

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 553 382**

51 Int. Cl.:

H04S 7/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.08.2009 E 11187023 (4)**

97 Fecha y número de publicación de la concesión europea: **09.09.2015 EP 2418877**

54 Título: **Un aparato y un método para generar datos de salida por ampliación de ancho de banda**

30 Prioridad:

13.08.2008 US 88505 P
28.10.2008 EP 08018793

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
09.12.2015

73 Titular/es:

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

72 Inventor/es:

DISCH, SASCHA;
PULKKI, VILLE;
LAITINEN, MIKKO-VILLE y
ERKUT, CUMHUR

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 553 382 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

Un aparato y un método para generar datos de salida por ampliación de ancho de banda

[0001] La presente invención es concerniente con el campo de procesamiento de audio, especialmente procesamiento de propiedades de audio espaciales.

[0002] El procesamiento y/o codificación de audio ha avanzado de muchas maneras. Se genera más y más demanda para aplicaciones de audio espaciales. En muchas aplicaciones, el procesamiento de señal de audio es utilizado para des-correlacionar o presentar señales. Tales aplicaciones pueden, por ejemplo llevar a cabo mezcla ascendente de mono-aural a estéreo, mezcla ascendente de mono/estéreo a multi-canal, reverberación artificial, ensanchamiento estéreo o mezcla/presentación interactiva del usuario.

[0003] Para ciertas clases de señales tales como por ejemplo señales semejantes a ruido, por ejemplo señales semejantes a aplauso, métodos y sistemas convencionales sufren ya sea de una calidad perceptual no satisfactoria o si se usa un procedimiento orientado al objeto, alta complejidad computacional debido al número de eventos auditivos a ser modelados o procesados. Otros ejemplos de la señal de audio que es problemático son en general material ambiental, por ejemplo el ruido que es emitido por una parvada de aves, una costa, caballos galopando, una división de soldados marchando, etc.

[0004] Los conceptos convencionales usan, por ejemplo, codificación estéreo paramétrica o codificación MPEG-surround (MPEG=grupo de expertos de películas). La Figura 6 muestra una aplicación típica de un des-correlacionador en un mezclador ascendente de mono-aural a estéreo. La Figura 6 muestra una señal de entrada mono-aural provista a un des-correlacionador 610, que provee una señal de entrada des-correlacionada en su salida. La señal de entrada original es provista a una matriz de mezcla ascendente 620 junto con la señal des-correlacionada. Dependiendo de los parámetros de control de mezcla ascendente 630, una señal de salida estéreo es presentada. El des-correlacionador de señal 610 genera una señal des-correlacionada D alimentada a la etapa de formación de matriz 620 junto con la señal mono-aural seca M. Al interior de la matriz de mezcla 620, los canales estéreo L (L = canal estéreo izquierdo) y R (R = canal estéreo derecho) son formados de acuerdo con una matriz de mezcla H. Los coeficientes en la matriz H pueden ser fijos, dependientes de la señal o controlados por el usuario.

[0005] Alternativamente, la matriz puede ser controlada mediante información lateral, transmitida junto con la mezcla descendente, que contiene una descripción paramétrica en cómo mezclar ascendentemente las señales de la mezcla descendente para formar la salida de multi-canal deseada. Esta información lateral espacial es generada usualmente por un codificador de señal antes del proceso de mezcla ascendente.

[0006] Esto se hace comúnmente en codificación de audio espacial paramétrica, por ejemplo, en estéreo paramétrico cf. J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates" en la 116a. Convención de la AES, Berlin, Pre-impresión 6072, Mayo 2004 y en MPEG Surround, cf. J. Herre, K. Kjörling, J. Breebaart, et. al., "MPEG Surround – the ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding" in Proceedings of the 122nd AES Convention, Viena, Austria, Mayo 2007. Una estructura típica de un des-codificador estéreo paramétrico es mostrada en la Figura 7. En este ejemplo, el proceso de des-correlación es efectuado en un dominio de transformada, que es indicado por el banco de filtro de análisis 710, que transforma una señal mono-aural de entrada a dominio de transformada, por ejemplo, el dominio de frecuencia en términos de un número de bandas de frecuencia.

[0007] En el dominio de frecuencia, el des-correlacionador 720 genera la señal des-correlacionada de conformidad, que va a ser mezclada ascendente en la matriz de mezcla ascendente 730. La matriz de mezcla ascendente 730 considera los parámetros de mezcla ascendente que son provistos por el bloque de modificación de parámetros 740, que es provisto con parámetros de entrada espaciales y acoplado a una etapa de control de parámetros 750. En el ejemplo mostrado en la Figura 7, los parámetros espaciales pueden ser modificados por un usuario o herramientas adicionales tales como por ejemplo post-procesamiento o presentación/proyección bin-aural. En este caso, los parámetros de mezcla ascendente pueden ser fusionados con los parámetros de los filtros bin-aurales para formar los parámetros de entrada para la matriz de mezcla ascendente 730. La medición de los parámetros se puede llevar a cabo por el bloque de modificación de parámetros 740. La salida de la matriz de mezcla ascendente 730 es luego provista a un banco de filtros de síntesis 760, que determina la señal de salida estéreo.

[0008] Como se describe anteriormente, la salida L/R de la matriz de mezcla H puede ser calculada de la señal de entrada mono-aural M y la señal des-correlacionada D, por ejemplo de acuerdo con

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} M \\ D \end{bmatrix}$$

[0009] En la matriz de mezcla, la cantidad de sonido des-correlacionado alimentado a la salida puede ser controlada en base a los parámetros transmitidos, por ejemplo ICC (ICC = correlación de inter-canal) y/o mezclada o ajustes definidos por el usuario.

5 [0010] Otro procedimiento convencional es establecido por el método de permutación temporal. Una propuesta dedicada en la des-correlación de señales semejantes a aplauso se puede encontrar, por ejemplo en Gerard Hotho, Steven van de Par, Jeroen Breebaart, "Multichannel Coding of Applause Signals," in EURASIP Journal on Advances in Signal Processing, Vol. 1, Art. 10, 2008. Aquí, una señal de audio monofónica es segmentada en segmentos de tiempo traslapantes que son permutados temporalmente pseudo-aleatoriamente dentro de un "super"-bloque para formar los canales de salida des-correlacionados. Las permutaciones son mutuamente independientes para un número de n canales de salida.

10 [0011] Otro procedimiento es el cambio de canal alternante del original y copia retardada con el fin de obtener una señal des-correlacionada, confróntese con la solicitud de patente alemana 102007018032.4-55.

15 [0012] En algunos sistemas orientados a objetos conceptuales convencionales, por ejemplo en Wagner, Andreas; Walther, Andreas; Melchoir, Frank; Strauß, Michael; "Generation of Highly Immersive Atmospheres for Wave Field Synthesis Reproduction" en la 116a. Convención Internacional de la EAS, Berlin, 2004, se describe cómo crear una escena inmersiva de muchos objetos, por ejemplo aplausos individuales, mediante la aplicación de una síntesis de campo de onda.

20 [0013] Todavía otro procedimiento es la llamada "codificación de audio direccional" (DirAc = codificación de audio direccional), que es un método para representación de sonido espacial, aplicable para diferentes sistemas de reproducción de sonido, confróntese con Pulkki, Ville, "Spatial Sound Reproduction with Directional Audio Coding" in J. Audio Eng. Soc., Vol. 55, No. 6, 2007. En la parte de análisis, la difusividad y dirección de llegada de sonido son estimados en un solo sitio dependiendo del tiempo y la frecuencia. En la parte de síntesis, las señales del micrófono son divididas primero en partes no difusas y partes difusas y luego son reproducidas utilizando estrategias diferentes.

25 [0014] Los procedimientos convencionales tienen una diversidad de desventajas. Por ejemplo, la mezcla ascendente guiada o sin guiar de las señales de audio que tiene contenido tales como aplauso puede requerir una des-correlación fuerte. Consecuentemente, por una parte, la des-correlación fuerte es necesaria para restaurar la sensación ambiental de estar, por ejemplo en una sala de conciertos. Por otra parte, filtros de des-correlación apropiados, por ejemplo filtros de todos los pasos, degradan la reproducción de calidad de eventos transitorios, como un solo aplauso al introducir efectos dañinos temporales pre y post – ecos y repique de filtro. Además, la toma panorámica espacial de eventos de un solo aplauso se tiene que hacer en una rejilla de tiempo más bien fina, en tanto que la des-correlación ambiental debe ser casi estacionaria con respecto al tiempo.

30 [0015] Los sistemas del estado del arte de acuerdo con J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates" en la 116a. Convención de la AES, Berlin, Pre-impresión 6072, Mayo 2004 y J. Herre, K. Kjörling, J. Breebaart, et. al., "MPEG Surround – the ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding" in Proceedings of the 122nd AES Convention, Viena, Austria, Mayo 2007 comprometen la resolución temporal contra la estabilidad ambiental y degradación de calidad transitoria contra des-correlación ambiental.

35 [0016] Un sistema que utiliza el método de permutación temporal, por ejemplo, exhibirá degradación perceptible del sonido de salida debido a una cierta calidad repetitiva en la señal de audio de salida. Esto es debido al hecho de que uno y el mismo segmento de la señal de entrada aparece sin alterar en cada canal de salida, aunque en un punto en el tiempo diferente. Además, para evitar la densidad de aplauso incrementada, algunos canales originales tienen que ser abandonados en la mezcla ascendente, y así, algunos eventos auditivos importantes se podrían perder en la mezcla ascendente resultante.

40 [0017] En los sistemas orientados a objetos, comúnmente tales eventos de sonido están espaciados como un gran grupo de fuentes semejantes a punto, lo que conduce a una implementación computacionalmente compleja.

45 [0018] GB 2 353 193 A describe un proceso de procesamiento de sonido, donde una señal de entrada que contiene una pluralidad de componentes de señal es separada en una pluralidad de componentes de señal mediante un separador de señal y cada componente de señal se somete a un procesamiento de sonido individual de manera que incluye un análisis espectral y la pluralidad de componentes de señal separadas son proporcionadas como al menos una señal de audio de salida por un controlador de salida. Se asume que la señal de audio de entrada contiene una mezcla de sonido de habla en-el-lugar y de sonido ambiente como pasa en la radiodifusión de deportes en vivo. La componente de sonido de habla en-el-lugar se extrae en primer lugar. Tras ello, la componente de sonido del habla extraída se sustrae de la señal original para obtener la componente de sonido ambiente.

50

55

60

65

[0019] Es el objeto de la presente invención proveer un concepto mejorado para el procesamiento de audio espacial.

[0020] Este objeto es obtenido por un aparato de acuerdo con la reivindicación 1 y un método de acuerdo con la reivindicación 12.

5 [0021] Es un hallazgo de la presente invención que una señal de audio puede ser descompuesta en varios componentes a los cuales una presentación espacial, por ejemplo, en términos de una des-correlación o en términos de un procedimiento de toma panorámica de amplitud, pueden ser adaptados. En otras palabras, la presente invención está basada en el descubrimiento de que, por ejemplo, en un escenario con múltiples fuentes de audio, las
10 fuentes de primer plano y el fondo pueden ser distinguidas y presentadas o des-correlacionadas diferentemente. Profundidades y/o extensiones espaciales en general diferentes de objetos de audio pueden ser distinguidos.

15 [0022] Uno de los puntos clave de la presente invención es la descomposición de señales, como el sonido que se origina de una audiencia que aplaude, una parvada de aves, una costa, caballos galopantes, una división de soldados marchando, etc. a una parte delantera y una parte del fondo, mediante lo cual el primer plano contiene eventos auditivos individuales originados de, por ejemplo fuentes cercanas y la parte del fondo contiene el ambiente de eventos lejanos perceptualmente fusionados. Antes de la mezcla final, estas dos partes de señal son procesadas separadamente, por ejemplo con el fin de sintetizar la correlación, presentar una escena, etc.

20 [0023] Las realizaciones no están limitadas a distinguir solamente las partes de primer plano y del fondo de la señal, pueden distinguir múltiples partes de audio diferentes, que pueden todas ser presentadas o des-correlacionadas diferentemente.

25 [0024] En general, las señales de audio pueden ser descompuestas en n partes semánticas diferentes mediante realizaciones que son procesadas separadamente. La descomposición/procesamiento separado de diferentes componentes semánticos se puede llevar a cabo en el dominio de tiempo y/o frecuencia por las realizaciones.

30 [0025] Las realizaciones pueden proveer la ventaja de calidad perceptual superior del sonido presentado a un costo computacional moderado. Las realizaciones de la presente proveen un nuevo método de des-correlación/proyección que ofrece alta calidad perceptual a costos moderados, especialmente para material de audio crítico semejante a aplauso u otro material ambiental similar, por ejemplo el ruido que es emitido por una parvada de aves, una costa, caballos galopando, una división de soldados marchando, etc.

35 [0026] Realizaciones de la presente invención serán detalladas con la ayuda de las figuras adjuntas, en las cuales:
La Figura 1a muestra una realización de un aparato para determinar una señal de audio de multi-canal de audio espacial;

40 la Figura 1b muestra un diagrama de bloques de otra realización;

la Figura 2 muestra una realización que ilustra una multiplicidad de señales descompuestas;

la Figura 3 ilustra una realización con una parte delantera y una descomposición semántica del fondo;

45 la Figura 4 ilustra un ejemplo de un método de separación de transitorios para obtener un componente de señal de fondo;

la Figura 5 ilustra una síntesis de fuente de sonido que tienen espacialmente una gran extensión;

50 la Figura 6 ilustra una aplicación del estado del arte de un des-correlacionador en el dominio de tiempo en un mezclador ascendente de mono a estéreo y

la Figura 7 muestra otra aplicación del estado del arte de un des-correlacionador en el dominio de frecuencia en un
55 escenario de mezclador ascendente de mono a estéreo.

[0027] La Figura 1 muestra una realización de un aparato 100 para determinar una señal de audio de multi-canal de salida espacial en base a una señal de audio de entrada. En algunas realizaciones, el aparato puede ser adaptado para basarse adicionalmente en la señal de audio de multi-canal de salida espacial en un parámetro de entrada. El parámetro de entrada puede ser generado localmente o provisto con la señal de audio de entrada, por ejemplo como
60 información lateral.

[0028] En la realización ilustrada en la Figura 1, el aparato 10 comprende un des-compositor 110 para descomponer la señal de audio de entrada para obtener una primera señal descompuesta que tiene una primera propiedad semántica y una segunda señal descompuesta que tiene una segunda propiedad semántica que es diferente de la
65 primera propiedad semántica.

[0029] El aparato 100 comprende además un presentador 120 para presentar la primera señal descompuesta utilizando una primera característica de presentación para obtener una primera señal presentada que tiene la primera propiedad semántica y para presentar la segunda señal descompuesta utilizando una segunda característica de presentación para obtener una segunda señal presentada que tiene la segunda propiedad semántica.

5 **[0030]** Una propiedad semántica puede corresponder a una propiedad espacial, tan cerca o lejos, enfocada o amplia y/o una propiedad dinámica, por ejemplo si una señal es tonal, estacionaria o transitoria y/o una propiedad de dominancia, por ejemplo si la señal es delantera o del fondo, una medida de la misma respectivamente.

10 **[0031]** Además, en una realización, el aparato 100 comprende un procesador 130 para procesar la primera señal presentada y la segunda señal presentada para obtener la señal de audio de multi-canal de salida espacial.

[0032] En otras palabras, el des-compositor 110 está adaptado para descomponer la señal de audio de entrada, en algunas realizaciones en base al parámetro de entrada. La descomposición de la señal de audio de entrada es adaptada a propiedades semánticas, por ejemplo espaciales de diferentes partes de la señal de audio de entrada. Además, la presentación llevada a cabo por el presentador 120 de acuerdo con las primeras y segundas características de presentación pueden también ser adaptadas a las propiedades espaciales lo que permite, por ejemplo en un escenario en donde la primera señal descompuesta corresponde a una señal de audio del fondo y la segunda señal descompuesta corresponde a una señal de audio de primer plano, presentación diferente o des-correlacionadores diferentes pueden ser aplicados, de una u otra manera respectivamente. En lo siguiente, el término "primer plano" se entiende que se refiere a un objeto de audio que es dominante en un medio ambiente de audio, de tal manera que un usuario que escucha potencial notaría un objeto de audio de primer plano. Un objeto o fuente de audio de primer plano puede ser distinguido o diferenciado de un objeto o fuente de audio del fondo. Un objeto o fuente de audio de fondo puede no ser notable por un escucha potencial en un medio ambiente de audio por ser menos dominante que un objeto o fuente de audio de primer plano. En algunas realizaciones, los objetos o fuentes de audio de primer plano pueden ser pero no están limitados a, una fuente de audio semejante a punto, en donde los objetos o fuentes de audio del fondo pueden corresponder a objetos o fuentes de audio espacialmente más amplios.

30 **[0033]** En otras palabras, en realizaciones la primera característica de presentación puede estar basada en o hacerse coincidir con la primera propiedad semántica y la segunda característica de presentación puede estar basada en o hacerse coincidir con la segunda propiedad semántica. En una realización, la primera propiedad semántica y la primera característica de presentación corresponden a una fuente u objeto de audio de primer plano y el presentador 120 puede estar adaptado para aplicar toma panorámica de amplitud a la primera señal descompuesta. El presentador 120 puede luego ser adaptado adicionalmente para proveer como la primera señal presentada dos versiones de toma panorámica de amplitud de la primera señal descompuesta. En esta realización, la segunda propiedad semántica y la segunda característica de presentación corresponden a un objeto o fuente de audio del fondo, una pluralidad de los mismos respectivamente y el presentador 120 puede ser adaptado para aplicar una des-correlación a la segunda señal descompuesta y proveer como segunda señal presentada la segunda señal descompuesta y la versión des-correlacionada de la misma.

45 **[0034]** En algunas realizaciones, el presentador 120 puede ser adaptado adicionalmente para presentar la primera señal descompuesta de tal manera que la primera característica de presentación no tiene una característica que introduce retardo. En otras palabras, puede no haber ninguna des-correlación de la primera señal descompuesta. En otra realización, la primera característica de presentación puede tener una característica que introduce retardo que tiene una primera cantidad de retardo y la segunda característica de presentación puede tener una segunda cantidad de retardo, la segunda cantidad de retardo es mayor que la primera cantidad de retardo. En otras palabras, en esta realización, tanto la primera señal descompuesta como la segunda señal descompuesta pueden ser des-correlacionadas, sin embargo, el nivel de des-correlación se puede escalar con la cantidad de retardo introducido a las versiones des-correlacionadas respectivas de las señales descompuestas. La des-correlación puede por consiguiente ser más fuerte para la segunda señal descompuesta que para la primera señal descompuesta.

60 **[0035]** En realizaciones, la primera señal descompuesta y la segunda señal descompuesta se pueden traslapar y/o pueden ser sincronizadas en el tiempo. En otras palabras, el procesamiento de señal se puede llevar a cabo por bloques, en donde un bloque de muestras de señal de audio de entrada puede ser subdividido por el des-compositor 110 en un número de bloques de señales descompuestas. En realizaciones, el número de señales descompuestas se puede por lo menos parcialmente traslapar en el dominio del tiempo, esto es, pueden representar muestras de dominio de tiempo traslapantes. En otras palabras, las señales descompuestas pueden corresponder a partes de la señal de audio de entrada que se superponen o traslapan, esto es, que representan señales de audio por lo menos parcialmente simultáneas. En realizaciones, las primeras y segundas señales descompuestas pueden representar versiones filtradas o transformadas de una señal de entrada original. Por ejemplo, pueden representar partes de señal que son extraídas de una señal espacial compuesta correspondiente por ejemplo a una fuente de sonido cercana o una fuente de sonido más distante. En otras realizaciones, pueden corresponder a componentes de señal transitorios y estacionarios, etc.

65 **[0036]** En realizaciones, el presentador 120 puede ser subdividido en un primer presentador y un segundo

presentador, en donde el primer presentador puede ser adaptado para presentar la primera señal descompuesta y el segundo presentador puede ser adaptado para presentar la segunda señal descompuesta. En realizaciones, el presentador 120 puede ser implementado en elementos de programación, por ejemplo como un programa almacenado en una memoria a ejecutarse en un procesador o un procesador de señales digitales que a su vez, es adaptado para presentar las señales descompuestas secuencialmente.

[0037] El presentador 120 puede estar adaptado para des-correlacionar la primera señal descompuesta para obtener una primera señal des-correlacionada y/o para des-correlacionar la segunda señal descompuesta para obtener una segunda señal des-correlacionada. En otras palabras, el presentador 120 puede ser adaptado para des-correlacionar tanto señales descompuestas, sin embargo, utilizando diferentes características de des-correlación o presentación. En algunas realizaciones, el presentador 120 puede ser adaptado para aplicar toma panorámica de amplitud ya sea a una u otra de las primeras o segundas señales descompuestas en lugar de esto o además de la des-correlación.

[0038] El presentador 120 puede ser adaptado para presentar las primeras y segundas señales presentadas cada una que tienen tantos componentes como canales en la señal de audio de multi-canal de salida espacial y el procesador 130 puede estar adaptado para combinar los componentes de las primeras y segundas señales presentadas para obtener la señal de audio de multi-canal de salida espacial. En otras realizaciones, el presentador 120 puede ser adaptado para presentar las primeras y segundas señales presentadas cada una que tiene menos componentes que la señal de audio de multi-canal de salida espacial y en donde el procesador 130 puede ser adaptado para mezclar ascendentemente los componentes de las primeras y segundas señales presentadas para obtener la señal de audio de multi-canal de salida espacial.

[0039] La Figura 1b muestra otra realización de un aparato 100 que comprende componentes similares como fueron presentados con la ayuda de la Figura 1a. Sin embargo, la Figura 1b muestra una realización que tiene más detalles. La Figura 1b muestra un des-compositor 110 que recibe la señal de audio de entrada y opcionalmente el parámetro de entrada. Como se puede ver en la Figura 1b, el des-compositor está adaptado para proveer una primera señal descompuesta y una segunda señal descompuesta a un presentador 120, que es indicado por las líneas discontinuas. En la realización mostrada en la Figura 1b, se supone que la primera señal descompuesta corresponde a una fuente de audio semejante a punto como la primera propiedad semántica y que el presentador 120 está adaptado para aplicar toma panorámica de amplitud como la primera característica de presentación a la primera señal descompuesta. En realizaciones, las primeras y segundas señales descompuestas son intercambiables, esto es, en otras realizaciones la toma panorámica de amplitud puede ser aplicada a la segunda señal descompuesta.

[0040] En la realización ilustrada en la Figura 1b, el presentador 120 muestra, en la trayectoria de señal de la primera señal descompuesta, dos amplificadores escalables 121 y 122, que están adaptados para amplificar dos copias de la primera señal descompuesta diferentemente. Los factores de amplificación diferentes usados pueden ser determinados, en realizaciones, a partir del parámetro de entrada, en otras realizaciones, pueden ser determinados a partir de la señal de audio de entrada, pueden ser pre-establecidos o pueden ser generados localmente, posiblemente también refiriéndose a una entrada de usuario. Las salidas de los dos amplificadores escalables 121 y 122 son provistas al procesador 130, para el cual detalles serán provistos posteriormente en la presente.

[0041] Como se puede ver de la Figura 1b, el des-compositor 110 provee una segunda señal descompuesta al presentador 120, que lleva a cabo una presentación diferente en la trayectoria de procesamiento de la segunda señal descompuesta. En otras realizaciones, la primera señal descompuesta puede ser procesada en la trayectoria actualmente descrita también o en lugar de la segunda señal descompuesta. Las primeras y segundas señales descompuestas pueden ser intercambiadas en realizaciones.

[0042] En la realización ilustrada en la Figura 1b, en la trayectoria de procesamiento de la segunda señal descompuesta, hay un des-correlacionador 123 seguido por un rotador o módulo estéreo paramétrico o módulo de mezcla ascendente 124 como segunda característica de presentación. El des-correlacionador 123 puede ser adaptado para des-correlacionar la segunda señal descompuesta $X[k]$ y para proveer una versión des-correlacionada $Q[k]$ de la segunda señal descompuesta al módulo estéreo paramétrico o mezcla de mezcla ascendente 124. En la Figura 1b, la señal mono $X[k]$ es alimentada a la unidad de des-correlacionador "D" 123 también como el módulo de mezcla ascendente 124. La unidad de des-correlacionador 123 puede crear la versión des-correlacionada $Q[k]$ de la señal de entrada, que tiene las mismas características de frecuencia y la misma energía a largo plazo. El módulo de mezcla ascendente 124 puede calcular una matriz de mezcla ascendente en base a los parámetros espaciales y sintetizar los canales de salida $Y_1[k]$ y $Y_2[k]$. El módulo de mezcla ascendente puede ser explicado de acuerdo con:

$$\begin{bmatrix} Y_1[k] \\ Y_2[k] \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_r \end{bmatrix} \begin{bmatrix} \cos(\alpha + \beta) & \text{sen}(\alpha + \beta) \\ \cos(-\alpha + \beta) & \text{sen}(-\alpha + \beta) \end{bmatrix} \begin{bmatrix} X[k] \\ Q[k] \end{bmatrix}$$

- 5 con los parámetros c_i , c_r , α y β que son constantes o valores variantes en tiempo o variantes en frecuencia estimados a partir de la señal de entrada $X[k]$ adaptablemente o transmitidos como información lateral junto con la señal de entrada $X[k]$ en forma de por ejemplo, parámetro de ILD (ILD = diferencia de nivel de inter-canal) y parámetros de ICC (ICC = correlación de inter-canal). La señal $X[k]$ es la señal mono-aural recibida, la señal $Q[k]$ es la señal des-correlacionada, que es una versión des-correlacionada de la señal de entrada $X[k]$. Las señales diferidas son denotadas por $Y_1[k]$ y $Y_2[k]$.
- 10 **[0043]** El des-correlacionar 123 puede ser implementado como un filtro de IIR (IIR = respuesta de impulso infinito), un filtro de FIR arbitrario (FIR = respuesta de impulso finita) o un filtro de FIR especial utilizando una sola derivación para retardar simplemente la señal.
- 15 **[0044]** Los parámetros c_i , c_r , α y β pueden ser determinados de maneras diferentes. En algunas realizaciones, son simplemente determinados mediante parámetros de entrada, que pueden ser provistos junto con la señal de audio de entrada, por ejemplo con los datos de mezcla descendente como información lateral. En otras realizaciones, pueden ser generados localmente o derivados de propiedades de la señal de audio de entrada.
- 20 **[0045]** En la realización mostrada en la Figura 1b, el presentador 120 es adaptado para proveer la segunda señal presentada en términos de las dos señales de salida $Y_1[k]$ y $Y_2[k]$ del módulo de mezcla ascendente 124 al procesador 130.
- 25 **[0046]** De acuerdo con la trayectoria de procesamiento de la primera señal descompuesta, las dos versiones de toma panorámica de amplitud de la primera señal descompuesta, disponibles de las salidas de los dos amplificadores escalables 121 y 122 son también provistas al procesador 130. En otras realizaciones, los amplificadores escalables 121 y 122 pueden estar presentes en el procesador 130, en donde solamente la primera señal descompuesta y un factor de toma panorámica pueden ser provistos por el presentador 120.
- 30 **[0047]** Como se puede ver en la Figura 1b, el procesador 130 puede ser adaptado para procesar o combinar la primera señal presentada y la segunda señal presentada, en esta realización simplemente al combinar las salidas con el fin de proveer una señal estéreo que tiene un canal izquierdo L y un canal derecho R correspondiente a la señal de audio de multi-canal de salida espacial de la Figura 1a.
- 35 **[0048]** En la realización de la Figura 1b, en ambas trayectorias de señalización, los canales izquierdo y derecho para una señal estéreo son determinados. En la trayectoria de la primera señal descompuesta, se lleva a cabo la toma panorámica de amplitud por los dos amplificadores escalables 121 y 122, por consiguiente, los dos componentes dan como resultado dos señales de audio en fase, que son escalables diferentemente. Esto corresponde a una impresión de una fuente de audio semejante a punto como una propiedad semántica o característica de presentación.
- 40 **[0049]** En la trayectoria de procesamiento de señal de la segunda señal descompuesta, las señales de salida $Y_1[k]$ y $Y_2[k]$ son provistas al procesador 130 correspondientes a canales izquierdo y derecho como se determina por el módulo de mezcla ascendente 124. Los parámetros c_i , c_r , α y β determinan la amplitud espacial de la fuente de audio correspondiente. En otras palabras, los parámetros c_i , c_r , α y β pueden ser escogidos de una manera o intervalo de tal manera que para los canales L y R cualquier correlación entre una correlación máxima y una correlación mínima puede ser obtenida en la segunda trayectoria de procesamiento de señal como segunda característica de presentación. Además, esto se puede llevar a cabo independientemente para bandas de frecuencia diferentes. En otras palabras, los parámetros c_i , c_r , α y β pueden ser escogidos de una manera o intervalo de tal manera que los canales L y R están en fase, modelando una fuente de audio semejante a punto como propiedad semántica.
- 50 **[0050]** Los parámetros c_i , c_r , α y β pueden también ser escogidos de una manera o intervalo de tal manera que los canales L y R en la segunda trayectoria de procesamiento de señal son des-correlacionados, modelando una fuente de audio más bien distribuida espacialmente como propiedad semántica, por ejemplo, modelado de una fuente de audio de fondo o espacialmente más amplia.
- 55 **[0051]** La Figura 2 ilustra otra realización que es más general. La Figura 2 muestra un bloque de descomposición semántica 210, que corresponde al des-compositor 110. La salida de la descomposición semántica 210 es la entrada de una etapa de presentación 220, que corresponde al presentador 120. La etapa de presentación 220 está compuesta de un número de presentadores individuales 221 a 222n, esto es, la etapa de descomposición semántica 210 está adaptada para descomposición de una señal de entrada mono/estéreo a n señales descompuestas, que tiene n propiedades semánticas. La descomposición se puede llevar a cabo en base a parámetros que controlan la descomposición, que pueden ser provistos junto con la señal de entrada mono/estéreo, ser pre-establecidos, ser generados localmente o ser introducidos por un usuario, etc.
- 60 **[0052]** En otras palabras, el des-compositor 110 puede ser adaptado para descomponer la señal de audio de entrada semánticamente en base al parámetro de entrada opcional y/o para determinar el parámetro de entrada a
- 65

partir de la señal de audio de entrada.

[0053] La salida de la etapa de des-correlación o presentación 220 es luego provista a un bloque de mezcla ascendente 230, que determina una salida de multi-canal en base a las señales des-correlacionadas o presentadas y opcionalmente basadas en parámetros controlados por la mezcla ascendente.

[0054] En general, las realizaciones pueden separar el material de sonido en n componentes semánticos diferentes y des-correlacionar cada componente separadamente con un des-correlacionador coincidente, que también son marcados D^1 a D^n en la Figura 2. En otras palabras, en realizaciones, las características de presentación se pueden hacer coincidir con las propiedades semánticas de las señales descompuestas. Cada uno de los des-correlacionadores o presentadores pueden ser adaptados a las propiedades semánticas del componente de señal descompuesta de conformidad. Subsecuentemente, los componentes procesados pueden ser mezclados para obtener la señal de multi-canal de salida. Los diferentes componentes podrían corresponder por ejemplo a objetos de modelado de primer plano y de fondo.

[0055] En otras palabras, el presentador 110 puede ser adaptado para combinar la primera señal descompuesta y la primera señal des-correlacionada para obtener la señal de mezcla ascendente estéreo o de multi-canal como la primera señal presentada y/o combinar la segunda señal descompuesta y la segunda señal des-correlacionada para obtener una señal de mezcla ascendente estéreo como la segunda señal presentada.

[0056] Además, el presentador 120 puede ser adaptado para presentar la primera señal descompuesta de acuerdo con una característica de audio de fondo y/o para presentar la segunda señal descompuesta de acuerdo con una característica de audio de primer plano o viceversa.

[0057] Puesto que, por ejemplo, las señales semejantes a aplauso pueden ser vistas como compuestas de aplausos individuales casi distintos y un ambiente semejante a ruido que se origina de aplausos lejanos muy densos, una descomposición apropiada de tales señales puede ser obtenida al distinguir entre eventos de aplausos de primer plano aislados como un componente y el fondo semejante a ruido como el otro componente. En otras palabras, en una realización, $n = 2$. En tal realización, por ejemplo, el presentador 120 puede ser adaptado para presentar la primera señal descompuesta mediante toma panorámica de amplitud de la primera señal descompuesta. En otras palabras, la correlación o presentación del componente de aplauso de primer plano puede en algunas realizaciones ser obtenida en D^1 mediante toma panorámica de amplitud de cada evento individual a su sitio original estimado.

[0058] En realizaciones, el presentador 120 puede ser adaptado para presentar la primera y/o segunda señal descompuesta por ejemplo, mediante filtración de todos los pasos de la primera o segunda señal descompuesta para obtener la primera o segunda señal des-correlacionada.

[0059] En otras palabras, en realizaciones, el fondo puede ser des-correlacionado o presentado mediante el uso de m filtros de todos los pasos mutuamente independientes $D^2_{1...m}$. En realizaciones, solamente el fondo casi estacionario puede ser procesado por los filtros de todos los pasos, los efectos de borrosidad temporal de los métodos de des-correlación del estado del arto pueden ser evitados de esta manera. Ya que la toma panorámica de amplitud puede ser aplicada a los eventos del objeto de primer plano, la densidad de aplauso de primer plano original puede aproximadamente ser restaurada en contra-posición al sistema del estado del arte tal como por ejemplo presentado en el párrafo J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates" en la 116a. Convención de la AES, Berlin, Pre-impresión 6072, Mayo 2004 y J. Herre, K. Kjörning, J. Breebaart, et. al., "MPEG Surround – the ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding" in Proceedings of the 122nd AES Convention, Viena, Austria, Mayo 2007.

[0060] En otras palabras, en realizaciones, el des-compositor 110 puede ser adaptado para descomponer la señal de audio de entrada semánticamente en base al parámetro de entrada, en donde el parámetro de entrada puede ser provisto junto con la señal de audio de entrada, por ejemplo una información lateral. En tal realización, el des-compositor 110 puede ser adaptado para determinar el parámetro de entrada a partir de la señal de audio de entrada. En otras realizaciones, el des-compositor 110 puede ser adaptado para determinar el parámetro de entrada como un parámetro de control independiente de la señal de audio de entrada, lo que puede ser generado localmente, pre-establecido o puede también ser introducido por un usuario.

[0061] En realizaciones, el presentador 120 puede ser adaptado para obtener una distribución espacial de la primera señal presentada o la segunda señal presentada al aplicar una toma panorámica de amplitud de banda ancha. En otras palabras, de acuerdo con la descripción de la Figura 1b anterior, en lugar de generar una fuente semejante a punto, la ubicación de toma panorámica de la fuente puede ser variada temporalmente con el fin de generar una fuente de audio que tiene una cierta distribución espacial. En realizaciones, el presentador 120 puede ser adaptado para aplicar el ruido de paso de bajos generado localmente para la toma panorámica de amplitud, esto es, los factores de escalamiento para la toma panorámica de amplitud para, por ejemplo los amplificadores escalables 121 y 122 en la Figura 1b corresponden a un valor de ruido generado localmente, esto es son variables en el tiempo con un cierto ancho de banda.

- 5 [0062] Las realizaciones pueden ser adaptadas para ponerse en operación en un modo guiado o en un modo sin guiar. Por ejemplo, en un escenario guiado, refiriéndose a las líneas discontinuas, por ejemplo en la Figura 2, la des-correlación puede ser llevada a cabo mediante la aplicación de filtros de des-correlación de tecnología estándar controladas en una rejilla de tiempo burda para por ejemplo, la parte de fondo o parte ambiental solamente y obtener la correlación mediante la re-distribución de cada evento individual en por ejemplo, la parte del primer plano vía posicionamiento espacial variable en el tiempo utilizando toma panorámica de amplitud de banda ancha en una rejilla de tiempo mucho más fina. En otras palabras, en realizaciones, el presentador 120 puede ser adaptado para poner en operación des-correlacionadores para diferentes señales descompuestas en diferentes rejillas de tiempo, por ejemplo basadas en escalas de tiempo diferentes, que pueden estar en términos de velocidades de muestras diferentes o diferente retardo para los respectivos des-correlacionadores. En una realización, al llevar a cabo la separación de primer plano y de fondo, la parte del primer plano puede usar toma panorámica de amplitud, en donde la amplitud es cambiada en una rejilla de tiempo mucho más fina que la operación para un des-correlacionador con respecto a la parte de fondo.
- 10
- 15 [0063] Además, se enfatiza que para la des-correlación de, por ejemplo señales semejantes a aplauso, esto es señales con calidad aleatoria casi estacionaria, la posición espacial exacta de cada aplauso de primer plano individual puede no ser tanto de importancia crucial, más bien la recuperación de la distribución global de la multitud de eventos de aplauso. Realizaciones pueden tomar ventaja de este hecho y pueden operar en un modo sin guía. En tal modo, el factor de toma panorámica de amplitud mencionada anteriormente podría ser controlado mediante el ruido de paso de bajos. La Figura 3 ilustra un sistema mono a estéreo que implementa el escenario. La Figura 3 muestra un bloque de descomposición semántico 310 correspondiente al des-compositor 110 para descomponer la señal de entrada mono a una parte de señal descompuesta de primer plano y de fondo.
- 20
- 25 [0064] Como se puede ver de la Figura 3, la parte descompuesta de fondo de la señal es presentada por el D¹ 320 de todos los pasos. Luego, la señal des-correlacionada es provista junto con la parte descompuesta de fondo sin presentar a la mezcla ascendente 330, correspondiente al procesador 130. La parte de señal descompuesta de primer plano es provista a una etapa D² de toma panorámica de amplitud 340, que corresponde al presentador 120. El ruido de paso de bajos generado localmente 350 es también provisto a la etapa de toma panorámica de amplitud 340, que puede luego proveer la señal descompuesta de primer plano en una configuración de toma panorámica de amplitud a la mezcla ascendente 330. La etapa de toma panorámica de amplitud D² 340 puede determinar su salida al proveer un factor de escalamiento para una selección de amplitud entre dos de un conjunto estéreo de canales de audio. El factor de escalamiento k puede estar basado en el ruido de paso de bajos.
- 30
- 35 [0065] Como se puede ver de la Figura 3, hay solamente una flecha entre la toma panorámica de amplitud 340 y la mezcla ascendente 330. Esta flecha puede también representar señales de toma panorámica de amplitud, esto es, en el caso de mezcla ascendente estéreo, ya el canal izquierdo y el canal derecho. Como se puede ver en la Figura 3, la mezcla ascendente 330 correspondiente al procesador 130 es luego adaptada para procesar o combinar las señales descompuestas de fondo o primer plano para derivar la salida estéreo.
- 40
- 45 [0066] Otras realizaciones pueden ser procesamiento natural con el fin de derivar señales descompuestas de fondo y de primer plano o parámetros de entrada para descomposición. El des-compositor 110 puede ser adaptado para determinar la primera señal descompuesta y/o la segunda señal descompuesta en base a un método de separación transitorio. En otras palabras, el des-compositor 110 puede ser adaptado para determinar la primera o segunda señal descompuesta en base a un método de separación y la otra señal descompuesta en base a la diferencia entre la primera señal descompuesta determinada y la señal de audio de entrada. En otras realizaciones, la primera o segunda señal descompuesta puede ser determinada en base al método de separación transitorio y la otra señal descompuesta puede estar basada en la diferencia entre la primera o segunda señal descompuesta y la señal de audio de entrada.
- 50
- 55 [0067] El des-compositor 110 y/o el presentador 120 y/o el procesador 130 pueden comprender una etapa de monosynth DirAC y/o una etapa de síntesis DirAC y/o una etapa de función de DirAC. En realizaciones, el des-compositor 110 puede ser adaptado para descomponer la señal de audio de entrada, el presentador 120 puede ser adaptado para presentar la primera y/o segunda señales descompuestas y/o el procesador 130 puede ser adaptado para procesar la primera y/o segunda señales presentadas en términos de diferentes bandas de frecuencia.
- 60
- 65 [0068] Realizaciones pueden usar la siguiente aproximación para las señales semejantes a aplauso. En tanto que los componentes de primer plano pueden ser obtenidos mediante métodos de detección o separación transitorios, cf. Pulkki, Ville; "Spatial Sound Reproduction with Directional Audio Coding" in J. Audio Eng. Soc., Vol. 55, No. 6, 2007, el componente de fondo puede ser dado por la señal residual. La Figura 4 ilustra un ejemplo en donde un método apropiado para obtener un componente de fondo $x'(n)$ de, por ejemplo, una señal semejante a aplauso $x(n)$ para implementar la descomposición semántica 310 en la Figura 3, esto es, una realización del des-compositor 120. La Figura 4 muestra una señal de entrada discreta en el tiempo $x(n)$, que es introducida a una DFT 410 (DFT = transformada de Fourier discreta). La salida del bloque de DFT 410 es provista a un bloque para el alisamiento del espectro 420 y a un bloque de blanqueo espectral 430 para el blanqueo espectral en base a la salida de la DFT 410 y la salida de la etapa de espectro liso 430.

[0069] La salida de la etapa de blanqueo espectral 430 es luego provista a una etapa de proyección de pico espectral 440, que separa el espectro y provee dos salidas, esto es, un ruido y señal residual transitoria y una señal tonal. El ruido y señal residual transitoria es provista a un filtro de LPC 450 (LPC = codificación de predicción lineal) de los cuales la señal de ruido residual es provista a la etapa de mezcla 460 junto con la señal tonal como salida de la etapa de proyección de pico espectral 440. La salida de la etapa de mezcla 460 es luego provista a una etapa de formación espectral 470 que forma el espectro en base al espectro alisado provisto por la etapa de espectro alisado 420. La salida de la etapa de formación espectral 470 es luego provista al filtro de síntesis 480, esto es, una transformada de Fourier discreta inversa con el fin de obtener $x'(n)$ que representa el componente de fondo. El componente de primer plano puede luego ser derivado como la diferencia entre la señal de entrada y la señal de salida, esto es, $x(n)-x'(n)$.

[0070] Realizaciones de la presente invención se pueden poner en operación en aplicaciones de realidad virtual tales como por ejemplo juegos en 3D. En tales aplicaciones, la síntesis de fuentes de ruido con una gran extensión espacial puede ser complicada y compleja cuando está basada en conceptos convencionales. Tales fuentes podrían ser, por ejemplo una costa, una parvada de aves, caballos galopando, la división de soldados marchando o una audiencia que aplaude. Comúnmente, tales eventos de sonido son espaciados como un gran grupo de fuentes semejantes a punto, lo que conduce a implementaciones computacionalmente complejas cf. Wagner, Andreas; Walther, Andreas; Melchoir, Frank; Strauß, Michael; "Generation of Highly Immersive Atmospheres for Wave Field Synthesis Reproduction" en la 116a. Convención de la EAS, Berlin, 2004.

[0071] Las realizaciones pueden llevar a cabo un método que efectúa la síntesis de la extensión de fuentes de sonido plausiblemente pero al mismo tiempo, que tienen complejidad estructural y computacional más baja. Las realizaciones pueden estar basadas en DirAC (DirAC = codificación de audio direccional), cf. Pulkki, Ville; "Spatial Sound Reproduction with Directional Audio Coding" in J. Audio Eng. Soc., Vol. 55, No. 6, 2007. En otras palabras, en realizaciones, el des-compositor 110 y/o el presentador 120 y/o el procesador 130 pueden ser adaptados para procesar señales de DirAC. En otras palabras, el des-compositor 110 puede comprender etapas de monosynth DirAC, el presentador 120 puede comprender una etapa de síntesis de DirAC y/o el procesador puede comprender una etapa de fusión de DirAC.

[0072] Las realizaciones pueden estar basadas en procesamiento de DirAC, por ejemplo, utilizando solamente dos estructuras de síntesis, por ejemplo, una para fuentes de sonido de primer plano y una para fuentes de sonido de fondo. El sonido de primer plano puede ser aplicado a una sola corriente de DirAC con datos direccionales controlados, dando como resultado la percepción de fuentes semejantes a punto cercanas. El sonido de fondo puede también ser producido al usar una sola corriente directa con datos direccionales controlados diferentemente, lo que conduce a la percepción de objetos de sonido esparcidos espacialmente. Luego las dos corrientes de DirAC pueden ser fusionadas y descodificadas para el voltaje de altavoz arbitrario o para audifonos, por ejemplo.

[0073] La Figura 5 ilustra una síntesis de fuentes de sonido que tienen una extensión espacialmente grande. La Figura 5 muestra un bloque de monosynth superior 610, que crea una corriente de mono-DirAC que conduce a una percepción de una fuente de sonido semejante a punto cercano, tales como los aplaudidores más cercanos de una audiencia. El bloque de monosynth inferior 620 es usado para crear una corriente de mono-DirAC que conduce a la percepción de sonido esparcido espacialmente que es por ejemplo para generar sonido de fondo como el sonido de aplauso de la audiencia. Las salidas de los dos bloques de monosynth DirAC 610 y 620 son luego fusionadas en la etapa de fusión de DirAC 630. La Figura 5 muestra que solamente dos bloques de síntesis de DirAC 610 y 620 son usados en esta realización. Uno de ellos es usado para crear los eventos de sonido, que están en el primer plano, tales como las aves más cercanas o personas más cercanas en una audiencia que aplaude y el otro genera un sonido de fondo, el sonido de la parvada de aves continuo, etc.

[0074] El sonido de primer plano es convertido a una corriente de mono-DirAC con el bloque de DirAC-monosynth 610 de una manera que los datos de azimuth se mantienen constantes con la frecuencia, sin embargo, cambiados aleatoriamente o controlados por un proceso en el tiempo externo. El parámetro de difusividad ψ es ajustado a cero, esto es, que representa una fuente semejante a punto. La entrada de audio al bloque 610 se supone que es consiste de sonidos no traslapantes temporalmente, tales como llamadas de aves distintas o aplausos de mano, lo que genera la percepción de fuentes de sonido cercanas, tales como aves o personas que aplauden. La extensión espacial de los eventos de sonido de primer plano es controlada al ajustar el θ y $\theta_{\text{intervalo-primer plano}}$, lo que significa que eventos de sonido individuales serán percibidos en las direcciones $\theta+\theta_{\text{intervalo-primer plano}}$, sin embargo, un solo evento puede ser percibido semejante a punto. En otras palabras, fuentes de sonido semejantes a punto son generadas en donde las posiciones posibles del punto están limitadas al intervalo $\theta+\theta_{\text{intervalo-primer plano}}$.

[0075] El bloque de fondo 620 toma como corriente de audio de entrada, una señal, que contiene todos los otros eventos de sonido no presentes en la corriente de audio de primer plano, que pretende incluir lotes de eventos de sonido temporalmente traslapantes, por ejemplo cientos de aves o un número mayor de aplaudidores lejanos. Los valores de azimuth anexados son luego ajustados aleatoriamente tanto en tiempo como frecuencia, dentro de valores de azimuth de restricción dados $\theta+\theta_{\text{intervalo-fondo}}$. La extensión espacial de los sonidos de fondo puede así ser sintetizada con baja complejidad computacional. La difusividad ψ puede también ser controlada. Si fuera agregado, el descodificador de DirAC aplicaría el sonido a todas direcciones, lo que puede ser usado cuando la fuente de

sonido rodea al usuario que escucha totalmente. Si no lo rodea, la difusividad puede ser mantenida baja o cercana a cero o cero en algunas realizaciones.

5 **[0076]** Las realizaciones de la presente invención pueden proveer la ventaja de que calidad perceptual superior de los sonidos presentados puede ser obtenida a un costo computacional moderado. Las realizaciones pueden permitir una implementación modular de presentación de sonido espacial como por ejemplo se muestra en la Figura 5.

10 **[0077]** Dependiendo de ciertos requerimientos de implementación de los métodos de la invención, los métodos de la invención pueden ser implementados en elementos físicos o elementos de programación. La implementación puede ser efectuada utilizando un medio de almacenamiento digital y particularmente, una memoria instantánea, un disco, un DVD o un CD que tiene señales de control que se pueden leer electrónicamente almacenadas en el mismo, que cooperan con el sistema de ordenador programable, de tal manera que los métodos de la invención son efectuados. En general, la presente invención es por consiguiente un producto de programas de ordenador con códigos de programa almacenados en un portador que se puede leer por la máquina, los códigos de programa son operativos para efectuar los métodos de la invención cuando el producto de programas de ordenador se ejecuta en un ordenador. En otras palabras, los métodos de la invención son por consiguiente un programa de ordenador que tiene códigos de programa para efectuar por lo menos uno de los métodos de la invención cuando el programa de ordenador se ejecuta en un ordenador.

REIVINDICACIONES

- 5 1. Un aparato (100) para determinar una señal de audio de multi-canal de salida espacial sobre la base de una señal de audio de entrada, que comprende:
- 10 un des-compositor (110) para descomponer la señal de audio de entrada para obtener una primera señal descompuesta que tiene una primera propiedad semántica, donde la primera señal descompuesta comprende una parte de señal de primer plano de la señal de audio de entrada, y una segunda señal descompuesta que tiene una
- 15 segunda propiedad semántica que es diferente de la primera propiedad semántica, donde la segunda señal descompuesta comprende una parte de señal de fondo de la señal de audio de entrada, donde el des-compositor (110) está adaptado para determinar la segunda señal descompuesta que comprende la parte de señal de fondo de la señal de audio de entrada mediante un método de separación de transitorios y la primera señal descompuesta que comprende la parte de señal de primer plano de la señal de audio de entrada sobre la base de una diferencia entre la segunda señal descompuesta y la señal de audio de entrada;
- 20 un presentador (120) para presentar la primera señal descompuesta utilizando una primera característica de presentación para obtener una primera señal presentada que tiene la primera propiedad semántica y para presentar la segunda señal descompuesta utilizando una segunda característica de presentación para obtener una segunda señal presentada que tiene la segunda propiedad semántica, donde la primera característica de presentación y la segunda característica de presentación son diferentes entre sí, donde el presentador (120) está adaptado para presentar la primera señal descompuesta según una característica de audio de primer plano como la primera característica de presentación y para presentar la segunda señal descompuesta según una característica de audio de fondo como la segunda característica de presentación; y
- 25 un procesador (130) para procesar la primera señal presentada y la segunda señal presentada para obtener la señal de audio de multi-canal de salida espacial.
- 30 2. El aparato (100) de acuerdo con la reivindicación 1, en el que el presentador (120) está adaptado para presentar la primera señal descompuesta de manera que la primera característica de presentación no tiene una característica de introducción de retardo o de manera que la primera característica de presentación tiene una característica de introducción de retardo que tiene una primera cantidad de retardo y en el que la segunda característica de presentación tiene una segunda cantidad de retardo que es mayor que la primera cantidad de retardo.
- 35 3. El aparato (100) de acuerdo con la reivindicación 1 ó 2, en el que el presentador (120) está adaptado para presentar la primera señal descompuesta mediante toma panorámica de amplitud como primera característica de presentación y para des-correlacionar la segunda señal descompuesta para obtener una segunda señal des-correlacionada como segunda característica de presentación.
- 40 4. El aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que el presentador (120) está adaptado para presentar las primeras y segundas señales presentadas, donde cada una tiene tantos componentes como canales en la señal de audio de multi-canal de salida espacial y el procesador (130) está adaptado para combinar los componentes de las primeras y segundas señales presentadas para obtener la señal de audio de multi-canal de salida espacial.
- 45 5. El aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que el presentador (120) está adaptado para presentar las primeras y segundas señales presentadas, cada una teniendo menos componentes que la señal de audio de multi-canal de salida espacial y en donde el procesador (130) está adaptado para la mezcla ascendente de los componentes de las primeras y segundas señales presentadas para obtener la señal de audio de
- 50 multi-canal de salida espacial.
- 55 6. El aparato (100) de acuerdo con una cualquiera de las reivindicaciones 3 a 5, en el que el presentador (120) está adaptado para presentar la segunda señal descompuesta mediante un filtrado paso-todo de la segunda señal para obtener la segunda señal des-correlacionada.
- 60 7. El aparato (100) de acuerdo con la reivindicación 1, en el que el des-compositor (110) está adaptado para determinar un parámetro de entrada como parámetro de control a partir de la señal de audio de entrada.
- 65 8. El aparato (100) de acuerdo con una cualquiera de las reivindicaciones 3 a 7, en el que el presentador (120) está adaptado para obtener una distribución espacial de la primera o segunda señal presentada mediante la aplicación de una toma panorámica de amplitud de banda ancha.
9. El aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en el que el presentador (120) está adaptado para presentar la primera señal descompuesta y la segunda señal descompuesta en base a diferentes rejillas de tiempo.

10. El aparato (100) de acuerdo con una cualquiera de las reivindicaciones 1 a 9, en el que el des-compositor (110) está adaptado para descomponer la señal de audio de entrada, el presentador (120) está adaptado para presentar la primera y/o segunda señales descompuestas y/o el procesador (130) está adaptado para procesar las primeras y/o segundas señales presentadas en términos de diferentes bandas de frecuencia.
- 5 11. El aparato (100) de acuerdo con la reivindicación 1, en el que el des-compositor (110) comprende:
- un bloque de DFT (410) para convertir la señal de audio de entrada en el dominio DFT;
- 10 un bloque para el alisamiento espectral (420) para alisar una salida del bloque DFT (410);
- un bloque de blanqueo espectral (430) para el blanqueo espectral de la salida del bloque DFT (410) en base a una salida del bloque de alisamiento espectral (420);
- 15 una etapa de proyección de pico espectral (440) para separar una salida de espectro del bloque de blanqueo espectral (430) y para proporcionar, como primera salida, un ruido y señal residual transitoria y, como segunda salida, una señal tonal;
- un filtro de LPC (450) para procesar el ruido y la señal residual transitoria para obtener una señal residual de ruido;
- 20 una etapa de mezcla (460) para mezclar la señal residual de ruido y la señal tonal;
- una etapa de formación espectral (470) para formar la salida de espectro de la etapa de mezcla (460) en base a la salida del bloque de espectro alisado (420); y
- 25 un filtro de síntesis (480) para realizar una transformada de Fourier discreta inversa para obtener la segunda señal descompuesta que comprende la parte de señal de fondo de la señal de audio de entrada.
12. Un método para determinar una señal de audio de multi-canal de salida espacial en base a una señal de audio de entrada y un parámetro de entrada, que comprende las etapas de:
- 30 descomponer la señal de audio de entrada para obtener una primera señal descompuesta que tiene una primera propiedad semántica, donde la primera señal descompuesta comprende una parte de señal de primer plano de la señal de audio de entrada, y una segunda señal descompuesta que tiene una segunda propiedad semántica que es diferente de la primera propiedad semántica, donde la segunda señal descompuesta comprende una parte de señal de fondo de la señal de audio de entrada; donde la segunda señal descompuesta que comprende la parte de señal de fondo de la señal de audio de entrada se determina mediante un método de separación de transitorios y la primera señal descompuesta que comprende la parte de señal de primer plano de la señal de audio de entrada se determina en base a una diferencia entre la segunda señal descompuesta y la señal de audio de entrada;
- 40 presentar la primera señal descompuesta utilizando una primera característica de presentación para obtener una primera señal presentada que tiene la primera propiedad semántica;
- presentar la segunda señal descompuesta utilizando una segunda característica de presentación para obtener una segunda señal presentada que tiene la segunda propiedad semántica, donde la primera característica de presentación y la segunda característica de presentación son diferentes entre sí, donde la primera señal descompuesta se presenta según una característica de audio de primer plano como la primera característica de presentación y la segunda señal descompuesta se presenta según una característica de audio de fondo como la segunda característica de presentación; y
- 45 procesar la primera señal presentada y la segunda señal presentada para obtener la señal de audio de multi-canal de salida espacial.
- 50 13. El método de acuerdo con la reivindicación 12, donde la etapa de descomponer comprende:
- 55 convertir la señal de audio de entrada en un dominio DFT utilizando una DFT;
- alisar espectralmente una salida de la etapa de conversión;
- 60 blanquear espectralmente una salida de la etapa de conversión en base a una salida de la etapa de alisamiento espectral;
- separar, mediante proyección de pico espectral, una salida de espectro de la etapa de blanqueo espectral y proporcionar, como primera salida, un ruido y señal residual transitoria y, como segunda salida, una señal tonal;
- 65 procesar, mediante filtrado LPC, el ruido y la señal residual transitoria para obtener una señal residual de ruido;

mezclar la señal residual de ruido y la señal tonal;

5 formar un salida de espectro de la etapa de mezcla en base a una salida de la etapa de alisado espectral; y

realizar una transformada de Fourier discreta inversa en una salida de la etapa de formar para obtener la segunda señal descompuesta que comprende la parte de señal de fondo de la señal de audio de entrada.

10 14. Un programa de ordenador que tiene un código de programa para efectuar el método de acuerdo con la reivindicación 13, donde el código de programa se ejecuta en un ordenador o un procesador.

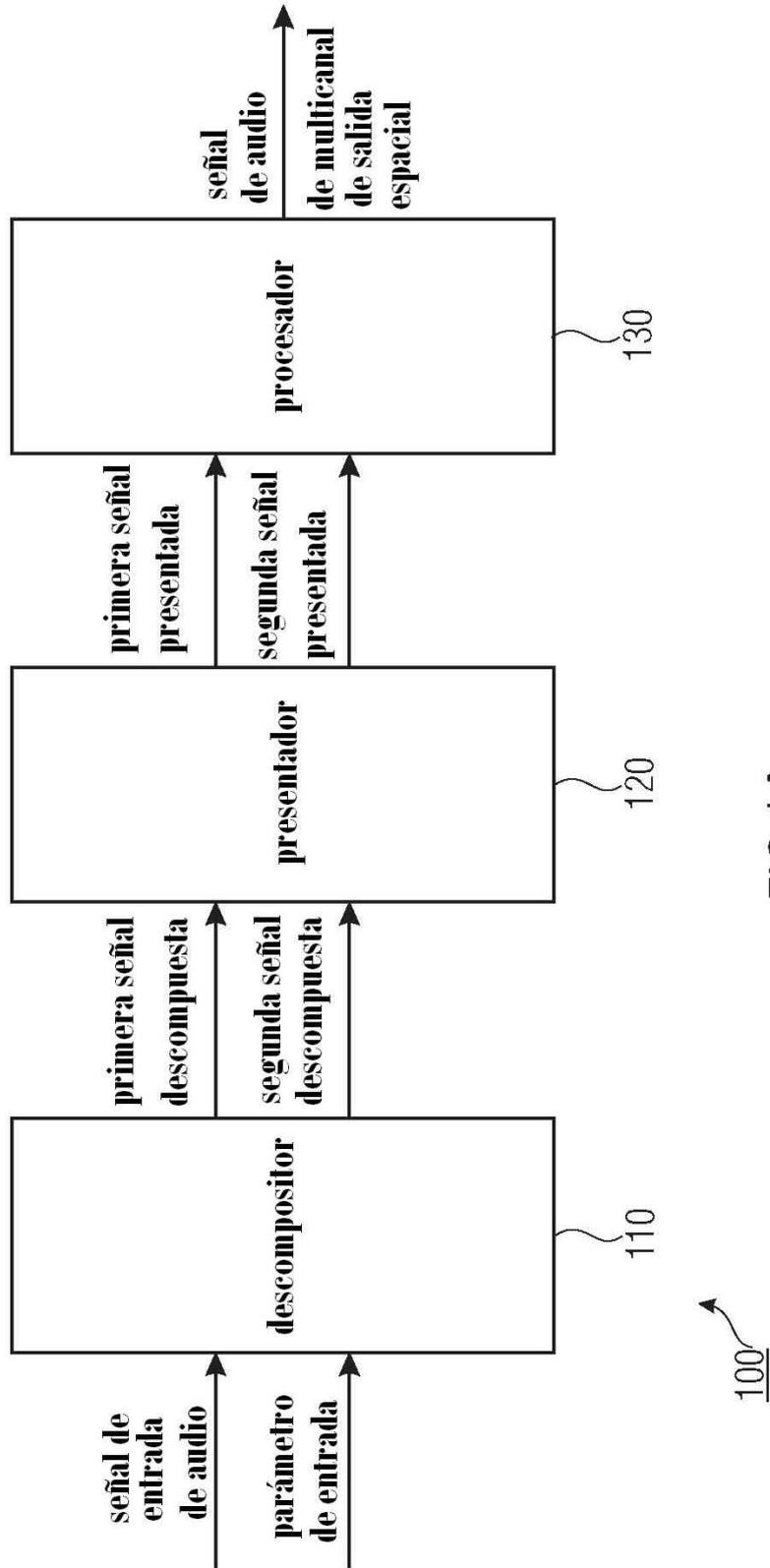


FIG 1A

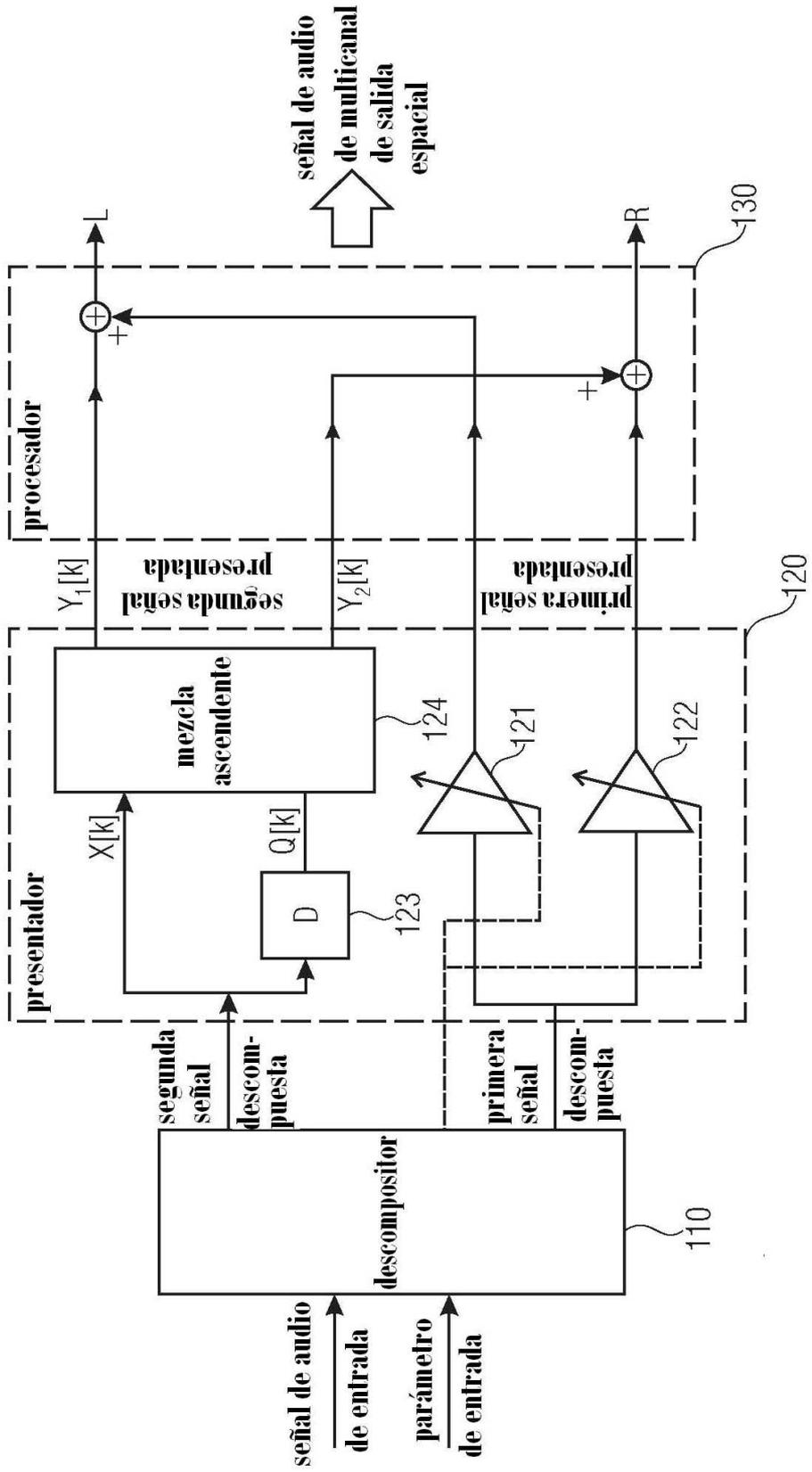


FIG 1B

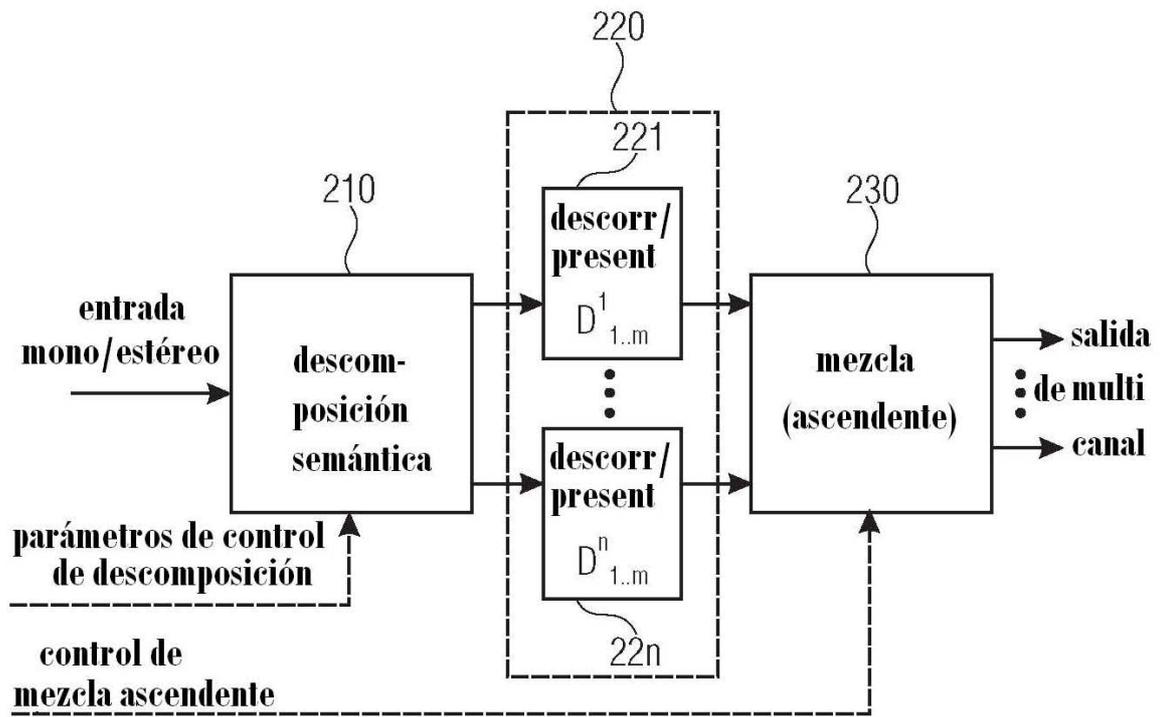


FIG 2

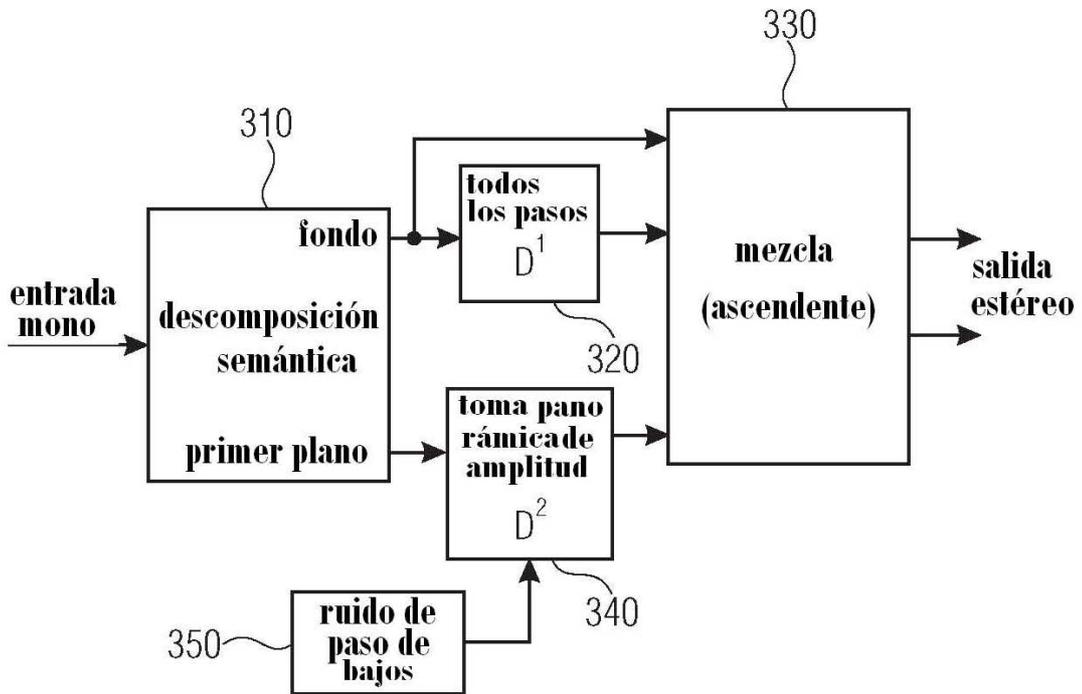


FIG 3

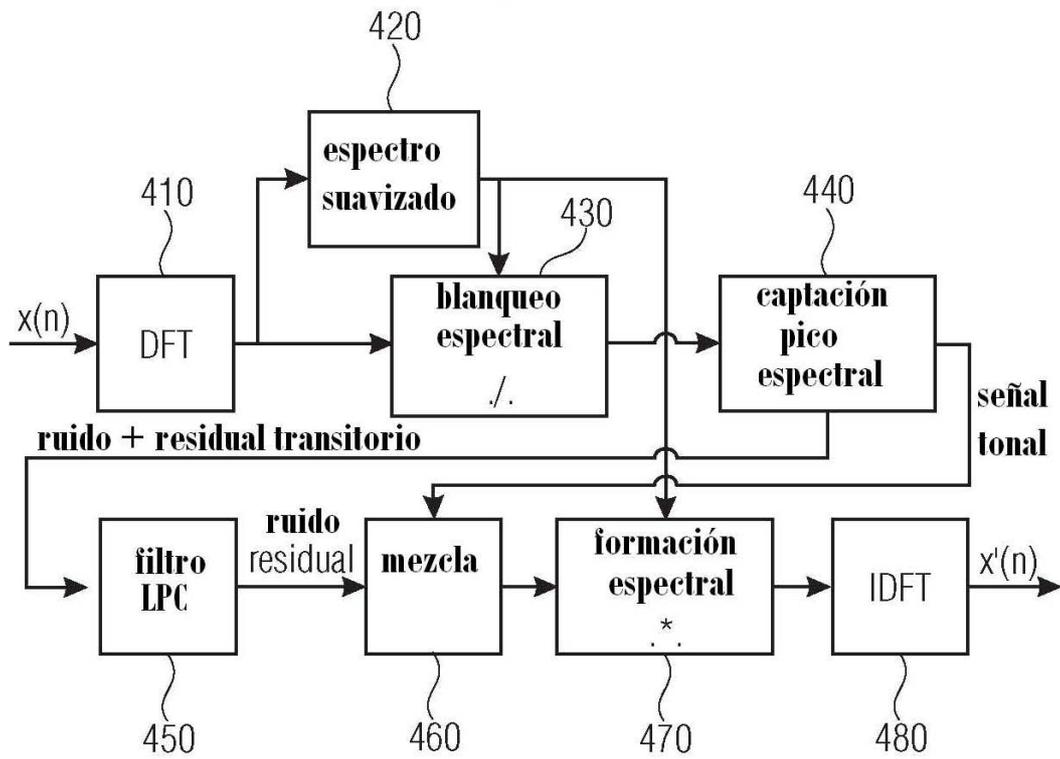


FIG 4



FIG 5

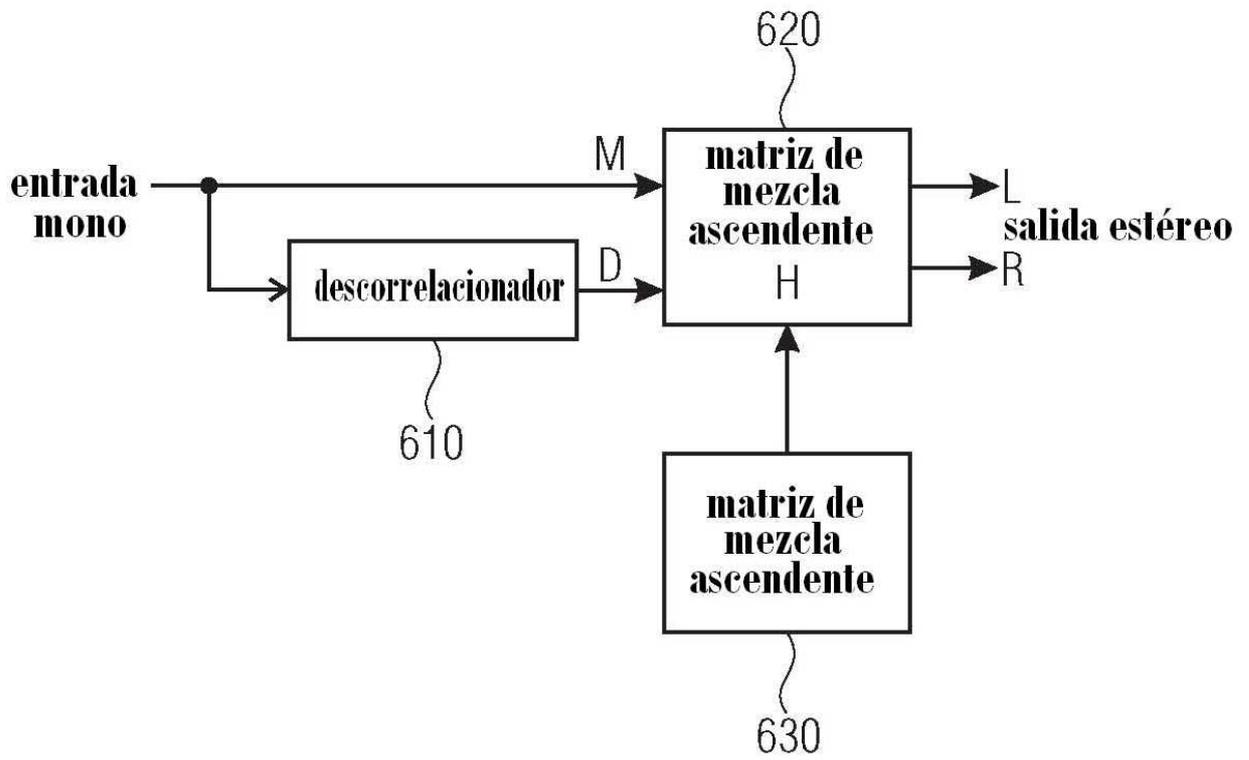


FIG 6
(ESTADO DE LA TÉCNICA)

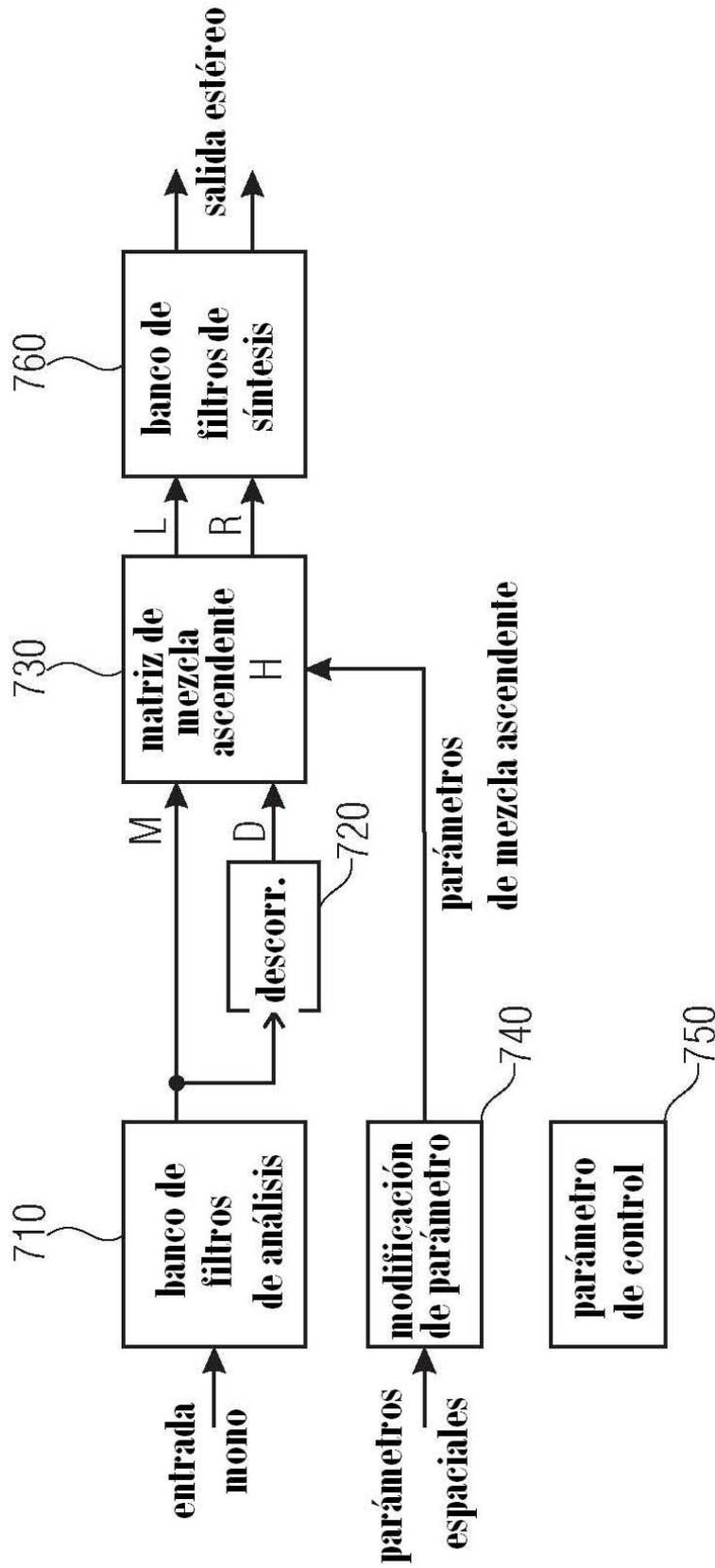


FIG 7

(ESTADO DE LA TÉCNICA)