonido difuso $B_{diff,l}^m$ después de transformar $B_{diff,l}^m$ de vuelta al dominio del tiempo. Esto puede ser conveniente en la práctica ya que algunos decorreladores operan señales en el dominio del tiempo.

[0119] Además, se debe observar que se puede agregar un bloque a la fig. 5, tal como un banco de filtros inverso antes del decorrelador, y el banco de filtros inverso se puede agregar en cualquier parte en el sistema.

[0120] Como se explicó en la Realización 3, el algoritmo en esta realización se puede configurar de tal manera que los componentes Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ se calculan para diferentes modos (órdenes) l . Por ejemplo, $B_{dir,l}^m(k, n)$ se puede calcular hasta el orden $l = 4$, mientras que $B_{diff,l}^m(k, n)$ se puede calcular solamente hasta el orden $l = 1$. Esto reduciría la complejidad computacional.

Realización 5

[0121] La figura 6 muestra otra realización de la invención que permite sintetizar un componente Ambisonics de un orden (nivel) l y modo m deseados a partir de las señales de múltiples (dos o más) micrófonos. La realización es similar a la Realización 4, pero la señal de sonido directo y la señal de sonido difuso se determinan a partir de la pluralidad de señales de micrófono y explotando la información de dirección de llegada.

[0122] Como en la Realización 4, la entrada a la invención son las señales de múltiples (dos o más) micrófonos. Los micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.

[0123] Como en la Realización 4, las múltiples señales de micrófono se transforman en el dominio del tiempo-frecuencia en el Bloque (101) utilizando, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT). La salida de la transformada de tiempo-frecuencia (101) son las señales de micrófono en el dominio del tiempo-frecuencia, que son denotadas por $P1...M(k, n)$. El siguiente procesamiento se lleva a cabo por separado para los cuadros de tiempo-frecuencia (k, n) .

[0124] Como en la Realización 4, se lleva a cabo una estimación de dirección de sonido en el Bloque (102) por tiempo y frecuencia utilizando dos o más de las señales de micrófono $P1...M(k, n)$. Los estimadores correspondientes se discuten en la Realización 1. La salida del estimador de dirección de sonido (102) es una dirección de sonido para un caso de tiempo n e índice de frecuencia k . La dirección de sonido se puede expresar por ejemplo, en términos de un vector normal unitario $n(k, n)$ o en términos de un ángulo de azimut $\varphi(k, n)$ y/o ángulo de elevación $\vartheta(k, n)$, los cuales se relacionan como se explicó en la Realización 1.

[0125] Como en la Realización 4, la respuesta de una función de base espacial del orden (nivel) l y modo m deseados se determina en el Bloque (103) por tiempo y frecuencia utilizando la información de dirección de sonido estimada. La respuesta de la función de base espacial se denota por $G_l^m(k, n)$. Por ejemplo, podemos considerar

armónicas esféricas de valor real con normalización N3D como función de base espacial y $G_l^m(k, n)$ se puede determinar como se explicó en la Realización 1.

[0126] Como en la Realización 4, una respuesta promedio de una función de base espacial del orden (nivel) l y modo m deseados, que es independiente del índice de tiempo n , se obtiene a partir del Bloque (106). Esta respuesta promedio se denota por $D_l^m(k)$ y describe la respuesta de una función de base espacial para los sonidos que llegan de todas las direcciones posibles (tal como sonidos difusos o sonidos ambientales). La respuesta promedio $D_l^m(k)$ se puede obtener como se describe en la Realización 3.

[0127] En esta realización, se determina una señal de sonido directo $P_{dir}(k, n)$ y una señal de sonido difuso $P_{diff}(k, n)$ en el Bloque (110) por índice de tiempo n e índice de frecuencia k a partir de las dos o más señales de micrófono disponibles $P_1 \dots M(k, n)$. Para este propósito, el Bloque (110) generalmente explota la información de dirección de sonido que se determinó en el Bloque (102). En lo sucesivo, se explican diferentes ejemplos del Bloque (110) que describen cómo determinar $P_{dir}(k, n)$ y $P_{diff}(k, n)$.

[0128] En un primer ejemplo del Bloque (110), se determina una señal de micrófono de referencia denotada por $P_{ref}(k, n)$ a partir de las múltiples señales de micrófono $P_1 \dots M(k, n)$ con base en la información de dirección de sonido proporcionada por el Bloque (102). La señal de micrófono de referencia $P_{ref}(k, n)$ se puede determinar seleccionando la señal de micrófono que sea la más cercana a la dirección de sonido estimada para el tiempo y frecuencia considerados. Este procedimiento de selección para determinar la señal de micrófono de referencia $P_{ref}(k, n)$ se explicó en la Realización 2. Después de determinar $P_{ref}(k, n)$, se pueden calcular una señal de sonido directo $P_{dir}(k, n)$ y una señal de sonido difuso $P_{diff}(k, n)$, por ejemplo, aplicando los filtros de un solo canal $W_{dir}(k, n)$ y $W_{diff}(k, n)$, respectivamente, a la señal de micrófono de referencia $P_{ref}(k, n)$. Esta estrategia y el cálculo de los filtros de un solo canal correspondientes se explicaron en la Realización 3.

[0129] En un segundo ejemplo del Bloque (110), determinamos una señal de micrófono de referencia $P_{ref}(k, n)$, como en el ejemplo previo y calculamos $P_{dir}(k, n)$ aplicando un filtro de un solo canal $W_{dir}(k, n)$ a $P_{ref}(k, n)$. Para determinar la señal difusa, sin embargo, seleccionamos una segunda señal de referencia $P_{ref, l}^m(k, n)$ y aplicamos un filtro de un solo canal $W_{diff}(k, n)$, a la segunda señal de referencia $P_{ref, l}^m(k, n)$, es decir,

$$P_{diff}(k, n) = W_{diff}(k, n)P_{ref, l}^m(k, n).$$

[0130] El filtro $W_{diff}(k, n)$ se puede calcular como se explicó, por ejemplo, en la Realización 3. La segunda señal de referencia $P_{ref, l}^m(k, n)$ corresponde a una de las señales de micrófono $P_1 \dots M(k, n)$ disponibles. Sin embargo, para diferentes órdenes l y modos m podemos utilizar diferentes señales de micrófono como segunda señal de referencia. Por ejemplo, para el nivel $l = 1$ y modo $m = -1$, podemos utilizar la primera señal de micrófono como segunda señal de referencia, es decir, $P_{ref, 1}^{-1}(k, n) = P_1(k, n)$. Para el nivel $l = 1$ y modo $m = 0$, podemos utilizar la segunda señal de micrófono, es decir, $P_{ref, 1}^0(k, n) = P_2(k, n)$. Para el nivel $l = 1$ y modo $m = 1$, podemos utilizar la tercera señal de micrófono, es decir, $P_{ref, 1}^1(k, n) = P_3(k, n)$. Las señales de micrófono $P_1 \dots M(k, n)$ disponibles se pueden asignar, por ejemplo, aleatoriamente a la segunda señal de referencia $P_{ref, l}^m(k, n)$ para los diferentes órdenes y modos. Esta es una estrategia razonable en la práctica ya que para situaciones de grabación difusa o ambiental, todas las señales de micrófono generalmente contienen potencia de sonido similar. Seleccionar diferentes señales de micrófono de referencia para diferentes órdenes y modos tiene la ventaja de que las señales de sonido difuso resultantes son a menudo (al menos parcialmente) mutuamente no correlacionadas para los diferentes órdenes y modos.

[0131] En un tercer ejemplo del Bloque (110), la señal de sonido directo $P_{dir}(k, n)$ se determina aplicando un filtro multicanal denotado por $w_{dir}(n)$ a las múltiples señales de micrófono $P_1 \dots M(k, n)$, es decir,

$$P_{dir}(k, n) = w_{dir}^H(n)p(k, n),$$

donde el filtro multicanal $w_{dir}(n)$ depende de la dirección de sonido estimada y el vector $p(k, n) = [P_1(k, n), \dots, P_M(k, n)]^T$ contiene las múltiples señales de micrófono. Existen muchos filtros multicanal $w_{dir}(n)$ óptimos diferentes en la bibliografía que se pueden utilizar para calcular $P_{dir}(k, n)$ a partir de la información de dirección de sonido, por ejemplo los filtros derivados en [InformedSF]. De manera similar, la señal de sonido difuso $P_{diff}(k, n)$ se determina aplicando un filtro multicanal denotado por $w_{diff}(n)$ a las múltiples señales de micrófono $P_1 \dots M(k, n)$, es decir,

$$P_{diff}(k, n) = w_{diff}^H(n)p(k, n),$$

donde el filtro multicanal $w_{diff}(n)$ depende de la dirección de sonido estimada. Existen muchos filtros multicanal $w_{diff}(n)$

óptimos diferentes en la bibliografía que se pueden utilizar para calcular $P_{diff}(k,n)$, por ejemplo el filtro que fue derivado en [DiffuseBF].

[0132] En un cuarto ejemplo del Bloque (110), determinamos $P_{dir}(k,n)$ y $P_{diff}(k,n)$ como en el ejemplo previo aplicando filtros multicanal $w_{dir}(n)$ y $w_{diff}(n)$, respectivamente, a las señales de micrófono $p(k,n)$. Sin embargo, utilizamos diferentes filtros $w_{diff}(n)$ para diferentes órdenes l y modos m de tal manera que las señales de sonido difuso $P_{diff}(k,n)$ resultantes para los diferentes órdenes l y modos m son mutuamente no correlacionadas. Estos diferentes filtros $w_{diff}(n)$ que minimizan la correlación entre las señales de salida se pueden calcular, por ejemplo, como se explica en [CovRender].

[0133] Como en la Realización 4, la señal de sonido directo $P_{dir}(k,n)$ determinada en el Bloque (105) se combina tal como multiplicada 115a por tiempo y frecuencia con la respuesta $G_l^m(k, n)$ de la función de base espacial determinada en el Bloque (103) que resulta en un componente Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ del orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k,n) .

[0134] Por otra parte, la señal de sonido difuso $P_{diff}(k,n)$ determinada en el Bloque (105) se combina tal como multiplicada 115b por tiempo y frecuencia con la respuesta promedio $D_l^m(k)$ de la función de base espacial determinada en el Bloque (106) que resulta en un componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ del orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k,n) .

[0135] Como en la Realización 3, el componente Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ calculados se combinan, por ejemplo, por medio de la operación de suma (109) para obtener el componente Ambisonics final $B_l^m(k, n)$ del orden (nivel) l y modo m deseados para el cuadro de tiempo-frecuencia (k,n) . Los componentes Ambisonics $B_l^m(k, n)$ resultantes eventualmente se pueden transformar de nuevo al dominio del tiempo utilizando un banco de filtros inverso o una STFT inversa, almacenar, transmitir, o utilizar, por ejemplo, para aplicaciones de reproducción de sonido espacial. En la práctica, se calcularían los componentes Ambisonics para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado. Como se explicó en la Realización 3, la transformación de vuelta al dominio del tiempo se puede llevar a cabo antes del cálculo de $B_l^m(k, n)$, es decir, antes de la operación (109).

[0136] Cabe observar que el algoritmo en esta realización se puede configurar de tal manera que los componentes Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ se calculan para diferentes modos (órdenes) l . Por ejemplo, $B_{dir,l}^m(k, n)$ se puede calcular hasta el orden $l = 4$, mientras que $B_{diff,l}^m(k, n)$ se puede calcular solamente hasta el orden $l = 1$ (en este caso, $B_{diff,l}^m(k, n)$ sería cero para órdenes mayores que $l = 1$). Si se desea, por ejemplo, calcular solamente $B_{dir,l}^m(k, n)$ pero no $B_{diff,l}^m(k, n)$ para un orden (nivel) l o modo m específicos, entonces, por ejemplo, el Bloque (110) se puede configurar de tal manera que la señal de sonido difuso $P_{diff}(k,n)$ se vuelve igual que cero. Esto se puede lograr, por ejemplo, ajustando el filtro $W_{diff}(k,n)$ en las ecuaciones previas a 0 y el filtro $W_{dir}(k,n)$ a 1. De manera similar, el filtro $w_{diff}^H(n)$, se podría establecer en cero.

Realización 6

[0137] La figura 7 muestra otra realización de la invención que permite sintetizar un componente Ambisonics de un orden (nivel) l y modo m deseados a partir de las señales de múltiples (dos o más) micrófonos. La realización es similar a la Realización 5, pero adicionalmente contiene decorreladores para los componentes Ambisonics difusos.

[0138] Como en la Realización 5, la entrada a la invención son las señales de múltiples (dos o más) micrófonos. Los micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.

[0139] Como en la Realización 5, las múltiples señales de micrófono se transforman en el dominio del tiempo-frecuencia en el Bloque (101) utilizando, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT). La salida de la transformada de tiempo-frecuencia (101) son las señales de micrófono en el dominio del tiempo-frecuencia, que son denotadas por $P_1 \dots M(k,n)$. El siguiente procesamiento es llevado a cabo por separado para los cuadros de tiempo-frecuencia (k,n) .

[0140] Como en la Realización 5, se lleva a cabo una estimación de dirección de sonido en el Bloque (102) por tiempo y frecuencia utilizando dos o más de las señales de micrófono $P_1 \dots M(k,n)$. Los estimadores correspondientes se discuten en la Realización 1. La salida del estimador de dirección de sonido (102) es una dirección de sonido para

un caso de tiempo n e índice de frecuencia k . La dirección de sonido se puede expresar por ejemplo, en términos de un vector normal unitario $\mathbf{n}(k,n)$ o en términos de un ángulo de azimut $\varphi(k,n)$ y/o ángulo de elevación $\vartheta(k,n)$, los cuales se relacionan como se explicó en la Realización 1.

5 **[0141]** Como en la Realización 5, la respuesta de una función de base espacial del orden (nivel) l y modo m deseados se determina en el Bloque (103) por tiempo y frecuencia utilizando la información de dirección de sonido estimada. La respuesta de la función de base espacial se denota por $G_l^m(k, n)$. Por ejemplo, podemos considerar armónicas esféricas de valor real con normalización N3D como función de base espacial y $G_l^m(k, n)$ se puede determinar como se explicó en la Realización 1.

10 **[0142]** Como en la Realización 5, una respuesta promedio de una función de base espacial del orden (nivel) l y modo m deseados, que es independiente del índice de tiempo n , se obtiene a partir del Bloque (106). Esta respuesta promedio se denota por $D_l^m(k)$ y describe la respuesta de una función de base espacial para los sonidos que llegan de todas las direcciones posibles (tal como sonidos difusos o sonidos ambientales). La respuesta promedio $D_l^m(k)$ se puede obtener como se describe en la Realización 3.

15 **[0143]** Como en la Realización 5, se determina una señal de sonido directo $P_{dir}(k,n)$ y una señal de sonido difuso $P_{diff}(k,n)$ en el Bloque (110) por índice de tiempo n e índice de frecuencia k a partir de las dos o más señales de micrófono disponibles $P_1 \dots M(k,n)$. Para este propósito, el Bloque (110) generalmente explota la información de dirección de sonido que se determinó en el Bloque (102). En la Realización 5 se explican diferentes ejemplos del Bloque (110).

20 **[0144]** Como en la Realización 5, la señal de sonido directo $P_{dir}(k,n)$ determinada en el Bloque (105) se combina tal como multiplicada 115a por tiempo y frecuencia con la respuesta $G_l^m(k, n)$ de la función de base espacial determinada en el Bloque (103) que resulta en un componente Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ del orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k,n) . Por otra parte, la señal de sonido difuso $P_{diff}(k,n)$ determinada en el Bloque (105) se combina tal como multiplicada 115b por tiempo y frecuencia con la respuesta promedio $D_l^m(k)$ de la función de base espacial determinada en el Bloque (106) que resulta en un componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ del orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k,n) .

30 **[0145]** Como en la Realización 4, el componente Ambisonics de sonido difuso $B_{diff,l}^m(k, n)$ calculado se decorrelaciona en el Bloque (107) utilizando un decorrelador que resulta en un componente Ambisonics de sonido difuso decorrelacionado, denotado por $\mathcal{B}_{diff,l}^m(k, n)$. El razonamiento y procedimientos detrás de la decorrelación se discuten en la Realización 4. Como en la Realización 4, el componente Ambisonics de sonido difuso $\mathcal{B}_{diff,l}^m(k, n)$ se puede transformar de nuevo al dominio del tiempo utilizando, por ejemplo, un filtro de bancos inverso o una STFT inversa antes de aplicar el decorrelador (107).

35 **[0146]** Como en la Realización 4, el componente Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $\mathcal{B}_{diff,l}^m(k, n)$ decorrelacionado se combinan, por ejemplo, por medio de la suma (109), para obtener el componente Ambisonics final $B_l^m(k, n)$ del orden (nivel) l y modo m deseados para el cuadro de tiempo-frecuencia (k,n) . Los componentes Ambisonics $B_l^m(k, n)$ resultantes eventualmente se pueden transformar de nuevo al dominio del tiempo utilizando un banco de filtros inverso o una STFT inversa, almacenar, transmitir, o utilizar, por ejemplo, para reproducción de sonido espacial. En la práctica, se calcularían los componentes Ambisonics para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado. 45 Como se explicó en la Realización 4, la transformación de vuelta al dominio del tiempo se puede llevar a cabo antes del cálculo de $B_l^m(k, n)$, es decir antes de la operación (109).

[0147] Como en la Realización 4, el algoritmo en esta realización se puede configurar de tal manera que los componentes Ambisonics de sonido directo $B_{dir,l}^m(k, n)$ y el componente Ambisonics de sonido difuso $\mathcal{B}_{diff,l}^m(k, n)$ se calculan para diferentes modos (órdenes) l . Por ejemplo, $B_{dir,l}^m(k, n)$ se puede calcular hasta el orden $l = 4$, mientras que $\mathcal{B}_{diff,l}^m(k, n)$ se puede calcular solamente hasta el orden $l = 1$.

Realización 7

55 **[0148]** La figura 8 muestra otra realización de la invención que permite sintetizar un componente Ambisonics de un orden (nivel) l y modo m deseados a partir de las señales de múltiples (dos o más) micrófonos. La realización es similar a la Realización 1, pero adicionalmente contiene un Bloque (111) que aplica una operación de suavizado a la respuesta $G_l^m(k, n)$ calculada de la función de base espacial.

[0149] Como en la realización 1, la entrada a la invención son las señales de múltiples (dos o más) micrófonos. Los micrófonos pueden estar dispuestos en una geometría arbitraria, por ejemplo, como una configuración coincidente, disposición lineal, disposición plana, o disposición tridimensional. Por otra parte, cada micrófono puede poseer una
5 directividad direccional omnidireccional o arbitraria. Las directividades de los diferentes micrófonos pueden diferir.

[0150] Como en la realización 1, las múltiples señales de micrófono se transforman en el dominio del tiempo-frecuencia en el Bloque (101) utilizando, por ejemplo, un banco de filtros o una transformada de Fourier de tiempo corto (STFT). La salida de la transformada de tiempo-frecuencia (101) son las señales de micrófono en el dominio del
10 tiempo-frecuencia, que son denotadas por $P1...M(k,n)$. El siguiente procesamiento se lleva a cabo por separado para los cuadros de tiempo-frecuencia (k,n) .

[0151] Como en la realización 1, sin pérdida de generalidad, la primera señal de micrófono se denomina como la señal de micrófono de referencia, es decir, $Pref(k,n) = P1(k,n)$.
15

[0152] Como en la realización 1, se lleva a cabo una estimación de dirección de sonido en el Bloque (102) por tiempo y frecuencia utilizando dos o más de las señales de micrófono $P1...M(k,n)$. Los estimadores correspondientes se discuten en la Realización 1. La salida del estimador de dirección de sonido (102) es una dirección de sonido para un caso de tiempo n e índice de frecuencia k . La dirección de sonido se puede expresar por ejemplo, en términos de
20 un vector normal unitario $n(k,n)$ o en términos de un ángulo de azimut $\varphi(k,n)$ y/o ángulo de elevación $\vartheta(k,n)$, los cuales se relacionan como se explicó en la Realización 1.

[0153] Como en la Realización 1, la respuesta de una función de base espacial del orden (nivel) l y modo m deseados se determina en el Bloque (103) por tiempo y frecuencia utilizando la información de dirección de sonido
25 estimada. La respuesta de la función de base espacial se denota por $G_l^m(k, n)$. Por ejemplo, podemos considerar armónicas esféricas de valor real con normalización N3D como función de base espacial y $G_l^m(k, n)$ se puede determinar como se explicó en la Realización 1.

[0154] En contraste a la Realización 1, la respuesta $G_l^m(k, n)$ se utiliza como entrada para el Bloque (111) que aplica una operación de suavizado a $G_l^m(k, n)$. La salida del Bloque (111) es una función de respuesta suavizada
30 denotada como $\overline{G}_l^m(k, n)$. El propósito de la operación de suavizado es reducir una varianza de estimación no deseada de los valores $G_l^m(k, n)$, lo cual puede ocurrir en la práctica, por ejemplo, si las direcciones de sonido $\varphi(k,n)$ y/o $\vartheta(k,n)$, estimadas en el Bloque (102) son ruidosas. El suavizado, aplicado a $G_l^m(k, n)$, se puede llevar a cabo, por ejemplo, a través del tiempo y/o frecuencia. Por ejemplo, se puede lograr un suavizado temporal utilizando el filtro
35 de promediado recursivo bien conocido

$$\overline{G}_l^m(k, n) = \alpha G_l^m(k, n) + (1 - \alpha) G_l^m(k, n - 1),$$

donde $G_l^m(k, n - 1)$ es la función de respuesta calculada en el cuadro de tiempo previo. Por otra parte, α es un
40 número de valor real entre 0 y 1 que controla la intensidad del suavizado temporal. Para valores de α cercanos a 0, se lleva a cabo un promediado temporal fuerte, mientras que para valores de α cercanos a 1, se lleva a cabo un promediado temporal corto. En aplicaciones prácticas, el valor de α depende de la aplicación y se puede establecer constante, por ejemplo, $\alpha = 0,5$. Alternativamente, también se puede llevar a cabo un suavizado espectral en el Bloque
45 suavizado espectral, por ejemplo, dentro de las llamadas bandas de ERB, se describe, por ejemplo, en [ERBsmooth].

[0155] En esta realización, la señal de micrófono de referencia $Pref(k,n)$ finalmente se combina tal como multiplicada 115 por el tiempo y la frecuencia con la respuesta suavizada $\overline{G}_l^m(k, n)$ de la función de base espacial
50 determinada en el Bloque (111) que resulta en el componente Ambisonics $B_l^m(k, n)$ deseado de orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k,n) . Los componentes Ambisonics $B_l^m(k, n)$ resultantes eventualmente se pueden transformar de nuevo al dominio del tiempo utilizando un banco de filtros inverso o una STFT inversa, almacenar, transmitir, o utilizar, por ejemplo, para reproducción de sonido espacial. En la práctica, se calcularían los componentes Ambisonics para todos los órdenes y modos deseados para obtener la señal Ambisonics deseada del orden (nivel) máximo deseado.
55

[0156] Claramente, el suavizado de ganancia en el Bloque (111) se puede aplicar también en todas las demás realizaciones de esta invención.

Realización 8

60

[0157] La presente invención se puede aplicar también en el caso llamado de onda múltiple, donde se considera más de una dirección de sonido por cuadro de tiempo-frecuencia. Por ejemplo, la Realización 2, que se ilustra en la fig. 3b, se puede realizar en el caso de onda múltiple. En este caso, el Bloque (102) estima J direcciones de sonido por tiempo y frecuencia, donde J es un valor entero mayor que uno, por ejemplo, J = 2. Para estimar múltiples direcciones de sonido, se pueden utilizar estimadores del estado de la técnica, por ejemplo ESPRIT o Root MUSIC, los cuales se describen en [ESPRIT,RootMUSIC1]. En este caso, la salida del Bloque (102) son múltiples direcciones de sonido, indicadas, por ejemplo, en términos de múltiples ángulos de azimut $\varphi_{1...J}(k,n)$ y/o ángulos de elevación $\theta_{1...J}(k,n)$.

[0158] Las múltiples direcciones de sonido se utilizan entonces en el Bloque (103) para calcular múltiples respuestas $G_{l,1...J}^m(k,n)$, una respuesta por cada dirección de sonido estimada como se discute, por ejemplo, en la Realización 1. Por otra parte, las múltiples direcciones de sonido calculadas en el Bloque (102) se utilizan en el Bloque (104) para calcular múltiples señales de referencia $P_{ref,1...J}(k,n)$, una por cada una de las múltiples direcciones de sonido. Cada una de las múltiples señales de referencia se puede calcular, por ejemplo, aplicando filtros multicanal $w_{1...J}(n)$ a las múltiples señales de micrófono, de manera similar a como se explicó en la Realización 2. Por ejemplo, la primera señal de referencia $P_{ref,1}(k,n)$ se puede obtener aplicando un filtro multicanal del estado de la técnica $w_1(n)$, el cual extraería los sonidos de la dirección $\varphi_1(k,n)$ y/o $\theta_1(k,n)$ mientras se atenúan los sonidos de todas las demás direcciones de sonido. Tal filtro se puede calcular, por ejemplo, como el filtro LCMV informado que se explica en [InformedSF]. Las múltiples señales de referencia $P_{ref,1...J}(k,n)$ se multiplican entonces con las múltiples respuestas $G_{l,1...J}^m(k,n)$ correspondientes para obtener múltiples componentes Ambisonics $B_{l,1...J}^m(k,n)$. Por ejemplo, el J-ésimo componente Ambisonics correspondiente a la J-ésima dirección de sonido y señal de referencia, respectivamente, se calcula como

$$B_{l,j}^m(k,n) = P_{ref,j}(k,n)G_{l,j}^m(k,n).$$

[0159] Finalmente, los componentes J Ambisonics se suman para obtener el componente Ambisonics final $B_l^m(k,n)$ deseado de orden (nivel) l y modo m para el cuadro de tiempo-frecuencia (k,n), es decir,

$$B_l^m(k,n) = \sum_{j=1}^J B_{l,j}^m(k,n).$$

[0160] Claramente, también las otras realizaciones antes mencionadas se pueden extender al caso de onda múltiple. Por ejemplo, en la Realización 5 y la Realización 6 podemos calcular múltiples sonidos directos $P_{dir,1...J}(k,n)$, uno para cada una de las múltiples direcciones de sonido, utilizando los mismos filtros multicanal que se mencionan en esta realización. Los múltiples sonidos directos se multiplican entonces con las múltiples respuestas $G_{l,1...J}^m(k,n)$ correspondientes llevando a múltiples componentes Ambisonics de sonido directo $B_{dir,l}^m(k,n)$ que se pueden sumar para obtener el componente Ambisonics de sonido directo final $B_{dir,l,1...J}^m(k,n)$ deseado.

[0161] Se debe observar que la invención no solamente se puede aplicar a las técnicas de Ambisonics bidimensional (cilíndricas) o tridimensional (esférica) sino también a cualquier otra técnica basada en funciones de base espacial para calcular cualquier componente de campo de sonido.

Realizaciones de la invención como una lista

[0162]

1. Transformar múltiples señales de micrófono en el dominio del tiempo-frecuencia.
2. Calcular una o más direcciones de sonido por tiempo y frecuencia a partir de las múltiples señales de micrófono.
3. Calcular para cada tiempo y frecuencia una o más funciones de respuesta dependiendo de dichas una o más direcciones de sonido.
4. Para cada tiempo y frecuencia, obtener una o más señales de micrófono de referencia.
5. Para cada tiempo y frecuencia, multiplicar dichas una o más señales de micrófono de referencia con dichas una o más funciones de respuesta para obtener uno o más componentes Ambisonics del orden y modo deseados.
6. Si múltiples componentes Ambisonics fueron obtenidos para el orden y modo deseados, sumar los componentes Ambisonics correspondientes para obtener el componente Ambisonics final deseado.
4. En algunas Realizaciones, calcular en la Etapa 4 uno o más sonidos directos y sonidos difusos a partir de las múltiples señales de micrófono en lugar de dichas una o más señales de micrófono de referencia.
5. Multiplicar dichos uno o más sonidos directos y sonidos difusos con una o más respuestas de sonido directo y respuestas de sonido difuso correspondientes para obtener uno o más componentes Ambisonics de sonido directo y componentes Ambisonics de sonido difuso para el orden y modo deseados.

6. Los componentes Ambisonics de sonido difuso se pueden decorrelacionar adicionalmente para diferentes órdenes y modos.

7. Sumar los componentes Ambisonics de sonido directo y componentes Ambisonics de sonido difuso para obtener el componente Ambisonics final deseada del orden y modo deseados.

5

Referencias

[0163]

- 10 [Ambisonics] R. K. Furness, "Ambisonics - An overview", en AES 8th International Conference, abril de 1990, pp. 181–189.
- [Ambix] C. Nachbar, F. Zotter, E. Deleflie, y A. Sontacchi, "AMBIX - A Suggested Ambisonics Format", Proceedings of the Ambisonics Symposium 2011.
- 15 [ArrayDesign] M. Williams y G. Le Du, "Multichannel Microphone Array Design", en Audio Engineering Society Convention 108, 2008.
- [CovRender] J. Vilkamo y V. Pulkki, "Minimization of Decorrelator Artifacts in Directional Audio Coding by Covariance Domain Rendering", J. Audio Eng. Soc, vol. 61, no. 9, 2013.
- [DiffuseBF] O. Thiergart y E. A. P. Habets, "Extracting Reverberant Sound Using a Linearly Constrained Minimum Variance Spatial Filter", IEEE Signal Processing Letters, vol. 21, no. 5, mayo de 2014.
- 20 [DirAC] V. Pulkki, "Directional audio coding in spatial sound reproduction and stereo upmixing", en Proceedings of The AES 28th International Conference, pp. 251-258, junio de 2006.
- [EigenMike] J. Meyer y T. Agnello, "Spherical microphone array for spatial sound recording", en Audio Engineering Society Convention 115, octubre de 2003.
- [ERBsmooth] A. Favrot y C. Faller, "Perceptually Motivated Gain Filter Smoothing for Noise Suppression", Audio Engineering Society Convention 123, 2007.
- 25 [ESPRIT] 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, EUA, abril de 1986.
- [FourierAcoust] E. G. Williams, "Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography", Academic Press, 1999.
- 30 [HARPEX] S. Berge y N. Barrett, "High Angular Resolution Planewave Expansion", en 2nd International Symposium on Ambisonics and Spherical Acoustics, mayo de 2010.
- [InformedSF] O. Thiergart, M. Taseska, y E. A. P. Habets, "An Informed Parametric Spatial Filter Based on Instantaneous Direction-of-Arrival Estimates", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 12, diciembre de 2014.
- 35 [MicSetup3D] H. Lee and C. Gribben, "On the optimum microphone array configuration for height channels", en 134 AES Convention, Roma, 2013.
- [MUSIC] R. Schmidt, "Multiple emitter location and signal parameter estimation", IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276–280, 1986.
- 40 [OptArrayPr] B. D. Van Veen y K. M. Buckley, "Beamforming: A versatile approach to spatial filtering", IEEE ASSP Magazine, vol. 5, no. 2, 1988.
- [RootMUSIC1] B. Rao and y K. Hari, "Performance analysis of root-MUSIC", en Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on, vol. 2, 1988, pp. 578–582.
- [RootMUSIC2] A. Mhamdi y A. Samet, "Direction of arrival estimation for nonuniform linear antenna", en Communications, Computing and Control Applications (CCCA), 2011 International Conference en marzo de 2011, pp. 1–5.
- 45 [RootMUSIC3] M. Zoltowski y C. P. Mathews, "Direction finding with uniform circular arrays via phase mode excitation and beamspace root-MUSIC", en Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference en, vol. 5, 1992, pp. 245–248.
- 50 [SDRestim] O. Thiergart, G. Del Galdo, y E. A. P. Habets, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation", The Journal of the Acoustical Society of America, vol. 132, no. 4, 2012.
- [SourceNum] J.-S. Jiang y M.-A. Ingram, "Robust detection of number of sources using the transformed rotational matrix", en Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE, vol. 1, marzo de 2004.
- 55 [SpCoherence] D. P. Jarrett, O. Thiergart, E. A. P. Habets, y P. A. Naylor, "Coherence-Based Diffuseness Estimation in the Spherical Harmonic Domain", IEEE 27th Convention of Electrical and Electronics Engineers in Israel (IEEEI), 2012.
- 60 [SphHarm] F. Zotter, "Analysis and Synthesis of Sound-Radiation with Spherical Arrays", tesis doctoral, University of Music and Performing Arts Graz, 2009.
- [VirtualMic] O. Thiergart, G. Del Galdo, M. Taseska, y E. A. P. Habets, "Geometry-based Spatial Sound Acquisition Using Distributed Microphone Arrays", IEEE Transactions on in Audio, Speech, and Language Processing, vol. 21, no. 12, De.
- 65

5 **[0164]** Aunque algunos aspectos se han descrito en el contexto de un aparato, es claro que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.

10 **[0165]** La señal inventiva se puede almacenar en un medio de almacenamiento digital o puede ser transmitida en un medio de transmisión tal como un medio inalámbrico de transmisión o un medio alámbrico de transmisión tal como Internet.

15 **[0166]** Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede llevar a cabo utilizando un medio de almacenamiento digital, por ejemplo un disco floppy, 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 son capaces de cooperar) con un sistema informático programable de tal manera que se lleve a cabo el procedimiento respectivo.

20 **[0167]** Algunas realizaciones según la invención comprenden un soporte de datos no transitorio que tiene señales de control legibles electrónicamente, las cuales son capaces de cooperar con un sistema informático programable, de tal manera que se lleve a cabo uno de los procedimientos descritos en esta invención.

25 **[0168]** Generalmente, las realizaciones 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 llevar a cabo uno de los procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa se puede almacenar, por ejemplo, en un soporte legible por máquina.

[0169] Otras realizaciones comprenden el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por máquina.

30 **[0170]** En otras palabras, una realización del procedimiento inventivo es, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los procedimientos descritos en esta invención, cuando el programa informático se ejecuta en un ordenador.

35 **[0171]** Una realización adicional de los procedimientos inventivos es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención.

40 **[0172]** Una realización adicional del procedimiento inventivo es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención. El flujo de datos o la secuencia de señales puede estar configurado, por ejemplo, para ser transferido por medio de una conexión de comunicación de datos, por ejemplo, a través de Internet.

45 **[0173]** Una realización adicional comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los procedimientos descritos en esta invención.

[0174] Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los procedimientos descritos en esta invención.

50 **[0175]** En algunas realizaciones, un dispositivo lógico programable (por ejemplo, una matriz de puerta programable en campo) se puede utilizar para llevar a cabo algunas o todas de las funcionalidades de los procedimientos descritos en esta invención. En algunas realizaciones, la matriz de puerta programable en campo puede cooperar con un microprocesador con el fin de llevar a cabo uno de los procedimientos descritos en esta invención. Generalmente, los procedimientos se llevan a cabo preferentemente por medio de cualquier aparato de hardware.

60 **[0176]** Las realizaciones descritas anteriormente son solamente ilustrativas para los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los detalles descritos en esta invención serán aparentes para otros expertos en la materia. Por lo tanto, la intención es estar limitados solamente por el alcance de las próximas reivindicaciones de patente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en esta invención.

REIVINDICACIONES

1. Aparato para generar una descripción de campo de sonido que tiene una representación de componentes de campo de sonido, que comprende:
 - 5 un determinador de dirección (102) para determinar una o más direcciones de sonido para cada cuadro de tiempo-frecuencia de una pluralidad de cuadros de tiempo-frecuencia de una pluralidad de señales de micrófono; donde el aparato está configurado para calcular, para cada cuadro de tiempo-frecuencia, una o más respuestas de una función de base espacial dependiendo de una o más direcciones de sonido,
 - 10 donde el aparato está configurado para obtener, para cada cuadro de tiempo-frecuencia, una o más señales de sonido de referencia a partir de la pluralidad de señales de micrófono, y el aparato comprende además un calculador de componente de campo de sonido (201) configurado para multiplicar, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, una o más señales de referencia con una o más respuestas de la función de base espacial para obtener uno o más componentes de campo de sonido,
 - 15 siendo los uno o más componentes de campo de sonido uno o más componentes de Ambisonics de un orden y modo deseados; o el aparato está configurado para calcular, para cada cuadro de tiempo-frecuencia, una o más respuestas de sonido directas de una función de base espacial, dependiendo de una o más direcciones de sonido, donde el aparato está configurado para obtener, para cada cuadro de tiempo-frecuencia, una o más señales de sonido directas y una o más señales de sonido difusas, a partir de la pluralidad de señales de micrófono, y el
 - 20 aparato comprende además un calculador de componentes de campo de sonido configurado para multiplicar, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, las una o más señales de sonido directas con las una o más respuestas de sonido directas de la función de base espacial y las una o más señales de sonido difusas con una o más respuestas de sonido difusas de la función de base espacial para obtener uno o más componentes de campo de sonido directo y uno o más componentes de campo de sonido difuso, siendo
 - 25 los uno o más componentes de campo de sonido directo y difuso uno o más componentes Ambisonics directos y difusos de un orden y modo deseados.

2. Aparato según la reivindicación 1, que comprende además un evaluador de función de base espacial
 - 30 (103) para evaluar, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, una o más funciones de base espacial utilizando una o más direcciones de sonido para obtener una o más funciones de respuesta.

3. Aparato según la reivindicación 1 ó 2, donde el calculador de componentes de campo de sonido (201)
 - 35 está configurado para calcular múltiples componentes Ambisonics para el orden o modo deseado, y donde el calculador de componentes de campo de sonido (201) está configurado para sumar los componentes Ambisonics correspondientes para obtener un componente Ambisonics final deseado.

4. Aparato según la reivindicación 1, donde el calculador de campo de sonido está configurado para
 - 40 decorrelacionar los uno o más componentes Ambisonics difusos para diferentes órdenes o modos.

5. Aparato según la reivindicación 1 ó 4, donde el calculador de componentes de campo de sonido (201)
 - 45 está configurado para sumar los componentes Ambisonics directos y difusos para obtener un componente Ambisonics final deseado del orden y modo deseados.

6. Aparato según una de las reivindicaciones anteriores, que comprende además un convertidor de tiempo-frecuencia (101) para convertir cada una de la pluralidad de señales de micrófono que están en un dominio de tiempo
 - 50 en una representación de tiempo-frecuencia que tiene la pluralidad de cuadros de tiempo-frecuencia.

7. Aparato según una de las reivindicaciones anteriores, que comprende además un convertidor de frecuencia-tiempo (20) para convertir los uno o más componentes del campo de sonido o una combinación de los uno
 - 55 o más componentes del campo de sonido directo y los uno o más componentes del campo de sonido difuso en una representación del dominio del tiempo de los componentes del campo de sonido.

8. Aparato según la reivindicación 7,
 - 60 donde el convertidor de frecuencia-tiempo (20) está configurado para procesar uno o más componentes de campo de sonido directo para obtener una pluralidad de componentes de campo de sonido directo de dominio de tiempo, donde el convertidor de frecuencia-tiempo (20) está configurado para procesar los componentes de campo de sonido difuso para obtener una pluralidad de componentes de campo de sonido difuso de dominio de tiempo, o donde un combinador (401) está configurado para realizar una combinación de los componentes de campo de sonido directo de dominio de tiempo y los componentes de campo de sonido difuso de dominio de tiempo en el dominio temporal; o donde un combinador (401) está configurado para combinar uno o más componentes de campo de sonido directo para un cuadro de frecuencia de tiempo y uno o más componentes de campo de sonido

difuso para el cuadro de tiempo-frecuencia correspondiente en el dominio de frecuencia, y donde el convertidor de tiempo-frecuencia (20) está configurado para procesar un resultado del combinador (401) para obtener los componentes de campo de sonido en el dominio de tiempo.

5 9. Aparato según una de las reivindicaciones anteriores, que comprende además un calculador de señal de referencia (104) para calcular una o más señales de sonido de referencia a partir de la pluralidad de señales de micrófono

utilizando una o más direcciones de sonido,

10 utilizando la selección de una señal de micrófono específica a partir de la pluralidad de señales de micrófono basándose en una o más direcciones de sonido, o

utilizando un filtro multicanal aplicado a dos o más señales de micrófono de la pluralidad de señales de micrófono, dependiendo el filtro multicanal de una o más direcciones de sonido y posiciones individuales de micrófonos, a partir de los cuales se obtienen la pluralidad de señales de micrófono.

15

10. Aparato según la reivindicación 2,

donde el evaluador de función de base espacial (103) está configurado para utilizar, para una función de base espacial, una representación parametrizada, donde un parámetro de la representación parametrizada es una dirección de sonido, y para insertar un parámetro correspondiente a la dirección de sonido en la representación parametrizada para obtener un resultado de evaluación para cada función de base espacial; o

20

donde el evaluador de función de base espacial (103) está configurado para utilizar una tabla de consulta para cada función de base espacial que tiene, como entrada, una identificación de función de base espacial y la dirección de sonido, y que tiene, como salida, un resultado de evaluación, y donde el evaluador de función de base espacial (103) está configurado para determinar, para la una o más direcciones de sonido determinadas por el determinador de dirección (102), una dirección de sonido correspondiente de la entrada de la tabla de consulta o para calcular una media ponderada o no ponderada entre dos entradas de la tabla de consulta vecinas a la una o más direcciones de sonido determinadas por el determinador de dirección (102); o

25

donde el evaluador de función de base espacial (103) está configurado para utilizar para una función de base espacial, una representación parametrizada, donde un parámetro de la representación parametrizada es una dirección de sonido, siendo la dirección de sonido unidimensional, tal como un ángulo de acimut, en una situación bidimensional o bidimensional, tal como un ángulo de acimut y un ángulo de elevación, en una situación tridimensional, y para insertar un parámetro correspondiente a la dirección de sonido en la representación parametrizada para obtener un resultado de evaluación para cada función de base espacial.

35

11. Aparato según la reivindicación 2, que comprende además:

un determinador de sonido directo o difuso (105) para determinar una porción directa o una porción difusa de la pluralidad de señales de micrófono, como señal de referencia,

40

donde el calculador de componentes de campo de sonido (201) está configurado para utilizar la porción directa únicamente en el cálculo de uno o más componentes de campo de sonido directo.

12. Aparato según la reivindicación 11, que comprende además:

45

un determinador de función de base de respuesta promedio (106) para determinar una respuesta de función de base espacial promedio, comprendiendo el determinador un proceso de cálculo o un proceso de acceso a una tabla de consulta; y

un calculador de componente difuso (301), para calcular uno o más componentes de campo de sonido difuso utilizando solamente la porción difusa como señal de referencia junto con la respuesta de función de base espacial promedio.

50

13. Aparato según la reivindicación 12, que comprende además:

un combinador (109, 401) para combinar un componente de campo de sonido directo; y

55

un componente de campo de sonido difuso para obtener el componente de campo de sonido.

14. Aparato según una de las reivindicaciones 12 a 13,

donde el calculador de componente difuso (301) está configurado para calcular componentes de sonido difusos hasta un primer número u orden predeterminado,

60

donde el calculador de componentes de campo de sonido (201) está configurado para calcular componentes de campo de sonido directos hasta un segundo número u orden predeterminado,

donde el segundo número u orden predeterminado es mayor que el primer número u orden predeterminado, y donde el primer número u orden predeterminado es 1 o mayor que 1.

65

15. Aparato según una de las reivindicaciones 12 a 14,

5 donde el determinador de sonido directo o difuso (105) comprende un decorrelacionador (107) para decorrelacionar un componente de sonido difuso antes o después de una combinación con una respuesta promedio de una función de base espacial en una representación de dominio de frecuencia o una representación de dominio de tiempo.

16. Aparato según la reivindicación 11,

10 comprendiendo además un calculador de componente difuso (301) para calcular, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de sonido difuso, donde el determinador de sonido directo o difuso (105) está configurado para calcular la porción directa y la porción difusa a partir de una única señal de micrófono de la pluralidad de señales de micrófono, y donde el calculador de componente difuso (301) está configurado para calcular uno o más componentes de sonido difuso utilizando la porción difusa como señal de referencia, y donde el calculador de componente de campo de sonido (201) está configurado para calcular uno o más componentes de campo de sonido directo utilizando la porción directa como señal de referencia; o

15 donde el determinador de sonido directo o difuso (105) está configurado para calcular una porción difusa a partir de una señal de micrófono de la pluralidad de señales de micrófono que es diferente de la señal de micrófono de la pluralidad de señales de micrófono, a partir de la cual se calcula la porción directa, y donde el calculador de componente difuso (301) está configurado para calcular uno o más componentes de sonido difuso utilizando la porción difusa como señal de referencia, y donde el calculador de componente de campo de sonido (201) está configurado para calcular uno o más componentes de campo de sonido directo utilizando la porción directa como señal de referencia; o

20 comprendiendo además un calculador de componente difuso (301) para calcular, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de sonido difuso, donde el determinador de sonido directo o difuso (105) está configurado para calcular una porción difusa para una función de base espacial diferente utilizando una señal de micrófono diferente de la pluralidad de señales de micrófono, y donde el calculador de componente difuso (301) está configurado para utilizar una primera porción difusa como la señal de referencia para una respuesta de función de base espacial promedio correspondiente a un primer número, y para utilizar una segunda porción difusa diferente como la señal de referencia correspondiente a una segunda respuesta de función de base espacial promedio numérica, donde el primer número es diferente del segundo número, y donde el primer número y el segundo número indican cualquier orden o nivel y modo de la una o más funciones de base espacial; o

25 comprendiendo además un calculador de componente difuso (301) para calcular, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de sonido difuso, donde el determinador de sonido directo o difuso (105) está configurado para calcular la porción directa utilizando un primer filtro multicanal aplicado a la pluralidad de señales de micrófono y calcular la porción difusa utilizando un segundo filtro multicanal aplicado a la pluralidad de señales de micrófono, siendo el segundo filtro multicanal diferente del primer filtro multicanal, y donde el calculador de componente difuso (301) está configurado para calcular uno o más componentes de sonido difuso utilizando la porción difusa como señal de referencia, y donde el calculador de componente de campo de sonido (201) está configurado para calcular uno o más componentes de campo de sonido directo utilizando la porción directa como señal de referencia; o

30 comprendiendo además un calculador de componente difuso (301) para calcular, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de sonido difuso, donde el determinador de sonido directo o difuso (105) está configurado para calcular las porciones difusas para diferentes funciones de base espacial utilizando diferentes filtros multicanal para las diferentes funciones de base espacial, y donde el calculador de componente difuso (301) está configurado para calcular uno o más componentes de sonido difuso utilizando la porción difusa como señal de referencia, y

35 donde el calculador de componente de campo de sonido (201) está configurado para calcular uno o más componentes de campo de sonido directo utilizando la porción directa como señal de referencia.

17. Aparato según la reivindicación 2,

55 donde el evaluador de función de base espacial (103) comprende un suavizador de ganancia (111) que opera en una dirección de tiempo o una dirección de frecuencia, para suavizar los resultados de la evaluación, y donde el calculador de componentes de campo de sonido (201) está configurado para utilizar resultados de evaluación suavizados al calcular uno o más componentes de campo de sonido o uno o más componentes de campo de sonido directo y uno o más componentes de campo de sonido difuso.

18. Aparato según la reivindicación 2,

60 donde el evaluador de función de base espacial (103) está configurado para utilizar una o más funciones de base espacial para Ambisonics en una situación bidimensional o tridimensional.61610202

65

19. Aparato según la reivindicación 18,

donde el evaluador de funciones de base espacial (103) está configurado para utilizar al menos las funciones de base espacial de al menos dos niveles u órdenes o al menos dos modos.

5

20. Aparato según la reivindicación 19,

donde el calculador de componentes del campo de sonido (201) está configurado para calcular el componente del campo de sonido para al menos dos niveles de un grupo de niveles que comprende el nivel 0, el nivel 1, el nivel 2, el nivel 3, el nivel 4, o

10

donde el calculador de componentes del campo de sonido (201) está configurado para calcular los componentes del campo de sonido para al menos dos modos del grupo de modos que comprende el modo -4, el modo -3, el modo -2, el modo -1, el modo 0, el modo 1, el modo 2, el modo 3, el modo 4.

15 21. Aparato según una de las reivindicaciones anteriores,

un calculador de componente difuso (301) para calcular, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, uno o más componentes de sonido difuso; y

20

un combinador (401) para combinar información de sonido difuso e información de campo de sonido directo para obtener una representación de dominio de frecuencia o una representación de dominio de tiempo de los componentes de campo de sonido,

25

donde el calculador de componente difuso (301) o el combinador (401) está configurado para calcular o combinar un componente difuso hasta un cierto orden o número, siendo el cierto orden o número menor que un orden o número hasta el cual el calculador de componente de campo de sonido (201) está configurado para calcular un componente de campo de sonido directo.

22. Aparato según la reivindicación 21, donde el orden o número determinado es uno o cero, y el orden o número hasta el cual el calculador de componentes de campo de sonido (201) está configurado para calcular un componente de campo de sonido es 2 o más.

30

23. Aparato según una de las reivindicaciones anteriores,

donde el calculador de componentes de campo de sonido (201) está configurado para multiplicar (115) una señal en un cuadro de tiempo-frecuencia de la señal de referencia por un resultado de evaluación obtenido a partir de una función de base espacial para obtener información sobre un componente de campo de sonido asociado con la función de base espacial, y para multiplicar (115) la señal en el cuadro de tiempo-frecuencia de la señal de referencia por un resultado de evaluación adicional obtenido a partir de una función de base espacial adicional para obtener información sobre un componente de campo de sonido adicional asociado con la función de base espacial adicional.

40

24. Procedimiento para generar una descripción de campo de sonido que tiene una representación de componentes de campo de sonido, que comprende:

determinar (102) una o más direcciones de sonido para cada cuadro de tiempo-frecuencia de una pluralidad de cuadros de tiempo-frecuencia de una pluralidad de señales de micrófono; y: ya sea calcular para cada cuadro de tiempo-frecuencia, una o más respuestas de una función de base espacial que depende de una o más direcciones de sonido, y

45

obtener, para cada cuadro de tiempo-frecuencia, una o más señales de sonido de referencia de la pluralidad de señales de micrófono, y

50

multiplicar, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, una o más señales de sonido de referencia con una o más respuestas de la función básica espacial para obtener uno o más componentes de campo de sonido, siendo uno o más componentes de campo de sonido uno o más componentes Ambisonics de un orden y modo deseados; o calcular, para cada cuadro de tiempo-frecuencia, una o más respuestas de sonido directas de una función de base espacial que depende de una o más direcciones de sonido, y obtener, para cada cuadro de tiempo-frecuencia, una o más señales de sonido directas y una o más señales de sonido difusas de la pluralidad de señales de micrófono, y multiplicar, para cada cuadro de tiempo-frecuencia de la pluralidad de cuadros de tiempo-frecuencia, las una o más señales de sonido directas con las una o más respuestas de sonido directas de la función de base espacial y las una o más señales de sonido difusas con una o más respuestas de sonido difusas de la función de base espacial, para obtener uno o más componentes de

55

campo de sonido directo y uno o más componentes de campo de sonido difuso, siendo los uno o más componentes de campo de sonido directos y difusos uno o más componentes Ambisonics directos y difusos de un orden y modo deseados.

60

25. Programa informático para realizar, cuando se ejecuta en un ordenador o procesador, el procedimiento de generación de una descripción de campo de sonido que tiene componentes de campo de sonido según la

reivindicación 24.

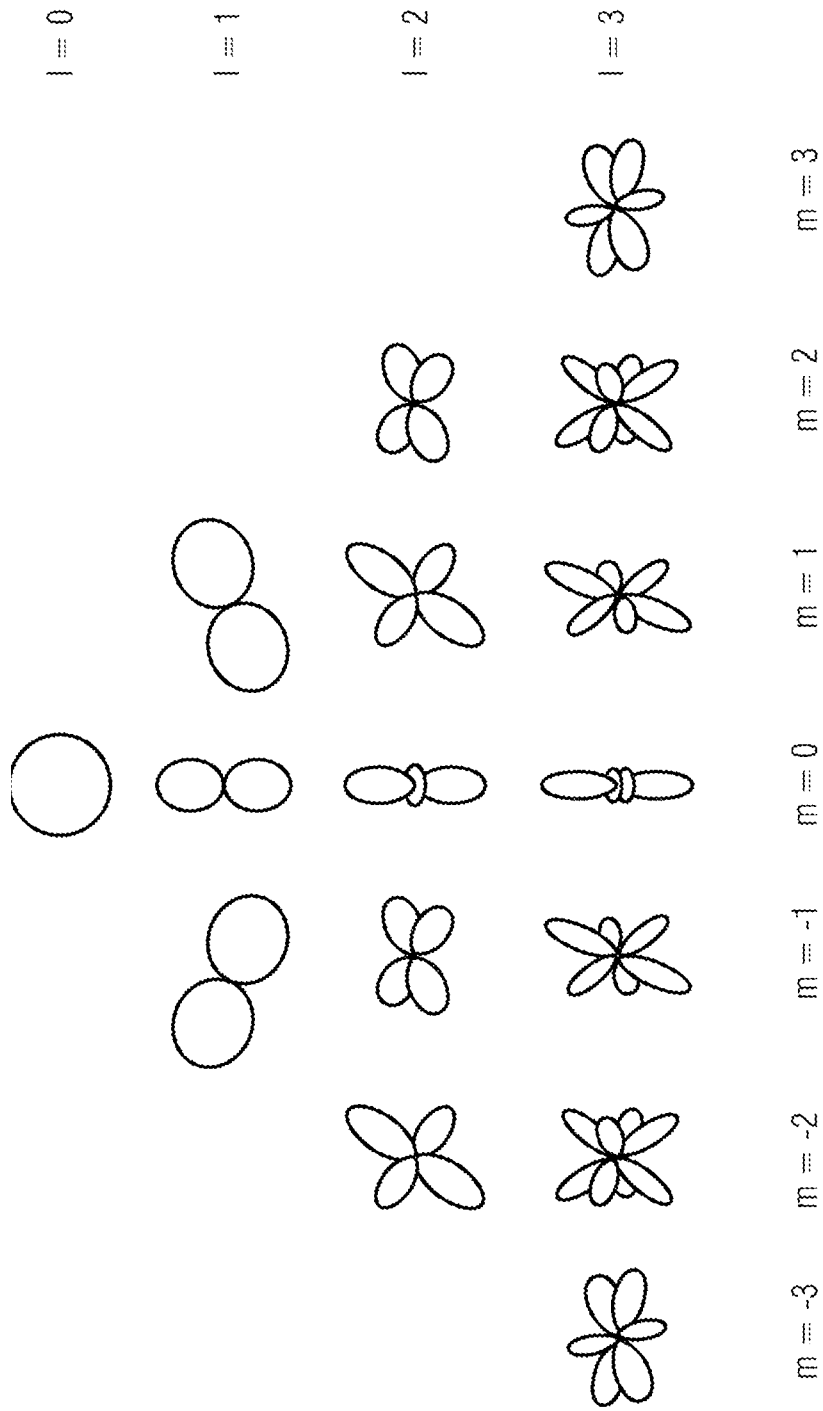


Fig. 1a

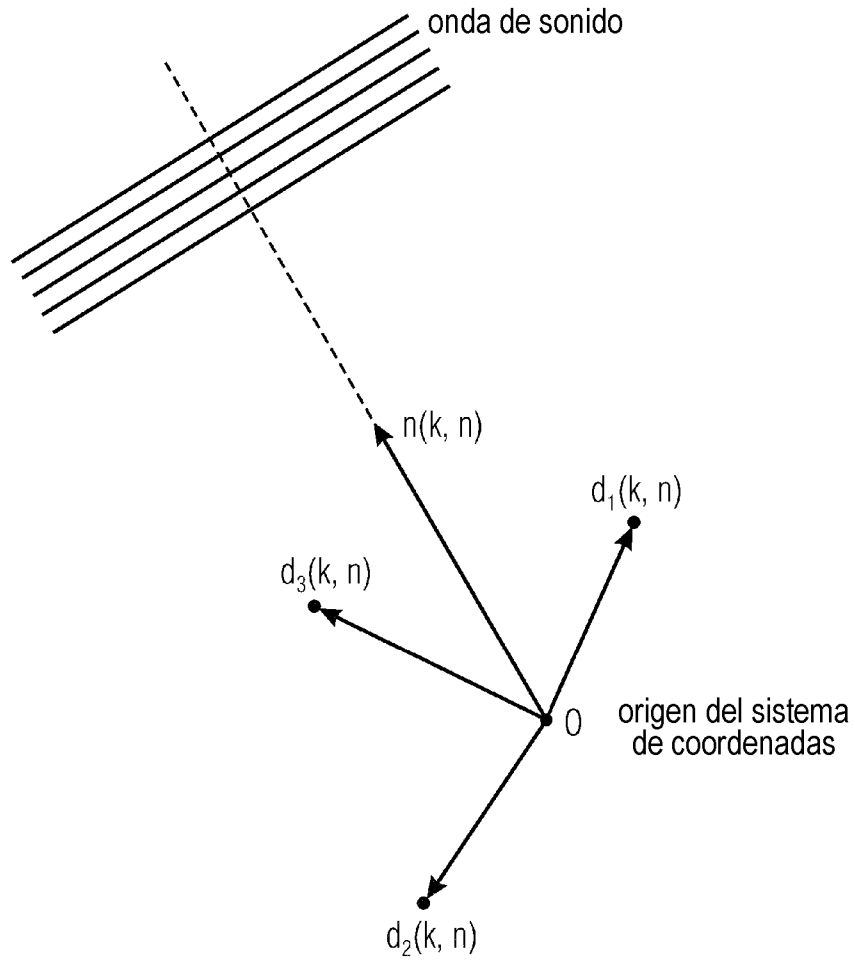


Fig. 1b

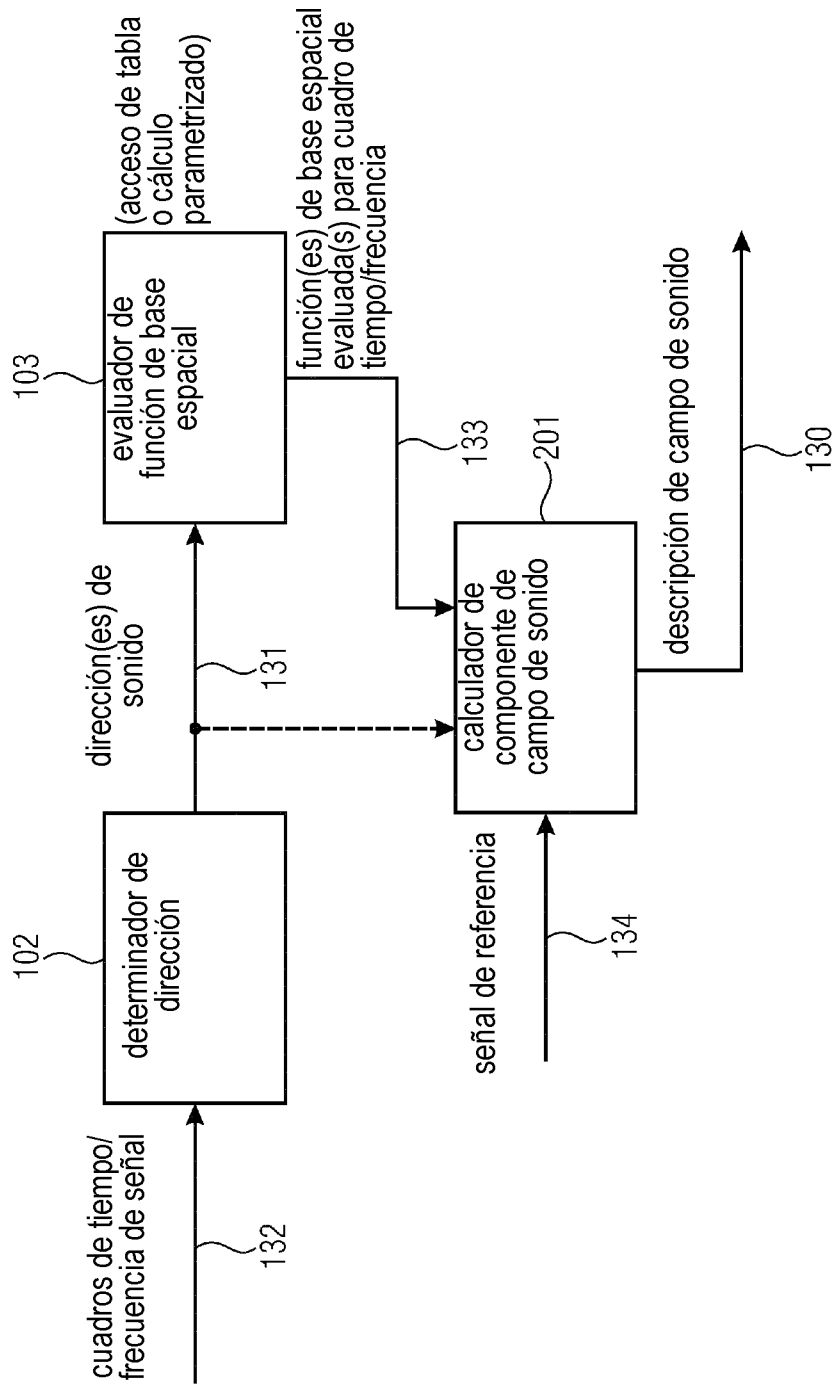


Fig. 1c

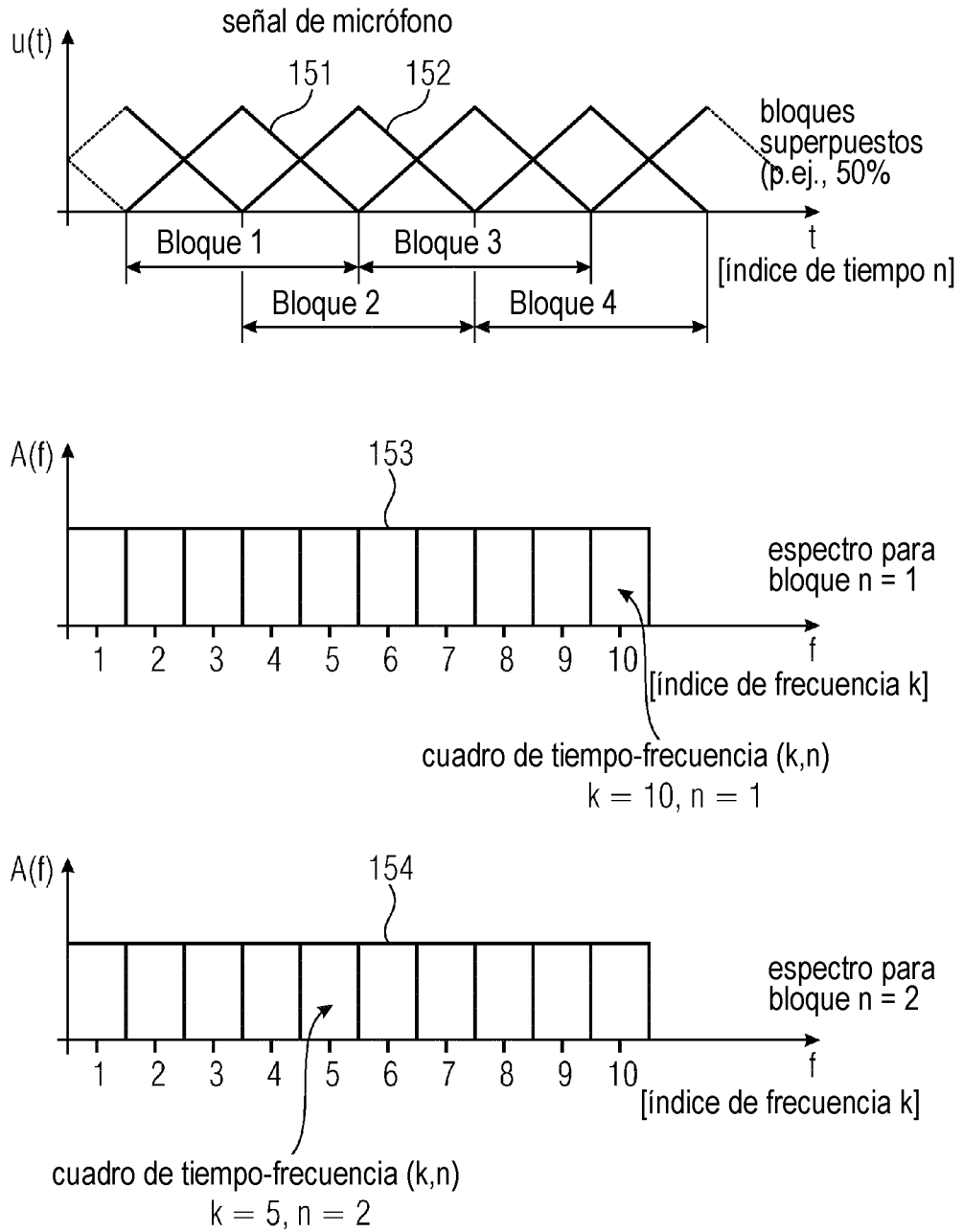


Fig. 1d

- cuadro de tiempo-frecuencia (10, 1)

tiene DOA $\vec{n} (10, 1)$ \vec{n} vector normal unitario

- cuadro de tiempo-frecuencia (5, 2) $\vec{n} = \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix}$ caso 3D

tiene DOA $\vec{n} (5, 2)$ $\vec{n} = \begin{pmatrix} n_x \\ n_y \end{pmatrix}$ caso 2D

- funciones de base espacial Y_i (DOA)

p.ej., $i = 1$: omnidireccional
 $i = 2$: direccional en dirección X
 $i = 3$: direccional en dirección Y
 $i = 4$: direccional en dirección Z

⋮



funciones de base espacial evaluadas $G_i(k,n)$

para cuadros de tiempo-frecuencia

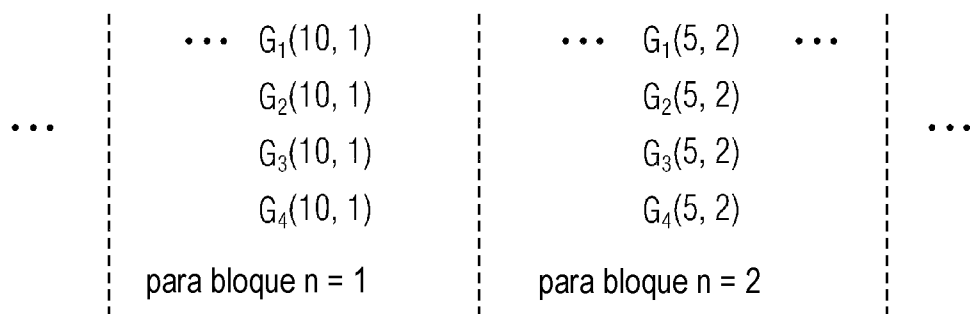
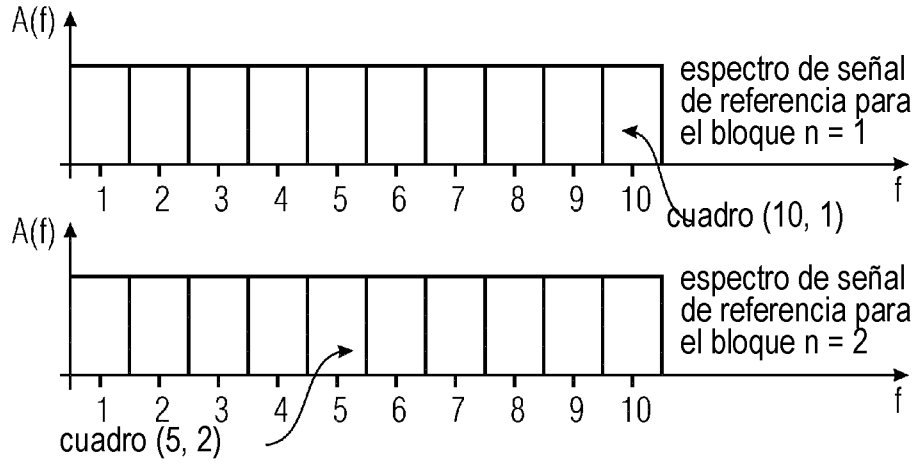


Fig. 1e

cálculo de componentes de campo de sonido

señal de referencia en representación de tiempo-frecuencia



$$B_i = f(P, G_i) \quad 155$$

representación (espectral) en el dominio de la frecuencia de los componentes de campo de sonido B_i

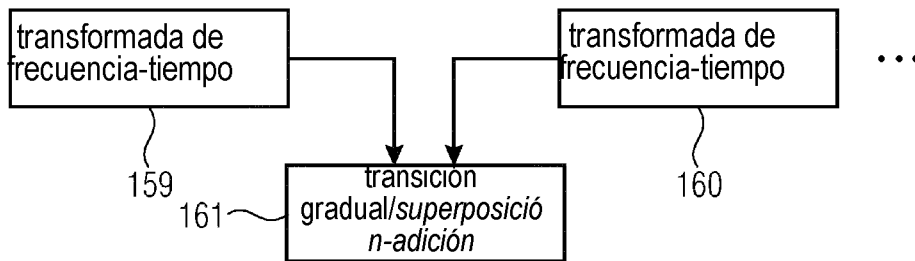
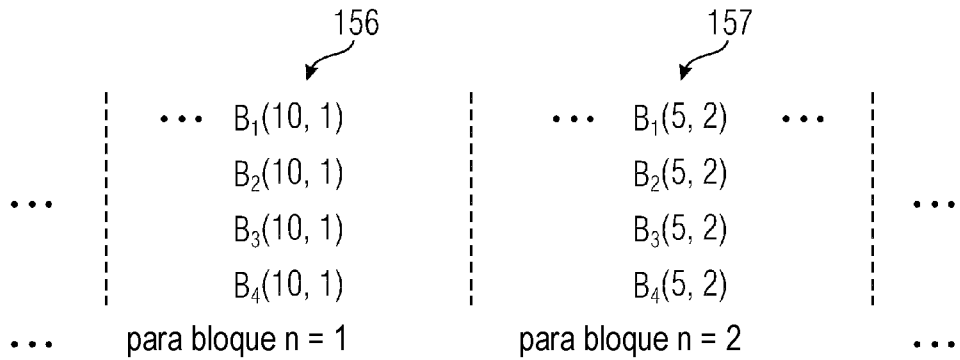
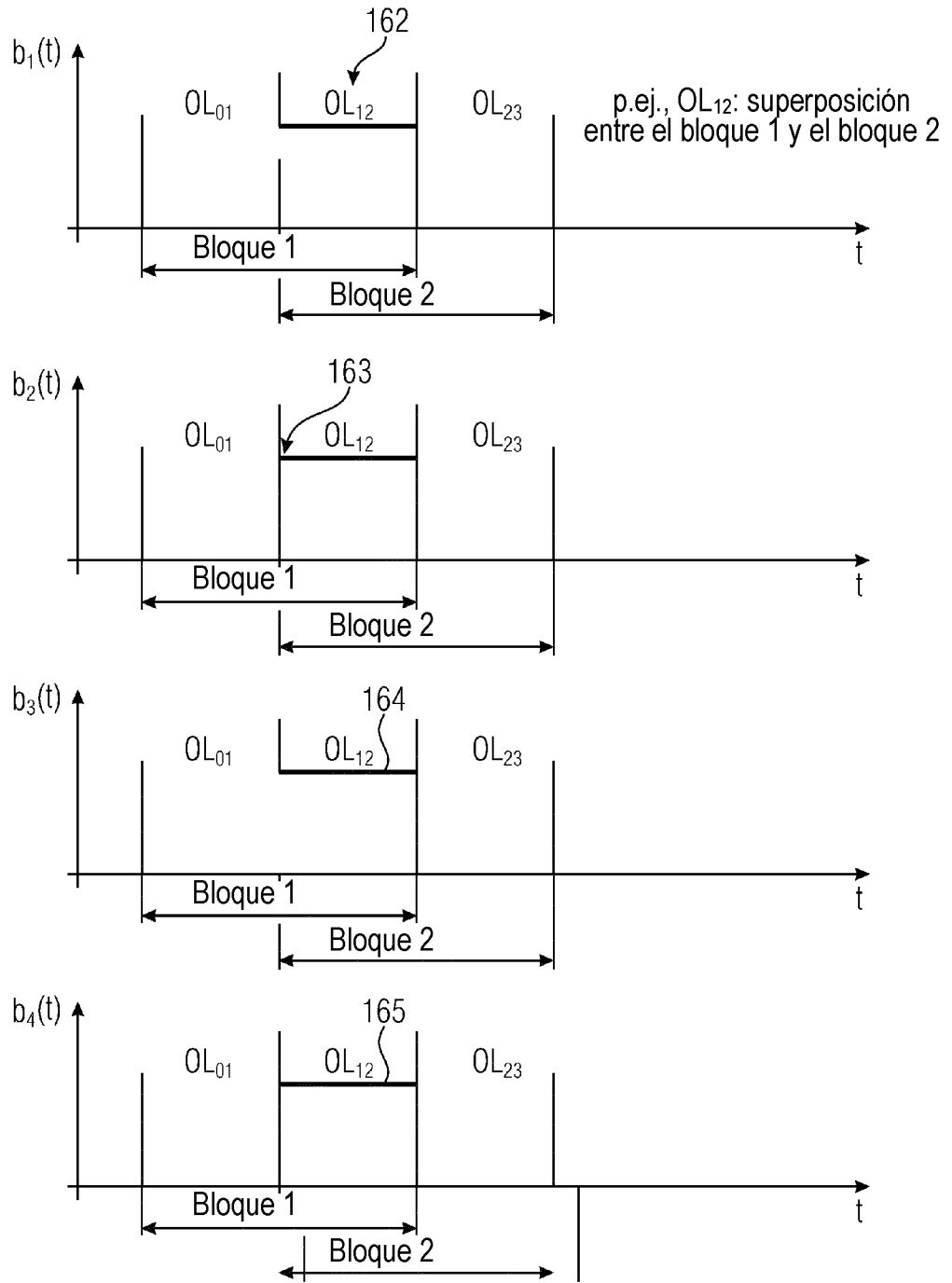


Fig. 1f



representación en el dominio del tiempo de los componentes de campo de sonido b_i

Fig. 1g

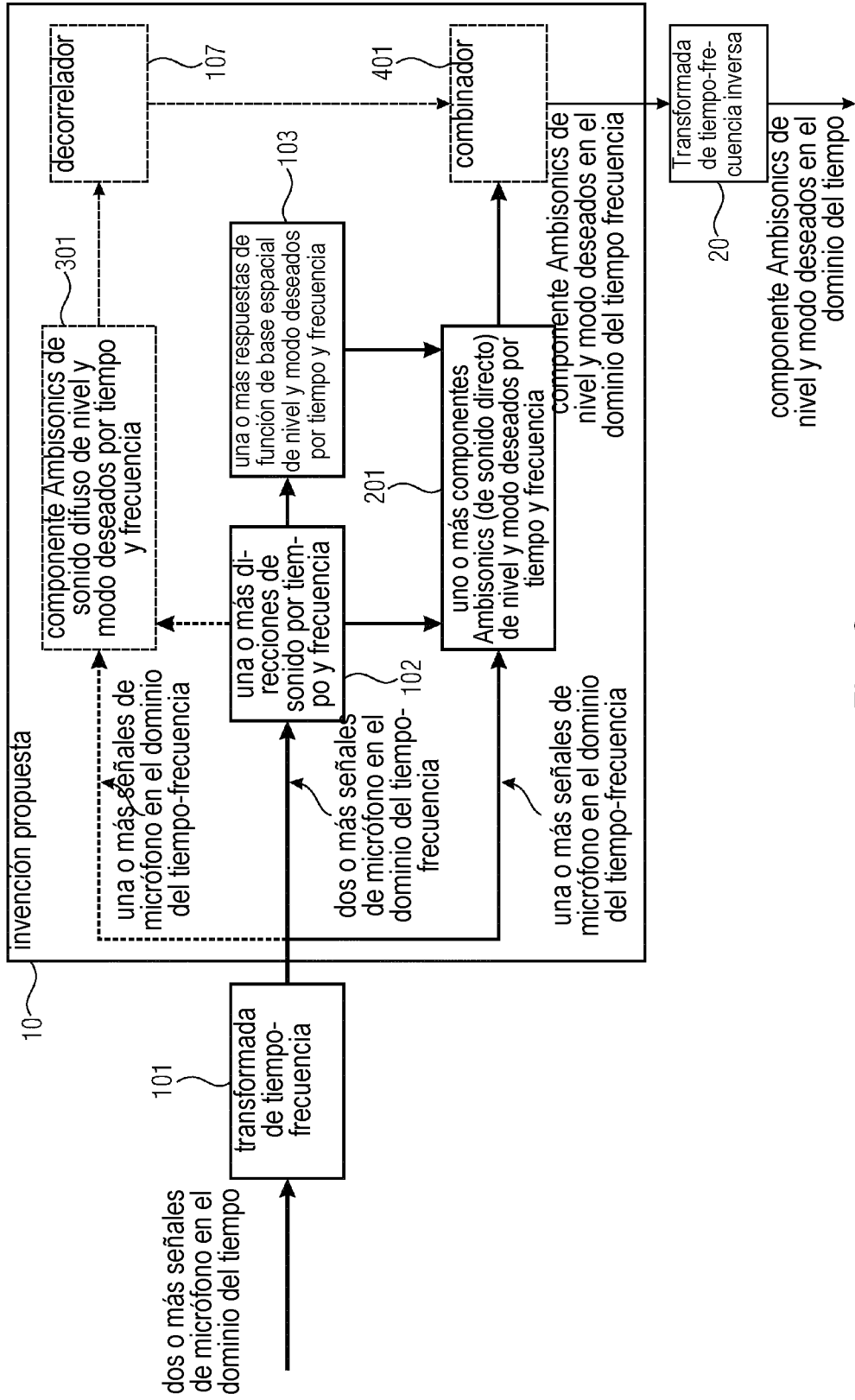


Fig. 2a

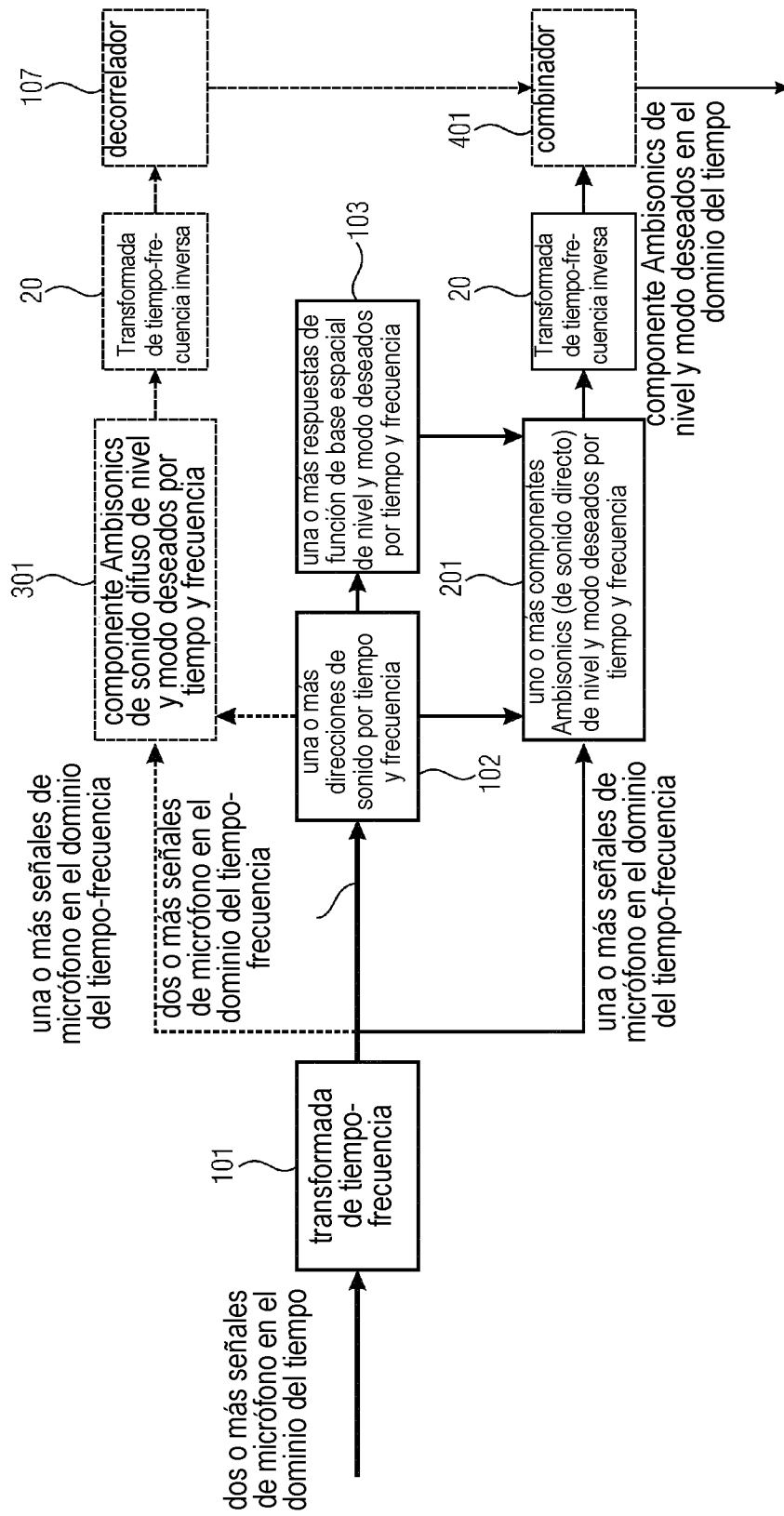


Fig. 2b

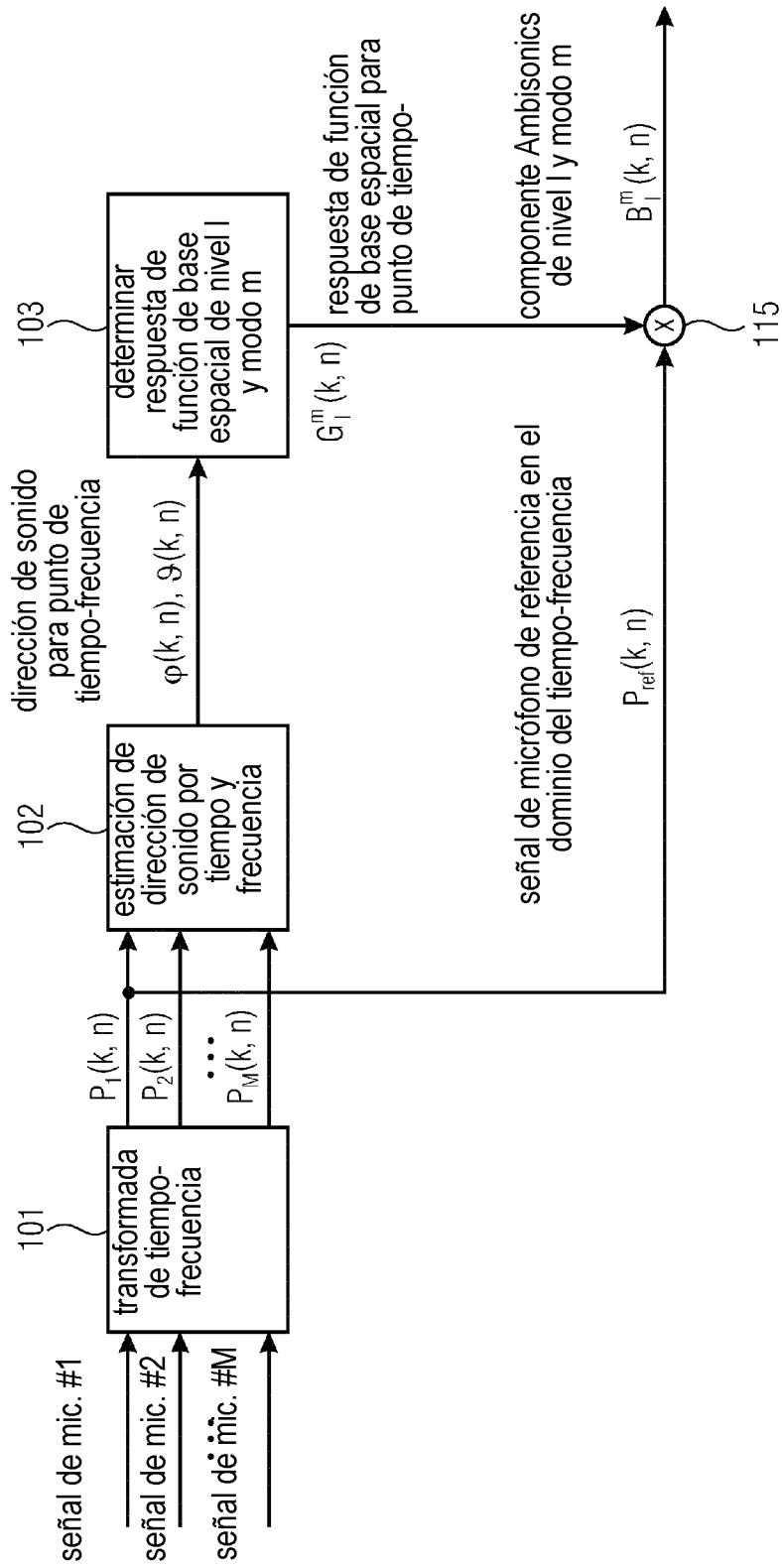


Fig. 3a

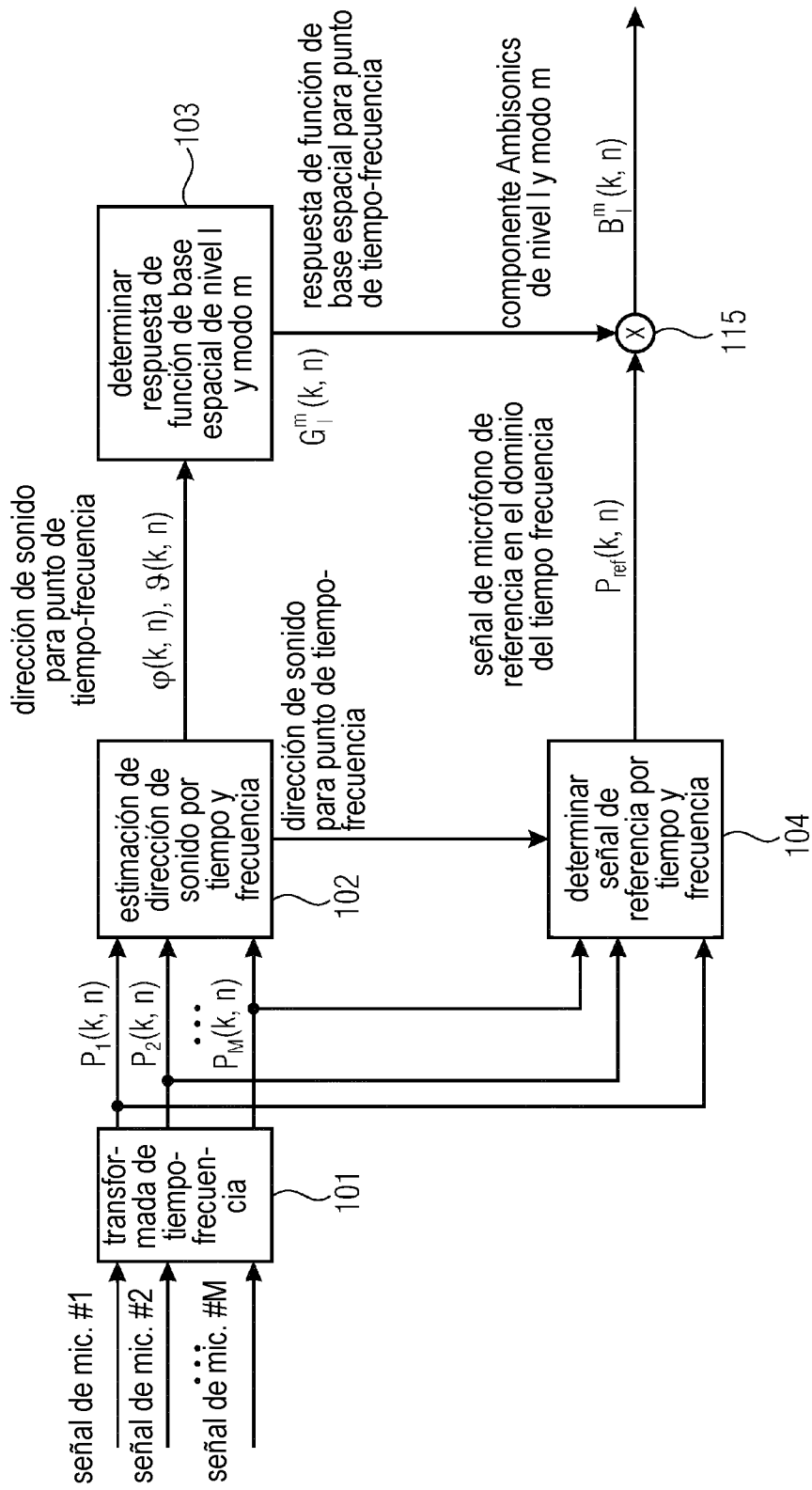


Fig. 3b

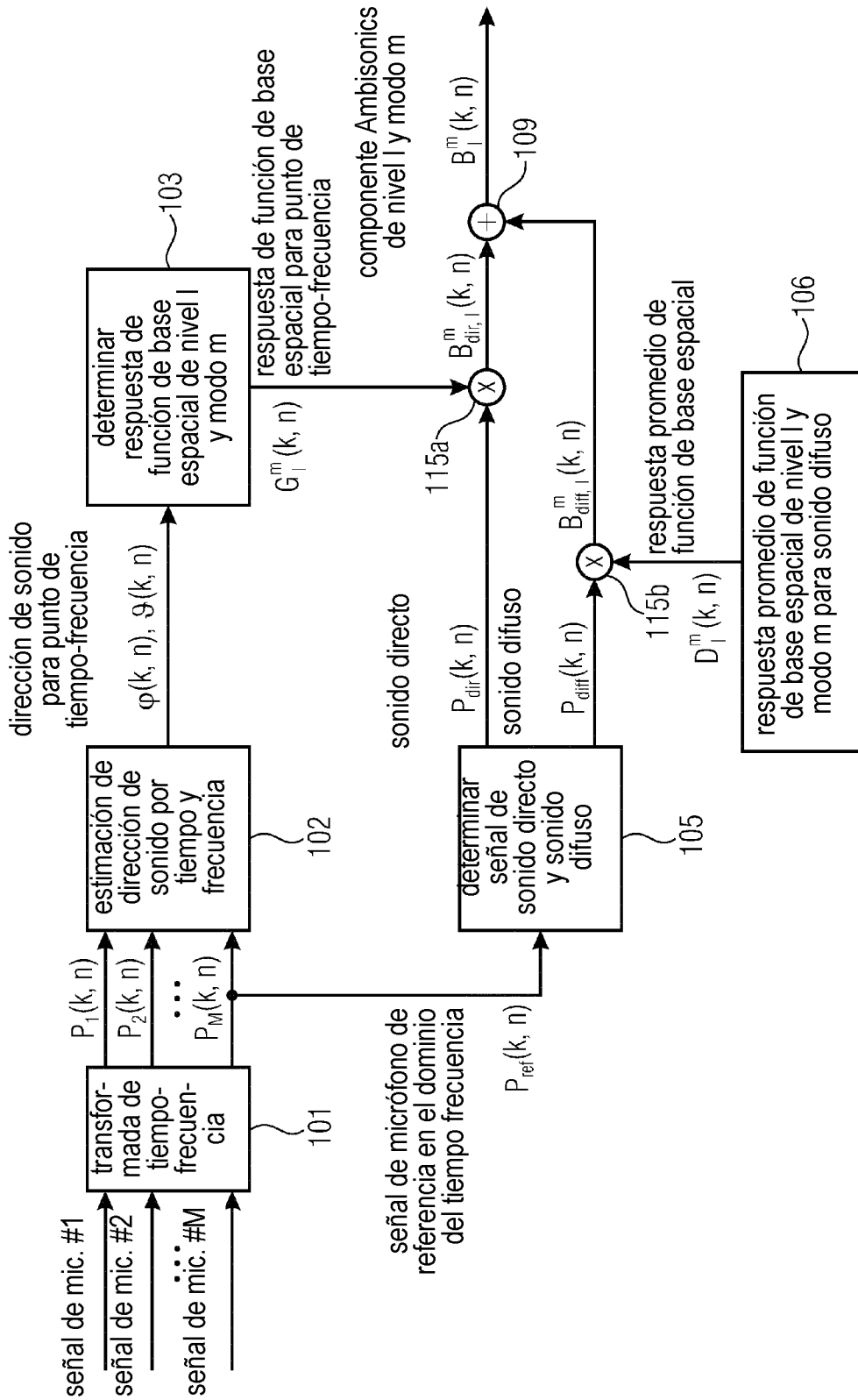


Fig. 4

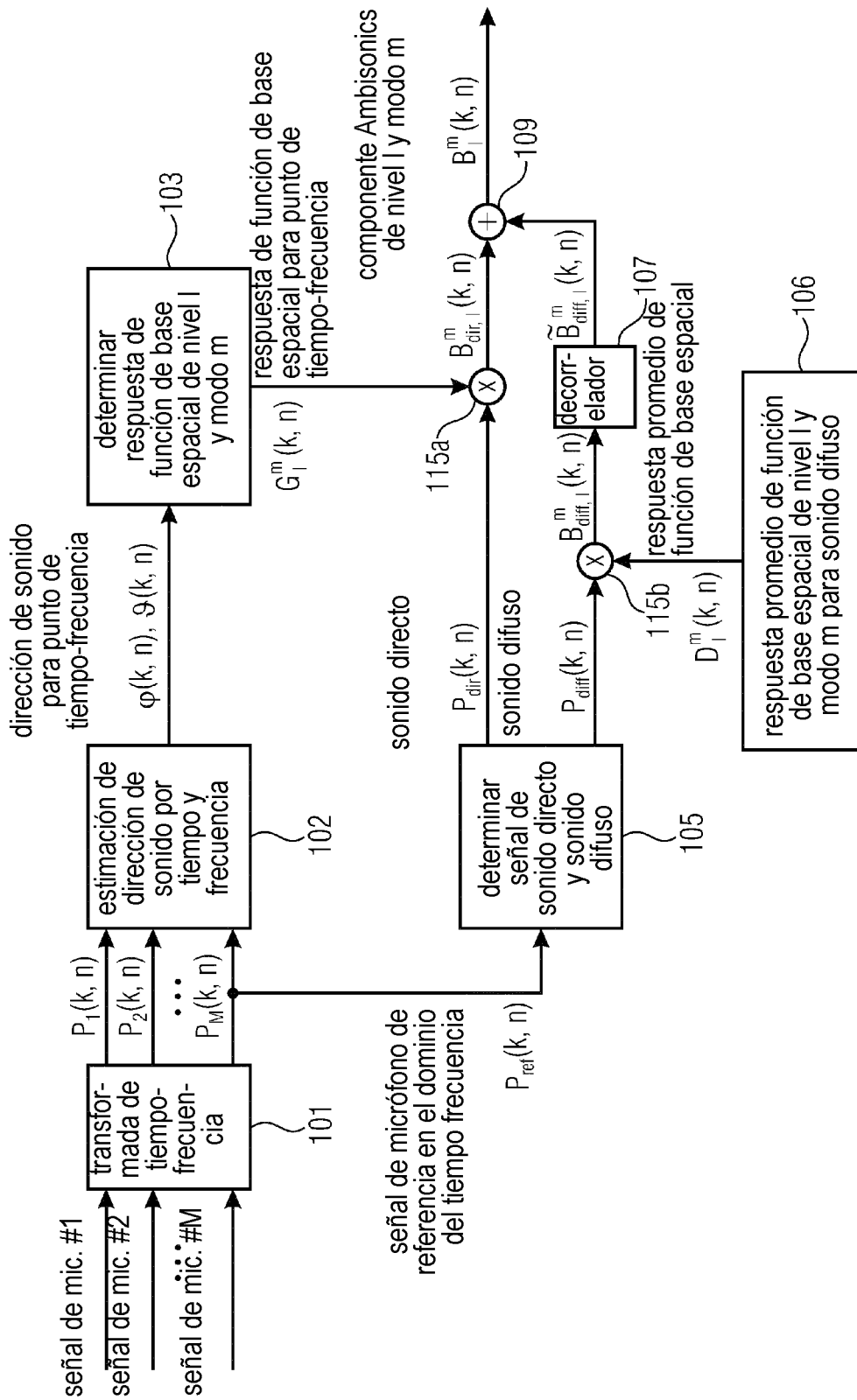


Fig. 5

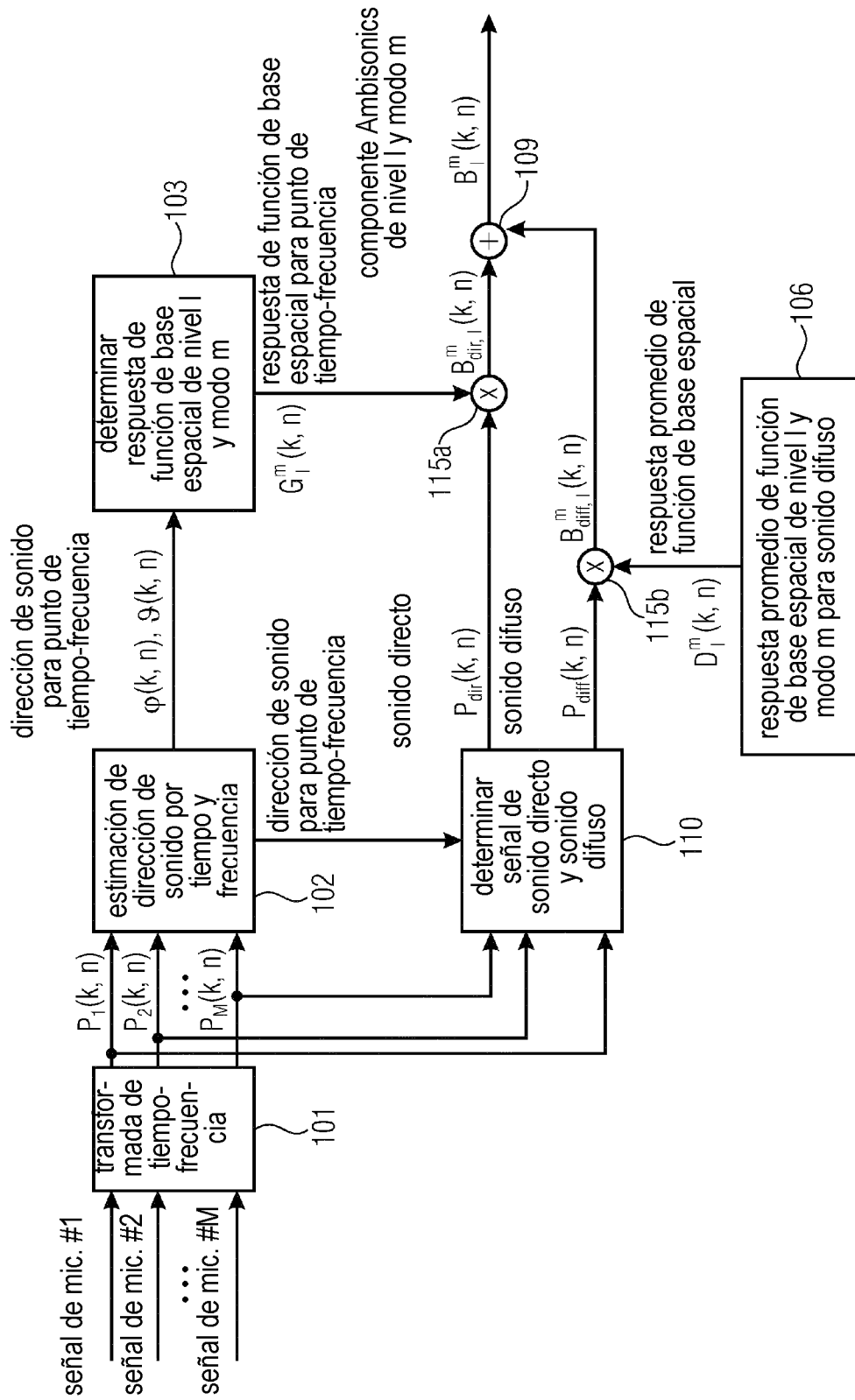


Fig. 6

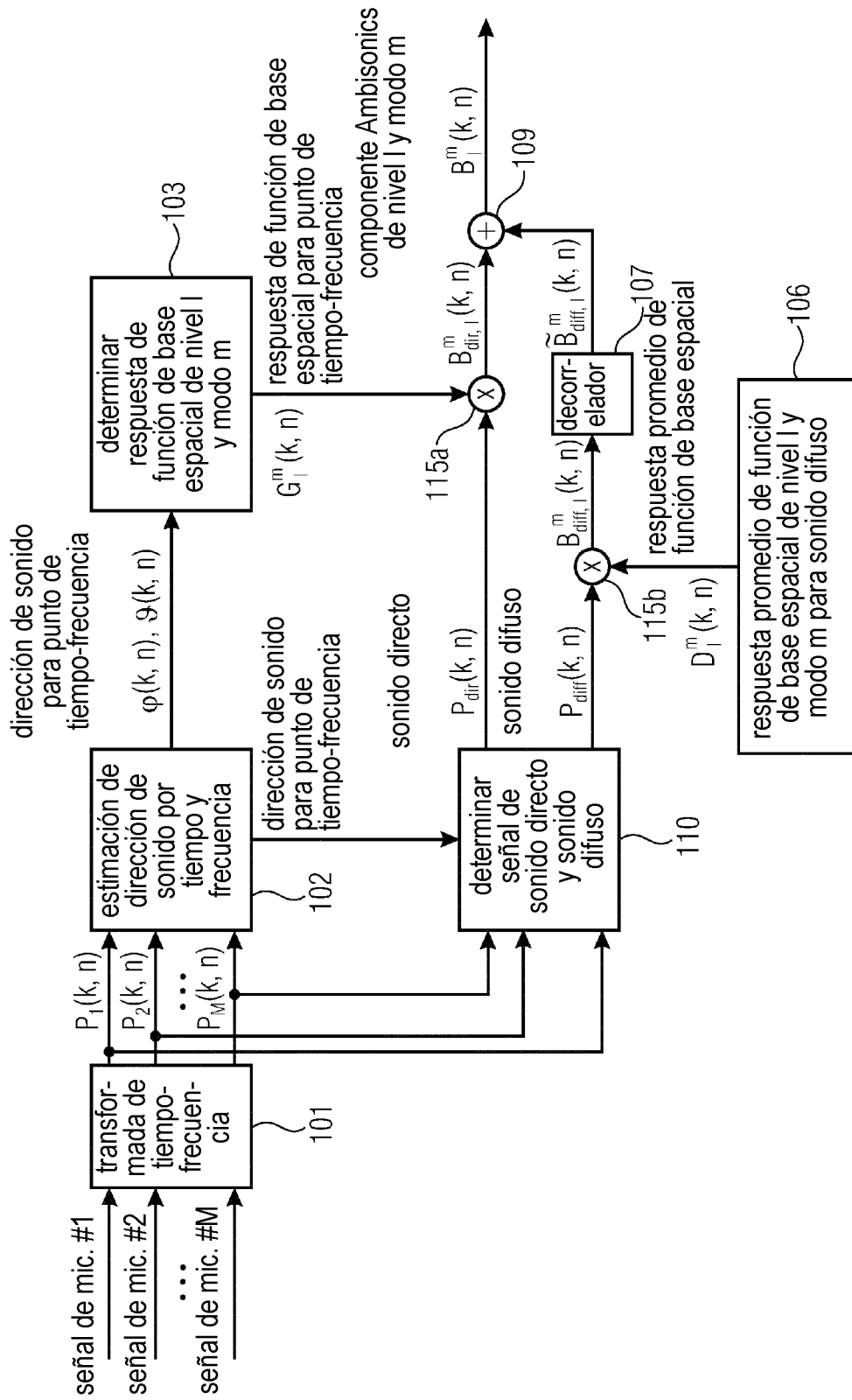


Fig. 7

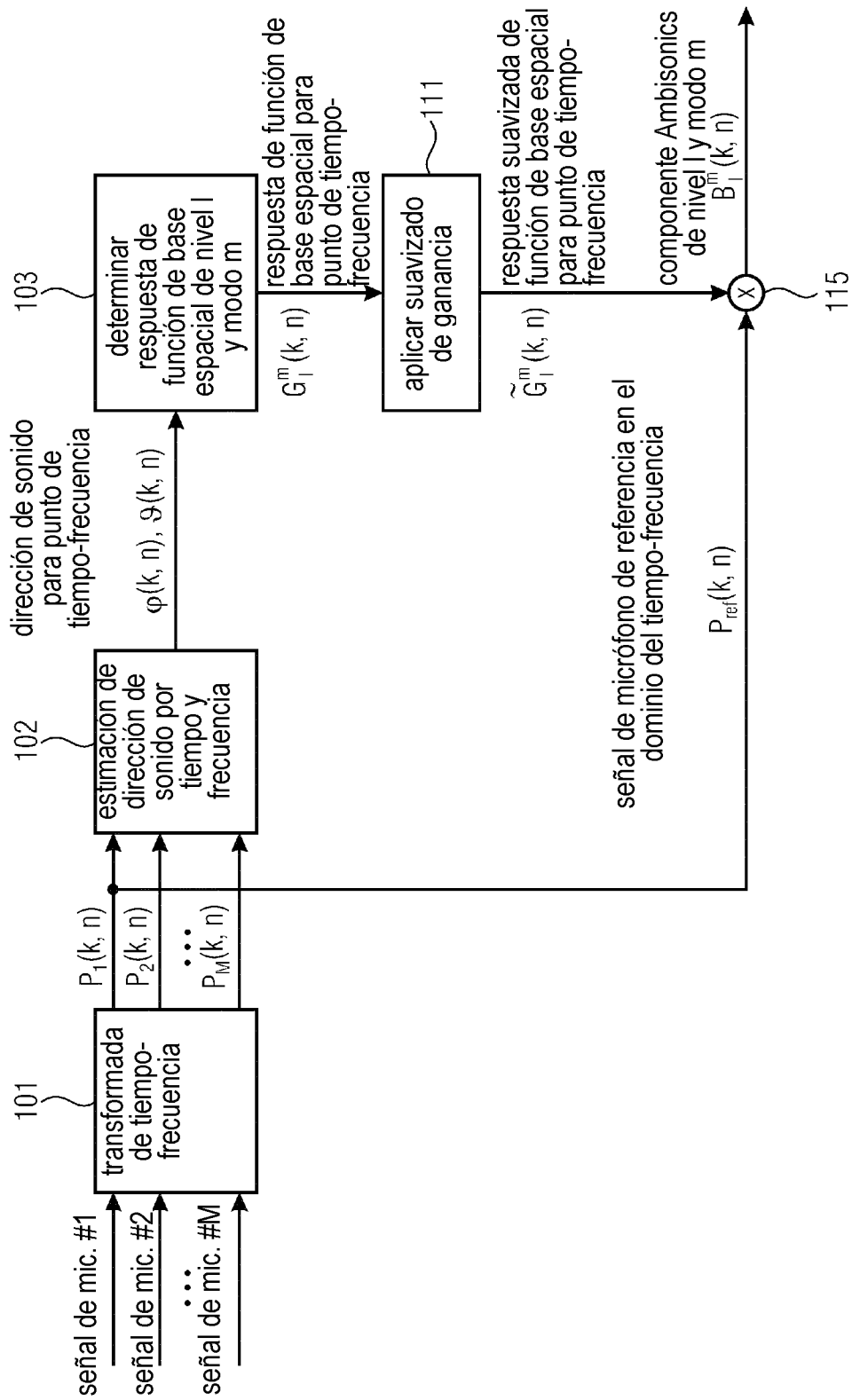


Fig. 8