(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 612 528**

(51) Int. CI.:

H04R 3/00 (2006.01)
G06F 3/16 (2006.01)
G10K 11/34 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **25.11.2013 PCT/EP2013/074650**

(87) Fecha y número de publicación internacional: **26.06.2014 WO2014095250**

(96) Fecha de presentación y número de la solicitud europea: **25.11.2013 E 13802283 (5)**

(97) Fecha y número de publicación de la concesión europea: **05.10.2016 EP 2936830**

(54) Título: **Filtro y procedimiento de filtrado espacial informado utilizando múltiples estimaciones instantáneas de la dirección de llegada**

(30) Prioridad:

21.12.2012 US 201261740866 P
24.05.2013 EP 13169163

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
17.05.2017

(73) Titular/es:

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

(72) Inventor/es:

HABETS, EMANUEL;
THIERGART, OLIVER;
BRAUN, SEBASTIAN y
TASESKA, MAJA

(74) Agente/Representante:

SALVA FERRER, Joan

DESCRIPCIÓN

Filtro y procedimiento de filtrado espacial informado utilizando múltiples estimaciones instantáneas de la dirección de llegada.

[0001] La presente invención se refiere al procesamiento de señal de audio, y, en particular, a un filtro y un procedimiento para el filtrado espacial informado utilizando múltiples estimaciones instantáneas de la dirección de llegada.

[0002] La extracción de las fuentes de sonido en condiciones de ruido y reverberantes se encuentra comúnmente en los sistemas de comunicación modernos. En las últimas cuatro décadas se han propuesto una gran variedad de técnicas de filtrado espacial para realizar esta tarea. Los filtros espaciales existentes son óptimos si las señales observadas se ajustan al modelo de señal y si la información necesaria para calcular los filtros es exacta. En la práctica, sin embargo, el modelo de la señal a menudo no se cumple y la estimación de la información necesaria es un reto importante.

[0003] Los filtros espaciales existentes se pueden clasificar, a grandes rasgos, en filtros espaciales lineales (véase, por ejemplo, [1, 2, 3, 4]) y filtros espaciales paramétricos (ver, por ejemplo, 5, 6, 7, 8). En general, los filtros espaciales lineales requieren la estimación de uno o más vectores de propagación o las estadísticas de segundo orden (SOS) de la fuente deseada o más, además de las estadísticas de segundo orden de la interferencia. Algunos filtros espaciales están diseñados para extraer una única fuente de señal, ya sea reverberante o no reverberante (véase, por ejemplo, 9, 10, 11, 12, 13, 14, 15, 16), mientras que otros filtros se han diseñado para extraer la suma de dos o más fuentes de señales reverberantes (véase, por ejemplo, 17, 18). Los procedimientos anteriormente mencionados requieren un conocimiento previo de la dirección de la o las fuentes deseadas, o un período en el que sólo las fuentes deseadas son activas, por separado o simultáneamente. El documento WO2005 004532 muestra un conformador de haz superdirectivo que utiliza un parámetro de regularización finito para calcularlos los coeficientes del filtro.

[0004] Un inconveniente de estos procedimientos es la incapacidad para adaptarse con suficiente rapidez a nuevas situaciones, por ejemplo, a movimientos de la fuente o a altavoces que compiten y se activan cuando la fuente deseada está activa. Los filtros espaciales paramétricos se basan a menudo en un modelo de señal relativamente simple, por ejemplo, la señal recibida en el dominio tiempo-frecuencia consiste en una única onda plana además de sonido difuso, y se calcula a partir de estimaciones instantáneas de los parámetros del modelo. Las ventajas de los filtros espaciales paramétricos son la respuesta direccional altamente flexible, la supresión comparativamente potente del sonido difuso y las fuentes de interferencia, y la capacidad de adaptarse rápidamente a las nuevas situaciones. Sin embargo, como se muestra en 19, el modelo de la señal de onda plana única subyacente se puede violar fácilmente en la práctica, lo que degrada altamente el funcionamiento de los filtros espaciales paramétricos. Cabe señalar que los filtros espaciales paramétricos del estado de la técnica utilizan todas las señales de micrófono disponibles para estimar los parámetros del modelo, mientras que para calcular la señal de salida final se utiliza solamente una única señal de micrófono y una ganancia con valor real. Una ampliación que combine las múltiples señales de micrófono disponibles para encontrar una señal de salida reforzada no es fácil.

[0005] Por lo tanto, sería muy apreciable si se proporcionara una mejora de los conceptos que permitiera obtener la respuesta espacial deseada a las fuentes de sonido.

[0006] Por consiguiente, el objetivo de la presente invención es proporcionar una mejora de los conceptos para extraer fuentes de sonido. El objetivo de la presente invención se resuelve con un filtro de acuerdo con la reivindicación 1, con un procedimiento de acuerdo con la reivindicación 14 y con un programa de ordenador de acuerdo con la reivindicación 15.

[0007] Las realizaciones proporcionan un filtro espacial para obtener la respuesta deseada para como máximo L fuentes de sonido activas simultáneas. El filtro espacial proporcionado se obtiene minimizando la potencia de difusión más el ruido en la salida del filtro sujeto a L restricciones lineales. A diferencia de los conceptos del estado de la técnica, las restricciones L se basan en estimaciones instantáneas de la dirección de llegada de banda estrecha. Además se proporcionan nuevos estimadores de la relación difusión-ruido/potencia de difusión que presentan una resolución espectral y temporal suficientemente altas para lograr tanto la reducción del ruido como la supresión del efecto de reverberación.

[0008] De acuerdo con algunas realizaciones, se facilitan conceptos para la obtención de una respuesta

espacial deseada y arbitraria para como máximo L fuentes de sonido, que están simultáneamente activas en un instante de tiempo-frecuencia. Para este propósito, la información paramétrica instantánea (IPI) de la escena acústica se incorpora en el diseño de un filtro espacial que da como resultado un "filtro espacial informado".

[0009] En algunas realizaciones, dicho filtro espacial informado, por ejemplo, combina todas las señales de micrófono disponibles a partir de pesos complejos que proporcionan una señal de salida mejorada.

[0010] Según realizaciones, el filtro espacial informado puede, por ejemplo, realizarse como un filtro espacial de varianza mínima restringida linealmente (LCMV) o como un filtro de Wiener paramétrico multicanal.

[0011] En algunas realizaciones, el filtro espacial informado que se proporciona, por ejemplo, se obtiene minimizando la potencia de difusión más el propio ruido sujeto a L restricciones lineales.

[0012] En algunas realizaciones, a diferencia de la técnica anterior, las L restricciones se basan en estimaciones instantáneas de la dirección de llegada (DOA) y las respuestas resultantes a las L direcciones de llegada corresponden a la direccionalidad deseada específica.

[0013] Por otra parte, se proporcionan nuevos estimadores de la señal requerida y estadísticas del ruido, por ejemplo, la relación difusión-ruido (DNR), que presentan una resolución espectral y temporal suficientemente altas, por ejemplo, para reducir la reverberación y el ruido.

[0014] En los párrafos siguientes se describen con mayor detalle realizaciones de la presente invención en relación con las figuras en las que:

La figura 1a ilustra un filtro de acuerdo con una realización,
La figura 1b ilustra el escenario de una posible aplicación de un filtro de acuerdo con una realización,
La figura 2 ilustra un filtro de acuerdo con una realización y una pluralidad de micrófonos,
La figura 3 ilustra un generador de pesos de acuerdo con una realización,
La figura 4 ilustra una magnitud de dos respuestas de ejemplo de acuerdo con realización,
La figura 5 ilustra un generador de pesos de acuerdo con otra realización que implementa una estrategia de varianza mínima restringida linealmente,
La figura 6 ilustra un generador de pesos de acuerdo con una realización adicional que implementa una estrategia de filtro de Wiener paramétrico multicanal,
La figura 7 ilustra una relación difusión-ruido real y estimada como función del tiempo y la frecuencia,
La figura 8 ilustra el índice de direccionalidad y la ganancia de ruido blanco de filtros espaciales comparados,
La figura 9 ilustra una dirección de llegada estimada y la ganancia resultante, y
La figura 10 ilustra un ejemplo para el caso de la reproducción de altavoces estéreo.

[0015] La figura 1a ilustra un filtro 100 que genera una señal de salida de audio, que comprende una pluralidad de muestras de la señal de salida de audio, basado en dos o más señales de entrada de micrófono que se proporcionan. La señal de salida de audio y las dos o más señales de entrada de micrófono están representadas en un dominio de tiempo-frecuencia, en el que cada una de las pluralidades de muestras de la señal de salida de audio se asigna a un contenedor de tiempo-frecuencia *(k, n)* de una pluralidad de contenedores de tiempo-frecuencia *(k, n).*

[0016] El filtro 100 comprende un generador de pesos 110 que se ha adaptado para recibir, para cada una de las pluralidades de contenedores de tiempo-frecuencia *(k, n)*, información de la dirección de llegada de uno o más componentes de sonido de una o más fuentes de sonido o información de la posición de una o más fuentes de sonido, y que se ha adaptado para generar información sobre la ponderación de cada una de las pluralidades de contenedores de tiempo-frecuencia *(k, n)*, en función de la información de la dirección de llegada del o los componentes de sonido de las más de una fuentes de sonido de dicho contenedor de tiempo-frecuencia *(k, n)*, o en función de la información de la posición de la o las fuentes de sonido de dicho contenedor de tiempo-frecuencia *(k, n)*.

[0017] Por otra parte, el filtro comprende un generador de la señal de salida 120 que genera la señal de salida de audio generando para cada una de las pluralidades de contenedores de tiempo-frecuencia *(k, n)* una pluralidad de las muestras de la señal de salida de audio, que se asigna a dicho contenedor de tiempo-frecuencia *(k, n)*, en función de la información sobre ponderación de dicho contenedor de tiempo-frecuencia *(k, n)* y en función de una muestra de entrada de audio, que se asigna a dicho contenedor de tiempo-frecuencia *(k, n)*, de cada una de las

dos o más señales de entrada de micrófono.

[0018] Por ejemplo, cada una de las dos o más señales de entrada de micrófono comprende una pluralidad de muestras de entrada de audio, en la que cada una de las muestras de entrada de audio se asigna a uno de los contenedores de tiempo-frecuencia *(k, n)* y el generador de señal de audio 120 se puede adaptar para generar una pluralidad de las muestras de la señal de salida de audio, que se asigna a dicho contenedor de tiempo-frecuencia *(k, n)*, en función de la información sobre ponderación de dicho contenedor de tiempo-frecuencia *(k, n)* y en función de una de las muestras de entrada de audio de cada una de las dos o más señales de entrada de micrófono, concretamente, por ejemplo, en función de una de las muestras de entrada de audio de cada una de las dos o más señales de entrada de micrófono, que están asignadas a dicho contenedor de tiempo-frecuencia *(k, n)*.

[0019] Para cada muestra de señal de salida de audio que se va generar para cada contenedor de tiempo-frecuencia *(k, n)*, el generador de pesos 110 genera la información sobre ponderación individualmente de nuevo. El generador de la señal de salida 120 genera entonces la muestra de la señal de salida de audio para el contenedor tiempo-frecuencia examinado *(k, n)* basándose en la información sobre ponderación generada para este contenedor de frecuencia tiempo. En otras palabras, el generador de pesos 110 calcula la nueva información sobre ponderación para cada contenedor de frecuencia tiempo para el que se va a generar una muestra de la señal de salida de audio.

[0020] Cuando se genera la información sobre ponderación, el generador de pesos 110 se adapta para tener en cuenta la información de una o más fuentes de sonido.

[0021] Por ejemplo, el generador de pesos 110 puede tener en cuenta una posición de una primera fuente de sonido. En una realización, el generador de pesos también puede en cuenta una posición de una segunda fuente de sonido.

[0022] O, por ejemplo, la primera fuente de sonido puede emitir una primera onda sonora con un primer componente de sonido. La primera onda sonora con el primer componente de sonido llega a un micrófono y el generador de pesos 110 puede tener en cuenta la dirección de llegada del primer componente de sonido 1 de la onda sonora. Por esto, el generador de pesos 110 tiene en cuenta información de la primera fuente de sonido. Por otra parte, la segunda fuente de sonido puede emitir una segunda onda sonora con un segundo componente de sonido. La segunda onda sonora con el segundo componente de sonido llega a un micrófono y el generador de pesos 110 puede tener en cuenta la dirección de llegada del segundo componente de sonido / de la segunda onda sonora. Por esto, el generador de pesos 110 tiene en cuenta información de la segunda fuente de sonido.

[0023] La figura 1b ilustra el escenario de una posible aplicación de un filtro 100 de acuerdo con una realización. Una primera onda sonora con un primer componente de sonido se emite por un primer altavoz 121 (una primera fuente de sonido) y llega a un primer micrófono 111. Se tiene en cuenta la dirección de llegada del primer componente de sonido (= la dirección de llegada de la primera onda sonora) en el primer micrófono 111. Por otra parte, una segunda onda sonora con un segundo componente de sonido se emite por un segundo altavoz 122 (una segunda fuente de sonido) y llega al primer micrófono 111. El generador de pesos 110 es capaz de tener en cuenta también la dirección de llegada del segundo componente de sonido en el primer micrófono 111 para determinar la información sobre ponderación. Por otra parte, la dirección de llegada de los componentes de sonido (= dirección de llegada de las ondas sonoras) en los otros micrófonos, tales como el micrófono 112 también se pueden tener en cuenta por el generador de pesos para determinar la información sobre ponderación.

[0024] Cabe señalar que las fuentes de sonido pueden, por ejemplo, ser fuentes de sonido físicas que existen físicamente en un entorno, por ejemplo, altavoces, instrumentos musicales o una persona que habla.

[0025] Sin embargo, hay que señalar que las fuentes de imágenes reflejadas son también fuentes de sonido. Por ejemplo, una onda sonora emitida por un altavoz 122 se puede reflejar en una pared 125 y luego parece que la onda sonora es emitida desde una posición 123 diferente de la posición del altavoz que, en realidad, emitió la onda sonora. Dicha fuente de imagen reflejada 123 también se considera una fuente de sonido. Un generador de pesos 110 se puede adaptar para generar la información sobre ponderación en función de la información de la dirección de llegada relativa a una fuente de imagen reflejada o en función de la información de la posición de una, dos o más fuentes de imágenes reflejadas.

[0026] La figura 2 ilustra un filtro 100 de acuerdo con una realización y una pluralidad de micrófonos 111, 112, 113, ... , 11n. En la realización de la fig. 2, el filtro 100 además comprende un banco de filtros 101. Por otra parte, en la realización de la fig. 2, el generador de pesos 110 comprende un módulo de cálculo de la información

102, un módulo de cálculo de los pesos 103 y un módulo de selección de la función de transferencia 104.

[0027] El procesamiento se lleva a cabo en un dominio tiempo-frecuencia en el que *k* indica el índice de frecuencia y *n* indica el índice de tiempo, respectivamente. En la entrada del aparato (el filtro 100) hay *M* señales de micrófono en el dominio del tiempo $x_1 \ldots {}_M(t)$ procedentes de los micrófonos 111, 112, 113, ... , 11n, que se transforman en un dominio tiempo-frecuencia mediante el banco de filtros 101. Las señales de micrófono transformadas vienen dadas por el vector

$$\mathbf{x}(k,n) = [\mathrm{X}_1(k, n)\, \mathrm{X}_2(k, n) \ldots \mathrm{X}_M(k, n)]^{\mathrm{T}}.$$

[0028] El filtro 100 genera una señal deseada *Y(k, n)* (la señal de salida de audio). La señal de audio de salida (señal deseada) *Y(k, n)* puede, por ejemplo, representar una señal mejorada para la reproducción mono, una señal de auricular para la reproducción de sonido binaural o una señal de altavoz para la reproducción de sonido espacial con una configuración de altavoces arbitraria.

[0029] La señal deseada *Y(k, n)* es generada por un generador de la señal de salida 120, por ejemplo, llevando a cabo una combinación lineal de las *M* señales de micrófono *x(k, n)* a partir de pesos complejos instantáneos $w(k, n) = [W_1(k, n)\, W_2(k, n) \ldots W_M(k, n)]^{\mathrm{T}}$, por ejemplo, empleando la fórmula

$$Y(k, n) = \mathbf{w}^{\mathrm{H}}(k, n)\, \mathbf{x}(k, n). \qquad (1)$$

[0030] Los pesos *w(k, n)* se determinan mediante el módulo de cálculo de pesos 103. Para cada *k* y cada *n*, los pesos *w(k, n)* se determinan nuevamente. En otras palabras, para cada contenedor de tiempo-frecuencia *(k, n),* se lleva a cabo una determinación de los pesos *w(k, n).* Más específicamente, los pesos *w(k, n)* son, por ejemplo, calculados a partir de la información paramétrica instantánea (IPI) $\Im$ *(k, n)* y a partir de la función de transferencia deseada correspondiente G (k, n).

[0031] El módulo de cálculo de la información 102 está configurado para calcular la IPI $\Im$ *(k, n)* a partir de las señales del micrófono x*(k, n).* La IPI describe características específicas de los componentes de señal y ruido comprendidos en las señales de micrófono x*(k, n)* para el instante de tiempo-frecuencia dado *(k, n).*

[0032] La figura 3 ilustra un generador de pesos 110 de acuerdo con una realización. El generador de pesos 110 comprende un módulo de cálculo de la información 102, un módulo de cálculo de los pesos 103 y un módulo de selección de la función de transferencia 104.

[0033] Como se muestra en el ejemplo de la fig. 3, la IPI comprende principalmente la dirección de llegada (DOA) instantánea de uno o más componentes de sonido direccional (por ejemplo, ondas planas), por ejemplo, calculada por un módulo de estimación de la DOA 201.

[0034] Como se explica a continuación, la información de la DOA se puede representar como un ángulo (por ejemplo, por [ángulo de acimut $\varphi(k, n)$, ángulo de elevación $\vartheta(k, n)$]), por una frecuencia espacial (por ejemplo, por μ [k | φ (*k, n*)]), por un desplazamiento de fase (por ejemplo, por a[k | $\varphi(k, n)$]) por un retardo temporal entre los micrófonos, por un vector de propagación (por ejemplo, por **a**[k | $\varphi(k, n)$]), o por una diferencia de nivel interaural (ILD) o por una diferencia de tiempo interaural (ITD).

[0035] Por otra parte, la IPI $\Im$ *(k, n)* puede, por ejemplo, comprender información adicional, por ejemplo, las estadísticas de segundo orden (SOS) de los componentes de señal o de ruido.

[0036] En una realización, el generador de pesos 110 se ha adaptado para generar la información sobre ponderación para cada una de las pluralidades de contenedores de tiempo-frecuencia *(k, n)* en función de la información estadística de los componentes de señal o de ruido de las dos o más señales de entrada de micrófono. Dicha información estadística son, por ejemplo, las estadísticas de segundo orden que se mencionan en este documento. La información estadística puede, por ejemplo, ser una potencia de un componente de ruido, una información de señal-difusión, una información señal-ruido, una información difusión-ruido, una potencia de un

componente de señal, una potencia de un componente de difusión o una matriz de densidades espectrales de la potencia de un componente de señal o de un componente de ruido de las dos o más señales de entrada de micrófono.

[0037] Las estadísticas de segundo orden se pueden calcular mediante un módulo de cálculo de estadísticas 205. Esta información de estadísticas de segundo orden puede, por ejemplo, comprender la potencia de un componente de ruido estacionario (por ejemplo, el propio ruido), la potencia de un componente de ruido no estacionario (por ejemplo, el sonido difuso), la relación señal-difusión (SDR), la relación señal-ruido (SNR) o la relación difusión-ruido (DNR). Esta información permite calcular los pesos óptimos $\mathbf{w}(k, n)$ en función de un criterio de optimización específico.

[0038] Un "componente de ruido estacionario"/"componente de ruido que varía lentamente" es, por ejemplo, un componente de ruido con estadísticas que no cambian o cambian lentamente con respecto al tiempo.

[0039] Un "componente de ruido no estacionario" es, por ejemplo, un componente de ruido con estadísticas que cambian rápidamente con el tiempo.

[0040] En una realización, el generador de pesos 110 se ha adaptado para generar la información sobre ponderación para cada una de las pluralidades de contenedores de tiempo-frecuencia (k, n) en función de la primera información de ruido que indica la información de los primeros componentes de ruido de las dos o más señales de entrada de micrófono y en función de la segunda información de ruido que indica la información de los segundos componentes de ruido de las dos o más señales de entrada de micrófono.

[0041] Por ejemplo, los primeros componentes de ruido pueden ser componentes de ruido no estacionario y la primera información de ruido puede ser información de componentes de ruido no estacionario.

[0042] Los segundos componentes de ruido pueden, por ejemplo, ser componentes de ruido estacionarios / componentes de ruido que varían lentamente y la segunda información de ruido puede ser información de los componentes de ruido estacionarios / que varían lentamente.

[0043] En una realización, el generador de pesos 110 está configurado para generar la primera información de ruido (por ejemplo, información sobre los componentes de ruido no estacionarios / que no varían lentamente) empleando, por ejemplo información estadística y predefinida (por ejemplo, información sobre una coherencia espacial entre dos o más señales de entrada de micrófono que resultan de los componentes de ruido no estacionarios) y en la que el generador de pesos 110 se ha configurado para generar la segunda información de ruido (por ejemplo, información sobre los componentes de ruido estacionarios / que varían lentamente) sin el empleo de la información estadística.

[0044] En cuanto a los componentes de ruido que cambian rápidamente, las señales de entrada de micrófono por sí solas no proporcionan información suficiente para determinar la información sobre dichos componentes de ruido. La información estadística es, por ejemplo, además, necesaria para determinar la información relativa a los componentes de ruido que cambian rápidamente.

[0045] Sin embargo, con respecto a los componentes de ruido que no cambian o cambian lentamente, la información estadística no es necesaria para determinar la información sobre estos componentes de ruido. En lugar de ello es suficiente evaluar las señales de micrófono.

[0046] Cabe señalar que la información estadística se puede calcular aprovechando la información estimada de la DOA como se muestra en la fig. 3. Además se debe señalar que la IPI también se puede proporcionar externamente. Por ejemplo, la DOA del sonido (la posición de las fuentes de sonido, respectivamente) se puede determinar mediante una cámara de vídeo junto con un algoritmo de reconocimiento de rostros suponiendo que emisores humanos constituyan la escena de sonido.

[0047] Un módulo de selección de la función de transferencia 104 se ha configurado para proporcionar una función de transferencia $G(k, n)$. La función de transferencia (potencialmente compleja) $G(k, n)$ de la fig. 2 y la fig. 3 describe la respuesta deseada del sistema dada el (por ejemplo, paramétrico actual) IPI $\vec{\Psi}(k, n)$. Por ejemplo, $G(k, n)$ puede describir un patrón de captación arbitrario de un micrófono espacial deseado para la mejora de la señal en la reproducción mono, una ganancia del altavoz dependiente de la DOA para la reproducción de altavoces o una función de transferencia relativa a la cabeza (HRTF) para la reproducción binaural.

[0048] Cabe señalar que, por lo general, las estadísticas de una escena de sonido grabado varían rápidamente en el tiempo y la frecuencia. En consecuencia, la IPI $\Im$ *(k, n)*.y los pesos óptimos **w***(k, n)* correspondientes son válidos sólo para un índice de tiempo-frecuencia específico y, por lo tanto, se vuelven a calcular para cada *k* y *n*. Así el sistema puede adaptarse de forma instantánea a la situación de grabación actual.

[0049] Además, debe tenerse en cuenta que los *M* micrófonos de entrada pueden, o bien formar una única matriz de micrófonos, o bien pueden estar distribuidos formando múltiples matrices en diferentes ubicaciones. Por otra parte, la IPI $\Im$ *(k, n)* puede comprender información de la posición en lugar de la información DOA, por ejemplo, las posiciones de las fuentes de sonido en una habitación de tres dimensiones. Por esto, se pueden definir filtros espaciales que no sólo filtran las direcciones específicas deseadas, sino también las zonas en un espacio tridimensional de la escena de grabación.

[0050] Todas las explicaciones dadas con respecto a las DOA son igualmente aplicables cuando una información de posición de una fuente de sonido está disponible. Por ejemplo, la información de posición se puede representar mediante una DOA (un ángulo) y una distancia. Cuando se emplea dicha representación de la posición, la DOA se puede obtener inmediatamente de la información de posición. O bien, la información de la posición se puede, por ejemplo, describir mediante las coordenadas *x, y, z*. A continuación, la DOA se puede calcular fácilmente a partir de la información de posición de la fuente de sonido y a partir de una posición del micrófono que registra la señal de micrófono de entrada respectiva.

[0051] En los párrafos siguientes se describen otras realizaciones.

[0052] Algunas realizaciones permiten la grabación de sonido espacialmente selectivo con reducción del ruido y supresión del efecto de reverberación. En este contexto se proporcionan realizaciones de la aplicación de filtrado espacial para la mejora de la señal en términos de extracción de la fuente, supresión del efecto de reverberación y reducción del ruido. El objetivo de tales realizaciones consiste en calcular una señal *Y(k, n)* que se corresponde con la salida de un micrófono direccional con un patrón de captación arbitrario. Esto significa que el sonido direccional (por ejemplo, una única onda plana) se atenúa o se conserva como se desee en función de su DOA, mientras que el sonido difuso o el propio ruido de un micrófono se suprime. De acuerdo con realizaciones, el filtro espacial proporcionado combina las ventajas de los filtros espaciales del estado de la técnica, entre otras cosas, proporcionando un alto índice de direccionalidad (DI) en situaciones con alta DNR y una ganancia máxima de ruido blanco (WNG) en caso contrario. De acuerdo con algunas realizaciones, el filtro espacial sólo puede ser linealmente acotado, lo que permite un cálculo rápido de los pesos. Por ejemplo, la función de transferencia *G(k, n)* de la fig. 2 y fig. 3 puede, por ejemplo, representar un patrón de captación deseado del micrófono direccional.

[0053] En los párrafos siguientes se proporciona una formulación del problema. A continuación se proporcionan realizaciones del módulo de cálculo de los pesos 103 y el módulo de cálculo de la IPI 102 para la grabación de sonido espacialmente selectivo con reducción del ruido y supresión del efecto de reverberación. Por otra parte, se describen realizaciones de un módulo de selección de la función de transferencia correspondiente 104.

[0054] En primer lugar se proporciona la formulación del problema. Se tiene en cuenta una matriz de micrófonos omnidireccionales *M* situados en $d_{1..M}$. Para cada *(k, n)* se supone que un campo sonoro se compone de *L < M* ondas planas (sonido direccional) que se propagan en un campo sonoro difuso isotrópico y espacialmente homogéneo. Las señales del micrófono ***x*** *(k, n)* pueden escribirse como

$$\mathbf{x}(k,n) = \sum_{l=1}^{L} \mathbf{x}_l(k,n) + \mathbf{x}_\mathrm{d}(k,n) + \mathbf{x}_\mathrm{n}(k,n). \qquad (2)$$

donde $x_l\ (k, n) = [X_l, (k, n, d_1) .. X_l\ (k, n, d_M)]^T$ comprende las señales de micrófono que son proporcionales a la presión sonora de la onda plana *l*, $x_d\ (k, n)$ es el ruido medido no estacionario (por ejemplo, sonido difuso) y $x_n\ (k, n)$ es el ruido estacionario / ruido que varía lentamente (por ejemplo, el propio ruido de un micrófono).

[0055] Suponiendo que los tres componentes en la fórmula (2) no presentan correlación mutua, la matriz de densidad espectral de potencia (PSD) de las señales del micrófono se pueden describir por

$$\begin{aligned}\mathbf{\Phi}(k,n) &= \mathrm{E}\left\{\mathbf{x}(k,n)\,\mathbf{x}^{\mathrm{H}}(k,n)\right\} \\ &= \sum_{l=1}^{L}\mathbf{\Phi}_l(k,n) + \mathbf{\Phi}_{\mathrm{d}}(k,n) + \mathbf{\Phi}_{\mathrm{n}}(k,n),\end{aligned} \tag{3}$$

con

$$\mathbf{\Phi}_{\mathrm{d}}(k,n) = \phi_{\mathrm{d}}(k,n)\;\mathbf{\Gamma}_{\mathrm{d}}(k). \tag{4}$$

[0056] Aquí $\phi_n(k, n)$ es la matriz de PSD del ruido estacionario / ruido que varía lentamente y $\phi_d(k, n)$ es la potencia esperada del ruido no estacionario, que puede variar rápidamente en el tiempo y la frecuencia. El elemento *ij* de la matriz de coherencia $T_d(k)$, representada por $\gamma_{ij}(k)$, es la coherencia entre el micrófono i y j resultante del ruido no estacionario. Por ejemplo, para un campo difuso isotrópico esféricamente $\gamma_{ij}(k) = \mathrm{sinc}(\chi, r_{ij})$ [20] con número de onda χ y $r_{ij} = ||d_j - d_i||$. El elemento $_{ij}$ de la matriz de coherencia $T_n(k)$es la coherencia entre el micrófono *i* y *j* resultante del ruido estacionario / ruido que varía lentamente. Para el propio ruido del micrófono $\phi_n(k, n) = \phi_n(k, n)$ **I**, donde **I** es una matriz de identidad y $\phi_n(k, n)$ es la potencia esperada del propio ruido.

[0057] El sonido direccional $x_l(k, n)$ de (2) se puede escribir como

$$\mathbf{x}_l(k,n) = \mathbf{a}\left[k \,|\, \varphi_l(k,n)\right] X_l(k,n,\mathbf{d}_1), \tag{5}$$

donde $\varphi(k, n)$ es el azimut de la DOA de la onda plana i ($\varphi = 0$ indica el costado de la matriz) y $a[k|\varphi_l(k, n)] = [a_1[k|\varphi_l(k, n)] \ldots a_M[k|\varphi_l(k, n)]]^T$ es el vector de propagación. El elemento *i* de $a[k|\varphi_l(k, n)]$,

$$a_i\left[k \,|\, \varphi_l(k,n)\right] = \exp\left\{j\,\kappa\, r_i \sin\varphi_l(k,n)\right\}. \tag{6}$$

describe el desplazamiento de fase de la onda plana *l* del primero al micrófono *i*. Cabe señalar que $r_i = ||di - dl||$. es igual a la distancia entre el primer micrófono y el micrófono *i*.

[0058] El ángulo $<a_i[k \mid \varphi(k, n)] = \mu_i[k \mid \varphi(k, n)]$ a menudo se conoce como frecuencia espacial. La DOA de la onda *l* se puede representar por $\varphi_l(k, n)$, $a_i[k \mid \varphi_l(k, n)]$, $a[k \mid \varphi(k, n)]$ o por $\mu_i[k \mid \varphi(k, n)]$

[0059] Como se ha explicado anteriormente, el objetivo de la realización es filtrar las señales de micrófono x(k, n) de manera que los sonidos direccionales que llegan de zonas espaciales específicas se atenúen o amplifiquen como se desee, mientras que el ruido estacionario y no estacionario es suprimido. Por tanto, la señal deseada puede expresarse como

$$Y(k,n) = \sum_{l=1}^{L} G\left[k \,|\, \varphi_l(k,n)\right] X_l(k,n,\mathbf{d}_1), \tag{7}$$

en la que $G[k \mid \varphi(k, n)]$ es una función de direccionalidad arbitraria, por ejemplo predefinida, de valores reales o complejos que puede ser dependiente de la frecuencia.

[0060] La fig. 4 se refiere a un escenario con dos funciones de direccionalidad y posiciones de la fuente arbitrarias de acuerdo con una realización. En particular, la fig. 4 muestra la magnitud de dos direccionalidades de ejemplo $G_1[k \mid \varphi(k, n)]$ y $G_2[k \mid \varphi(k, n)]$. Al utilizar $G_1[k \mid \varphi(k, n)]$ (véase la línea continua en la fig. 4), el sonido direccional que llega desde φ <45 ° se atenúa en 21 dB, mientras que el sonido direccional procedente de otras direcciones no se atenúa. En principio, se pueden diseñar direccionalidades arbitrarias, incluso funciones tales como $G_2[k \mid \varphi(k, n)]$. (véase la línea discontinua en la fig. 4). Por otra parte, $G[k \mid \varphi(k, n)]$. se puede diseñar variante en el tiempo, por ejemplo, para extraer fuentes de sonido en movimiento o emergentes una vez que han sido localizadas.

[0061] Una estimación de la señal Y(k, n) se obtiene mediante una combinación lineal de las señales de micrófono *x(k, n*), por ejemplo, por

$$\hat{Y}(k, n) = \mathbf{w}^{\mathrm{H}}(k, n)\, \mathbf{x}(k, n), \qquad (8)$$

donde *w(k, n)* es un vector de ponderación complejo de longitud *M*. El correspondiente vector de ponderación óptimo w(k, n) se obtiene en los párrafos siguientes. En los párrafos siguientes, la dependencia de los pesos *w(k, n)* en *k* y *n* se omite por razones de brevedad.

[0062] En este punto se describen dos realizaciones del módulo de cálculo de pesos 103 en las fig. 2 y fig. 3.

[0063] A partir de 5 y 7 se deduce que *w(k, n)* debe satisfacer las restricciones lineales

$$\mathbf{w}^{\mathrm{H}}(k, n)\, \mathbf{a}[k \mid \varphi_l(k, n)] = G[k \mid \varphi_l(k, n)], \quad l \in \{1, 2, \ldots, L\}. \qquad (9)$$

[0064] Por otra parte, se debe minimizar la potencia del ruido no estacionario y estacionario/que varía lentamente en la salida del filtro.

[0065] La fig. 5 representa una realización de la invención para la aplicación de filtrado espacial. En particular, la figura 5 ilustra un generador de pesos 110 de acuerdo con otra realización. Otra vez, el generador de pesos 110 comprende un módulo de cálculo de la información 102, un módulo de cálculo de los pesos 103 y un módulo de selección de la función de transferencia 104.

[0066] Más particularmente, la fig. 5 ilustra un estrategia de varianza mínima restringida linealmente (LCMV). En esta realización (véase la fig. 5), los pesos *w(k, n)* se calculan a partir de la IPI *I(k, n)* que comprende la DOA de *L* ondas planas y estadísticas de ruido estacionario y no estacionario. Posteriormente, la información puede comprender la DNR, las potencias separadas $\phi_n(k, n)$ y $\phi_d(k, n)$ de los dos componentes de ruido o las matrices de PSD ϕ_n y ϕ_d de los dos componentes de ruido.

[0067] Por ejemplo, ϕ_d se puede considerar como una primera información de ruido en un primer componente de ruido de los dos componentes de ruido y ϕ_n se puede considerar como una segunda información de ruido en un segundo componente de ruido de los dos componentes de ruido.

[0068] Por ejemplo, el generador de pesos 110 se puede configurar para determinar la primera información de ruido ϕ_d en función de una o más coherencias entre al menos algunos de los primeros componentes de ruido de la o las señales de entrada de micrófono. Por ejemplo, el generador de pesos 110 se puede configurar para determinar la primera información de ruido en función de una matriz de coherencia $T_d(k)$, que indica las coherencias resultantes procedentes de los primeros componentes de ruido de la o las señales de entrada de micrófono, por ejemplo, aplicando la fórmula

$$\mathbf{\Phi}_{\mathrm{d}}(k, n) = \phi_{\mathrm{d}}(k, n)\ \mathbf{\Gamma}_{\mathrm{d}}(k).$$

[0069] Los pesos *w(k, n)* para resolver el problema en 8 se encuentran minimizando la suma de la potencia del propio ruido (ruido estacionario / ruido que varía lentamente) y la potencia de sonido difuso (ruido no estacionario) en la salida del filtro, es decir,

$$\mathbf{w}_{\mathrm{nd}} = \arg\min_{\mathbf{w}} \; \mathbf{w}^{\mathrm{H}} \underbrace{\left[\mathbf{\Phi}_{\mathrm{d}}(k,n) + \mathbf{\Phi}_{\mathrm{n}}(k,n)\right]}_{\mathbf{\Phi}_{\mathrm{u}}(k,n)} \mathbf{w} \tag{10}$$

$$= \arg\min_{\mathbf{w}} \; \mathbf{w}^{\mathrm{H}} \underbrace{\left[\phi_{\mathrm{d}}(k,n)\,\mathbf{\Gamma}_{\mathrm{d}}(k) + \mathbf{\Phi}_{\mathrm{n}}(k,n)\right]}_{\mathbf{\Phi}_{\mathrm{u}}(k,n)} \mathbf{w} \tag{11}$$

[0070] Usando 4 y suponiendo que $\phi_n(k, n) = \phi_n(k, n)\, I$, el problema de optimización se puede expresar como

$$\mathbf{w}_{\mathrm{nd}} = \arg\min_{\mathbf{w}} \; \mathbf{w}^{\mathrm{H}} \underbrace{\left[\Psi(k,n)\,\mathbf{\Gamma}_{\mathrm{d}}(k) + \mathbf{I}\right]}_{\mathbf{C}(k,n)} \mathbf{w} \tag{12}$$

donde

$$\Psi(k,n) = \frac{\phi_{\mathrm{d}}(k,n)}{\phi_{\mathrm{n}}(k,n)} \tag{13}$$

es la DNR de entrada variable con el tiempo en los micrófonos. La solución a 10 y 12 dadas las restricciones 9 es [21]

$$\mathbf{w}_{\mathrm{nd}} = \mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}\left[\mathbf{A}^{\mathrm{H}}\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}\right]^{-1}\mathbf{g} \tag{14}$$

$$= \mathbf{C}^{-1}\mathbf{A}\left[\mathbf{A}^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{A}\right]^{-1}\mathbf{g}. \tag{15}$$

donde **A***(k, n)*=[a[k | φ*(k, n)*]... a[k | φ$_l$*(k, n)*]] comprende información de la DOA de las *L* ondas planas en términos de vectores de propagación. La ganancia deseada correspondiente viene dada por

$$\mathbf{g}(k,n) = \left[G\left[k \mid \varphi_1(k,n)\right] \; \ldots \; G\left[k \mid \varphi_l(k,n)\right]\right]^{\mathrm{T}}. \tag{16}$$

[0071] Las realizaciones de la estimación de ψ(k, n) y la otra IPI requerida se describen a continuación.

[0072] Otras realizaciones se basan en un filtro de Wiener paramétrico multicanal. En tales realizaciones, como se ilustra en la fig. 6, la IPI comprende además información sobre las estadísticas de la señal, por ejemplo, la matriz de PSD de la señal ϕ_s(k, n) que comprende las potencias de las L ondas planas (sonido direccional). Además, se tienen en cuenta los parámetros de control opcional λ1…L *(k, n)* para controlar la cantidad de distorsión de señal

para cada una de las ondas planas L.

[0073] La fig. 6 ilustra una realización para la aplicación de filtrado espacial implementando un generador de pesos 110 que emplea un filtro de Wiener paramétrico multicanal. Otra vez, el generador de pesos 110 comprende un módulo de cálculo de la información 102, un módulo de cálculo de los pesos 103 y un módulo de selección de la función de transferencia 104.

[0074] Los pesos *w(k, n)* se calculan a través de una estrategia de filtro de Wiener multicanal. El filtro de Wiener minimiza la potencia de la señal residual en la salida, es decir,

$$\mathbf{w}_{\mathrm{nd}} = \arg\min_{\mathbf{w}} \underbrace{\mathrm{E}\left\{|\widehat{Y}(k,n) - Y(k,n)|^2\right\}}_{C(k,n)} \tag{17}$$

[0075] La función de coste *C(k, n)* a minimizar se puede escribir como

$$C(k,n) = \mathrm{E}\left\{|\widehat{Y}(k,n) - Y(k,n)|^2\right\} \tag{18}$$

$$= \left[\mathbf{g} - \mathbf{A}^{\mathrm{H}}(k,n)\,\mathbf{w}\right]^{\mathrm{H}} \mathbf{\Phi}_{\mathrm{s}}(k,n) \left[\mathbf{g} - \mathbf{A}^{\mathrm{H}}(k,n)\,\mathbf{w}\right] + \mathbf{w}^{\mathrm{H}}\,\mathbf{\Phi}_{\mathrm{u}}(k,n)\,\mathbf{w}, \tag{19}$$

donde $\phi_S(k, n) = \boldsymbol{E}\{\boldsymbol{x}_S(k, n)\, \boldsymbol{x}_S(k, n)^H\}$ comprende los sonido direccionales PSD y $\boldsymbol{x}_S(k, n = [X_I(k, n, d_1)\ X_I(k, n, d_1) \ldots X_L(k, n, d_1]$ comprende las señales proporcionales a las presiones sonoras de las *L* ondas planas en el micrófono de referencia. Tenga en cuenta que $\phi_S(k, n)$ es una matriz diagonal, donde los elementos diagonales diag{ $\phi_S(k, n)$}= $[\phi_1(k, n) \ldots \phi_L(k, n)]$T son las potencias de las ondas planas que llegan. Con el fin de tener el control sobre las distorsiones de señal introducidas, se puede incluir una matriz diagonal A(k, n) que comprende parámetros de control dependientes del tiempo y la frecuencia diag{Δ} = $[\lambda_1(k, n)\ \lambda_2(k, n) \ldots \lambda_L(k, n)]^T$, es decir,

$$C_{\mathrm{PW}}(k,n) = \left[\mathbf{g} - \mathbf{A}^{\mathrm{H}}(k,n)\,\mathbf{w}\right]^{\mathrm{H}} \mathbf{\Lambda}(k,n)\,\mathbf{\Phi}_{\mathrm{s}}(k,n) \left[\mathbf{g} - \mathbf{A}^{\mathrm{H}}(k,n)\,\mathbf{w}\right] + \mathbf{w}^{\mathrm{H}}\,\mathbf{\Phi}_{\mathrm{u}}(k,n)\,\mathbf{w}. \tag{20}$$

[0076] La solución al problema de minimización de 17 dado $C_{PW}(k, n)$ es

$$\mathbf{w} = \left[\mathbf{A}^{\mathrm{H}}\mathbf{\Lambda}(k,n)\mathbf{\Phi}_{\mathrm{s}}(k,n)\mathbf{A} + \mathbf{\Phi}_{\mathrm{u}}\right]^{-1}\mathbf{A}\mathbf{\Lambda}(k,n)\mathbf{\Phi}_{\mathrm{s}}(k,n)\mathbf{g}. \tag{21}$$

[0077] Esto es idéntico a

$$\mathbf{w} = \mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}\left[\mathbf{\Lambda}^{-1}\mathbf{\Phi}_{\mathrm{s}}^{-1} + \mathbf{A}^{\mathrm{H}}\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}\right]^{-1}\mathbf{g} \tag{22}$$

[0078] Cabe señalar que para $\Delta^{-1} = 0$, se obtiene la solución de LCMV en (14). Para $\Delta^{-1} = I$, se obtiene el filtro de Wiener multicanal. Para otros valores $\lambda_{1\ldots L}$ *(k, n),* se puede controlar la cantidad de distorsión de la señal de la fuente correspondiente y la cantidad de supresión de ruido residual, respectivamente. Por lo tanto, normalmente uno define $\mathbf{\Lambda}$ en función de la información paramétrica disponible, es decir,

$$\lambda_l^{-1}(k,n) = f\left(\mathcal{J}(k,n)\right), \qquad (23)$$

donde f (.) es una función arbitraria definida por el usuario. Por ejemplo, uno puede elegir $\lambda_{1...L}$ *(k, n)* de acuerdo con

$$\lambda_l^{-1}(k,n) = \frac{1}{1+\frac{\phi_l(k,n)}{\phi_u(k,n)}}, \qquad (24)$$

donde φ_l*(k, n)* es la potencia de la señal l (onda plana l) y φ_n*(k, n)* = φ_n*(k, n)* + φ_d*(k, n)* es la potencia de la señal no deseada (ruido estacionario / ruido que varía lentamente más ruido no estacionario). Por esto, el filtro de Wiener paramétrico depende de la información estadística de un componente de la señal de las dos o más señales de entrada de micrófono y, por lo tanto, el filtro de Wiener paramétrico depende de la información estadística de un componente de ruido de las dos o más señales de entrada de micrófono.

[0079] Si la fuente *l* es fuerte en comparación con el ruido, se obtiene una $\mathcal{J}$ cercano a cero lo que significa que se obtiene la solución LCMV (sin distorsión de la señal de la fuente). Si el ruido es fuerte en comparación con la potencia de la fuente, se obtiene una $\mathcal{J}$ cercana a uno lo que significa que se obtiene el filtro de Wiener multicanal (fuerte supresión del ruido).

[0080] A continuación se describe la estimación de φ_s*(k, n)* y φ_u*(k, n)*

[0081] En los párrafos siguientes, se describen realizaciones del módulo de estimación de parámetros instantáneos 102.

[0082] Se necesitan estimar diferentes IPI antes de que se puedan calcular los pesos. Las DOA de las ondas planas *L* calculadas en el módulo 201 se pueden obtener con conocidos estimadores de DOA de banda estrecha tales como ESPRIT [22] o root MUSIC [23] u otros estimadores del estado de la técnica. Estos algoritmos pueden proporcionar, por ejemplo, el ángulo de azimut φ*(k, n)*, la frecuencia espacial μ [k | φ (*k, n*)], el desplazamiento de fase a[k | φ*(k, n)*] o el vector de propagación **a**[k | φ*(k, n)*] de una o más ondas que llegan a la matriz. La estimación de DOA no se analiza más ya que la propia estimación de DOA es bien conocida en la técnica.

[0083] En los párrafos siguientes se describe la estimación de la relación difusión-ruido (DNR). En particular, la estimación de la DNR de entrada Y(k, n), es decir, se analiza una realización del módulo 202 en la fig. 5. La estimación de la DNR aprovecha la información de DOA obtenida en el módulo 201. Para estimar Ψ*(k, n)* se puede utilizar un filtro espacial adicional que anula las ondas planas L de manera que sólo se capta el sonido difuso. Los pesos de este filtro espacial se encuentran, por ejemplo, maximizando la WNG de la matriz, es decir,

$$\mathbf{w}_\Psi = \arg\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}}\mathbf{w} \qquad (25)$$

sujetos a

$$\mathbf{w}^{\mathrm{H}}\,\mathbf{a}[k \mid \varphi_l(k,n)] = 0, \quad l \in \{1,2,\ldots,L\}, \qquad (26)$$

$$\mathbf{w}^{\mathrm{H}}\,\mathbf{a}[k \mid \varphi_0(k,n)] = 1. \qquad (27)$$

[0084] La restricción 27 asegura que los pesos $\mathbf{w}_\psi$ no son cero. El vector de propagación **a**[k | $\varphi(k, n)$] se corresponde con una dirección específica $\varphi_0(k, n)$ que es diferente de las DOAs $\varphi_l(k, n)$ de las *L* ondas planas. En los párrafos siguientes, para φ_o(k, n) se elige la dirección que tiene la distancia más grande con todos los $\varphi_l(k, n)$ es decir,

$$\varphi_0(k,n) = \arg\max_{\varphi}\left(\min_{l}|\varphi - \varphi_l(k,n)|\right). \tag{28}$$

donde dados los pesos $\mathbf{w}_\psi$, la potencia de salida del filtro espacial adicional viene dada por

$$\mathbf{w}_\Psi^{\mathrm{H}}\,\mathbf{\Phi}(k,n)\,\mathbf{w}_\Psi = \phi_{\mathrm{d}}(k,n)\,\mathbf{w}_\Psi^{\mathrm{H}}\,\mathbf{\Gamma}_{\mathrm{d}}(k)\,\mathbf{w}_\Psi + \phi_{\mathrm{n}}(k,n)\,\mathbf{w}_\Psi^{\mathrm{H}}\,\mathbf{w}_\Psi. \tag{29}$$

[0085] La DNR de entrada ahora se puede calcular con 13 y 29, es decir,

$$\Psi(k,n) = \frac{\mathbf{w}_\Psi^{\mathrm{H}}\,\mathbf{\Phi}(k,n)\,\mathbf{w}_\Psi - \phi_{\mathrm{n}}(k,n)\,\mathbf{w}_\Psi^{\mathrm{H}}\,\mathbf{w}_\Psi}{\phi_{\mathrm{n}}(k,n)\,\mathbf{w}_\Psi^{\mathrm{H}}\,\mathbf{\Gamma}_{\mathrm{d}}(k)\,\mathbf{w}_\Psi}. \tag{30}$$

[0086] La potencia esperada necesaria del propio ruido del micrófono $\phi_n(k, n)$ se puede, por ejemplo, estimar en los periodos de silencio suponiendo que la potencia es constante o que varía lentamente con el tiempo. Tenga en cuenta que el estimador de la DNR propuesta no proporciona necesariamente la varianza de estimación más baja en la práctica debido a los criterios de optimización elegidos 45, pero proporciona resultados imparciales.

[0087] En los párrafos siguientes, se analiza la estimación de PSD no estacionario $\phi_u(k, n)$ es decir, otra realización del módulo 202 en la fig. 5. La potencia (PSD) del ruido no estacionario se puede estimar con

$$\phi_{\mathrm{d}}(k,n) = \frac{\mathbf{w}_\Psi^{\mathrm{H}}[\mathbf{\Phi}(k,n) - \mathbf{\Phi}_{\mathrm{n}}(k,n)]\mathbf{w}_\Psi}{\mathbf{w}_\Psi^{\mathrm{H}}\mathbf{\Gamma}_{\mathrm{d}}(k)\mathbf{w}_\Psi}, \tag{31}$$

donde w_ψ se define en el párrafo anterior. Cabe señalar que la matriz PSD del ruido estacionario/que varía lentamente $\phi_n(k, n)$ se puede estimar en los periodos de silencio (es decir, en la ausencia de la señal y el ruido no estacionario), es decir,

$$\mathbf{\Phi}_{\mathrm{n}}(k,n) = \mathrm{E}\left\{\mathbf{x}(k,n)\mathbf{x}^{\mathrm{H}}(k,n)\right\}. \tag{32}$$

donde la esperanza se aproxima haciendo el promedio de *n* tramas silenciosas. Las tramas silenciosas se pueden detectar con métodos del estado de la técnica.

[0088] En los párrafos siguientes se analiza la estimación de la matriz de PSD de señal no deseada (véase el módulo 203).

[0089] La matriz de PSD de la señal no deseada (ruido estacionario/que varía lentamente más ruido no estacionario) ϕ_u(k, n) se puede obtener con

$$\mathbf{\Phi}_{\mathbf{u}}(k,n) = \phi_{\mathrm{n}}(k,n)\left(\Psi(k,n)\,\mathbf{\Gamma}_{\mathrm{d}}(k) + \mathbf{\Gamma}_{\mathrm{n}}(k)\right), \tag{33}$$

o más en general con

$$\mathbf{\Phi}_{\mathbf{u}}(k,n) = \phi_{\mathbf{d}}(k,n)\mathbf{\Gamma}_{\mathbf{d}}(k) + \mathbf{\Phi}_{\mathbf{n}}(k,n). \qquad (34)$$

donde $\Gamma_d(k)$ y $\Gamma_n(k)$ están disponibles como información a priori (véase más arriba). La DNR Y(k, n), la potencia de ruido estacionario/que varía lentamente $\phi_n(k, n)$ y otras cantidades requeridas se pueden calcular como se ha explicado anteriormente. Por lo tanto, la estimación de os(k, n) aprovecha la información de la DOA obtenida por el módulo 201.

[0090] En los párrafos siguientes se analiza la estimación de la matriz de PSD de la señal (véase el módulo 204).

[0091] La potencia $\phi_{1...L}(k, n)$ de las ondas planas que llegan, necesaria para calcular $\phi_s(k, n)$, se puede calcular con

$$\begin{bmatrix} \phi_1(k,n) \\ \vdots \\ \phi_L(k,n) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1(k,n) \\ \vdots \\ \mathbf{w}_L(k,n) \end{bmatrix}^{\mathbf{H}} [\mathbf{\Phi}_x(k,n) - \mathbf{\Phi}_{\mathbf{u}}(k,n)] \begin{bmatrix} \mathbf{w}_1(k,n) \\ \vdots \\ \mathbf{w}_L(k,n) \end{bmatrix}. \qquad (35)$$

donde los pesos w_l suprimen todas las ondas planas que llegan menos la onda i, es decir,

$$\mathbf{w}_l(k,n)^{\mathbf{H}} \mathbf{a}_{l'}(k,n) = \begin{cases} 1 & \text{si } l = l' \\ 0 & \text{en caso contrario} \end{cases} \qquad (36)$$

[0092] Por ejemplo,

$$\mathbf{w}_l = \arg\min_{\mathbf{w}} \ \mathbf{w}^{\mathbf{H}}\mathbf{w} \qquad (37)$$

sujeto a 36. La estimación de $\phi_s(k, n)$ aprovecha la información de DOA obtenida en el módulo 201. La matriz PSD necesaria de las señales no deseadas $\phi_u(k, n)$ se puede calcular como se explica en el párrafo anterior.

[0093] En este punto se describe un módulo de selección de la función de transferencia 104 de acuerdo con una realización.

[0094] En esta aplicación, se puede encontrar la ganancia G[k I ϕ,(k, n)] para la onda plana 1 correspondiente en función de la información de DOA $\phi_l(k, n)$. La función de transferencia G[k I ϕ (k, n)] para los diferentes DOAs ϕ (k, n) están disponibles en el sistema, por ejemplo, como una información a priori definida por el usuario. La ganancia también se puede calcular a partir del análisis de una imagen, por ejemplo, utilizando las posiciones de rostros que se han detectado. En la fig. 4 se representan dos ejemplos. Estas funciones de transferencia se corresponden con los patrones de captación deseados del micrófono direccional. La función de transferencia G[kl ϕ (k, n)] se puede proporcionar por ejemplo, como una tabla de consulta, es decir, para un ϕ(k, n) estimado se selecciona la ganancia correspondiente G[kl ϕ (k, n)] de la tabla de consulta. Tenga en cuenta que la función de transferencia también se puede definir como una función de la frecuencia espacial µ[k I ϕ (k, n)] en lugar del azimut ϕ (k,n), es decir, G(k, µ) en lugar de G[kl ϕ (k, n)]. Las ganancias también se pueden calcular a partir de la información de la posición de la fuente en lugar de la información de la DOA.

[0095] En este punto se proporcionan los resultados experimentales. Los siguientes resultados de la simulación demuestran la aplicabilidad práctica de las realizaciones descritas anteriormente. El sistema propuesto se compara con los sistemas del estado de la técnica, lo que se explicará más adelante. A continuación se analiza la configuración experimental y se proporcionan los resultados.

[0096] Al principio se consideran los filtros espaciales existentes.

[0097] Mientras que la PSD $\phi_n(k, n)$ se puede estimar en los periodos de silencio, $\phi_d(k, n)$ comúnmente se supone desconocida y no observable. Por lo tanto, se consideran dos filtros espaciales existentes que se pueden calcular sin este conocimiento.

[0098] El primer filtro espacial es conocido como un conformador de haz de retardo y suma y reduce al mínimo la potencia del propio ruido en la salida del filtro [es decir, maximiza la WNG] [1]. El vector de ponderación óptimo que minimiza el error cuadrático medio (ECM) entre 7 y 8 sujeto a 9 se obtiene con

$$\mathbf{w_n} = \arg\min_{\mathbf{w}} \underbrace{\mathbf{w}^H \mathbf{\Phi_n}(k, n)\, \mathbf{w}}_{\mathbf{w}^H\mathbf{w}} \qquad (38)$$

[0099] Existe una solución de forma cerrada a 38 [1] que permite un cálculo rápido de w_n. Cabe señalar que este filtro no proporciona necesariamente el DI más grande.

[0100] El segundo filtro espacial es conocido como el robusto conformador de haz superdirectivo (SD) y reduce al mínimo la potencia del sonido difuso en la salida del filtro [es decir, maximiza el DI] con el límite inferior de la WNG [24]. El límite inferior de la WNG incrementa la robustez frente a errores del vector de propagación y limita la amplificación del propio ruido [24]. El vector de ponderación óptimo que minimiza el ECM entre 7 y 8 sujeto a 9 y cumple el límite inferior de la WNG se obtiene con

$$\mathbf{w_d} = \arg\min_{\mathbf{w}} \underbrace{\mathbf{w}^H \mathbf{\Phi_d}(k, n)\, \mathbf{w}}_{\mathbf{w}^H \mathbf{\Gamma_d}(k,n)\, \mathbf{w}} \qquad (39)$$

y sujeto a una restricción cuadrática $w^H w < \beta$. El parámetro β^{-1} define la WNG mínima y determina el DI alcanzable del filtro. En la práctica, a menudo es difícil encontrar un equilibrio óptimo entre una WNG suficiente en situaciones de baja SNR y un nivel suficiente de DI en situaciones de alta SNR. Por otra parte, la resolución de 39 conduce a un problema de optimización no convexa debido a la restricción cuadrática que requiere mucho tiempo de resolución. Esto es especialmente problemático ya que el vector de ponderación tiene que recalcularse para cada k y n debido a las restricciones variables con el tiempo 9.

[0101] En este punto se considera una configuración experimental. Suponiendo que L = 2 ondas planas en el modelo de 2 y una matriz uniforme lineal (ULA) con M = 4 micrófonos con una separación entre micrófonos de 3 cm, se simuló una habitación como caja de zapatos (7,0 x 5,4 x 2,4 m3, RT 60" caja de zapatos" 380 ms) utilizando el procedimiento de la imagen fuente [25, 26] con dos fuentes de habla en <PA = 86° y <Ps = 11°, respectivamente (distancia 1,75 m, véase fig. 4). Las señales consistían en silencios de 0,6 s seguidas de dos voces superpuestas. Se añadió ruido gaussiano blanco a las señales de micrófono lo que resultó en una relación señal-ruido de segmentos (SSNR) de 26 dB. El sonido se muestreó a 16 kHz y se transformó en el dominio de tiempo-frecuencia usando una STFT de 512 puntos con 50% de superposición.

[0102] Se tiene en cuenta la función de direccionalidad $G_1(\varphi)$ de la fig. 4, es decir, la fuente A se extrae sin distorsiones mientras que se atenúa la potencia de la fuente B en 21 dB. Se tienen en cuenta los dos filtros espaciales anteriores y el filtro espacial proporcionado. Para el robusto conformador de haz SD (39), la WNG mínima se establece en -12 dB. Para el filtro espacial proporcionado (12), la DNR Y(k, n) se calcula como se ha explicado anteriormente. La potencia del propio ruido $\phi_n(k, n)$ se calcula a partir de la parte de señal silenciosa en el comienzo. La esperanza de 3 se aproxima mediante un promedio temporal recursivo sobre τ = 50 ms.

[0103] En los párrafos siguientes se tienen en cuenta las restricciones direccionales que no varían con el tiempo.

[0104] Para esta simulación se supone un conocimiento previo acerca de las dos posiciones de la fuente φ_A y φ_B. En todas las etapas de procesamiento hemos utilizado utilizó $\varphi_1(k, n) = \varphi_A$ y $\varphi_2(k, n) = \varphi_B$. Por lo tanto las restricciones direccionales de 9 y 26 no varían con el tiempo.

[0105] La fig. 7 ilustra la DNR real y estimada $\Psi(k, n)$. Las dos zonas marcadas indican, respectivamente, una parte silenciosa y activa de la señal. En particular la fig. 7 representa la DNR real y estimada Ψ (k, n) como función del tiempo y la frecuencia. Obtenemos una DNR relativamente alta durante la actividad de voz debido al ambiente reverberante. La DNR estimada en la fig. 7(b) posee una resolución temporal limitada debido al proceso de promediado temporal incorporado. Sin embargo, las estimaciones de Ψ (k, n) son lo suficientemente precisas como lo demuestran los resultados siguientes.

[0106] La fig. 8(a) representa el DI promedio para w_n y w_d (ambos son independientes de la señal) y para el filtro espacial propuesto w_{nd} (que es dependiente de la señal). Para el filtro espacial propuesto mostramos el DI de una parte silenciosa de la señal y durante la actividad de voz [ambas partes de la señal están marcadas en la fig. 7(b)]. En los periodos de silencio, el filtro espacial propuesto (línea discontinua w_{nd}) proporciona el mismo bajo DI que w_n. Durante la actividad de voz (línea continua w_{nd}) el DI obtenido es tan alto como para el robusto conformador de haz SD (w_d). La fig. 8(b) muestra las WNGs correspondientes. En los periodos de silencio el filtro espacial propuesto (línea discontinua w_{nd}) alcanza una WNG alta, mientras que durante la actividad de la señal la WNG es relativamente baja.

[0107] La fig. 8: DI y WNG de los filtros espaciales comparados. Para $w_{d'}$ la WNG mínima se estableció en -12 dB para que el filtro espacial fuera robusto frente al propio ruido del micrófono.

[0108] En general, la fig. 8 muestra que el filtro espacial propuesto combina las ventajas de ambos filtros espaciales existentes: en los periodos de silencio, se proporciona una WNG máxima que conduce a una amplificación mínima del propio ruido propio, es decir, alta robustez.

[0109] Durante la actividad de la señal y alta reverberación, donde el propio ruido es generalmente enmascarado, se proporciona un DI alto (a costa de una WNG baja) que conduce a una reducción óptima del sonido difuso. En este caso incluso WNGs más bien pequeñas son tolerables.

[0110] Tenga en cuenta que para las frecuencias más altas (f > 5 kHz), todos los filtros espaciales trabajan casi de forma idéntica ya que la matriz de coherencia $r_d(k)$ en 39 y 12 es aproximadamente igual a una matriz de identidad.

[0111] En los párrafos siguientes se tienen en cuenta las restricciones direccionales instantáneas.

[0112] Para esta simulación se supone que no hay información a priori de φ_A y φ_B disponible. Las DOAs $\varphi_1(k, n)$ y $\varphi 2(k, n)$ se estiman con ESPRIT. Por lo tanto las restricciones 9 varían con el tiempo. Para el robusto conformador de haz SD (w_d) solo se emplea una sola restricción invariante con el tiempo 9 que se corresponde con una dirección de observación fija de $\varphi_A = 860$. Este conformador de haz sirve de referencia.

[0113] La fig. 9 representa la DOA estimada $\varphi_1(k, n)$ y las ganancias resultantes $G[k|\varphi 1(k, n)]$. En particular la fig. 9 ilustra la DOA estimada $\varphi_1(k, n)$ y la ganancia resultante $IG[k\ I\varphi_1(k, n)]I^2$. La onda plana que llega no se atenúa si la DOA está dentro de la ventana espacial de la fig. 4 (línea continua). De lo contrario la potencia de la onda es atenuada en 21 dB.

[0114] La tabla 1 ilustra el rendimiento de todos los filtros espaciales [* sin procesar]. Los valores entre paréntesis se refieren a las restricciones direccionales invariables con el tiempo, los valores que no están entre paréntesis se refieren a las restricciones direccionales instantáneas. Las señales fueron ponderadas en A antes de calcular la SIR, SRR y SSNR.

Tabla 1

	SIR	[dB]	SRR	[dB]	SSNR	[dB]	PESQ	
*	11	(11)	-7	(-7)	26	(26)	1,5	(1,5)
w_n	21	(32)	-2	(-3)	33	(31)	2,0	(1,7)
w_d	26	(35)	0	(-1)	22	(24)	2,1	(2,0)
w_{nd}	25	(35)	1	(-1)	28	(26)	2,1	(2,0)

[0115] En particular la tabla 1 resume el rendimiento global de los filtros espaciales en términos de relación señal-interferencia (SIR), la relación señal-reverberación (SRR) y la SSNR en la salida del filtro. En términos de SIR y SRR (separación de fuentes, supresión del efecto de reverberación), la estrategia propuesta (w_{nd}) y el robusto conformador de haz SD (w_d) proporcionan el rendimiento el más alto. No obstante, la SSNR del w_d propuesto es 6 dB más alto que el SSNR de wd' lo que representó un beneficio claramente audible. El mejor rendimiento en términos de SSNR se obtiene utilizando w_n. En términos de PESQ w_{nd} y w_d superan a w_n. El uso de las restricciones direccionales instantáneas en lugar de restricciones invariantes con el tiempo (valores entre paréntesis) redujo principalmente la SIR alcanzable, pero proporciona una adaptación rápida en el caso de posiciones de fuente variables. Cabe señalar que el tiempo de cálculo de todos los pesos complejos necesarios por marco de tiempo fue superior a 80 s con w_d (caja de herramientas CVX [27,28]) e inferior a 0,08 s con la estrategia propuesta (MATLAB R2012b, MacBook Pro 2008).

[0116] En los párrafos siguientes se describen realizaciones para la reproducción sonora espacial. El objetivo de las realizaciones es captar una escena sonora por ejemplo, con una matriz de micrófonos y reproducir el sonido espacial con un sistema de reproducción de sonido arbitrario (por ejemplo, configuración de altavoces 5.1, reproducción de auriculares) de manera que se vuelva a recrear la impresión espacial original. Suponemos que el sistema de reproducción de sonido comprende N canales, es decir, calculamos N señales de salida Y(k, n).

[0117] En primer lugar se proporciona la formulación del problema. Se tiene en cuenta el modelo de la señal (véase la fórmula 2 anterior) y se formula un problema similar. El ruido estacionario/que varía lentamente se corresponde con el propio ruido del micrófono no deseada, mientras que el ruido no estacionario se corresponde con el sonido difuso deseado. En esta aplicación se desea el sonido difuso, ya que es de gran importancia para reproducir la impresión espacial original de la escena de grabación.

[0118] En los párrafos siguientes, se consigue la reproducción del sonido direccional $X_l(k, n, dl)$ sin distorsiones de la DOA correspondiente $\varphi_l k, n)$. Por otra parte, se reproducirá el sonido difuso con la energía correcta de todas las direcciones, mientras que se suprime el propio ruido del micrófono. Por lo tanto, la señal deseada Y (k, n) en 7 se expresa en este punto como

$$Y_i(k,n) = \sum_{l=1}^{L} G_i[k \mid \varphi_l(k,n)]\, X_l(k,n,\mathbf{d}_1) + G_{\mathrm{d}}(k,n) X_{\mathrm{d},i}(k,n,\mathbf{d}), \tag{40}$$

donde $Y_l(k, n)$ es la señal del canal i del sistema de reproducción del sonido (i = {1, ... N}), $X_{d,i}(k, n, d)$ es el sonido difuso medido en un punto arbitrario (por ejemplo, en la primer micrófono d_1) que se va a reproducir desde el altavoz i, y G_d (k, n) es una función de ganancia para el sonido difuso que asegura una potencia correcta del sonido difuso durante la reproducción (por lo general $\mathcal{J}$ l) . Idealmente las señales $X_{d,\ i}$ (k, n) tienen la potencia del sonido difuso correcta y son mutuamente no correlacionadas en los canales i, es decir,

$$\mathrm{E}\left\{X_{\mathrm{d},i}(k,n) X_{\mathrm{d},j}^{*}(k,n)\right\} = \begin{cases} \phi_{\mathbf{d}}(k,n) & \text{si } i=j \\ 0 & \text{en caso contrario} \end{cases} \tag{41}$$

[0119] Las funciones de transferencia $G_l[kl\varphi_l(k, n)]$ de los componentes de sonido direccionales se corresponden con una función de ganancia altavoz dependiente de DOA. Un ejemplo para el caso de la

reproducción en altavoz estéreo se representa en la fig. 10. Si la onda *l* llega desde $\varphi_l(k, n) = 300$, G1 = 1 y G2 = O.

[0120] Esto significa que este sonido direccional se reproduce sólo desde el canal *i* = 1 del sistema de reproducción (canal izquierdo). Para $\varphi,(k, n) = 0^0$ tenemos $\mathcal{J}$, es decir, el sonido direccional se reproduce con la misma potencia desde ambos altavoces. Alternativamente, Gi[k I $\varphi_1(k, n)$] se puede corresponder con una HRTF si se desea una reproducción binaural.

[0121] Las señales Y,(k, n) son estimadas mediante una combinación lineal de las señales de micrófono a partir de los pesos complejos w(k, n) como se ha explicado anteriormente, es decir,

$$\widehat{Y}_i(k, n) = \mathbf{w}_i^{\mathrm{H}}(k, n)\, \mathbf{x}(k, n), \qquad (42)$$

sujetas a restricciones específicas. Las restricciones y el cálculo de los pesos $w_i(k, n)$ se explican en la subsección siguiente.

[0122] En los párrafos siguientes se tiene en cuenta el módulo de cálculo de pesos 103 de acuerdo con las realizaciones correspondientes. En este contexto se proporcionan dos realizaciones del módulo de cálculo de pesos 103 de la fig. 2. Se deduce de la fórmula 5 y la fórmula 40 que w,(k, n) debe cumplir las restricciones lineales

$$\mathbf{w}_i^{\mathrm{H}}(k, n)\, \mathbf{a}[k \mid \varphi_l(k, n)] = G_i[k \mid \varphi_l(k, n)], \quad l \in \{1, 2, \ldots, L\}, i \in \{1, 2 \ldots, N\}. \qquad (43)$$

[0123] Por otra parte la potencia del sonido difuso se debe mantener. Por lo tanto $w_l(k, n)$ puede cumplir la restricción cuadrática

$$\mathbf{w}_i^{\mathrm{H}}\, \mathbf{\Gamma}_{\mathrm{d}}(k, n) \mathbf{w}_i = |G_{\mathrm{d}}(k, n)|^2, \quad \forall i. \qquad (44)$$

[0124] Por otra parte se debe minimizar la potencia del propio ruido en la salida del filtro. Así, los pesos óptimos se pueden calcular como

$$\mathbf{w}_i = \arg\min_{\mathbf{w}} \ \mathbf{w}^{\mathrm{H}} \mathbf{w} \qquad (45)$$

sujetos a la fórmula 43 y la fórmula 44. Esto conduce a un problema de optimización convexa que se puede resolver, por ejemplo, con procedimientos numéricos conocidos [29].

[0125] Con respecto al módulo de estimación de parámetros instantáneos 102, de acuerdo con las realizaciones correspondientes, las DOAs $\varphi_l(k, n)$ de las L ondas planas se pueden obtener con estimadores de DOA de banda estrecha bien conocidos, tales como ESPRIT [22] o root MUSIC [23], u otros estimadores del estado de la técnica.

[0126] En este punto se tiene en cuenta el módulo de selección de la función de transferencia 104 de acuerdo con las realizaciones correspondientes. En esta aplicación, la ganancia $G_l[k \mid \varphi(k, n)]$ del canal i se encuentra para el sonido direccional correspondiente 1 en función de la información de la DOA $\varphi(k, n)$. La función de transferencia G;[k $\varphi(k, n)$] para las diferentes DOAs $\varphi(k, n)$ y canales i está disponible en el sistema, por ejemplo, como una información a priori definida por el usuario. Las ganancias también se puede calcular a partir del análisis de una imagen, por ejemplo, utilizando las posiciones de rostros que se han detectado.

[0127] Las funciones de transferencia $G_l[k \mid \varphi(k, n)]$ se proporcionan normalmente como una tabla de consulta, es decir, para un $\varphi(k, n)$ estimado seleccionamos las ganancias correspondientes $G_l[k\ \varphi(k, n)]$ de la tabla de consulta. Tenga en cuenta que la función de transferencia también se puede definir como una función de la frecuencia espacial $\mu[k \mid \varphi(k, n)]$ en lugar del azimut $\varphi(k, n)$, es decir, $G(k, \mu)$ en lugar de G;[k I $\varphi(k, n)$]. Tenga en cuenta, además, que la función de transferencia también se puede corresponder con una HRTF que permite una

reproducción de sonido binaural. En este caso $G_l[k\ \varphi\ (k, n)]$ suele ser compleja. Tenga en cuenta que las ganancias o las funciones de transferencia también se pueden calcular a partir de la información de la posición de la fuente en lugar de la información de la DOA.

[0128] En la fig. 10 se describe un ejemplo de reproducción con altavoz estéreo. En particular la fig. 10 ilustra las funciones de ganancia para la reproducción estéreo.

[0129] Aunque se han descrito algunos aspectos en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo se corresponde con una etapa del procedimiento o una característica de una etapa del procedimiento. De forma análoga los aspectos que se describen en el contexto de un etapa del procedimiento también representan una descripción de un bloque correspondiente o un elemento o característica del aparato correspondiente.

[0130] La señal descompuesta de la invención se puede almacenar en un medio de almacenamiento digital o se puede transmitir con un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable como Internet.

[0131] En función de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación puede realizarse utilizando un medio de almacenamiento digital, por ejemplo un disquete, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control legibles electrónicamente y almacenadas en el mismo, que coopera (o es capaz de cooperar) con un sistema informático programable de manera que se lleve a cabo el procedimiento respectivo.

[0132] De acuerdo con la invención algunas realizaciones comprenden un soporte de datos no transitorio que tiene señales de control legibles electrónicamente y que son capaces de cooperar con un sistema informático programable, de manera que se lleve a cabo uno de los procedimientos descritos en este documento.

[0133] En general 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.

[0134] Otras realizaciones comprenden el programa de ordenador para llevar a cabo uno de los procedimientos descritos en este documento almacenado en un soporte legible por máquina.

[0135] En otras palabras, una realización del procedimiento de la invención es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en este documento cuando el programa informático se ejecuta en un ordenador.

[0136] Una realización adicional de los procedimientos de la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los procedimientos descritos en este documento.

[0137] Una realización adicional del procedimiento de la invención 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 este documento. El flujo de datos o la secuencia de señales pueden, por ejemplo, estar configurados para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

[0138] Una realización adicional comprende además 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 este documento.

[0139] 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 este documento.

[0140] En algunas realizaciones se puede utilizar un dispositivo lógico programable (por ejemplo, un campo de matriz de puertas programables) para llevar a cabo algunas o todas las funcionalidades de los procedimientos descritos en el este documento. En algunas realizaciones un campo de matriz de puertas programables podrá

cooperar con un microprocesador para llevar a cabo uno de los procedimientos descritos en este documento. En general los procedimientos se llevan a cabo, preferentemente, por cualquier aparato de hardware.

[0141] Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en este documento serán evidentes para otros expertos en la técnica. Es la intención, por lo tanto, de estar limitado sólo por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a modo de descripción y las explicaciones de las realizaciones de este documento.

Referencias

[0142]

[1] J. Benesty, J. Chen, and Y. Huang, Microphone Array Signal Processing. Berlin, Germany: Springer-Verlag, 2008.

[2] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in Handbook on Array Processing and Sensor Networks, S. Haykin and K. Ray Liu, Eds. Wiley, 2008, ch. 9.

[3] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in Springer Handbook of Speech Processing, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2008, ch. 47.

[4] J. Benesty, J. Chen, and E. A. P. Habets, Speech Enhancement in the STFT Domain, ser. SpringerBriefs in Electrical and Computer Engineering. Springer-Verlag, 2011.

[5]I. Tashev, M. Seltzer, and A. Acero, "Microphone array for headset with spatial noise suppressor," in Proc. Ninth International Workshop on Acoustic, Echo and Noise Control (IWAENC), Eindhoven, The Netherlands, 2005.

[6] M. Kallinger, G. Del Galdo, F. Kuech, D. Mahne, and R. Schultz-Amling, "Spatial filtering using directional audio coding parameters," in Proc. IEEE IntI. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2009, pp. 217-220.

[7] M. Kallinger, G. D. Galdo, F. Kuech, and O. Thiergart, "Dereverberation in the spatial audio coding domain," in Audio Engineering Society Convention 130, London UK, May 2011.

[8] G. Del Galdo, O. Thiergart, T. Weller, and E. A. P. Habets, "Generating virtual microphone signals using geo metrical information gathered by distributed arrays," in Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA), Edinburgh, United Kingdom, May 2011.

[9] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of handsfree speaker input in cars," IEEE Trans. Veh. Technol., vol. 42, no. 4, pp. 514-518, Nov. 1993.

[10] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," IEEE Trans. Signal Process., vol. 47, no. 10, pp. 2677-2684, Oct. 1999.

[11] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Trans. Signal Process., vol. 49, no. 8, pp. 1614-1626, Aug. 2001.

[12] W. Herbordtand W. Kellermann, "Adaptive beamforming for audio signal acquisition," in Adaptive Signal Processing: Applications to real-world problems, ser. Signals and Communication Technology, J. Benesty and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2003, ch. 6, pp. 155-194.

[13] R. Talmon, I. Cohen, and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 7, pp. 1420-1434, Sep. 2009.

[14] A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 1, pp. 206-219, Jan. 2011.

[15] E. A. P. Habets and J. Benesty, "Joint dereverberation and noise reduction using a two-stage beamforming approach," in Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA), 2011, pp. 191-195.

[16] M. Taseska and E. A. P. Habets, "MMSE-based blind source extraction in diffuse noise fields using a complex coherence-based a priori SAP estimator," in Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC), Sep. 2012.

[17] G. Reuven, S. Gannot, and I. Cohen, "Dual source transfer-function generalized sidelobe canceller," IEEE Trans. Speech Audio Process., vol. 16, no. 4, pp. 711-727, May 2008.

[18] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 6, pp. 1071-1086, Aug. 2009.

[19] O. Thiergart and E. A. P. Habets, "Sound field model violations in parametric spatial sound processing," in Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC), Sep. 2012.

[20] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson Jr., "Measurement of correlation coefficients in reverberant sound fields," The Journal of the Acoustical Society of America, vol. 27, no. 6, pp. 1072-1077,1955.

[21] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," Proc. IEEE, vol. 60, no. 8, pp. 926-935, Aug. 1972.

[22] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 37, no. 7, pp. 984-995, July 1989.

[23] B. Rao and K. Hari, "Performance analysis of root-music"," in Signals, Systems and Computers, 1988. Twenty Second Asilomar Conference on, vol. 2, 1988, pp. 578-582.

[24] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process., vol. 35, no. 10, pp. 1365-1376, Oct. 1987.

[25] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am., vol. 65, no. 4, pp. 943-950, Apr. 1979.

[26] E. A. P. Habets. (2008, May) Room impulse response (RIR) generator. [Online]. Available: http://home.tiscali.nl/ehabets/rirgenerator.html; see also: http://web. archive.org/web/20120730003147/http://home.tiscali.nl/ehabets/rir_generator.html

[27] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http:llcvxr.com/cvx, September 2012.

[28] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in Recent Advances in Learning and Control, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95-110.

[29] H. L. Van Trees, Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing. John Wiley & Sons, April 2002, vol. 1.

REIVINDICACIONES

1. Un filtro (100) para generar una señal de salida de audio que comprende una pluralidad de muestras de la señal de salida de audio, a partir de dos o más señales de entrada de micrófono, en el que la señal de salida de audio y las dos o más señales de entrada de micrófono están representadas en dominio tiempo-frecuencia, en el que cada una de la pluralidad de muestras de la señal de salida de audio se asigna a un contenedor de tiempo-frecuencia ((k, n)) de una pluralidad de contenedores de tiempo-frecuencia ((k, n)) y en el que el filtro (100) comprende:

un generador de pesos (110) que se ha adaptado para recibir, para cada una de las pluralidades de contenedores de tiempo-frecuencia ((k, n)), información de la dirección de llegada de uno o más componentes de sonido de una o más fuentes de sonido, o información de la posición de una o más fuentes de sonido, y que se ha adaptado para generar información sobre ponderación para cada una de las pluralidades de contenedores de tiempo-frecuencia ((k, n)) en función de la información de la dirección de llegada del o los componentes de sonido de una o más fuentes de sonido de dicho contenedor de tiempo-frecuencia ((k, n)), o en función de la información de la posición de la o las fuentes de sonido de dicho contenedor de tiempo-frecuencia ((k, n)); en el que el generador de pesos (110) se ha adaptado para generar la información sobre ponderación de cada una de las pluralidades de contenedores de tiempo-frecuencia ((k, n)) en función de una primera información de ruido que indica información sobre una primera matriz de coherencia de los primeros componentes de ruido de las dos o más señales de entrada de micrófono y en función de una segunda información de ruido que indica información sobre una segunda matriz de coherencia de los segundos componentes de ruido de las dos o más señales de micrófono de entrada; y un generador de la señal de salida (120) que genera la señal de salida de audio generando para cada una de la pluralidades de contenedores de tiempo-frecuencia ((k, n)) una de las pluralidades de muestras de la señal de salida de audio, que se asigna a dicho contenedor de tiempo-frecuencia ((k, n)), en función de la información sobre ponderación de dicho contenedor de tiempo-frecuencia ((k, n)) y en función de una muestra de entrada de audio, asignándose a dicho contenedor de tiempo-frecuencia ((k, n)) de cada una de las dos o más señales de micrófono de entrada.

2. Un filtro (100) según la reivindicación 1, en el que el generador de pesos (110) se ha configurado para generar la primera información de ruido empleando información estadística, y en el que el generador de pesos (110) se ha configurado para generar la segunda información de ruido sin emplear información estadística, en el que la información estadística está predefinida.

3. Un filtro (100) según la reivindicación 1 o 2, en el que el generador de pesos (110) se ha adaptado para generar la información sobre ponderación para cada una de las pluralidades de contenedores de tiempo-frecuencia ((k, n)) en función de la fórmula:

$$\mathbf{w_{nd}} = \mathbf{\Phi_u^{-1}A}\left[\mathbf{A^H\Phi_u^{-1}A}\right]^{-1}\mathbf{g},$$

en el que

$$\mathbf{\Phi_u} = \mathbf{\Phi_d} + \mathbf{\Phi_n},$$

en el que Φ_d es una primera matriz de densidad espectral de potencia de los primeros componentes de ruido de las dos o más señales de entrada de micrófono,
en el que Φ_n es una segunda matriz de densidad espectral de potencia de los segundos componentes de ruido de las dos o más señales de entrada de micrófono,
en el que A indica la información de la dirección de llegada, en el que w_{nd} es un vector que indica la información sobre ponderación,

en el que

$$\mathbf{g}(k,n) = \left[G[k\,|\,\varphi_1(k,n)]\ \ldots\ G[k\,|\,\varphi_l(k,n)]\right]^{\mathrm{T}},$$

en el que $G[k|\varphi_1(k,n)]$ es una primera función de direccionalidad predefinida, de valores reales o complejos, en

función de la información de la dirección de llegada, y
en el que G[klφ_l(k,n)] es una función de direccionalidad predefinida adicional, de valores reales o complejos, en función de la información de la dirección de llegada.

4. Un filtro (100) de acuerdo con una de las reivindicaciones anteriores, en el que el generador de pesos (110) se ha configurado para determinar la primera información de ruido en función de una o más coherencias entre al menos algunos de los primeros componentes de ruido de las dos o más señales de entrada de micrófono, en el que una o más coherencias están predefinidas.

5. Un filtro (100) de acuerdo con una de las reivindicaciones anteriores, en el que el generador de pesos (110) se ha configurado para determinar la primera información de ruido en función de una matriz de coherencia $r_d(k)$ que indica las coherencias resultantes de los primeros componentes de ruido de las dos o más señales de entrada de micrófono, en el que la matriz de coherencia $r_d(k)$ está predefinida.

6. Un filtro (100) de acuerdo con la reivindicación 5, en el que el generador de pesos (110) se ha configurado para determinar la primera información de ruido de acuerdo con la fórmula:

$$\mathbf{\Phi_d}(k,n) = \phi_{\mathbf{d}}(k,n)\ \mathbf{\Gamma_d}(k),$$

en la que $r_d(k)$ es la matriz de coherencia, en el que la matriz de coherencia está predefinida,
en la que $\Phi f(k, n)$ es la primera información de ruido, y
en la que Φ_d(k, n) es una potencia esperada de los primeros componentes de ruido de las dos o más señales de entrada de micrófono.

7. Un filtro (100) de acuerdo con una de las reivindicaciones anteriores, en el que el generador de pesos (110) se ha configurado para determinar la primera información de ruido en función de la segunda información de ruido y en función de la información de la dirección de llegada.

8. Un filtro (100) de acuerdo con una de las reivindicaciones anteriores,
en el que el generador de pesos (110) se ha configurado para generar la información sobre ponderación como una primera información sobre ponderación w_Ψ y
en el que el generador de pesos (110) se ha configurado para generar la primera información sobre ponderación determinando una segunda información sobre ponderación,
en el que el generador de pesos (110) se ha configurado para generar la información sobre ponderación w_Ψ aplicando la fórmula

$$\mathbf{w}_\Psi = \arg\min_{\mathbf{w}}\ \mathbf{w}^{\mathrm{H}}\mathbf{w}$$

de manera que la fórmula

$$\mathbf{w}^{\mathrm{H}}\,\mathbf{a}[k \mid \varphi_l(k,n)] = 0,$$

se cumpla,

en el que φ_l(k,n) indica la información de la dirección de llegada,
en el que $a[k\varphi_l(k,n)]$ indica un vector de propagación y
en el que w indica la segunda información sobre ponderación.

9. Un filtro (100) de acuerdo con la reivindicación 8, en el que el generador de pesos (110) se ha configurado para generar una información difusión-ruido o una potencia de un componente difuso en función de la segunda información sobre ponderación y en función de las dos o más señales de entrada de micrófono para determinar la primera información sobre ponderación.

10. Un filtro (100) de acuerdo con una de las reivindicaciones 1 a 3, en el que el generador de pesos (110)

se ha configurado para determinar la información sobre ponderación mediante la aplicación de un filtro de Wiener paramétrico, en el que el filtro de Wiener paramétrico depende de la información estadística de un componente de la señal de las dos o más señales de entrada de micrófono, y en el que el filtro de Wiener paramétrico depende de la información estadística de un componente de ruido de las dos o más señales de entrada de micrófono.

11. Un filtro (100) de acuerdo con una de las reivindicaciones anteriores, en el que el generador de pesos (110) se ha configurado para determinar la información sobre ponderación en función de la información de la dirección de llegada que indica una dirección de llegada de una o más ondas planas.

12. Un filtro (100) de acuerdo con una de las reivindicaciones anteriores,
en el que el generador de pesos (110) comprende un módulo de selección de la función de transferencia (104) para proporcionar una función de transferencia predefinida y
en el que el generador de pesos (110) se ha configurado para generar la información sobre ponderación en función de la información de la dirección de llegada y en función de la función de transferencia predefinida.

13. Un filtro (100) de acuerdo con la reivindicación 12, en el que el módulo de selección de la función de transferencia (104) se ha configurado para proporcionar la función de transferencia predefinida de manera que la función de transferencia predefinida indica un patrón de captación arbitrario en función de la información de la dirección de llegada, de manera que la función de transferencia predefinida indica la ganancia de un altavoz en función de la información de la dirección de llegada, o de manera que la función de transferencia predefinida indica una función de transferencia relativa a la cabeza en función de la información de la dirección de llegada.

14. Un procedimiento para generar una señal de salida de audio que comprende una pluralidad de muestras de la señal de salida de audio, a partir de dos o más señales de entrada de micrófono, en el que la señal de salida de audio y las dos o más señales de entrada de micrófono están representadas en dominio tiempo-frecuencia, en el que cada una de las pluralidades de muestras de la señal de salida de audio se asigna a un contenedor de tiempo-frecuencia *((k, n))* de una pluralidad de contenedores de tiempo-frecuencia *((k, n))* y en el que el procedimiento comprende:

la recepción, para cada una de las pluralidades de contenedores de tiempo-frecuencia *((k, n)),* de la información de la dirección de llegada de uno o más componentes de sonido de una o más fuentes de sonido o información de la posición de una o más fuentes de sonido, la generación de información sobre ponderación para cada una de las pluralidades de contenedores de tiempo-frecuencia *((k, n))* en función de la información de la dirección de llegada del o los componentes de sonido de la o las fuentes de sonido de dicho contenedor de tiempo-frecuencia *((k, n)),* o en función de la información de la posición de la o las fuentes de sonido de dicho contenedor de tiempo-frecuencia *((k, n))*; en el que la generación de la información sobre ponderación para cada una de las pluralidades de contenedores de tiempo-frecuencia *((k, n))* se lleva a cabo en función de primera información de ruido que indica la información de una primera matriz de coherencia de los primeros componentes de ruido de las dos o más señales de entrada de micrófono y en función de una segunda información de ruido que indica la información de una segunda matriz de coherencia de los segundos componentes de ruido de las dos o más señales de entrada de micrófono; y
la generación de la señal de salida de audio generando para cada una de las pluralidades de contenedores de tiempo-frecuencia *((k, n))* una pluralidad de las muestras de la señal de salida de audio, que se asigna a dicho contenedor de tiempo-frecuencia *((k, n)),* en función de la información sobre ponderación de dicho contenedor de tiempo-frecuencia (k, n) y en función de una muestra de entrada de audio, que se asigna a dicho contenedor de tiempo-frecuencia *((k, n))*, de cada una de las dos o más señales de entrada de micrófono.

15. Un programa informático para implementar el procedimiento de la reivindicación 14 cuando se ejecuta en un ordenador o procesador de señal.

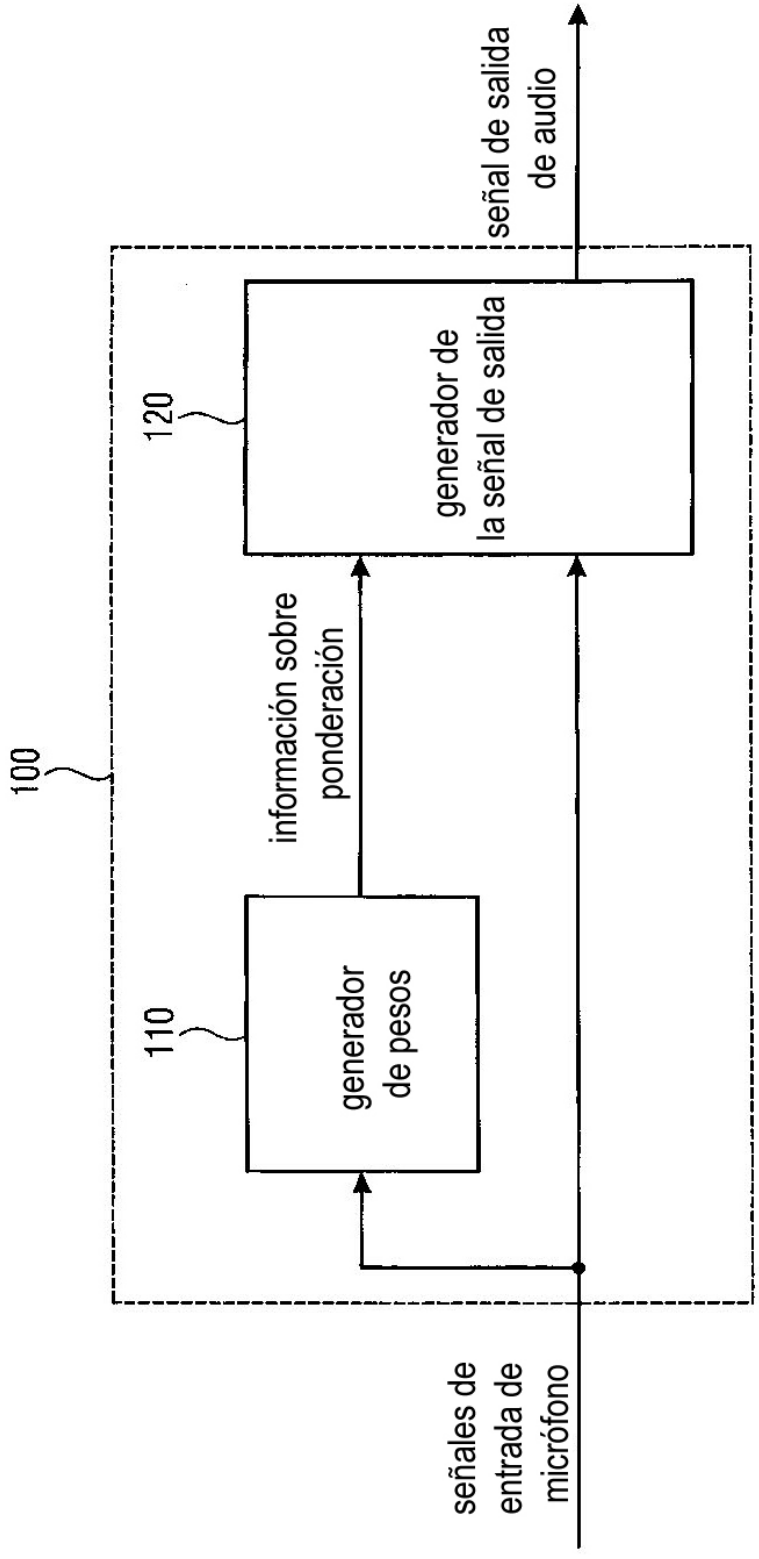
señal de salida de audio
120
generador de la señal de salida
información sobre ponderación
100
110
generador de pesos
señales de entrada de micrófono

FIG 1A

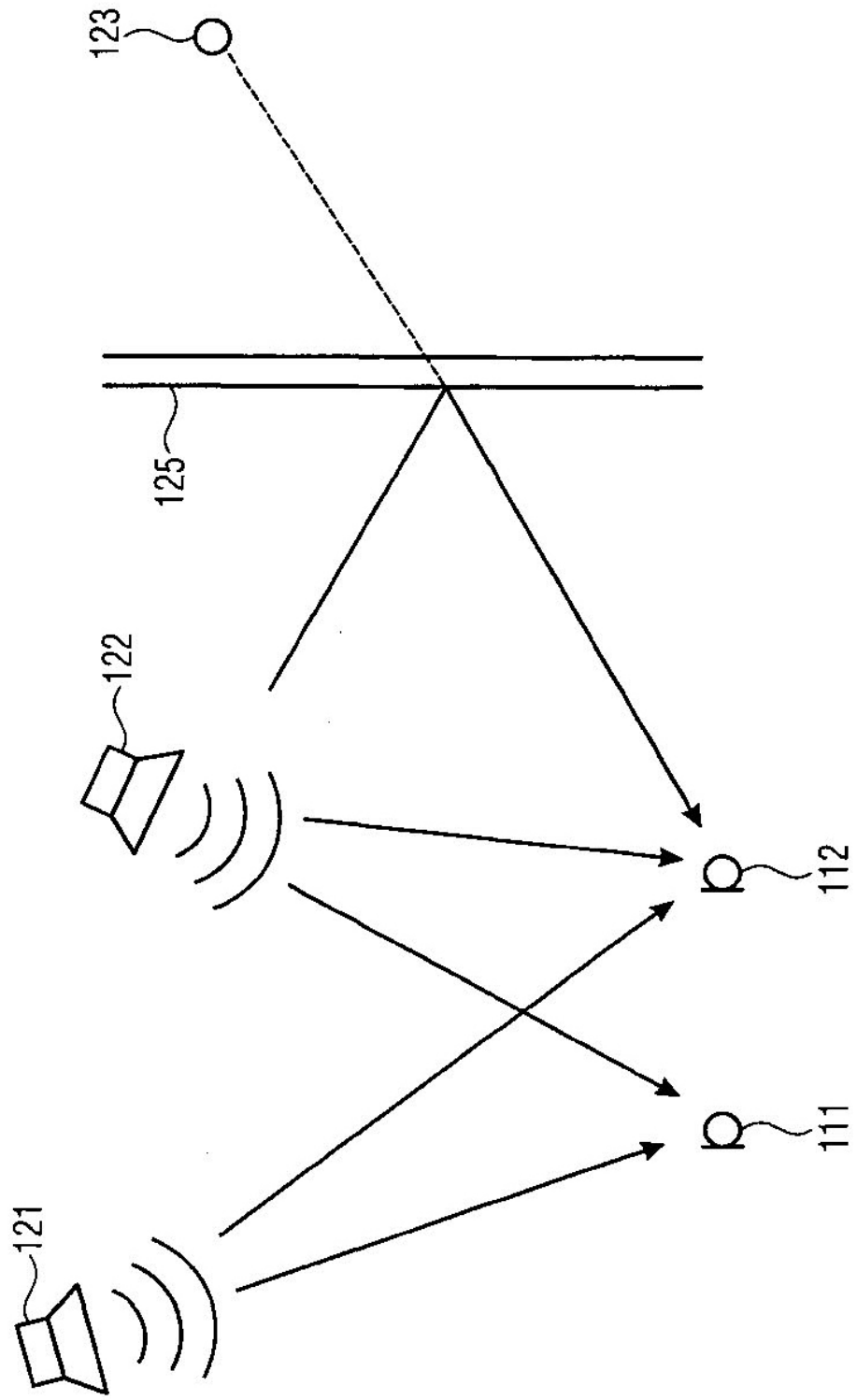

FIG 1B

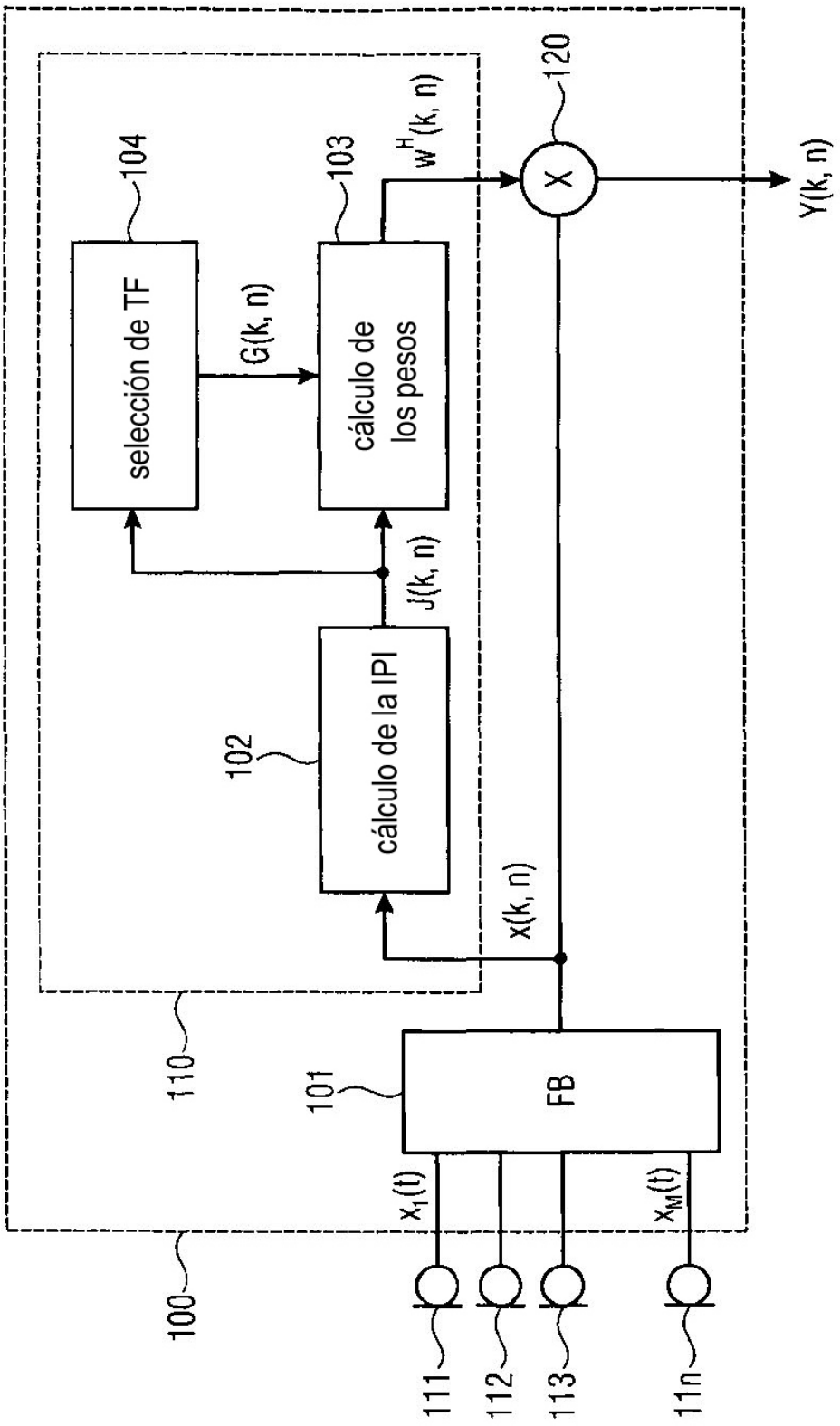
100
110
101
FB
102
cálculo de la IPI
103
cálculo de los pesos
104
selección de TF
120
X
$x_1(t)$
$x_M(t)$
111
112
113
11n
x(k, n)
J(k, n)
G(k, n)
$w^H(k, n)$
Y(k, n)

FIG 2

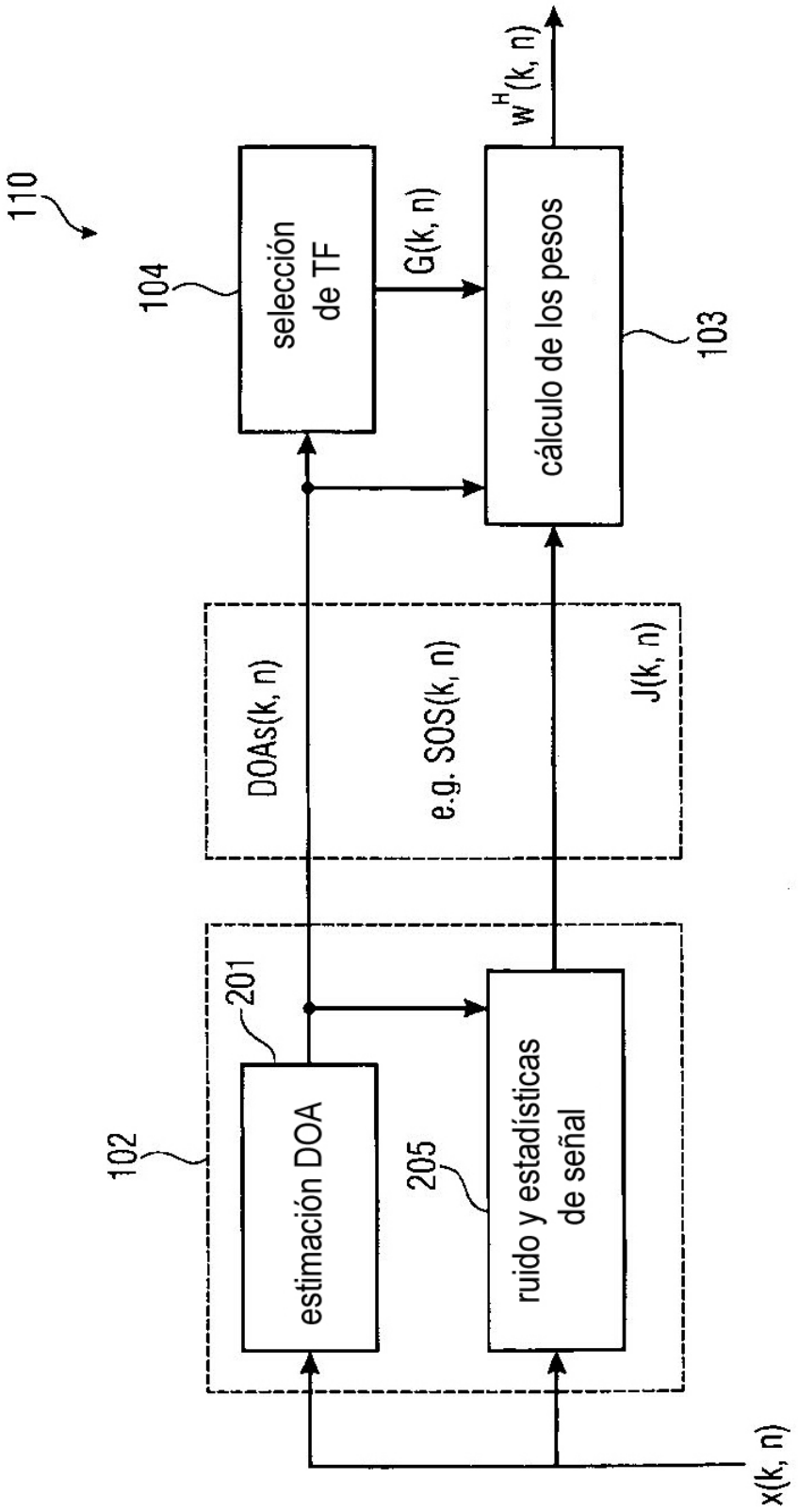
110
104
selección de TF
G(k, n)
cálculo de los pesos
103
$w^H(k, n)$
DOAs(k, n)
e.g. SOS(k, n)
J(k, n)
201
estimación DOA
205
ruido y estadísticas de señal
102
x(k, n)

FIG 3

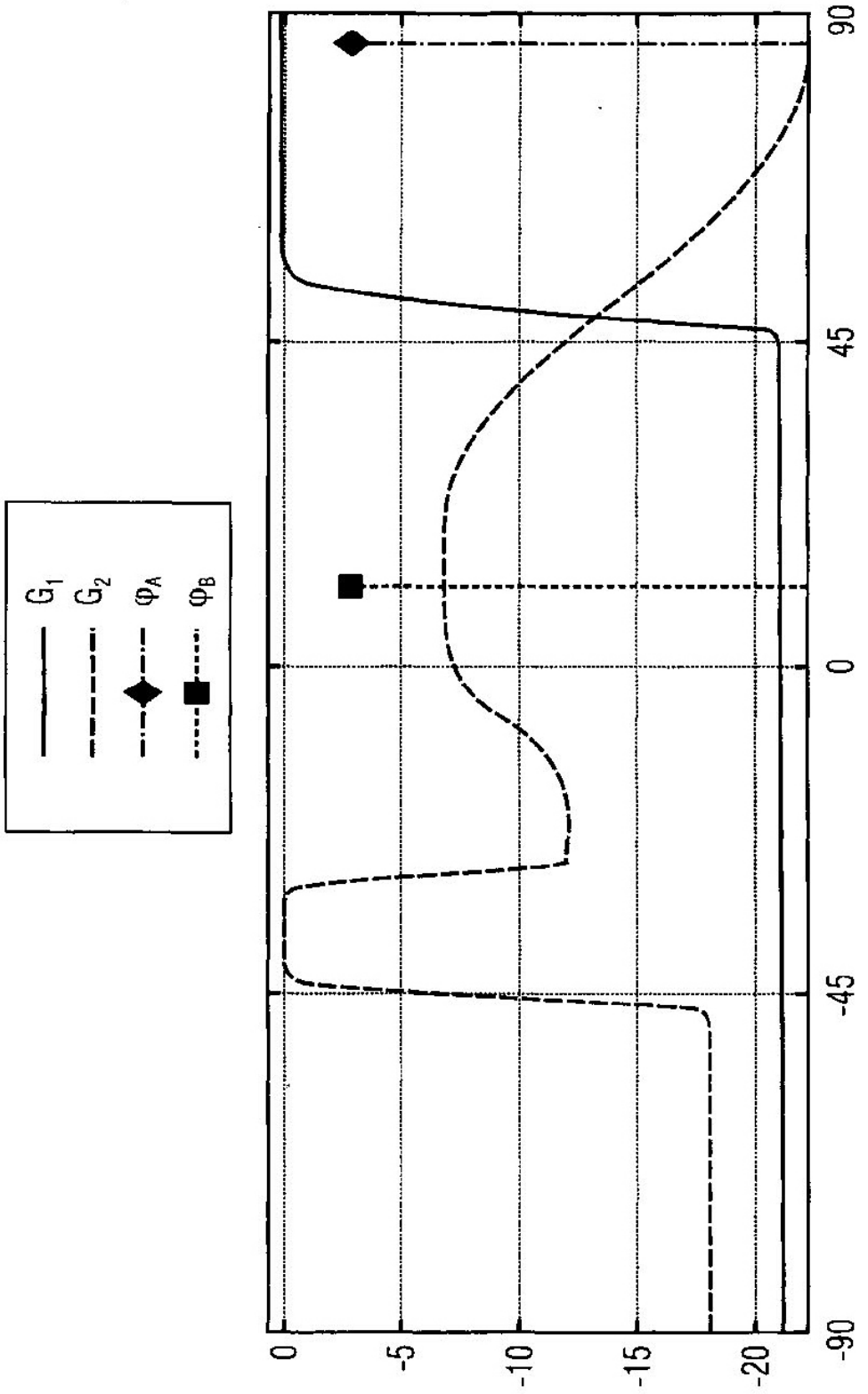

FIG 4

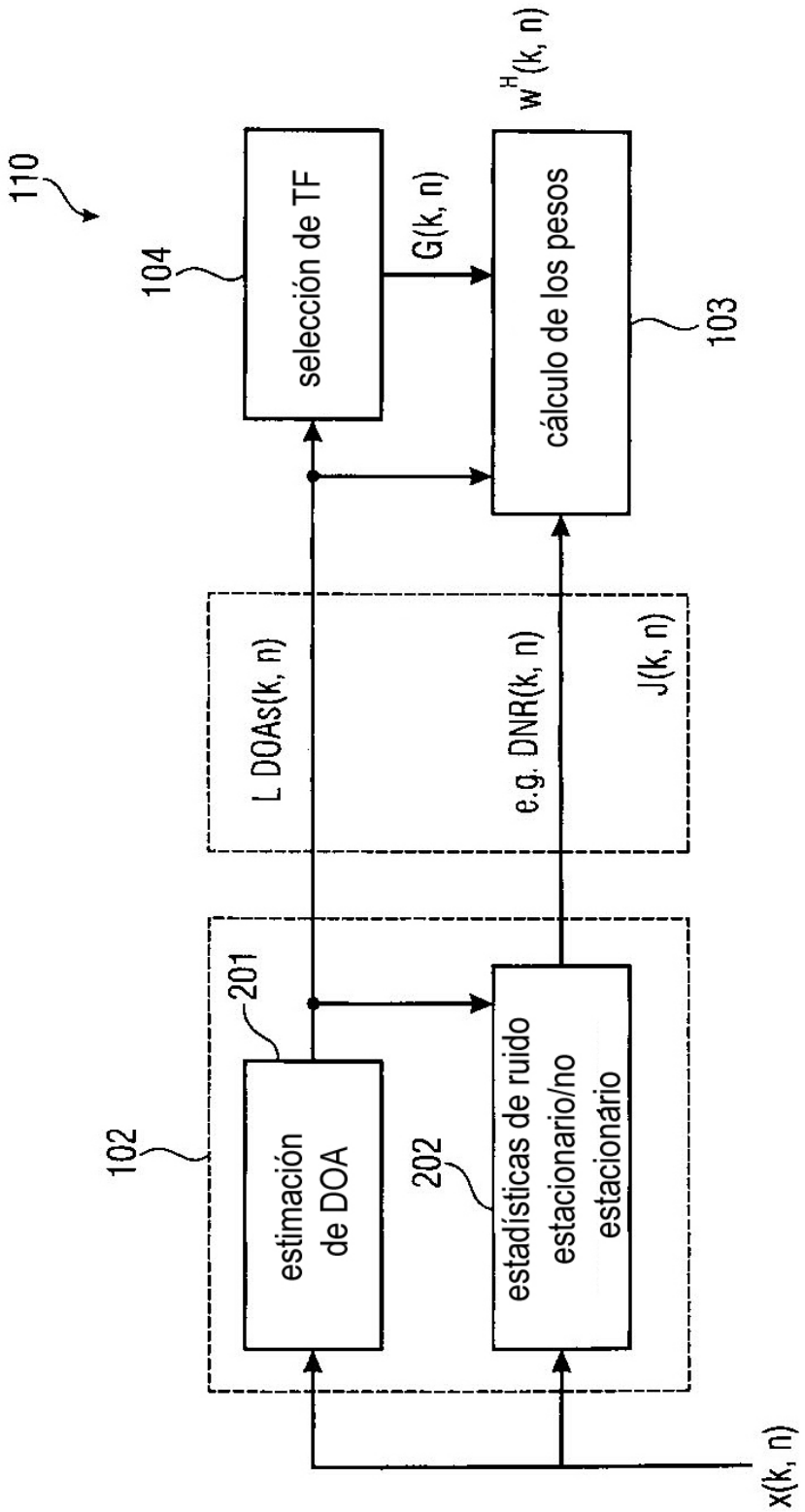
110
104
selección de TF
G(k, n)
$w^H(k, n)$
cálculo de los pesos
103
L DOAs(k, n)
e.g. DNR(k, n)
J(k, n)
102
201
estimación de DOA
202
estadísticas de ruido estacionario/no estacionario
x(k, n)

FIG 5

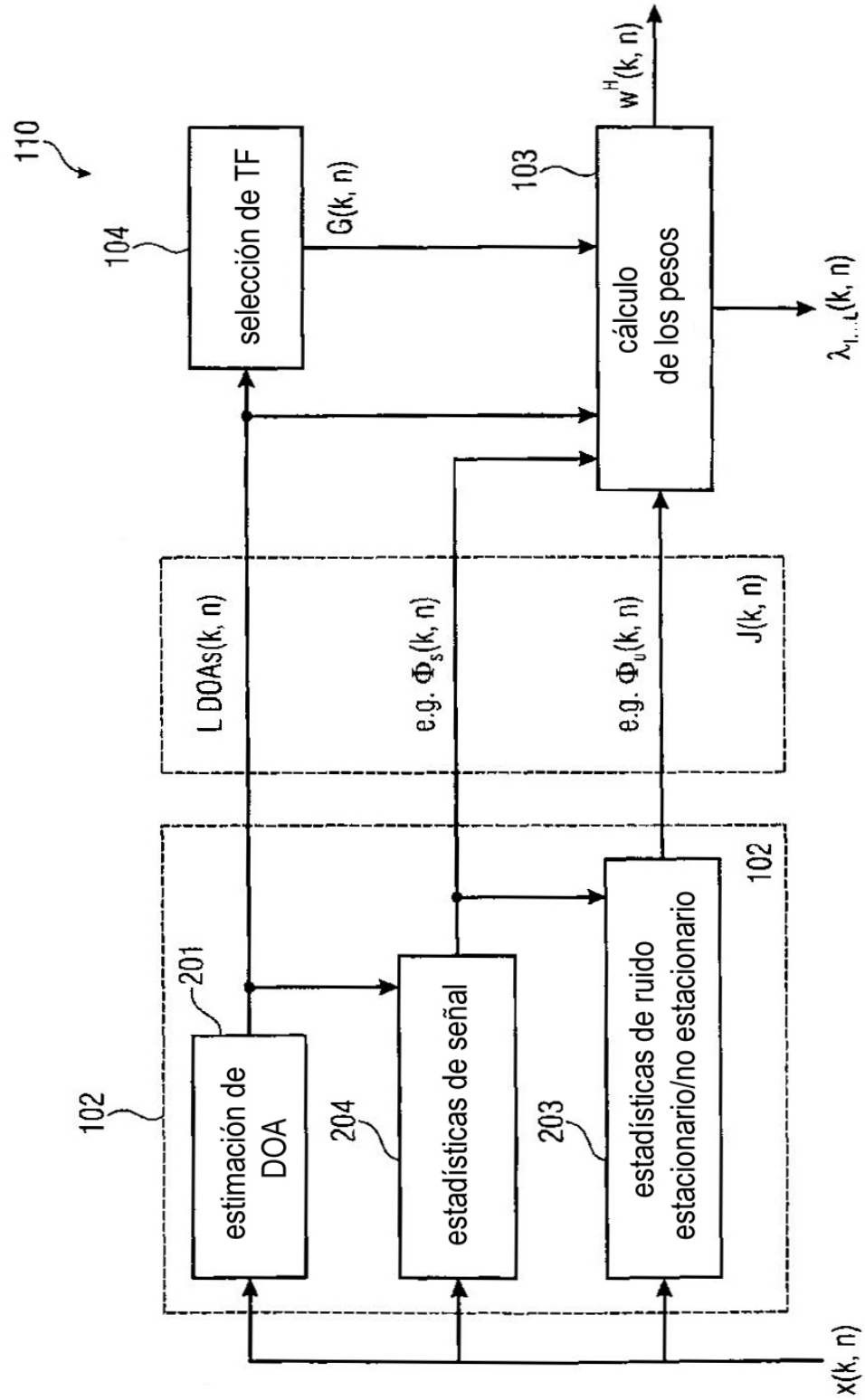
110
104
selección de TF
G(k, n)
103
cálculo
de los pesos
$w^H(k, n)$
$\lambda_{1...L}(k, n)$
L DOAs(k, n)
e.g. $\Phi_s(k, n)$
e.g. $\Phi_u(k, n)$
J(k, n)
102
201
estimación de DOA
204
estadísticas de señal
203
estadísticas de ruido estacionario/no estacionario
102
x(k, n)

FIG 6

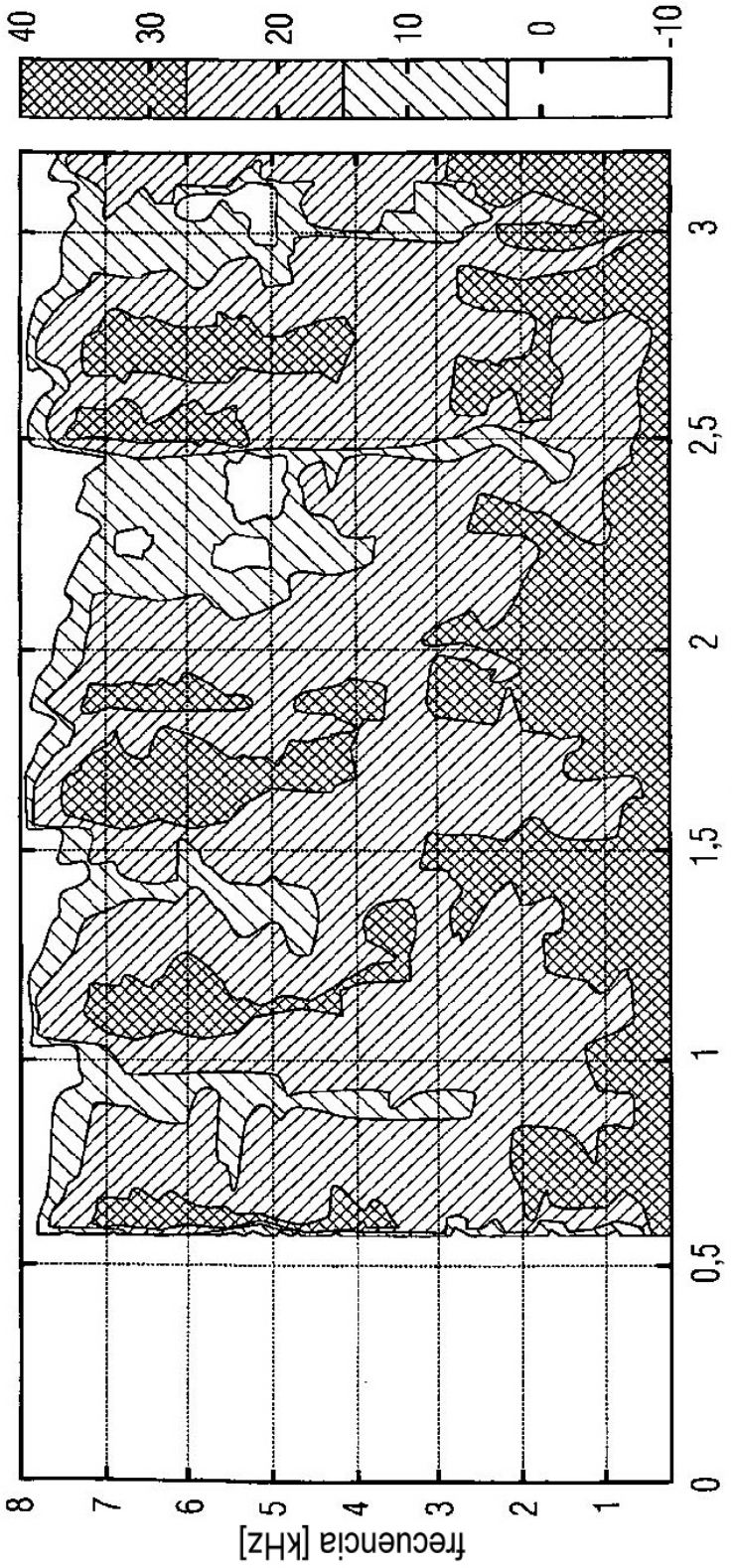

FIG 7A

REAL Ψ (k, n) [dB]

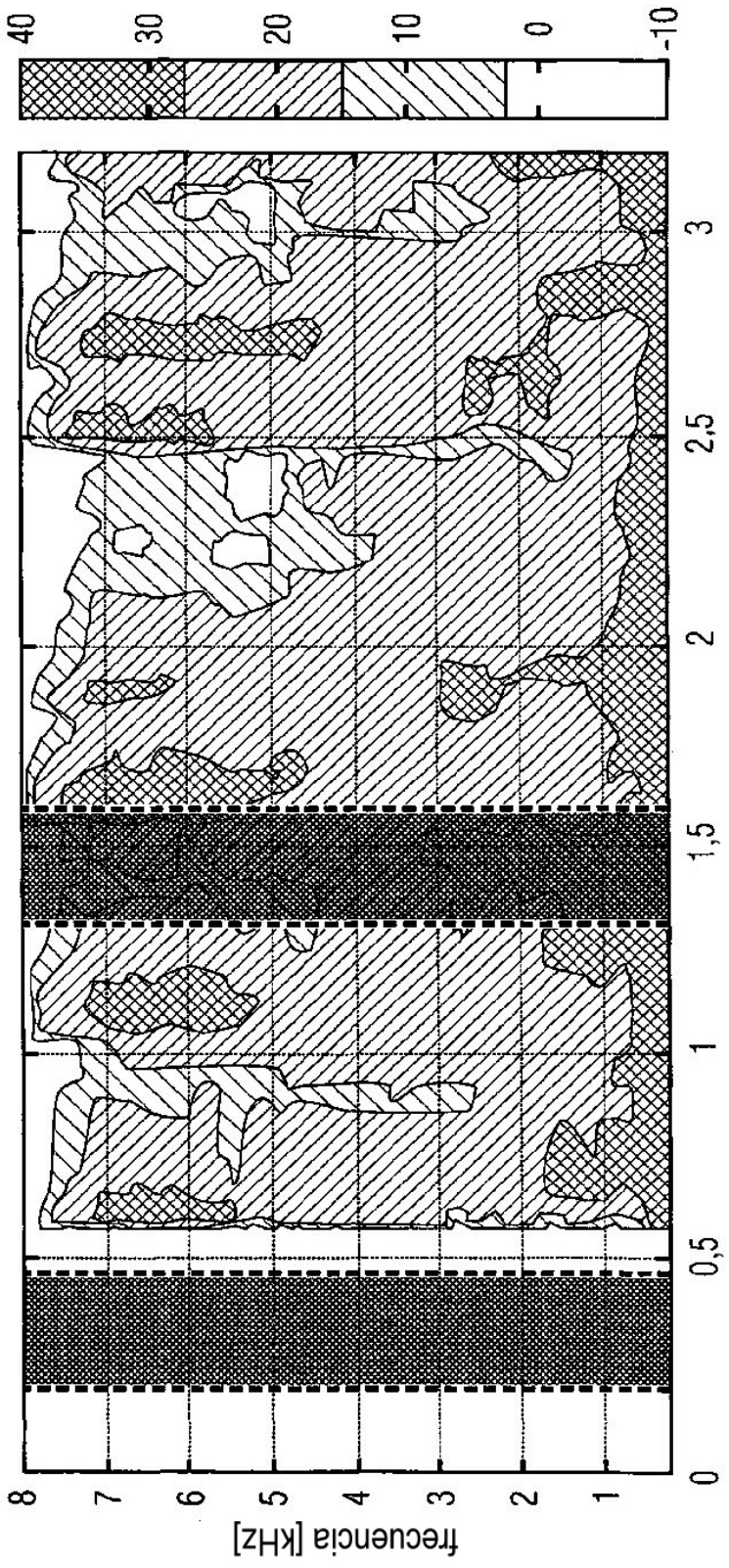

FIG 7B
ESTIMADO Ψ (k, n)[dB]

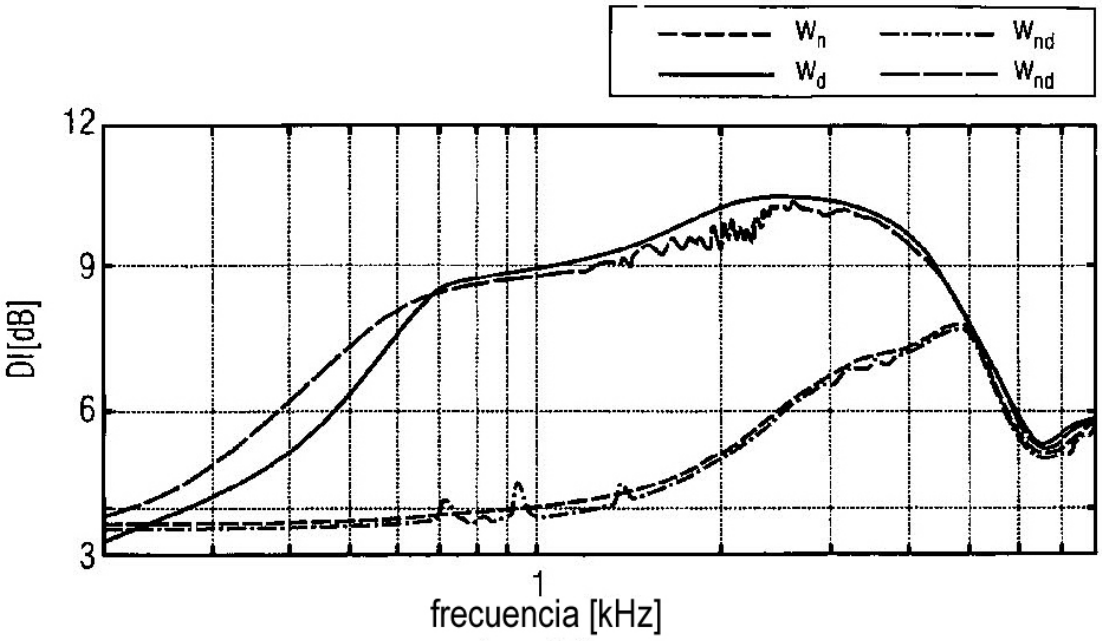

FIG 8A
DI PROMEDIO

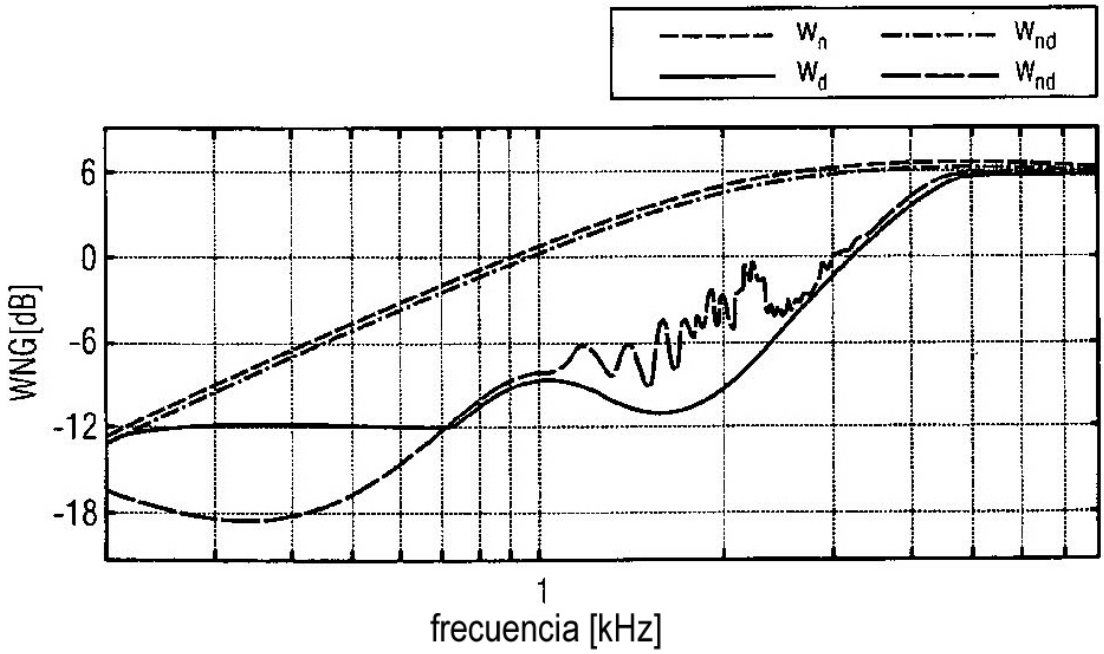

FIG 8B
WNG PROMEDIO

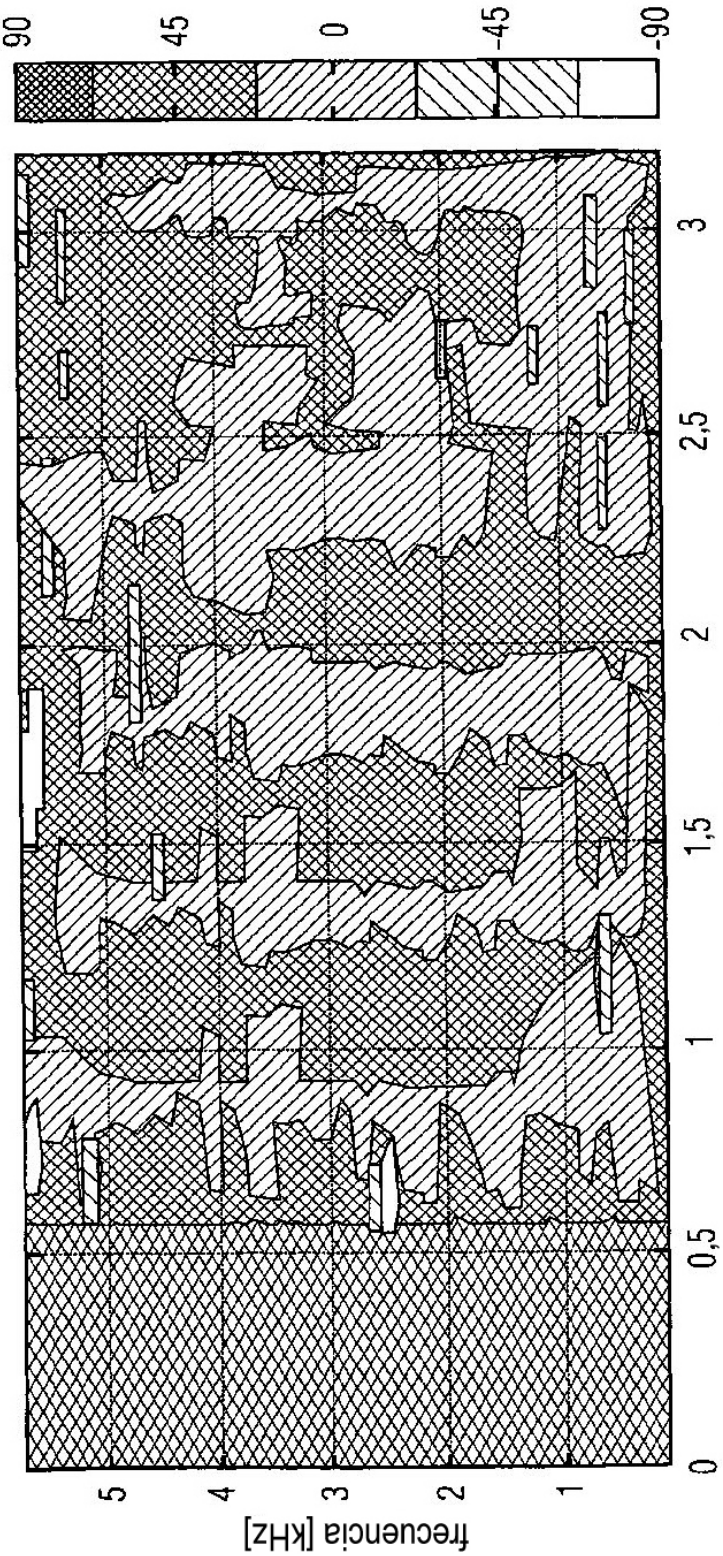

FIG 9A

DOA φ_1 (k, n)[dB] [°]

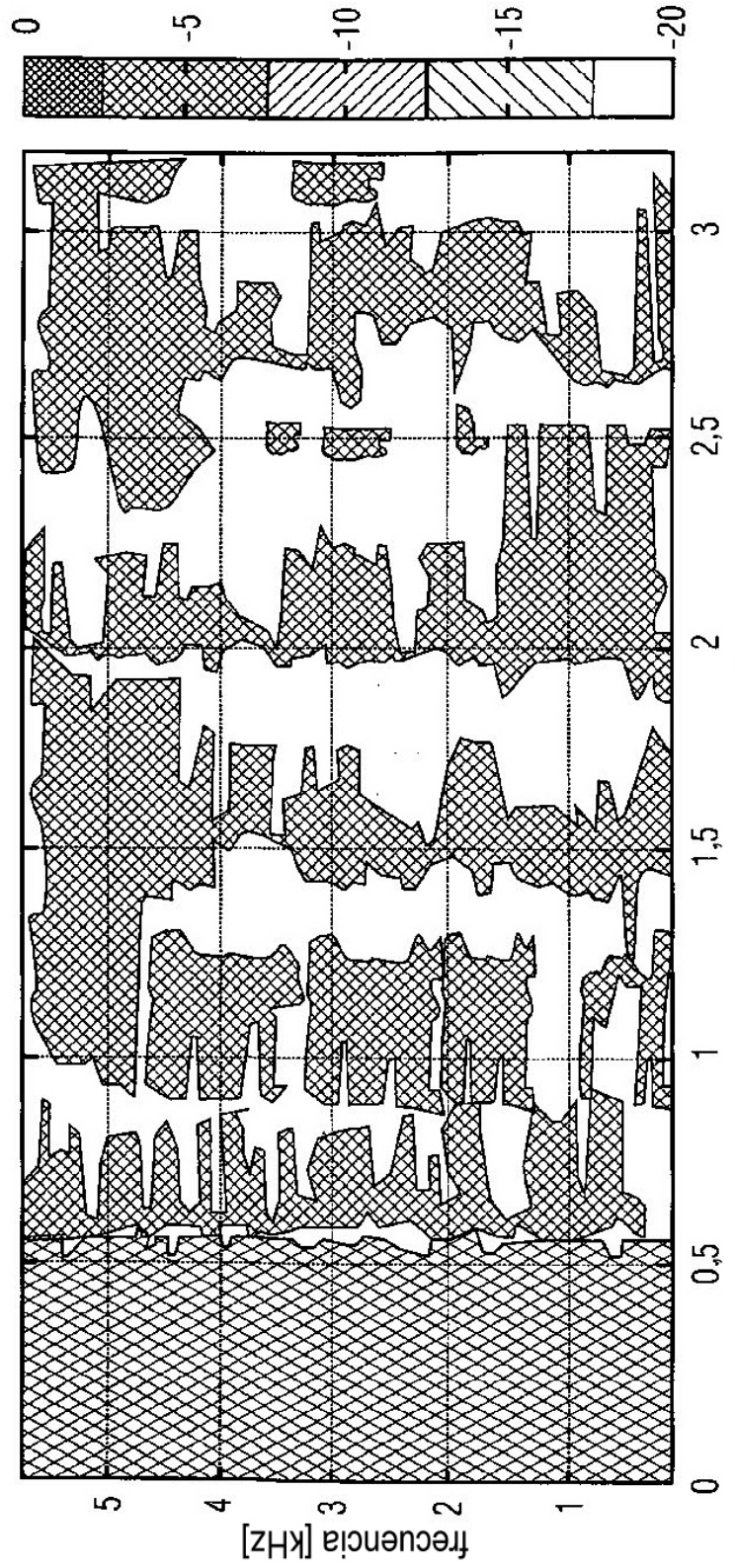

FIG 9B

$|G[k]\varphi_1(k, n)]|^2$ [dB]

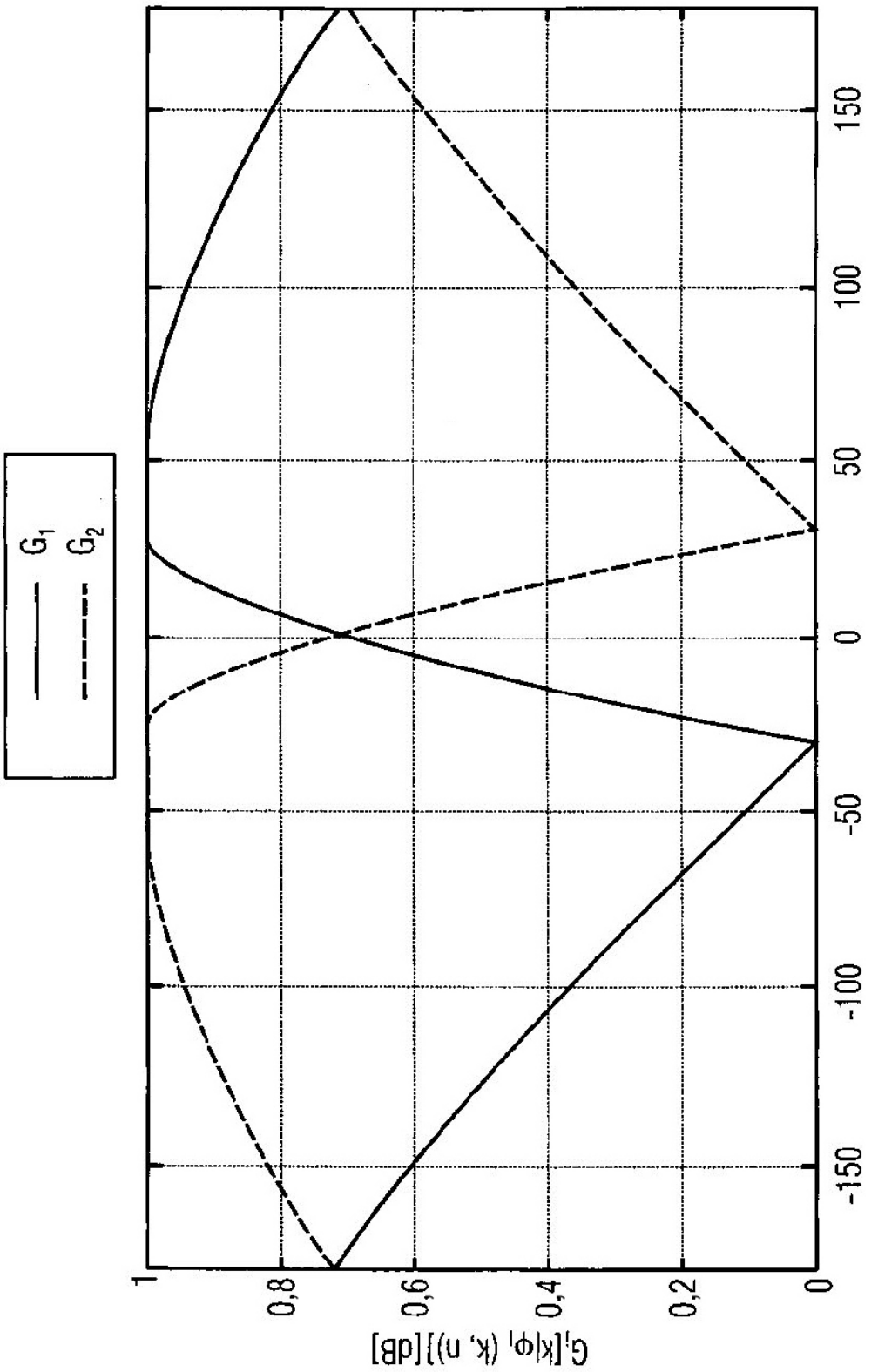

FIG 10