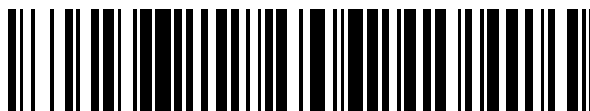


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 742 853**

51 Int. Cl.:

H04S 3/00 (2006.01)

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.10.2013 PCT/EP2013/072170**

87 Fecha y número de publicación internacional: **12.09.2014 WO14135235**

96 Fecha de presentación y número de la solicitud europea: **23.10.2013 E 13788708 (9)**

97 Fecha y número de publicación de la concesión europea: **22.05.2019 EP 2965540**

54 Título: **Aparato y procedimiento para la descomposición directa-ambiental de multicanal para el procesamiento de señales de audio**

30 Prioridad:

05.03.2013 US 201361772708 P

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

17.02.2020

73 Titular/es:

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

72 Inventor/es:

**UHLE, CHRISTIAN;
HABETS, EMANUEL;
GAMPP, PATRICK y
KRATZ, MICHAEL**

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 742 853 T3

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

DESCRIPCIÓN

Aparato y procedimiento para la descomposición directa-ambiental de multicanal para el procesamiento de señales de audio

5

[0001] La presente invención se refiere a un aparato y procedimiento para la descomposición directa-ambiental de multicanal para el procesamiento de señales de audio.

[0002] El procesamiento de señales de audio está cobrando cada vez más importancia. En este campo, la separación de señales de sonido en señales de sonido directo y ambiental cumple una función importante.

10

[0003] En general, los sonidos acústicos consisten en una mezcla de sonidos directos y sonidos ambientales (o difusos). Los sonidos directos son emitidos por fuentes de sonido, por ejemplo, un instrumento musical, un vocalista o un altavoz, y llegan por el trayecto más corto posible al receptor, por ejemplo, la entrada del oído de un oyente o un micrófono.

15

[0004] Al escuchar un sonido directo, se percibe como procedente de la dirección de la fuente de sonido. Las claves auditivas relevantes para la localización y para otras propiedades de sonido espacial son la diferencia de nivel interaural, la diferencia de tiempo interaural y la coherencia interaural. Las ondas de sonido directo que evocan la diferencia de nivel interaural y la diferencia de tiempo interaural idénticas se perciben como procedentes de la misma dirección. En ausencia de sonido difuso, las señales que llegan al oído izquierdo y al derecho, o a cualquier otra multitud de sensores son coherentes.

20

[0005] Por el contrario, los sonidos ambientales son emitidos por numerosas fuentes de sonido separadas o límites de reflexión de sonido que contribuyen al mismo sonido ambiental. Cuando una onda de sonido llega a la pared de un recinto, una porción de ésta se refleja y la superposición de todas las reflexiones en un recinto, la reverberación, es un ejemplo destacado para el sonido ambiental. Otros ejemplos son los sonidos de audiencias (por ejemplo, aplausos), sonidos del medio ambiente (por ejemplo, lluvia) y otros sonidos de fondo (por ejemplo, ruido de murmullo). Los sonidos ambientales son percibidos por el oyente como difusos, sin poderse ubicar, y evocan una impresión de envolvimiento (de estar "inmerso en el sonido"). Al capturar un campo de sonido ambiental utilizando una multitud de sensores espaciados, las señales registradas son al menos parcialmente incoherentes.

25

30

[0006] Diversas aplicaciones de post-producción y reproducción de sonido se benefician con la descomposición de las señales de audio en componentes de señal directa y componentes de señal ambiental. El principal desafío para ese tipo de procesamiento de señales consiste en obtener una alta separación manteniendo una alta calidad de sonido para un número arbitrario de señales de canal de entrada y para todas las posibles características de la señal de entrada. La descomposición directa-ambiental (DAD), es decir, la descomposición de las señales de audio en componentes de señal directa y componentes de señal ambiental, habilita la reproducción o modificación separada de los componentes de señal, lo que resulta conveniente, por ejemplo, para la mezcla ascendente de las señales de audio.

40

[0007] El término mezcla ascendente se refiere al procedimiento de crear una señal con P canales dada una señal de entrada con N canales, donde $P > N$. Su principal aplicación es la reproducción de señales de audio mediante el uso de configuraciones de sonido envolvente con más canales que los disponibles en la señal de entrada. La reproducción del contenido mediante el uso de algoritmos avanzados de procesamiento de señales permite al oyente utilizar todos los canales disponibles de la configuración de reproducción de sonido de multicanal. Tal procesamiento puede descomponer la señal de entrada en componentes significativos de señal (por ejemplo, basándose en su posición percibida en la imagen estéreo, los sonidos directos versus sonidos ambientales, instrumentos individuales) o en señales en las que estos componentes de señal están atenuados o intensificados.

50

[0008] Se conocen ampliamente dos conceptos de mezcla ascendente.

1. Mezcla ascendente guiada: mezcla ascendente con información adicional que guía el procedimiento de mezcla ascendente. La información adicional puede estar "codificada" de una manera específica en la señal de entrada o puede ser almacenada adicionalmente.

55

2. Mezcla ascendente no guiada: se obtiene la señal de salida de la señal de entrada de audio exclusivamente sin información adicional alguna.

[0009] Los procedimientos avanzados de mezcla ascendente se pueden categorizar adicionalmente con respecto al posicionamiento de las señales directas y ambientales. Se distingue entre la "estrategia de directo/ambiente" y la estrategia "en la banda". El componente central de las técnicas basadas en el concepto directo/ambiente es la extracción de una señal ambiental que es alimentada, por ejemplo, a los canales traseros o los canales de altavoz alto de una configuración de sonido envolvente de multicanal. La reproducción del ambiente

60

mediante el uso de los canales traseros o de altavoz alto evoca en el oyente una impresión de involucrimiento (de estar “inmerso en el sonido”). Además, las fuentes de sonido directo se pueden distribuir entre los canales delanteros según su posición percibida en el panorama estéreo. Por el contrario, concepto estrategia “en la banda” tiene como fin colocar todos los sonidos (sonido directo, así como los sonidos ambientales) alrededor del oyente utilizando todos los 5 altavoces disponibles.

[0010] En La descomposición de una señal de audio en señales directas y ambientales también habilita la modificación separada de los sonidos ambientales o sonidos directos, por ejemplo, mediante su escalado o filtrado. Un caso habitual es el procesamiento de una grabación de una interpretación musical que ha sido capturada con una 10 cantidad demasiado elevada de sonido ambiental. Otro caso habitual es la producción de audio (por ejemplo, para el sonido o música de películas), donde se combinan las señales de audio capturadas en diferentes ubicaciones y, por lo tanto, con diferentes características de sonido ambiental.

[0011] En cualquier caso, los requisitos para ese procesamiento de señales consisten en obtener una alta 15 separación manteniendo al mismo tiempo una alta calidad del sonido de un número arbitrario de señales de canal de entrada y para todas las posibles características de las señales de entrada.

[0012] En la técnica anterior se han dado a conocer varios conceptos para DAD o para atenuar o intensificar los componentes de señal directa o los componentes de señal ambiental y se revisan brevemente a continuación. 20

[0013] Los conceptos conocidos se relacionan con el procesamiento de señales de voz con el propósito de eliminar el ruido de fondo perjudicial de las grabaciones con micrófono.

[0014] En [1] se describe un procedimiento para atenuar la reverberación de las grabaciones de voz con dos 25 canales de entrada. Los componentes de señal de reverberación se reducen atenuando los componentes de señal no correlacionados (o difusos) en la señal de entrada. El procesamiento se implementa en el dominio de tiempo-frecuencia de tal manera que las señales de subbanda se procesen por medio del procedimiento de ponderación espectral. Los factores de ponderación de valor real se calculan empleando las densidades espectrales de potencia (PSD) 30

$$\phi_{xx}(m, k) = E \{X(m, k)X^*(m, k)\} \quad (1)$$

$$\phi_{yy}(m, k) = E \{Y(m, k)Y^*(m, k)\} \quad (2)$$

35
$$\phi_{xy}(m, k) = E \{X(m, k)Y^*(m, k)\} \quad (3)$$

donde $X(m, k)$ e $Y(m, k)$ denotan representaciones en el dominio de tiempo-frecuencia de las señales de entrada en el dominio de tiempo $x_i[n]$ e $y_i[n]$, $E\{\cdot\}$ es la operación de expectativa y X^* es el conjugado complejo de X .

40 **[0015]** Los autores originales señalan que diferentes funciones de ponderación espectral son viables cuando son proporcionales a $\phi_{xy}(m, k)$, por ejemplo, cuando se utilizan ponderaciones iguales a la función de correlación cruzada normalizada (o función de coherencia)

$$\rho(m, k) = \frac{|\Phi_{xy}(m, k)|}{\sqrt{\Phi_{xx}(m, k)\Phi_{yy}(m, k)}} \quad (4)$$

45 **[0016]** Siguiendo una línea de razonamiento similar, la descripción del procedimiento en [2] extrae una señal ambiental utilizando ponderación espectral con ponderaciones derivadas de la función de correlación cruzada normalizada calculada en bandas de frecuencia, véase la Fórmula (4) (o con los términos textuales de los autores originales, “la función de coherencia intercanal de corto tiempo”). La diferencia comparada con [1] es que en lugar de 50 atenuar los componentes de señal difusa, se atenúan los componentes de señal directa utilizando las ponderaciones espectrales que son una función estable monótona de $(1 - \rho(m, k))$.

[0017] La descomposición para la aplicación de la mezcla ascendente de las señales de entrada con dos 55 canales mediante el uso del filtrado Wiener de multicanal se ha descrito en [3]. El procesamiento se realiza en el dominio de tiempo-frecuencia. La señal de entrada se modela como mezcla de la señal ambiental y una fuente directa activa (por banda de frecuencia), donde la señal directa en un canal se limita a representar una copia en escala del componente de señal directa en el segundo canal, es decir, paneo en amplitud. El coeficiente de paneo y las potencias

de la señal directa y la señal ambiental se estiman utilizando la correlación cruzada normalizada y las potencias de las señales de entrada en ambos canales. La señal de salida directa y las señales de salida ambiental se derivan de las combinaciones lineales de las señales de entrada, con coeficientes de valor real. Se aplica un post-escalado adicional de tal manera que la potencia de las señales de salida sea igual a las cantidades estimadas.

5
[0018] El procedimiento descrito en [4] extrae una señal ambiental mediante el uso de ponderación espectral, sobre la base de un cálculo estimativo de la potencia ambiente. La potencia ambiente se estima sobre la base de presunciones de que los componentes de señal directa en ambos canales estén completamente correlacionados, de que las señales ambientales de canal no estén correlacionadas entre sí y con las señales directas, y que las potencias ambiente en ambos canales sean iguales.

10
[0019] En [5] se describe un procedimiento para la mezcla ascendente de señales estéreo basadas en la Codificación de Audio Direccional (DirAC). DirAC tiene como fin analizar y reproducir la dirección de llegada, la difusión y el espectro de un campo sonoro. Para la mezcla ascendente de señales de entrada estéreo, se simulan grabaciones anecoicas en formato B de las señales de entrada.

15
[0020] En [6] se describe un procedimiento para extraer reverberación no correlacionada de señal de audio estéreo mediante el uso de un algoritmo de filtro adaptativo que tiene como fin predecir el componente de señal directa en una señal de canal mediante el uso de la otra señal de canal por medio de un algoritmo de Mínimos Cuadrados (LMS). Posteriormente se derivan las señales ambientales mediante la sustracción de las señales directas estimadas de las señales de entrada. El fundamento de esta estrategia es que la predicción solo funciona para señales correlacionadas y que el error de predicción se asemeja a la señal no correlacionada. Existen diversos algoritmos de filtro adaptativo basados en el principio de LMS y son viables, por ejemplo, el algoritmo de LMS o de LMS Normalizados (NLMS).

20
[0021] Para la descomposición de las señales de entrada con más de dos canales, se describe un procedimiento en [7] donde en primer lugar se realiza la mezcla descendente de las señales de multicanal para obtener una señal estéreo de 2 canales y posteriormente se aplica un procedimiento para el procesamiento de señales de entrada estéreo presentado en [3].

25
[0022] En cuanto al procesamiento de señales monoaurales, el procedimiento descrito en [8] extrae una señal ambiental mediante el uso de ponderación espectral, donde las ponderaciones espectrales se calculan empleando la extracción de características y aprendizaje supervisado.

30
[0023] Otro procedimiento para extraer una señal ambiental de grabaciones monoaurales para la aplicación de la mezcla ascendente obtiene la representación en el dominio de tiempo-frecuencia de la diferencia de la representación en el dominio de tiempo-frecuencia de la señal de entrada y una versión comprimida de la misma, preferentemente calculada utilizando factorización no negativa de matrices [9].

35
[0024] En [10] se describe un procedimiento para extraer y cambiar los componentes de señal reverberante de una señal de audio sobre la base de una estimación de la función de transferencia de magnitud del sistema reverberante que ha generado la señal reverberante. Se deriva una estimación de las magnitudes de la representación en el dominio de frecuencia de los componentes de señal por medio de filtrado recursivo y puede ser modificada.

40
[0025] El documento WO 2011/104146 A1 describe un aparato para generar una señal de mezcla descendente mejorada sobre la base de una señal de micrófono multicanal que comprende un analizador espacial configurado para calcular un conjunto de parámetros de señal espacial que comprende una información de dirección que describe una dirección de llegada de un sonido directo, una información de potencia de sonido directo y una información de potencia de sonido difuso sobre la base de la señal de micrófono multicanal.

45
[0026] Iain McCowan et al: "Post-filtro de matriz de micrófonos para ruido difuso archivado", Conferencia Internacional de IEEE sobre acústica, habla y procesamiento de señales 2002. Procedimientos. (ICASSP) Orlando, FL, del 13 al 17 de mayo de 2002; [IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Conferencia Internacional de IEEE sobre acústica, habla y procesamiento de señales)], Nueva York, NY, IEEE, EE.UU., 13 de mayo de 2002, páginas 1905, describe una técnica para estimar la densidad espectral de potencia de señal que se utilizará en la función de transferencia de un post-filtro de matriz de micrófonos. La técnica es una generalización del post-filtro Zelinski existente, que utiliza las densidades espectrales automáticas y cruzadas de las entradas de matriz para estimar las densidades espectrales de señal y ruido.

50
[0027] Andreas Walther y col. al.: "Descomposición ambiental directa y mezcla ascendente de señales envolventes", Aplicaciones del procesamiento de señales para audio y acústica (WASPAA), Taller IEEE 2011 sobre, IEEE, 16 de octubre de 2011, páginas 277 - 280, describe un procedimiento, que estima componentes de señal ambiental y directa en una mezcla descendente de una señal envolvente dada. Esas estimaciones se utilizan entonces para calcular filtros para obtener señales envolventes directas y ambientales. También se describen estrategias sobre

cómo usar las señales descompuestas para la mezcla ascendente de señales de audio envolvente en dos y tres dimensiones.

[0028] El objetivo de la presente invención es proporcionar conceptos mejorados para descomposición directa-ambiental de multicanal para el procesamiento de señales de audio. El objetivo de la presente invención se soluciona mediante un aparato según la reivindicación 1, mediante un procedimiento según la reivindicación 14 y mediante un programa informático según la reivindicación 15.

[0029] Se proporciona un aparato para generar una o más señales de canal de salida de audio dependiendo de dos o más señales de canal de entrada de audio según la reivindicación 1. Cada una de las dos o más señales de canal de entrada de audio comprende porciones de señal directa y porciones de señal ambiental. El aparato comprende una unidad de determinación de filtros para determinar un filtro mediante la estimación de primera información de densidad espectral de potencia y mediante la segunda estimación de información de densidad espectral de potencia. Además, el aparato comprende un procesador de señales para generar una o más señales de canal de salida de audio mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio. La primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio. O bien la primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio. O bien, la primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio.

[0030] Las formas de realización proporcionan conceptos para descomponer señales de entrada de audio en componentes de señal directa y componentes de señal ambiental, que se pueden aplicar a la post-producción y reproducción de sonido. El principal desafío en ese tipo de procesamiento de señales consiste en obtener una alta separación y al mismo tiempo mantener una alta calidad de sonido con respecto a un número arbitrario de señales de canal de entrada y con respecto a todas las características posibles de las señales de entrada. Los conceptos proporcionados se basan en el procesamiento de señales de multicanal en el dominio de tiempo-frecuencia, que lleva a una solución óptima limitada en el sentido del error cuadrático medio y, por ejemplo sujeto a restricciones en cuanto a la distorsión de las señales estimadas pretendidas o a la reducción de la interferencia residual.

[0031] Se proporcionan formas de realización para la descomposición de señales de entrada de audio en componentes de señal directa y componentes de señal ambiental. Además, se describe una derivación de filtros para calcular los componentes de señal ambiental y asimismo, formas de realización para la aplicación de los filtros.

[0032] Algunas formas de realización se relacionan con la mezcla ascendente no guiada siguiendo la estrategia de directo/ambiente con señales de entrada que tienen más de un canal.

[0033] Para las aplicaciones contempladas de la descomposición descrita, el interés está en el cálculo de señales de salida con el mismo número de canales que las señales de entrada. Para esta aplicación, las formas de realización ofrecen muy buenos resultados en términos de separación y calidad del sonido, puesto que puede hacer frente a señales de entrada donde las señales directas tienen retardo de tiempo entre los canales de entrada. A diferencia de otros conceptos, por ejemplo los conceptos presentados en [3], las formas de realización no asumen que los sonidos directos incluidos en las señales de entrada se panean solo mediante escalado (paneo en amplitud), sino también mediante la introducción de diferencias de tiempo entre las señales directas de cada canal.

[0034] Además, las formas de realización pueden operar sobre señales de entrada con un número arbitrario de canales, a diferencia de todos los demás conceptos de la técnica anterior (véase más arriba), que solo pueden procesar señales de entrada con uno o dos canales.

[0035] Otras ventajas de las formas de realización son el uso de parámetros de control, la estimación de la matriz PSD ambiente y otras modificaciones del filtro, como se describe más adelante.

[0036] Algunas formas de realización proporcionan sonidos ambientales constantes para todos los objetos sonoros de entrada. Al descomponer las señales de entrada en sonidos directos y ambientales, algunas formas de realización adaptan las características del sonido ambiental por medio del procesamiento apropiado de las señales de audio y otras formas de realización reemplazan los componentes de señal ambiental por medio de reverberación artificial y otros sonidos ambientales artificiales.

[0037] Según una forma de realización, el aparato puede comprender además un banco de filtros de análisis que está configurado para transformar las dos o más señales de canal de entrada de audio del dominio de tiempo a un dominio de tiempo-frecuencia. La unidad de determinación de filtros puede estar configurada para determinar el filtro mediante la estimación de la primera información de densidad espectral de potencia y la segunda información de densidad espectral de potencia dependiendo de las señales de canal de entrada de audio, que están representadas en el dominio de tiempo-frecuencia. El procesador de señales puede estar configurado para generar una o más señales de canal de salida de audio, que están representadas en un dominio de tiempo-frecuencia, mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio, que están representadas en el dominio de tiempo-frecuencia. Asimismo, el aparato puede comprender además un banco de filtros de síntesis que está configurado para transformar una o más señales de canal de salida de audio, que están representadas en un dominio de tiempo-frecuencia, del dominio de tiempo-frecuencia al dominio de tiempo.

[0038] Además, se proporciona un procedimiento para generar una o más señales de canal de salida de audio dependiendo de dos o más señales de canal de entrada de audio según la reivindicación 14. Cada una de las dos o más señales de canal de entrada de audio comprende porciones de señal directa y porciones de señal ambiental. El procedimiento comprende:

- Determinación de un filtro mediante la estimación de la primera información de densidad espectral de potencia y mediante la segunda estimación de información de densidad espectral de potencia y;
- Generación de una o más señales de canal de salida de audio mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio.

[0039] La primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio. O bien la primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio. O bien la primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio.

[0040] Asimismo, se proporciona un programa informático para implementar el procedimiento antes descrito al ejecutarse en un ordenador o un procesador de señales según la reivindicación 15.

[0041] A continuación se describen las formas de realización de la presente invención en forma más detallada con referencia a las figuras, en las cuales:

la figura 1 ilustra un aparato para generar una o más señales de canal de salida de audio dependiendo de dos o más señales de canal de entrada de audio según una forma de realización,

la figura 2 ilustra señales de entrada y salida de la descomposición de una grabación con 5 canales de música clásica, con señales de entrada (columna izquierda), señales de salida ambiental (columna central) y señales de salida directa (columna derecha) según una forma de realización,

la figura 3 ilustra un resumen general básico de la descomposición que utiliza la estimación de señales ambientales y la estimación de señales directas según una forma de realización,

la figura 4 ilustra un resumen general básico de la descomposición que utiliza la estimación de señales directas según una forma de realización,

la figura 5 ilustra un resumen general básico de la descomposición que utiliza la estimación de señales ambiente según una forma de realización,

la figura 6a ilustra un aparato según otra forma de realización, en la que el aparato comprende además un banco de filtros de análisis y un banco de filtros de síntesis, y

la figura 6b representa un aparato según otra forma de realización, que muestra la extracción de los componentes de señal directa, en la que el bloque AFB es una serie de N bancos de filtros de análisis (uno por cada canal), y en la que SFB es una serie de bancos de filtros de síntesis.

[0042] La figura 1 ilustra un aparato para generar una o más señales de canal de salida de audio dependiendo

de dos o más señales de canal de entrada de audio según una forma de realización. Cada una de las dos o más señales de canal de entrada de audio comprende porciones de señal directa y porciones de señal ambiental.

[0043] El aparato comprende una unidad de determinación de filtros 110 para determinar un filtro mediante la estimación de primera información de densidad espectral de potencia y mediante la segunda estimación de información de densidad espectral de potencia.

[0044] Asimismo, el aparato comprende un procesador de señales 120 para generar una o más señales de canal de salida de audio mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio.

[0045] La primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio.

[0046] bien, la primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio.

[0047] bien, la primera información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio y la segunda información de densidad espectral de potencia indica la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio.

[0048] Se describen formas de realización que proporcionan conceptos para descomponer señales de entrada de audio en componentes de señal directa y componentes de señal ambiental que se pueden aplicar para la post-producción y reproducción de sonido. El principal desafío para ese tipo de procesamiento de señales consiste en obtener una alta separación manteniendo una alta calidad de sonido para un número de señales de canal de entrada y para todas las posibles características de la señal de entrada. Las formas de realización presentadas se basan en el procesamiento de señales de multicanal en el dominio de tiempo-frecuencia y ofrecen una solución óptima en el sentido del error cuadrático medio sometido a las restricciones sobre la distorsión de las señales estimadas deseadas o sobre la reducción de la interferencia residual.

[0049] En primer lugar se describen los conceptos de la invención, en los cuales se basan las formas de realización de la presente invención.

[0050] Se asume que se reciben N señales de canal de entrada $\mathbf{y}[n]$:

$$\mathbf{y}_t[n] = [y_1[n] \dots y_N[n]]^T. \quad (5)$$

[0051] Por ejemplo, $N \geq 2$. El objetivo de los conceptos presentados es descomponer las señales de canal de entrada $y_1[n] \dots y_M[n]$ ($= [\mathbf{y}[n]]^T$) en N componentes de señal directa indicados por $\mathbf{d}[n] = [d_1[n] \dots d_M[n]]^T$ y/o N componentes de señal ambiental indicados por $\mathbf{a}[n] = [a_1[n] \dots a_M[n]]^T$. El procesamiento se puede aplicar para todos los canales de entrada, o se pueden dividir los canales de la señal de entrada en subconjuntos de canales que se procesan por separado.

[0052] Según las formas de realización, se estima uno o más de los componentes de señal directa $d_1[n], \dots, d_M[n]$ y/o uno o más de los componentes de señal ambiental $a_1[n], \dots, a_M[n]$ sobre la base de las dos o más señales de canal de entrada $y_1[n], \dots, y_M[n]$ para obtener una o más estimaciones ($\hat{d}_1[n], \dots, \hat{d}_N[n], \hat{a}_1[n], \dots, \hat{a}_N[n]$) de los componentes de señal directa $d_1[n], \dots, d_M[n]$ y/o de los componentes de señal ambiental $a_1[n], \dots, a_M[n]$ como una o más señales de canal de salida.

[0053] En la figura 2 se ilustra un ejemplo de las salidas proporcionadas en algunas formas de realización, para $N = 5$. Una o más señales de canal de salida de audio $\hat{d}_1[n], \dots, \hat{d}_N[n]$ ($= [\hat{\mathbf{d}}_t[n]]^T$), $\hat{a}_1[n], \dots, \hat{a}_N[n]$ ($= [\hat{\mathbf{a}}_t[n]]^T$) se obtienen estimando los componentes de señal directa y los componentes de señal ambiental de forma independiente, como se ilustra en la figura 3. Por otro lado, se hace un cálculo estimativo ($\hat{\mathbf{d}}_t[n]$ o $\hat{\mathbf{a}}_t[n]$) con respecto a una de las dos señales ($\mathbf{d}[n]$ o $\mathbf{a}[n]$) y la otra señal se obtiene restando el primer resultado de la señal de entrada. La figura 4 ilustra el procesamiento para estimar los componentes de señal directa $\mathbf{d}[n]$ en primer lugar y

derivar los componentes de señal ambiental $\mathbf{a}[n]$ restando el cálculo estimativo de las señales directas de la señal de entrada. Con un razonamiento similar, se puede derivar la estimación de los componentes de señal ambiental en primer lugar según lo ilustrado en el diagrama de bloques de la figura 5.

5 **[0054]** Según las formas de realización, el procesamiento se puede realizar, por ejemplo, en el dominio de tiempo-frecuencia. Se puede obtener una representación en el dominio de tiempo-frecuencia de la señal de entrada de audio, por ejemplo, por medio de un banco de filtros (el banco de filtros de análisis), por ejemplo la transformada de Fourier de Tiempo Reducido (STFT).

10 **[0055]** Según una forma de realización ilustrada en la figura 6a, un banco de filtros de análisis 605 transforma las señales de canal de entrada de audio $\mathbf{y}[n]$ del dominio de tiempo al dominio de tiempo-frecuencia. Asimismo, en la figura 6a, un banco de filtros de síntesis 625 transforma la estimación de los componentes de señal directa $\hat{d}[m,1], \dots, \hat{d}[m,k]$ del dominio de tiempo-frecuencia al dominio de tiempo, para obtener las señales de canal de salida de audio $\hat{d}_1[n], \dots, \hat{d}_N[n]$ ($= [\hat{\mathbf{d}}_t[n]]^T$).

15 **[0056]** En la forma de realización de la figura 6a, el banco de filtros de análisis 605 está configurado para transformar las dos o más señales de canal de entrada de audio del dominio de tiempo a un dominio de tiempo-frecuencia. La unidad de determinación de filtros 110 está configurada para determinar el filtro mediante la estimación de la primera información de densidad espectral de potencia y la segunda información de densidad espectral de potencia dependiendo de las señales de canal de entrada de audio, que están representadas en el dominio de tiempo-frecuencia. El procesador de señales 120 está configurado para generar una o más señales de canal de salida de audio, que están representadas en un dominio de tiempo-frecuencia, mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio, que están representadas en el dominio de tiempo-frecuencia. El banco de filtros de síntesis 625 está configurado para transformar una o más señales de canal de salida de audio, que están representadas en un dominio de tiempo-frecuencia, del dominio de tiempo-frecuencia al dominio de tiempo.

20 **[0057]** Una representación en el dominio de tiempo-frecuencia comprende un cierto número de señales de subbanda que evolucionan con el tiempo. Opcionalmente se pueden combinar las subbandas adyacentes en forma lineal formando señales de subbanda más amplias para reducir la complejidad informática. Cada subbanda de las 30 señales de entrada es procesada por separado, como se describe en detalle a continuación. Se obtienen señales de salida en el dominio de tiempo aplicando el procesamiento inverso del banco de filtros, es decir el banco de filtros de síntesis, respectivamente. Se asume que todas las señales tienen media cero, las señales en el dominio de tiempo-frecuencia pueden ser modeladas en forma de variables aleatorias complejas.

35 **[0058]** A continuación se proporcionan definiciones y supuestos.

[0059] Las siguientes definiciones se utilizan en toda la descripción del procedimiento ideado. La representación en el dominio de tiempo-frecuencia de una señal de entrada de multicanal con N canales está dada por

$$\mathbf{y}(m, k) = [Y_1(m, k) Y_2(m, k) \cdots Y_N(m, k)]^T, \quad (6)$$

donde el índice de tiempo m y el índice de subbanda k , $k = 1 \dots K$ y se presume que es una mezcla aditiva del componente de señal directa $\mathbf{d}(m, k)$ y el componente de señal ambiental $\mathbf{a}(m, k)$, es decir

$$\mathbf{y}(m, k) = \mathbf{d}(m, k) + \mathbf{a}(m, k), \quad (7)$$

donde

$$\mathbf{d}(m, k) = [D_1(m, k) D_2(m, k) \cdots D_N(m, k)]^T \quad (8)$$

$$\mathbf{a}(m, k) = [A_1(m, k) A_2(m, k) \cdots A_N(m, k)]^T, \quad (9)$$

donde $D_i(m, k)$ denota el componente directo y $A_i(m, k)$ el componente ambiental del i° canal.

55 **[0060]** El objetivo de la descomposición directa-ambiental es estimar $\mathbf{d}(m, k)$ y $\mathbf{a}(m, k)$. Las señales de salida se calculan utilizando las matrices de filtros $\mathbf{H}_D(m, k)$ o $\mathbf{H}_A(m, k)$ o ambas. Las matrices de filtros son de un tamaño $N \times N$ y son de valor complejo o, en algunas formas de realización, pueden ser, por ejemplo, de valor real. Se obtiene un

cálculo estimativo de las señales de N -canales de los componentes de señal directa y los componentes de señal ambiental de

$$\hat{\mathbf{d}}(m, k) = \mathbf{H}_D^H(m, k)\mathbf{y}(m, k) \quad (10)$$

$$\hat{\mathbf{a}}(m, k) = \mathbf{H}_A^H(m, k)\mathbf{y}(m, k), \quad (11)$$

[0061] Por otro lado, solo se puede emplear una matriz de filtros, y la substracción ilustrada en la figura 4 se puede expresar de la siguiente manera

$$\hat{\mathbf{d}}(m, k) = \mathbf{H}_D^H(m, k)\mathbf{y}(m, k) \quad (12)$$

$$\hat{\mathbf{a}}(m, k) = [\mathbf{I} - \mathbf{H}_D(m, k)]^H \mathbf{y}(m, k), \quad (13)$$

15 donde \mathbf{I} es la matriz de identidad del tamaño $N \times N$, o, como se ilustra en la figura 5, como

$$\hat{\mathbf{a}}(m, k) = \mathbf{H}_A^H(m, k)\mathbf{y}(m, k) \quad (14)$$

$$\hat{\mathbf{d}}(m, k) = [\mathbf{I} - \mathbf{H}_A(m, k)]^H \mathbf{y}(m, k), \quad (15)$$

20 respectivamente. En este caso, el superíndice H denota la transposición conjugada de una matriz o un vector. La matriz de filtros $\mathbf{H}_D(m, k)$ se utiliza para ejecutar cálculos estimativos correspondientes a las señales directas $\hat{\mathbf{d}}(m, k)$. La matriz de filtros $\mathbf{H}_A(m, k)$ se utiliza para hacer cálculos estimativos correspondientes a las señales ambiente $\hat{\mathbf{a}}(m, k)$.

25 **[0062]** En las Fórmulas anteriores (10) – (15), $\mathbf{y}(m, k)$ indica las dos o más señales de canal de entrada de audio. $\hat{\mathbf{a}}(m, k)$ indica una estimación de las porciones de señal ambiental y $\hat{\mathbf{d}}(m, k)$ indica una estimación de las porciones de señal directa de las señales de canal de entrada de audio, respectivamente. $\hat{\mathbf{a}}(m, k)$ y/o $\hat{\mathbf{d}}(m, k)$ o uno más componentes de vector de $\hat{\mathbf{a}}(m, k)$ y/o $\hat{\mathbf{d}}(m, k)$ pueden ser una o más señales de canal de salida de audio.

30 **[0063]** El procesador de señales 120 de la figura 1 y la figura 6a pueden emplear una, algunas o todas las Fórmulas (10), (11), (12), (13), (14) y (15) 120 de la figura 1 y la figura 6a para aplicar el filtro de la figura 1 y la figura 6a a las señales de canal de entrada de audio. El filtro de la figura 1 y la figura 6a puede ser, por ejemplo, $\mathbf{H}_D(m, k)$, $\mathbf{H}_A(m, k)$, $\mathbf{H}_D^H(m, k)$, $\mathbf{H}_A^H(m, k)$, $[\mathbf{I} - \mathbf{H}_D(m, k)]$ o $[\mathbf{I} - \mathbf{H}_A(m, k)]$. En otras formas de realización, sin embargo, el filtro, 35 determinado por la unidad de determinación de filtros 110 y empleado por el procesador de señales 120, puede no ser una matriz sino otro tipo de filtro. Por ejemplo, en otras formas de realización, el filtro puede comprender uno o más vectores que definen el filtro. En otras formas de realización, el filtro puede comprender una pluralidad de coeficientes que definen el filtro.

40 **[0064]** Las matrices de filtros se calculan a partir de los cálculos estimativos de las estadísticas de la señal, como se describe más adelante.

[0065] En particular, la unidad de determinación de filtros 110 está configurada para determinar el filtro mediante la estimación de la primera información de densidad espectral de potencia (PSD) y la segunda información 45 PSD.

[0066] Define:

$$\phi_{X_i X_j}(m, k) = E\{X_i(m, k)X_j^*(m, k)\}, \quad (16)$$

50 donde $E\{\cdot\}$ es el operador de expectativa y X^* denota el conjugado complejo de X . En el caso de $i = j$ la PSD y para i

$\neq j$ se obtienen las PSD cruzadas.

[0067] Las matrices de covarianza correspondientes a $\mathbf{y}(m, k)$, $\mathbf{d}(m, k)$ y $\mathbf{a}(m, k)$ son

$$\Phi_{\mathbf{y}}(m, k) = E\{\mathbf{y}(m, k)\mathbf{y}^H(m, k)\} \quad (17)$$

$$\Phi_{\mathbf{d}}(m, k) = E\{\mathbf{d}(m, k)\mathbf{d}^H(m, k)\} \quad (18)$$

$$\Phi_{\mathbf{a}}(m, k) = E\{\mathbf{a}(m, k)\mathbf{a}^H(m, k)\}. \quad (19)$$

[0068] Las matrices de covarianza $\Phi_{\mathbf{y}}(m, k)$, $\Phi_{\mathbf{d}}(m, k)$ y $\Phi_{\mathbf{a}}(m, k)$ comprenden estimaciones de la PSD correspondiente a todos los canales de la diagonal principal, mientras que los elementos fuera de la diagonal son estimaciones de la PSD cruzada de las señales de los respectivos canales. Por consiguiente, cada una de las matrices $\Phi_{\mathbf{y}}(m, k)$, $\Phi_{\mathbf{d}}(m, k)$ y $\Phi_{\mathbf{a}}(m, k)$ representa una estimación de la información de densidad espectral de potencia.

[0069] En las Fórmulas (17) – (19), $\Phi_{\mathbf{y}}(m, k)$ indica una información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio. $\Phi_{\mathbf{d}}(m, k)$ indica una información de densidad espectral de potencia sobre los componentes de señal directa de las dos o más señales de canal de entrada de audio. $\Phi_{\mathbf{a}}(m, k)$ indica una información de densidad espectral de potencia sobre los componentes de señal ambiental de las dos o más señales de canal de entrada de audio.

[0070] Cada una de las matrices $\Phi_{\mathbf{y}}(m, k)$, $\Phi_{\mathbf{d}}(m, k)$ y $\Phi_{\mathbf{a}}(m, k)$ de las Fórmulas (17), (18) y (19) se puede considerar información de densidad espectral de potencia. No obstante, se debe tener en cuenta que en otras formas de realización, la primera y la segunda información de densidad espectral de potencia no son una matriz, sino que pueden estar representadas en cualquier otro tipo de formato adecuado. Por ejemplo, según ciertas formas de realización, la primera y/o la segunda información de densidad espectral de potencia pueden estar representadas en forma de uno o más vectores. En otras formas de realización, la primera y/o la segunda información de densidad espectral de potencia pueden estar representadas en forma de una pluralidad de coeficientes.

[0071] Se asume que

• $D_i(m, k)$ y $A_j(m, k)$ no están mutuamente correlacionadas:

$$E\{D_i(m, k)A_j^*(m, k)\} = 0 \quad \forall i, j,$$

• $A_i(m, k)$ y $A_j(m, k)$ no están mutuamente correlacionadas:

$$E\{A_i(m, k)A_j^*(m, k)\} = 0 \quad \forall i \neq j,$$

• La potencia ambiente es igual en todos los canales:

$$E\{A_i(m, k)A_j^*(m, k)\} = \phi_A(m, k) \quad \forall i = j.$$

[0072] En consecuencia, surge que

$$\Phi_{\mathbf{y}}(m, k) = \Phi_{\mathbf{d}}(m, k) + \Phi_{\mathbf{a}}(m, k), \quad (20)$$

$$\Phi_{\mathbf{a}}(m, k) = \phi_A(m, k) \mathbf{I}_{N \times N}, \quad (21)$$

[0073] Como consecuencia de la Fórmula (20) se deduce que cuando se determinan dos matrices de las matrices $\Phi_{\mathbf{y}}(m, k)$, $\Phi_{\mathbf{d}}(m, k)$ y $\Phi_{\mathbf{a}}(m, k)$, entonces la tercera matriz surge inmediatamente. Además, en consecuencia, se deduce que es suficiente determinar únicamente:

- la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio, o

- la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de audio y la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio o

5

- la información de densidad espectral de potencia sobre las porciones de señal directa de las dos o más señales de canal de entrada de audio y la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio,

10 puesto que la información de densidad espectral de la tercera potencia (que no ha sido estimada) se torna inmediatamente evidente de la relación de los tres tipos de información de densidad espectral de potencia (por ejemplo, según la Fórmula (20) o según cualquier otra reformulación de la relación de los tres tipos de información de densidad espectral de potencia (PSD de la señal de entrada completa, PSD de los componentes ambiente y PSD de los componentes directos), cuando estos tres tipos de información PSD no están representados en forma de matrices, 15 sino cuando están disponibles en otro tipo de representación adecuada, por ejemplo, en forma de uno o más vectores o, por ejemplo, en forma de una pluralidad de coeficientes, etc.

[0074] Para evaluar la eficiencia del procedimiento ideado, se definen las siguientes señales:

20 • Distorsión de la señal directa:

$$\mathbf{q}_d(m, k) = [\mathbf{I} - \mathbf{H}_D(m, k)]^H \mathbf{d}(m, k),$$

• Señal ambiental residual:

25

$$\mathbf{r}_a(m, k) = \mathbf{H}_D^H(m, k) \mathbf{a}(m, k),$$

• Distorsión de la señal ambiental:

30

$$\mathbf{q}_a(m, k) = [\mathbf{I} - \mathbf{H}_A(m, k)]^H \mathbf{a}(m, k),$$

• Señal directa residual:

35

$$\mathbf{r}_d(m, k) = \mathbf{H}_A^H(m, k) \mathbf{d}(m, k),$$

[0075] A continuación se describe la derivación de las matrices de filtro según la figura 4 y según la figura 5. Para mayor facilidad de lectura se omiten los índices de subbanda y los índices de tiempo.

40 **[0076]** En primer lugar se describen formas de realización correspondientes a la estimación de los componentes de señal directa.

[0077] El fundamento del procedimiento ideado es calcular los filtros de tal manera que la señal residual ambiente \mathbf{r}_a se minimice y al mismo tiempo se limite la distorsión de la señal directa \mathbf{q}_d . Esto lleva al problema de 45 optimización restringida

$$\mathbf{H}_D(\beta_i) = \arg \min_{\mathbf{H}_D} E\{\|\mathbf{r}_a\|^2\} \quad (22)$$

50

$$\text{subject to } E\{\|\mathbf{q}_d\|^2\} \leq \sigma_{d,\max}^2,$$

sujeto a

donde $\sigma_{d,\max}^2$ es la distorsión máxima admisible de la señal directa. La solución está dada por

55

$$\mathbf{H}_D(\beta_i) = [\Phi_d + \beta_i \Phi_a]^{-1} \Phi_d. \quad (23)$$

[0078] El filtro para el cálculo de la señal directa de salida del i° canal es igual a

$$\mathbf{h}_{D,i}(\beta_i) = [\Phi_d + \beta_i \Phi_a]^{-1} \Phi_d \mathbf{u}_i. \quad (24)$$

5 donde \mathbf{u}_i es un vector nulo de longitud N con 1 en la i^{a} posición. El parámetro β_i habilita una compensación entre la
 10 reducción de señal ambiental residual y la distorsión de señal ambiental. En el caso del sistema ilustrado en la figura
 4, los niveles más bajos de ambiente residual en la señal de salida directa llevan a niveles ambientes más altos en las
 señales de salida ambiental. Una menor cantidad de distorsión de las señales directas lleva a una atenuación de los
 componentes de señal directa de las señales de salida ambiental. El parámetro dependiente del tiempo y la frecuencia
 β_i puede ser fijado de manera independiente por cada canal y puede ser controlado por las señales de entrada o las
 10 señales derivadas de las mismas, como se describe a continuación.

[0079] Cabe señalar que se puede obtener una solución similar formulando el problema de optimización restringida de la siguiente manera

15

$$\mathbf{H}_D(\beta_i) = \arg \min_{\mathbf{H}_D} E\{\|\mathbf{q}_d\|^2\} \quad (25)$$

$$\text{sujeto a } E\{\|\mathbf{r}_a\|^2\} \leq \sigma_{a,\max}^2,$$

Cuando Φ_d es de rango uno, la relación entre $\sigma_{d,\max}^2$ y β_i correspondiente a la señal del i° canal se deriva de la
 20 siguiente manera

$$\sigma_{d,\max}^2 = \left(\frac{\beta_i}{\beta_i + \lambda} \right)^2 \phi_{D_i D_i}. \quad (26)$$

donde $\phi_{D_i D_i}$ es la PSD de la señal directa del i° canal, y λ es la relación directa a ambiente de multicanal (DAR)

25

$$\lambda = \text{tr}\{\Phi_a^{-1} \Phi_d\} \quad (27)$$

$$= \text{tr}\{\Phi_a^{-1} \Phi_y\} - N, \quad (28)$$

30 donde la traza de una matriz cuadrada A es igual a la suma de los elementos situados en la diagonal principal, $\text{tr}\{\mathbf{K}\} =$

$$\sum_{i=1}^N k_{ii}(m, k).$$

[0080] Se debe tener en cuenta que la afirmación, de que Φ_d es de rango uno, es solo un supuesto. Sin
 importar si en realidad este supuesto es cierto o no, las formas de realización de la presente invención emplean las
 35 Fórmulas expuestas (26), (27) y (28), incluso en situaciones en que, en realidad, el resultado exacto de Φ_d es tal
 que Φ_d no es de rango uno. En esas situaciones, las formas de realización de la presente invención también dan
 buenos resultados incluso cuando el supuesto de que Φ_d es de rango uno en realidad no es verdad.

[0081] A continuación se describe una estimación de los componentes de señal ambiental.

40

[0082] El razonamiento detrás del procedimiento ideado es calcular los filtros de tal manera que la señal
 residual directa \mathbf{r}_d se minimiza y a la vez se restringe la distorsión de la señal ambiental \mathbf{q}_a . Esto lleva al problema de
 optimización restringida

45

$$\mathbf{H}_A(\beta_i) = \arg \min_{\mathbf{H}_A} E\{\|\mathbf{r}_d\|^2\} \quad (29)$$

$$\text{subject to } E\{\|\mathbf{q}_a\|^2\} \leq \sigma_{a,\max}^2,$$

50

sujeto a

donde $\sigma_{a,\max}^2$ es la distorsión máxima de la señal ambiental. La solución está dada por

$$\mathbf{H}_A(\beta_i) = [\beta_i \Phi_d + \Phi_a]^{-1} \Phi_a, \quad (30)$$

[0083] El filtro para calcular la señal ambiental de salida del i º canal es igual a

$$\mathbf{h}_{A,i}(\beta_i) = [\beta_i \Phi_d + \Phi_a]^{-1} \Phi_a \mathbf{u}_i. \quad (31)$$

[0084] A continuación, se proporcionan en detalle formas de realización que incorporan los conceptos de la presente invención.

[0085] Para determinar la información de densidad espectral de potencia, por ejemplo, se podría estimar la matriz PSD de las señales de canal de entrada de audio Φ_y directamente utilizando la obtención del promedio móvil de tiempo reducido o promedio recursivo. La matriz de PSD ambiente Φ_a , puede ser estimada de la manera descrita a continuación. Entonces se puede obtener la matriz PSD directa Φ_d , por ejemplo, utilizando la Fórmula (20).

[0086] A continuación se asume de nuevo que no hay más de una fuente de sonido directo activa por tiempo en cada subbanda (fuente directa única) y que, en consecuencia, Φ_d es de rango uno.

[0087] Se debe tener en cuenta que las afirmaciones de que no hay más de una fuente de sonido directo activa y que Φ_d es de rango uno son solo supuestos. Independientemente de si estos supuestos son ciertos o no, las formas de realización de la presente invención emplean las siguientes Fórmulas, en particular, las Fórmulas (32) y (33), incluso en situaciones en que, en realidad, hay más de una fuente de sonido directo activa e incluso cuando, en realidad, el resultado exacto de Φ_d es tal que Φ_d no es de rango uno. En esas situaciones, las formas de realización de la presente invención también ofrecen buenos resultados incluso cuando los supuestos de que no hay más de una fuente activa de sonido directo y que Φ_d es de rango uno no son ciertos en realidad.

[0088] Por consiguiente, suponiendo que no hay más de una fuente activa de sonido directo y que Φ_d es de rango uno, la Fórmula (23) se puede expresar de la siguiente manera

$$\mathbf{H}_D(\beta_i) = \frac{\Phi_a^{-1} \Phi_d}{\beta_i + \lambda} \quad (32)$$

$$= \frac{\Phi_a^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda}. \quad (33)$$

[0089] La Fórmula (33) ofrece una solución para el problema de optimización restringida de la Fórmula (22).

[0090] En las anteriores Fórmulas (32) y (33), Φ_a^{-1} es la matriz inversa de Φ_a . Es evidente que Φ_a^{-1} también indica la información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio.

[0091] Para determinar $\mathbf{H}_D(\beta_i)$, se tiene que determinar Φ_a^{-1} y Φ_d . Una vez que Φ_a está disponible, se puede determinar Φ_a^{-1} . λ se define según las Fórmulas (27) y (28) y su valor está disponible cuando Φ_a^{-1} y Φ_d están disponibles. Además de determinar Φ_a^{-1} , Φ_d y λ , se debe elegir un valor adecuado para β_i .

[0092] Asimismo, la Fórmula (33) puede ser reformulada (véase la Fórmula (20)), de tal manera que:

$$\mathbf{H}_D(\beta_i) = \frac{(\Phi_y - \Phi_d)^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33a)$$

y, por consiguiente, solo se tiene que determinar la información PSD Φ_y sobre las señales de canal de entrada de audio y la información PSD Φ_d sobre las porciones de señal directa de las señales de canal de entrada de audio.

[0093] Asimismo, la Fórmula (33) puede ser reformulada (véase la Fórmula (20)), de tal manera que:

5

$$\mathbf{H}_D(\beta_i) = \frac{\Phi_a^{-1}(\Phi_d + \Phi_a) - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33b)$$

y, por consiguiente, de manera que solo se tiene que determinar la información PSD Φ_a^{-1} sobre las porciones ambiente de señal de las señales de canal de entrada de audio y la información PSD Φ_d sobre las porciones de señal directa de las señales de canal de entrada de audio.

[0094] Además, la Fórmula (33) puede ser reformulada de tal manera que:

15

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{\Phi_a^{-1}\Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33c)$$

y, por consiguiente, de manera que se determine $\mathbf{H}_A^+(\beta_i)$.

20

[0095] La Fórmula (33c) proporciona una solución para el problema de optimización restringida de la Fórmula (29).

[0096] De manera similar, las Fórmulas (33a) y (33b) pueden ser reformuladas en:

25

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{(\Phi_y - \Phi_d)^{-1}\Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33d)$$

30

o:

35

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{\Phi_a^{-1}(\Phi_d + \Phi_a) - \mathbf{I}_{N \times N}}{\beta_i + \lambda} \quad (33e)$$

40

[0097] Se debe tener en cuenta que determinando $\mathbf{H}_D(\beta_i)$, se obtiene inmediatamente el filtro $\mathbf{H}_A(\beta_i)$ de la siguiente manera: $\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \mathbf{H}_D(\beta_i)$.

[0098] Además, se debe tener en cuenta que determinando $\mathbf{H}_D(\beta_i)$, se obtiene inmediatamente el filtro

45 $\mathbf{H}_D(\beta_i)$ de la siguiente manera: $\mathbf{H}_D(\beta_i) = \mathbf{I}_{N \times N} - \mathbf{H}_A(\beta_i)$.

[0099] Como se señala anteriormente, para determinar $\mathbf{H}_D(\beta_i)$, por ejemplo, según la Fórmula (33), se puede determinar Φ_y y Φ_a :

La matriz PSD de las señales de audio $\Phi_y(m, k)$ puede ser estimada directamente, por ejemplo, por obtención de promedio recursivo

50

$$\Phi_y(m, k) = (1 - \alpha) \mathbf{y}(m, k) \mathbf{y}^H(m, k) + \alpha \Phi_y(m - 1, k), \quad (34a)$$

donde α es el coeficiente de filtro que determina el tiempo de integración o por ejemplo, utilizando el promedio ponderado móvil de tiempo reducido

$$5 \quad \Phi_{\mathbf{y}}(m, k) = b_0 \cdot \mathbf{y}(m, k) \mathbf{y}^H(m, k) + b_1 \cdot \mathbf{y}(m-1, k) \mathbf{y}^H(m-1, k) \\ + b_2 \cdot \mathbf{y}(m-2, k) \mathbf{y}^H(m-2, k) + \dots + b_L \cdot \mathbf{y}(m-L, k) \mathbf{y}^H(m-L, k) \quad (34b)$$

donde L es, por ejemplo, el número de valores pasados utilizados para el cálculo de la PSD, y $b_0 \dots b_L$ son los coeficientes de filtro que están, por ejemplo, en el rango [0 1] (por ejemplo, $0 \leq$ coeficiente de filtro ≤ 1), o

por ejemplo, utilizando el promedio ponderado móvil de tiempo reducido, según la Ecuación (34b) pero con $b_i = \frac{1}{L+1}$ para todos los $i = 0 \dots L$.

15 **[0100]** Se describe ahora la estimación de la matriz PSD ambiente $\Phi_{\mathbf{a}}$ según las formas de realización.

[0101] La matriz PSD ambiente $\Phi_{\mathbf{a}}$ está dada por

$$20 \quad \Phi_{\mathbf{a}} = \hat{\phi}_A \mathbf{I}_{N \times N}, \quad (35)$$

donde $\mathbf{I}_{N \times N}$ es la matriz de identidad de tamaño $N \times N$. $\hat{\phi}_A$ es, por ejemplo, un número.

[0102] Una solución según una forma de realización se obtiene, por ejemplo, utilizando un valor constante, recurriendo a la Fórmula (21) y estableciendo $\hat{\phi}_A$ en una constante positiva real ϵ . La ventaja de esta estrategia es que la complejidad informática es ínfima.

[0103] En algunas formas de realización, la unidad de determinación de filtros 110 está configurada para determinar $\hat{\phi}_A$ dependiendo de las dos o más señales de canal de entrada de audio.

30 **[0104]** Una opción de muy baja complejidad informática consiste, según una forma de realización, en usar una fracción de la potencia de entrada y establecer $\hat{\phi}_A$ en el valor medio o el valor mínimo de la PSD de entrada o una fracción del mismo, por ejemplo

$$35 \quad \hat{\phi}_A = \frac{g}{N} \text{tr}\{\Phi_{\mathbf{y}}\}, \quad (36)$$

donde el parámetro g controla la cantidad de potencia del ambiente, y $0 < g < 1$.

40 **[0105]** Según otra forma de realización, se lleva a cabo una estimación basada en la media aritmética. Dado el supuesto que lleva a la Fórmula (20) y la Fórmula (21), se puede demostrar que se puede calcular la PSD $\hat{\phi}_A$ utilizando

$$45 \quad \hat{\phi}_A = \frac{1}{N} \text{tr}\{\Phi_{\mathbf{y}} - \Phi_{\mathbf{d}}\} \quad (37)$$

$$= \frac{1}{N} (\text{tr}\{\Phi_{\mathbf{y}}\} - \text{tr}\{\Phi_{\mathbf{d}}\}). \quad (38)$$

50 Aunque $\text{tr}\{\Phi_{\mathbf{y}}\}$ puede ser calculado directamente utilizando por ejemplo la integración recursiva de la Fórmula (34a), o, por ejemplo, se estima el promedio ponderado móvil de tiempo reducido de la Fórmula (34b), $\text{tr}\{\Phi_{\mathbf{d}}\}$ de la siguiente

manera

$$\text{tr}\{\Phi_d\} = \frac{1}{N-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^N [(\phi_{Y_i Y_i} - \phi_{Y_j Y_j})^2 \quad (39)$$

$$+ 4\text{Re}\{\phi_{Y_i Y_j}\}^2]^{\frac{1}{2}}. \quad (40)$$

[0106] Por otro lado, se puede calcular la PSD $\hat{\phi}_A(m, k)$ para $N > 2$ eligiendo dos señales de canal de entrada y estimando $\hat{\phi}_A(m, k)$ solo para un par de canales de las señales. Se obtienen resultados más precisos cuando se aplica este procedimiento a más de un par de señales de canal de entrada y combinando los resultados, por ejemplo promediando los cálculos estimativos totales. Se pueden elegir los subconjuntos aprovechando los canales con potencia ambiente similar, por ejemplo mediante la estimación de la potencia ambiente por separado en todos los canales posteriores y los canales anteriores de una grabación 5.1.

[0107] Asimismo, se debe tener en cuenta que de las Fórmulas (20) y (35), surge que

$$\Phi_d = \Phi_y - \hat{\phi}_A \mathbf{I}_{N \times N}. \quad (35a)$$

[0108] Según algunas formas de realización, se determina Φ_d mediante la determinación de $\hat{\phi}_A$ (por ejemplo, según la Fórmula (35), o la Fórmula (36) o según las Fórmulas (37) – (40)) y empleando la Fórmula (35a) para obtener la información de densidad espectral de potencia sobre las porciones de señal ambiente de las señales de canal de entrada de audio. A continuación, se puede determinar $\mathbf{H}_D(\beta_i)$, por ejemplo, empleando la Fórmula (33a).

[0109] A continuación, se considera la elección del parámetro β_i .

[0110] β_i es un parámetro de compensación. El parámetro de compensación β_i es un número.

[0111] En algunas formas de realización, solo se determina un parámetro de compensación β_i que es válido para todas las señales de canal de entrada de audio y, a continuación, se considera este parámetro de compensación como información de compensación de las señales de canal de entrada de audio.

[0112] En otras formas de realización, se determina un parámetro de compensación β_i para cada una de las dos o más señales de canal de entrada de audio y, a continuación, estos dos o más parámetros de compensación de las señales de canal de entrada de audio forman, juntos, la información de compensación.

[0113] En otras formas de realización, la información de compensación puede no estar representada en forma de parámetro sino que puede ser representada en un tipo diferente de formato adecuado.

[0114] Como se ha señalado anteriormente, el parámetro β_i habilita una compensación de compensación entre la reducción de la señal y la distorsión de la señal directa. Puede ser elegido para ser constante, o dependiente de la señal, como se ilustra en la figura 6b.

[0115] La figura 6b ilustra un aparato según otra forma de realización. El aparato comprende un banco de filtros de análisis 605 para transformar las señales de canal de entrada de audio $\mathbf{y}_i[n]$ del dominio de tiempo al dominio de tiempo-frecuencia. Asimismo, el aparato comprende un banco de filtros de síntesis 625 para transformar una o más señales de canal de salida de audio, (por ejemplo, los componentes de señal directa estimados $\hat{d}_1[n], \dots, \hat{d}_N[n]$ de las señales de canal de entrada de audio) del dominio de tiempo-frecuencia al dominio de tiempo.

[0116] Una pluralidad de K unidades de determinación de beta 1111, ..., 11K1 ("calcular Beta") determina los parámetros β_i . Asimismo, una pluralidad de K unidades de cálculo de subfiltros 1112, ..., 11K2 determina los subfiltros $\mathbf{H}_D^H(m, 1), \dots, \mathbf{H}_D^H(m, K)$. La pluralidad de unidades de determinación de beta 1111, ..., 11K1 y la pluralidad de unidades de cálculo de subfiltros 1112, ..., 11K2 juntas, forman la unidad de determinación de filtros 110 de la figura

1 y la figura 6a según una forma de realización específica. La pluralidad de subfiltros $\mathbf{H}_D^H(m,1), \dots, \mathbf{H}_D^H(m, K)$ juntos forman el filtro de la figura 1 y la figura 6a según una forma de realización particular.

5 **[0117]** Asimismo, la figura 6b ilustra una pluralidad de subprocesadores de señales 121, ..., 12K, en la que cada subprocesador de señales 121, ..., 12K está configurado para aplicar uno de los subfiltros $\mathbf{H}_D^H(m,1), \dots, \mathbf{H}_D^H(m, K)$ a una de las señales de canal de entrada de audio para obtener una de las señales de canal de salida de audio. La pluralidad de subprocesadores de señales 121, ..., 12K forman juntos el procesador de señales de la figura 1 y la figura 6a según una forma de realización particular.

10 **[0118]** En lo sucesivo se describen diferentes usos para controlar el parámetro β_i por medio de análisis de señales.

[0119] En primer lugar se consideran las señales transitorias.

15 **[0120]** Según una forma de realización, la unidad de determinación de filtros 110 está configurada para determinar la información de compensación (β_i, β_j) dependiendo de si hay un transitorio presente en al menos una de las dos o más señales de canal de entrada de audio.

20 **[0121]** La estimación de la matriz PSD de entrada da mejor resultado para una señal estacionaria. Por otro lado, la descomposición de la señal de entrada transitoria puede dar lugar a la fuga del componente de señal transitoria hacia la señal ambiental de salida. El control de β_i por medio de un análisis de señales con respecto al grado de probabilidad de presencia no estacionaria o transitoria de manera que β_i sea menor cuando la señal comprende transitorios y mayor en las porciones sostenidas lleva a señales de salida más constantes cuando se aplican los filtros $\mathbf{H}_D(\beta_i)$. El control de β_i por medio de un análisis de señales con respecto al grado de probabilidad de presencia no estacionaria o transitoria de tal manera que β_i sea mayor cuando la señal comprende transitorios y menor en las porciones sostenidas lleva a señales de salida más constantes cuando se aplican los filtros $\mathbf{H}_A(\beta_i)$.

[0122] Se consideran ahora las señales ambientales desfavorables.

30 **[0123]** En una forma de realización, la unidad de determinación de filtros 110 está configurada para determinar la información de compensación (β_i, β_j) dependiendo de la presencia de ruido aditivo en al menos un canal de señal a través del cual se transmite una de las dos o más señales de canal de entrada de audio.

35 **[0124]** El procedimiento propuesto descompone las señales de entrada independientemente del carácter de los componentes de señal ambiental. Cuando las señales de entrada se han transmitido por canales de señales ruidosas, es ventajoso estimar la probabilidad de la presencia de ruido aditivo perjudicial y controlar β_i de tal manera que la DAR de salida (relación directa a ambiente) se incremente.

[0125] Se describe ahora el control de los niveles de las señales de salida.

40 **[0126]** Para controlar los niveles de las señales de salida, se puede configurar β_i por separado para el i° canal. Los filtros para el cálculo de la señal ambiental de salida del i° canal están dados por la Fórmula (31).

45 **[0127]** Para el caso de cualquiera de dos canales, se puede calcular β_i dado β_j de tal manera que las PSD de las señales residuales ambientales $r_{a,i}$ y $r_{a,j}$ en el i° y j° canal de salida sean iguales, es decir,

$$\mathbf{h}_{A,i}^H(\beta_i) \Phi_a \mathbf{h}_{A,i}(\beta_i) = \mathbf{h}_{A,j}^H(\beta_j) \Phi_a \mathbf{h}_{A,j}(\beta_j). \quad (41)$$

o

50

$$\begin{aligned} & (\mathbf{u}_i - \mathbf{h}_{D,i}(\beta_i))^H \Phi_a (\mathbf{u}_i - \mathbf{h}_{D,i}(\beta_i)) \\ & = (\mathbf{u}_j - \mathbf{h}_{D,j}(\beta_j))^H \Phi_a (\mathbf{u}_j - \mathbf{h}_{D,j}(\beta_j)). \end{aligned} \quad (42)$$

[0128] Por otro lado, β_i puede ser calculado de tal manera que las PSD de las señales de salida ambiental \hat{a}_i y \hat{a}_j sean iguales para todos los pares de i y j .

55

[0129] Se considera ahora el uso de la información de paneo.

[0130] Para el caso de dos canales de entrada, la información de paneo cuantifica las diferencias de nivel entre ambos canales por cada subbanda. La información de paneo se puede aplicar para controlar β_i a fin de controlar el ancho percibido de las señales de salida.

5

[0131] A continuación se considera la ecualización de las señales del canal ambiental de salida.

[0132] El procesamiento descrito no garantiza que todas las señales de canal ambiental de salida tengan iguales potencias de subbanda. Para garantizar que todas las señales de canal ambiental de salida tengan iguales potencias de subbanda, se modifican los filtros según lo descrito a continuación con respecto a la forma de realización que utiliza los filtros \mathbf{H}_D antes descritos. La matriz de covarianza de la señal ambiental de salida (que comprende las auto-PSD de cada canal situado en la diagonal principal) se puede obtener de la siguiente manera

10

$$\Phi_{\hat{\mathbf{a}}} = (\mathbf{I} - \mathbf{H}_D)^H \Phi_{\mathbf{y}} (\mathbf{I} - \mathbf{H}_D). \quad (43)$$

15

[0133] Para garantizar que las PSD de todos los canales ambientes de salida sean iguales, se reemplazan los filtros \mathbf{H}_D por $\tilde{\mathbf{H}}_D$:

$$\tilde{\mathbf{H}}_D = \mathbf{I} - \mathbf{G}(\mathbf{I} - \mathbf{H}_D) = \mathbf{I} - \mathbf{G} + \mathbf{G}\mathbf{H}_D \quad (44)$$

20

donde \mathbf{G} es una matriz diagonal cuyos elementos situados en la diagonal principal son

$$g_{ii} = \sqrt{\frac{\text{tr}\{\Phi_{\hat{\mathbf{a}}}\}}{N\phi_{\hat{A}_i, \hat{A}_i}}}, \quad 1 \leq i \leq N. \quad (45)$$

25

[0134] En el caso de la forma de realización que utiliza los filtros \mathbf{H}_A antes descrita, se puede obtener la matriz de covarianza de la señal ambiental de salida (que comprende las auto-PSD de cada canal situado en la diagonal principal) se puede obtener de la siguiente manera

30

$$\Phi_{\hat{\mathbf{a}}} = \mathbf{H}_A^H \Phi_{\mathbf{y}} \mathbf{H}_A. \quad (46)$$

[0135] Para garantizar que las PSD de todos los canales ambiente de salida sean iguales, se reemplazan los filtros \mathbf{H}_A por $\tilde{\mathbf{H}}_A$:

35

$$\tilde{\mathbf{H}}_A = \mathbf{G}\mathbf{H}_A. \quad (47)$$

[0136] Aunque algunos aspectos han sido descritos en el contexto de un aparato, está claro que estos aspectos también representan una descripción del procedimiento correspondiente, en el cual un bloque o dispositivo corresponde a una etapa del procedimiento o a una característica de una etapa del procedimiento. De manera análoga, los aspectos descritos en el contexto de una etapa del procedimiento también representan una descripción de un bloque o elemento correspondiente o de una característica de un aparato correspondiente.

40

[0137] La señal descompuesta de la presente invención puede ser almacenada en un medio digital o se puede transmitir por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable tal como Internet.

45

[0138] Dependiendo de ciertos requisitos de implementación, las formas de realización de la invención pueden ser implementadas en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en el mismo señales de control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema informático programable de tal manera que se ejecute el procedimiento respectivo.

50

[0139] Algunas formas de realización comprenden un soporte de datos no transitorios que tiene señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se ejecute uno de los procedimientos descritos en esta invención.

55

[0140] En general, las formas de realización de la presente invención pueden ser implementadas en forma de producto de programa informático con un código de programa, donde el código de programa cumple la función de ejecutar uno de los procedimientos al ejecutarse el programa informático en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un soporte legible por una máquina.

5

[0141] Otras formas de realización comprenden el programa informático para ejecutar uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por una máquina.

[0142] En otras palabras, una forma de realización del procedimiento de la invención consiste, por lo tanto, en un programa informático que consta de un código de programa para realizar uno de los procedimientos descritos en esta invención al ejecutarse el programa informático en un ordenador.

[0143] Una forma de realización adicional de los procedimientos de la invención consiste, por lo tanto, en un soporte de datos (o medio de almacenamiento digital, o medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para ejecutar uno de los procedimientos descritos en esta invención.

[0144] Una forma de realización adicional del procedimiento de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para ejecutar uno de los procedimientos descritos en esta invención. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferidos a través de una conexión de comunicación de datos, por ejemplo a través de Internet.

[0145] Una forma de realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los procedimientos descritos en esta invención.

25

[0146] Una forma de realización adicional comprende un ordenador en el que se ha instalado el programa informático para ejecutar uno de los procedimientos descritos en esta invención.

[0147] En algunas formas de realización, se puede utilizar un dispositivo lógico programable (por ejemplo una matriz de puertas programables en el campo) para ejecutar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas formas de realización, una matriz de puertas programables en el campo puede cooperar con un microprocesador para ejecutar uno de los procedimientos descritos en esta invención. Por lo general, los procedimientos son ejecutados preferentemente por cualquier aparato de hardware.

[0148] Las formas de realización antes descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles descritos en esta invención serán evidentes para las personas expertas en la técnica. Por lo tanto, solo es intención limitarse al alcance de las siguientes reivindicaciones de patente y no a los detalles específicos presentados a modo de descripción y explicación de las formas de realización de esta invención.

40

Referencias:

[0149]

[1] J.B. Allen, D.A. Berkeley, y J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals", J.Acoust.Soc. Am., vol. 62, 1977.

[2] C. Avendano y J.-M. Jot, "A frequency-domain approach to multi-channel upmix", J. Audio Eng. Soc., vol. 52, 2004.

[3] C. Faller, "Multiple-loudspeaker playback of stereo signals", J. Audio Eng. Soc., vol. 54, 2006.

[4] J. Merimaa, M. Goodwin, y J.-M. Jot, "Correlation-based ambience extraction from stereo recordings", in Proc. of the AES 123a Conv., 2007.

[5] Ville Pulkki, "Directional audio coding in spatial sound reproduction and stereo upmixing", in Proc. of the AES 28th Int. Conf., 2006.

[6] J. Usher y J. Benesty, "Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer", IEEE Tram. on Audio, Speech, and Language Processing, vol.15, pp. 2141-2150, 2007.

60

[7] A. Walther y C. Faller, "Direct-ambient decomposition and upmix of surround sound signals", in Proc. of IEEE WASPAA, 2011.

[8] C. Uhle, J. Herre, S. Geyersberger, F. Ridderbusch, A. Walter; y O. Moser, "Apparatus y method for extracting an

ambient signal in an: apparatus y method for obtaining weighting coefficients for extracting an ambient signal y computer program”, Solicitud de Patente de Estados Unidos 2009/0080666, 2009.

[9] C. Uhle, J. Herre, A. Walther, O. Hellmuth, y C. Janssen, “Apparatus and method for generating an ambient signal
5 from an audio signal, apparatus y method for deriving a multi-channel audio signal from an audio signal and computer program”, Solicitud de Patente de Estados Unidos 2010/0030563, 2010.

[10] G. Soulodre, “System for extracting and changing the reverberant content of an audio input signal”, Patente de Estados Unidos 8,036,767, Fecha de presentación de la Patente: 11 de octubre de 2011.

10

REIVINDICACIONES

1. Un aparato para generar una o más señales de canal de salida de audio dependiendo de dos o más señales de canal de entrada de audio, en el que cada una de las dos o más señales de canal de entrada de audio comprende porciones de señal directa y porciones de señal ambiental, donde el aparato comprende:
- una unidad de determinación de filtro (110) para determinar un filtro mediante la estimación de primera información de densidad espectral de potencia y mediante la segunda estimación de información de densidad espectral de potencia, en la que el filtro depende de la primera información de densidad espectral de potencia y de la segunda información de densidad espectral de potencia, y
- un procesador de señales (120) para generar una o más señales de canal de salida de audio mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio, en el que una o más señales de canal de salida de audio depende del filtro,
- en el que la unidad de determinación de filtro (110) está configurada para estimar la primera información de densidad espectral de potencia mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, información de densidad espectral de potencia sobre dicha señal de canal de entrada de audio y la unidad de determinación de filtro (110) está configurada para estimar la segunda información de densidad espectral de potencia mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, información de densidad espectral de potencia sobre las porciones de señal ambiental de dicha señal de canal de entrada de audio, o
- en el que la unidad de determinación de filtro (110) está configurada para estimar la primera información de densidad espectral de potencia mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, información de densidad espectral de potencia en dicha señal de canal de entrada de audio, y la unidad de determinación de filtro (110) está configurada para estimar la segunda información de densidad espectral de potencia mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, información de densidad espectral de potencia en las porciones de señal directa de dicha señal de canal de entrada de audio, o
- en el que la unidad de determinación de filtro (110) está configurada para estimar la primera información de densidad espectral de potencia mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, información de densidad espectral de potencia en las porciones de señal directa de dicha señal de canal de entrada de audio, y la unidad de determinación de filtro (110) está configurada para estimar la segunda información de densidad espectral de potencia mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, información de densidad espectral de potencia en las porciones de señal ambiental de dicha señal de canal de entrada de audio.
2. Un aparato según la reivindicación 1, en el que el aparato comprende además un banco de filtros de análisis (605) para transformar las dos o más señales de canal de entrada de audio del dominio de tiempo a un dominio de tiempo-frecuencia, en el que la unidad de determinación de filtro (110) está configurada para determinar el filtro mediante la estimación de la primera información de densidad espectral de potencia y la segunda información de densidad espectral de potencia dependiendo de las señales de canal de entrada de audio, que está representada en el dominio de tiempo-frecuencia, en el que el procesador de señales (120) está configurado para generar una o más señales de canal de salida de audio, que están representadas en un dominio de tiempo-frecuencia, mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio, que está representada en el dominio de tiempo-frecuencia, y en el que el aparato comprende además un banco de filtros de síntesis (625) para transformar una o más señales de canal de salida de audio, que están representadas en un dominio de tiempo-frecuencia, del dominio de tiempo-frecuencia al dominio de tiempo.
3. Un aparato según la reivindicación 1 o 2, en el que la unidad de determinación de filtro (110) está configurada para determinar el filtro mediante la estimación de la primera información de densidad espectral de potencia, mediante la estimación de la segunda información de densidad espectral de potencia, y mediante la determinación de una información de señal de canal de entrada de audio (β_i, β_j) dependiendo de al menos una de las dos o más señales de canal de entrada de audio.
4. Un aparato según la reivindicación 3, en el que la unidad de determinación de filtro (110) está configurada para determinar la información de señal de canal de entrada de audio (β_i, β_j) dependiendo de la presencia o no de un transitorio en al menos una de las dos o más señales de canal de entrada de audio.
5. Un aparato según la reivindicación 3 o 4, en el que la unidad de determinación de filtro (110) está configurada para determinar la información de señal de canal de entrada de audio (β_i, β_j) dependiendo de la presencia de ruido aditivo en al menos un canal de señal a través del cual se transmite una de las dos o más señales de canal de entrada de audio.

6. Un aparato según una de las reivindicaciones 3 a 5,
 en el que la unidad de determinación de filtro (110) está configurada para determinar la información de densidad
 espectral de potencia sobre las dos o más señales de canal de entrada de audio dependiendo de una primera matriz
 (Φ_y), comprendiendo la primera matriz (Φ_y) una estimación de la densidad espectral de potencia correspondiente
 5 a cada señal de canal de las dos o más señales de canal de entrada de audio ubicadas en la diagonal principal de la
 primera matriz (Φ_y), y está configurada para determinar la información de densidad espectral de potencia sobre las
 porciones de señal ambiental de las dos o más señales de canal de entrada de audio dependiendo de la segunda
 matriz (Φ_a) o dependiendo de una matriz inversa (Φ_a^{-1}) de la segunda matriz (Φ_a), comprendiendo la segunda
 matriz (Φ_a) una estimación de la densidad espectral de potencia correspondiente a las porciones de señal ambiental
 10 de cada señal de canal de las dos o más señales de canal de entrada de audio ubicadas en la diagonal principal de la
 segunda matriz (Φ_a), o
 en el que la unidad de determinación de filtro (110) está configurada para determinar la información de densidad
 espectral de potencia sobre las dos o más señales de canal de entrada de audio dependiendo de la primera matriz (Φ_y), y está configurada para determinar la información de densidad espectral de potencia sobre las porciones de
 15 señal directa de las dos o más señales de canal de entrada de audio dependiendo de una tercera matriz (Φ_d) o
 dependiendo de una matriz inversa (Φ_d^{-1}) de la tercera matriz (Φ_d), comprendiendo la tercera matriz (Φ_d) una
 estimación de la densidad espectral de potencia correspondiente a las porciones de señal directa de cada canal de
 las dos o más señales de canal de entrada de audio en la diagonal principal de la tercera matriz (Φ_d), o
 en el que la unidad de determinación de filtro (110) está configurada para determinar la información de densidad
 20 espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio
 dependiendo de la segunda matriz (Φ_a) o dependiendo de una matriz inversa (Φ_a^{-1}) de la segunda matriz (Φ_a), y
 está configurada para determinar la información de densidad espectral de potencia sobre las porciones de señal directa
 de las dos o más señales de canal de entrada de audio dependiendo de la tercera matriz (Φ_d) o dependiendo de
 una matriz inversa (Φ_d^{-1}) de la tercera matriz (Φ_d).
- 25
7. Un aparato según la reivindicación 6,
 en el que la unidad de determinación de filtro (110) está configurada para determinar la primera matriz (Φ_y) para
 determinar la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de
 audio, y está configurada para determinar la segunda matriz (Φ_a) o una matriz inversa (Φ_a^{-1}) de la segunda matriz
 30 (Φ_a) para determinar la información de densidad espectral de potencia sobre las porciones de señal ambiental de
 las dos o más señales de canal de entrada de audio, o
 en el que la unidad de determinación de filtro (110) está configurada para determinar la primera matriz (Φ_y) para
 determinar la información de densidad espectral de potencia sobre las dos o más señales de canal de entrada de
 audio, y está configurada para determinar la tercera matriz (Φ_d) o una matriz inversa (Φ_d^{-1}) de la tercera matriz (Φ_d)
 35 (Φ_d) para determinar la información de densidad espectral de potencia sobre las porciones de señal directa de las
 dos o más señales de canal de entrada de audio, o
 en el que la unidad de determinación de filtro (110) está configurada para determinar la segunda matriz (Φ_a) o una
 matriz inversa (Φ_a^{-1}) de la segunda matriz (Φ_a) para determinar la información de densidad espectral de potencia
 sobre las porciones de señal ambiental de las dos o más señales de canal de entrada de audio, y está configurada
 40 para determinar la tercera matriz (Φ_d) o una matriz inversa (Φ_d^{-1}) de la tercera matriz (Φ_d) para determinar la
 información de densidad espectral de potencia sobre las porciones de señal ambiental de las dos o más señales de
 canal de entrada de audio.

8. Un aparato según la reivindicación 6 o 7,

en el que la unidad de determinación de filtro (110) está configurada para determinar el filtro $\mathbf{H}_D(\beta_i)$ dependiendo de la fórmula

5

$$\mathbf{H}_D(\beta_i) = \frac{\Phi_a^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda}$$

o dependiendo de la fórmula

10

$$\mathbf{H}_D(\beta_i) = \frac{(\Phi_y - \Phi_d)^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda}$$

15

o dependiendo de la fórmula

20

$$\mathbf{H}_D(\beta_i) = \frac{\Phi_a^{-1}(\Phi_d + \Phi_a) - \mathbf{I}_{N \times N}}{\beta_i + \lambda}$$

25

o en el que la unidad de determinación de filtro (110) está configurada para determinar el filtro $\mathbf{H}_A(\beta_i)$ dependiendo de la fórmula

30

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{\Phi_a^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda}$$

35

o dependiendo de la fórmula

40

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{(\Phi_y - \Phi_d)^{-1} \Phi_y - \mathbf{I}_{N \times N}}{\beta_i + \lambda}$$

o dependiendo de la fórmula

45

$$\mathbf{H}_A(\beta_i) = \mathbf{I}_{N \times N} - \frac{\Phi_a^{-1}(\Phi_d + \Phi_a) - \mathbf{I}_{N \times N}}{\beta_i + \lambda}$$

50

en la que Φ_y es la primera matriz,

en la que Φ_a es la segunda matriz,

en la que Φ_a^{-1} es la matriz inversa de la segunda matriz,

en la que Φ_d es la tercera matriz,

55 en la que $\mathbf{I}_{N \times N}$ es una matriz unitaria del tamaño $N \times N$,

en la que N indica el número de señales de canal de entrada de audio,

en la que β_i es la información de señal de canal de entrada de audio que es un número, y

en la que $\lambda = \text{tr}\{\Phi_a^{-1}\Phi_d\}$,
 en la que tr es el operador de traza.

9. Un aparato según una de las reivindicaciones 3 a 8, en el que la unidad de determinación de filtro (110) está configurada para determinar un parámetro de señal de canal de entrada (β_i, β_j) para cada una de las dos o más señales de canal de entrada de audio como información de señal de canal de entrada de audio (β_i, β_j), en el que el parámetro de señal de canal de entrada (β_i, β_j) de cada una de las señales de canal de entrada de audio depende de dicha señal de canal de entrada de audio.

10. Un aparato según la reivindicación 8, en el que la unidad de determinación de filtro (110) está configurada para determinar un parámetro de señal de canal de entrada (β_i, β_j) para cada una de dos o más señales de canal de entrada de audio como información de señal de canal de entrada de audio (β_i, β_j), de tal manera que, por cada par de una primera señal de canal de entrada de audio de las señales de canal de entrada de audio y otra segunda señal de canal de entrada de audio de las señales de canal de entrada de audio

$$\mathbf{h}_{A,i}^H(\beta_i)\Phi_a\mathbf{h}_{A,i}(\beta_i) = \mathbf{h}_{A,j}^H(\beta_j)\Phi_a\mathbf{h}_{A,j}(\beta_j)$$

es cierto,

en la que β_i es el parámetro de señal de canal de entrada de dicha primera señal de canal de entrada de audio, en la que β_j es el parámetro de señal de canal de entrada de dicha segunda señal de canal de entrada de audio, en el que

$$\mathbf{h}_{A,i}(\beta_i) = [\beta_i\Phi_d + \Phi_a]^{-1}\Phi_a\mathbf{u}_i,$$

en la que $\mathbf{h}_{A,i}^H(\beta_i)$ es la matriz de transposición conjugada de $\mathbf{h}_{A,i}(\beta_i)$, y en la que \mathbf{u}_i es un vector nulo de longitud N con 1 en la i^{a} posición.

11. Un aparato según la reivindicación 8 o 10, en el que la unidad de determinación de filtro (110) está configurada para determinar la segunda matriz Φ_a según la fórmula

$$\Phi_a = \hat{\phi}_A \mathbf{I}_{N \times N},$$

en la que la unidad de determinación de filtro (110) está configurada para determinar la tercera matriz Φ_d según la fórmula

$$\Phi_d = \Phi_y - \hat{\phi}_A \mathbf{I}_{N \times N},$$

en la que $\hat{\phi}_A$ es un número.

12. Un aparato según la reivindicación 11, en el que la unidad de determinación de filtro (110) está configurada para determinar $\hat{\phi}_A$ dependiendo de las dos o más señales de canal de entrada de audio.

13. Un aparato según una de las reivindicaciones 1 a 7, en el que la unidad de determinación de filtro (110) está configurada para determinar una matriz de filtros intermedia \mathbf{H}_D para proporcionar una estimación de componentes de señal directa de las dos o más señales de canal de entrada de audio mediante la estimación de una primera información de densidad espectral de potencia y mediante la estimación de una segunda información de densidad espectral de potencia, y

en el que la unidad de determinación de filtro (110) está configurada para determinar el filtro $\tilde{\mathbf{H}}_D$ dependiendo de la matriz de filtros intermedia \mathbf{H}_D según la fórmula

$$\tilde{\mathbf{H}}_D = \mathbf{I} - \mathbf{G} + \mathbf{G}\mathbf{H}_D,$$

- en la que \mathbf{I} es una matriz unitaria, y
 5 en la que \mathbf{G} es una matriz diagonal,
 en la que el procesador de señales (120) está configurado para generar una o más señales de canal de salida de audio mediante la aplicación del filtro $\tilde{\mathbf{H}}_D$ a las dos o más señales de canal de entrada de audio.
14. Un procedimiento para generar una o más señales de canal de salida de audio dependiendo de dos o más señales de canal de entrada de audio, en el que cada una de las dos o más señales de canal de entrada de audio comprende porciones de señal directa y porciones de señal ambiental, en el que el procedimiento comprende:
- la determinación de un filtro mediante la estimación de una primera información de densidad espectral de potencia y mediante la estimación de una segunda información de densidad espectral de potencia, en el que el filtro depende de
 15 la primera información de densidad espectral de potencia y de la segunda información de densidad espectral de potencia, y
 la generación de una o más señales de canal de salida de audio mediante la aplicación del filtro a las dos o más señales de canal de entrada de audio, en el que una o más señales de canal de salida de audio dependen del filtro, en el que la estimación de la primera información de densidad espectral de potencia se realiza mediante la estimación,
 20 para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, de la información de densidad espectral de potencia en dicha señal de canal de entrada de audio, y la estimación de la segunda información de densidad espectral de potencia se realiza mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, de la información de densidad espectral de potencia en porciones de señal ambiental de dicha señal de canal de entrada de audio, o
 25 en el que la estimación de la primera información de densidad espectral de potencia se realiza mediante la estimación, para cada señal del canal de entrada de audio de las dos o más señales del canal de entrada de audio, de la información de densidad espectral de potencia en dicha señal del canal de entrada de audio, y la estimación de la segunda información de densidad espectral de potencia se realiza mediante la estimación, para cada señal de canal de entrada de audio de los dos o más audio señales del canal de entrada, de la información de densidad espectral de
 30 potencia en las porciones de señal directa de dicha señal de canal de entrada de audio, o
 en el que la estimación de la primera información de densidad espectral de potencia se realiza mediante la estimación, para cada señal de canal de entrada de audio de las dos o más señales de canal de entrada de audio, de la información de densidad espectral de potencia en las porciones de señal directa de dicha señal de canal de entrada de audio y la estimación de la segunda información de densidad espectral de potencia se realiza mediante la estimación, para cada
 35 señal del canal de entrada de audio de las dos o más señales del canal de entrada de audio, de la información de densidad espectral de potencia en las porciones de señal ambiental de dicha señal del canal de entrada de audio.
15. Un programa informático para implementar el procedimiento de la reivindicación 14 cuando se ejecuta en un ordenador o un procesador.
 40

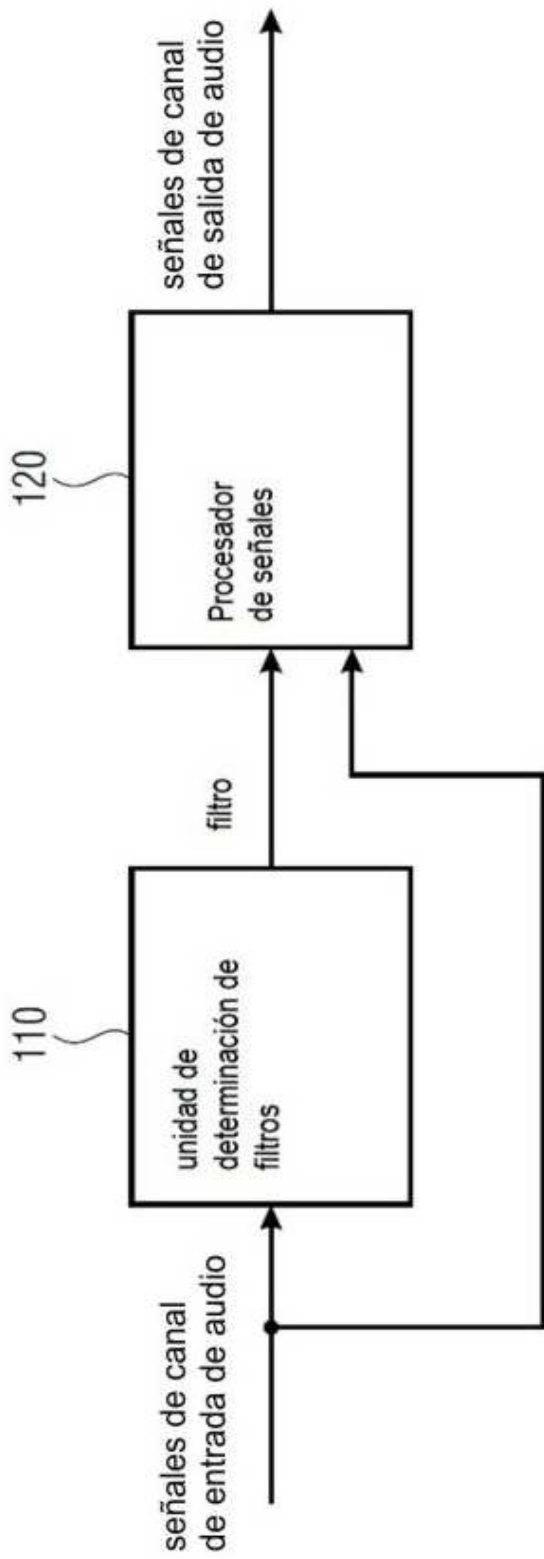


FIG 1

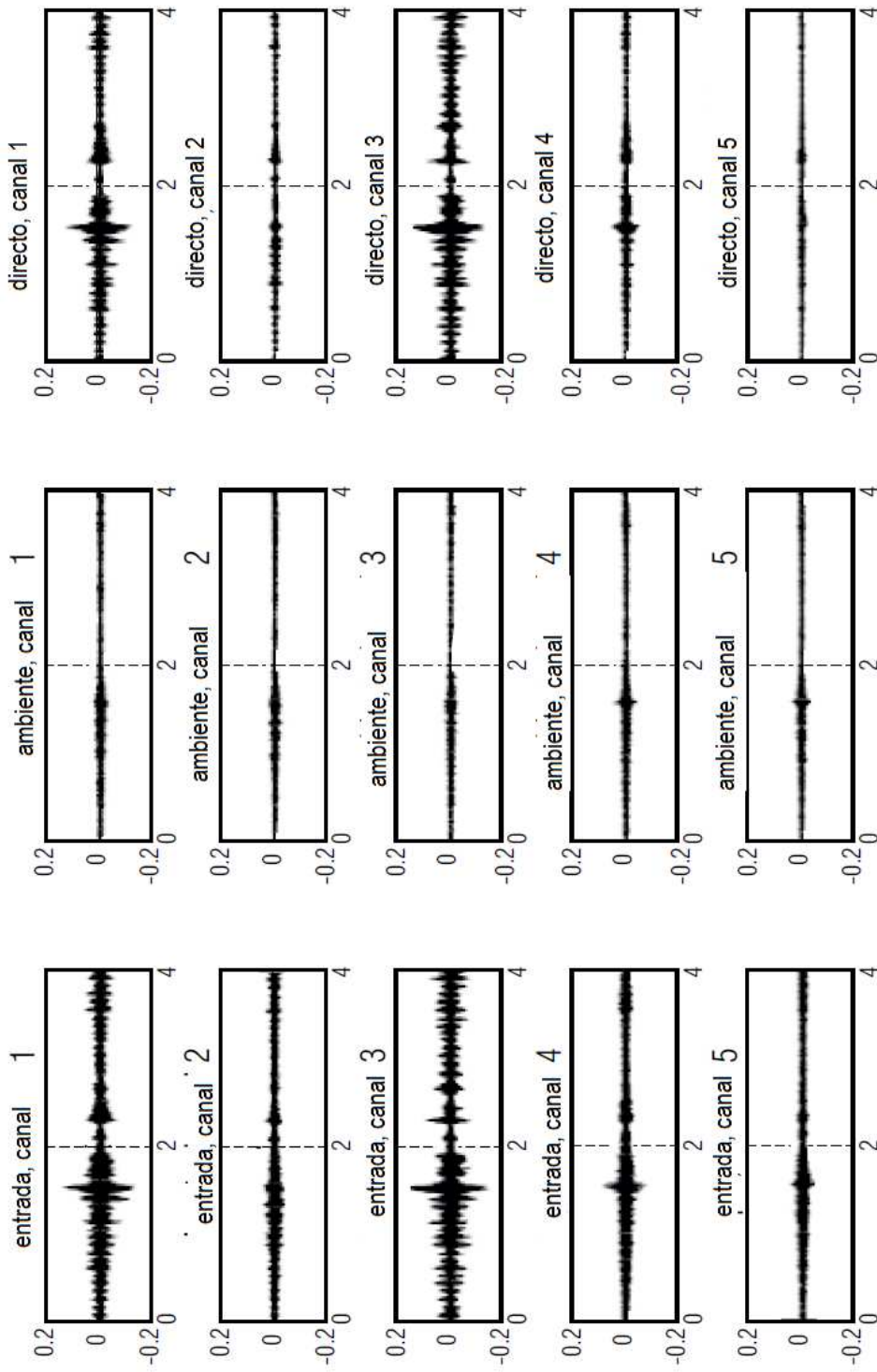


FIG 2

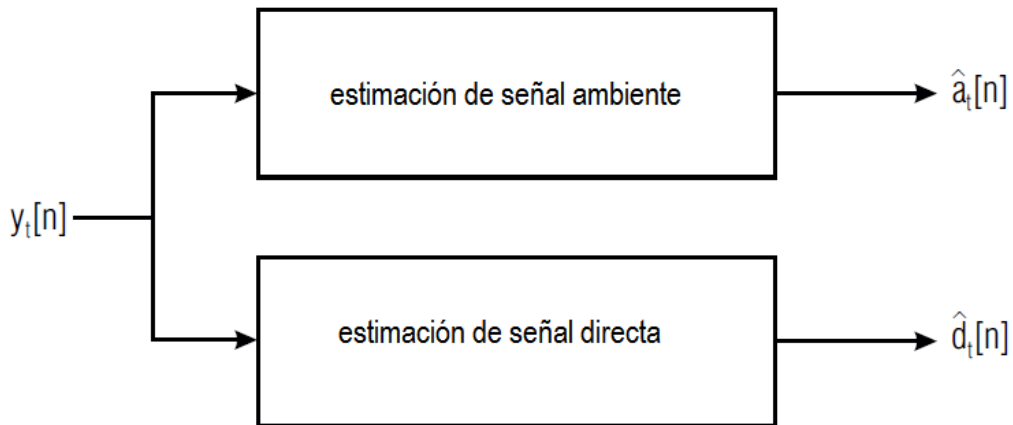


FIG 3

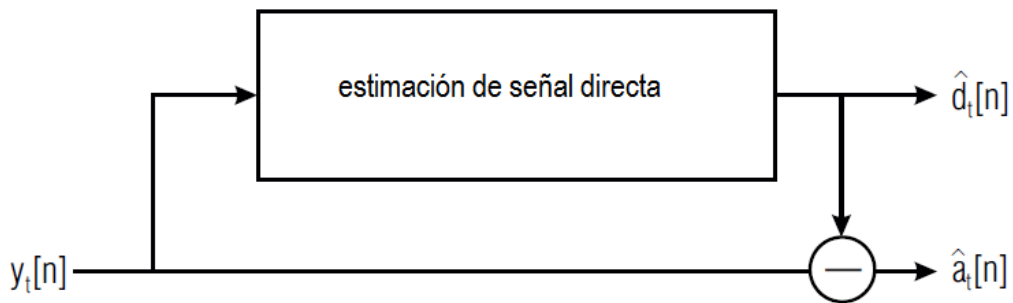


FIG 4

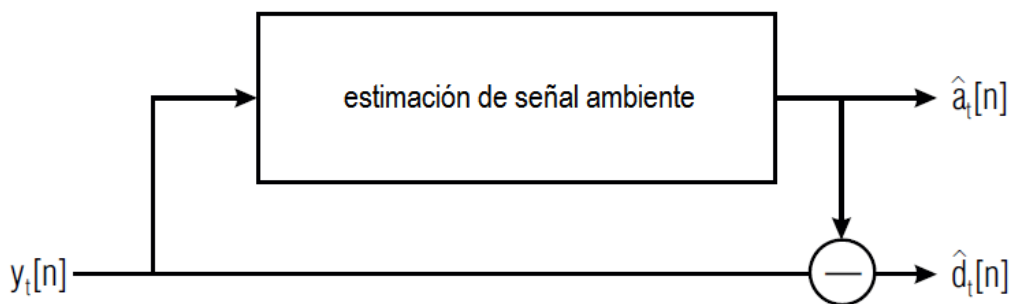


FIG 5

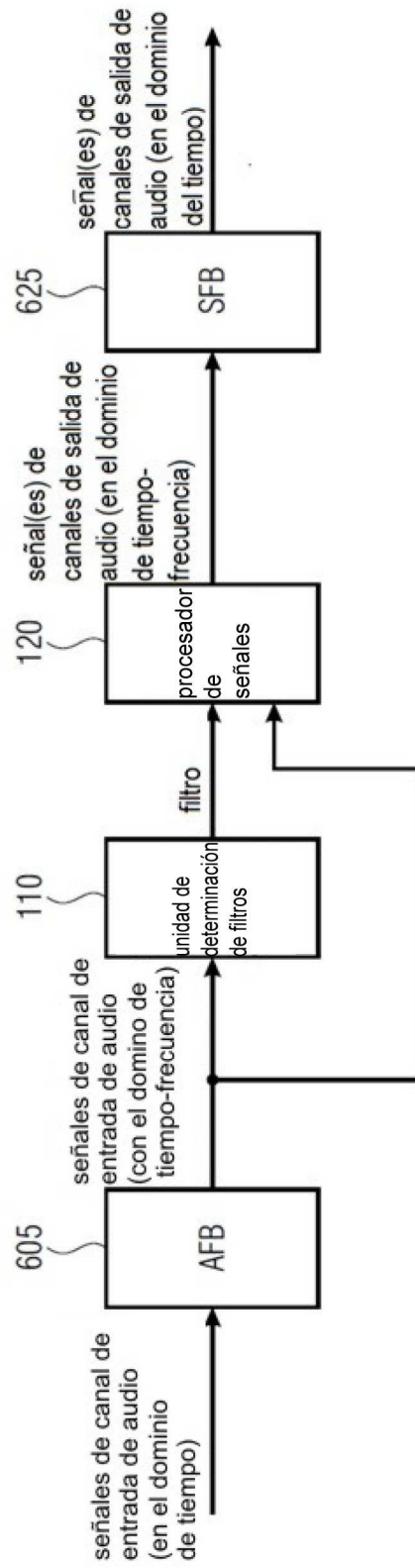


FIG 6A

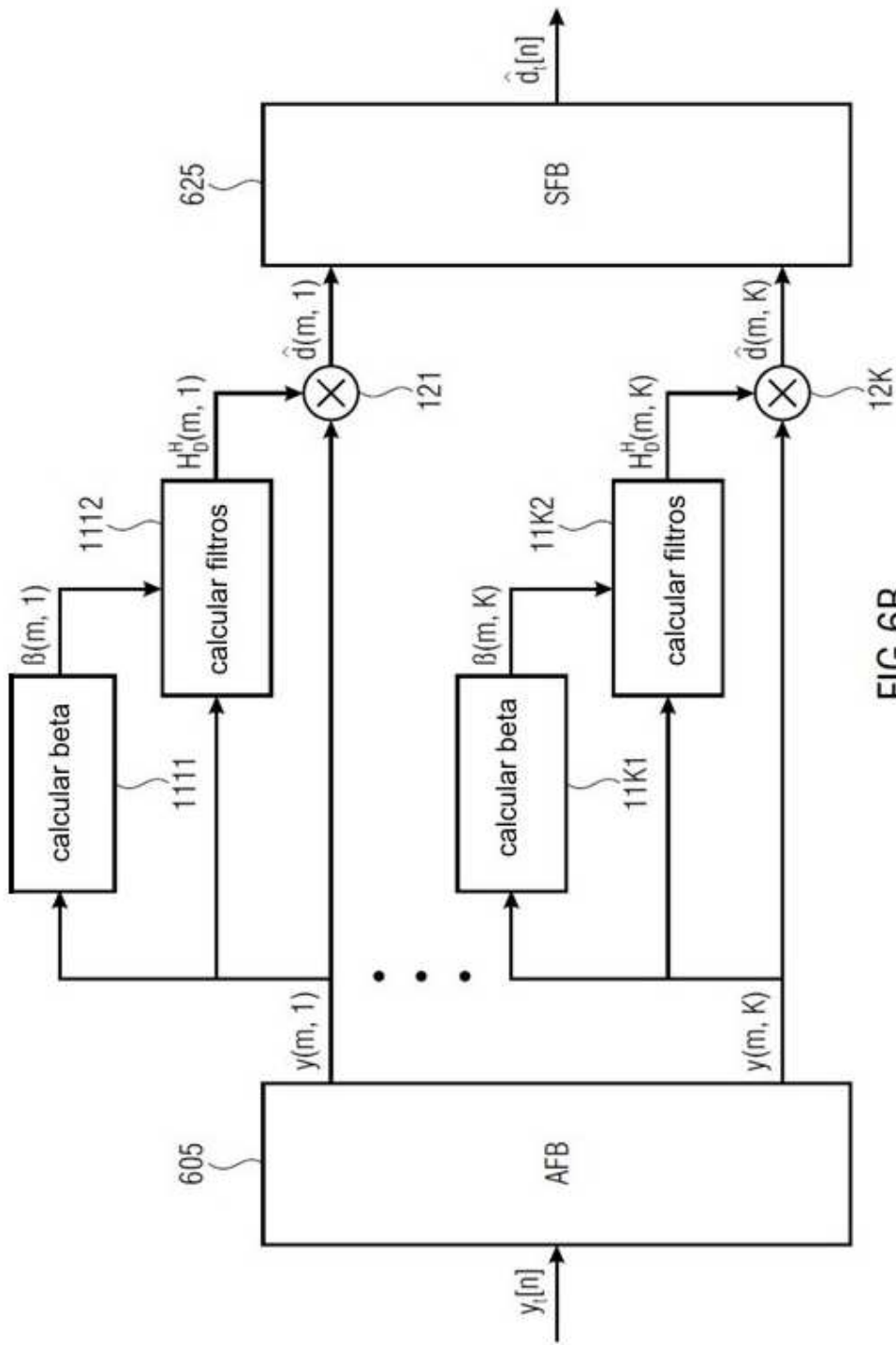


FIG 6B