

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 392 609**

51 Int. Cl.:

H04S 7/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **09777815 .3**

96 Fecha de presentación: **11.08.2009**

97 Número de publicación de la solicitud: **2311274**

97 Fecha de publicación de la solicitud: **20.04.2011**

54

Título: **Aparato para determinar una señal de audio multicanal de salida espacial**

30

Prioridad:

13.08.2008 US 88505 P

28.10.2008 EP 08018793

45

Fecha de publicación de la mención BOPI:

12.12.2012

45

Fecha de la publicación del folleto de la patente:

12.12.2012

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:

ARIZTI ACHA, Monica

ES 2 392 609 T3

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

DESCRIPCIÓN

Aparato para determinar una señal de audio multicanal de salida espacial

5 La presente invención pertenece al campo del procesamiento de audio, especialmente el procesamiento de propiedades de audio espaciales.

10 El procesamiento y/o la codificación de audio han avanzado de muchas maneras. Cada vez se genera más demanda para aplicaciones de audio espaciales. En muchas aplicaciones el procesamiento de señales de audio se utiliza para decorrelacionar o transformar señales. Tales aplicaciones pueden llevar a cabo, por ejemplo, un mezclado en sentido ascendente de mono a estéreo, un mezclado de mono/estéreo a multicanal, reverberación artificial, ampliación de estéreo o mezclado/transformación interactivo de usuario.

15 Para determinadas clases de señales como por ejemplo señales similares a ruido, como por ejemplo señales similares a aplausos, los métodos y sistemas convencionales se ven afectados por o bien una calidad de percepción insatisfactoria o bien, si se usa un enfoque orientado al objeto, una elevada complejidad computacional debido al número de eventos auditivos que deben modelizarse o procesarse. Otros ejemplos de material de audio, que es problemático, son generalmente material de ambiente tal como, por ejemplo, el ruido que se emite por una bandada de pájaros, una orilla del mar, caballos galopando, una división de soldados marchando, etc.

20 Los conceptos convencionales usan, por ejemplo, codificación de estéreo paramétrica o de sonido envolvente MPEG (MPEG = *Moving Pictures Expert Group*, Grupo de expertos de imágenes en movimiento). La figura 6 muestra una aplicación típica de un decorrelacionador en una mezcladora en sentido ascendente de mono a estéreo. La figura 6 muestra una señal de entrada mono proporcionada a un decorrelacionador 610, que proporciona una señal de entrada decorrelacionada en su salida. La señal de entrada original se proporciona a una matriz 620 de mezclado en sentido ascendente junto con la señal decorrelacionada. Dependiendo de parámetros 630 de control de mezclado en sentido ascendente, se transforma en una señal de salida estéreo. El decorrelacionador 610 de señales genera una señal *D* decorrelacionada alimentada a la fase 620 de formación de matriz junto con la señal *M* mono sin modificar. Dentro de la matriz 620 de mezclado, los canales estéreo L (L = canal estéreo izquierdo) y R (R = canal estéreo derecho) están formados según una matriz *H* de mezclado. Los coeficientes en la matriz *H* pueden fijarse de manera dependiente de la señal o controlarse por un usuario.

30 Alternativamente, la matriz puede controlarse mediante información complementaria, transmitida junto con la mezcla en sentido descendente, que contiene una descripción paramétrica de cómo mezclar en sentido ascendente las señales de la mezcla en sentido descendente para formar la salida multicanal deseada. Esta información complementaria espacial se genera habitualmente mediante un codificador de señales antes del proceso de mezclado en sentido ascendente.

35 Esto se realiza normalmente en la codificación de audio espacial paramétrica como, por ejemplo, en estéreo paramétrico, véase J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates" en 116ª Convención de AES, Berlín, preimpresión 6072, mayo de 2004 y en sonido envolvente MPEG, véase J. Herre, K. Kjörling, J. Breebaart, *et al.*, "MPEG Surround - the ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding" en Procedimientos de la 122ª Convención de AES, Viena, Austria, mayo de 2007. Una estructura típica de un decodificador estéreo paramétrico se muestra en la figura 7. En este ejemplo, el proceso de decorrelación se realiza en un dominio de transformada, que está indicado por el banco 710 de filtros de análisis, que transforma una señal mono de entrada en el dominio de transformada como, por ejemplo, el dominio de frecuencia en cuanto a un número de bandas de frecuencia.

40 En el dominio de frecuencia, el decorrelacionador 720 genera la señal decorrelacionada correspondiente, que debe mezclarse en sentido ascendente en la matriz 730 de mezclado en sentido ascendente. La matriz 730 de mezclado en sentido ascendente considera parámetros de mezclado en sentido ascendente, que se proporcionan mediante la caja 740 de modificación de parámetros, a la que se le proporcionan parámetros de entrada espaciales y que está acoplada a una fase 750 de control de parámetros. En el ejemplo mostrado en la figura 7, los parámetros espaciales pueden modificarse por un usuario o mediante herramientas adicionales tales como, por ejemplo, postprocesamiento para transformación/presentación binaural. En este caso, los parámetros de mezclado en sentido ascendente pueden fusionarse con los parámetros procedentes de los filtros binaurales para formar los parámetros de entrada para la matriz 730 de mezclado en sentido ascendente. La medición de los parámetros puede llevarse a cabo mediante el bloque 740 de modificación de parámetros. La salida de la matriz 730 de mezclado en sentido ascendente se proporciona entonces a un banco 760 de filtros de síntesis, que determina la señal de salida estéreo.

50 Como se describió anteriormente, la salida *L/R* de la matriz *H* de mezclado puede calcularse a partir de la señal *M* de salida mono y la señal *D* decorrelacionada, por ejemplo según

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

En la matriz de mezclado, la cantidad de sonido decorrelacionado alimentado a la salida puede controlarse basándose en los parámetros transmitidos tales como, por ejemplo, configuraciones de ICC (ICC = *Interchannel Correlation*, correlación intercanal) y/o mixtas o definidas por el usuario.

- 5 Otro enfoque convencional se establece mediante el método de permutación temporal. Una propuesta dedicada sobre la decorrelación de señales similares a aplausos puede encontrarse, por ejemplo, en Gerard Hotho, Steven van de Par, Jeroen Breebaart, "Multichannel Coding of Applause Signals," en *EURASIP Journal on Advances in Signal Processing*, vol. 1, art. 10, 2008. En este caso, una señal de audio monofónica se segmenta en segmentos de tiempo solapantes, que se permitan temporalmente de manera pseudoaleatoria dentro de un "superbloque" para formar los canales de salida decorrelacionados. Las permutaciones son independientes entre sí para un número n de canales de salida.

Otro enfoque es el intercambio de canal alternante de copia original y retardada con el fin de obtener una señal decorrelacionada, véase la solicitud de patente alemana 102007018032.4-55.

- 15 En algunos sistemas orientados al objeto 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 EAS, Berlín, 2004, se describe cómo crear una escena inmersiva a partir de muchos objetos tales como por ejemplo palmadas individuales, mediante la aplicación una síntesis de campo de ondas.

- 20 Aún otro enfoque es la denominada "codificación de audio direccional" (DirAC = *Directional Audio Coding*), que es un método para la representación de sonido espacial, aplicable para diferentes sistemas de reproducción de sonido, véase Pulkki, Ville, "Spatial Sound Reproduction with Directional Audio Coding" en *J. Audio Eng. Soc.*, vol. 55, n.º 6, 2007. En la parte de análisis, la difusividad y la dirección de llegada del sonido se estiman en una única ubicación dependiendo del tiempo y de la frecuencia. En la parte de síntesis, señales de micrófono se dividen en primer lugar en partes no difusas y difusas y entonces se reproducen usando diferentes estrategias.

- 25 Los enfoques convencionales tienen varias desventajas. Por ejemplo, el mezclado en sentido ascendente guiado o no guiado de señales de audio que tienen contenido, tal como un aplauso, puede requerir una decorrelación fuerte. Por consiguiente, por un lado, es necesaria una decorrelación fuerte para restaurar la sensación ambiente de estar, por ejemplo, en una sala de conciertos. Por otro lado, los filtros de decorrelación adecuados tales como, por ejemplo, los filtros todo paso, degradan una reproducción de calidad de eventos transitorios, como una única palmada con las manos introduciendo efectos de difuminado temporales tales como ecos previos y posteriores y llamada de filtro. Además, la panorámica espacial de eventos de palmadas individuales debe realizarse en una cuadrícula de tiempo bastante fina, mientras que la decorrelación ambiente debe ser cuasi estacionaria a lo largo del tiempo.

- 30 Los sistemas del estado de la técnica según J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates" en la 116ª Convención de AES, Berlín, preimpresión 6072, mayo de 2004 y J. Herre, K. Kjörling, J. Breebaart, *et al.*, "MPEG Surround - the ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding" en Procedimientos de la 122ª Convención de AES, Viena, Austria, mayo de 2007, comprometen la resolución temporal frente a la estabilidad ambiente y la degradación de calidad transitoria frente a la decorrelación ambiente.

- 35 Un sistema que utilice el método de permutación temporal, por ejemplo, presentará una degradación perceptible del sonido de salida debido a una determinada calidad repetitiva en la señal de audio de salida. Esto se debe al hecho de que un mismo segmento de la señal de entrada aparece inalterado en cada canal de salida, aunque en un punto de tiempo diferente. Además, para evitar un aumento de la densidad de aplausos, deben omitirse algunos canales originales en el mezclado en sentido ascendente y, por tanto, puede perderse algún evento auditivo importante en la mezcla en sentido ascendente resultante.

- 45 En sistemas orientados al objeto, normalmente tales eventos de sonido están espaciados como un gran grupo de fuentes de tipo puntual, lo que conduce a una implementación computacionalmente compleja.

- 50 El documento GB 2353193 A da a conocer un método y un aparato de procesamiento de sonido que pueden realizar un procesamiento de sonido sobre señales de audio de entrada que contienen una pluralidad de componentes de señal que son diferentes en condiciones de procesamiento de sonido deseadas, de manera que permite la reproducción del sonido natural. La señal de audio de entrada se separa en una pluralidad de componentes de señal separadas, y cada componente de señal se somete a un procesamiento de sonido individual, incluyendo por ejemplo análisis espectral, y la pluralidad de componentes de señal separadas se emiten como al menos una señal de audio. La pluralidad de componentes de señal separadas se sintetizan para dar una señal de audio sintetizada, o la pluralidad de componentes

de señal separadas se emiten por separado como señales de audio. El método es útil para la radiodifusión en directo, en la que la señal de audio de entrada contiene una componente de sonido ambiente y una componente de habla en el acto. También es útil para ayudar a personas con una mala función auditiva.

El objeto de la presente invención es proporcionar un concepto mejorado para el procesamiento de audio espacial.

5 Este objeto se consigue mediante un aparato según la reivindicación 1 y un método según la reivindicación 11.

Es un hallazgo de la presente invención que una señal de audio puede descomponerse en varias componentes a las que puede estar adaptada una transformación espacial, por ejemplo, en cuanto a una decorrelación o en cuanto a un enfoque de panorámica de amplitud. Dicho de otro modo, la presente invención se basa en el hallazgo de que, por ejemplo, en un escenario con múltiples fuentes de audio, las fuentes de primer plano y de segundo plano pueden distinguirse y transformarse o decorrelacionarse de manera diferente. Generalmente pueden distinguirse diferentes extensiones y/o profundidades espaciales de objetos de audio.

Uno de los puntos clave de la presente invención es la descomposición de señales, como el sonido que se origina de un público que aplaude, una bandada de pájaros, una orilla del mar, caballos galopando, una división de soldados marchando, etc. en una parte de primer plano y una de segundo plano, de modo que la parte de primer plano contiene eventos auditivos individuales originados de, por ejemplo, fuentes cercanas y la parte de segundo plano contiene el ambiente de los eventos lejanos fusionados perceptivamente. Antes del mezclado final, estas dos partes de señal se procesan por separado, por ejemplo, con el fin de sintetizar la correlación, transformar una escena, etc.

Las realizaciones no se limitan a distinguir sólo las partes de primer plano y de segundo plano de la señal, puede distinguir múltiples partes de audio diferentes, que pueden transformarse o decorrelacionarse todas de manera diferente.

En general, las señales de audio pueden descomponerse en n partes semánticas diferentes mediante las realizaciones, que se procesan por separado. La descomposición/el procesamiento separado de componentes semánticas diferentes puede llevarse a cabo en el dominio de tiempo y/o de frecuencia mediante las realizaciones.

Las realizaciones pueden proporcionar la ventaja de una calidad de percepción superior del sonido transformado a un coste de cálculo moderado. Las realizaciones proporcionan con ello un método de decorrelación/transformación novedoso que ofrece una calidad de percepción elevada a costes moderados, especialmente para material de audio crítico similar a aplausos u otro material de ambiente similar como, por ejemplo, el ruido que se emite por una bandada de pájaros, una orilla del mar, caballos galopando, una división de soldados marchando, etc.

Las realizaciones de la presente invención se detallarán con la ayuda de las figuras adjuntas, en las que

- 30 la figura 1a muestra un aparato para determinar una señal de audio multicanal de audio espacial;
- la figura 1b muestra un diagrama de bloques de otro aparato;
- la figura 2 muestra un aparato que ilustra una multiplicidad de señales descompuestas;
- 35 la figura 3 ilustra una realización de la invención con una descomposición semántica de primera plano y una de segundo plano;
- la figura 4 ilustra un ejemplo de un método de separación transitoria para obtener una componente de señal de segundo plano;
- 40 la figura 5 ilustra una síntesis de fuentes de sonido que tienen espacialmente una gran extensión;
- la figura 6 ilustra una aplicación del estado de la técnica de un decorrelacionador en dominio de tiempo en una mezcladora en sentido ascendente de mono a estéreo; y
- la figura 7 muestra otra aplicación del estado de la técnica de un decorrelacionador en dominio de frecuencia en un escenario de mezcladora en sentido ascendente de mono a estéreo.

La figura 1 muestra un aparato 100 para determinar una señal de audio multicanal de salida espacial basándose en una señal de audio de entrada. En algunas realizaciones, el aparato puede estar adaptado para basar adicionalmente la señal de audio multicanal de salida espacial en un parámetro de entrada. El parámetro de entrada puede generarse localmente o proporcionarse con la señal de audio de entrada, por ejemplo, como información complementaria.

En la figura 1, el aparato 100 comprende un descomponedor 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 primera propiedad semántica.

5 El aparato 100 comprende además un transformador 120 para transformar la primera señal descompuesta usando una primera característica de transformación para obtener una primera señal transformada que tiene la primera propiedad semántica y para transformar la segunda señal descompuesta usando una segunda característica de transformación para obtener una segunda señal transformada que tiene la segunda propiedad semántica.

10 Una propiedad semántica puede corresponder a una propiedad espacial, tal como cerca o lejos, enfocada o amplia, y/o a una propiedad dinámica tal como por ejemplo si una señal es tonal, estacionaria o transitoria y/o a una propiedad de dominancia tal como por ejemplo si la señal es de primer plano o de segundo plano, a una medida de las mismas respectivamente.

Además, el aparato 100 comprende un procesador 130 para procesar la primera señal transformada y la segunda señal transformada para obtener la señal de audio multicanal de salida espacial.

15 Dicho de otro modo, el descomponedor 110 está adaptado para descomponer la señal de audio de entrada, En algunas realizaciones, basándose en el parámetro de entrada. La descomposición de la señal de audio de entrada está adaptada a las propiedades semánticas, por ejemplo espaciales, de diferentes partes de la señal de audio de entrada. Además, la transformación llevada a cabo por el transformador 120 según las características de transformación primera y segunda también puede estar adaptada a las propiedades espaciales, lo que permite, por ejemplo en un escenario en el que la primera señal descompuesta corresponde a una señal de audio de segundo plano y la segunda señal descompuesta corresponde a una señal de audio de primer plano, una transformación diferente o pueden aplicarse decorrelacionadores, al revés respectivamente. En lo sucesivo el término “de primer plano” se entiende que se refiere a un objeto de audio que es dominante en un entorno de audio, de modo que un posible oyente notaría un objeto de audio de primer plano. Una fuente u objeto de audio de primer plano puede distinguirse o diferenciarse de una fuente u objeto de audio de segundo plano. Una fuente u objeto de audio de segundo plano puede no poder notarse por un posible oyente en un entorno de audio ya que es menos dominante que una fuente u objeto de audio de primer plano. En las realizaciones las fuentes u objetos de audio de primer plano pueden ser, pero no se limitan a, una fuente de audio de tipo puntual, pudiendo corresponder las fuentes u objetos de audio de segundo plano a fuentes u objetos de audio espacialmente más lejanos.

20 Dicho de otro modo, en las realizaciones la primera característica de transformación puede basarse en o hacerse coincidir con la primera propiedad semántica y la segunda característica de transformación puede basarse en o hacer coincidir con la segunda propiedad semántica. En una realización, la primera propiedad semántica y la primera característica de transformación corresponden a una fuente u objeto de audio de primer plano y el transformador 120 puede estar adaptado para aplicar una panorámica de amplitud a la primera señal descompuesta. El transformador 120 puede estar adaptado entonces adicionalmente para proporcionar como primera señal transformada dos versiones con panorámica de amplitud de la primera señal descompuesta. En esta realización, la segunda propiedad semántica y la segunda característica de transformación corresponden a una fuente u objeto de audio de segundo plano, una pluralidad de los mismos respectivamente, y el transformador 120 puede estar adaptado para aplicar una decorrelación a la segunda señal descompuesta y proporcionar como segunda señal transformada la segunda señal descompuesta y la versión decorrelacionada de la misma.

30 El transformador 120 puede estar adaptado adicionalmente para transformar la primera señal descompuesta de modo que la primera característica de transformación no tenga una característica de introducción de retardo. Dicho de otro modo, puede no haber decorrelación de la primera señal descompuesta. En otro caso, la primera característica de transformación puede tener una característica de introducción de retardo que tiene una primera cantidad de retardo y la segunda característica de transformación puede tener una segunda cantidad de retardo, siendo la segunda cantidad de retardo mayor que la primera cantidad de retardo. Dicho de otro modo, tanto la primera señal descompuesta como la segunda señal descompuesta pueden estar decorrelacionadas, sin embargo, el nivel de decorrelación puede ajustarse a escala con la cantidad de retardo introducido a las respectivas versiones decorrelacionadas de las señales descompuestas. Por tanto, la decorrelación puede ser más fuerte para la segunda señal descompuesta que para la primera señal descompuesta.

35 En las realizaciones, la primera señal descompuesta y la segunda señal descompuesta pueden solaparse y/o pueden ser sincronas en el tiempo. Dicho de otro modo, el procesamiento de señales puede llevarse a cabo por bloques, pudiendo subdividirse un bloque de muestras de señal de audio de entrada por el descomponedor 110 en varios bloques de señales descompuestas. En las realizaciones, el número de señales descompuestas pueden solaparse al menos parcialmente en el dominio de tiempo, es decir pueden representar muestras de dominio de tiempo solapantes.

40 Dicho de otro modo, las señales descompuestas pueden corresponder a partes de la señal de audio de entrada, que se solapan, es decir que representan al menos parcialmente señales de audio simultáneas. En las realizaciones, las señales descompuestas primera y segunda pueden representar versiones filtradas o transformadas de una señal de entrada original. Por ejemplo, pueden representar partes de señal que se extraen de una señal espacial compuesta

correspondiente por ejemplo a una fuente de sonido cercada o una fuente de sonido más distante. En otras realizaciones, pueden corresponder a componentes de señal transitorias y estacionarias, etc.

5 En las realizaciones, el transformador 120 puede subdividirse en un primer transformador y un segundo transformador, pudiendo estar adaptado el primer transformador para transformar la primera señal descompuesta y pudiendo estar adaptado el segundo transformador para transformar la segunda señal descompuesta. En las realizaciones, el transformador 120 puede implementarse en software, por ejemplo, como programa almacenado en una memoria que debe ejecutarse en un procesador o un procesador de señal digital que, a su vez, está adaptado para transformar las señales descompuestas secuencialmente.

10 El transformador 120 puede estar adaptado para decorrelacionar la primera señal descompuesta para obtener una primera señal decorrelacionada y/o para decorrelacionar la segunda señal descompuesta para obtener una segunda señal decorrelacionada. Dicho de otro modo, el transformador 120 puede estar adaptado para decorrelacionar ambas señales descompuestas, sin embargo, usando diferentes características de decorrelación o transformación. En las realizaciones, el transformador 120 puede estar adaptado para aplicar una panorámica de amplitud a una cualquiera de las señales descompuestas primera o segunda en lugar o además de una decorrelación.

15 El transformador 120 puede estar adaptado para transformar las señales transformadas primera y segunda que tienen cada una tantos componentes como canales en la señal de audio multicanal de salida espacial y el procesador 130 puede estar adaptado para combinar las componentes de las señales transformadas primera y segunda para obtener la señal de audio multicanal de salida espacial. En otras realizaciones, el transformador 120 puede estar adaptado para transformar las señales transformadas primera y segunda que tienen cada una menos componentes que la señal de audio multicanal de salida espacial y pudiendo estar adaptado el procesador 130 para mezclar en sentido ascendente las componentes de las señales transformadas primera y segunda para obtener la señal de audio multicanal de salida espacial.

25 La figura 1b muestra otro aparato 100, que comprende componentes similares a las introducidas con la ayuda de la figura 1a. Sin embargo, la figura 1b muestra un aparato que tiene más detalles. La figura 1b muestra un descomponedor 110 que recibe la señal de audio de entrada y opcionalmente el parámetro de entrada. Como puede observarse a partir de la figura 1b, el descomponedor está adaptado para proporcionar una primera señal descompuesta y una segunda señal descompuesta a un transformador 120, lo que se indica mediante las líneas discontinuas. En el aparato mostrado en la figura 1b, se supone que la primera señal descompuesta corresponde a una fuente de audio de tipo puntual como primera propiedad semántica y que el transformador 120 está adaptado para aplicar una panorámica de amplitud como primera característica de transformación a la primera señal descompuesta. En las realizaciones, las señales descompuestas primera y segunda son intercambiables, es decir en otras realizaciones puede aplicarse una panorámica de amplitud a la segunda señal descompuesta.

35 En el aparato representado en la figura 1b, el transformador 120 muestra, en el trayecto de señal de la primera señal descompuesta, dos amplificadores 121 y 122 ajustables a escala, que están adaptados para amplificar dos copias de la primera señal descompuesta de manera diferente. Los diferentes factores de amplificación usados pueden determinarse, en las realizaciones, a partir del parámetro de entrada, en otras realizaciones, pueden determinarse a partir de la señal de audio de entrada, ya esté preestablecida o se genere localmente, posiblemente también haciendo referencia a una entrada de usuario. Las salidas de los dos amplificadores 121 y 122 ajustables a escala se proporcionan al procesador 130, para el que se proporcionan detalles más adelante.

40 Como puede observarse a partir de la figura 1b, el descomponedor 110 proporciona una segunda señal descompuesta al transformador 120, que lleva a cabo una transformación diferente en el trayecto de procesamiento de la segunda señal descompuesta. En otras realizaciones, la primera señal descompuesta puede procesarse en el trayecto descrito en este momento además o en lugar de la segunda señal descompuesta. Las señales descompuestas primera y segunda pueden intercambiarse en las realizaciones.

45 En el aparato representado en la figura 1b, en el trayecto de procesamiento de la segunda señal descompuesta, hay un decorrelacionador 123 seguido de un módulo 124 de rotador o estéreo paramétrico o de mezcla en sentido ascendente como segunda característica de transformación. El decorrelacionador 123 puede estar adaptado para decorrelacionar la segunda señal descompuesta $X[k]$ y para proporcionar una versión decorrelacionada $Q[k]$ de la segunda señal descompuesta al módulo 124 estéreo paramétrico o de mezcla en sentido ascendente. En la figura 1b, la señal mono $X[k]$ se alimenta a la unidad 123 de decorrelacionador "D" así como el módulo 124 de mezcla en sentido ascendente. La unidad 123 de decorrelacionador puede crear la versión decorrelacionada $Q[k]$ de la señal de entrada, que tiene las mismas características de frecuencia y la misma energía a largo plazo. El módulo 124 de mezcla en sentido ascendente puede calcular una matriz de mezcla ascendente basándose en los parámetros espaciales y sintetizar los canales de salida $Y_1[k]$ y $Y_2[k]$. El módulo de mezcla en sentido ascendente puede explicarse según

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

5 siendo los parámetros c_l , c_r , α y β constantes, o valores variables en tiempo y en frecuencia estimados a partir de la señal de entrada $X[k]$ de manera adaptativa, o transmitidos como información complementaria junto con la señal de entrada $X[k]$ en forma de por ejemplo parámetros de ILD (ILD = *Inter Channel Level Difference*, diferencia de nivel intercanal) y parámetros de ICC (ICC = *Inter Channel Correlation*). La señal $X[k]$ es la señal mono recibida, la señal $Q[k]$ es la señal decorrelacionada, siendo una versión decorrelacionada de la señal de entrada $X[k]$. Las señales de salida se designan con $Y_1[k]$ y $Y_2[k]$.

10 El decorrelacionador 123 puede implementarse como un filtro de IIR (IIR = *Infinite Impulse Response*, respuesta infinita al impulso), un filtro de FIR arbitrario (FIR = *Finite Impulse Response*, respuesta finita al impulso) o un filtro de FIR especial que usa una única toma para retardar de manera simple la señal.

Los parámetros c_l , c_r , α y β pueden determinarse de diferentes maneras. En algunas realizaciones, se determinan de manera simple mediante parámetros de entrada, que pueden proporcionarse junto con la señal de audio de entrada, por ejemplo, con los datos de mezcla en sentido descendente como información complementaria. En otras realizaciones, pueden generarse localmente o derivarse de propiedades de la señal de audio de entrada.

15 En el aparato mostrado en la figura 1b, el transformador 120 está adaptado para proporcionar la segunda señal transformada en cuanto a las dos señales de salida $Y_1[k]$ y $Y_2[k]$ del módulo 124 de mezcla en sentido ascendente al procesador 130.

20 Según el trayecto de procesamiento de la primera señal descompuesta, las dos versiones con panorámica de amplitud de la primera señal descompuesta, disponibles de las salidas de los dos amplificadores 121 y 122 ajustables a escala, también se proporcionan al procesador 130. En otras realizaciones, los amplificadores 121 y 122 ajustables a escala pueden estar presentes en el procesador 130, pudiendo proporcionarse sólo la primera señal descompuesta y un factor de panorámica por el transformador 120.

25 Como puede observarse en la figura 1b, el procesador 130 puede estar adaptado para procesar o combinar la primera señal transformada y la segunda señal transformada, en esta realización combinando simplemente las salidas con el fin de proporcionar una señal estéreo que tiene un canal izquierdo L y un canal derecho R que corresponde a la señal de audio multicanal de salida espacial de la figura 1a.

30 En el aparato en la figura 1b, en ambos trayectos de señalización, se determinan los canales izquierdo y derecho para una señal estéreo. En el trayecto de la primera señal descompuesta, se lleva a cabo una panorámica de amplitud mediante los dos amplificadores 121 y 122 ajustables a escala, por tanto, las dos componentes dan como resultado dos señales de audio en fase, que se ajustan a escala de manera diferente. Esto corresponde a una impresión de una fuente de audio de tipo puntual como propiedad semántica o característica de transformación.

35 En el trayecto de procesamiento de señales de la segunda señal descompuesta, las señales de salida $Y_1[k]$ y $Y_2[k]$ se proporcionan al procesador 130 que corresponde a los canales izquierdo y derecho tal como se determina mediante el módulo 124 de mezcla en sentido ascendente. Los parámetros c_l , c_r , α y β determinan la lejanía espacial de la fuente de audio correspondiente. Dicho de otro modo, los parámetros c_l , c_r , α y β pueden elegirse de un modo o intervalo tal, que para los canales L y R cualquier correlación entre una correlación máxima y una correlación mínima pueda obtenerse en el segundo trayecto de procesamiento de señales como segunda característica de transformación. Además, esto puede llevarse a cabo independientemente para diferentes bandas de frecuencia. Dicho de otro modo, los parámetros c_l , c_r , α y β pueden elegirse de un modo o intervalo tal, que los canales L y R estén en fase, modelizando una fuente de audio de tipo puntual como propiedad semántica.

40 Los parámetros c_l , c_r , α y β también pueden elegirse de un modo o intervalo tal, que los canales L y R en el segundo trayecto de procesamiento de señales estén decorrelacionados, modelizando una fuente de audio espacialmente bastante distribuida como propiedad semántica, por ejemplo modelizando una fuente de sonido de segundo plano o espacialmente más lejana.

45 La figura 2 ilustra otro aparato, que es más general. La figura 2 muestra un bloque 210 de descomposición semántica, que corresponde al descomponedor 110. La salida de la descomposición 210 semántica es la entrada de una fase 220 de transformación, que corresponde al transformador 120. La fase 220 de transformación se compone de varios transformadores 221 a 22n individual, es decir la fase 210 de descomposición semántica está adaptada para descomponer una señal de entrada mono/estéreo en n señales descompuestas, que tienen n propiedades semánticas. La descomposición puede llevarse a cabo basándose en parámetros que controlan la descomposición, que pueden proporcionarse junto con la señal de entrada mono/estéreo, ya esté preestablecida, se genere localmente o se introduzca por un usuario, etc.

Dicho de otro modo, el descomponedor 110 puede estar adaptado para descomponer la señal de audio de entrada semánticamente basándose en el parámetro de entrada opcional y/o para determinar el parámetro de entrada a partir de la señal de audio de entrada.

5 La salida de la fase 220 de decorrelación o de transformación se proporciona entonces a un bloque 230 de mezcla en sentido ascendente, que determina una salida multicanal basándose en las señales decorrelacionadas o transformadas y opcionalmente basándose en parámetros controlados mediante mezcla en sentido ascendente.

10 Generalmente, las realizaciones pueden separar el material sonoro en n componentes semánticas diferentes y decorrelacionar cada componente por separado con un decorrelacionador asociado, que también se indican con D^1 a D^n en la figura 2. Dicho de otro modo, en las realizaciones las características de transformación pueden hacerse coincidir con las propiedades semánticas de las señales descompuestas. Cada uno de los decorrelacionadores o transformadores puede estar adaptado para las propiedades semánticas de la componente de señal descompuesta de manera correspondiente. Posteriormente, las componentes procesadas pueden mezclarse para obtener la señal multicanal de salida. Las diferentes componentes pueden corresponder, por ejemplo, a objetos de modelizado de primer plano y de segundo plano.

15 Dicho de otro modo, el transformador 110 puede estar adaptado para combinar la primera señal descompuesta y la primera señal decorrelacionada para obtener una señal de mezcla en sentido ascendente estéreo o multicanal como primera señal transformada y/o para combinar la segunda señal descompuesta y la segunda señal decorrelacionada para obtener una señal de mezcla en sentido ascendente estéreo como segunda señal transformada.

20 Además, el transformador 120 puede estar adaptado para transformar la primera señal descompuesta según una característica de audio de segundo plano y/o para transformar la segunda señal descompuesta según una característica de audio de primer plano o viceversa.

25 Puesto que, por ejemplo, señales similares a aplausos pueden observarse como compuestas de palmadas cercanas individuales distintas y un ambiente similar a ruido originado a partir de palmadas lejanas muy densas, una descomposición adecuada de tales señales puede obtenerse distinguiendo entre eventos de palmada de primer plano aislados como una componente y un segundo plano similar a ruido como la otra componente. Dicho de otro modo, en una realización, $n=2$. En una realización de este tipo, por ejemplo, el transformador 120 puede estar adaptado para transformar la primera señal descompuesta mediante una panorámica de amplitud de la primera señal descompuesta. Dicho de otro modo, la correlación o transformación de la componente de palmada de primer plano puede conseguirse, en las realizaciones, en D^1 mediante una panorámica de amplitud de cada evento individual hasta su ubicación original estimada.

30 En las realizaciones, el transformador 120 puede estar adaptado para transformar la primera y/o segundo señal descompuesta, por ejemplo, sometiendo a filtrado todo paso la primera o segunda señal descompuesta para obtener la primera o segunda señal decorrelacionada.

35 Dicho de otro modo, en las realizaciones, los antecedentes puede decorrelacionarse o transformarse mediante el uso de m filtros $D^2_{1...m}$ todo paso independientes entre sí. En las realizaciones, sólo el segundo plano cuasi estacionario puede procesarse mediante los filtros todo paso, pudiendo evitarse de esta manera los efectos de difuminado temporal de los métodos de decorrelación del estado de la técnica. Dado que la panorámica de amplitud puede aplicarse a los eventos del objeto de primer plano, la densidad de aplausos de primera plano original puede restaurarse aproximadamente a diferencia del sistema del estado de la técnica tal como se presenta, por ejemplo, en el artículo corto J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates" en 116ª Convención de AES, Berlín, preimpresión 6072, mayo de 2004 y J. Herre, K. Kjörling, J. Breebaart, *et al.*, "MPEG Surround - the ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding" en Procedimientos de la 122ª Convención de AES, Viena, Austria, mayo de 2007.

45 Dicho de otro modo, en las realizaciones, el descomponedor 110 puede estar adaptado para descomponer la señal de audio de entrada semánticamente basándose en el parámetro de entrada, pudiendo proporcionarse el parámetro de entrada junto con la señal de audio de entrada como, por ejemplo, información complementaria. En una realización de este tipo, el descomponedor 110 puede estar adaptado para determinar el parámetro de entrada a partir de la señal de audio de entrada. En otras realizaciones, el descomponedor 110 puede estar adaptado para determinar el parámetro de entrada como parámetro de control independiente de la señal de audio de entrada, que puede generarse localmente, estar preestablecida o también puede introducirse por un usuario.

50 En las realizaciones, el transformador 120 puede estar adaptado para obtener una distribución espacial de la primera señal transformada o la segunda señal transformada aplicando una panorámica de amplitud de banda. Dicho de otro modo, según la descripción de la figura 1b anterior, en lugar de generar una fuente de tipo puntual, la ubicación de panorámica de la fuente puede variarse temporalmente con el fin de generar una fuente de audio que tenga una determinada distribución espacial. En las realizaciones, el transformador 120 puede estar adaptado para aplicar el ruido de paso bajo generado localmente para una panorámica de amplitud, es decir los factores de ajuste a escala para la panorámica de amplitud para, por ejemplo, los amplificadores 121 y 122 ajustables a escala en la figura 1b

corresponden a un valor de ruido generado localmente, es decir son variables en el tiempo con un determinado ancho de banda.

Las realizaciones pueden estar adaptadas para hacerse funcionar en un modo guiado o uno no guiado. Por ejemplo, en un escenario guiado, haciendo referencia a las líneas discontinuas, por ejemplo en la figura 2, la decorrelación puede llevarse a cabo aplicando filtros de decorrelación de tecnología convencional controlados en una cuadrícula de tiempo gruesa a , por ejemplo, sólo la parte de segundo plano o ambiente y obtener la correlación mediante la redistribución de cada evento individual en, por ejemplo, la parte de primer plano por medio de posicionamiento espacial variable en el tiempo usando una panorámica de amplitud de banda ancha en una cuadrícula de tiempo mucho más fina. Dicho de otro modo, en las realizaciones, el transformador 120 puede estar adaptado para hacer funcionar los decorrelacionadores para diferentes señales descompuestas en diferentes cuadrículas de tiempo, por ejemplo basándose en diferentes escalas de tiempo, lo que puede ser en cuanto a diferentes tasas de muestreo o un retardo diferente para los respectivos decorrelacionadores. En una realización, a la hora de llevar a cabo la separación de primera plano y de segundo plano, la parte de primer plano puede usar una panorámica de amplitud, mientras que la amplitud se cambia en una cuadrícula de tiempo mucho más fina que el funcionamiento para un decorrelacionador con respecto a la parte de segundo plano.

Además, se enfatiza que para la decorrelación de, por ejemplo, señales similares a aplausos, es decir señales con una calidad aleatoria cuasi estacionaria, la posición espacial exacta de cada palmada de primer plano individual puede no ser de importancia tan crucial, como la recuperación de la distribución global de la multitud de eventos de palmada. Las realizaciones pueden aprovechar este hecho y pueden funcionar en un modo no guiado. En un modo de este tipo, el factor de panorámica de amplitud mencionado anteriormente puede controlarse mediante el ruido de paso bajo según la invención. La figura 3 ilustra un sistema de mono a estéreo que implementa el escenario. La figura 3 muestra un bloque 310 de descomposición semántica que corresponde al descomponedor 110 para descomponer la señal de entrada mono en una parte de señal descompuesta de primera plano y de segundo plano.

Como puede observarse a partir de la figura 3, la parte descompuesta de segundo plano de la señal se transforma mediante el filtro 320 todo paso D^1 . La señal decorrelacionada se proporciona entonces junto con la parte descompuesta de segundo plano no transformada a la mezcla 330 en sentido ascendente, que corresponde al procesador 130. La parte de señal descompuesta de primer plano se proporciona a una fase 340 de panorámica de amplitud D^2 , que corresponde al transformador 120. El ruido 250 de paso bajo generado localmente también se proporciona a la fase 340 de panorámica de amplitud, que puede proporcionar entonces la señal descompuesta de primer plano en una configuración con panorámica de amplitud a la mezcla 330 en sentido ascendente. La fase 340 de panorámica de amplitud D^2 puede determinar su salida proporcionando un factor de ajuste a escala k para una selección de amplitud entre dos de un conjunto estéreo de canales de audio. El factor de ajuste a escala k puede basarse en el ruido de paso bajo.

Como puede observarse a partir de la figura 3, hay una flecha entre la panorámica 340 de amplitud y la mezcla 330 en sentido ascendente. Esta flecha también puede representar señales con panorámica de amplitud, es decir en el caso de la mezcla en sentido ascendente estéreo, ya el canal izquierdo y el derecho. Como puede observarse a partir de la figura 3, la mezcla 330 en sentido ascendente que corresponde al procesador 130 está adaptada entonces para procesar o combinar las señales descompuestas de segundo plano y de primer plano para derivar la salida estéreo.

Otras realizaciones pueden usar procesamiento nativo con el fin de derivar señales descompuestas de segundo plano y de primer plano o parámetros de entrada para la descomposición. El descomponedor 110 puede estar adaptado para determinar la primera señal descompuesta y/o la segunda señal descompuesta basándose en un método de separación transitoria. Dicho de otro modo, el descomponedor 110 puede estar adaptado para determinar la primera o segunda señal descompuesta basándose en un método de separación y la otra señal descompuesta basándose en 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 determinarse basándose en el método de separación transitoria y la otra señal descompuesta puede basarse en la diferencia entre la primera o segunda señal descompuesta y la señal de audio de entrada.

El descomponedor 110 y/o el transformador 120 y/o el procesador 130 pueden comprender una fase de monosíntesis de DirAC y/o una fase de síntesis de DirAC y/o una fase de fusión de DirAC. En las realizaciones, el descomponedor 110 puede estar adaptado para descomponer la señal de audio de entrada, el transformador 120 puede estar adaptado para transformar las señales descompuestas primera y/o segunda, y/o el procesador 130 puede estar adaptado para procesar las señales transformadas primera y/o segunda en cuanto a diferentes bandas de frecuencia.

Las realizaciones pueden usar la siguiente aproximación para señales similares a aplausos. Aunque las componentes de primer plano pueden obtenerse mediante métodos de separación o detección transitoria, véase Pulkki, Ville; "Spatial Sound Reproducción with Directional Audio Coding" en J. Audio Eng. Soc., vol. 55, n.º 6, 2007, la componente de segundo plano puede venir dada por la señal residual. La figura 4 representa un ejemplo de un método adecuado para obtener una componente de segundo plano $x'(n)$ de, por ejemplo, una señal similar a aplausos $x(n)$ para implementar la descomposición 310 semántica en la figura 3, es decir una realización del descomponedor 120. La figura 4 muestra una señal de entrada diferenciada en el tiempo $x(n)$, que es una entrada para una DFT 410 (DFT = *Discrete Fourier*

Transform, transformada discreta de Fourier). La salida del bloque 410 de DFT se proporciona a un bloque para suavizar el espectro 420 y a un bloque 430 de blanqueamiento espectral para el blanqueamiento espectral basándose en la salida de la DFT 410 y la salida de la fase 430 de espectro suave.

5 La salida de la fase 430 de blanqueamiento espectral se proporciona entonces a una fase 440 de detección de picos espectrales, que separa el espectro y proporciona dos salidas, es decir una señal residual transitoria y de ruido y una señal tonal. La señal residual transitoria y de ruido se proporciona a un filtro 450 de LPC (LPC = *Linear Prediction Coding*, codificación predictiva lineal) de la que se proporciona la señal de ruido residual a la fase 460 de mezclado junto con la señal tonal como salida de la fase 440 de detección de picos espectrales. La salida de la fase 460 de mezclado se proporciona entonces a una fase 470 de conformación espectral, que conforma el espectro basándose en el espectro suavizado proporcionado por la fase 420 de espectro suavizado. La salida de la fase 470 de conformación espectral se proporciona entonces al filtro 480 de síntesis, es decir una inversa de la transformada discreta de Fourier con el fin de obtener $x'(n)$ que representa la componente de segundo plano. La componente de primer plano puede derivarse entonces como la diferencia entre la señal de entrada y la señal de salida, es decir como $x(n)-x'(n)$.

15 Las realizaciones de la presente invención pueden hacerse funcionar en aplicaciones de realidad virtual tales como, por ejemplo, juegos en 3D. En tales aplicaciones, la síntesis de fuentes de sonido con una gran extensión espacial puede ser complicada y compleja cuando se basa en conceptos convencionales. Tales recursos pueden, por ejemplo, ser una orilla del mar, una bandada de pájaros, caballos galopando, la división de soldados marchando o un público que aplaude. Normalmente, tales eventos de sonido están espaciados como un grupo grande de fuentes de tipo puntual, lo que conduce a implementaciones computacionalmente complejas, véase Wagner, Andreas; Walther, Andreas; Melchoir, Frank; Strauß, Michael; "Generation of Highly Immersive Atmospheres for Wave Field Synthesis Reproduction" en la 116ª Convención internacional de EAS, Berlín, 2004.

25 Las realizaciones pueden llevar a cabo un método, que realiza la síntesis de la extensión de fuentes de sonido de manera plausible pero que tiene, al mismo tiempo, una complejidad estructural y computacional inferior. Las realizaciones pueden basarse en DirAC (DirAC = *Directional Audio Coding*), véase Pulkki, Ville; "Spatial Sound Reproduction with Directional Audio Coding" en *J. Audio Eng. Soc.*, vol. 55, n.º 6, 2007. Dicho de otro modo, en las realizaciones, el descomponedor 110 y/o el transformador 120 y/o el procesador 130 pueden estar adaptados para procesar señales de DirAC. Dicho de otro modo, el descomponedor 110 puede comprender fases de monosíntesis de DirAC, el transformador 120 puede comprender una fase de síntesis de DirAC y/o el procesador puede comprender una fase de fusión de DirAC.

30 Las realizaciones pueden basarse en procesamiento de DirAC, por ejemplo, usando sólo dos estructuras de síntesis, por ejemplo, una para fuentes de sonido de primer plano y una para fuentes de sonido de segundo plano. El sonido de primer plano puede aplicarse a un flujo de DirAC individual con datos direccionales controlados, dando como resultado la percepción de fuentes de tipo puntual cercanas. El sonido de segundo plano también puede reproducirse usando un único flujo directo con datos direccionales controlados de manera diferente, lo que conduce a la percepción de objetos de sonido diseminados espacialmente. Los dos flujos de DirAC pueden entonces fusionarse y decodificarse para una configuración de altavoces arbitraria o para auriculares, por ejemplo.

35 La figura 5 ilustra una síntesis de fuentes de sonido que tiene una espacialmente grande. La figura 5 muestra un bloque 610 de monosíntesis superior, que crea un flujo de mono-DirAC que conduce a una percepción de una fuente de sonido de tipo puntual cercana, tal como las palmadas más cercanas de un público. El bloque 620 de monosíntesis inferior se usa para crear un flujo de mono-DirAC que conduce a la percepción de sonido diseminado espacialmente, lo que es adecuado, por ejemplo, para generar sonido de segundo plano como el sonido de palmadas del público. Las salidas de los dos bloques 610 y 620 de monosíntesis de DirAC se fusionan entonces en la fase 630 de fusión de DirAC. La figura 5 muestra que sólo se usan dos bloques 610 y 620 de síntesis de DirAC. Uno de ellos se usa para crear los eventos de sonidos que están en el primer plano, tal como los pájaros cercanos o más cercanos o las personas cercanas o más cercanas en un público que aplaude y el otro genera un sonido de segundo plano, el sonido continuo de la bandada de pájaros, etc.

40 El sonido de primer plano se convierte en un flujo de mono-DirAC con el bloque 610 de monosíntesis de DirAC de una manera tal que los datos acimutales se mantienen constantes con la frecuencia, sin embargo, cambian aleatoriamente o se controlan mediante un proceso externo en el tiempo. El parámetro de difusividad Ψ se establece a 0, es decir representando una fuente de tipo puntual. Se supone que la entrada de audio al bloque 610 son sonidos temporalmente no solapantes, tales como distintos reclamos de pájaros o palmadas con las manos, que generan la percepción de fuentes de sonido cercanas, tales como pájaros o personas que aplauden. La extensión espacial de los eventos de sonido de primer plano se controla ajustando θ y $\theta_{\text{intervalo_primer plano}}$, lo que significa que los eventos de sonido individuales se percibirán en la dirección $\theta + \theta_{\text{intervalo_primer plano}}$, sin embargo, un único evento puede percibirse como puntual. Dicho de otro modo, se generan fuentes de sonido de tipo puntual en las que las posibles posiciones del punto están limitadas al intervalo $\theta \pm \theta_{\text{intervalo_primer plano}}$.

45 El bloque 620 de segundo plano adopta como flujo de audio de entrada una señal, que contiene todos los demás eventos de sonido no presentes en el flujo de audio de primer plano, lo que pretende incluir muchos eventos de sonido temporalmente solapantes, por ejemplo cientos de pájaros o un gran número de personas lejanas que aplauden. Los

5 valores acimutales asociados se fijan entonces aleatoriamente tanto en tiempo como en frecuencia, dentro de los valores acimutales restringidos facilitados $\theta \pm \theta_{\text{intervalo_primer plano}}$. Por tanto, la extensión espacial de los sonidos de segundo plano puede sintetizarse con una complejidad computacional baja. También puede controlarse la difusividad Ψ . Si se añadió, el decodificador de DirAC aplicaría el sonido a todas las direcciones, lo que puede usarse cuando la fuente de sonido envuelve totalmente al oyente. Si no lo envuelve, la difusividad puede mantenerse baja o próxima a cero, o cero en las realizaciones.

Las realizaciones de la presente invención pueden proporcionar la ventaja de que puede conseguirse una calidad de percepción superior de sonidos transformados a un coste computacional moderado. Las realizaciones pueden permitir una implementación modular de transformación de sonido espacial tal como se muestra, por ejemplo, en la figura 5.

10 Dependiendo de determinados requisitos de implementación de los métodos inventivos, los métodos inventivos pueden implementarse en hardware o en software. La implementación can puede realizarse usando un medio de almacenamiento digital y, particularmente, una memoria flash, un disco, un DVD o un CD que tiene señales de control legibles electrónicamente almacenadas en el mismo, que coopera con el sistema informático programable, de modo que se realicen los métodos inventivos. Generalmente, la presente invención es, por tanto, un producto de programa informático con un código de programa almacenado en una portadora legible por máquina, siendo el código de programa operativo para realizar los métodos inventivos cuando se ejecuta en un ordenador el producto de programa informático. Dicho de otro modo, los métodos inventivos son, por tanto, un programa informático que tiene un código de programa para realizar al menos uno de los métodos inventivos cuando se ejecuta el programa informático en un ordenador.

20

REIVINDICACIONES

1. Aparato (100) para determinar una señal de audio multicanal de salida espacial basándose en una señal de audio de entrada, que comprende:
 - 5 un descomponedor (110) semántico configurado para descomponer la señal de audio de entrada para obtener una primera señal descompuesta que tiene una primera propiedad semántica, siendo la primera señal descompuesta una parte de señal de primer plano, y una segunda señal descompuesta que tiene una segunda propiedad semántica que es diferente de la primera propiedad semántica, siendo la segunda señal descompuesta una parte de señal de segundo plano;
 - un transformador (120) configurado
 - 10 para transformar la parte de señal de primer plano usando la panorámica de amplitud como primera característica de transformación para obtener una primera señal transformada que tiene la primera propiedad semántica, comprendiendo el transformador (120) una fase (221, 340) de panorámica de amplitud para procesar la parte de señal de primer plano, en la que se proporciona ruido (350) de paso bajo generado localmente a la fase (340) de panorámica de amplitud para variar temporalmente una ubicación de panorámica de una fuente de audio en la parte de señal de primer plano; y
 - 15 para transformar la parte de señal de segundo plano decorrelacionando la segunda señal descompuesta como segunda característica de transformación para obtener una segunda señal transformada que tiene la segunda propiedad semántica; y
 - 20 un procesador (130, 330) configurado para procesar la primera señal transformada y la segunda señal transformada para obtener la señal de audio multicanal de salida espacial.
2. Aparato (100) según la reivindicación 1, en el que la primera característica de transformación se basa en la primera propiedad semántica y la segunda característica de transformación se basa en la segunda propiedad semántica.
3. Aparato (100) según la reivindicación 1 ó 2, en el que el transformador (120) está adaptado para transformar las señales transformadas primera y segunda que tienen cada una tantos componentes como canales en la señal de audio multicanal de salida espacial y el procesador (130) está adaptado para combinar las componentes de las señales transformadas primera y segunda para obtener la señal de audio multicanal de salida espacial.
4. Aparato (100) según la reivindicación 1 ó 2, en el que el transformador (120) está adaptado para transformar las señales transformadas primera y segunda que tienen cada una menos componentes que la señal de audio multicanal de salida espacial y en el que el procesador (130) está adaptado para mezclar en sentido ascendente las componentes de las señales transformadas primera y segunda para obtener la señal de audio multicanal de salida espacial.
5. Aparato (100) según la reivindicación 1, en el que el descomponedor (110) está adaptado para determinar un parámetro de entrada como parámetro de control a partir de la señal de audio de entrada.
6. Aparato (100) según una de las reivindicaciones 1 a 5, en el que el transformador (120) está adaptado para transformar la primera señal descompuesta y la segunda señal descompuesta basándose en diferentes cuadrículas de tiempo.
7. Aparato (100) según una de las reivindicaciones 1 a 8, en el que el descomponedor (110) está adaptado para determinar la primera señal descompuesta y/o la segunda señal descompuesta basándose en un método de separación transitoria.
8. Aparato (100) según la reivindicación 7, en el que el descomponedor (110) está adaptado para determinar una de la primera señal descompuesta o la segunda señal descompuesta mediante un método de separación transitoria y la otra basándose en la diferencia entre ésta y la señal de audio de entrada.
9. Aparato (100) según una de las reivindicaciones 1 a 8, en el que el descomponedor (110) está adaptado para descomponer la señal de audio de entrada, el transformador (120) está adaptado para transformar las señales descompuestas primera y/o segunda, y/o el procesador (130) está adaptado para procesar las señales transformadas primera y/o segunda en cuanto a diferentes bandas de frecuencia.
10. Aparato según la reivindicación 1, en el que el procesador está configurado para procesar la primera señal transformada, la segunda señal transformada y la parte de señal de segundo plano para obtener la señal de audio multicanal de salida espacial.

11. Método para determinar una señal de audio multicanal de salida espacial basándose en una señal de audio de entrada y un parámetro de entrada que comprende las etapas de:

5 descomponer semánticamente la señal de audio de entrada para obtener una primera señal descompuesta que tiene una primera propiedad semántica, siendo la primera señal descompuesta una parte de señal de primer plano, y una segunda señal descompuesta que tiene una segunda propiedad semántica que es diferente de la primera propiedad semántica, siendo la segunda señal descompuesta una parte de señal de segundo plano;

10 transformar la parte de señal de primer plano usando la panorámica de amplitud para obtener una primera señal transformada que tiene la primera propiedad semántica, procesando la parte de señal de primer plano en una fase (221, 340) de panorámica de amplitud, en la que se proporciona ruido (350) de paso bajo generado localmente a la fase (340) de panorámica de amplitud para variar temporalmente una ubicación de panorámica de una fuente de audio en la parte de señal de primer plano;

transformar la parte de señal de segundo plano decorrelacionando la segunda señal descompuesta para obtener una segunda señal transformada que tiene la segunda propiedad semántica; y

15 procesar la primera señal transformada y la segunda señal transformada para obtener la señal de audio multicanal de salida espacial.

12. Programa informático que tiene un código de programa para realizar el método según la reivindicación 11, cuando el código de programa se ejecuta en un ordenador o un procesador.

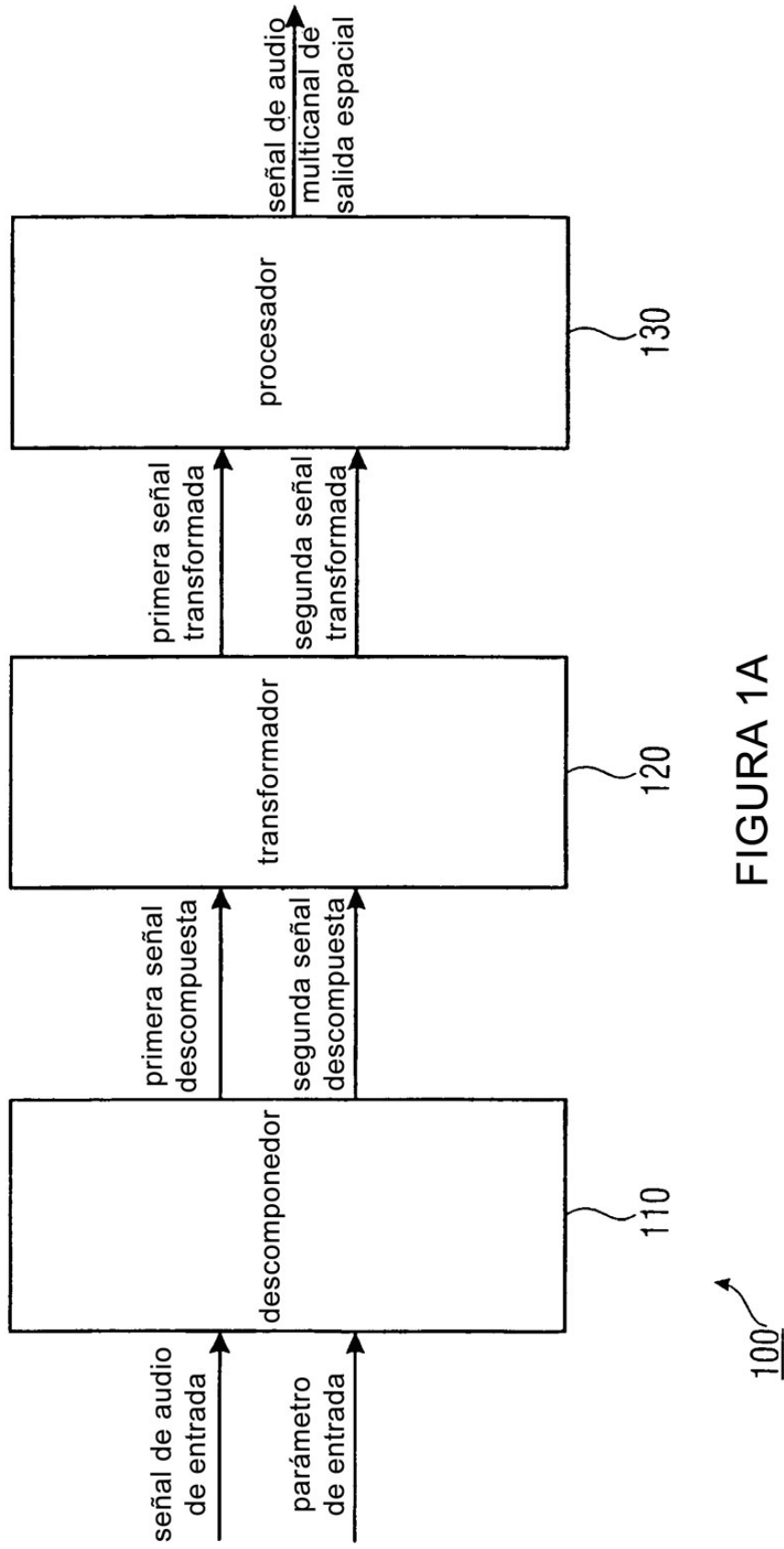


FIGURA 1A

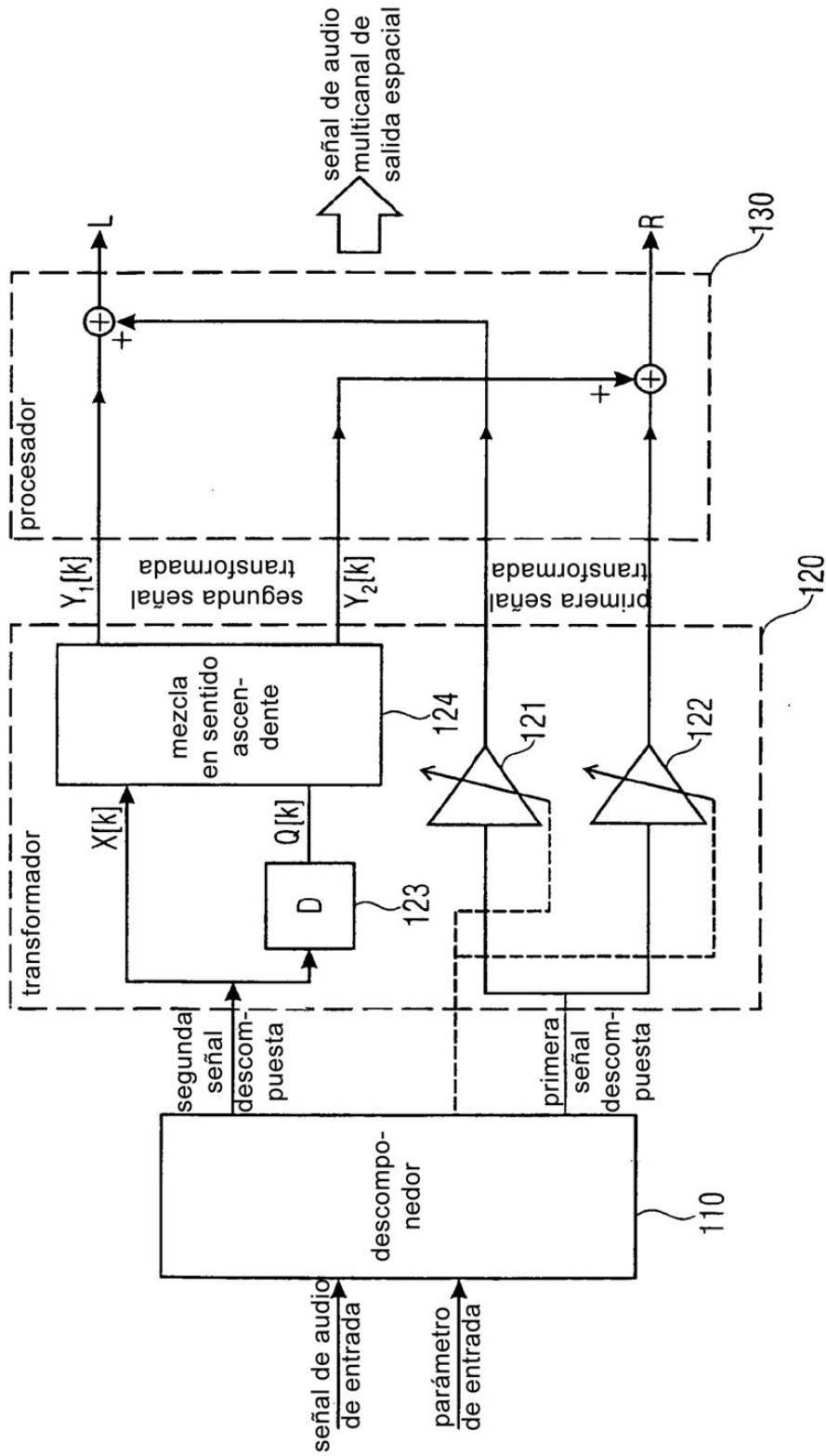


FIGURA 1B

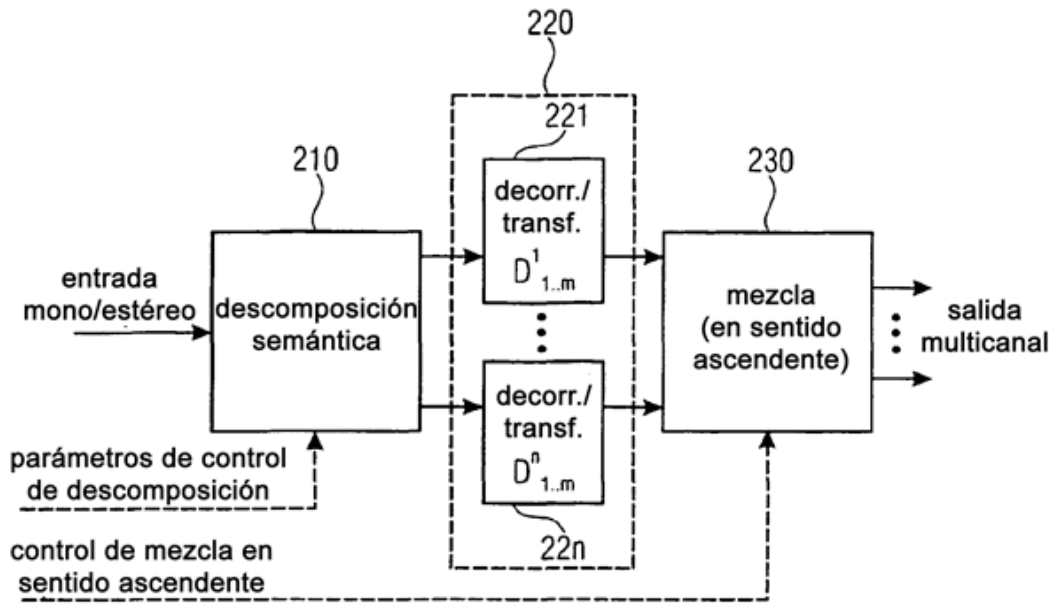


FIGURA 2

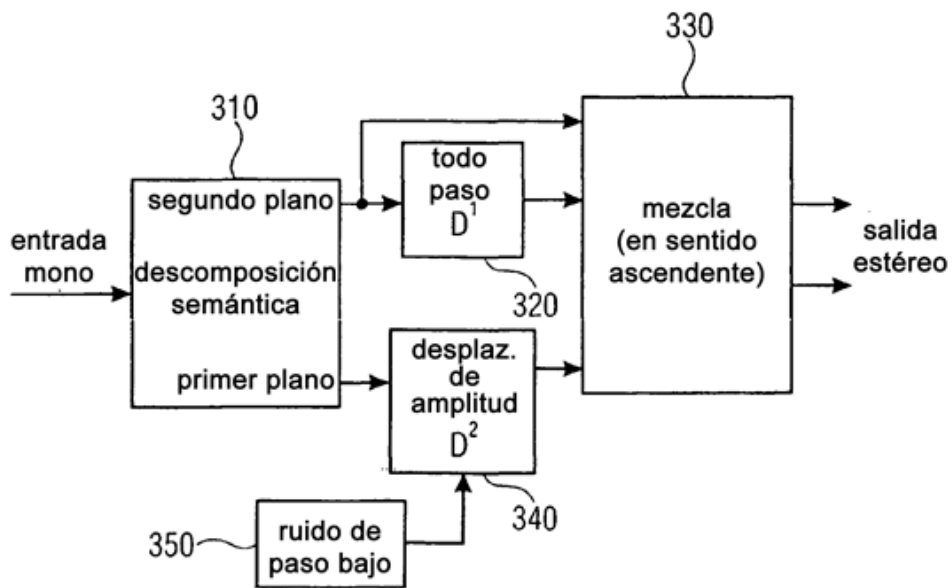


FIGURA 3

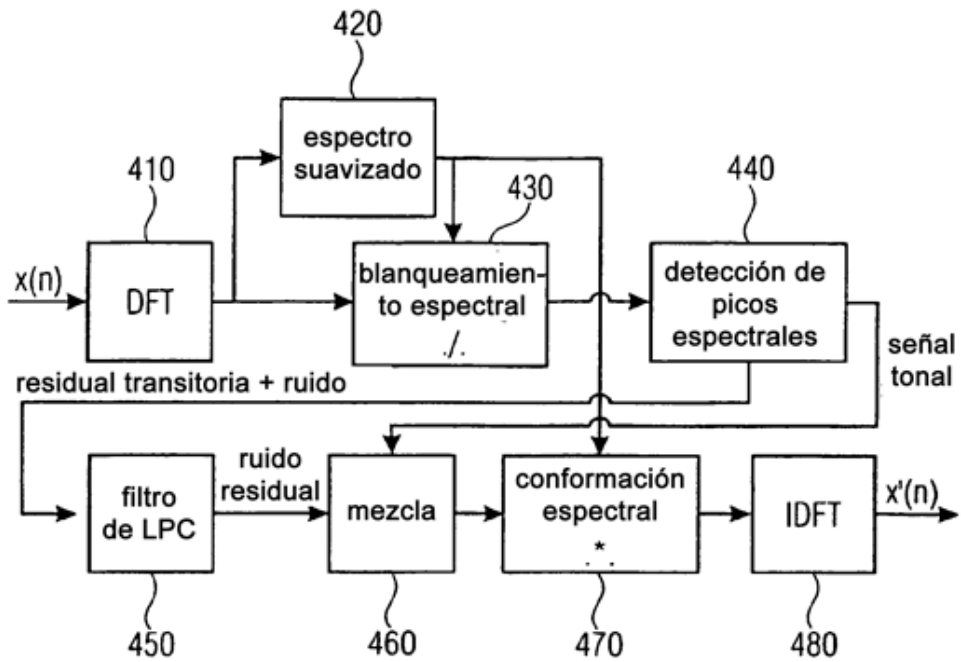


FIGURA 4

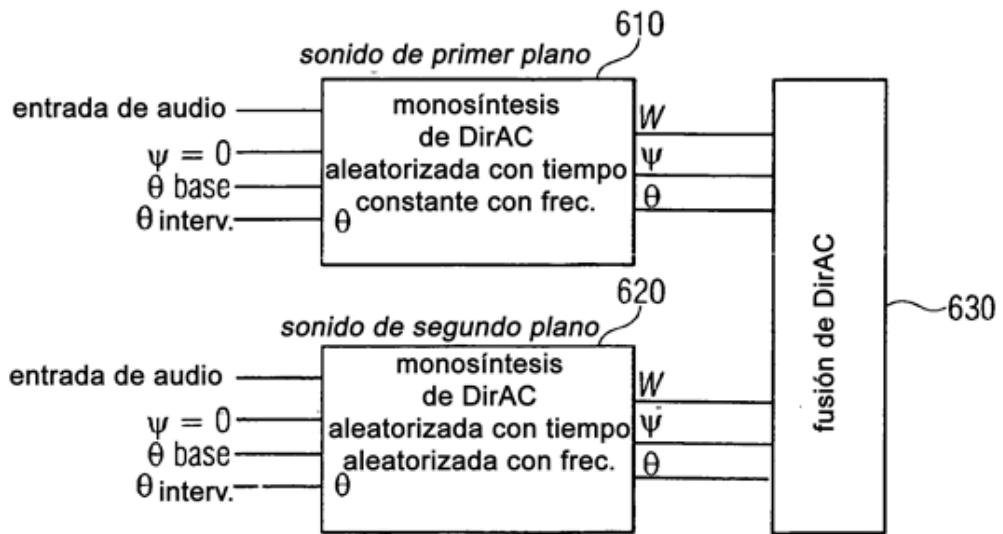


FIGURA 5

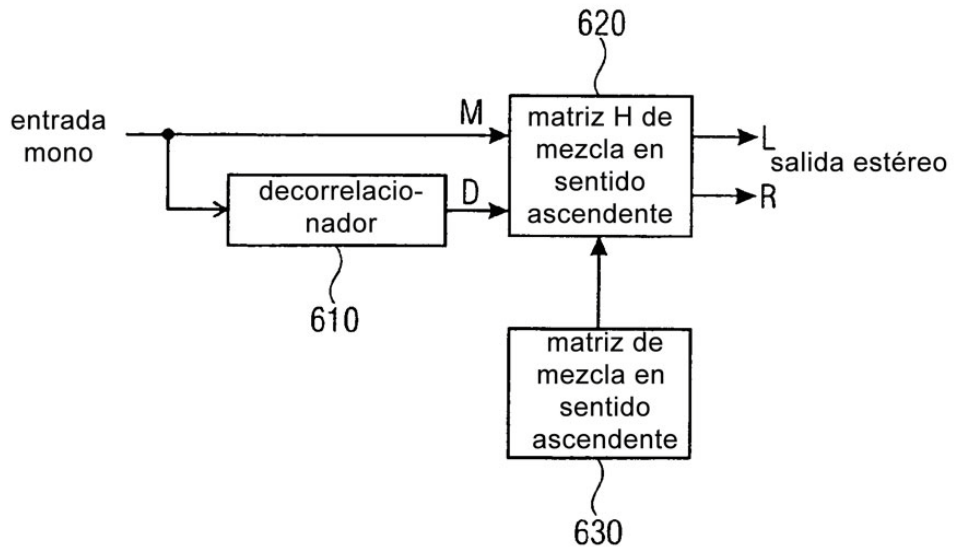


FIGURA 6
(ESTADO DE LA TÉCNICA)

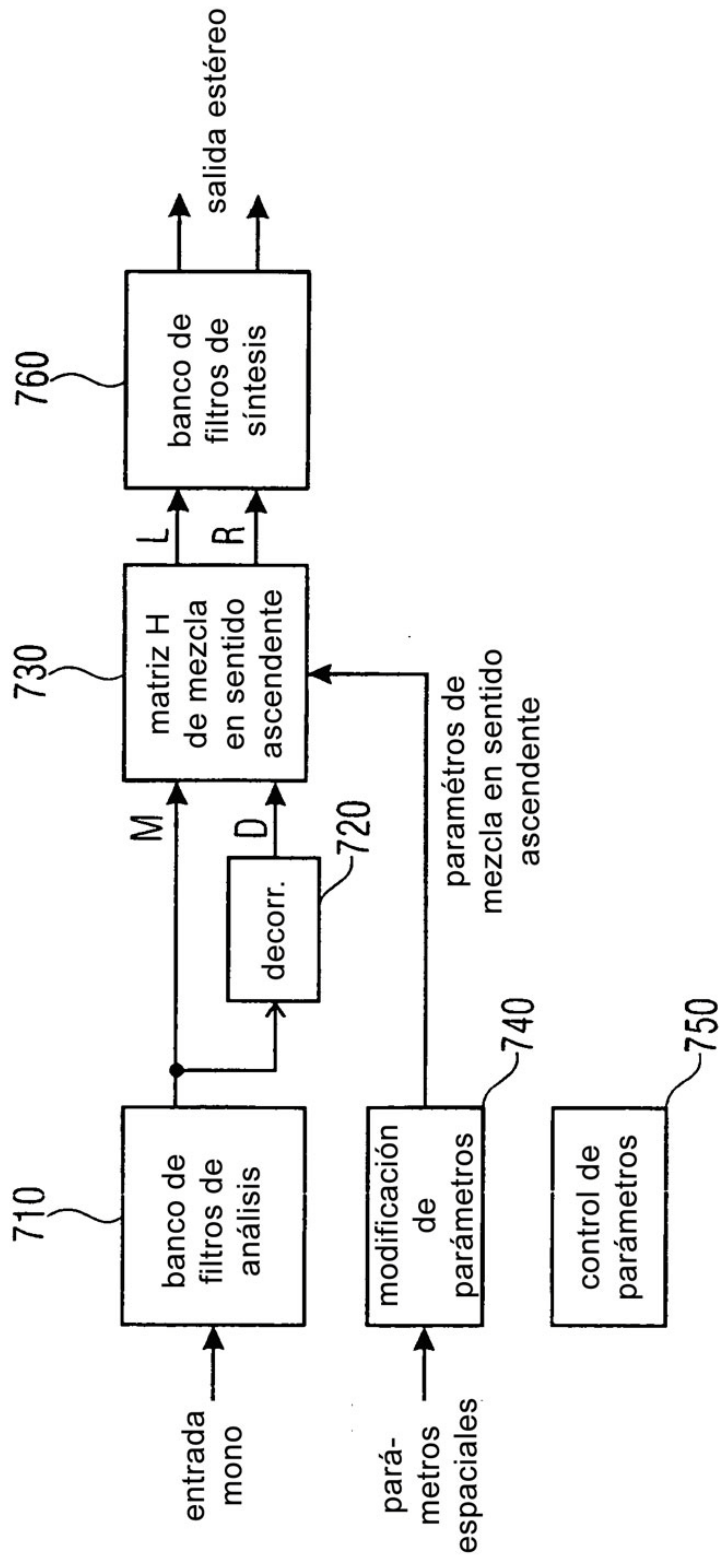


FIGURA 7
(ESTADO DE LA TÉCNICA)