

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 530 960**

51 Int. Cl.:

H04S 3/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.11.2011 E 11787858 (7)**

97 Fecha y número de publicación de la concesión europea: **14.01.2015 EP 2649814**

54 Título: **Aparato y método para descomponer una señal de entrada utilizando un mezclador descendente**

30 Prioridad:

10.12.2010 US 421927 P
11.05.2011 EP 11165742

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
09.03.2015

73 Titular/es:

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

72 Inventor/es:

WALTHER, ANDREAS

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 530 960 T3

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

APARATO Y MÉTODO PARA DESCOMPONER UNA SEÑAL DE ENTRADA UTILIZANDO UN MEZCLADOR
DESCENDENTE

DESCRIPCIÓN

5

La presente invención se relaciona con el procesamiento de audio y, en particular, con la descomposición de una señal de audio en diferentes componentes tales como componentes perceptualmente distintos.

10

El sistema auditivo humano detecta el sonido de todas las direcciones. El entorno auditivo percibido (el adjetivo auditivo indica lo que se percibe, en tanto que la palabra sonido se utiliza para describir fenómenos físicos) genera una impresión de las propiedades acústicas del espacio circundante y los eventos de sonido que se suscitan. La impresión auditiva percibida en un campo sonoro específico (por lo menos parcialmente) se puede modelar considerando tres tipos diferentes de señales en las entradas al oído: El sonido *directo*, *las reflexiones tempranas* y *las reflexiones difusas*. Estas señales contribuyen a la formación de la imagen espacial auditiva percibida.

15

El sonido directo denota las ondas de cada evento de sonido que llega al oyente en primer lugar directamente desde una fuente de sonido sin alteraciones. Es característico de la fuente de sonido y ofrece la información menos comprometida acerca de la dirección de incidencia del evento de sonido. Los indicios primarios para estimar la dirección de una fuente de sonido en el plano horizontal son las diferencias entre las señales de entrada del oído izquierdo y derecho, es decir las *diferencias interaurales de tiempo* (ITDs) y las *diferencias interaurales de intensidad* (ILDs). Seguidamente, una multitud de reflexiones del sonido directo llegan a los oídos de diferentes direcciones y con diferentes retardos de tiempo e intensidades relativas. Con el aumento del retardo de tiempo, con respecto al sonido directo, la densidad de las reflexiones aumenta hasta que constituyen un desorden estadístico.

20

El sonido reflejado contribuye a la percepción de la distancia y a la impresión auditiva espacial, que está compuesta por al menos dos componentes: *el ancho aparente de la fuente* (ASW) (Otro término utilizado comúnmente para ASW es la espaciosidad auditiva) y la sensación de sonido envolvente del oyente (LEV). El ASW se define como ensanchamiento del ancho aparente de una fuente de sonido y se determina fundamentalmente por las reflexiones laterales tempranas. LEV se refiere a la sensación del oyente de estar envuelto por el sonido y se determina fundamentalmente por las reflexiones que llegan tardíamente. La meta de la reproducción de sonido estereofónico electroacústico consiste en evocar la percepción de una imagen espacial auditiva agradable. Esta puede tener una referencia natural o arquitectónica (por ej. la grabación de un concierto en una sala) o puede ser un campo sonoro que no existe en la realidad (por ej. música electroacústica).

25

Desde el campo de la acústica de salas de concierto, se sabe que – para obtener un campo sonoro subjetivamente agradable – una fuerte sensación de impresión auditiva espacial es importante, donde la LEV constituye una parte integral. La capacidad de las disposiciones de parlantes para reproducir un campo sonoro envolvente por medio de la reproducción de un campo sonoro difuso es un tema de interés. En un campo de sonido sintético no es posible reproducir todas las reflexiones de origen natural utilizando transductores especializados. Esto es especialmente cierto en el caso de las reflexiones tardías difusas. Las propiedades de oportunidad e intensidad de las reflexiones difusas se pueden simular empleando señales “reverberadas” como alimentaciones de los parlantes. Si ellas están suficientemente no relacionadas, el número y ubicación de los parlantes utilizados para la reproducción determina si el campo de sonido se percibe como difuso. La meta consiste en evocar la percepción de un campo sonoro continuo y difuso utilizando sólo un número discreto de transductores. En otras palabras, la generación de campos sonoros donde no se puede estimar la dirección de llegada del sonido y, especialmente donde no se puede localizar ningún transductor individual. La difusividad subjetiva de los campos sonoros sintéticos puede ser evaluada en pruebas subjetivas.

35

Las reproducciones de sonido estereofónico apuntan a evocar la percepción de un campo sonoro continuo utilizando sólo un número discreto de transductores. Las características más deseables son la estabilidad direccional de las fuentes localizadas y la renderización realista del entorno auditivo circundante. La mayoría de los formatos utilizados hoy en día para almacenar o transportar grabaciones estereofónicas se basan en canales. Cada canal transporta una señal que se pretende reproducir a través de un parlante asociado ubicado en una posición específica. Se diseña una imagen auditiva específica durante el proceso de grabación o mezcla. Esta imagen se recrea con precisión si la configuración de parlantes utilizada para la reproducción se asemeja a la configuración objetivo para la cual la grabación estuviera destinada.

40

El número de canales factibles de transmisión y reproducción crece constantemente y con cada formato de reproducción que surge llega el deseo de renderizar el contenido de formato heredado por el sistema de reproducción real. Los algoritmos de mezcla ascendente son una solución a esta necesidad, computando una señal con más canales a partir de una señal heredada. Se ha propuesto un número de algoritmos de mezcla ascendente estéreo en la literatura, por ej. Carlos Avendano y Jean-Marc Jot, “A frequency-domain approach to multichannel upmix”, *Journal of the Audio Engineering Society*, vol. 52, no. 7/8, pp. 740–749, 2004; Christof Faller, “Multiple-loudspeaker playback of stereo signals”, *Journal of the Audio Engineering Society*, vol. 54, no. 11, pp. 1051–1064, noviembre de 2006; John Usherand Jacob Benesty, “Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 7, pp. 2141–2150, septiembre de 2007. La mayoría de estos algoritmos se basan en una descomposición de señal directa/ambiental seguida por la renderización adaptada a la configuración de parlantes objetivo.

55

60

Las descomposiciones de señales directas/ambientales descritas no son fácilmente aplicables a las señales envolventes de canales múltiples. No es fácil formular un modelo de señal y filtrado para obtener, de N canales de

audio, los correspondientes N sonidos directos y N canales de sonido ambiental. El modelo de señal simple utilizada en el caso del estéreo, ver por ej. Christof Faller, "Multiple-loudspeaker playback of stereo signals," *Journal of the Audio Engineering Society*, vol. 54, no. 11, pp. 1051–1064, noviembre de 2006, suponiendo que el sonido directo se correlaciona entre todos los canales, no captura la diversidad de las relaciones de canales que pueden existir entre canales de señal envolvente.

El objetivo general de la reproducción de sonido estereofónico consiste en evocar la percepción de un campo sonoro continuo utilizando sólo un número limitado de canales de transmisión y transductores. Dos parlantes son el requisito mínimo para la reproducción de sonido espacial. Los sistemas modernos del consumidor con frecuencia ofrecen un gran número de canales de reproducción. Básicamente, las señales estereofónicas (independientemente del número de canales) se graban o mezclan de tal manera que por cada fuente el sonido directo se dirige de manera coherente (=dependiente) a un número de canales con pases de señal específicos direccionales y los sonidos reflejados *independientes* se dirigen a un número de canales que determinan los pases de señal correspondientes al ancho aparente de la fuente y a la envolvente del oyente. La percepción correcta de la imagen auditiva pretendida habitualmente sólo es posible en el punto de observación ideal en la configuración de reproducción a la cual estaba destinada la grabación. La adición de más parlantes a una determinada configuración de parlantes habitualmente habilita una reconstrucción/ simulación más realista de un campo sonoro natural. Para aprovechar la ventaja plena de una configuración de parlantes extendida si las señales de entrada se dan en otro formato, o para manipular las partes perceptualmente distintas de la señal de entrada, se debe poder acceder a ellas por separado. A continuación, esta memoria descriptiva presenta un método para separar los componentes dependientes e independientes de las grabaciones estereofónicas que comprenden un número arbitrario de canales de entrada.

Una descomposición de las señales de audio en componentes perceptualmente diferenciados es necesaria para la modificación, realce, reproducción adaptativa y codificación perceptual de alta calidad de las señales de audio. Últimamente se ha propuesto un número de métodos que permiten la manipulación y/o extracción de componentes de señal perceptualmente diferenciados de señales de entrada de dos canales. Dado que las señales de entrada con más de dos canales se están tornando cada vez más comunes, las manipulaciones descritas son convenientes sólo para señales de entrada multicanales. Sin embargo, la mayoría de los conceptos descritos con respecto a la entrada de dos canales pueden no hacerse extensivos fácilmente al trabajo con señales de entrada con un número arbitrario de canales.

Si se ejecutara un análisis de señal en sus partes directa y ambiental con, por ejemplo, una señal envolvente de 5.1 canales con un canal izquierdo, un canal central, un canal derecho y un realce de baja frecuencia (sub-bajo), no es totalmente claro cómo se aplicaría un análisis de señal directa/ ambiental. Se podría pensar en la comparación de cada par de los seis canales, lo que da lugar a un procesamiento jerárquico que consta, a la larga, de 15 operaciones de comparación diferentes. Luego, una vez realizadas todas estas 15 operaciones de comparación, en que cada canal ha sido comparado con todos los demás canales, habría que determinar cómo evaluar los 15 resultados. Esto lleva mucho tiempo, los resultados son difíciles de interpretar y, debido a la considerable cantidad de recursos de procesamiento, es imposible de utilizar, por ej., para aplicaciones en tiempo real de separación de señales directas/ ambientales o, en general, descomposición de señales que pueden utilizarse, por ejemplo, en el contexto de operaciones de audio de mezcla ascendente o de otro tipo.

En la obra de M. M. Goodwin y J. M. Jot, "Primary–ambient signal decomposition and vector–based localization for spatial audio coding and enhancement," en *Proc. Of ICASSP 2007*, 2007, se aplica un *análisis de los componentes principales* a las señales de entrada a fin de realizar la descomposición de señales primarias (= directas) y ambientales. Los modelos utilizados en el trabajo de Christof Faller, "Multiple-loudspeaker playback of stereo signals," *Journal of the Audio Engineering Society*, vol. 54, no. 11, pp. 1051–1064, noviembre de 2006 y C. Faller, "A highly directive 2–capsule based microphone system," en *Preprint 123rd Conv. Aud. Eng. Soc.*, oct. 2007 asumen el sonido no correlacionado o parcialmente correlacionado en las señales estéreo y de micrófono, respectivamente. Derivan filtros para extraer la señal difusa/ambiental, dada esta suposición. Estos enfoques se limitan a señales de audio de uno y dos canales.

Otra referencia es la de C. Avendano y J.–M. Jot, "A frequency–domain approach to multichannel upmix", *Journal of the Audio Engineering Society*, vol. 52, no. 7/8, pp. 740–749, 2004. La referencia de M. M. Goodwin y J. M. Jot, "Primary–ambient signal decomposition and vector–based localization for spatial audio coding and enhancement," en *Proc. Of ICASSP 2007*, 2007, presenta los siguientes comentarios sobre la referencia de Avendano, Jot. La referencia presenta una estrategia que conlleva la creación de una máscara de tiempo–frecuencia para extraer el componente ambiental de una señal de audio estéreo de entrada. No obstante, la máscara se basa en la correlación cruzada entre las señales de los canales izquierdo y derecho, por lo que este enfoque no es inmediatamente aplicable al problema de extraer el componente de ambiente de una entrada multicanal arbitraria. Para usar ese método basado en la correlación en este caso de alto orden se necesitaría un análisis de correlación jerárquica por pares, lo que conllevaría un significativo costo informático, o alguna medida alternativa de correlación entre los múltiples canales.

La Renderización Espacial de Respuesta a los Impulsos (SIRR) (Juha Merimaa y Ville Pulkki, "Spatial impulse response rendering", en *Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFx'04)*, 2004) estima el sonido directo con dirección y sonido difuso en las respuestas a impulsos en Formato B. De manera muy similar a SIRR, la Codificación de Audio Direccional (DirAC) (Ville Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, junio de 2007) implementa un análisis similar de sonido directo y difuso a las señales de audio continuas de Formato B.

La estrategia presentada por Julia Jakka, *Binaural to Multichannel Audio Upmix*, para su tesis de licenciatura y doctoral, Helsinki University of Technology, 2005 describe señales binaurales que utilizan una mezcla ascendente como entrada.

La referencia Boaz Rafaely, "Spatially Optimal Wiener Filtering in a Reverberant Sound field, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2001, octubre 21 a 24, 2001, New Paltz, New York," describe la derivación de filtros Wiener que son espacialmente óptimos para los campos sonoros reverberantes. Se da una aplicación a la cancelación de ruido en dos micrófonos en recintos reverberantes. Los filtros óptimos que se derivan de la correlación espacial de campos sonoros difusos capturan el comportamiento local de los campos sonoros y, por lo tanto, son de menor orden y potencialmente más robustos espacialmente que los filtros adaptativos de cancelación de ruido convencionales en recintos reverberantes. Se presentan formulaciones para filtros óptimos libres y causalmente restringidos y se demuestra una aplicación ilustrativa al realce de voz con dos micrófonos utilizando un simulacro por computadora.

El documento US 2007/165868 se considera la técnica anterior más próxima. Muestra un aparato para descomponer una señal de entrada que tiene un número de al menos tres canales de entrada, que comprende:

un mezclador para mezclar la señal de entrada para obtener una señal mezclada, donde el mezclador está configurado para efectuar la mezcla de manera que un número de canales de mezcla de la señal mezclada sea al menos 2;

un analizador para analizar la señal mezclada a fin de derivar un resultado del análisis; y un procesador de señales para procesar una señal derivada de la señal de entrada utilizando el resultado del análisis, donde el procesador de señales está configurado para aplicar el resultado del análisis a los canales de entrada de la señal de entrada o a canales de la señal derivada de la señal de entrada para obtener una señal descompuesta.

Un objetivo de la presente invención es presentar un concepto mejorado para descomponer una señal de entrada.

Este objetivo se logra por medio de un aparato para descomponer una señal de entrada de acuerdo con la reivindicación 1, un método para descomponer una señal de entrada de acuerdo con la reivindicación 14 o un programa de computación de acuerdo con la reivindicación 15.

La presente invención se basa en el hallazgo de que, para descomponer una señal multicanal, una estrategia ventajosa consiste en no realizar el análisis con respecto a los diferentes componentes de la señal con la señal de entrada directamente, es decir con la señal que tiene por lo menos tres canales de entrada. Por el contrario, se procesa la señal de entrada multicanal que consta de por lo menos tres canales de entrada por medio de un mezclador descendente para efectuar la mezcla descendente de la señal de entrada a fin de obtener una señal con mezcla descendente. La señal con mezcla descendente consta de un número de canales de mezcla descendente que es menor que el número de canales de entrada y es, preferentemente, de dos. Luego, se ejecuta el análisis de la señal de entrada sobre la señal con mezcla descendente y no sobre la señal de entrada directamente y el análisis da lugar a un resultado del análisis. Sin embargo, este resultado del análisis no se aplica a la señal con mezcla descendente, sino que se aplica a la señal de entrada o, de lo contrario, a una señal derivada de la señal de entrada donde esta señal derivada de la señal de entrada puede ser una señal de mezcla ascendente o, dependiendo del número de canales de las señales de entrada, también una señal de mezcla descendente, aunque esta señal derivada de la señal de entrada suele ser diferente de la señal con mezcla descendente, sobre la cual se ha realizado el análisis. Cuando se considera el caso, por ejemplo, de que la señal de entrada es una señal de 5,1 canales, luego la señal de mezcla descendente, sobre la cual se realiza el análisis, podría ser una mezcla descendente estéreo con dos canales. A continuación se aplican los resultados del análisis a la señal de entrada 5.1 directamente, a una mezcla ascendente mayor tal como una señal de salida 7.1 o a una mezcla descendente multicanal de la señal de entrada que tiene, por ejemplo, sólo tres canales, que son el canal izquierdo, el canal central y el canal derecho, cuando sólo se cuenta con un aparato de renderización de audio de tres canales. En todo caso, no obstante, la señal sobre la cual el procesador de señales aplica los resultados del análisis es diferente de la señal de mezcla descendente sobre la cual se ha realizado el análisis y, por lo general, tiene más canales que la señal con mezcla descendente, sobre la cual se realiza el análisis con respecto a los componentes de la señal.

El denominado análisis/ procesamiento "indirecto" es posible debido a que se puede presumir que todo componente de la señal en los canales de entrada individuales también aparecen en los canales con mezcla descendente, puesto que la mezcla descendente consiste típicamente en una adición de canales de entrada de diferentes maneras. Una mezcla descendente sencilla consiste, por ejemplo, en la ponderación de los canales de entrada individuales necesaria por una norma de mezcla descendente o una matriz de mezcla descendente y luego la suma de los mismos una vez ponderados. Una mezcla descendente alternativa consiste en filtrar los canales de entrada con ciertos filtros tales como filtros HRTF y en realizar la mezcla descendente utilizando señales filtradas, es decir las señales filtradas por filtros HRTF como se conoce en la técnica. En el caso de una señal de entrada de cinco canales se requieren 10 filtros HRTF, y las salidas de los filtro HRTF correspondientes a la parte izquierda /oído izquierdo se suman y las salidas de los filtros HRTF correspondientes a los filtros de los canales derechos se suman para el oído derecho. Se pueden aplicar otras mezclas descendentes a fin de reducir el número de canales que deben ser procesados en el analizador de señales.

Por ende, las realizaciones de la presente invención describen un novedoso concepto para extraer componentes perceptualmente diferenciados de señales de entrada arbitrarias mediante la consideración de una señal de análisis, en tanto que el resultado del análisis se aplica a la señal de entrada. Dicha señal de análisis se puede obtener, por ej. considerando un modelo de propagación de los canales o señales de parlantes a los oídos. Esto está motivado, en parte, por el hecho de que el sistema auditivo humano también utiliza únicamente dos sensores (el oído izquierdo y derecho) para evaluar los campos sonoros. Por consiguiente, la extracción de componentes perceptualmente distintos se reduce básicamente a la consideración de una señal de análisis que en lo sucesivo se indica como mezcla

descendente. En todo este documento, el término mezcla descendente se utiliza para indicar cualquier procesamiento previo de la señal multicanal que dé lugar a una señal de análisis (esto puede incluir, por ej. un modelo de propagación, HRTFs, BRIRs, una simple mezcla descendente de factores cruzados).

Conociendo el formato de la entrada dada y las características deseadas de la señal a extraer, se pueden definir las relaciones ideales entre canales para el formato con mezcla descendente y, por ello, basta con un análisis de esta señal de análisis para generar una máscara de ponderación (o múltiples máscaras de ponderación) para la descomposición de señales multicanal.

En una forma de realización, el problema de los múltiples canales se simplifica utilizando una mezcla descendente estéreo de una señal envolvente y aplicando un análisis directo/ambiente a la mezcla descendente. Basándose en el resultado, es decir en los cálculos de espectros de potencia de corto tiempo de los sonidos directos y ambientales, se derivan filtros para descomponer una señal de N canales a N canales de sonido directo y N de sonido ambiental.

La presente invención es ventajosa por el hecho de que se aplica el análisis de señales a un menor número de canales, lo que reduce significativamente el tiempo de procesamiento necesario, por lo que el concepto de la invención se puede aplicar incluso a aplicaciones en tiempo real para la mezcla ascendente o la mezcla descendente de cualquier otra operación de procesamiento de señales en que se necesiten los componentes diferentes tales como los componentes perceptualmente diferentes de una señal.

Otra ventaja de la presente invención es que, si bien se realiza una mezcla descendente, se ha encontrado que ésta no deteriora la detectabilidad de los componentes perceptualmente diferentes de la señal de entrada. En otras palabras, aun cuando los canales de entrada han sido objeto de mezcla descendente, de todas maneras es muy posible separar los componentes individuales de la señal. Más aun, la mezcla descendente funciona como una suerte de "colección" de todos los componentes de la señal de todos los canales de entrada en dos canales y el único análisis aplicado a estas señales de mezcla descendente "conectadas" produce un único resultado que ya no tiene que ser interpretado y que se puede utilizar directamente para el procesamiento de la señal.

En una realización preferida, se obtiene una eficiencia particular con el fin de descomponer la señal cuando el análisis de señales se ejecuta sobre la base de la curva de similitud dependiente de la frecuencia precalculada como curva de referencia. El término similitud incluye la correlación y la coherencia, donde – en un sentido estricto matemático – la correlación se calcula entre dos señales sin desplazamiento en el tiempo adicional y se calcula la coherencia desplazando las dos señales en tiempo/ fase de manera que las señales tengan una correlación máxima y la correlación real en la frecuencia y luego se la calcula con el desplazamiento en el tiempo/fase aplicado. En el marco de este texto, se considera que similitud, correlación y coherencia tienen el mismo significado, es decir un grado cuantitativo de similitud entre dos señales, por ej., cuando un valor absoluto más elevado de similitud significa que las dos señales son más similares y un valor absoluto más bajo de similitud significa que las dos señales son menos similares.

Se ha demostrado que el uso de una curva de similitud como curva de referencia da lugar a un análisis de muy eficiente implementación, puesto que se puede utilizar la curva para sencillas operaciones de comparación y/o para cálculos de factores de ponderación. El uso de una curva de similitud dependiente de la frecuencia previamente calculada permite realizar sólo cálculos sencillos en lugar de operaciones de filtrado de Wiener más complejas. Más aun, la aplicación de la curva de similitud dependiente de la frecuencia es de especial utilidad debido a que no se encara el problema desde un punto de vista estadístico sino que se lo enfrenta de manera más analítica, dado que se introduce la mayor cantidad de información posible desde la configuración actual a fin de obtener una solución al problema. Además, la flexibilidad de este procedimiento es muy elevada, puesto que la curva de referencia puede ser obtenida de muchas maneras diferentes. Una manera consiste en medir realmente las dos o más señales en una configuración determinada y luego calcular la curva de similitud en la frecuencia a partir de las señales medidas. Por lo tanto, se pueden emitir señales independientes de parlantes o señales diferentes que tienen un cierto grado de dependencia conocido de antemano.

La otra alternativa preferida consiste simplemente en calcular la curva de similitud bajo la presunción de señales independientes. En este caso, no hay ninguna señal realmente necesaria, puesto que el resultado es independiente de las señales.

La descomposición de señales empleando una curva de referencia para el análisis de las señales se puede aplicar al procesamiento estéreo, es decir, para descomponer una señal estéreo. Por otro lado, este procedimiento se puede implementar asimismo junto con un mezclador descendente para descomponer señales multicanales. Por otro lado, también se puede implementar este procedimiento para señales multicanales sin utilizar un mezclador descendente cuando se contempla la evaluación por pares de señales de manera jerárquica.

A continuación se describen las realizaciones preferidas de la presente invención con respecto a las figuras que la acompañan, en las cuales:

La Fig. 1 es un diagrama de bloques para ilustrar un aparato para descomponer una señal de entrada que utiliza un mezclador descendente;

La Fig. 2 es un diagrama de bloques que ilustra una implementación de un aparato para descomponer una señal que tiene un número de por lo menos tres canales de entrada que utiliza un analizador con una curva de correlación dependiente de la frecuencia previamente calculada de acuerdo con otro aspecto de la invención;

La Fig. 3 ilustra otra implementación preferida de la presente invención con un procesamiento en el dominio de la frecuencia para la mezcla descendente, el análisis y el procesamiento de señales;

La Fig. 4 ilustra una curva de correlación dependiente de la frecuencia previamente calculada correspondiente a una curva de referencia para el análisis indicado en la Fig. 1 o la Fig. 2;

La Fig. 5 exhibe un diagrama de bloques que ilustra otro procesamiento para extraer los componentes independientes;

La Fig. 6 ilustra otra implementación de un diagrama de bloques para el procesamiento posterior en que se extraen los componentes difusos independientes, directos independientes y directos;

La Fig. 7 ilustra un diagrama de bloques que implementa el mezclador descendente como generador de señales de análisis;

5 La Fig. 8 ilustra un gráfico de flujo para indicar una manera preferible de procesamiento en el analizador de señales de la Fig. 1 o la Fig. 2;

Las Figs. 9a–9e ilustran diferentes curvas de correlación dependientes de la frecuencia precalculadas que se pueden utilizar como curvas de referencia para varias configuraciones diferentes con diferentes números y posiciones de las fuentes de sonido (tales como parlantes);

10 La Fig. 10 ilustra a un diagrama de bloques para describir otra forma de realización para una estimación de la difusividad cuando los componentes difusos son los componentes que se han de descomponer y

Las Figs. 11A y 11B ilustran ejemplos de ecuaciones para aplicar un análisis de señales sin una curva de correlación dependiente de la frecuencia, sino que basándose en la estrategia de filtrado de Wiener.

15 La Fig. 1 ilustra un aparato para descomponer una señal de entrada 10 que consta de un número de por lo menos tres canales de entrada o, en general, N canales de entrada. Estos canales de entrada son ingresados a un mezclador descendente 12 para efectuar la mezcla descendente de la señal de entrada a fin de obtener una señal con mezcla descendente 14, donde el mezclador descendente 12 está configurado para efectuar la mezcla descendente de manera que un número de canales de mezcla descendente de la señal con mezcla descendente 14, que está indicado por "m", sea por lo menos dos y menor que el número de canales de entrada de la señal de entrada 10. Los m canales de mezcla descendente son ingresados a un analizador 16 para analizar la señal con mezcla descendente a fin de derivar un resultado del análisis 18. El resultado del análisis 18 es ingresado a un procesador de señales 20, donde el procesador de señales está dispuesto para procesar la señal de entrada 10 o una señal derivada de la señal de entrada por medio de un derivador de señales 22 utilizando el resultado del análisis, donde el procesador de señales 20 está configurado para aplicar los resultados del análisis a los canales de entrada o a los canales de la señal 24 derivada de la señal de entrada para obtener una señal descompuesta 26.

20 En la realización ilustrada en la Fig. 1, un número de canales de entrada es n, el número de canales de mezcla descendente es m, el número de canales derivados es l y el número de canales de salida es igual l, cuando el procesador de señales procesa la señal derivada en lugar de la señal de entrada. Por otro lado, cuando el derivador de señales 22 no existe, luego la señal de entrada es procesada directamente por el procesador de señales y luego el número de canales de la señal descompuesta 26 indicado por "l" en la Fig. 1 ha de ser igual a n. Por ende, la Fig. 1 ilustra dos ejemplos diferentes. Un ejemplo no cuenta con el derivador de señales 22 y la señal de entrada se aplica directamente al procesador de señales 20. En el otro ejemplo se implementa el derivador de señales 22 y luego la señal derivada 24, en lugar de la señal de entrada 10, es procesada por el procesador de señales 20. El derivador de señales puede ser, por ejemplo, un mezclador de canales de audio tal como un mezclador ascendente para generar más canales de salida. En este caso l sería mayor que n. En otra forma de realización, el derivador de señales podría ser otro procesador de audio que ejecuta la ponderación, retraso o cualquier otro procedimiento de los canales de entrada y, en este caso, el número de canales de salida l del derivador de señales 22 sería igual al número n de canales de entrada. En una implementación adicional, el derivador de señales podría ser un mezclador descendente que reduce el número de canales de la señal de entrada a la señal derivada. En esta implementación, es preferible que el número l sea aun mayor que el número m de canales con mezcla descendente a fin de obtener una de las ventajas de la presente invención, es decir que se aplica el análisis de señales a un menor número de señales de canales.

30 El analizador cumple la función de analizar la señal con mezcla descendente con respecto a los componentes perceptualmente distintos. Estos componentes perceptualmente distintos pueden ser componentes independientes de los canales individuales por un lado, y componentes dependientes por el otro. Los componentes de señal alternativos a analizar mediante la presente invención son componentes directos por un lado y componentes ambientales por el otro. Hay muchos componentes más que pueden ser separados por la presente invención, como por ejemplo componentes de voz de componentes de música, componentes de ruido de componentes de voz, componentes de ruido de componentes de música, componentes de ruido de alta frecuencia con respecto a componentes de ruido de baja frecuencia, en señales de múltiples tonos los componentes producidos por los diferentes instrumentos, etc. Esto se debe a que existen potentes herramientas de análisis tales como el filtrado de Wiener descrito en el contexto de las Figs. 11A, 11B u otros procedimientos de análisis tales como el uso de una curva de correlación dependiente de la frecuencia como se describe, por ejemplo, en el contexto de la Fig. 8 de acuerdo con la presente invención.

35 La Fig. 2 ilustra otro aspecto, en que se implementa el analizador para usar una curva de correlación dependiente de la frecuencia precalculada 16. Por consiguiente, el aparato para descomponer una señal 28 que consta de una pluralidad de canales comprende el analizador 16 para analizar una correlación entre dos canales de una señal de análisis idéntica a la señal de entrada o relacionada con la señal de entrada, por ejemplo, mediante una operación de mezcla descendente ilustrada en el contexto de la Fig. 1. La señal de análisis analizada por el analizador 16 tiene por lo menos dos canales de análisis, y el analizador 16 está configurado para usar una curva de correlación dependiente de la frecuencia precalculada como curva de referencia para determinar el resultado del análisis 18. El procesador de señales 20 puede operar de la misma manera descrita en el contexto de la Fig. 1 y está configurado para procesar la señal de análisis o una señal derivada de la señal de análisis mediante un derivador de señales 22, donde el derivador de señales 22 puede ser implementado de manera similar a la descrita en el contexto del derivador de señales 22 de la Fig. 1. Por otro lado, el procesador de señales puede procesar una señal, a partir de la cual se deriva la señal de

análisis y el procesamiento de señales utiliza el resultado del análisis para obtener una señal descompuesta. Por ende, en la realización de la Fig. 2 la señal de entrada puede ser idéntica a la señal de análisis y, en este caso, la señal de análisis también puede ser una señal estéreo con sólo dos canales, como se ilustra en la Fig. 2. Por otro lado, además, la señal de análisis puede ser derivada de una señal de entrada mediante cualquier tipo de procesamiento, como por ejemplo la mezcla descendente, como se describe en el contexto de la Fig. 1 o por cualquier otro tipo de procesamiento tal como la mezcla ascendente o similar. Además, el procesador de señales 20 puede ser de utilidad para aplicar el procesamiento de señales a la misma señal que ha sido ingresada al analizador o el procesador de señales puede aplicar un procesamiento de señales a una señal, de la cual se ha derivado la señal de análisis, como se indica en el contexto de la Fig. 1, o bien el procesador de señales puede aplicar un procesamiento de señales a una señal que ha sido derivada de la señal de análisis, por ejemplo por mezcla ascendente o similar.

Por consiguiente, existen diferentes posibilidades para el procesador de señales y todas estas posibilidades son ventajosas debido a la singular operación del analizador que utiliza una curva de correlación dependiente de la frecuencia precalculada como curva de referencia para determinar el resultado del análisis.

A continuación se describen otras formas de realización. Se ha de notar que, como se describe en el contexto de la Fig. 2, se considera incluso el uso de una señal de análisis de dos canales (sin mezcla descendente) el uso de una señal de análisis de dos canales (sin mezcla descendente). Por ende, la presente invención descrita en sus diferentes aspectos en el contexto de las Fig. 1 y Fig. 2, que se pueden utilizar juntos o como aspectos separados, la mezcla descendente puede ser procesada por el analizador o una señal de dos canales, que probablemente no ha sido generada por una mezcla descendente, puede ser procesada por el analizador de señales usando la curva de referencia previamente calculada.

En este contexto, se debe tener en cuenta que la siguiente descripción de los aspectos de implementación se pueden aplicar a ambos aspectos ilustrados esquemáticamente en la Fig. 1 y en la Fig. 2, aun cuando ciertas características se describan sólo con respecto a un aspecto no a ambos. Si, por ejemplo, se considera la Fig. 3, queda claro que las características de dominio de la frecuencia de la Fig. 3 se describen en el contexto del aspecto ilustrado en la Fig. 1, aunque es obvio que seguidamente se describe una transformación de tiempo/frecuencia con respecto a la Fig. 3 y que también se puede aplicar la transformada inversa a la implementación de la Fig. 2, que no incluye un mezclador descendente, pero que consta de un analizador estipulado que utiliza una curva de correlación precalculada dependiente de la frecuencia.

En particular, se colocaría el convertidor de tiempo/frecuencia para convertir la señal de análisis antes de ingresar la señal de análisis al analizador, y se colocaría el convertidor de frecuencia/tiempo a la salida del procesador de señales para convertir la señal procesada de regreso al dominio del tiempo. En caso de existir un derivador de señales, el convertidor de tiempo/frecuencia podría estar situado a la entrada del derivador de señales de manera que el derivador de señales, el analizador y el procesador de señales operen en el dominio de subbandas de frecuencia. En este contexto, frecuencia y subbanda se refieren básicamente a una porción de la frecuencia de una representación de frecuencia.

Resulta claro además que el analizador de la Fig. 1 puede ser implementado de numerosas maneras diferentes, aunque este analizador también se implementa, en una forma de realización, como el analizador descrito en la Fig. 2, es decir como analizador que utiliza una curva de correlación dependiente de la frecuencia precalculada como alternativa al filtrado de Wiener o cualquier otro método de análisis.

La realización de la Fig. 3 aplica un procedimiento de mezcla descendente a una señal de entrada arbitraria para obtener una representación de dos canales. Se ejecuta un análisis en el dominio de tiempo-frecuencia y se calculan máscaras de ponderación que se multiplican con la representación de tiempo frecuencia de la señal de entrada, como se ilustra en la Fig. 3.

En la figura, T/F indica una transformada de tiempo frecuencia, por lo general una Transformada de Fourier de Corto Tiempo (STFT). iT/F denota la respectiva transformada inversa. $[x_1(n), \dots, x_N(n)]$ son las señales de entrada en el dominio del tiempo, donde n es el índice de tiempo. $[X_1(m, i), \dots, X_N(m, i)]$ denotan los coeficientes de la descomposición de frecuencia, donde m es el índice de tiempo de descomposición e i es el índice de frecuencia de descomposición. $[D_1(m, i), D_2(m, i)]$ son los dos canales de la señal con mezcla descendente.

$$\begin{pmatrix} D_1(m, i) \\ D_2(m, i) \end{pmatrix} = \begin{pmatrix} H_{11}(i) & H_{12}(i) & \cdots & H_{1N}(i) \\ H_{21}(i) & H_{22}(i) & \cdots & H_{2N}(i) \end{pmatrix} \begin{pmatrix} X_1(m, i) \\ X_2(m, i) \\ \vdots \\ X_N(m, i) \end{pmatrix} \quad (1)$$

$W(m, i)$ es la ponderación calculada. $[Y_1(m, i), \dots, Y_N(m, i)]$ son las descomposiciones de frecuencia calculadas de cada canal. $H_{ij}(i)$ son los coeficientes de mezcla descendente, que deben tener valor real o valor complejo y los

coeficientes pueden ser constantes en el tiempo o variantes en el tiempo. Por ende, los coeficientes de mezcla descendente pueden ser sólo constantes o filtros tales como filtros HRTF, filtros de reverberación o filtros similares.

$$Y_j(m,i) = W_j(m,i) \cdot X_j(m,i), \text{ donde } j = (1, 2, \dots, N) \quad (2)$$

5 En la Fig. 3 se ilustra el caso de aplicación de la misma ponderación a todos los canales.

$$Y_j(m,i) = W(m,i) \cdot X_j(m,i) \quad (3)$$

10 $[y_1(n), \dots, y_N(n)]$ son las señales de salida en el dominio del tiempo que comprenden los componentes de señal extraídos. (La señal de entrada puede tener un número arbitrario de canales (N), producido por una configuración de parlantes de reproducción objetivo arbitraria. La mezcla descendente puede incluir HRTFs para obtener señales de entrada al oído, simulación de filtros auditivos, La mezcla descendente también se puede llevar a cabo en el dominio del tiempo.).

15 En una forma de realización, la diferencia entre una correlación de referencia (En todo este texto, se utiliza el término correlación como sinónimo de la similitud entre canales y, por consiguiente, también puede incluir evaluaciones de desplazamientos en el tiempo, para lo cual se utiliza habitualmente el término coherencia. Aun en el caso de evaluarse desplazamientos en el tiempo, el valor resultante puede tener un signo. (Comúnmente, la coherencia se define como sólo con valores positivos) en función de la frecuencia ($c_{ref}(\omega)$), y se computa la correlación real de la señal de

20 entrada con mezcla descendente ($c_{sig}(\omega)$). Dependiendo de la desviación de la curva real con respecto a la curva de referencia, se calcula un factor de ponderación por cada mosaico de tiempo–frecuencia, indicando que comprende componentes dependientes o independientes. La ponderación de tiempo–frecuencia obtenida indica los componentes independientes y ya se la puede aplicar a cada canal de la señal de entrada para producir una señal multicanal (un número de canales igual al número de canales de entrada) que incluye partes independientes que se pueden percibir como diferenciadas o difusas.

25 La curva de referencia se puede definir de diferentes maneras. Los Ejemplos son:

- Curva de referencia teórica ideal correspondiente a un campo sonoro difuso bi- o tridimensional difuso idealizado compuesto por componentes independientes.
- La curva ideal que se obtiene con la configuración de parlantes objetivo de referencia para la señal de entrada dada (por ej., la configuración estéreo standard con ángulos azimutales ($\pm 30^\circ$), o la configuración standard de cinco canales de acuerdo con ITU–R BS.775 con ángulos azimutales ($0^\circ, \pm 30^\circ, \pm 110^\circ$)).

30 • La curva ideal para la configuración de parlantes que en realidad está presente (las posiciones reales se podrían medir o se las podría conocer mediante entradas del usuario. La curva de referencia se puede calcular suponiendo la reproducción de señales independientes por los parlantes dados).

• La potencia de corto tiempo real dependiente de la frecuencia de cada señal de entrada se puede incorporar al cálculo de la referencia.

35 Dada una curva de referencia dependiente de la frecuencia ($c_{ref}(\omega)$), se puede definir un umbral superior ($c_{hi}(\omega)$) y un umbral inferior ($c_{lo}(\omega)$) (ver la Fig. 4). Las curvas umbral pueden coincidir con la curva de referencia ($c_{ref}(\omega) = c_{hi}(\omega) = c_{lo}(\omega)$), o se las puede definir presumiendo los umbrales de detectabilidad, o se las puede derivar en forma heurística.

40 Si la desviación de la curva real con respecto a la curva de referencia está dentro de los límites establecidos por los umbrales, la caja real obtiene una ponderación que indica los componentes individuales. Por encima del umbral superior o por debajo del umbral inferior, la caja está indicada como dependiente. Esta indicación puede ser binaria, o gradual (es decir, siguiendo una función de decisión suave). En particular, si el umbral superior y el inferior coinciden con la curva de referencia, la ponderación aplicada está directamente relacionada con la desviación de la curva de referencia.

45 Con referencia a la Fig. 3, el número de referencia 32 ilustra un convertidor de tiempo/frecuencia que puede ser implementado en forma de transformada de Fourier de corto tiempo o como cualquier tipo de banco de filtros que genere señales de subbandas tales como un banco de filtros QMF o similar. Independientemente de la implementación detallada del convertidor de tiempo/frecuencia 32, la salida del convertidor de tiempo/frecuencia es, por cada canal de entrada x_i un espectro por cada periodo de tiempo de la señal de entrada. Por consiguiente, el procesador de tiempo/frecuencia 32 puede ser implementado para tomar siempre un bloque de muestras de entrada de una señal de canal individual y para calcular la representación de la frecuencia como espectro FFT con líneas espectrales que se proyectan desde una frecuencia más baja a una frecuencia más elevada. Luego, para el siguiente bloque de tiempo, se ejecuta el mismo procedimiento para que, al final, se calculen los espectros de corto tipo por cada señal de canal de entrada. Se dice que un cierto rango de frecuencias de un determinado espectro que se relaciona con un determinado bloque de muestras de entrada de un canal de entrada es un “mosaico de tiempo/frecuencia” y, preferentemente el

análisis que se ejecuta en el analizador 16 se realiza basándose en estos mosaicos de tiempo/frecuencia. Por lo tanto el analizador recibe, como entrada para un mosaico de tiempo/frecuencia, el valor espectral de una primera frecuencia correspondiente a un determinado bloque de muestras de entrada del primer canal con mezcla descendente D_1 y recibe el valor correspondiente a la misma frecuencia y el mismo bloque (en el tiempo) del segundo canal con mezcla descendente D_2 .

A continuación, como se ilustra por ejemplo en la Fig. 8, el analizador 16 está configurado para determinar (80) un valor de correlación entre los dos canales de entrada por subbanda y bloque de tiempo, es decir un valor de correlación correspondiente un mosaico de tiempo/frecuencia. Seguidamente, el analizador 16 obtiene, en la forma de realización ilustrada con respecto a la Fig. 2 o la Fig. 4, un valor de correlación (82) para la correspondiente subbanda de la curva de correlación de referencia. Cuando, por ejemplo, la subbanda es la subbanda indicada en 40 en la Fig. 4, luego el paso 82 da lugar al valor 41 que indica una correlación entre -1 y $+1$, y el valor 41 es entonces el valor de correlación obtenido. A continuación, en el paso 83, se ejecuta el resultado correspondiente a la subbanda que utiliza el valor de correlación determinado del paso 80 y el valor de correlación obtenido 41 obtenido en el paso 82 mediante la ejecución de una comparación y la decisión posterior, o se lo hace calculando una diferencia real. El resultado puede ser, como se señalara anteriormente, un resultado binario que dice que el mosaico real de tiempo/frecuencia considerado en la señal de análisis /mezcla descendente tiene componentes independientes. Esta decisión se toma cuando el valor de correlación determinado en realidad (en el paso 80) es igual al valor de correlación de referencia o es bastante aproximado al valor de correlación de referencia.

Cuando, sin embargo, se determina que el valor de correlación determinado, indica una correlación absoluta más elevada que el valor de correlación de referencia, luego se determina que el mosaico de tiempo/frecuencia en cuestión comprende componentes dependientes. Por consiguiente, cuando la correlación de un mosaico de tiempo/frecuencia de la señal de mezcla descendente o de análisis indica un valor de correlación absoluto más alto que la curva de referencia, entonces se puede decir que los componentes de este mosaico de tiempo/frecuencia son dependientes entre sí. No obstante, cuando se indica que la correlación está muy cerca de la curva de referencia, en ese caso se puede decir que los componentes son independientes. Los componentes dependientes pueden recibir un primer valor de ponderación un primer valor de ponderación tal como 1 y los componentes independientes pueden recibir un segundo valor de ponderación tal como 0. De preferencia, como se ilustra en la Fig. 4, se utilizan umbrales alto y bajo que están espaciados a una distancia de la línea de referencia para obtener un mejor resultado que es más adecuado que el uso de la curva de referencia sola.

Más aun, en lo que respecta a la Fig. 4, cabe señalar que la correlación puede variar entre -1 y $+1$. Además, una correlación con un signo negativo indica un desplazamiento de fases de 180° entre las señales. Por lo tanto, también se podrían aplicar otras correlaciones que sólo se extienden entre 0 y 1, en las cuales la parte negativa de la correlación simplemente se torna positiva. En este procedimiento, se ignoraría entonces un desplazamiento en el tiempo o un desplazamiento de fases con el fin de determinar la correlación.

La forma alternativa de calcular el resultado consiste en calcular en realidad la distancia entre el valor de correlación determinado en el bloque 80 y el valor de correlación obtenido en el bloque 82 y en determinar luego una métrica entre 0 y 1 como factor de ponderación basado en la distancia. Si bien la primera alternativa (1) de la Fig. 8 sólo da origen a valores de 0 o 1, la posibilidad (2) da lugar a valores de entre 0 y 1 y, en algunas implementaciones, es preferible.

El procesador de señales 20 de la Fig. 3 está ilustrado en forma de multiplicadores y los resultados del análisis son sólo un factor de ponderación determinado que es reenviado desde el analizador al procesador de señales ilustrado en 84 en la Fig. 8 y luego es aplicado al correspondiente mosaico de tiempo/frecuencia de la señal de entrada 10. Cuando, por ejemplo, el espectro considerado en realidad es el 20° espectro de la secuencia de espectros y cuando la caja de frecuencia considerada en realidad es la 5a frecuencia de este 20° espectro, luego el mosaico de tiempo/frecuencia puede estar indicado como (20, 5) donde el primer número indica el número del bloque en el tiempo y el segundo número indica la caja de frecuencia de este espectro. A continuación, se aplica el resultado del análisis correspondiente al mosaico de tiempo/frecuencia (20, 5) al correspondiente mosaico de tiempo/frecuencia (20, 5) de cada canal de la señal de entrada en la Fig. 3 o, cuando se implementa un derivador de señales como el ilustrado en la Fig. 1, al correspondiente mosaico de tiempo/frecuencia de cada canal de la señal derivada.

Seguidamente, se describe el cálculo de una curva de referencia en forma más detallada. Sin embargo, para la presente invención, básicamente no es importante la forma en que se derivara la curva de referencia. Puede ser una curva arbitraria o, por ejemplo, valores en una tabla de búsqueda que indica una relación ideal o ventajosa de las señales de entrada x_i de la señal de mezcla descendente D_0 , y en el contexto de la Fig. 2 de la señal de análisis. La siguiente derivación es ilustrativa.

Se puede evaluar la *difusión física* de un campo sonoro mediante un método dado a conocer por Cook et al. (Richard K. Cook, R. V. Waterhouse, R. D. Berendt, Seymour Edelman, y Jr. M.C. Thompson, "Measurement of correlation coefficients in reverberant sound fields," *Journal Of The Acoustical Society Of America*, vol. 27, no. 6, pp. 1072-1077, noviembre de 1955), que utiliza el *coeficiente de correlación* (r) de la presión del sonido en estado estable de ondas planas en dos puntos espacialmente separados, como se ilustra en la siguiente ecuación (4)

$$r = \frac{\langle p_1(n) \cdot p_2(n) \rangle}{\left[\langle p_1^2(n) \rangle \cdot \langle p_2^2(n) \rangle \right]^{1/2}}$$

donde $p_1(n)$ y $p_2(n)$ son las mediciones de presión de sonido en dos puntos, n es el índice de tiempo, y $\langle \cdot \rangle$ denota el promedio de tiempo. En un campo sonoro en estado estable, se pueden derivar las siguientes relaciones:

$$r(\mathbf{k}, d) = \frac{\sin(kd)}{kd} \text{ (en el caso de campos sonoros tridimensionales) y} \quad (5)$$

$$r(\mathbf{k}, d) = J_0(kd) \text{ (en el caso de campos sonoros bidimensionales)} \quad (6)$$

5 donde d es la distancia entre dos puntos de medición y $k = \frac{2\pi}{\lambda}$ es el número de onda, donde λ es la longitud de onda. (La curva física de referencia $r(k, d)$ ya puede ser utilizada como c_{ref} para el procesamiento posterior).

Una medida de la *difusividad perceptual* de un campo sonoro es el *coeficiente de correlación cruzada interaural* (ρ), medido en un campo sonoro. La medición ρ implica que el radio entre los sensores de presión (resp. de los oídos) es

10 fijo. Incluyendo esta restricción, r se convierte en una función de la frecuencia con la frecuencia de radián $\omega = kc$, donde c es la velocidad del sonido en el aire. Más aun, las señales de presión difieren de las señales de campo libre anteriormente consideradas debido a la reflexión, difracción y efectos de curvatura causados por los pabellones auriculares, la cabeza y torso del oyente. Esos efectos, esenciales para la audición espacial, son descritos por funciones de transferencia relacionadas con la cabeza (HRTFs). Tomando en cuenta esas influencias, las señales de presión consiguientes en las entradas a los oídos son $p_L(n, \omega)$ y $p_R(n, \omega)$. Para el cálculo, se pueden utilizar los

15 datos de HRTF medidos o se pueden obtener aproximaciones utilizando un modelo analítico (por ej. Richard O. Duda y William L. Martens, "Range dependence of the response of a spherical head model," *Journal Of The Acoustical Society Of America*, vol. 104, no. 5, pp. 3048–3058, noviembre de 1998).

Dado que el sistema auditivo humano actúa como analizador de frecuencia con limitada selectividad de frecuencia, se puede agregar también esta selectividad de frecuencia. Se presume que los filtros auditivos se comportan como filtros

20 pasa banda superpuestos. En la siguiente explicación ilustrativa, se utiliza una estrategia de bandas críticas para aproximar estos pasa banda superpuestos mediante filtros rectangulares. Se puede calcular el ancho de banda rectangular equivalente (ERB) en función de la frecuencia central (Brian R. Glasberg y Brian C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, pp. 103–138, 1990). teniendo en cuenta que el procesamiento binaural se realiza después del filtrado auditivo, se debe calcular ρ para canales de frecuencia

25 separados, para dar las siguientes señales de presión dependientes de la frecuencia

$$p_{\hat{L}}(n, \omega) = \frac{1}{b(\omega)} \int_{\omega - \frac{b(\omega)}{2}}^{\omega + \frac{b(\omega)}{2}} p_L(n, \omega) d\omega \quad (7)$$

$$p_{\hat{R}}(n, \omega) = \frac{1}{b(\omega)} \int_{\omega - \frac{b(\omega)}{2}}^{\omega + \frac{b(\omega)}{2}} p_R(n, \omega) d\omega, \quad (8)$$

donde los límites de integración están dados por los límites de la banda crítica de acuerdo con la frecuencia central real ω . Los factores $1/b(\omega)$ pueden o no utilizarse en las ecuaciones (7) y (8).

30 Si se anticipa o retarda una de las mediciones de presión de sonido por una diferencia de tiempo independiente de la frecuencia, se puede evaluar la coherencia de las señales. El sistema auditivo humano puede hacer uso de esa propiedad de alineamiento. Habitualmente, la coherencia interaural se calcula dentro del ± 1 ms. Dependiendo de la potencia de procesamiento existente, se pueden implementar cálculos que utilizan sólo el valor de retardo cero (en el caso de la baja complejidad) o la coherencia con una anticipación y retardo de tiempo (si es posible la alta complejidad).

35 A continuación, no se hace distinción entre ambos casos.

Se obtiene el comportamiento ideal considerando un campo sonoro difuso ideal, que puede ser idealizado en forma de campo de ondas que se compone de ondas planas no correlacionadas igualmente fuertes que se propagan en todas direcciones (es decir, una superposición de un número infinito de ondas planas que se propagan con relaciones de fase aleatorias y direcciones de propagación uniformemente distribuidas). Se puede considerar una señal radiada por un

40 parlante como onda plana para un oyente situado a una distancia suficiente. Esta presunción de ondas planas es común en la reproducción estereofónica por medio de altoparlantes. Por consiguiente, un campo sonoro sintético reproducido por altoparlantes consiste en ondas planas contribuyentes desde un número limitado de direcciones.

Dada una señal de entrada con N canales, producidos para la reproducción por medio de una disposición con las siguientes posiciones de parlantes $[l_1, l_2, l_3, \dots, l_N]$. (En el caso de una configuración de sólo reproducción horizontal,

45 l_i indica el ángulo azimutal. En el caso general, $l_i = (\text{azimut}, \text{elevación})$ indica la posición del parlante con respecto a la cabeza del oyente. Si la configuración presente en el recinto auditorio difiere de la configuración de referencia, l_i puede representar, por otro lado, las posiciones de los parlantes de la configuración de reproducción real). Con esta

información, se puede calcular una curva de referencia de coherencia interaural ρ_{ref} correspondiente a un simulacro de campo difuso para esta configuración bajo la presunción de que se alimentan señales independientes a cada parlante. Se puede incluir la potencia de la señal contribuida por cada canal de entrada en cada mosaico de tiempo-frecuencia en el cálculo de la curva de referencia. En el ejemplo de implementación, ρ_{ref} se utiliza como c_{ref} .

5 Se ilustran diferentes curvas de referencia como ejemplos de curvas de referencia o curvas de correlación dependientes de la frecuencia en las Figs. 9a a 9e correspondientes a un número diferente de fuentes de sonido en diferentes posiciones de las fuentes de sonido y diferentes orientaciones de la cabeza, como se indica en las Figs.

A continuación se describe en forma más detallada el cálculo de los resultados del análisis en el contexto de la Fig. 8 sobre la base de las curvas de referencia.

10 La meta consiste en derivar una ponderación que es igual a 1, si la correlación de los canales de mezcla descendente es igual a la correlación de referencia calculada bajo la presunción de que se están reproduciendo señales independientes desde todos los parlantes. Si la correlación de la mezcla descendente es igual a +1 o -1, la ponderación derivada debe ser 0, lo que indica que no hay presencia de componentes independientes. Entre esos casos extremos, la ponderación representaría una transición razonable entre la indicación como independiente ($W=1$) o completamente dependiente ($W=0$).

15 Dada la curva de correlación de referencia $c_{ref}(\omega)$ y el cálculo de la correlación / coherencia de la señal de entrada real reproducida por la configuración de reproducción real ($c_{sig}(\omega)$) (c_{sig} es la correlación resp. de la coherencia de la mezcla descendente), se puede calcular la desviación de $c_{sig}(\omega)$ de $c_{ref}(\omega)$. Esta desviación (que posiblemente incluye un umbral superior e inferior) se mapea con respecto al rango [0;1] para obtener una ponderación ($W(m, i)$) que se aplica a todos los canales de entrada a los componentes independientes separados.

20 El siguiente ejemplo ilustra un posible mapeo cuando los umbrales corresponden a la curva de referencia:

La magnitud de la desviación (indicada como Δ) de la curva real c_{sig} de la referencia c_{ref} está dada por

$$\Delta(\omega) = |c_{sig}(\omega) - c_{ref}(\omega)| \quad (9)$$

25 Dado que la correlación / coherencia está limitada entre [-1;+1], la desviación máxima posible hacia +1 o -1 por cada frecuencia está dada por

$$\bar{\Delta}_+(\omega) = 1 - c_{ref}(\omega) \quad (10)$$

$$\bar{\Delta}_-(\omega) = c_{ref}(\omega) + 1 \quad (11)$$

Por consiguiente, la ponderación correspondiente a cada frecuencia se obtiene de

$$W(\omega) = \begin{cases} 1 - \frac{\Delta(\omega)}{\bar{\Delta}_+(\omega)} & c_{sig}(\omega) \geq c_{ref}(\omega) \\ 1 - \frac{\Delta(\omega)}{\bar{\Delta}_-(\omega)} & c_{sig}(\omega) < c_{ref}(\omega) \end{cases} \quad (13)$$

30 Considerando la dependencia del tiempo y la resolución de frecuencia limitada de la descomposición de frecuencia, los valores de ponderación se derivan de la siguiente manera (En este caso, se da la situación general de una curva de referencia que puede cambiar en el tiempo. También es posible una curva de referencia independiente del tiempo (es decir $c_{ref}(i)$):

$$W(m, i) = \begin{cases} 1 - \frac{\Delta(m, i)}{\bar{\Delta}_+(m, i)} & c_{sig}(m, i) \geq c_{ref}(m, i), \\ 1 - \frac{\Delta(m, i)}{\bar{\Delta}_-(m, i)} & c_{sig}(m, i) < c_{ref}(m, i) \end{cases} \quad (14)$$

35 Dicho procesamiento se puede llevar a cabo en una descomposición en la frecuencia con coeficientes de frecuencia agrupados en subbandas perceptualmente motivadas por razones de complejidad informática y para obtener filtros con respuestas al impulso más cortas. Más aun, se podrían aplicar filtros de alisado y también se pueden aplicar funciones de compresión (es decir, la distorsión de la ponderación de manera ventajosa, introduciendo además valores mínimos y /o máximos.

40 La Fig. 5 ilustra una implementación adicional de la presente invención, en la cual se implementa el mezclador descendente usando HRTF y filtros auditivos de acuerdo con lo ilustrado. Más aun, la Fig. 5 ilustra asimismo que los

resultados del análisis producidos por el analizador 16 son los factores de ponderación por cada caja de tiempo/frecuencia y se ilustra el procesador de señales 20 como extractor para extraer componentes independientes. Seguidamente, la salida del procesador 20 es, una vez más, de N canales, aunque cada canal incluye ahora sólo los componentes independientes y no incluye más componentes dependientes. En esta implementación, el analizador

5

calcularía las ponderaciones de manera que, en la primera implementación de la Fig. 8, un componente independiente recibiría un valor de ponderación de 1 y un componente dependiente recibiría un valor de ponderación de 0. Luego, los mosaicos de tiempo/frecuencia de los N canales originales procesados por el procesador 20 que tienen componentes dependientes se ajustaría a 0.

10

En la otra alternativa, en que hay valores de ponderación de entre 0 y 1 en la Fig. 8, el analizador calcularía la ponderación de manera que un mosaico de tiempo/frecuencia que tiene una pequeña distancia hasta la curva de referencia recibiría un valor elevado (más cerca de 1), y un mosaico de tiempo/frecuencia con una gran distancia hasta la curva de referencia recibiría un factor de ponderación pequeño (que se acercaría más a 0). En la ponderación posterior ilustrada, por ejemplo, en la Fig. 3 en 20, los componentes independientes serían entonces amplificados, en tanto que los componentes dependientes se atenuarían.

15

Sin embargo, en caso de implementarse el procesador de señales 20 para no extraer los componentes independientes, sino para extraer los componentes dependientes, las ponderaciones serían asignadas de manera opuesta de manera que, cuando la ponderación se ejecuta en los multiplicadores 20 ilustrados en la Fig. 3, los componentes independientes se atenúan y los componentes dependientes se amplifican. Es por eso que se puede aplicar cada procesador de señales para extraer los componentes de señal, puesto que la determinación de los componentes de señal extraídos en realidad la realiza por la asignación real de valores de ponderación.

20

La Fig. 6 ilustra otra implementación del concepto de la invención, aunque en este caso con una implementación diferente del procesador 20. En la realización de la Fig. 6, el procesador 20 está implementado para extraer partes difusas independientes, partes directas independientes y partes/ componentes directos per se.

25

Para obtener, a partir de los componentes independientes separados (Y_1, \dots, Y_N), las partes que contribuyen a la percepción de un campo sonoro envolvente / ambiente, hay que tener en cuenta otras restricciones. Una de esas restricciones puede ser la presunción de que el sonido ambiente envolvente es igualmente fuerte desde cada dirección. Así, por ej. se puede extraer la energía mínima de cada mosaico de tiempo-frecuencia en cada canal de las señales sonoras independientes para obtener una señal ambiente envolvente (que puede ser procesada a su vez para obtener un alto número de canales ambiente). Ejemplo:

30

$$\tilde{Y}_j(m,i) = g_j(m,i) \cdot Y_j(m,i), \text{ with } g_j(m,i) = \sqrt{\frac{\min_{1 \leq k \leq N} \{P_{Y_k}(m,i)\}}{P_{Y_j}(m,i)}}, \quad (15)$$

donde P denota un cálculo de potencia de corto tiempo (Este ejemplo ilustra el caso más sencillo. Un caso excepcional obvio, en que esto no es aplicable es cuando uno de los canales incluye pausas de señal durante las cuales la potencia de este canal sería muy baja o cero.)

35

En algunos casos es ventajoso extraer la partes de energía iguales de todos los canales de entrada y calcular la ponderación utilizando sólo estos espectros extraídos.

$$\tilde{X}_j(m,i) = g_j(m,i) \cdot X_j(m,i), \text{ with } g_j(m,i) = \sqrt{\frac{\min_{1 \leq k \leq N} \{P_{X_k}(m,i)\}}{P_{X_j}(m,i)}}, \quad (16)$$

Los extraídos dependientes (los que se pueden derivar, por ej., como partes $Y_{\text{dependiente}} = Y_j(m,i) - X_j(m,i)$) se pueden utilizar para detectar dependencias de los canales y así estimar los signos direccionales inherentes a la señal de entrada, dando lugar a otros procesos tales como, por ej., el nuevo paneo.

40

La Fig. 7 ilustra una variante del concepto general. Se alimenta la señal de entrada de N canales a un generador de señales de análisis (ASG). La generación de la señal de análisis del canal M puede incluir, por ej. un modelo de propagación desde los canales / parlantes a los oídos u otros métodos indicados como mezcla descendente en todo este documento. La indicación de los componentes diferenciados se basa en la señal de análisis. Las máscaras que indican los diferentes componentes se aplican a las señales de entrada (A extracción / D extracción (20a, 20b)). Las señales de entrada ponderadas pueden ser procesadas posteriormente (A post / D post (70a, 70b) para dar señales de salida con carácter específico, donde en este ejemplo, se eligieron los designadores "A" y "D" para indicar que los componentes a extraer pueden ser "Ambientales" y "de Sonido Directo".

45

A continuación se describe la Fig. 10. Un campo sonoro fijo se denomina *difuso* si la distribución direccional de la energía de sonido no depende de la dirección. La distribución de energía direccional se puede evaluar midiendo todas las direcciones utilizando un micrófono sumamente direccional. En la acústica del recinto, el campo sonoro reverberante de un ambiente cerrado con frecuencia está modelado en forma de campo difuso. Un campo sonoro difuso puede ser idealizado como campo de ondas que se compone de ondas plantas no correlacionadas igualmente fuertes, que se propagan en todas direcciones. Ese tipo de campo sonoro es isotrópico y homogéneo.

50

Si la uniformidad de la distribución de energía es de particular interés, se puede utilizar el coeficiente de correlación de punto a punto

$$r = \frac{\langle p_1(t) \cdot p_2(t) \rangle}{\left[\langle p_1^2(t) \rangle \cdot \langle p_2^2(t) \rangle \right]^{\frac{1}{2}}}$$

de las presiones sonoras en estado estable $p_1(t)$ y $p_2(t)$ en dos puntos

espacialmente separados para evaluar la *difusión física* de un campo sonoro. En el caso de supuestos campos sonoros difusos tridimensionales y bidimensionales en estado estable ideales inducidos por una fuente sinusoidal, se pueden derivar las siguientes relaciones:

$$r_{3D} = \frac{\sin(kd)}{kd}$$

y

$$r_{2D} = J_0(kd),$$

$$k = \frac{2\pi}{\lambda} \text{ (donde } \lambda = \text{longitud de onda)}$$

donde es el número de ondas y d es la distancia entre los puntos de medición.

Dadas estas relaciones, se puede evaluar la difusión de un campo sonoro comparando los datos de medición con las curvas de referencia. Dado que las relaciones ideales son sólo condiciones necesarias, aunque no suficientes, se puede contemplar un número de mediciones con diferentes orientaciones del eje que conecta los micrófonos.

Considerando un oyente en un campo sonoro, las mediciones de presión sonora están dadas por las señales de entrada al oído $p_l(t)$ y $p_r(t)$. Por consiguiente, la distancia presunta d entre los puntos de medición se fija y r se convierte en una

función de sólo frecuencia con $f = \frac{kc}{2\pi}$, donde c es la velocidad del sonido en el aire. Las señales de entrada al oído

difieren de las señales de campo libre antes estudiadas en la influencia de los efectos causados por los pabellones auriculares, la cabeza y el torso del oyente. Esos efectos sustanciales para la audición espacial, se describen por funciones de transferencia relacionadas con la cabeza (HRTFs). Se pueden utilizar los datos HRTF medidos para incorporar estos efectos. Utilizamos un modelo analítico para simular una aproximación a las HRTFs. Se modela la cabeza en forma de esfera rígida con un radio de 8,75 cm y ubicaciones de los oídos en el azimut $\pm 100^\circ$ y elevación de 0° . Dado el comportamiento teórico de r en un campo sonoro difuso ideal y la influencia de las HRTFs, es posible determinar una curva de referencia de correlación cruzada interaural dependiente de la frecuencia correspondiente a los campos sonoros difusos.

La estimación de la difusividad se basa en la comparación de indicios simulados con indicios de referencia del campo difuso presunto. Esta comparación está sujeta a las limitaciones de la audición humana. En el sistema auditivo el procesamiento binaural sigue la periferia auricular que consiste el oído externo, el oído medio y el oído interno. No se tienen en cuenta los efectos del oído externo a los que no se aproxima el modelo esférico (por ej. la forma de los pabellones auriculares, el canal auditivo) y los efectos del oído medio. La selectividad espectral del oído interno se modela en forma de banco de filtros pasa banda superpuestos (indicado como *filtros auditivos* en la Fig. 10). Se utiliza un enfoque crítico para aproximar estos pasa bandas superpuestos por filtros rectangulares. Se calcula el ancho de banda rectangular equivalente (ERB) en función de la frecuencia central de conformidad con,

$$b(f_c) = 24.7 \cdot (0.00437 \cdot f_c + 1)$$

Se presume que el sistema auditivo humano tiene capacidad para ejecutar una alineación temporal para detectar componentes de señal coherentes y que se utiliza el análisis de correlación cruzada para el cálculo del tiempo de alineamiento τ (correspondiente a ITD) en presencia de sonidos complejos. Se evalúan desplazamientos en el tiempo de hasta aproximadamente 1– 1,5 kHz, de la señal portadora utilizando la correlación de relación cruzada de forma de onda, en tanto que, a frecuencias mayores, la correlación cruzada de envolvente se convierte en el indicio relevante. En lo sucesivo no hacemos esta distinción. Se modela la estimación de coherencia interaural (CI) como valor máximo absoluto de la función de correlación cruzada interaural normalizada

$$IC = \max_{\tau} \left| \frac{\langle p_L(t) \cdot p_R(t + \tau) \rangle}{\left[\langle p_L^2(t) \rangle \cdot \langle p_R^2(t) \rangle \right]^{\frac{1}{2}}} \right|$$

Algunos modelos de percepción binaural tienen en cuenta un análisis de correlación cruzada interaural en ejecución. Dado que consideramos señales estacionarias, no tomamos en cuenta la dependencia del tiempo. Para modelar la influencia del procesamiento de bandas críticas, computamos la función de correlación cruzada normalizada dependiente de la frecuencia de la siguiente manera

$$IC(f_c) = \frac{\langle A \rangle}{[\langle B \rangle \cdot \langle C \rangle]^{\frac{1}{2}}}$$

donde A es la función de correlación cruzada por banda crítica y B y C son las funciones de autocorrelación por banda crítica. Su relación con el dominio de la frecuencia por el espectro cruzado de pasa banda y los autoespectros de pasa banda se puede formular de la siguiente manera:

$$A = \max_{\tau} \left| 2 \operatorname{Re} \left(\int_{f_-}^{f_+} L^*(f) R(f) e^{j2\pi f(t-r)} df \right) \right|,$$

$$B = \left| 2 \left(\int_{f_-}^{f_+} L^*(f) L(f) e^{j2\pi f t} df \right) \right|,$$

$$C = \left| 2 \left(\int_{f_-}^{f_+} R^*(f) R(f) e^{j2\pi f t} df \right) \right|,$$

donde $L(f)$ y $R(f)$ son las transformadas de Fourier de las señales de entrada al oído, $f^{\pm} = f_c \pm \frac{b(f_c)}{2}$ son los límites de integración superior e inferior de la banda crítica de acuerdo con la frecuencia central real y * denota un conjugado complejo.

Si las señales de dos o más fuentes en diferentes ángulos se superponen, se evocan indicios fluctuantes ILD e ITD. Dichas variaciones de ILD e ITD en función del tiempo y/o la frecuencia pueden generar espaciosidad. Sin embargo, en el promedio a largo plazo, puede no haber ILDs y ITDs en un campo sonoro difuso. Una ITD promedio de cero significa que la correlación entre las señales puede no incrementarse por el alineamiento temporal. Las ILD pueden ser evaluadas, en principio, en todo el rango completo de frecuencia audible. Dado que la cabeza no constituye un obstáculo a frecuencias bajas, las ILD son más eficientes a las frecuencias media y alta.

Seguidamente se describe la Fig. 11A y 11B para ilustrar una implementación alternativa del analizador sin utilizar una curva de referencia como se describe en el contexto de la Fig. 10 o Fig. 4.

Se aplica una transformada de Fourier de corto tiempo (STFT) a los canales de audio envolventes $x_1(n)$ a $x_N(n)$, para dar los espectros de corto tiempo $X_1(m, i)$ a $X_N(m, i)$, respectivamente, donde m es el índice del espectro (tiempo) e i el índice de frecuencia. Se computan los espectros de una mezcla descendente estéreo de la señal de entrada envolvente, indicados por $\bar{X}_1(m, i)$ y $\bar{X}_2(m, i)$. En el caso de envolvente 5.1, es adecuada una mezcla descendente ITU como ecuación (1). $X_1(m, i)$ a $X_5(m, i)$ corresponden, en este orden, a los canales izquierdo (L), derecho (R), central (C), envolvente izquierdo (LS), y envolvente derecho (RS). En lo sucesivo se omiten los índices de tiempo y frecuencia la mayor parte del tiempo para abreviar la notación.

Basándose en la señal estéreo con mezcla descendente, se computan los filtros W_D y W_A para obtener los cálculos de señal envolvente de sonido directo y ambiental en las ecuaciones (2) y (3).

Dada la presunción de que la señal de sonido ambiental no se correlaciona entre todos los canales de entrada, elegimos los coeficientes de mezcla descendente para que esta presunción se aplique también a los canales de mezcla descendente. Por consiguiente, podemos formular el modelo de la señal de mezcla descendente en la ecuación 4.

D_1 y D_2 representan los espectros STFT de sonido directo correlacionados y A_1 y A_2 representan el sonido ambiental no correlacionado. Se presume además que el sonido directo y ambiental en cada canal no están mutuamente correlacionados.

La estimación del sonido directo, en el sentido del mínimo del cuadrado medio, se obtiene aplicando un filtro de Wiener a la señal envolvente original para suprimir la ambiente. Para derivar un filtro único que se pueda aplicar a todos los canales de entrada, estimamos los componentes directos de la mezcla descendente usando el mismo filtro para el canal izquierdo y derecho como en la ecuación (5).

La función de error cuadrático medio conjunta para esta estimación está dada por la ecuación (6).

$E\{\cdot\}$ es el operador de expectativa y P_D y P_A son las sumas de los cálculos de potencia a corto plazo de los componentes directo y ambiente, (ecuación 7).

La función de error (6) se reduce al mínimo fijando su derivada en cero. El filtro así obtenido para la estimación del sonido directo está en la ecuación 8.

5 De modo similar, el filtro de estimación para el sonido ambiente se puede derivar de acuerdo con la ecuación 9.

A continuación se derivan los cálculos estimativos correspondientes a P_D y P_A , necesarios para computar W_D y W_A . La correlación cruzada de la mezcla descendente está dada por la ecuación 10.

donde, dado el modelo de señal de mezcla descendente (4), se hace referencia a (11).

10 Suponiendo además que los componentes ambiente de la mezcla descendente tienen la misma potencia en el canal de mezcla descendente izquierdo y derecho, se puede escribir la ecuación 12.

Sustituyendo por la ecuación 12 la última línea de la ecuación 10 y considerando la ecuación 13 se llega a las ecuaciones (14) y (15).

15 Como se describiera en el contexto de Fig. 4, se puede imaginar la generación de las curvas de referencia correspondientes a una correlación mínima colocando dos o más fuentes de sonido directo en una configuración de reproducción y colocando una cabeza de oyente en una determinada posición en esta configuración de reproducción. A continuación, los parlantes emiten señales completamente independiente. En el caso de una configuración de dos parlantes, los dos canales tendrían que estar completamente no relacionados con una correlación igual a 0 en caso de que no haya ningún producto de entrecruzamiento. Sin embargo, estos productos entrecruzados se producen debido al acople cruzado del lado izquierdo al lado derecho de un sistema auditivo humano, y también tienen lugar otros entrecruzamientos debido a las reverberaciones del recinto, etc.. Por lo tanto, las curvas de referencia obtenidas

20 ilustradas en la Fig. 4 o en las Figs. 9a a 9d no siempre son de 0, sino que tienen valores particularmente diferentes de 0 aunque las señales de referencia imaginadas en esta situación fueran completamente independientes. Sin embargo, es importante comprender que en realidad estas señales no son necesarias.

También es suficiente presumir una independencia total entre las dos o más señales al calcular la curva de referencia. En este contexto, cabe señalar, sin embargo, que se pueden calcular otras curvas de referencia correspondientes a otras situaciones, como por ejemplo usando o suponiendo señales que no son completamente independientes, sino que tienen una cierta dependencia o grado de dependencia mutua previamente conocido. Al calcular esa curva de referencia diferente, la interpretación o producción de los factores de ponderación sería diferente con respecto a una curva de referencia en la que se presumen señales completamente independientes.

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

30 La señal descompuesta de la invención puede ser almacenada en un medio de almacenamiento digital o se la puede transmitir por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión conectado por cables tal como la Internet.

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

40 Algunas realizaciones de acuerdo con la invención comprenden un transportador de datos no transitorio que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema de computación programable de tal manera que se ejecute uno de los métodos descritos en la presente

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

50 Otras realizaciones comprenden el programa de computación para ejecutar uno de los métodos aquí descritos, almacenado en un portador legible por una máquina.

55 En otras palabras, una realización del método de la invención consiste, por lo tanto, en un programa de computación que consta de un código de programa para realizar uno de los métodos aquí descritos al ejecutarse el programa de computación en una computadora.

Una realización adicional de los métodos de la presente invención consiste, por lo tanto, en un portador de datos (o medio de almacenamiento digital, o medio legible por computadora) que comprende, grabado en el mismo, el programa de computación para ejecutar uno de los métodos aquí descritos.

60 Otra realización del método de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa de computación para ejecutar uno de los métodos aquí descritos. El flujo de datos o la secuencia de señales puede estar configurada, por ejemplo, para ser transferida a través de una conexión de comunicación de datos, por ejemplo por la Internet.

Otra de las realizaciones comprende un medio de procesamiento, por ejemplo una computadora, un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los métodos aquí descritos

Otra de las realizaciones comprende una computadora en la que se ha instalado el programa de computación para ejecutar uno de los métodos aquí descritos.

5 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 métodos aquí descritos. En algunas realizaciones, una matriz de puertas programables en el campo puede cooperar con un microprocesador para ejecutar uno de los métodos aquí descritos. Por lo general, los métodos son ejecutados preferentemente por cualquier aparato de hardware.

10 Las realizaciones precedentemente descritas son meramente ilustrativas de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles aquí descritos han de ser evidentes para las personas con capacitación en la técnica. Por lo tanto, sólo es intención limitarse al alcance de las siguientes reivindicaciones de patente y no a los detalles específicos presentados a manera de descripción y explicación de las realizaciones aquí presentadas

15

REIVINDICACIONES

1. Un aparato para descomponer una señal de entrada (10) que tiene un número de por lo menos tres canales de entrada, que comprende:
- 5 un mezclador descendente (12) para efectuar la mezcla descendente de la señal de entrada a fin de obtener una señal con mezcla descendente, donde el mezclador descendente (12) está configurado para efectuar la mezcla descendente de manera que un número de canales con mezcla descendente de la señal de mezcla descendente (14) sea por lo menos 2 y menor que el número de canales de entrada;
- 10 un analizador (16) para analizar la señal de mezcla descendente para derivar un resultado del análisis (18) y un procesador de señales (20) para procesar la señal de entrada (10) o una señal (24) derivada de la señal de entrada, utilizando el resultado del análisis (18), donde el procesador de señales (20) está configurado para aplicar el resultado del análisis a los canales de entrada de la señal de entrada o los canales de la señal derivada de la señal de entrada para obtener una señal descompuesta (26), donde la señal derivada de la señal de entrada es diferente de la señal de mezcla descendente.
- 15 2. El aparato de acuerdo con la reivindicación 1, que además comprende un convertidor de tiempo /frecuencia (32) para convertir los canales de entrada en una secuencia de tiempo de representaciones de frecuencias de los canales, donde cada representación de frecuencia de los canales de entrada tiene una pluralidad de subbandas, o donde el mezclador descendente (12) comprende un convertidor de tiempo /frecuencia para convertir la señal de mezcla descendente, donde el analizador (16) está configurado para generar un resultado del análisis (18) correspondiente a las subbandas individuales y
- 20 donde el procesador de señales (20) está configurado para aplicar los resultados individuales del análisis a las subbandas correspondientes de la señal de entrada o de la señal derivada de la señal de entrada.
3. El aparato de acuerdo con la reivindicación 1 o 2, en el cual el analizador (16) está configurado para producir, como resultado del análisis, factores de ponderación ($W(m, i)$) y
- 25 donde el procesador de señales (20) está configurado para aplicar los factores de ponderación a la señal de entrada o a la señal derivada de la señal de entrada mediante la ponderación con los factores de ponderación.
4. El aparato de acuerdo con una de las reivindicaciones anteriores, en el cual el mezclador descendente está configurado para sumar canales de entrada ponderados o no ponderados de acuerdo con una norma de mezcla descendente que implica que por lo menos dos canales con mezcla descendente sean diferentes entre sí.
- 30 5. El aparato de acuerdo con una de las reivindicaciones anteriores, en el cual el mezclador descendente (12) está configurado para filtrar la señal de entrada (10) usando filtros basados en respuestas al impulso del recinto, filtros basados en respuestas al impulso binaural del recinto, (BRIR) o filtros basados en HRTF.
- 35 6. El aparato de acuerdo con una de las reivindicaciones anteriores, en el cual el procesador (20) está configurado para aplicar un filtro Wiener a la señal de entrada o a la señal derivada de la señal de entrada y en el cual el analizador (16) está configurado para calcular el filtro Wiener utilizando valores de expectativa derivados de los canales de mezcla descendente.
7. El aparato de acuerdo con una de las reivindicaciones anteriores, que además comprende un derivador de señales (22) para derivar la señal de la señal de entrada de manera que la señal derivada de la señal de entrada tenga un número diferente de canales en comparación con la señal de mezcla descendente o la señal de entrada.
- 40 8. El aparato de acuerdo con una de las reivindicaciones anteriores, en el cual el analizador (20) está configurado para usar una curva de similitud dependiente de la frecuencia previamente guardada que indica una similitud dependiente de la frecuencia entre dos señales que pueden ser generadas por señales de referencia conocidas de antemano.
- 45 9. El aparato de acuerdo con cualquiera de las reivindicaciones 1 a 8, en el cual el analizador está configurado para usar una curva de similitud dependiente de la frecuencia previamente guardada que indica una similitud dependiente de la frecuencia entre dos o más señales en una posición del oyente bajo la presunción de que las señales tienen una característica de similitud conocida y de que las señales pueden ser emitidas por los parlantes en las posiciones conocidas de los parlantes.
- 50 10. El aparato de acuerdo con una de las reivindicaciones 1 a 7, en el cual el analizador está configurado para calcular una curva de similitud dependiente de la frecuencia utilizando la potencia de corto tiempo dependiente de la frecuencia de los canales de entrada.
- 55 11. El aparato de acuerdo con cualquiera de las reivindicaciones 8 a 10, en el cual el analizador (16) está configurado para calcular una similitud del canal de mezcla descendente en una subbanda de frecuencia (80), a fin de comparar el resultado de similitud con una similitud indicada por la curva de referencia (82, 83) y generar el factor de ponderación sobre la base del resultado de la compresión como resultado del análisis, o bien
- 60 para calcular una distancia entre el correspondiente resultado y una similitud indicada por la curva de referencia correspondiente a la misma subbanda de frecuencia y para calcular, además, un factor de ponderación basándose en la distancia como resultado del análisis.

12. El aparato de acuerdo con una de las reivindicaciones anteriores, en el cual el analizador (16) está configurado para analizar los canales de mezcla descendente en subbandas determinadas por una resolución en la frecuencia del oído humano.

5 13. El aparato de acuerdo con una de las reivindicaciones 1 a 12, en el cual el analizador (16) está configurado para analizar la señal de mezcla descendente a fin de generar un resultado del análisis que da lugar a una descomposición de la ambiente directa y en el cual el procesador de señales (20) está configurado para extraer la parte directa o la parte ambiente utilizando el resultado del análisis.

10 14. Un método para descomponer una señal de entrada (10) que consta de un número de por lo menos tres canales de entrada, que comprende:

ejecutar la mezcla descendente (12) de la señal de entrada para obtener una señal de mezcla descendente, de manera que un número de canales de mezcla descendente de la señal de mezcla descendente (14) sea por lo menos 2 y menor que el número de canales de entrada;

15 analizar (16) la señal de mezcla descendente para derivar un resultado del análisis (18) y procesar (20) la señal de entrada (10) o una señal (24) derivada de la señal de entrada, utilizando el resultado del análisis (18), donde el resultado del análisis se aplica a los canales de entrada de la señal de entrada o a los canales de la señal derivada de la señal de entrada para obtener una señal descompuesta (26), donde la señal derivada de la señal de entrada es diferente de la señal de mezcla descendente.

20 15. Un programa de computación para poner en práctica el método de acuerdo con la reivindicación 14, al ejecutarse el programa de computación en una computadora o procesador.

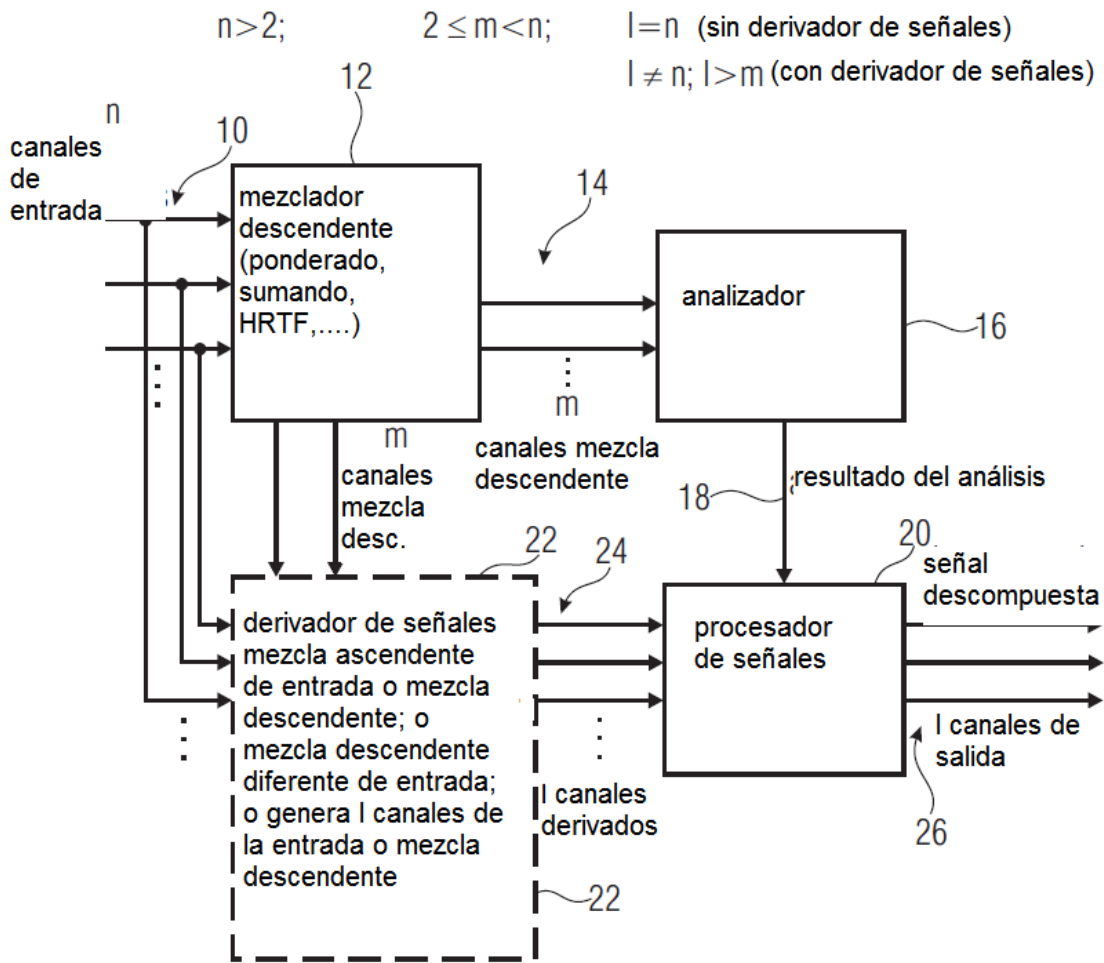


FIG 1

$$i \geq 2; \quad i \neq j; \quad k \geq j$$

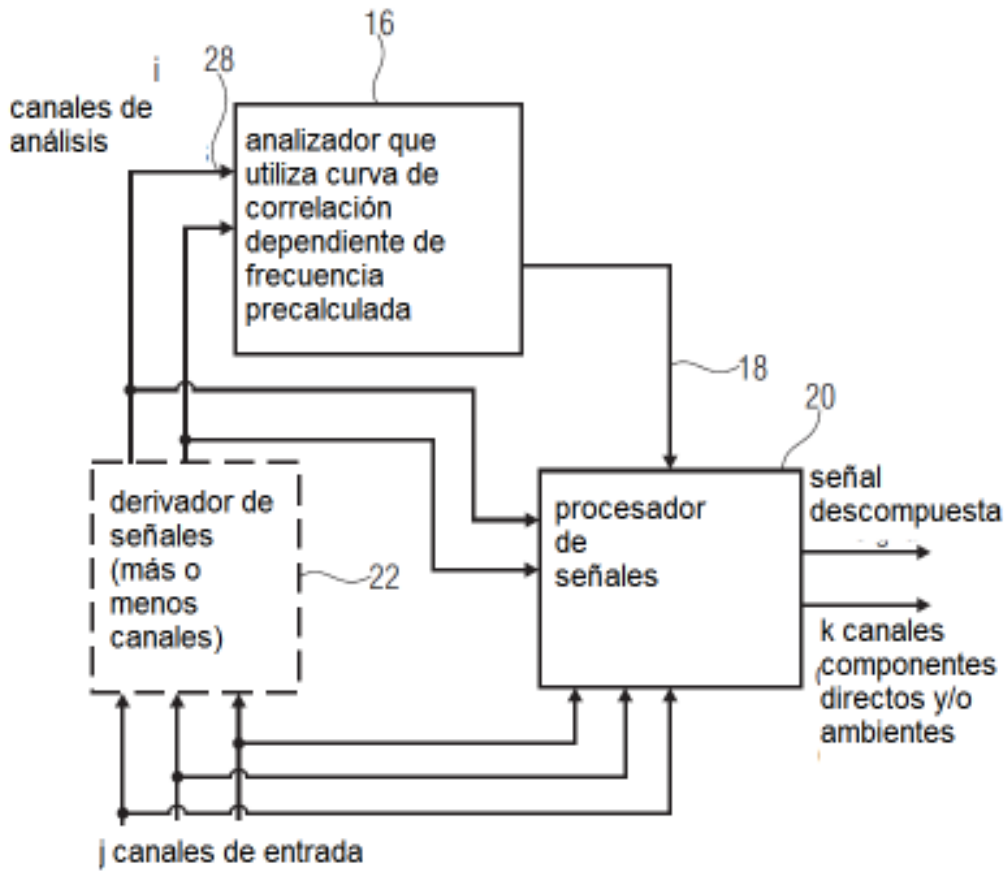


FIG 2

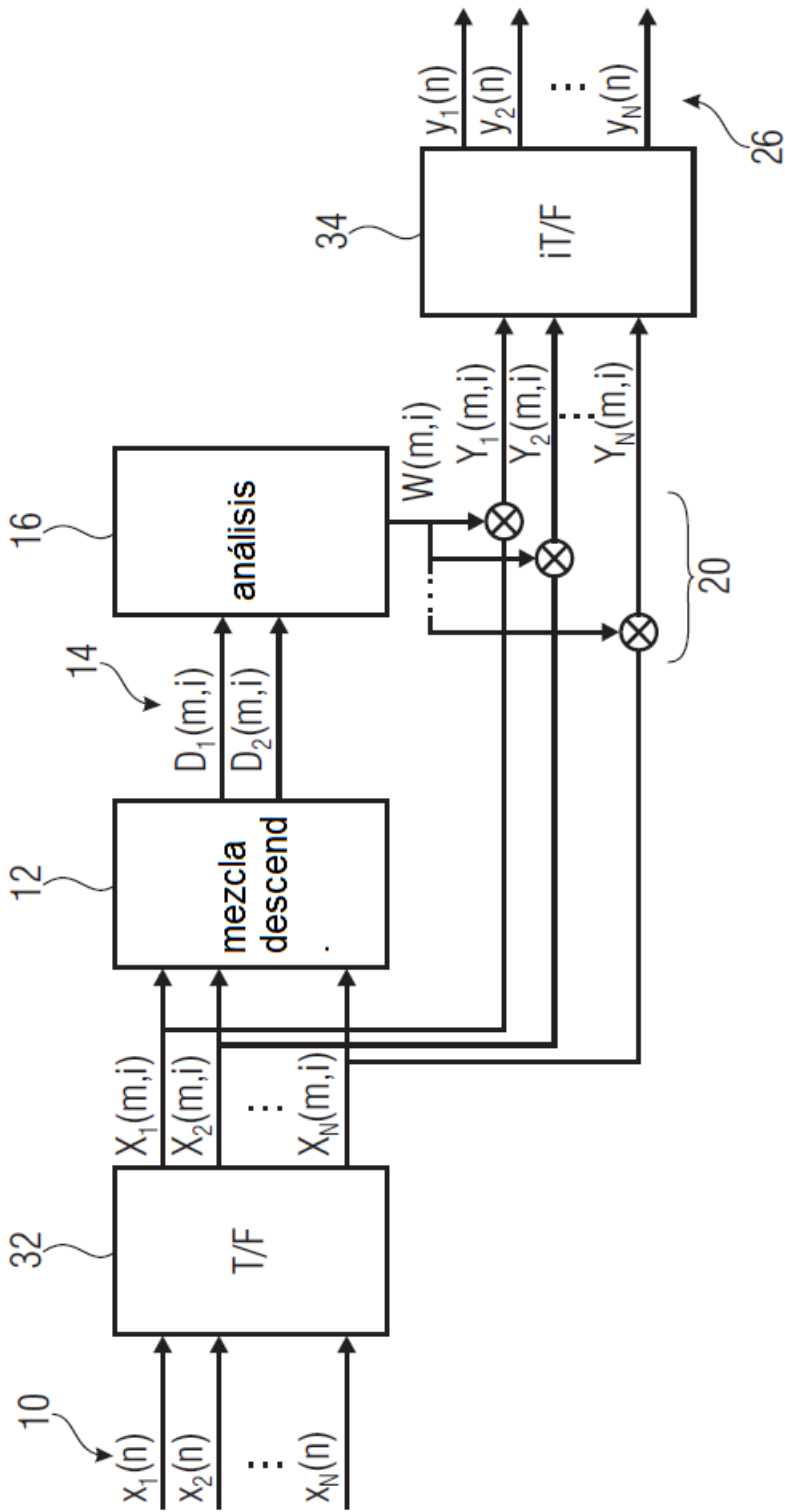


FIG 3

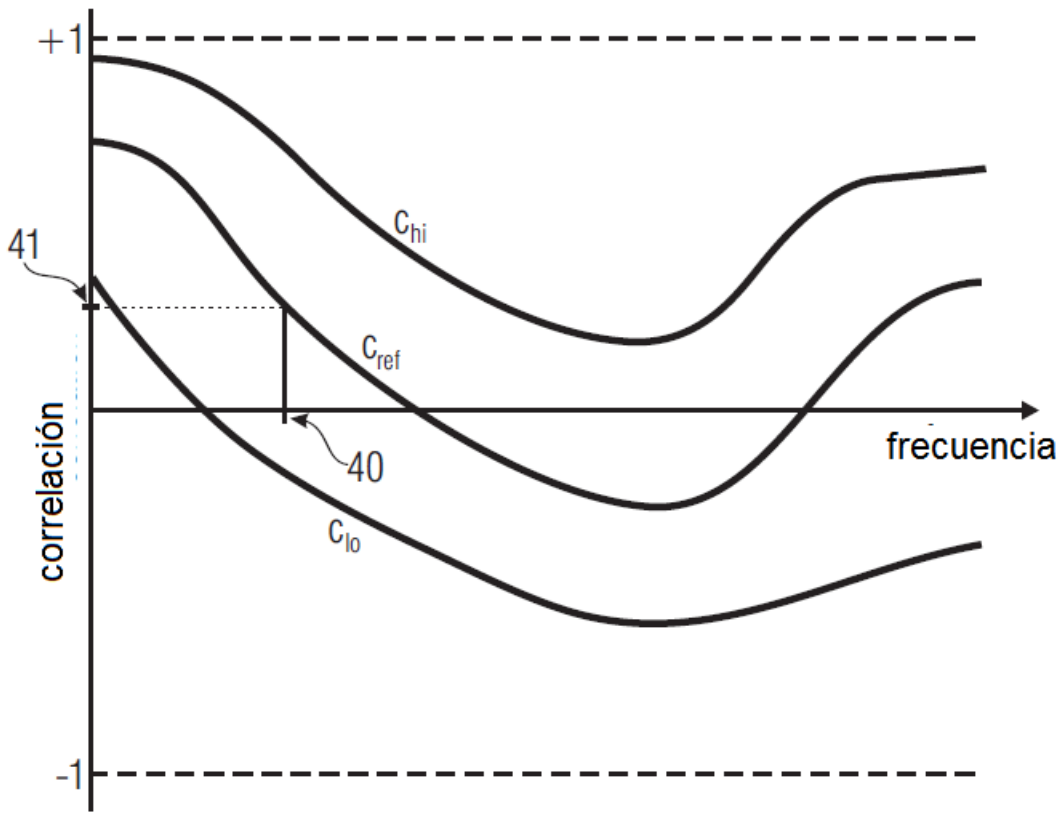


FIG 4

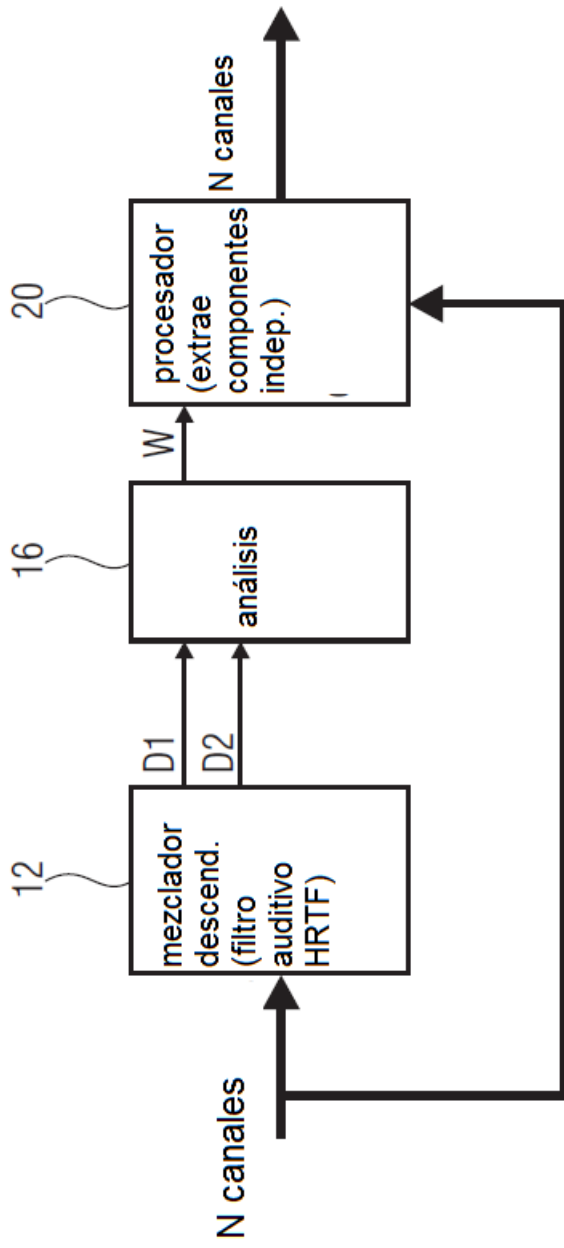


FIG 5

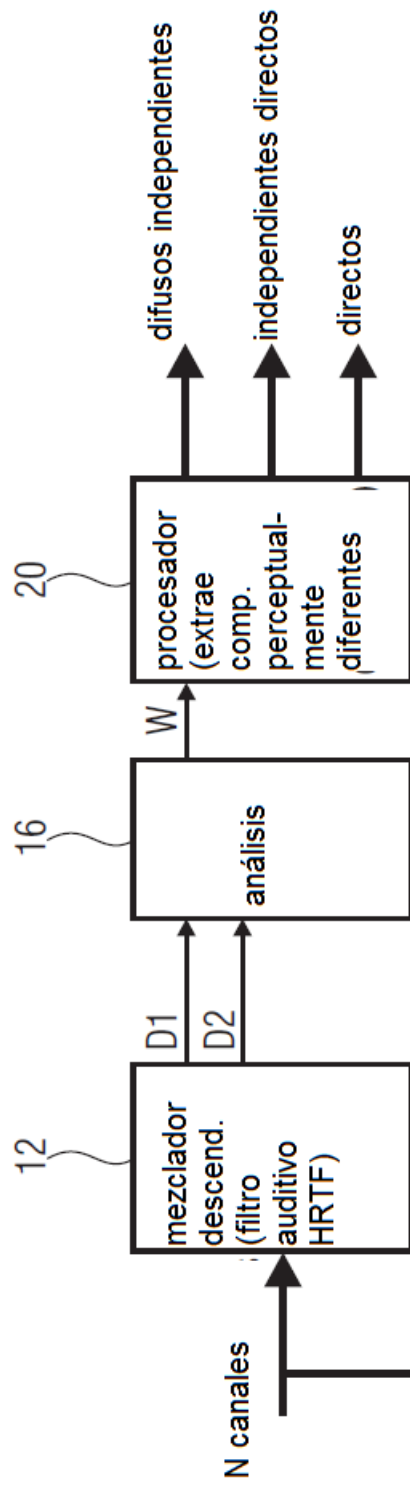


FIG 6

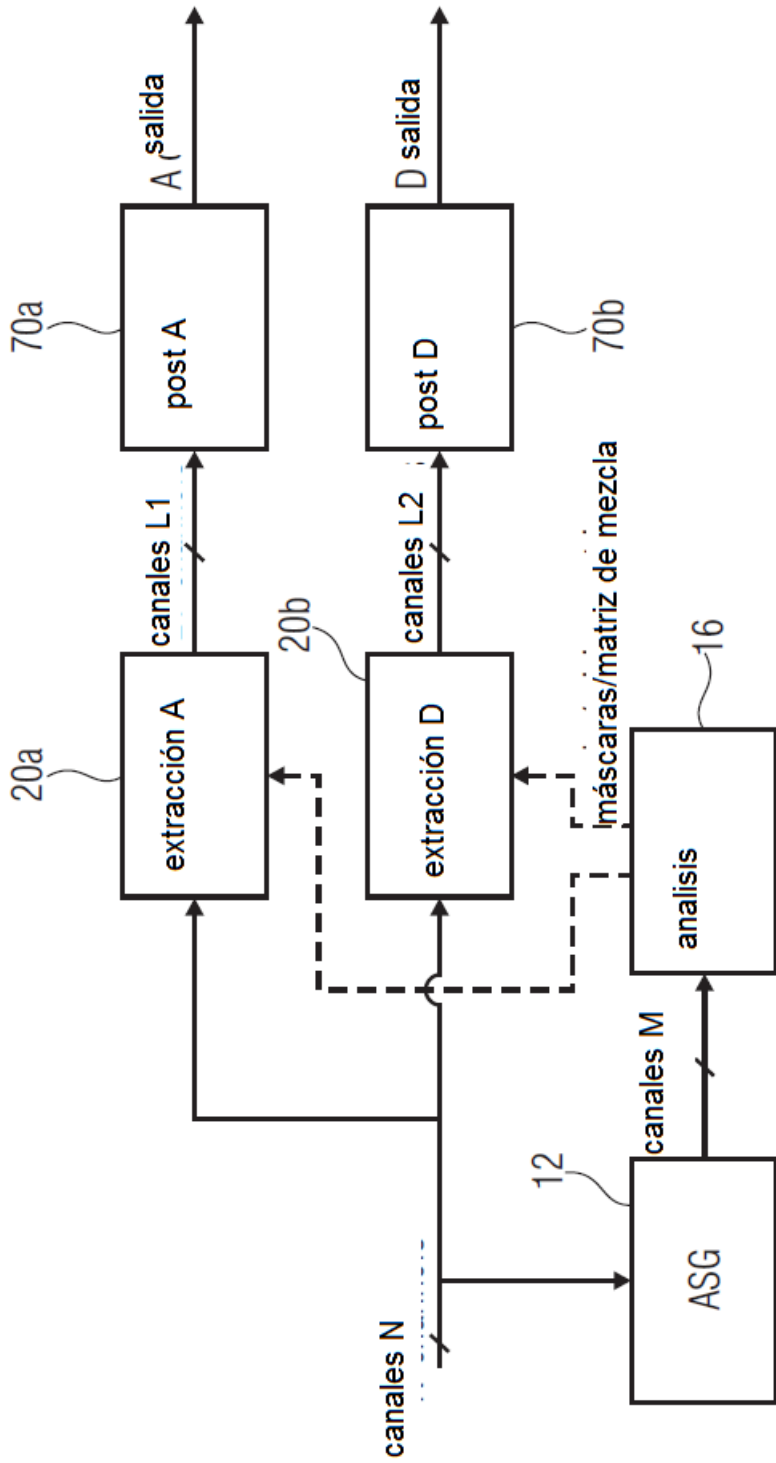


FIG 7

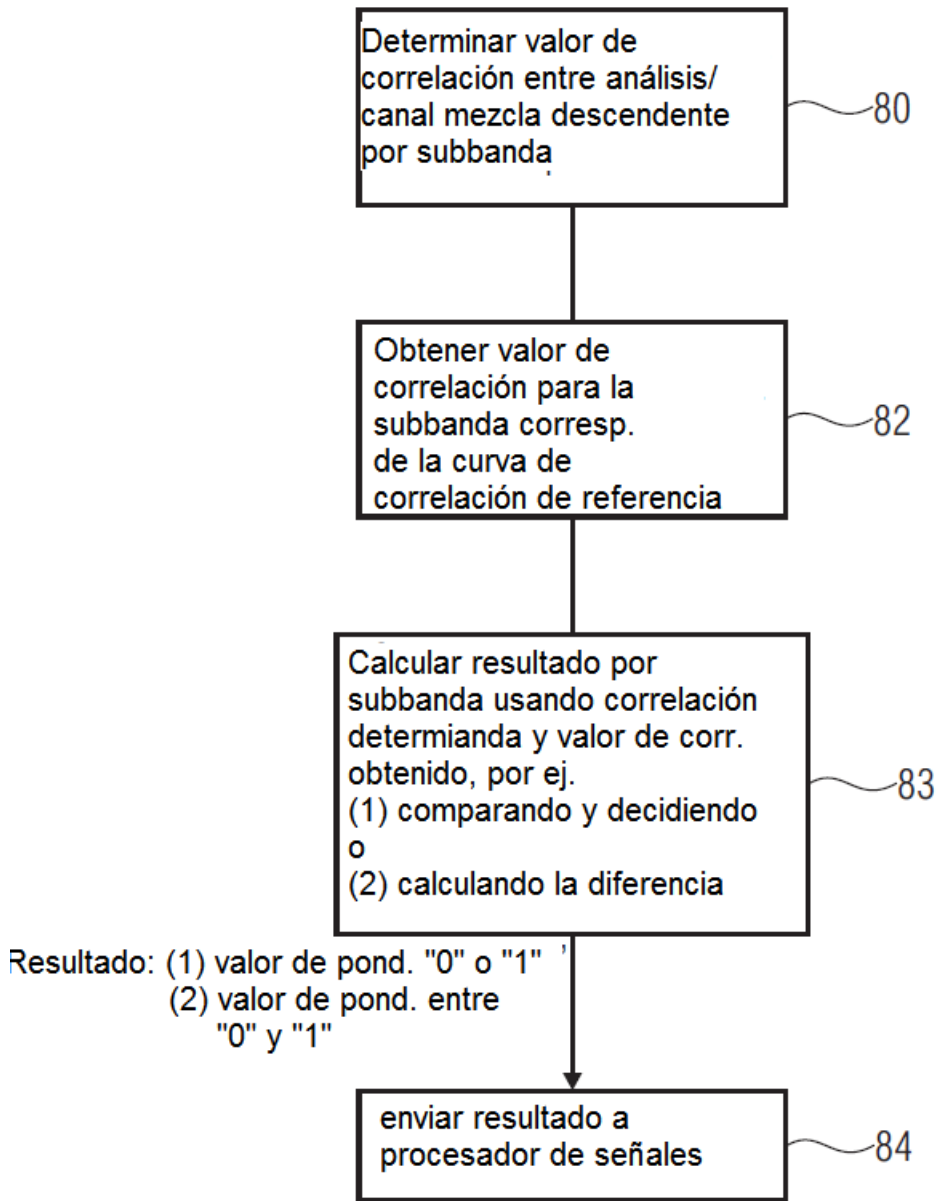
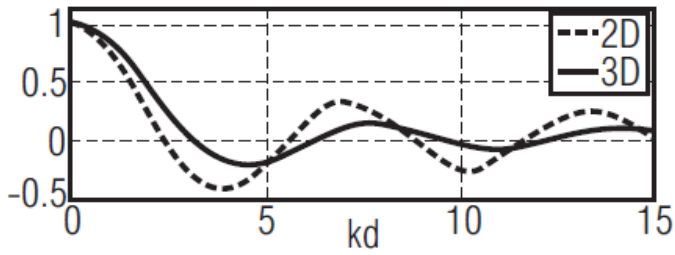
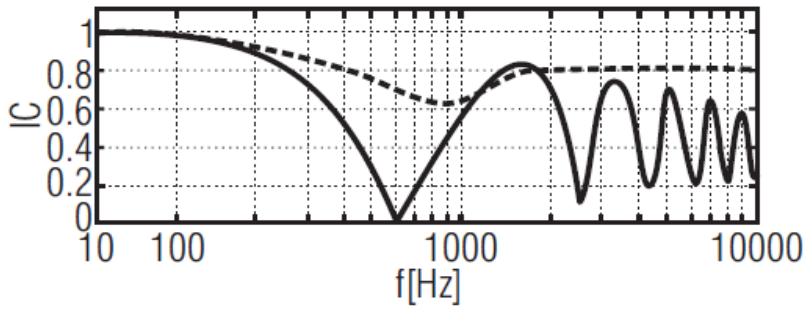


FIG 8



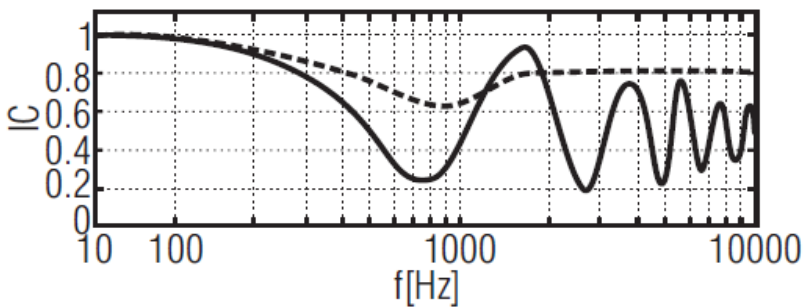
Corr. cruzada (r) en función de kd en campos sonoros difusos ideales bidimensionales y tridimensionales

FIG 9A



$IC(f_c)$ corresp. a dos fuentes sonoras azimut $\pm 30^\circ$, con orientación de cabeza 0°

FIG 9B



Dos fuentes sonoras a azimut $\pm 30^\circ$ con orientación de cabeza 25°

FIG 9C

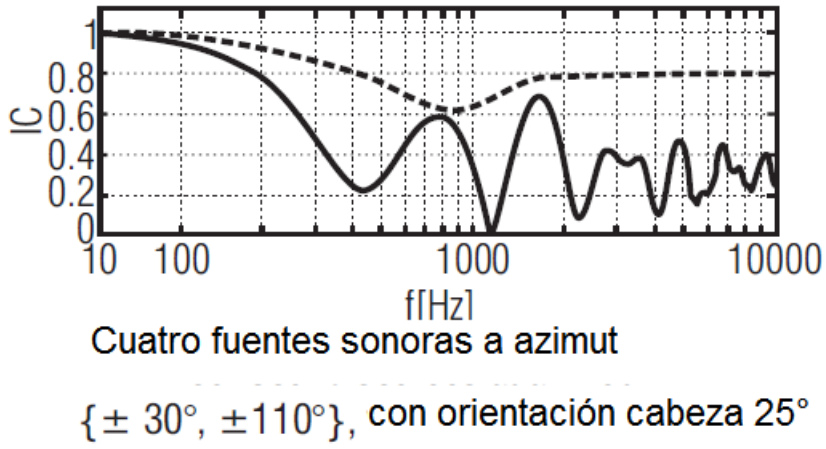


FIG 9D

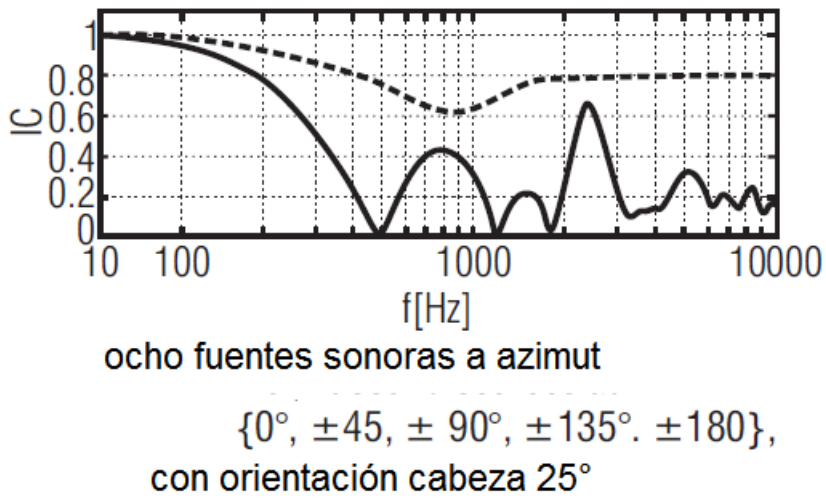


FIG 9E

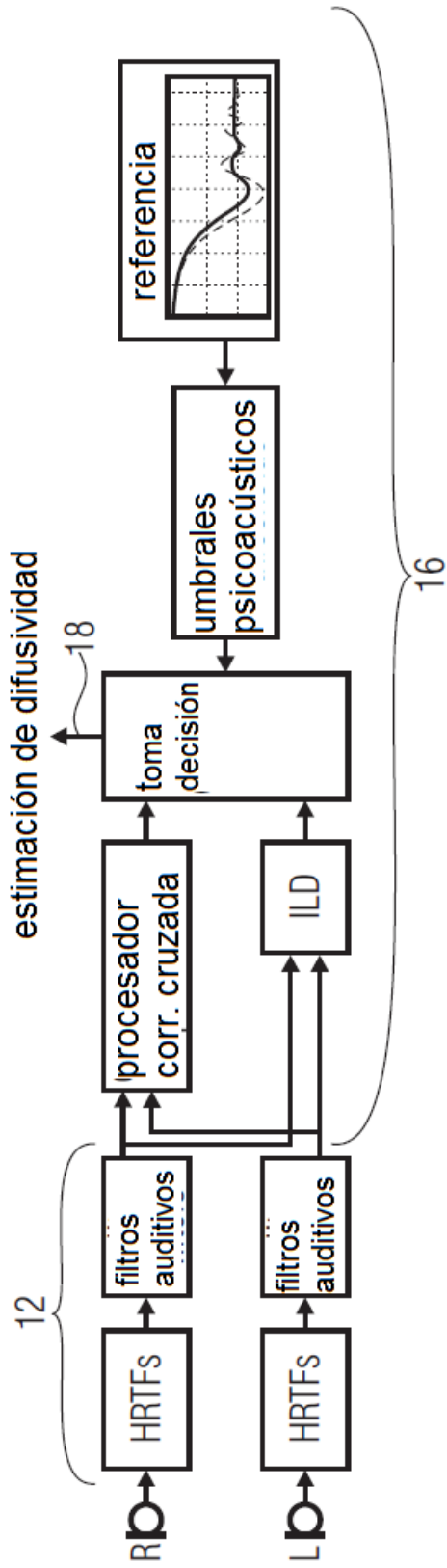


FIG 10

$$\begin{pmatrix} \bar{X}_1(m,i) \\ \bar{X}_2(m,i) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 1 & \frac{1}{\sqrt{2}} & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1(m,i) \\ \vdots \\ X_5(m,i) \end{pmatrix}, \quad (1)$$

$$\begin{pmatrix} Y_{1D}(m,i) \\ \vdots \\ Y_{ND}(m,i) \end{pmatrix} = W_D(m,i) \begin{pmatrix} X_1(m,i) \\ \vdots \\ X_N(m,i) \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} Y_{1A}(m,i) \\ \vdots \\ Y_{NA}(m,i) \end{pmatrix} = W_A(m,i) \begin{pmatrix} X_1(m,i) \\ \vdots \\ X_N(m,i) \end{pmatrix}, \quad (3)$$

$$\begin{aligned} \bar{X}_1 &= D_1 + A_1 \\ \bar{X}_2 &= D_2 + A_2, \end{aligned} \quad (4)$$

$$\begin{aligned} \hat{D}_1 &= W_D \cdot \bar{X}_1 \\ \hat{D}_2 &= W_D \cdot \bar{X}_2. \end{aligned} \quad (5)$$

$$\begin{aligned} J &= E\{|D_1 - \hat{D}_1|^2\} + E\{|D_2 - \hat{D}_2|^2\} \\ &= (W_d - 1)^2 P_D + W_d^2 P_A, \end{aligned} \quad (6)$$

$$\begin{aligned} P_D &= E\{|D_1|^2\} + E\{|D_2|^2\} \\ P_A &= E\{|A_1|^2\} + E\{|A_2|^2\}. \end{aligned} \quad (7)$$

$$W_D = \frac{P_D}{P_D + P_A}. \quad (8)$$

$$W_A = \frac{P_A}{P_D + P_A} = 1 - W_d. \quad (9)$$

FILTRADO DE WIENER PARA ANÁLISIS

FIG 11A

$$\begin{aligned}
\operatorname{Re}(E\{\check{X}_1\check{X}_2^*\}) &= \operatorname{Re}(E\{D_1+A_1\}(D_2^*+A_2^*)) \\
&= \operatorname{Re}(E\{D_1D_2^*\}) \\
&= \sqrt{E\{|D_1|^2\}E\{|D_2|^2\}}. \tag{10}
\end{aligned}$$

$$\begin{aligned}
E\{|D_1|^2\} &= E\{|\bar{X}_1|^2\} - E\{|A_1|^2\} \\
E\{|D_2|^2\} &= E\{|\bar{X}_2|^2\} - E\{|A_2|^2\}. \tag{11}
\end{aligned}$$

$$\begin{aligned}
E\{|D_1|^2\} &= E\{|\bar{X}_1|^2\} - \frac{P_A}{2} \\
E\{|D_2|^2\} &= E\{|\bar{X}_2|^2\} - \frac{P_A}{2}. \tag{12}
\end{aligned}$$

$$P_D + P_A = E\{|\bar{X}_1|^2\} + E\{|\bar{X}_2|^2\}, \tag{13}$$

$$P_D = \sqrt{(E\{|\bar{X}_1|^2\} - E\{|\bar{X}_2|^2\})^2 + 4\operatorname{Re}(E\{\bar{X}_1\bar{X}_2^*\})^2} \tag{14}$$

$$P_A = E\{|\bar{X}_1|^2\} + E\{|\bar{X}_2|^2\} - P_D. \tag{15}$$

FILTRADO DE WIENER PARA ANÁLISIS

FIG 11B