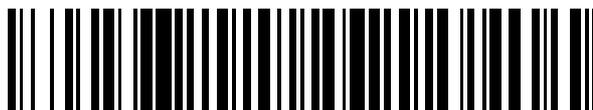


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 426 136**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 21/0272 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.05.2010** **E 10718175 (2)**

97 Fecha y número de publicación de la concesión europea: **31.07.2013** **EP 2427880**

54 Título: **Tanscodificador de formatos de audio**

30 Prioridad:

08.05.2009 EP 09006291

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
21.10.2013

73 Titular/es:

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

72 Inventor/es:

**THIERGART, OLIVER;
FALCH, CORNELIA;
KÜCH, FABIAN;
DEL GALDO, GIOVANNI;
HERRE, JÜRGEN y
KALLINGER, MARKUS**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 426 136 T3

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

DESCRIPCIÓN

Tanscodificador de formatos de audio

5 **[0001]** La presente invención es con el campo de transcodificación de formatos de audio, especialmente la transcodificación de formatos de codificación paramétricos.

[0002] Recientemente, varias técnicas paramétricas para la codificación de señales de audio de multicanal/multiobjetos han sido propuestas. Cada sistema tiene ventajas y desventajas únicas, sus características tales como el tipo de caracterización paramétrica, dependencia/independencia de una instalación de altavoces específica, etc. Diferentes técnicas, paramétricas son optimizadas para diferentes estrategias de codificación.

10 **[0003]** Como un ejemplo el formato de codificación de audio direccional (DIRAC) para la representación del sonido multi-canal está basado en una señal de mezcla descendente e información lateral que contienen los parámetros de dirección y difusividad para un número de sub-bandas de frecuencia. Debido a esta parametrización, el sistema de DIRAC puede ser usado para implementar fácilmente, por ejemplo filtración direccional y de esta manera para aislar sonido que se origina de una dirección particular en relación con un arreglo de micrófono usado para captar el
15 sonido. De esta manera, DIRAC puede también ser considerado como un extremo frontal acústico que es apto de cierto procesamiento espacial.

[0004] Como un ejemplo adicional, Spatial Audio Object Coding (SAOC) ISO/IEC, "MPEG audio technologies – Part. 2: Spatial Audio Object Coding (SAOC)", ISO/IEC JTC1/SC29/WG11 (MPEG) FCD 23003-2, J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC to SAOC – Recent Developments in Parametric Coding of Spatial Audio", 22nd
20 Regional UK AES Conference, Cambridge, UK, April 2007, J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers and W. Oomen: "Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124th AES Convention, Amsterdam 2008, Preprint 7377, es un sistema de codificación paramétrico que representa escenas de audio que contienen múltiples objetos de audio de una manera eficiente en velocidad de bits.

25 **[0005]** Aquí, la representación está basada en una señal de mezcla descendente e información lateral paramétrica. En contraste con DIRAC, que tiene como objetivo representar la escena de sonido espacial original como fue captada por el arreglo de micrófonos, SAOC no tiene como objetivo para reconstruir una escena de sonido natural. En lugar de esto, un número de objetos de audio (fuentes de sonido) son transmitidos y son combinados en un descodificador de SAOC a una escena de sonido objetivo de acuerdo con las preferencias del usuario en la terminal
30 del descodificador, esto es, el usuario puede colocar y manipular libre e interactivamente cada uno de los objetos de sonido.

[0006] En general, en la reproducción y escucha de multi-canal, un usuario está rodeado por múltiples altavoces. Existen varios procedimientos para capturar señales de audio para instalaciones específicas. Un objeto general en la reproducción es reproducir la composición espacial de una señal grabada originalmente, esto es, el origen de una
35 fuente de audio individual, tal como la ubicación de una trompeta en una orquesta. Varias instalaciones de altavoces son bastante comunes y pueden crear diferentes impresiones espaciales. Sin usar técnicas de post-producción especiales, los montajes estéreo de dos canales comúnmente conocidos pueden solamente recrear eventos auditivos en una línea entre los dos altavoces. Esto es obtenido principalmente mediante la llamada "toma panorámica de amplitud", en donde la amplitud de la señal asociada con una fuente de audio es distribuida entre los dos altavoces dependiendo de la posición de la fuente de audio con respecto a los altavoces. Esto se hace usualmente durante la grabación o mezcla subsecuente. Esto es, una fuente de audio procedente de la izquierda
40 lejana con respecto a la posición de escucha será reproducida principalmente por el altavoz izquierdo, mientras que una fuente de audio enfrente de la posición de escucha será reproducida con amplitud (nivel) idéntica por ambos altavoces. Sin embargo, el sonido que emana de otras direcciones no puede ser reproducido.

45 **[0007]** Consecuentemente, al usar más altavoces que están colocados alrededor del usuario, se pueden cubrir más direcciones y se puede crear una impresión espacial más natural. La disposición física de altavoz de multi-canal probablemente más bien conocida es el estándar 5.1 (ITU-R775-1), que consiste de 5 altavoces, cuyos ángulos acimutales con respecto a la posición de escucha son predeterminados para ser de 0° , $\pm 30^\circ$ y $\pm 110^\circ$. Eso significa que durante la grabación o mezcla, la señal es confeccionada a aquella configuración de altavoces específica y
50 desviaciones de un montaje de reproducción del estándar dará como resultado calidad de reproducción disminuida.

[0008] También se han propuesto numerosos otros sistemas con números variables de altavoces ubicados en direcciones diferentes. Sistemas profesionales, especialmente en cines e instalaciones de sonido, también incluyen altavoces a diferentes alturas.

55 **[0009]** De acuerdo con los diferentes montajes de reproducción, varios procedimientos de grabación diferentes han sido diseñados y propuestos para los sistemas de altavoz mencionados previamente, con el fin de grabar y reproducir la impresión espacial en la situación de escucha como habría sido percibida en el medio ambiente de grabación. Una manera teóricamente ideal de grabar sonido espacial para un sistema de altavoces multi-canal

escogido sería usar el mismo número de micrófonos como altavoces. En tal caso, los patrones de directividad de los micrófonos deben también corresponder a la disposición física de altavoces, de tal manera que el sonido de cualquier dirección individual solamente grabado con un número pequeño de micrófonos (1, 2 o más). Cada micrófono está asociado a un altavoz específico. Mientras más altavoces son usados en la reproducción, más estrechos son los patrones de directividad de los micrófonos que tiene que ser. Sin embargo, los micrófonos direccionales estrechos son más bien caros y comúnmente tienen una respuesta de frecuencia no plana, degradando la calidad del sonido grabado de manera indeseable. Además, usando varios micrófonos con patrones de directividad demasiado amplia como entrada a la reproducción de multi-canal da como resultado una percepción auditiva coloreada y borrosa debido al hecho de que el sonido que emana de una sola dirección siempre sería reproducido con más altavoces que lo que es necesario ya que sería grabado con micrófonos asociados a diferentes altavoces. En general, los micrófonos actualmente disponibles son mejor apropiados para grabaciones y reproducciones de dos canales, esto es, están diseñados sin el objetivo de una reproducción de una impresión espacial de los alrededores.

[0010] Desde el punto de vista de diseño de micrófono, se han discutido varios procedimientos para adaptar los patrones de directividad de micrófonos a las demandas en reproducción de audio espacial. En general, todos los micrófonos capturan sonido diferentemente dependiendo de la dirección de llegada del sonido al micrófono. Esto es, los micrófonos tienen una sensibilidad diferente, dependiendo de la dirección de llegada del sonido grabado. En algunos micrófonos, este efecto es menor, ya que capturan sonido casi independientemente de la dirección. Estos micrófonos son llamados en general micrófonos omnidireccionales. En un diseño de micrófono típico, un diafragma secular es anexado a una envolvente hermética al aire pequeña. Si el diafragma no es anexado al envolvente y el sonido llega al mismo igualmente desde cada lado, su patrón direccional tiene dos lóbulos. Esto es, tal micrófono captura sonido con igual sensibilidad de la parte frontal como de la parte posterior del diafragma, sin embargo con polaridades inversas. Tal micrófono no captura sonido procedente de la dirección coincidente al plano del diafragma, esto es, perpendicular a la dirección de sensibilidad máxima. Tal patrón direccional es llamado dipolo o de figura de ocho.

[0011] Los micrófonos omnidireccionales pueden también ser modificados a micrófonos direccionales, utilizando una envolvente no hermética al aire para el micrófono. La envolvente está construida especialmente de tal manera que se permite que las ondas de sonido se propaguen a través de la envolvente y lleguen al diafragma, en donde algunas direcciones de propagación son preferidas, de tal manera que el patrón direccional de tal micrófono se convierte en un patrón entre omnidireccional y dipolos. Estos patrones pueden tener por ejemplo dos lóbulos. Sin embargo, los lóbulos pueden tener resistencia diferente. Algunos micrófonos comúnmente conocidos tienen patrones que tienen solamente un lóbulo. El ejemplo más importante es el patrón cardioide, en donde la función direccional D puede ser expresada como $D = 1 + \cos(\theta)$, θ es la dirección de llegada del sonido. La función direccional tal cuantifica, que fracción de amplitud del sonido entrante es capturada, dependiendo de la dirección diferente.

[0012] Los patrones omnidireccional discutidos previamente también son llamados patrones de orden 0 y los otros patrones mencionados previamente (dipolo y cardioide) son llamados patrones de primer orden. Todos los diseños de micrófono discutidos previamente no permiten la formación arbitraria de los patrones de directividad, puesto que su patrón de directividad es determinado completamente por la construcción mecánica.

[0013] Para superar parcialmente el problema, algunas estructuras acústicas especializadas han sido diseñadas, que pueden ser usadas para crear patrones direccionales más estrechos que aquellos de los micrófonos de primer orden. Por ejemplo, cuando un tubo con agujeros en el mismo es anexado a un micrófono omnidireccional, se puede crear un micrófono con patrón direccional estrecho. Estos micrófonos son llamados micrófonos de escopeta o rifle. Sin embargo, comúnmente no tienen una respuesta de frecuencia plana, esto es, el patrón de directividad es estrechado al costo de la calidad del sonido grabado. Además, el patrón de directividad es predeterminado por la construcción geométrica y así, el patrón de directividad de una grabación efectuada con tal micrófono no puede ser controlado después de la grabación.

[0014] Por consiguiente, otros procedimientos han sido propuestos para parcialmente permitir o alterar el patrón de directividad después de la grabación real. En general, esto depende de la idea básica de grabación de sonido con un arreglo de micrófonos omnidireccionales o direccionales y aplicar el procesamiento de señal después de esto. Varias de tales técnicas han sido propuestas recientemente. Un ejemplo bastante simple es grabar sonido con dos micrófonos omnidireccionales, que son colocados cercanos entre sí y restar ambas señales entre sí. Esto crea una señal de micrófono virtual que tiene un patrón direccional equivalente a un dipolo.

[0015] En otros esquemas más sofisticados, las señales de micrófono pueden también ser retardadas o filtradas antes de sumarlas. Utilizando la formación, una señal correspondiente a un haz estrecho es formada mediante la filtración de cada señal del micrófono con un filtro diseñado especialmente y sumar las señales después de la filtración (formación de haz de suma de filtros). Sin embargo, estas técnicas son ciegas a la señal misma, esto es, no están conscientes de la dirección de llegada del sonido. Así, un patrón direccional predeterminado puede ser definido, que es independiente de la presencia real de una fuente de sonido en la dirección predeterminada. En general, la estimación de la "dirección de llegada" del sonido es una tarea por sí misma.

[0016] En general, numerosas características direccionales espaciales diferentes pueden ser formadas con las técnicas anteriores. Sin embargo, la formación de patrones de sensibilidad selectivos especialmente de manera arbitraria (esto es, formación de patrones direccionales estrechos) requiere un gran número de micrófonos.

5 **[0017]** Una manera alternativa para crear grabaciones de multi-canal es colocar un micrófono cerca de cada fuente de sonido (por ejemplo, un instrumento) a ser grabada y recrear la impresión espacial al controlar los niveles de las señales de micrófono de acercamiento en la mezcla final. Sin embargo, tal sistema demanda un gran número de micrófonos y mucha interacción del usuario para crear la mezcla descendente final.

10 **[0018]** Un procedimiento para superar el problema anterior es DIRAC, que puede ser usado con diferentes sistemas de micrófono y que es apto de grabar sonido para reproducción con montajes de altavoces arbitrarios. El propósito de DIRAC es reproducir la impresión espacial de un medio ambiente acústico tan precisamente como sea posible, usando un sistema de altavoces de multicanal que tiene una instalación geométrica arbitraria. En este medio ambiente de grabación, las respuestas del medio ambiente (que puede ser sonido grabado continuo o respuesta de impulso) son medidas con un micrófono omnidireccional (w) y con un conjunto de micrófonos que permiten medir la dirección de llegada de sonido y la difusividad del sonido.

15 **[0019]** En los siguientes párrafos y en la solicitud, se comprenderá que el término “difusividad” es una medida para la no directividad de sonido. Esto es, el sonido que llegue en la posición de grabación o escucha con igual intensidad desde todas las direcciones, es difundido máximamente. Una manera común de cuantificar la difusión es usar valores de difusividad del intervalo $[0, \dots, 1]$, en donde un valor de 1 describe el sonido difundido máximamente y un valor de 0 describe sonido perfectamente direccional, esto es, sonido que llega desde una dirección claramente distinguibles solamente. Un procedimiento comúnmente conocido para medir la dirección de llegada del sonido es aplicar tres micrófonos de figura de 8 (X, Y, Z) alineados con ejes de coordenadas cartesianas. Micrófonos especiales, llamados “micrófonos de formato B”, han sido diseñados, que producen directamente todas las respuestas deseadas. Sin embargo, como se menciona anteriormente, las señales W, X, Y y Z pueden también ser calculadas de un conjunto de micrófonos omnidireccionales discretos.

25 **[0020]** En el análisis de DIRAC, una señal de sonido grabada es dividida en canales de frecuencia, que corresponden a la selectividad de frecuencias de la percepción auditiva humana. Esto es, la señal es por ejemplo procesada por un banco de filtros o una transformada de Fourier para dividir la señal en numerosos canales de frecuencia, que tienen un ancho de banda adaptado a la selectividad de frecuencia de la audición humana. Luego, las señales de banda de frecuencia son analizadas para determinar la dirección de origen de sonido y un valor de difusividad para cada canal de frecuencia con una resolución en el tiempo predeterminada. Esta resolución en el tiempo no tiene que ser fija y por supuesto, puede ser adaptada al medio ambiente de grabación. En DIRAC, uno o más canales de audio son grabados o transmitidos, junto con la dirección analizada y datos de difusividad.

35 **[0021]** En la síntesis o descodificación, los canales de audio finalmente aplicados a los altavoces pueden estar basados en el canal omnidireccional W (grabado con una alta calidad debido al patrón de directividad omnidireccional del micrófono usado) o el sonido para cada altavoz puede ser calculado como suma ponderada de W, X, Y y Z , formando así una señal que tiene una cierta característica direccional para cada altavoz. Correspondiente a la codificación, cada canal de audio es dividido en canales de frecuencia, que son opcionalmente divididos de manera adicional en corrientes difusas y no difusas, dependiendo de la difusividad analizada. Si la difusividad tiene que ser medida para ser alta, una corriente difusa puede ser reproducida utilizando una técnica que produce una percepción difusa del sonido, tal como las técnicas de descorrelación también usadas en codificación de pista binaural.

45 **[0022]** El sonido no difuso el sonido es reproducido utilizando una técnica como objetivo reproducir una fuente de audio virtual semejante a punto, ubicada en la dirección indicada por los datos de dirección encontrados en el análisis, esto es, la generación de la señal de DIRAC. Esto es, la reproducción espacial no es confeccionada a un montaje de altavoces “ideal” específico, como en las técnicas del arte previo (por ejemplo 5.1). Este es particularmente el caso ya que el origen del sonido es determinado como parámetros de dirección (esto es, descrito por un vector) utilizando el conocimiento a cerca de los patrones de directividad en los micrófonos utilizados en la grabación. Como ya se discutió, el origen del sonido en el espacio tridimensional es parametrizado de manera frecuencia-selectiva. Como tal, la impresión direccional puede ser reproducida con alta calidad para montajes de altavoces arbitrarios, ya que la geometría del montaje de altavoces es conocida. DIRAC por consiguiente no está limitado a geometrías de altavoces especiales y permite en general, una reproducción espacial más flexible del sonido.

55 **[0023]** DIRAC, cf. Pulkki, V., provee un sistema para representar señales de audio espaciales en base a una o más señales de mezcla descendente más información lateral adicional. La información lateral describe, entre otros aspectos posibles la dirección de llegada del campo de sonido en el grado de su difusividad en un número de bandas de frecuencias, como se muestra en la figura 5. Otro ejemplo de un DirAc conocido se describe en MARKUS KALLINGER Y OTROS: "Spatial filtering using directional audio coding parameters", ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2009. ICASSP 2009. IEEE INTERNATIONAL CONFERENCE ON, PISCATAWAY, NJ, USA, 19 de Abril de 2009, páginas 217-220, XP031459205

[0024] La figura 5 ejemplifica una señal de DIRAC, que está compuesta de tres componentes direccionales como por ejemplo, señales de micrófono de figura de ocho X, Y, Z más una señal omnidireccional W. Cada una de las señales está disponible en el dominio de frecuencia, que es ilustrado en la figura 5 por múltiples planos apilados para cada una de las señales. En base a las cuatro señales, se puede llevar a cabo una estimación de la dirección y difusividad en los bloques 510 y 520, que ejemplifican dicha estimación de la dirección y la difusividad para cada uno de los canales de frecuencia. El resultado de estas estimaciones es dado por los parámetros $\theta(t,f)$, $\phi(t,f)$ y $\psi(t,f)$ que representan el ángulo de acimut, el ángulo de elevación y la difusividad para cada una de las capas de frecuencia.

[0025] La parametrización de DIRAC puede ser usada para implementar fácilmente un filtro espacial con una característica espacial deseada, por ejemplo, solamente hacer pasar sonido de la dirección de un parlante particular. Esto se puede obtener mediante la aplicación de una dirección/difusividad y opcionalmente ponderación dependiente de frecuencia a las señales de mezcla descendente como se ilustra en las figuras 6 y 7.

[0026] La figura 6 muestra un descodificador 620 para reconstrucción de una señal de audio. El descodificador 620 un comprende un selector de dirección 622 y un procesador de audio 624. De acuerdo con el ejemplo de la figura 6, una entrada de audio de multi-canal 626 grabada por varios micrófonos es analizada por un analizador de dirección 628 que deriva parámetros de dirección que indican una dirección de origen de una porción de las señales de audio, esto es, la dirección de origen de la porción de señal analizada. La dirección, de la cual la mayor parte de la energía es incidente al micrófono es escogida y la posición de grabación es determinada para cada porción de señal específica. Esto se puede hacer por ejemplo, también usando la técnica de micrófono de DIRAC descritas previamente. Otros procedimientos de análisis direccional basados en información de audio grabada puede ser usado para implementar el análisis. Como resultado, el analizador de dirección 628 deriva parámetros de dirección 630, que indica la dirección de origen de una porción de un canal de audio o de la señal de multi-canal 626. Además, el analizador direccional 628 puede ser operativo para derivar un parámetro de difusividad 632 para cada porción de señal, por ejemplo para cada intervalo de frecuencia o para cada cuadro de tiempo de la señal.

[0027] El parámetro de dirección 630 y opcionalmente, el parámetro de difusividad 632 son transmitidos al selector de dirección 620, que es implementado para seleccionar la dirección deseada para el origen con respecto a una posición de grabación o una porción reconstruida de la señal de audio reconstruida. La información en cuanto a la dirección deseada es transmitida al procesador de audio 624. El procesador de audio 624 recibe por lo menos un canal de audio 634, que tiene una porción para lo cual los parámetros de dirección han sido derivados. El por lo menos un canal modificado por el procesador de audio puede por ejemplo ser un mezcla descendente de la señal de multi-canal 626, generada mediante algoritmos de mezcla descendente multicanal convencionales. Un caso extremadamente simple sería la suma directa de las señales de la entrada de audio de multi-canal 626. Sin embargo, ya que el concepto no está limitado por el número de canales de entrada, todos los canales de entrada de audio 626 pueden ser procesados simultáneamente por el descodificador de audio 620.

[0028] El procesador de audio 624 modifica la porción de audio para derivar la porción reconstruida de la señal de audio reconstruida, en donde la modificación comprende incrementar una intensidad de una porción del canal de audio que tiene parámetros de dirección que indican una dirección de origen cercana a la dirección deseada de origen con respecto a otra porción del canal de audio que tiene parámetros de dirección que indican una dirección de origen alejada de la dirección de origen deseada. En el ejemplo de la figura 6, la modificación es efectuada al multiplicar un factor de escalamiento 636 (q) con la porción del canal de audio a ser modificada. Esto es, si la porción del canal de audio es analizada para ser originaria de una dirección cercana a la dirección deseada seleccionada, un factor de escalamiento grande 636 es multiplicado con la porción de audio. Así, en su salida 638, el procesador de audio emite una porción reconstruida de la señal de audio reconstruida correspondiente a la porción del canal de audio provisto en su entrada. Como se indica además por la línea discontinua en la salida 638 del procesador de audio 624, esto puede no solamente ser efectuado para una señal de mono-salida, sino también para señales de salida de multi-canal, para los cuales el número de canales de salida no es fijo o predeterminado.

[0029] En otras palabras, el descodificador de audio 620 toma su entrada de tal análisis direccional por ejemplo como se usa en DIRAC. Las señales de audio 626 de un arreglo de micrófonos pueden ser divididas en bandas de frecuencia de acuerdo con la resolución de frecuencia del sistema auditivo humano. La dirección del sonido y opcionalmente la difusividad de sonido es analizada dependiendo del tiempo en cada canal de frecuencia. Estos atributos son proporcionados además, por ejemplo ángulos de dirección de azimuth (azi) y elevación (ele) y como índice difusividad (Ψ), que varía entre cero y uno.

[0030] Luego, la característica direccional propuesta o seleccionada es impuesta sobre las señales adquiridas al usar una operación de ponderación sobre ellas, que depende de los ángulos de dirección (azih y ele) y opcionalmente, de la difusividad (Ψ). Evidentemente, esta ponderación puede ser especificada diferentemente para diferentes bandas de frecuencia y en general, variarán con respecto al tiempo.

[0031] La figura 7 muestra un ejemplo adicional basado en síntesis de DIRAC. En este sentido, el ejemplo de la figura 7 podría ser interpretado como una mejora de la reproducción de DIRAC, que permite controlar el nivel del sonido dependiendo de la dirección analizada. Esto hace posible enfatizar el sonido procedente de una o múltiples direcciones o suprimir el sonido de una o múltiples direcciones. Cuando es aplicado en reproducción de multicanal,

se obtiene un post-procesamiento de la imagen del sonido reproducida. Si solamente un canal es usado como salida, el efecto es equivalente al uso de un micrófono direccional con patrones direccionales arbitrarios durante la grabación de la señal. Como se muestra en la figura 7, la derivación de parámetros de dirección, también como la derivación de un canal de audio transmitida es ilustrada. El análisis es efectuado en base a canales de micrófono de formato BW, X, Y y Z, como por ejemplo como es grabado por un micrófono de campo de sonido.

[0032] El procesamiento es efectuado de cuadro en cuadro. Por consiguiente, las señales de audio continuas son divididas en cuadros, que son escalados por una en función de ventana para evitar discontinuidades en la frontera de cuadro. Los cuadros de señal en ventana son sometidos a una transformada de Fourier en un bloque de transformada de Fourier 740, que divide las señales del micrófono en N bandas de frecuencia. Por propósito de simplicidad, el procesamiento de una banda de frecuencia arbitraria será descrito en los siguientes párrafos, ya que las bandas de frecuencia restantes son procesadas equivalentemente. El bloque de transformada de Fourier 740 deriva coeficientes que describen la fuerza o intensidad de los componentes de frecuencia presentes en cada uno de los canales de micrófono de formato BW, X, Y y Z dentro del cuadro de ventana analizado. Estos parámetros de frecuencia 742 son introducidos al codificador de audio 744 para derivar un canal de audio y parámetros de dirección asociados. En el ejemplo mostrado en la figura 7, el canal de audio transmitido es escogido como el canal omnidireccional 746 de información en cuanto a la señal de todas direcciones. En base a los coeficientes 742 para las porciones omnidireccional y direccional de los canales de micrófono de formato B, se efectúa un análisis de dirección y difusividad por un bloque de análisis de dirección 748.

[0033] La dirección de origen del sonido para la porción analizada del canal de audio es transmitida a un decodificador de audio 750 para reconstruir la señal de audio junto con el canal omnidireccional 746. Cuando los parámetros de difusividad 752 están presentes, la trayectoria de señal es dividida en una trayectoria no difusa 754a y una trayectoria difusa 754b. La trayectoria no difusa 754a es escalada de acuerdo con el parámetro de difusividad, de tal manera que, cuando la difusividad Ψ es baja, la mayor parte de la energía o de la amplitud permanecerá en la trayectoria no difusa. Inversamente, cuando la difusividad es alta, la mayor parte de la energía será desplazada a la trayectoria difusa 754b. En la trayectoria difusa 754b, la señal es descorrelacionada o difundida utilizando descorrelacionadores 756a o 756b. La descorrelación puede ser efectuada utilizando técnicas convencionalmente conocidas, tales como convolución con una señal de ruido blanco, en donde la señal de ruido blanco puede diferir del canal de frecuencia a canal de frecuencia. En tanto que la descorrelación esté conservando energía, una salida final puede ser regenerada al simplemente agregar las señales de la trayectoria de señal no difusa 754a y la trayectoria de señal difusa 754b en la salida, puesto que las señales en las trayectorias de señal ya han sido, como se indica por el parámetro de difusividad Ψ .

[0034] Cuando la reconstrucción es efectuada para un montaje de multi-canal, la trayectoria de señal directa 754a también como la trayectoria de señal difusa 754b son divididas en un número de sub-trayectorias correspondientes a las señales de altavoz individuales en las posiciones divididas 758a y 758b. Para este fin, la división en la posición dividida 758a y 758b puede ser interpretada como equivalente a una mezcla ascendente de por lo menos un canal de audio a múltiples canales para una reproducción vía un sistema de altavoces que tiene múltiples altavoces.

[0035] Por consiguiente, cada uno de los múltiples canales tiene una porción de canal del canal de audio 746. La dirección de origen de las porciones de audio individuales es reconstruida por el bloque de regresión 760 que incrementa o disminuye adicionalmente la intensidad o la amplitud de las porciones de canal correspondientes a los altavoces usados para la reproducción. Para este fin, el bloque de redirección 760 requiere en general conocimiento acerca del montaje de altavoces usado para la reproducción. La redistribución (redirección) real y la derivación de los factores de ponderación asociados puede ser implementada por ejemplo utilizando técnicas que utilizan un panorámica de amplitud a base de vector. Al suministrar diferentes montajes de altavoces geométricos al bloque de redistribución 760, configuraciones arbitrarias de altavoces de reproducción pueden ser usadas en realizaciones sin pérdida de calidad de reproducción. Después del procesamiento, se efectúan múltiples transformadas de Fourier inversas sobre las señales de dominio de frecuencia por los bloque de transformada de Fourier inversa 762 para derivar una señal de dominio del tiempo, que puede ser reproducida por los altavoces individuales. Antes de la reproducción, se efectúa una técnica de superposición o traslape y adición por las unidades de suma 764 para concatenar los cuadros de audio individuales para derivar señales de dominio de tiempo continua, listas para ser reproducidas por los altavoces.

[0036] De acuerdo con el ejemplo mostrado en la figura 7, el procesamiento de señal de DIRAC es enmendado en que se introduce un procesador de audio 766 que es introducido para modificar la porción del canal de audio realmente procesada y que permite incrementar la intensidad de una porción del canal de audio que tiene parámetros direccional que indica una dirección de origen cercana a una dirección deseada. Esto se obtiene mediante la aplicación de un factor de ponderación adicional a la trayectoria de señal directa. Esto es, si la porción de frecuencia procesada se origina de la dirección deseada, la señal es enfatizada mediante al aplicar una ganancia adicional a aquella porción de señal específica. La aplicación de la ganancia puede ser efectuada antes del punto de división 758a, ya que el efecto contribuirá a todas las porciones del canal igualmente.

[0037] La aplicación del factor de ponderación adicional puede ser implementada dentro del bloque de redistribución 760 que, en aquel caso, aplica factores de ganancia de redistribución implementados por el factor de ponderación adicional.

[0038] Cuando se usa mejora direccional en la reconstrucción de una señal de multi-canal, la reproducción puede por ejemplo, ser efectuada al estilo de presentación de DIRAC, como se muestra en la figura 7. El canal de audio que ser reproducido es dividido en bandas de frecuencia igual a aquellas usadas en el análisis direccional. Estas bandas de frecuencias son luego divididas en corrientes cursos de agua, corriente difusa y una corriente no difusa. La corriente difusa es reproducida, por ejemplo, al aplicar el sonido a cada altavoz después de convolución con ráfagas de sonido blanco 30 ms. Las ráfagas de sonido son diferentes para cada altavoz. La corriente no difusa es aplicada a la dirección entregada del análisis direccional que e, por supuesto dependiente del tiempo. Para obtener una percepción direccional en sistemas de altavoces de multi-canal, se puede usar toma panorámica de amplitud de par en par simple o de triplete. Además cada canal de frecuencia es multiplicado por un factor de ganancia o factor de escalamiento, que depende de la dirección analizada. En términos generales, una función puede ser especificada, que define un patrón direccional deseada para reproducción. Este puede por ejemplo ser una sola dirección, que será enfatizada. Sin embargo, patrones direccionales arbitrarios pueden ser fácilmente implementados en línea con la figura 7.

[0039] En el siguiente procedimiento, se describe un ejemplo adicional como una lista de etapas de procesamiento. La lista está basada en la suposición de que el sonido es grabado con un micrófono de formato B y es luego procesado para escuchar con montajes de altavoz multi-canal o monofónicos utilizando presentación estilo DIRAC o prestación que suministra parámetros direccionales, que indican la dirección de origen de porciones del canal de audio.

[0040] En primer lugar, las señales de micrófono pueden ser divididas en bandas de frecuencia y ser analizadas en dirección y opcionalmente difusividad en cada banda dependiendo de la frecuencia. Como un ejemplo, la dirección puede ser parametrizada por un ángulo de acimut y un ángulo de elevación (azi, ele). En segundo lugar, se puede especificar una función F, que describe el patrón direccional deseado. La función puede tener una forma arbitraria. Depende comúnmente de la dirección. Además, también puede depender de la difusividad, si la información de difusividad está disponible. La función puede ser diferente para frecuencias diferentes y puede también ser alterada dependiendo del tiempo. En cada banda de frecuencia, un factor direccional q de la función F puede ser derivado para cada instancia de tiempo, que es usado para la ponderación (escalamiento) subsecuente de la señal de audio.

[0041] En tercer lugar, los valores de muestras de audio pueden ser multiplicados con los valores q de factores direccionales correspondientes a cada porción de tiempo y frecuencia para formar la señal de salida. Esto se puede hacer en una representación de dominio de tiempo y/o frecuencia. Además, este procesamiento puede por ejemplo sr implementado como parte de una presentación de DIRAC a cualquier número de canales de salida deseados.

[0042] Como se describió previamente, el resultado puede ser escuchado al usar un sistema de altavoces de multi-canal o m un sistema de altavoces monofónicos. Recientemente, técnicas paramétricas par la transmisión/almacenamiento eficiente en velocidad de bits de escenas de audio que contienen múltiples objetos de audio han sido propuestas, por ejemplo, Binaural Cue Coding (Type 1), cf. C. Faller and F. Baumgarte, "Binaural Cue Coding – Part II: Schemes and applications", IEEF Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003, or Joint Source Coding, cf. C. Faller, "Parametric Joint-Coding of Audio Sources", 120th AES Convention, Paris, 2006, Preprint 6752, and MPEG Spatial Audio Object Coding (SAOC) , cf. J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC to SAOC – Recent Developments in Parametric Coding of Spatial Audio", 22nd Regional UK AES Conference, Cambridge, UK, April 2007, J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers and W. Oomen: "Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124th AES Convention, Amsterdam 2008, Preprint 7377).

[0043] Estas técnicas tienen como objetivo reconstruir perceptualmente la escena de audio de salida deseada en lugar de por una coincidencia de forma de onda. La figura 8 muestra una vista general del tal sistema (aquí: MPEG SAOC). La figura 8 muestra una vista general del sistema de SAOC MPEG. El sistema comprende un codificador de SAOC 810, un descodificador de SAOC 820 opcionalmente y un procesador 830. El procesamiento general puede ser llevado a cabo de una manera selectiva en frecuencia, en donde el procesamiento definido en los siguientes se puede llevar a cabo en cada una de las bandas de frecuencia individuales. El codificador de SAOC es introducido con un número de (N) señales de objeto de audio de entrada, que son sometidos como mezcla descendente como parte del procesamiento del codificador de SAOC. El codificador de SAOC 810 emite la señal de mezcla descendente e información lateral. La información lateral extraída por el codificador de SAOC 810 representa las características de los objetos de audio de entrada. Para SAOC de MPEG, el objeto energizado para todos los objetos de audio son los componentes más significativos de la información lateral. En la práctica, en lugar de potencias de objeto absoluta, potencias relativas, llamada diferencias de nivel de objeto (OLD) son transmitidas. La coherencia/correlación entre pares de objetos es llamada coherencia de inter-objetos (IOC) y puede ser usada para describir las propiedades de los objetos de audio de entrada adicionales.

[0044] La señal de mezcla descendente y la información lateral pueden ser transmitidas o almacenadas. Para este fin, la señal de audio de mezcla descendente puede ser comprimida utilizando codificadores de audio perceptuales bien conocidos, tales como capar 2 o 3 de MPEG-1, también conocidos como MP3, codificación de audio avanzada de MPEG (AAC), etc.

[0045] En el extremo de recepción, el decodificador SAOC 820 trata conceptualmente de restaurar las señales de objeto originales, a las cuales es también referido como separación de objeto, utilizando la información lateral transmitida. Estas señales de objeto aproximadas son luego mezcladas a una escena objetivo representada por M canales de salida de audio utilizando una matriz de representación, que es aplicada por el procesador 830. Efectivamente, la separación de las señales de objeto nunca es ejecutada puesto que tanto la etapa de separación como la etapa de mezcla son combinadas a una sola etapa de transcodificación, que da como resultado una reducción enorme en complejidad de cálculo.

[0046] Tal esquema puede ser muy eficiente, tanto en términos de velocidad de bits de transmisión, solamente necesita transmitir unos pocos canales mezcla descendente más algo de información lateral en lugar de N señales de audio de objetos más información de representación o un sistema discreto y complejidad computacional, la complejidad de procesamiento se relaciona principalmente con el número de canales de salida en lugar del número de objetos de audio. Ventajas adicionales para el usuario en el extremo de recepción incluye la libertad de escoger un montaje de prestación de su elección, por ejemplo, reproducción mono, estéreo, surround, reproducción y la función de audífonos virtualizada, etc., y el elemento de interactividad del usuario: La matriz de representación y la escena de salida, puede se ajustados y cambiados interactivamente por el usuario de acuerdo con el deseo, preferencia personal u otros criterios, por ejemplo, ubicar los parlantes de un grupo conjuntamente en un área espacial para maximizar la discriminación de otros parlantes restantes. Esta interactividad es obtenida al proveer una interfaz de usuario del decodificador.

[0047] Un concepto de transcodificación convencional para transcodificación de SAOC a MPEG surround (MPS) para prestación de multicanal es considerado en lo siguiente. En general, la descodificación de SAOC se puede hacer al usar un proceso de transcodificación. MPEG SAOC presenta la escena de audio objetivo, que está compuesta de todos los objetos de audio individuales, a un montaje de reproducción de sonido de multi-canal al transcodificarlo al formato surround de MPEG relacionado, confróntese con cf. J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén, W. Oomen, K. Linzmeier, K.S. Chong: "MPEG Surround – The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding", 122nd AES Convention, Vienna, Austria, 2007, Preprint 7084.

[0048] De acuerdo con la figura 9, la información lateral de SAOC es analizada sintácticamente 910 y luego transcodificada 920 junto con datos suministrados por el usuario acerca de la configuración de reproducción y parámetros de presentación de objetos. Adicionalmente, los parámetros mezcla descendente de SAOC son acondicionados por un pre-procesador de mezcla descendente 930. Tanto la mezcla descendente procesada como la información lateral de MPS se pueden luego hacer pasar al decodificador de MPS 940 para la presentación final.

[0049] Los conceptos convencionales tienen la desventaja de que son ya sea fáciles de implementar como, por ejemplo para el caso de DIRAC, pero la información del usuario o prestación individual de usuario no puedan ser aplicados o son más complejos para implementar, sin embargo, proveen la ventaja de que la información del usuario puede ser considerada como, por ejemplo para SAOC.

[0050] Es el objeto de la presente invención proveer un concepto de codificación de audio que se pueden ser implementado fácilmente y permite manipulación individual del usuario.

[0051] El objeto es obtenido por un transcodificador de formato de audio de acuerdo con la reivindicación 1 y un procedimiento para transcodificación de formato de audio de acuerdo con la reivindicación 11.

[0052] Es un hallazgo de la presente invención que las capacidades de codificación de audio direccional y codificación objeto de audio espacial pueden ser combinados. Es también un hallazgo de la presente invención que componentes de audio direccionales pueden ser convertidos a medidas o señales de fuente de audio separadas. Realizaciones pueden proveer medios para combinar eficientemente las capacidades del DIRAC y el sistema de SAOC, Así creando un procedimiento que usa DIRAC como un extremo frontal acústico con su capacidad de filtración espacial integrada y usa este sistema para separar el audio entrante en objetos de audio, que son luego representados y presentado utilizando SAOC. Además, realizaciones pueden proveer la ventaja de que la conversión de una representación de DIRAC a una representación SAOC puede ser efectuada de una manera extremadamente eficiente al convertir los dos tipos de información lateral y preferiblemente en algunas realizaciones, dejando la señal de mezcla descendente sin tocar.

[0053] Realizaciones de la presente invención serán detalladas utilizando las figuras adjuntas en las cuales:

la figura 1 muestra una realización de un transcodificador de formato de audio;

la figura 2 muestra otra realización de un transcodificador de formato de audio;

la figura 3 muestra todavía otra realización de un transcodificador de formato de audio;

la figura 4a muestra una superposición de componentes de audio direccionales;

la figura 4b ilustra una función de peso ejemplar usada en una realización;

la figura 4c muestra una función de ventana ejemplar usada en una realización;

la figura 5 ilustra el estado del DIRAC del arte;

la figura 6 ilustra el análisis direccional del estado de la técnica;

la figura 7 ilustra ponderación direccional del estado de la técnica combinado con presentación de DIRAC;

5 la figura 8 muestra una vista general del sistema de SAOC de MPEG y

la figura 9 ilustra una transcodificación de SAOC a MPS del estado de la técnica.

[0054] La figura 1 muestra un transcodificador de formato de audio 100 para transcodificar una señal de audio de entrada, la señal de audio de entrada tiene por lo menos dos componentes de audio direccionales. El transcodificador de formato de audio 100 comprende un convertidor 110 para convertir la señal de entrada a una
 10 señal convertida, la señal convertida tiene una representación de señal convertida y una dirección de llegada de señal convertida. Además, el transcodificador de formato de audio 100 comprende un proveedor de posición 120 para proveer por lo menos dos posiciones espaciales de por lo menos dos fuentes de audio espaciales. Las por lo menos dos posiciones espaciales pueden ser conocidas a priori, esto es, por ejemplo dadas o introducidas por un usuario o determinadas o detectadas en base a la señal convertida. Además, el transcodificador de formato de audio
 15 100 comprende un procesador 130 para procesar la representación de señal convertida en base a por lo menos dos posiciones espaciales para obtener por lo menos dos medidas fuente de audio separadas.

[0055] Realizaciones pueden proveer medios para combinar eficientemente las capacidades de los sistemas de DIRAC y de SAOC. Otra realización de la presente invención es ilustrada en la figura 2. La figura 2 muestra otro transcodificador de formato de audio 100, en donde el convertidor 110 es implementado como una etapa de análisis
 20 de DIRAC 301. En realizaciones, el transcodificador de formato de audio 100 puede ser adaptado para transcodificación de una señal de entrada de acuerdo con una señal de DIRAC, una señal de formato B o una señal de un arreglo de micrófonos. De acuerdo con la realización ilustrada en la figura 2, DIRAC puede ser usado como un extremo frontal acústico para adquirir una escena de audio espacial utilizando un micrófono de formato B o alternativamente, un arreglo de micrófonos, como se muestra por la etapa o bloque de de análisis de DIRAC 301.

[0056] Como ya se mencionó anteriormente, en realizaciones, el transcodificador de formato de audio 100, el convertidor 110, el proveedor de posición 120 y/o el procesador 130 pueden ser adaptados para convertir la señal de entrada en términos de un número de sub-bandas de frecuencia y/o segmentos de tiempo o cuadros de tiempo.

[0057] En las realizaciones, el convertidor 110 puede ser adaptado para convertir la señal de entrada a la señal convertida que comprende además una medida de difusividad y/o confiabilidad por subbanda de frecuencia.

[0058] En la figura 2, la representación de la señal convertida es también marcada "señales de mezcla descendente". En la realización ilustrada en la figura 2, la parametrización de DIRAC fundamental de la señal acústica a dirección y opcionalmente medida de difusividad y confiabilidad dentro de cada sub-banda de frecuencia puede ser usada por el proveedor de posición 120, esto es, el bloque de "número de fuentes y cálculo de posición"
 30 304 para detectar las posiciones espaciales en las cuales las fuentes de audio están activas. De acuerdo con la línea punteada marcada "potencia de mezcla descendente" en la figura 2, las potencias de mezcla descendente puede ser provistas al proveedor de la posición 120.
 35

[0059] En la realización ilustrada en la figura 2, el procesador 130 puede usar las posiciones espaciales, opcionalmente otro conocimiento a priori, para implementar un conjunto de filtros espaciales 311, 312, 31N para los cuales factores de ponderación son calculados en el bloque 303 con el fin de aislar o separar cada fuente de audio.

[0060] En otras palabras, en realizaciones, el procesador 130 puede ser adaptado para determinar un factor de ponderación para cada uno de por lo menos dos fuentes de audio separadas. Además, en realizaciones, el procesador 130 puede ser adaptado para procesamiento de la representación de la señal convertida en términos de por lo menos dos filtros espaciales para aproximar por lo menos dos fuentes de audio aisladas con por lo menos dos señales fuente de audio separadas como las por lo menos dos medidas fuente de audio separadas. La medida de
 40 fuente de audio puede por ejemplo corresponder a respectivas señales o energía de señal.
 45

[0061] En la realización ilustrada en la figura 2, las por lo menos dos fuentes de audio son representadas más en general por N fuentes de audio y las señales correspondientes. Así, en la figura 2, se muestran N filtros o etapas de síntesis, esto es 311, 312,..., 31N. En estos N filtros espaciales, la mezcla descendente de DIRAC, esto es, los componentes omnidireccionales, da como resultado un conjunto de fuentes de audio separadas aproximadas, que pueden ser usadas como entrada a un codificador SAOC. En otras palabras, en realizaciones, las fuentes de audio separadas pueden ser interpretadas como objetos de audio distintos y subsecuentemente codificadas en un codificador de SAOC. Así, realizaciones del transcodificador de formato de audio 100 pueden comprender un codificador SAOC para codificar por lo menos dos señales fuente de audio separada para obtener una señal codificada de SAOC que comprende un componente de mezcla descendente de SAOC y un componente de
 50 información lateral de SAOC.
 55

[0062] Las realizaciones descritas anteriormente pueden llevar a cabo una secuencia discreta de filtración direccional de DIRAC y subsecuente codificación de SAOC, para los cuales, en el siguiente, una mejora estructural será introducida, que conduce a una reducción en complejidad de cálculo. Como se explica anteriormente, en general, N señales de fuente de audio separadas pueden ser reconstruidas en realizaciones que usan N bancos de filtros de síntesis de DIRAC, 311 a 31N y luego subsecuentemente ser analizadas utilizando los bancos de filtros de análisis de SAOC en el codificador de SAOC. Luego, el codificador de SAOC puede calcular una señal de suma/mezcla descendente otra vez a partir de las señales de objetos separadas. Además, el procesamiento de las muestras de señal reales puede ser computacionalmente más complejo que llevar a cabo cálculos en el dominio de parámetros, lo que puede suceder a una velocidad de toma de muestras mucho más baja y la cual será establecida en realizaciones adicionales.

[0063] Realizaciones pueden proveer con la ventaja de procesamiento extremadamente eficiente. Realizaciones pueden comprender las siguientes dos simplificaciones. En primer lugar, tanto DIRAC como SAOC se pueden poner en operación utilizando bancos de filtros que permiten sub-bandas de frecuencia esencialmente idénticas para ambos esquemas en algunas realizaciones. Preferiblemente, en algunas realizaciones, uno y el mismo banco de filtros es usado para ambos esquemas. En este caso, los bancos de filtros de síntesis de DIRAC bancos de filtros de análisis de SAOC pueden ser evitados dando como resultado complejidad computacional reducida y retardo algorítmico reducido. Alternativamente, las realizaciones pueden usar dos bancos de filtros diferentes, que proporcionan parámetros en una rejilla de subbanda de frecuencia comparable. Los ahorros en cálculos de bancos de filtros de tales realizaciones puede no ser tan alta.

[0064] En segundo lugar, en realizaciones, en lugar de calcular explícitamente las señales fuentes separadas, el efecto de la separación puede ser obtenido por cálculos de dominio de parámetros solamente. En otras palabras, en realizaciones, el procesador 130 puede ser adaptado para estimar una información de energía, por ejemplo una energía o energía normalizada, para cada uno de por lo menos dos fuentes de audio separadas como las por lo menos dos medidas de fuente de audio separadas. En realizaciones, la energía mezcla descendente de DIRAC puede ser calculada.

[0065] En realizaciones, para cada posición de fuente de audio deseada/detectada, la ponderación direccional/peso de filtración puede ser determinada dependiendo de la dirección y posiblemente difusividad y características de separación propuestas. En realizaciones, la energía para cada fuente de audio de las señales separadas puede ser estimada del producto de la energía de mezcla descendente y el factor de ponderación de energía. En realizaciones, el procesador 130 puede ser adaptado para convertir las energías de las por lo menos dos fuentes de audio separadas a OLD de SAOC.

[0066] Algunas realizaciones pueden llevar a cabo el procedimiento de procesamiento aerodinámico descrito anteriormente sin involucrar ningún procesamiento de las señales de mezcla descendente reales. Adicionalmente, en algunas realizaciones, también se pueden calcular las coherencias de inter-objeto (IOC). Esto se puede obtener al considerar la ponderación direccional y las señales de mezcla descendente todavía en el dominio transformado.

[0067] En realizaciones, el procesador 130 puede ser apto para calcular la IOC para por lo menos dos fuentes de audio separadas. En general, el procesador (130) puede ser apto para calcular la IOC para dos de cada una de las por lo menos dos fuentes de audio separadas. En realizaciones, el proveedor de posición 120 puede comprender un detector que es apto para detectar las por lo menos dos posiciones espaciales de las por lo menos dos fuentes de audio espaciales en base a la señal convertida. Además, el proveedor/detector de posición 120 puede ser apto para detectar las por lo menos dos posiciones espaciales mediante una combinación de múltiples segmentos de tiempo de señal de entrada subsecuentes. El proveedor/detector de posición 120 puede también ser apto para detectar las por lo menos dos posiciones espaciales en base a una estimación de probabilidad máxima en cuanto a la densidad espacial de energía. El proveedor/detector de posición 120 puede ser apto para detectar una multiplicidad de posiciones de fuentes de audio espaciales en base a la señal convertida.

[0068] La figura 3 ilustra otra realización de un transcodificador de formato de audio 100. Similar a la realización ilustrada en la figura 2, el convertidor 110 es implementado como una etapa de "análisis de DirAC" 401. Además, el proveedor/detector de posición 120 es implementado como la etapa de "cálculo de número y posición de fuentes" 404. El procesador 130 comprende la etapa de "cálculo del factor de ponderación" 403, una etapa para calcular energías de fuentes separadas 402 y una etapa 405 para calcular OLD de SAOC y la corriente de bits.

[0069] Otra vez, en la realización ilustrada en la figura 3, la señal es adquirida usando un arreglo de micrófonos o alternativamente, un micrófono de formato B y es alimentada a la etapa de "análisis de DirAC" 401. Este análisis proporciona una o más señales de mezcla descendente e información de subbanda de frecuencia para cada cuadro de tiempo de procesamiento que incluye valores estimativos de la energía de mezcla descendente instantánea y dirección. Adicionalmente, la etapa de "análisis de DirAC" 401 puede proveer una medida de difusividad y/o una medida de la confiabilidad de los valores estimativos de dirección. A partir de esta información y posiblemente otros datos, tales como la energía de mezcla descendente instantánea, se pueden calcular valores estimativos del número de fuentes de audio y su posición por el proveedor/detector de posición 120, la etapa 404, respectivamente, por ejemplo, al combinar mediciones de varios cuadros de tiempo de procesamiento que son subsecuentes en el tiempo.

- 5 **[0070]** El procesador 130 puede ser apto para derivar un factor de ponderación direccional para cada fuente de audio y su posición en la etapa 403 a partir de la posición de fuente estimada y la dirección y opcionalmente, los valores de difusividad y/o confiabilidad del cuadro de tiempo procesado. Al combinar primero los valores estimativos de energía de mezcla descendente y los factores de ponderación en 402, los OLD de SAOC pueden ser derivados en 405. También, se puede generar una corriente de bits de SAOC completa en algunas realizaciones. Adicionalmente, el procesador 130 puede ser apto para calcular los IOC de SAOC al considerar la señal de mezcla descendente y utilizar el bloque de procesamiento 405 en la realización ilustrada en la figura 3. En realizaciones, las señales de mezcla descendente y la información lateral de SAOC pueden luego ser almacenadas o transmitidas conjuntamente para la descodificación de SAOC o presentación.
- 10 **[0071]** La “medida de difusividad” es un parámetro que describe para cada binario de tiempo – frecuencia, que tan “difuso” está el campo de sonido. Sin pérdida de generalidad, es definido en el intervalo de [0, 1], en donde difusividad = 0 indica un campo de sonido perfectamente coherente, por ejemplo, una onda plana ideal, mientras que difusividad = 1 indica un campo de sonido plenamente difuso, por ejemplo, el obtenido con un gran número de fuentes de audio esparcidas espacialmente que emiten ruido mutuamente sin correlacionar. Se pueden emplear varias expresiones matemáticas como medida de difusividad. Por ejemplo, en Pulkki, V., “Directional audio coding in spatial sound reproduction and stereo upmixing,” in Proceedings of the AES 28th International Conference, pp. 251-258, Piteå, Sweden, June 30 – July 2, 2006, la difusividad es calculada por medio de un análisis de energía de las señales de entrada, que comparan la intensidad activa con la energía de campo de sonido.
- 15 **[0072]** En lo siguiente, se describirá la medida de confiabilidad. Dependiendo del estimador de la dirección de llegada utilizado, es posible derivar una métrica que expresa que tan confiable es cada valor estimativo de dirección en cada binario de tiempo – frecuencia. Esta información puede ser aprovechada tanto en la determinación del número como posición de las fuentes, también como en el cálculo de los factores de ponderación en las etapas 403 y 404 respectivamente.
- 20 **[0073]** En lo siguiente, se detallarán realizaciones del procesador 130, esto es, también la etapa de “cálculo del número y posición de las fuentes” 404. El número y posición de las fuentes de audio para cada cuadro de tiempo pueden ya sea ser un conocimiento a priori, esto es una entrada externa o estimados automáticamente. Para el último caso, varios procedimientos son posibles. Por ejemplo, en algunas realizaciones se puede usar un estimador de probabilidad máxima en la densidad espacial de energía. El último puede calcular la densidad de energía de la señal de entrada con respecto a la dirección. Al suponer que las fuentes de sonido exhiben una distribución de von Mises, es posible estimar cuantas fuentes existen y en donde están ubicadas al escoger la solución con probabilidad más alta. Una distribución espacial de energía ejemplar es ilustrada en la figura 4a.
- 25 **[0074]** La figura 4a ilustra una vista gráfica de una densidad espacial de energía, ejemplificada por dos fuentes de audio. La figura 4a muestra la energía relativa en dB en las ordenadas y el ángulo azimuth en las abscisas. Además, la figura 4a ilustra tres señales diferentes, una representa la densidad espacial de energía real, que es caracterizada por una línea delgada y por ser ruidosa. Además, la línea gruesa ilustra la densidad espacial de energía teórica de una primera fuente y la línea de puntos ilustra lo mismo para una segunda fuente. El modelo que se ajusta mejor a la observación comprende dos fuentes de audio ubicadas a +45° y -135°, respectivamente. En otros modelos, la elevación puede también estar disponible. En tales realizaciones, la densidad espacial de energía se convierte en una función tridimensional.
- 30 **[0075]** En lo siguiente, se proveen más detalles en cuanto a una implementación de una realización adicional del procesador 130, especialmente en la etapa de cálculo de peso 403. Este bloque de procesamiento calcula los pesos para cada objeto a ser extraído. Los pesos son calculados en base a los datos provistos por en análisis de DirAC en 401, junto con la información en cuanto al número de fuentes y su posición de 404. La información puede ser procesada conjuntamente para todas las fuentes o separadamente, de tal manera que los pesos para cada objeto son calculados independientemente de los otros.
- 35 **[0076]** Los pesos para los i-ésimos objetos son definidos para cada binario de tiempo y frecuencia, de tal manera que si $g_i(k,n)$ denota el peso para el índice de frecuencia k y el índice de tiempo n , el espectro complejo de la señal de mezcla descendente para el i-ésimo objeto puede ser calculado simplemente por
- $$W_i(k,n) = W(k,n) \times \gamma_i(k,n).$$
- 40 **[0077]** Como ya se mencionó, las señales obtenidas de tal manera podrían ser enviadas a un codificador de SAOC. Sin embargo, las realizaciones pueden evitar totalmente esta etapa al calcular los parámetros de SAOC de los pesos $\gamma_i(k,n)$ directamente.
- 45 **[0078]** En lo siguiente, se explicará brevemente como se pueden calcular los pesos $\gamma_i(k,n)$ en algunas realizaciones. Si no se especifica de otra manera, todas las cantidades en lo siguiente dependen de (k, n) , es decir los índices de frecuencia y tiempo.
- 50
- 55

[0079] Se puede suponer que la difusividad Y o medida de confiabilidad, está definida en el intervalo $[0, 1]$, en donde $Y = 1$ corresponde a una señal totalmente difusa. Además, q denota la dirección de llegada, en el siguiente ejemplo denota el ángulo azimuth. La extensión al espacio 3D es directa.

5 **[0080]** Además, γ_i denota el peso con el cual la señal de mezcla descendente es escalada para extraer la señal de audio del i -ésimo objeto, $W(k, n)$ denota el espectro complejo de la señal de mezcla descendente y $W_i(k, n)$ denota el espectro complejo del i -ésimo objeto extraído.

[0081] En una primera realización, se define una función bidimensional en el dominio (θ, Ψ) . Una realización simple utiliza una función Gaussiana 2D $g(\theta, \Psi)$, de acuerdo con

$$g(\theta, \Psi) = A e^{-\left(\frac{(\theta-\alpha)^2}{2\sigma_\theta^2} + \frac{(\Psi)^2}{2\sigma_\Psi^2}\right)},$$

10 en donde α es la dirección en donde el objeto está ubicado y σ_θ^2 y σ_Ψ^2 son parámetros que determinan el ancho de la función Gaussiana, esto es sus variancias con respecto a ambas dimensiones. A es un factor de amplitud que se puede suponer igual a 1 en lo siguiente.

15 **[0082]** El peso $\gamma_i(k, n)$ puede ser determinado al calcular la ecuación anterior para los valores de $\theta(k, n)$ y $\Psi(k, n)$ obtenidos del procesamiento de DirAC, esto es

$$\gamma_i(k, n) = g(\theta(k, n), \Psi(k, n)).$$

[0083] Una función ejemplar es mostrada en la figura 4b. En la figura 4b, se puede ver que se presentan pesos significativos para valores de difusividad bajos. Para la figura 4b, se ha supuesto $\alpha = -\pi/4$ rad (o -45deg), y $\sigma_\theta^2 = 0,25$ y $\sigma_\Psi^2 = 0,2$.

20 **[0084]** El peso es más grande para $\Psi(k, n) = 0$ y $\theta = \alpha$. Para direcciones más alejadas de α , también como para una difusividad más alta, el peso disminuye. Al cambiar los parámetros de $g(\theta(k, n), \Psi(k, n))$ varias funciones $g(\theta(k, n), \Psi(k, n))$ pueden ser diseñadas, que extraen objetos de diferentes direcciones.

[0085] Si los pesos obtenidos de diferentes objetos conducen a una energía total que es más grande que la presente en la señal de mezcla descendente, esto es, si

$$25 \sum_{i=1}^N \gamma_i^2 > 1$$

entonces es posible actuar sobre los factores de multiplicación A en la función $g(\theta(k, n), \Psi(k, n))$ para forzar que la suma de los cuadrados sea menor o igual a 1.

30 **[0086]** En una segunda realización, la ponderación para la parte difusa y parte no difusa de la señal de audio se puede llevar a cabo con diferentes ventanas de ponderación. Se pueden encontrar más detalles en Markus Kallinger, Giovanni Del Galdo, Fabian Kuech, Dirk Mahne, Richard Schultz-Amling, "SPATIAL FILTERING USING DIRECTIONAL AUDIO CODING PARAMETERS", ICASSP 09.

[0087] El espectro del i -ésimo objeto puede ser obtenido mediante

$$W_i = \gamma_{i,di} \sqrt{\Psi} \cdot W + \gamma_{i,co} \sqrt{1-\Psi} \cdot W,$$

35 en donde $\gamma_{i,di}$ y $\gamma_{i,co}$ son los pesos para la parte difusa y no difusa (coherente), respectivamente. La ganancia para la parte no difusa puede ser obtenida de una ventana unidimensional tal como la siguiente

$$g(\theta) = \sqrt{0.5 \cdot \left(1 + \cos\left(\frac{\pi \cdot (\theta - \alpha)}{B/2}\right)\right)} \quad \text{for } \alpha - B/2 \leq \theta \text{ para } :$$

= 0 en los demás casos

en donde B es el ancho de la ventana. Una ventana ejemplar para $\alpha = -\pi/4$, $B = \pi/4$ se ilustra en la figura 4c.

[0088] La ganancia para la parte difusa $\gamma_{i, di}$, puede ser obtenida de una manera similar. Ventanas apropiadas son por ejemplo cardioides, subcardioides dirigidas hacia a o simplemente un patrón omnidireccional. Una vez calculadas las ganancias $\gamma_{i, di}$ y $\gamma_{i, co}$, el peso γ_i puede ser obtenido simplemente como

$$\gamma_i = \gamma_{i, di} \sqrt{\Psi} + \gamma_{i, co} \sqrt{1 - \Psi}$$

de tal manera que

$$W_i = \gamma_i \cdot W$$

[0089] Si los pesos obtenidos de diferentes objetos conducen a una energía total que es más grande que la presente en la señal de mezcla descendente, esto es, si

$$\sum_{i=1}^N \gamma_i^2 > 1,$$

entonces es posible reescalar las ganancias de conformidad. Este bloque de procesamiento puede también proveer los pesos para un objeto de fondo (residual) adicional, para el cual la energía es luego calculada en el bloque 402. El objeto de fondo contiene la energía restante que no ha sido asignada a ningún otro objeto. La energía puede ser asignada al objeto de fondo también para reflejar la incertidumbre de los valores estimativos de dirección. Por ejemplo, la dirección de llegada para un cierto binario de tiempo – frecuencia es estimado para ser exactamente dirigido hacia un cierto objeto. Sin embargo, ya que el valor estimativo no está libre de error, una parte pequeña de energía puede ser asignada al objeto de fondo.

[0090] En los siguiente, se proveen detalles adicionales en cuanto a una realización adicional del procesador 130, especialmente en la etapa de "calcular energía de fuentes separadas" – 402. Este bloque de procesamiento toma los pesos calculados por 403 y los usa para calcular las energías de cada objeto. Si $\gamma_i(k, n)$ denota el peso del i-ésimo objeto para el binario de tiempo-frecuencia definido por (k, n), entonces la energía $E_i(k, n)$ es simplemente

$$E_i(k, n) = |W(k, n)|^2 \gamma_i^2(k, n),$$

[0091] En donde $W(k, n)$ es la representación de tiempo-frecuencia compleja de la señal de mezcla descendente.

[0092] Idealmente, la suma de las energías de todos los objetos es igual a la energía presente en la señal de mezcla descendente, es decir

$$|W(k, n)|^2 = \sum_{i=1}^N E_i(k, n),$$

en donde N es el número de objetos.

[0093] Esto se puede obtener de diferentes maneras. Una realización puede comprender usar un objeto residual, como ya se mencionó en el contexto de cálculo del factor de ponderación. La función del objeto residual es representar cualquier energía faltante en el equilibrio global de energía de los objetos de salida, de tal manera que su energía total es igual a la energía de mezcla descendente en cada mosaico de tiempo/frecuencia.

[0094] En otras palabras, en algunas realizaciones, el procesador 130 puede ser adaptado para determinar adicionalmente un factor de ponderación para un objeto de fondo adicional, en donde los factores de ponderación son de tal manera que la suma de las energías asociadas con las por lo menos dos fuentes de audio separadas y el objeto de fondo adicional igualan la energía de la representación de señal convertida.

[0095] Un mecanismo relacionado es definido en el estándar ISO/IEC, "MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC)," ISO/IEC JTC1/SC29/WG11 (MPEG) FCD 23003-2), en cuanto a como asignar cualquier energía faltante. Otra estrategia ejemplar puede comprender reescalar los pesos apropiadamente para obtener el equilibrio de energía global deseado.

[0096] En general, si la etapa 403 provee pesos para el objeto de fondo, esta energía puede ser mapeada al objeto residual. En lo siguiente se proveen más detalles en cuanto al cálculo de los OLD de SAOC y opcionalmente IOC en la etapa de corriente de bits 405, como se pueden llevar a cabo en algunas realizaciones.

5 **[0097]** Este bloque de procesamiento procesa además la energía de los objetos de audio y los convierte a parámetros compatibles con SAOC, esto es OLD. Para este fin, las energías de objetos son normalizadas con respecto a la energía del objeto con la energía más alta resultante en valores de energía relativos para cada
10 mosaico de tiempo/frecuencia. Estos parámetros pueden ya sea ser utilizados directamente para el procesamiento del descodificador de SAOC subsecuente o pueden ser cuantificados y transmitidos/almacenados como parte de una corriente de bits de SAOC. Similarmente, los parámetros de IOC pueden ser emitidos o transmitidos/almacenados como parte de una corriente de bits de SAOC.

15 **[0098]** Dependiendo de ciertos requerimientos de implementación de los procedimientos de la invención, los procedimientos de la invención pueden ser implementados en elementos físicos o en elementos de programación. La implementación puede ser efectuada usando un medio de almacenamiento digital, en particular un disco, un DVD o un CD que tiene señales de control que se pueden leer electrónicamente almacenadas en el mismo, que cooperan con un sistema de computadora programable, de tal manera que los procedimientos de la invención son efectuados. En general, por consiguiente, la presente invención es un producto de programa de computadora con códigos de programa almacenados en un portador que se puede leer por la máquina, los códigos de programa son operativos para efectuar los procedimientos de la invención cuando el producto de programa de computadora se ejecuta en una computadora. En otras palabras, los procedimientos de la invención son por consiguiente un programa de
20 computadora que tiene códigos de programa para efectuar por lo menos uno de los procedimientos de la invención, cuando el programa de computadora se ejecuta en una computadora.

[0099] En tanto que lo anterior ha sido particularmente mostrado y descrito con referencia a realizaciones particulares de la misma, se comprenderá por aquellos experimentados en el arte que varios otros cambios en forma y detalle se pueden hacer sin desviarse del alcance de la misma.

25 **[0100]** El alcance de protección se define mediante las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un transcodificador de formato de audio (100) para transcodificar una señal de audio de entrada, la señal de audio de entrada tiene por lo menos dos componentes de audio direccionales, caracterizado porque comprende:
- 5 un convertidor (110) para convertir la señal de audio de entrada a una señal convertida, la señal convertida tiene una representación de señal convertida y una dirección de llegada de señal convertida;
- un proveedor de posición (120) para proveer por lo menos dos posiciones espaciales de por lo menos dos fuentes de audio espaciales y
- 10 un procesador (130) para procesar la representación de señal convertida en base a las por lo menos dos posiciones espaciales y la dirección de llegada de señal convertida para obtener por lo menos dos medidas de fuente de audio separadas,
- en donde el procesador (130) es apto para determinar un factor de ponderación para cada una de las por lo menos dos fuentes de audio separadas y
- 15 en donde el procesador (130) es apto para procesar la representación de señal convertida en términos de por lo menos dos filtros espaciales (311, 322, 31N) que dependen de los factores de ponderación para aproximar por lo menos dos fuentes de audio aisladas con por lo menos dos señales fuente de audio separadas como las por lo menos dos medidas de audio separadas o en donde el procesador (130) es apto para estimar (402) una información de energía para cada una de las por lo menos dos fuentes de audio separadas dependiendo de la representación de señal convertida y de los factores de ponderación como las por lo menos dos medidas fuente de audio separadas.
2. El transcodificador de formato de audio (100) de la reivindicación 1 para la transcodificación de una señal de entrada de acuerdo con una señal codificada de audio direccional (DirAc), una señal de formato B o una señal de un arreglo de micrófonos.
- 20 3. El transcodificador de formato de audio (100) de una de las reivindicaciones precedentes, en el que el convertidor (110) es apto para convertir la señal de entrada en términos de un número de bandas/subbandas de frecuencia y/o segmentos/cuadros de tiempo.
- 25 4. El transcodificador de formato de audio (100) de la reivindicación 3, en el que el convertidor (110) es apto para convertir la señal de audio de entrada a la señal convertida que comprende además una medida de difusividad y/o confiabilidad por banda de frecuencia.
5. El transcodificador de formato de audio (100) de la reivindicación 1, que comprende además un codificador de codificación de objeto de audio espacial (SAOC) para codificar las por lo menos dos señales fuente de audio separadas para obtener una señal codificada de SAOC que comprende un componente de mezcla descendente de SAOC y un componente de información lateral de SAOC.
- 30 6. El transcodificador de formato de audio (100) de la reivindicación 1, en el que el procesador (130) es apto para convertir las energías de las por lo menos dos fuentes de audio separadas a diferencias a nivel de objeto de SAOC (SAOC-OLD).
- 35 7. El transcodificador de formato de audio (100) de la reivindicación 6, en el que el procesador (130) es apto para calcular una coherencia de inter-objeto (IOC) para las por lo menos dos fuentes de audio separadas.
8. El transcodificador de formato de audio (100) de una de las reivindicaciones 3 a 7, en el que el proveedor de posición (120) comprende un detector para detectar las por lo menos dos posiciones espaciales de las por lo menos dos fuentes de audio espaciales en base a la señal convertida, en donde el detector es apto para detectar las por lo menos dos posiciones espaciales mediante una combinación de múltiples segmentos/cuadros de tiempo de señal de entrada subsecuentes.
- 40 9. El transcodificador de formato de audio (100) de la reivindicación 8, en el que el detector es apto para detectar las por lo menos dos posiciones espaciales en base a una estimación de probabilidad máxima en cuanto a una densidad espacial de energía de la señal convertida.
- 45 10. El transcodificador de formato de audio (100) de una de las reivindicaciones 1 a 9, en el que el procesador (130) es apto para determinar además un factor de ponderación para un objeto de fondo adicional, en donde los factores de ponderación son de tal manera que la suma de las energías asociadas con las por lo menos dos fuentes de audio separadas y el objeto de fondo adicional igualan la energía de la representación de señal convertida.
- 50 11. Un procedimiento para la transcodificación de una señal de audio de entrada, la señal de audio de entrada tiene por lo menos dos componentes de audio direccionales, caracterizado porque comprende las etapas de:

convertir la señal de audio de entrada a una señal convertida, la señal convertida tiene una representación de señal convertida y la dirección de llegada de señal convertida;

proveer por lo menos dos posiciones espaciales de las por lo menos dos fuentes de audio espaciales y

5 procesar la representación de señal convertida en base a las por lo menos dos posiciones espaciales para obtener las por lo menos dos medidas fuente de audio separadas,

en donde la etapa de procesamiento comprende:

determinar (303) un factor de ponderación para cada una de las por lo menos dos fuentes de audio separadas y

10 procesar la representación de señal convertida usando por lo menos dos filtros espaciales (311, 312, 31N) que dependen de los factores de ponderación para aproximar por lo menos dos fuentes de audio aisladas con por lo menos dos señales fuente de audio separadas como las por lo menos dos medidas fuente de audio separadas o estimar (402) una información de energía para cada una de las por lo menos dos fuentes de audio separadas dependiendo de la representación de la señal convertida y de los factores de ponderación como las por lo menos dos medidas fuente de audio separadas.

15 12. Un programa de computadora adaptado para efectuar el procedimiento de la reivindicación 11, cuando el programa de computadora se ejecuta en una computadora o un procesador.

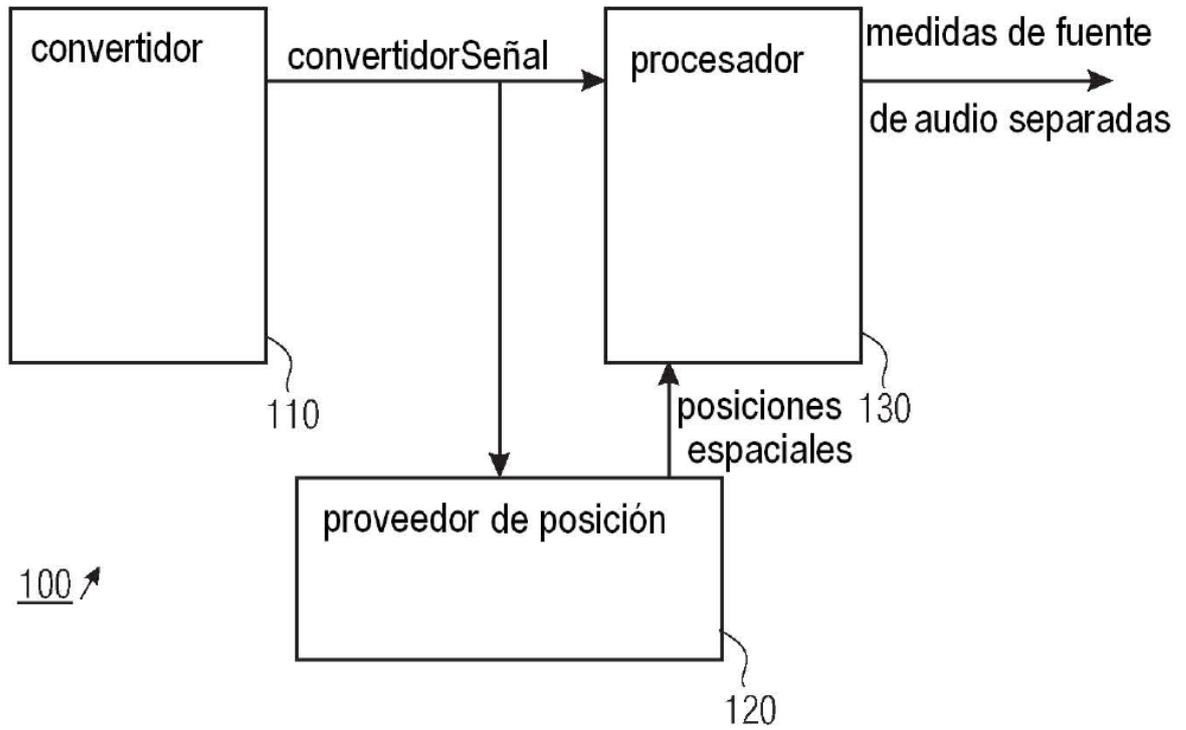


FIG 1

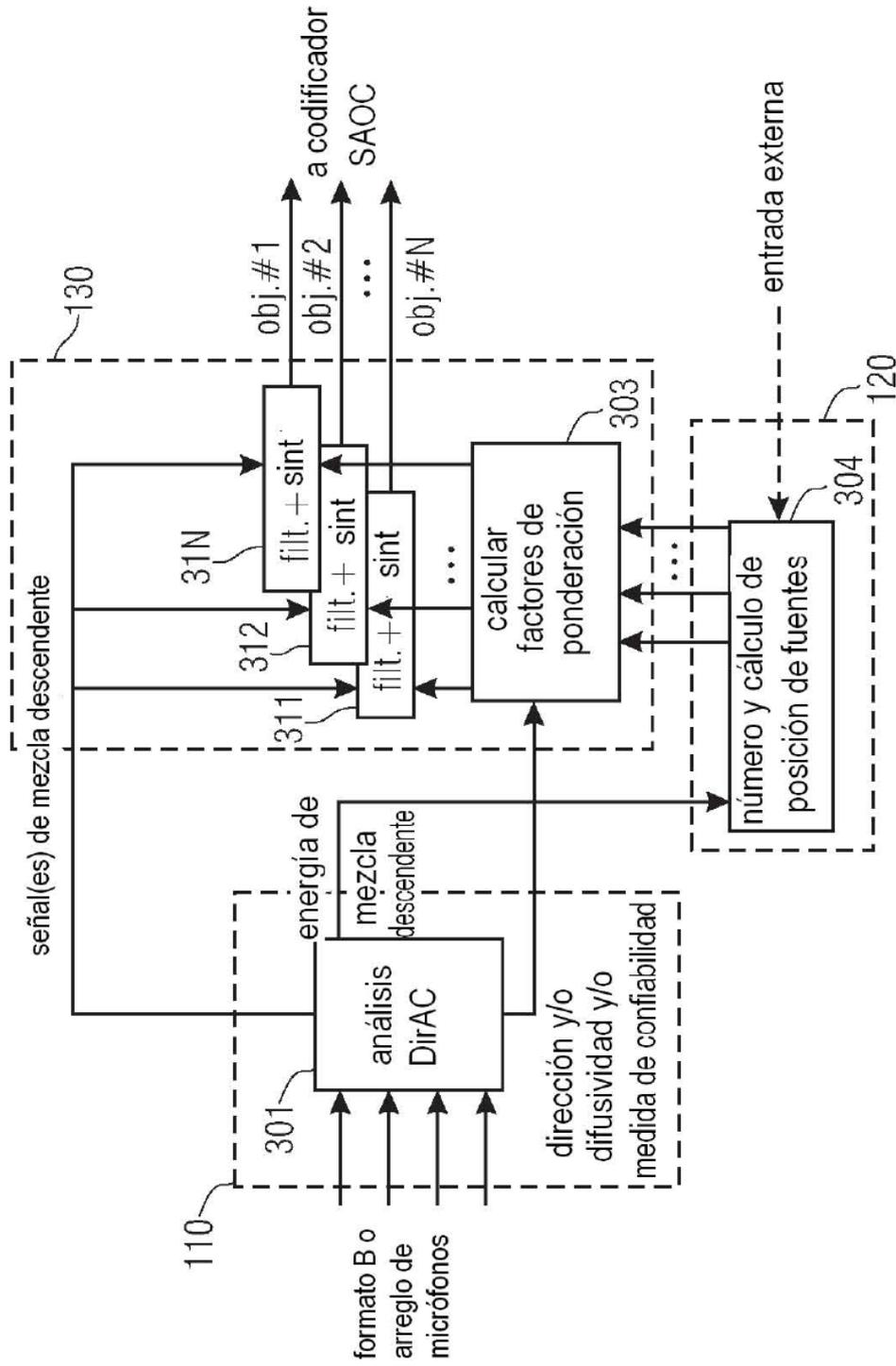


FIG 2

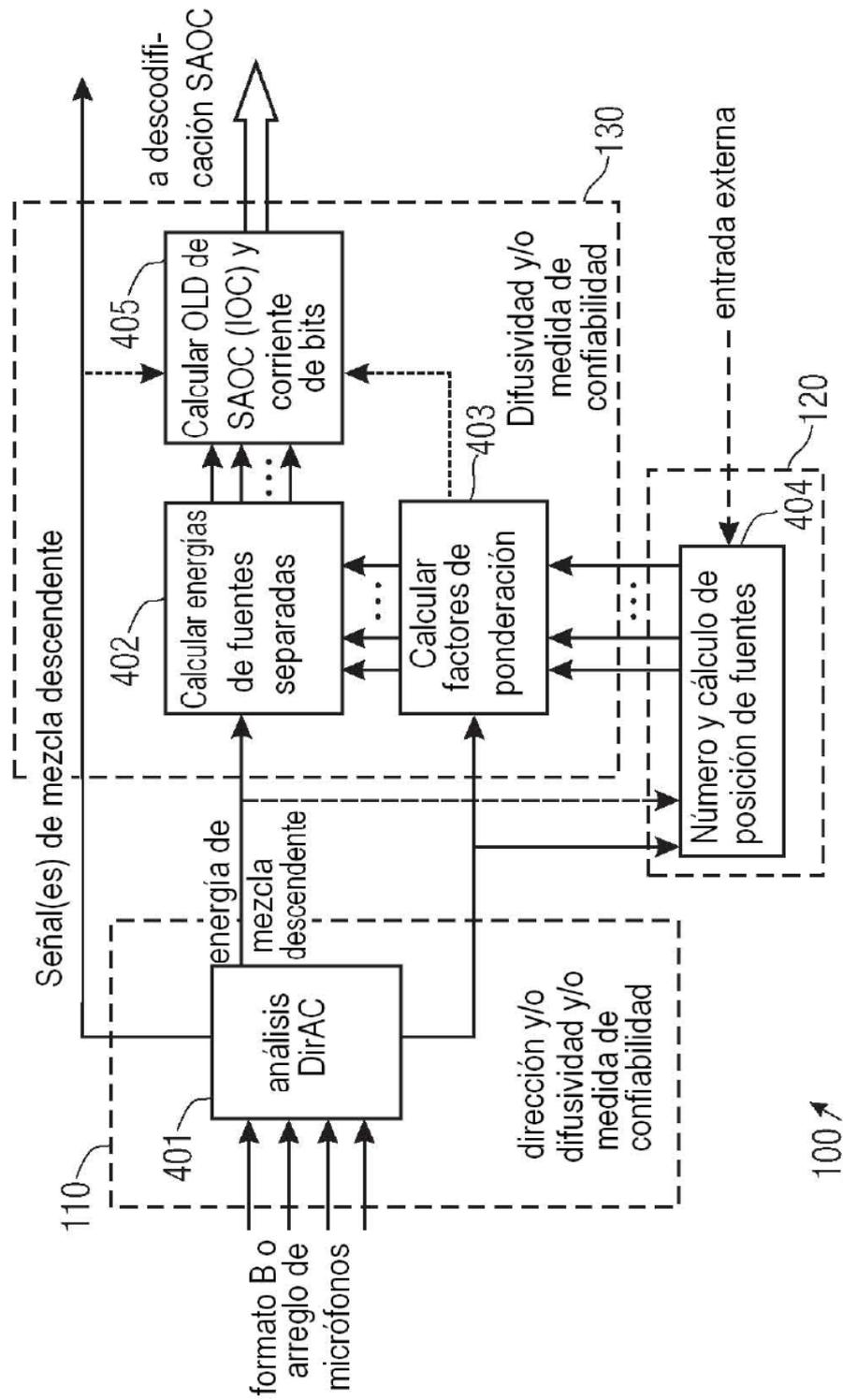


FIG 3

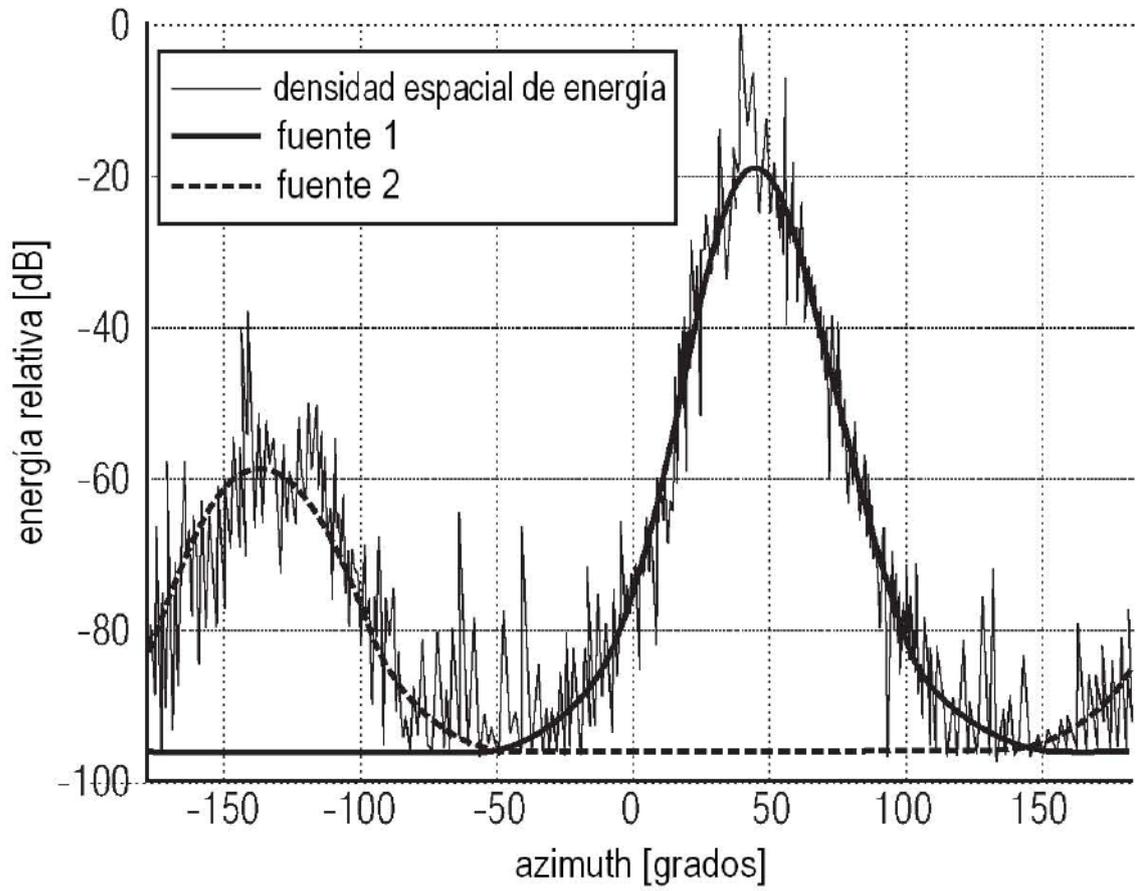


FIG 4A

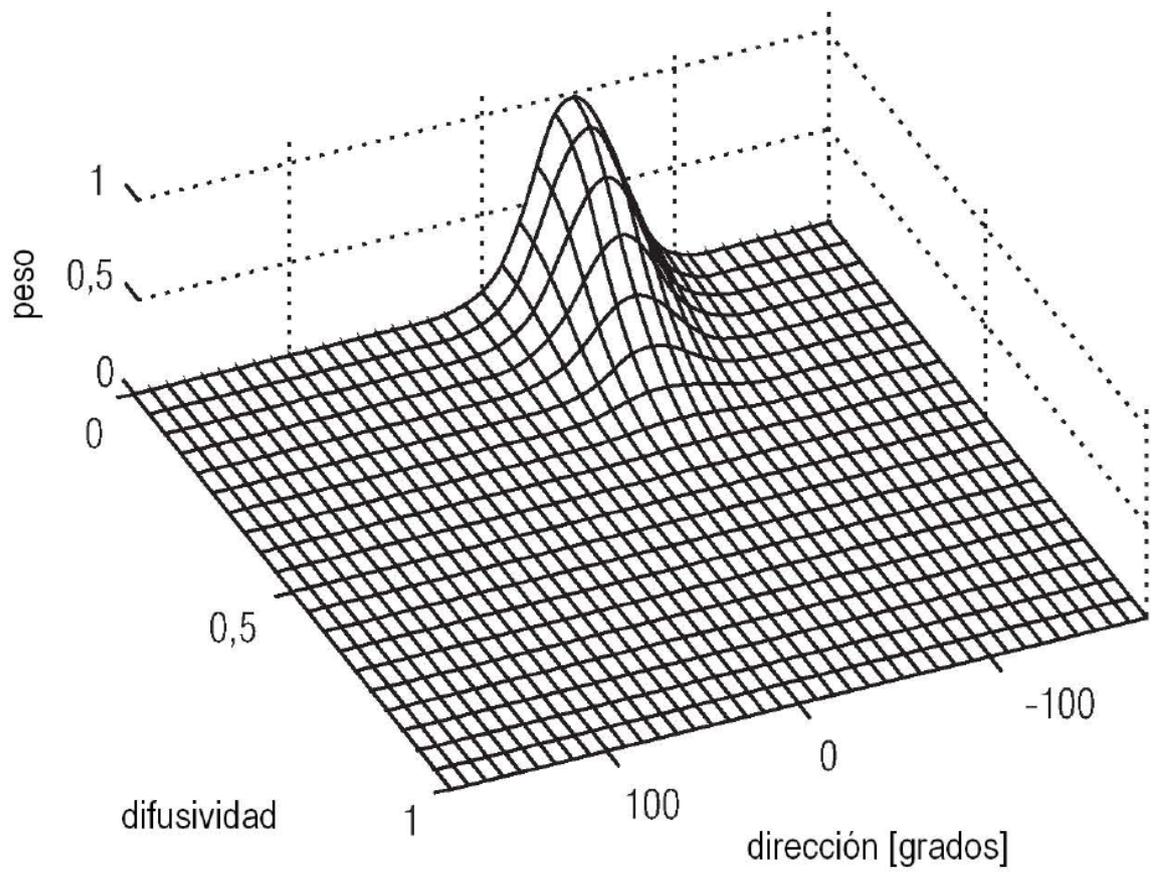


FIG 4B

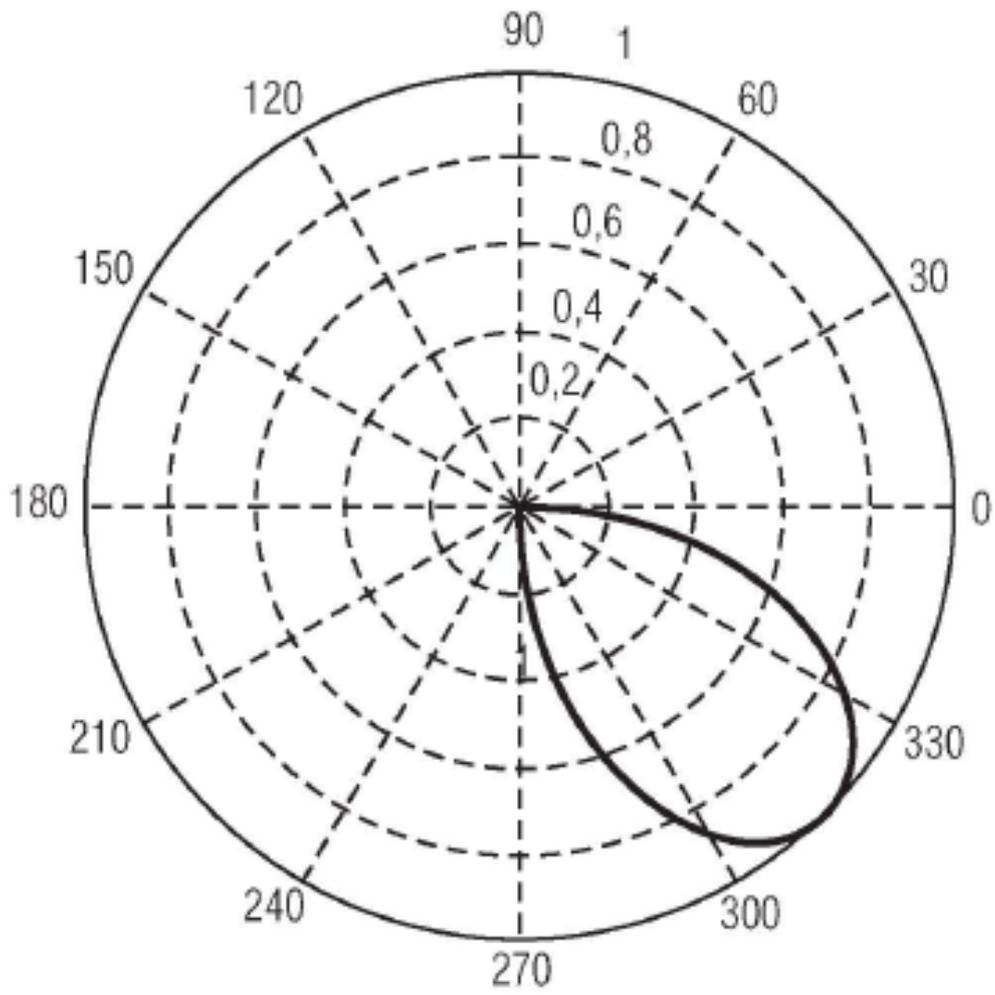


FIG 4C

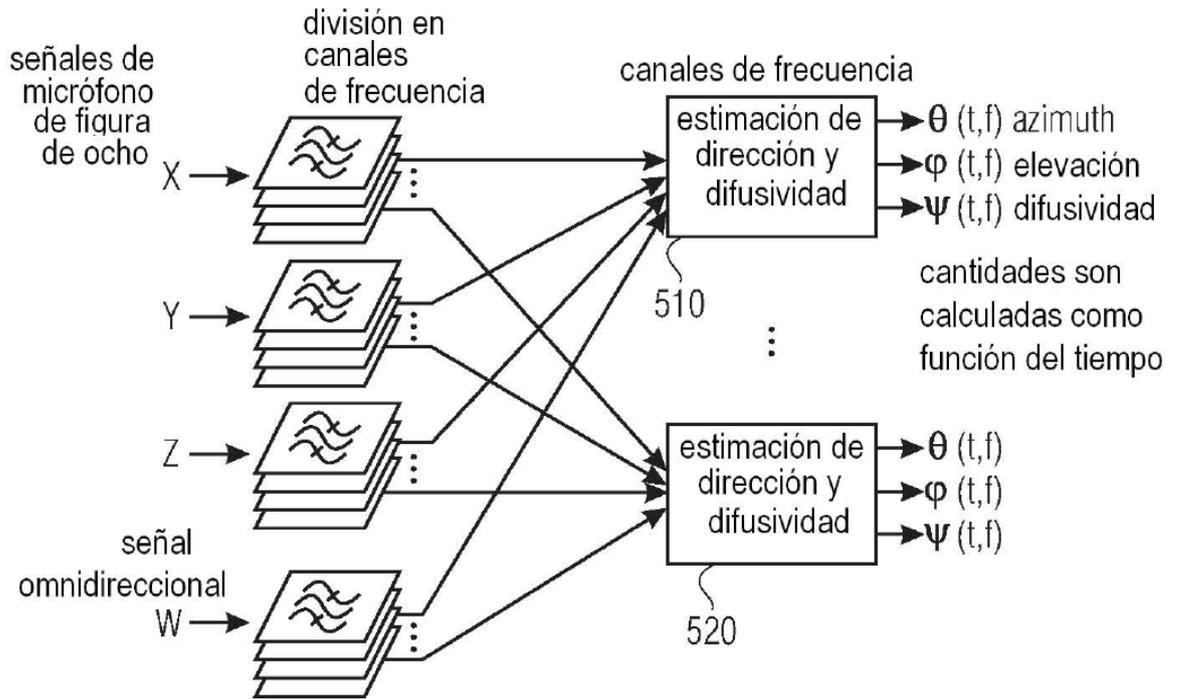


FIG 5
(TECNICA PREVIA)

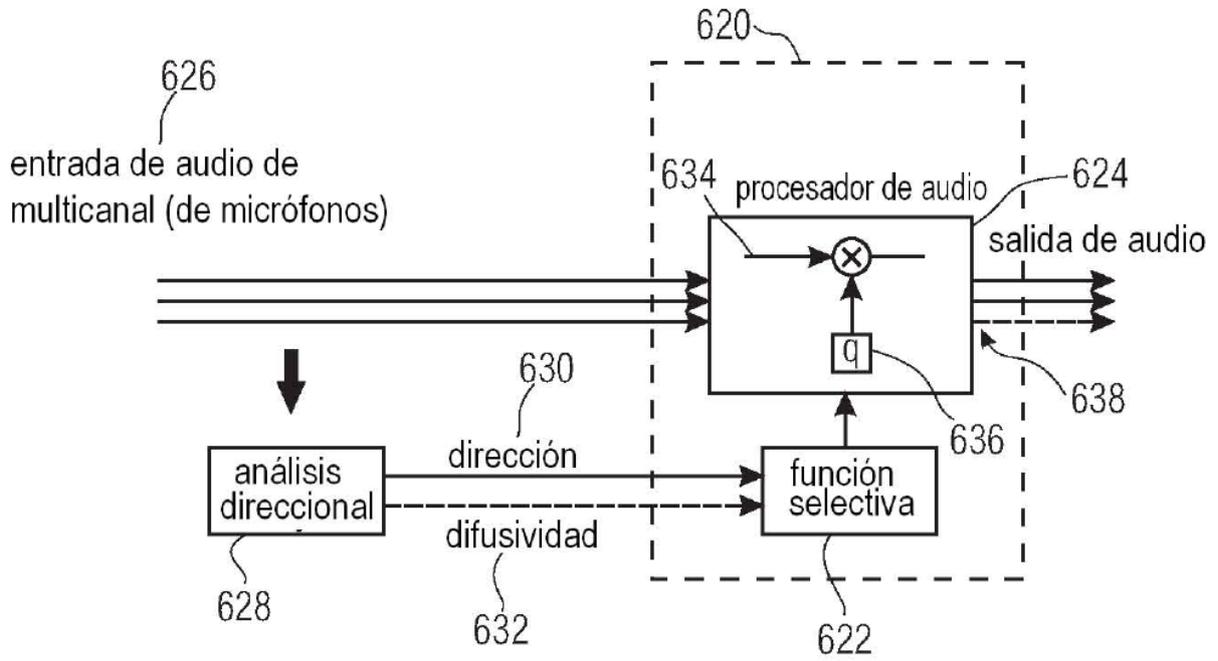


FIG 6
 (TECNICA PREVIA)

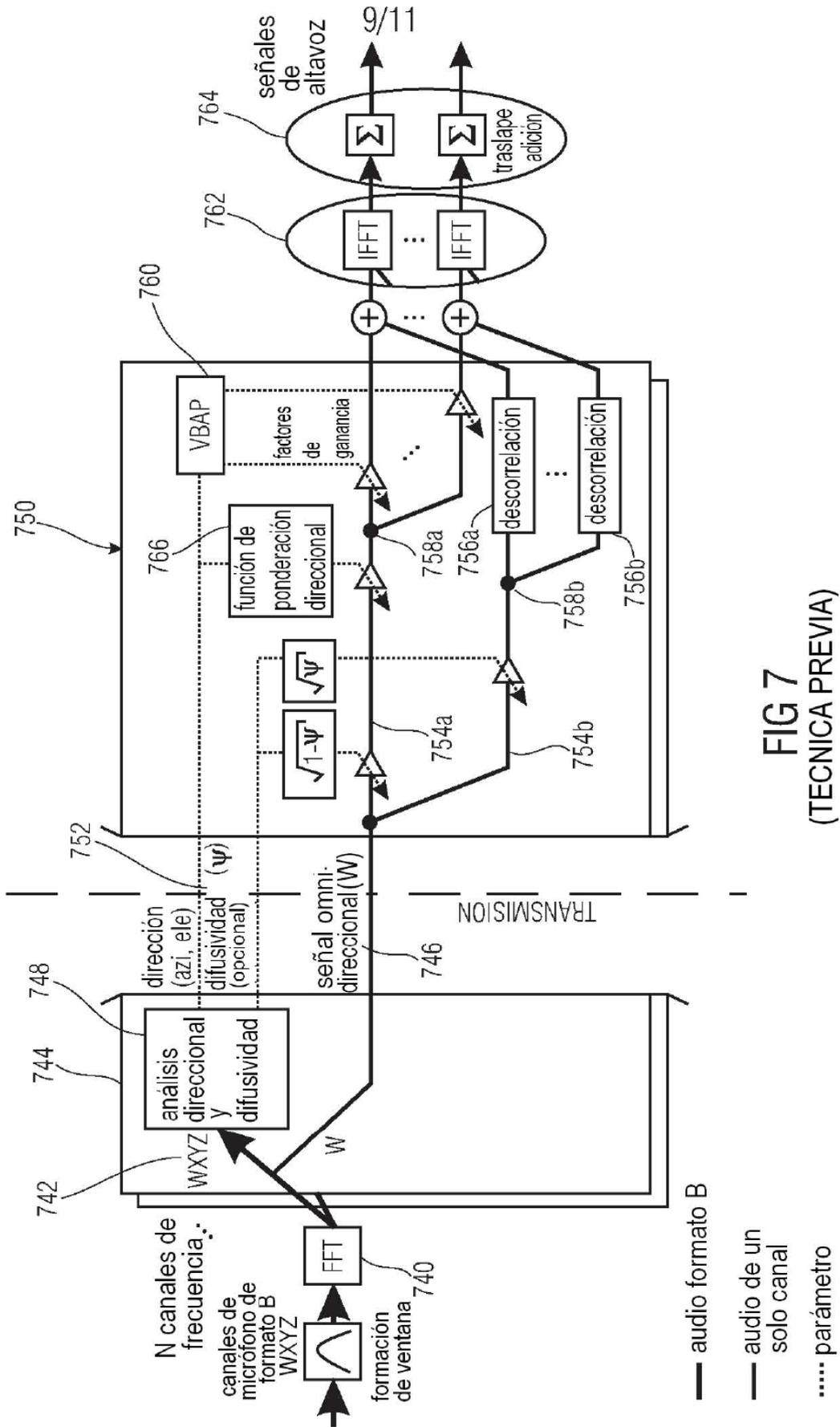


FIG 7
(TECNICA PREVIA)

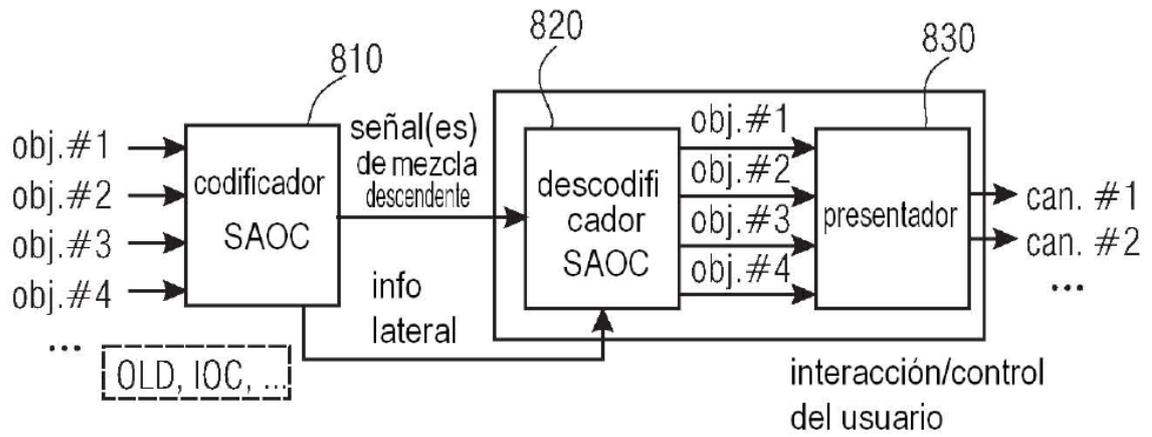


FIG 8
(TECNICA PREVIA)

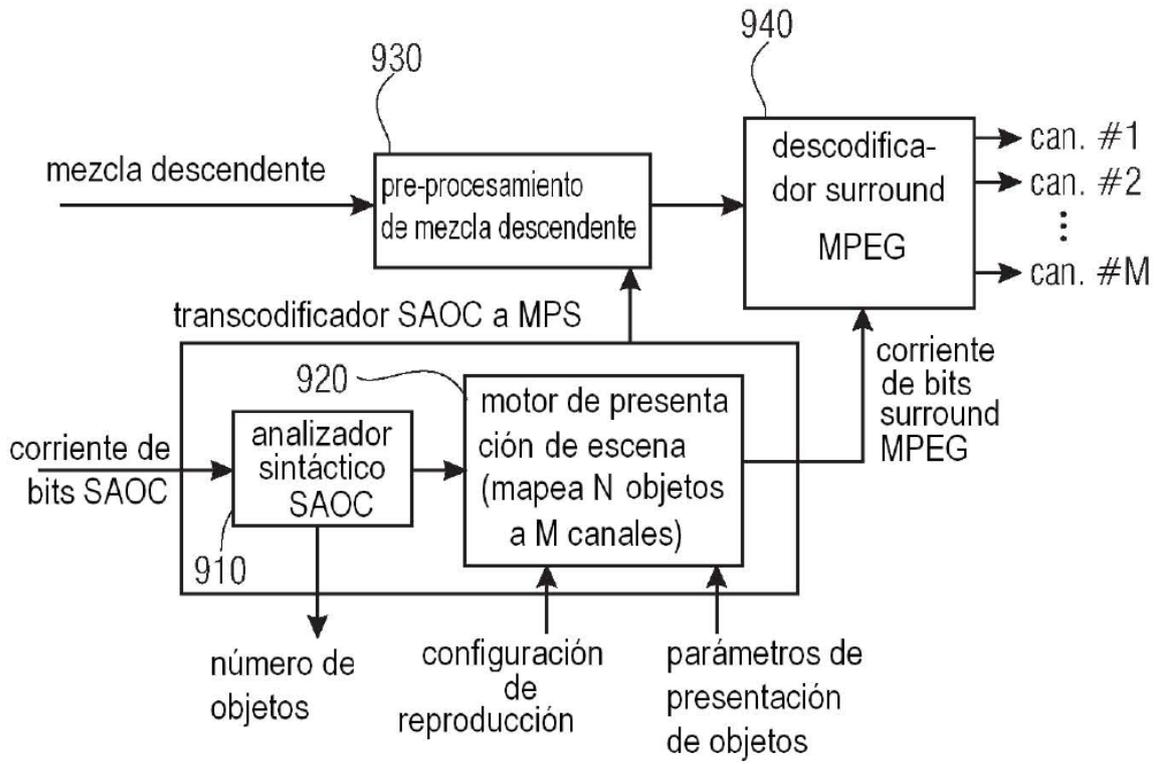


FIG 9
(TECNICA PREVIA)