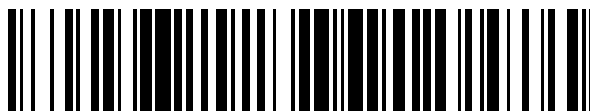


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 461 601**

51 Int. Cl.:

**H04S 3/02** (2006.01)

**G10L 19/00** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.09.2008 E 08802724 (8)**

97 Fecha y número de publicación de la concesión europea: **19.03.2014 EP 2198632**

54 Título: **Procedimiento y aparato para generar una señal de audio binaural**

30 Prioridad:

**09.10.2007 EP 07118107**

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

**20.05.2014**

73 Titular/es:

**KONINKLIJKE PHILIPS N.V. (50.0%)  
High Tech Campus 5  
5656 AE Eindhoven, NL y  
DOLBY INTERNATIONAL AB (50.0%)**

72 Inventor/es:

**BREEBART, DIRK JEROEN y  
VILLEMOES, LARS FALCK**

74 Agente/Representante:

**PONTI SALES, Adelaida**

**ES 2 461 601 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**

Procedimiento y aparato para generar una señal de audio binaural

## 5 CAMPO DE LA INVENCION

**[0001]** La invención se refiere a un procedimiento y aparato para generar una señal de audio binaural y en particular, pero no exclusivamente, a la generación de una señal de audio binaural a partir de una señal mono de mezcla reductora.

## 10 ANTECEDENTES DE LA INVENCION

**[0002]** En la última década, ha existido una tendencia hacia el audio multicanal y específicamente hacia el audio espacial, que se extiende más allá de las señales estéreo convencionales. Por ejemplo, las grabaciones tradicionales en estéreo comprenden solamente dos canales, mientras que los modernos sistemas de audio avanzados utilizan típicamente cinco o seis canales, como en los populares sistemas de sonido envolvente (surround sound) 5.1. Esto proporciona una experiencia auditiva más envolvente en donde el usuario puede encontrarse rodeado por las fuentes de sonido.

**[0003]** Se han desarrollado diversas técnicas y estándares para la comunicación de tales señales multicanal. Por ejemplo, seis canales separados, que representan un sistema envolvente 5.1, pueden transmitirse de acuerdo con estándares tales como los estándares Advanced Audio Coding (Codificación de audio avanzada) (AAC) o Dolby Digital (Digital Dolby).

**[0004]** Sin embargo, a fin de proporcionar una compatibilidad inversa, se conoce el mezclador reductor del más alto número de canales a un número menor, y específicamente, se utiliza frecuentemente el mezclador reductor de una señal de sonido envolvente 5.1 a una señal estéreo permitiendo que una señal estéreo se reproduzca por medio de decodificadores tradicionales (estéreo) y una señal 5.1 por medio de decodificadores de sonido envolvente.

**[0005]** Un ejemplo es el procedimiento de codificación compatible a la inversa MPEG2. Una señal multicanal se reduce por mezcla a una señal estéreo. Las señales adicionales se codifican en la porción de datos auxiliares permitiendo que el decodificador multicanal MPEG2 genere una representación de la señal multicanal. Un decodificador MPEG1 desestimará los datos auxiliares, y por lo tanto, decodificará solamente la mezcla reductora en estéreo.

**[0006]** Existen varios parámetros que pueden utilizarse para describir las propiedades espaciales de las señales de audio. Uno de tales parámetros es la correlación cruzada intercanales, tal como la correlación cruzada entre el canal izquierdo y el canal derecho para señales estéreo. Otro parámetro es la amplificación de los canales. En los llamados codificadores de audio espaciales (paramétricos), estos y otros parámetros se extraen de la señal de audio original a fin de producir un señal de audio que tiene un número de canales reducido, por ejemplo, solamente un canal, más un conjunto de parámetros que describen las propiedades espaciales de la señal de audio original. En los llamados decodificadores de audio espaciales (paramétricos), se reinstalan las propiedades espaciales, como se describen por los parámetros espaciales transmitidos.

**[0007]** Actualmente gana interés el posicionamiento de la fuente de sonido 3D, especialmente en el dominio móvil. La reproducción de música y efectos de sonido en los juegos móviles puede agregar un valor significativo a la experiencia del consumidor al posicionarse en 3D, creando efectivamente un efecto 3D 'delirante'. Específicamente, se conoce la grabación y reproducción de señales de audio binaurales que contienen información direccional específica a la cual el oído humano es sensible. Las grabaciones binaurales se producen típicamente utilizando dos micrófonos instalados en una cabeza humana simulada de manera que el sonido grabado corresponde al sonido capturado por el oído humano e incluyen cualquier influencia debido a la forma de la cabeza y los oídos. Las grabaciones binaurales difieren de las grabaciones en estéreo (es decir, estereofónicas), en que la reproducción de una grabación binaural generalmente se destina a auriculares o audífonos, mientras que una grabación en estéreo se produce generalmente para su reproducción mediante altavoces. Aunque una grabación binaural permite la reproducción de toda la información espacial utilizando solamente dos canales, una grabación en estéreo no proporcionaría la misma percepción espacial.

**[0008]** Las grabaciones normales en canal dual (estereofónicas) o en canales múltiples (por ejemplo, 5.1) pueden transformarse en grabaciones binaurales convolucionando cada señal normal con un conjunto de funciones de transferencia perceptual. Tales funciones de transferencia perceptual modelan la influencia de la cabeza humana, y posiblemente de otros objetos en la señal. Un tipo muy conocido de función de transferencia perceptual espacial es la llamada función de transferencia relacionada con la cabeza (HRTF). Un tipo alternativo de función de transferencia perceptual espacial, que también toma en cuenta las reflexiones ocasionadas por las paredes, el techo y el piso de una habitación, es la respuesta binaural al impulso ambiental (BRIR).

- 5 [0009] Típicamente, los algoritmos de posicionamiento 3D emplean HRTFs (o BRIRs), que describen la transferencia desde una cierta posición de la fuente de sonido hasta los tímpanos por medio de una respuesta al impulso. El posicionamiento de la fuente de sonido 3D puede aplicarse a señales multicanal por medio de HRTFs permitiendo así que una señal binaural proporcione información del sonido espacial a un usuario, por ejemplo, utilizando un par de audífonos.
- 10 [0010] Un algoritmo convencional de síntesis binaural se perfila en la Figura 1. Un conjunto de canales de entrada se filtra mediante un conjunto de HRTFs. Cada señal de entrada se divide en dos señales (un componente izquierdo 'L' y uno derecho 'R'); cada una de estas señales se filtra subsecuentemente mediante un HRTF que corresponde a la posición de la fuente de sonido deseada. Todas las señales del oído izquierdo se suman subsecuentemente para generar la señal de salida binaural izquierda y las señales del oído derecho se suman para generar la señal de salida binaural derecha.
- 15 [0011] Se conocen sistemas codificadores que pueden recibir una señal codificada de sonido envolvente y generar una experiencia de sonido envolvente a partir de una señal binaural. Por ejemplo, se conocen sistemas de audífonos que permiten que la señal de sonido envolvente se convierta en una señal binaural de sonido envolvente proporcionando una experiencia de sonido envolvente al usuario de los audífonos.
- 20 [0012] La Figura 2 ilustra un sistema en donde un decodificador envolvente MPEG recibe una señal en estéreo con datos espaciales paramétricos. La corriente de bits de entrada se des-multiplexa mediante un des-multiplexor (201) dando como resultado parámetros espaciales y una corriente de bits de mezcla reductora. Esta última corriente de bits se decodifica utilizando un decodificador convencional mono o estéreo (203). La mezcla reductora decodificada se decodifica mediante un decodificador espacial (205) que genera una salida multicanal en base a los parámetros espaciales transmitidos. Finalmente, la salida multicanal se procesa entonces por medio de una etapa de síntesis binaural (207) (similar a la de la Figura 1) dando como resultado una señal de salida binaural que proporciona una experiencia de sonido envolvente al usuario.
- 25 [0013] Sin embargo, tal procedimiento es complejo y requiere sustanciales recursos computacionales y puede reducir además la calidad de audio e introducir artefactos audibles.
- 30 [0014] A fin de superar algunas de estas desventajas, se ha propuesto que un decodificador de audio multicanal paramétrico pueda combinarse con un algoritmo de síntesis binaural, de tal manera que una señal multicanal pueda emitirse en audífonos sin que se requiera que la señal multicanal se genere primero desde la señal de mezcla reductora transmitida seguida por la mezcla reductora de la señal multicanal utilizando filtros HRTF.
- 35 [0015] En tales decodificadores, los parámetros espaciales de mezclado ascendente para recrear la señal multicanal se combinan con los filtros HRTF a fin de generar parámetros combinados que pueden aplicarse directamente a la señal de mezcla reductora para generar la señal binaural. A fin de hacerlo así, los filtros HRTF se parametrizan.
- 40 [0016] Un ejemplo de tal decodificador se ilustra en la Figura 3 y se describe adicionalmente en Breebaart J., "Analysis and synthesis of binaural parameters for efficient 3D audio rendering in MPEG Surround" (Análisis y síntesis de parámetros binaurales para la emisión eficiente de audio 3D en sonido envolvente MPEG), Proc. ICME, Beijing, China (2007) y Breebaart J., Faller C., "Spatial audio processing: MPEG Surround and other applications" (Procesamiento de audio espacial: sonido envolvente MPEG y otras aplicaciones), Wiley & Sons, New York (2007).
- 45 [0017] Una corriente de bits de entrada que contiene parámetros espaciales y una señal de mezcla reductora se reciben por un des-multiplexor 301. La señal de mezcla reductora se decodifica mediante un decodificador convencional 303 dando como resultado una mezcla reductora mono o estéreo.
- 50 [0018] Adicionalmente, los datos HRTF se convierten en el dominio de parámetro por medio de una unidad de extracción de parámetro HRTF 305. Los parámetros HRTF resultantes se combinan en una unidad de conversión 307 para generar parámetros combinados referidos como parámetros binaurales. Estos parámetros describen el efecto combinado de los parámetros espaciales y el procesamiento HRTF.
- 55 [0019] El decodificador espacial sintetiza la señal de salida binaural modificando la señal de mezcla reductora decodificada dependiendo de los parámetros binaurales. Específicamente, la señal de mezcla reductora se transfiere a un dominio de banco de transformación o filtro mediante una unidad de transformación 309 (o el decodificador convencional 303 puede proporcionar directamente la señal de mezcla reductora decodificada como una señal de transformación). La unidad de transformación 309 puede comprender específicamente un banco de filtro QMF para generar sub-bandas QMF. La señal de mezcla reductora de sub-banda se alimenta a una unidad matriz 311 que lleva a
- 60

cabo una operación de matriz de 2 x 2 en cada sub-banda.

**[0020]** Si la mezcla reductora transmitida es una señal estéreo, las dos señales de entrada a la unidad matriz 311 son dos señales estéreo. Si la mezcla reductora transmitida es una señal mono, una de las señales de entrada a la unidad matriz 311 es la señal mono y la otra señal es una señal decorrelacionada (similar al mezclado ascendente convencional de una señal mono a una señal estéreo).

**[0021]** Para las mezclas descendentes tanto mono como estéreo, la unidad matriz 311 lleva a cabo la operación:

$$\begin{bmatrix} y_{L_B}^{n,k} \\ y_{R_B}^{n,k} \end{bmatrix} = \begin{bmatrix} h_{11}^{n,k} & h_{12}^{n,k} \\ h_{21}^{n,k} & h_{22}^{n,k} \end{bmatrix} \begin{bmatrix} y_{L_0}^{n,k} \\ y_{R_0}^{n,k} \end{bmatrix}$$

en donde  $k$  es el número del índice de la sub-banda,  $n$  el número de índice de ranura (intervalo de transformación),

$h_{ij}^{n,k}$

los elementos de matriz para la sub-banda  $k$ ,  $y_{L_0}^{n,k}$ ,  $y_{R_0}^{n,k}$  las dos señales de entrada para la sub-banda  $k$  y

$y_{L_B}^{n,k}$ ,  $y_{R_B}^{n,k}$  las muestras de la señal de salida binaural.

**[0022]** La unidad matriz 311 alimenta las muestras de la señal de salida binaural a una unidad inversa de transformación 313 que transforma la señal de nuevo en el dominio de tiempo. La señal binaural de dominio de tiempo resultante puede entonces alimentarse a los audífonos para proporcionar una experiencia de sonido envolvente.

**[0023]** El procedimiento descrito tiene un número de ventajas:

**[0024]** El procesamiento HRTF puede llevarse a cabo en el dominio de transformación lo cual, en muchos casos, puede reducir el número de transformaciones requerido, dado que puede utilizarse el mismo dominio de transformación para decodificar la señal de mezcla reductora.

**[0025]** La complejidad del procesamiento es muy baja (solamente utiliza multiplicación por matrices de 2 x 2) y es virtualmente independiente del número de canales de audio simultáneos. Puede aplicarse a mezclas descendentes tanto mono como estéreo;

Los HRTFs se representan en una manera muy compacta y, por tanto, pueden transmitirse y almacenarse muy eficientemente.

**[0026]** Sin embargo, el procedimiento tiene también algunas desventajas. Específicamente, el procedimiento solamente es adecuado para HRTFs que tienen respuestas de impulso relativamente cortas (generalmente menores que el intervalo de transformación) dado que no pueden representarse respuestas de impulso más largas por medio de los valores HRTF de sub-banda parametrizados. Por tanto, el procedimiento no puede utilizarse para entornos de audio que tienen grandes ecos o reverberaciones. Específicamente, el procedimiento no funciona típicamente con HRTFs ecoicos o con respuestas binaurales al impulso ambiental (BRIRs) que pueden ser largas y, por tanto, muy difíciles de moldear correctamente con el procedimiento paramétrico.

**[0027]** De aquí que sería ventajoso un sistema mejorado para generar una señal de audio binaural y, en particular, sería ventajoso un sistema que permita una flexibilidad incrementada, desempeño mejorado, implementación facilitada, uso de recursos reducido y/o aplicabilidad mejorada a diferentes entornos de audio.

**[0028]** El documento WO 2007/031896 A da a conocer una unidad de decodificador espacial dispuesta para la transformación de uno o más canales de audio en un par de canales de salida binaurales. El dispositivo comprende una unidad de conversión de parámetros para convertir los parámetros espaciales en parámetros binaurales que contienen información binaural. El dispositivo comprende, además, una unidad de síntesis espacial para la transformación de los canales de audio en un par de señales binaurales, durante el uso de los parámetros binaurales. La unidad de síntesis espacial opera en un dominio de transformada, tal como el dominio de QMF.

## SUMARIO DE LA INVENCION

**[0029]** Por consiguiente, la invención busca preferentemente mitigar, aliviar o eliminar una o más de las desventajas anteriormente mencionadas solas o en cualquier combinación.

5 **[0030]** De acuerdo con un primer aspecto de la invención, se proporciona un aparato para generar una señal de audio binaural, comprendiendo el aparato: medios para recibir datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales; medios de datos de parámetros para convertir los parámetros espaciales de los datos de parámetros espaciales en primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptual binaural; medios de conversión para convertir la señal de audio de M canales en una primera señal estéreo en respuesta a los primeros parámetros binaurales; un filtro de estéreo para generar la señal de audio binaural filtrando la primera señal de estéreo; y medios de coeficiente para determinar los coeficientes del filtro para el filtro de estéreo en respuesta a la función de transferencia perceptual binaural.

15 **[0031]** La invención puede permitir que se genere una señal de audio binaural mejorada. En particular, las modalidades de la invención pueden utilizar una combinación de procesamiento de frecuencia y tiempo para generar señales binaurales que reflejan entornos de audio ecoicos y/o HRTF o BRIRs con largas respuestas al impulso. Puede lograrse una implementación de baja complejidad. El procesamiento puede implementarse con bajas demandas de recursos computacionales y/o de memoria.

20 **[0032]** La señal de audio de M canales puede ser específicamente una señal mono o estéreo que comprende una mezcla reductora de un número mayor de canales espaciales, tal como una mezcla reductora de una señal envolvente 5.1 o 7.1. Los datos de parámetros espaciales pueden comprender específicamente diferencias intercanales y/o diferencias de correlación cruzada para la señal de audio de N canales. La(s) función(es) de transferencia perceptual binaural puede(n) ser HRTF o una función(es) de transferencia BRIR.

25 **[0033]** De acuerdo con una característica opcional de la invención, el aparato comprende además medios de transformación para transformar la señal de audio de M canales de un dominio de tiempo en un dominio de sub-banda y en donde los medios de conversión y el filtro de estéreo se encuentran dispuestos para procesar individualmente cada sub-banda del dominio de sub-banda.

30 **[0034]** La característica puede proporcionar implementación facilitada, demandas de recursos reducidas y/o compatibilidad con muchas aplicaciones de procesamiento de audio tales como los algoritmos de decodificación convencionales.

35 **[0035]** De acuerdo con una característica opcional de la invención, la duración de una respuesta al impulso de la función de transferencia perceptual binaural excede el intervalo de actualización de transformación.

40 **[0036]** La invención puede permitir que se genere una señal binaural mejorada y/o puede reducir la complejidad. Len particular, la invención puede generar señales binaurales correspondientes a entornos de audio con características de largo eco o reverberación.

**[0037]** De acuerdo con una característica opcional de la invención, el medio de conversión se encuentra dispuesto para generar, para cada sub-banda, muestras de salida de estéreo sustancialmente como:

45 
$$\begin{bmatrix} L_o \\ R_o \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} L_I \\ R_I \end{bmatrix}$$

50 en donde al menos uno de  $L_I$  y  $R_I$  es una muestra de un canal de audio de la señal de audio de M canales en la sub-banda y el medio de conversión se encuentra dispuesto para determinar coeficientes de matriz  $h_{xy}$  en respuesta tanto a los datos de parámetros espaciales como a la al menos una función de transferencia perceptual binaural.

**[0038]** La característica puede permitir que se genere una señal binaural mejorada y/o puede reducir la complejidad.

55 **[0039]** De acuerdo con una característica opcional de la invención, el medio de coeficiente comprende: medios para proporcionar representaciones de sub-banda de respuestas al impulso de una pluralidad de funciones de transferencia perceptual binaural correspondientes a diferentes fuentes de sonido en la señal de N canales; medios para determinar los coeficientes de filtro mediante una combinación ponderada de los coeficientes correspondientes de las representaciones de sub-banda; y medios para determinar los pesos de las representaciones de sub-banda para la combinación ponderada en respuesta a los datos de parámetros espaciales.

60 **[0040]** La invención puede permitir que se genere una señal binaural mejorada y/o puede reducir la complejidad. En

particular, puede determinarse la baja complejidad y, sin embargo, coeficientes de filtro de alta calidad.

**[0041]** De acuerdo con una característica opcional de la invención, los primeros parámetros binaurales comprenden parámetros de coherencia indicativos de una correlación entre los canales de la señal de audio binaural.

5

**[0042]** La característica puede permitir que se genere una señal binaural mejorada y/o puede reducir la complejidad. En particular, puede proporcionarse eficientemente la correlación deseada mediante una operación de baja complejidad previo al filtrado. Específicamente, puede llevarse a cabo una multiplicación de matriz de sub-banda de baja complejidad para introducir las propiedades de correlación o coherencia deseadas en la señal binaural. Tales propiedades pueden introducirse previo al filtrado y sin que se requiera modificar los filtros. Por tanto, la característica puede permitir controlar eficientemente y con baja complejidad las características de correlación o coherencia.

10

**[0043]** De acuerdo con una característica opcional de la invención, los primeros parámetros binaurales no comprenden al menos uno de parámetros de localización indicativos de la localización de cualquier fuente de sonido de la señal de audio binaural y parámetros de reverberación indicativos de una reverberación de cualquier componente de sonido de la señal de audio binaural.

15

**[0044]** La característica puede permitir que se genere una señal binaural mejorada y/o puede reducir la complejidad. En particular, la característica puede permitir controlar la localización de los parámetros de información y/o reverberación exclusivamente mediante los filtros, facilitando así la operación y/o proporcionando una calidad mejorada. La coherencia o correlación de los canales estéreo binaurales puede controlarse mediante el medio de conversión permitiendo así controlar la correlación/coherencia y la localización y/o reverberación independientemente y cuando sea más práctico o eficiente.

20

**[0045]** De acuerdo con una característica opcional de la invención, el medio de coeficiente se encuentra dispuesto para determinar los coeficientes de filtro para reflejar al menos una de las claves de localización y las claves de reverberación para la señal de audio binaural.

25

**[0046]** La característica puede permitir que se genere una señal binaural mejorada y/o puede reducir la complejidad. En particular, las propiedades de localización o reverberación deseadas pueden proporcionarse eficientemente mediante el filtrado de la sub-banda, proporcionando así una calidad mejorada y, en particular, permitiendo estimular eficientemente, por ejemplo, los entornos de audio ecoicos.

30

**[0047]** De acuerdo con una característica opcional de la invención, la señal de audio de M canales es una señal de audio mono y el medio de conversión se encuentra dispuesto para generar una señal decorrelacionada a partir de la señal de audio mono y para generar la primera señal estéreo mediante una multiplicación de matriz aplicada a las muestras de una señal estéreo que comprende la señal decorrelacionada y la señal de audio mono.

35

**[0048]** La característica puede permitir que se genere una señal binaural mejorada a partir de una señal mono y/o puede reducir la complejidad. En particular, la invención puede permitir que se generen todos los parámetros requeridos para generar una señal de audio binaural de alta calidad a partir de los parámetros espaciales típicamente disponibles.

40

**[0049]** De acuerdo con otro aspecto de la invención, se proporciona un procedimiento para generar una señal de audio binaural, comprendiendo el procedimiento: recibir los datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y los datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales; convertir los parámetros espaciales de los datos de parámetros espaciales en los primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptual binaural; convertir la señal de audio de M canales en la primera señal estéreo en respuesta a los primeros parámetros binaurales; generar la señal de audio binaural filtrando la primera señal estéreo; y determinar los coeficientes de filtro para el filtro de estéreo en respuesta a la al menos una función de transferencia perceptual binaural.

45

50

**[0050]** De acuerdo con otro aspecto de la invención, se proporciona un transmisor para transmitir una señal de audio binaural, comprendiendo el transmisor: medios para recibir los datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y los datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales; medios de datos de parámetros para convertir los parámetros espaciales de los datos de parámetros espaciales en los primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptual binaural; medios de conversión para convertir la señal de audio de M canales en la primera señal estéreo en respuesta a los primeros parámetros binaurales; un filtro de estéreo para generar la señal de audio binaural filtrando la primera señal estéreo; medios de coeficiente para determinar los coeficientes de filtro para el filtro de estéreo en respuesta a la al menos una función de transferencia

55

60

perceptual binaural; y medios para transmitir la señal de audio binaural.

5 **[0051]** De acuerdo con otro aspecto de la invención, se proporciona un sistema de transmisión para transmitir una señal de audio, incluyendo el sistema de transmisión, un transmisor que comprende: medios para recibir los datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y los datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales; medios de datos de parámetros para convertir los parámetros espaciales de los datos de parámetros espaciales en los primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptual binaural; medios de conversión para convertir la señal de audio de M canales en la primera señal estéreo en respuesta a los primeros parámetros binaurales; un filtro de estéreo para generar la señal de audio binaural filtrando la primera señal estéreo; medios de coeficiente para determinar los coeficientes de filtro para el filtro de estéreo en respuesta a la función de transferencia perceptual binaural; medios para transmitir la señal de audio binaural; y un receptor para recibir la señal de audio binaural.

15 **[0052]** De acuerdo con otro aspecto de la invención, se proporciona un dispositivo de grabación de audio para grabar una señal de audio binaural, comprendiendo el dispositivo de grabación de audio medios para recibir los datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y los datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales; medios de datos de parámetros para convertir los parámetros espaciales de los datos de parámetros espaciales en los primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptual binaural; medios de conversión para convertir la señal de audio de M canales en la primera señal estéreo en respuesta a los primeros parámetros binaurales; un filtro de estéreo para generar la señal de audio binaural filtrando la primera señal estéreo; medios de coeficiente (419) para determinar los coeficientes de filtro para el filtro de estéreo en respuesta a la función de transferencia perceptual binaural; y medios para grabar la señal de audio binaural.

25 **[0053]** De acuerdo con otro aspecto de la invención, se proporciona un procedimiento para transmitir una señal de audio binaural, comprendiendo el procedimiento: recibir los datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y los datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales; convertir los parámetros espaciales de los datos de parámetros espaciales en los primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptual binaural; convertir la señal de audio de M canales en la primera señal estéreo en respuesta a los primeros parámetros binaurales; generar la señal de audio binaural filtrando la primera señal estéreo en un filtro de estéreo; determinar los coeficientes de filtro para el filtro de estéreo en respuesta a la función de transferencia perceptual binaural; y transmitir la señal de audio binaural.

35 **[0054]** De acuerdo con otro aspecto de la invención, se proporciona un procedimiento para transmitir y recibir una señal de audio binaural, comprendiendo el procedimiento: un transmisor que lleva a cabo las etapas de: recibir los datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y los datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales; convertir los parámetros espaciales de los datos de parámetros espaciales en los primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptual binaural; convertir la señal de audio de M canales en la primera señal estéreo en respuesta a los primeros parámetros binaurales; generar la señal de audio binaural filtrando la primera señal estéreo en un filtro de estéreo; determinar los coeficientes de filtro para el filtro de estéreo en respuesta a la función de transferencia perceptual binaural, y transmitir la señal de audio binaural; y un receptor que lleva a cabo la etapa de recibir la señal de audio binaural.

**[0055]** De acuerdo con otro aspecto de la invención, se proporciona un producto de programa de computadora para ejecutar el procedimiento de cualquiera de los procedimientos anteriormente descritos.

50 **[0056]** Estos y otros aspectos, características y ventajas de la invención serán aparentes a partir de, y elucidados con referencia a, la(s) modalidad(es) descrita(s) en adelante.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

55 Las modalidades de la invención se describirán solamente a modo de ejemplo, con referencia a los dibujos, en los cuales:

La Figura 1 es una ilustración de un procedimiento para la generación de una señal binaural de acuerdo con la técnica anterior;

60 La Figura 2 es una ilustración de un procedimiento para la generación de una señal binaural de acuerdo con la técnica

anterior;

La Figura 3 es una ilustración de un procedimiento para la generación de una señal binaural de acuerdo con la técnica anterior;

5

La Figura 4 ilustra un dispositivo para generar una señal de audio binaural de acuerdo con algunas modalidades de la invención;

10

La Figura 5 ilustra un diagrama de flujo de un ejemplo de un procedimiento para generar una señal de audio binaural de acuerdo con algunas modalidades de la invención; y

La Figura 6 ilustra un ejemplo de un sistema de transmisión para la comunicación de una señal de audio de acuerdo con algunas modalidades de la invención.

15

#### DESCRIPCIÓN DETALLADA DE LAS MODALIDADES

**[0058]** La siguiente descripción se enfoca en las modalidades de la invención aplicables a la síntesis de una señal estéreo binaural a partir de una mezcla reductora de una pluralidad de canales espaciales. En particular, la descripción será apropiada para la generación de una señal binaural para su reproducción en audífonos a partir de una corriente de bits de sonido envolvente MPEG codificada utilizando una configuración llamada '5151' que tiene 5 canales como entrada (indicados por el primer '5'), una mezcla reductora mono (la 'primera'), una reconstrucción de 5 canales (el segundo '5') y una parametrización de acuerdo con la estructura de árbol de árbol '1'. Puede encontrarse información detallada acerca de diferentes estructuras de árbol en Herre, J., Kjörling K., Breebaart J., Faller C., Disch S., Purnhagen H., Koppens J., Hilpert J., Rödén J., Oomen W., Linzmeier K., Chong K.S., "MPEG Surround – The ISO/MPEG standard for efficient and compatible multi-channel audio coding" (Sonido envolvente MPEG – El estándar ISO/MPEG para la codificación eficiente y compatible de audio multicanal), Proc. 122 AES convention, Viena, Austria (2007) y Breebaart J., Hotho G, Koppens J., Schuijers E., Oomen W., van de Par S., "Background, concept and architecture of the recent MPEG Surround standard on multi-channel audio compression" (Antecedentes, concepto y arquitectura del estándar reciente de sonido envolvente MPEG en la compresión del audio multicanal), J. Audio Engineering Society, 55 p. 331-351 (2007). Sin embargo, se apreciará que la invención no se limita a esta aplicación, sino que puede aplicarse, por ejemplo, a muchas otras señales de audio incluyendo, por ejemplo, señales de sonido envolvente en mezcla reductora a una señal de estéreo.

20

25

30

35

40

**[0059]** En dispositivos de la técnica anterior, tales como el de la Figura 3, no pueden representarse eficientemente los HRTFs o BRIRs largos por medio de los datos parametrizados y la operación de matriz llevada a cabo por la unidad matriz 311. En efecto, las multiplicaciones de matriz de sub-banda se limitan a representar respuestas al impulso del dominio de tiempo que tienen una duración que corresponde al intervalo de tiempo de transformación utilizado para la transformación al dominio de tiempo de sub-banda. Por ejemplo, si la transformación es una transformación Fourier rápida (FFT) cada intervalo FFT de las muestras N se transfiere a las muestras de sub-banda N que se alimentan a la unidad matriz. Sin embargo, las respuestas al impulso más largas que las muestras N no se representarán adecuadamente.

45

**[0060]** Una solución a este problema es utilizar un procedimiento de filtrado del dominio de sub-banda en donde la operación de matriz se reemplaza por un procedimiento de filtrado de matriz en donde se filtran las sub-bandas individuales. Por tanto, en tales modalidades, el procesamiento de sub-banda, en lugar de una simple multiplicación de matriz, puede proporcionarse como:

50

$$\begin{bmatrix} y_{L_B}^{n,k} \\ y_{R_B}^{n,k} \end{bmatrix} = \sum_{i=0}^{N_q-1} \begin{bmatrix} h_{11}^{n-i,k} & h_{12}^{n-i,k} \\ h_{21}^{n-i,k} & h_{22}^{n-i,k} \end{bmatrix} \begin{bmatrix} y_{L_0}^{n-i,k} \\ y_{R_0}^{n-i,k} \end{bmatrix}$$

en donde  $N_q$  es el número de derivaciones utilizadas por el filtro para representar la(s) función(es) HRTF/BRIR.

55

**[0061]** Tal procedimiento corresponde efectivamente a aplicar cuatro filtros a cada sub-banda (uno por cada permutación del canal de entrada y del canal de salida de la unidad matriz 311).

60

**[0062]** Aunque tal procedimiento puede ser ventajoso en algunas modalidades también tiene algunas desventajas asociadas. Por ejemplo, el sistema requiere cuatro filtros para cada sub-banda lo cual incrementa significativamente la complejidad y los requerimientos de recursos para el procesamiento. Además, en muchos casos puede ser complicado, difícil o incluso imposible generar los parámetros que corresponden precisamente a las respuestas al impulso



HRTF/BRIR deseadas.

5 **[0063]** Específicamente, para la simple multiplicación de matriz de la Figura 3, la coherencia de la señal binaural puede estimarse con la ayuda de parámetros HRTF y de parámetros espaciales transmitidos debido a que ambos tipos de parámetro existen en el mismo dominio (parámetro). La coherencia de la señal binaural depende de la coherencia entre las señales individuales de la fuente de sonido (como se describe mediante los parámetros espaciales) y de la trayectoria acústica desde las posiciones individuales hasta los auriculares (descrita mediante los HRTFs). Si los niveles de señal relativos, los valores de coherencia en pares y las funciones de transferencia HRTF se describen todos de una manera estadística (paramétrica), la coherencia neta que resulta del efecto combinado de la emisión espacial y el procesamiento HRTF puede estimarse directamente en el dominio de parámetro. Este proceso se describe en Breebaart J., "Analysis and synthesis of binaural parameters for efficient 3D audio rendering in MPEG Surround" (Análisis y síntesis de parámetros binaurales para la emisión eficiente de audio 3D en sonido envolvente MPEG), Proc. ICME, Beijing, China (2007) y Breebaart J., Faller C., "Spatial audio processing: MPEG Surround and other applications" (Procesamiento de audio espacial: sonido envolvente MPEG y otras aplicaciones), Wiley & Sons, New York (2007). Si se conoce la coherencia deseada, puede obtenerse una señal de salida con una coherencia de acuerdo con el valor especificado mediante una combinación de una señal decorrelacionada y la señal mono, por medio de una operación de matriz. Este proceso se describe en Breebaart J., van de Par S., Koohrausch A., Schuijers E., "Parametric coding of stereo audio" (Codificación paramétrica de audio en estéreo), EURASIP J. Applied Signal Proc. 9, p. 1305-1322 (2005) y Engdegård J., Purnhagen H., Rödén J., Liljeryd L., "Synthetic ambience in parametric stereo coding" (Ambientación sintética en la codificación paramétrica en estéreo), Proc. 116<sup>th</sup> AES convention, Berlin, Alemania (2004).

25 **[0064]** Como resultado, las entradas de matriz de señal decorrelacionada ( $h_{12}$  y  $h_{22}$ ) resultan de las relaciones relativamente simples entre los parámetros espaciales y HRTF. Sin embargo, para respuestas de filtro tales como las descritas anteriormente, es significativamente más difícil calcular la coherencia resultante de la decodificación espacial y de la síntesis binaural debido a que el valor de coherencia deseado es diferente para la primera parte (el sonido directo) de la BRIR que para la parte restante (la reverberación tardía).

30 **[0065]** Específicamente, para BRIRs, las propiedades requeridas pueden cambiar considerablemente con el tiempo. Por ejemplo, la primera parte de una BRIR puede describir el sonido directo (sin efectos de ambiente). En consecuencia, esta parte es altamente direccional (con distintas propiedades de localización reflejadas por ejemplo, por las diferencias de nivel y las diferencias de tiempo de llegada, y una alta coherencia). Las reflexiones tempranas y la reverberación tardía, por otra parte, son, frecuentemente, relativamente menos direccionales. Por tanto, las diferencias de nivel entre los oídos son menos pronunciadas, las diferencias de tiempo de llegada son difíciles de determinar con precisión debido a la naturaleza estoquiástica de estas y la coherencia, en muchos casos, es bastante baja. Este cambio de propiedades de localización es bastante importante para capturar con precisión, pero esto puede dificultarse debido a que requeriría que la coherencia de las respuestas de filtro se cambiaran dependiendo de la posición dentro de la respuesta de filtro real, mientras que, al mismo tiempo, la respuesta total de filtro debería depender de los parámetros espaciales y los coeficientes HRTF. Esta combinación de requerimientos es muy difícil de cumplir con un número limitado de etapas de procesamiento.

40 **[0066]** En suma, determinar la coherencia correcta entre las señales de salida binaural y asegurar su comportamiento temporal correcto es muy difícil para una mezcla reductora mono y típicamente es imposible utilizando los procedimientos conocidos para el procedimiento de multiplicación de matriz de la técnica anterior.

45 **[0067]** La Figura 4 ilustra un dispositivo para generar una señal de audio binaural de acuerdo con algunas modalidades de la invención. En el procedimiento descrito, la multiplicación paramétrica de matriz se combina con filtración de baja complejidad para permitir emular entornos de audio con largo eco o reverberación. En particular, el sistema permite utilizar largos HRTFs/BRIRs mientras se mantiene la baja complejidad y la práctica implementación.

50 **[0068]** El dispositivo comprende un desmultiplexor 401 que recibe una corriente de bits de datos de audio que comprende una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales. Además, los datos comprenden datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales. En el ejemplo específico, la señal de mezcla reductora es una señal mono, i.e., M=1 y la señal de audio de N canales es una señal envolvente 5.1, i.e., N=6. Los datos de audio son específicamente una codificación de sonido envolvente MPEG de una señal envolvente y los datos espaciales comprenden diferencias inter-niveles (ILDs) y parámetros de correlación cruzada intercanales (ICC).

60 **[0069]** Los datos de audio de la señal mono se alimentan a un decodificador 403 acoplado al desmultiplexor 401. El decodificador 403 decodifica la señal mono utilizando un algoritmo de decodificación convencional adecuado, como lo sabrá bien la persona experta en la técnica. por tanto, en el ejemplo, la salida del decodificador 403 es una señal de audio mono decodificada.

5 **[0070]** El decodificador 403 se encuentra acoplado a un procesador de transformación 405 que es operable para convertir la señal mono decodificada del dominio de tiempo a un dominio de sub-banda de frecuencia. En algunas modalidades, el procesador de transformación 405 puede encontrarse dispuesto para dividir la señal en intervalos de transformación (correspondiendo a los bloques de muestra que comprenden un número de muestras adecuado) y para llevar a cabo una transformación Fourier rápida (FFT) en cada intervalo de tiempo de transformación. Por ejemplo, la FFT puede ser una FFT de 64 puntos siendo divididas las muestras de audio mono en 64 bloques de muestra a los cuales se aplica la FFT para generar 64 muestras de sub-banda complejas.

10 **[0071]** En el ejemplo específico, el procesador de transformación 405 comprende un banco de filtros QMF que opera con un intervalo de transformación de 64 muestras. Por tanto, para cada bloque de 64 muestras de dominio de tiempo, se generan 64 muestras de sub-banda en el dominio de frecuencia.

15 **[0072]** En el ejemplo, la señal recibida es una señal mono que va a mezclarse de manera ascendente con una señal estéreo binaural. Por consiguiente, la señal mono de sub-banda de frecuencia se alimenta a un descorrelacionador 407 que genera una versión decorrelacionada de la señal mono. Se apreciará que puede utilizarse cualquier procedimiento adecuado para generar una señal decorrelacionada sin detracción de la invención.

20 **[0073]** Las salidas del procesador de transformación 405 y el descorrelacionador 407 se alimentan a un procesador de matriz 409. Por tanto, el procesador de matriz 409 se alimenta a la representación de sub-banda de la señal mono así como a la representación de sub-banda de la señal decorrelacionada generada. El procesador de matriz 409 procede para convertir la señal mono en una primera señal estéreo. Específicamente, el procesador de matriz 409 lleva a cabo una multiplicación de matriz en cada sub-banda, proporcionada por:

25

$$\begin{bmatrix} L_o \\ R_o \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} L_l \\ R_l \end{bmatrix}$$

30 en donde  $L_l$  y  $R_l$  son la muestra de las señales de entrada al procesador de matriz 409, i.e., en el ejemplo específico,  $L_l$  y  $R_l$  son las muestras de sub-banda de la señal mono y de la señal decorrelacionada.

35 **[0074]** La conversión efectuada por el procesador de matriz 409 depende de los parámetros binaurales generados en respuesta a los HRTFs/BRIRs. En el ejemplo, la conversión depende también de los parámetros espaciales que relacionan la señal mono recibida y los canales espaciales (adicionales).

40 **[0075]** Específicamente, el procesador de matriz 409 se encuentra acoplado a un procesador de conversión 411 que además se encuentra acoplado al desmultiplexor 401 y un almacén de HRTF 413 que comprende los datos que representan el(los) HRTF(s) deseado(s) (o equivalentemente, el(los) BRIR(s) deseado(s)). Por brevedad, lo siguiente se referirá solamente a HRTF(s), pero se apreciará que pueden utilizarse BRIR(s) en lugar de (o además de) HRTF(s). El procesador de conversión o411 recibe los datos espaciales del desmultiplexor y los datos que representan el HRTF del almacén de HRTF 413. El procesador de conversión 411 procede entonces para generar los parámetros binaurales utilizados por el procesador de matriz 409 convirtiendo los parámetros espaciales en los primeros parámetros binaurales en respuesta a los datos HRTF.

45 **[0076]** Sin embargo, en el ejemplo, no se calcula la parametrización total del HRTF ni los parámetros necesarios para generar una señal binaural de salida. Por el contrario, los parámetros binaurales utilizados en la multiplicación de matriz reflejan solamente parte de la respuesta HRTF deseada. En particular, se estiman los parámetros binaurales solamente para la parte directa (excluyendo las reflexiones tempranas y la reverberación tardía) del HRTF/BRIR. Esto se logra utilizando el proceso convencional de estimación de parámetro utilizando el primer pico de la respuesta al impulso del dominio de tiempo de HRTF solamente durante el proceso de parametrización de HRTF. Solamente se utiliza subsecuentemente la coherencia resultante para la parte directa (excluyendo las claves de localización tales como las diferencias en nivel y/o tiempo) en la matriz de  $2 \times 2$ . De hecho, en el ejemplo específico, se generan los coeficientes de matriz solamente para reflejar la coherencia o correlación deseada de la señal binaural y no se incluye la consideración de las características de localización o reverberación.

50 **[0077]** Por tanto, la multiplicación de matriz solamente lleva a cabo parte del procesamiento deseado y la salida del procesador de matriz 409 no es la señal binaural final, sino, por el contrario, es una señal intermedia (binaural) que refleja la coherencia deseada del sonido directo entre los canales.

60 **[0078]** Los parámetros binaurales en forma de los coeficientes de matriz  $h_{xy}$  se encuentran en el ejemplo generado

calculando primeramente las potencias relativas de la señal en los diferentes canales de audio de la señal de N canales en base a los datos espaciales y, específicamente, en base a los parámetros de diferencia de nivel contenidos en los mismos. Las potencias relativas en cada uno de los canales binaurales se calculan entonces en base a estos valores y a los HRTFs asociados con cada uno de los canales N. También, se calcula un valor esperado para la correlación cruzada entre las señales binaurales en base a las potencias de la señal en cada uno de los canales N y de los HRTFs. En base a la correlación cruzada y la potencia combinada de la señal binaural, se calcula subsecuentemente una medición de coherencia para el canal y se determinan los parámetros de matriz para proporcionar esta correlación. Los detalles específicos de cómo pueden generarse los parámetros binaurales se describirán más adelante.

5  
10 **[0079]** El procesador de matriz 409 se encuentra acoplado a dos filtros 415, 417 que son operables para generar la señal de audio binaural de salida filtrando la señal estéreo generada por el procesador de matriz 409. Específicamente, cada una de las dos señales se filtra individualmente como una señal mono y no se introduce ningún acoplamiento cruzado de ninguna señal de un canal al otro. Por consiguiente, solamente se emplean dos filtros mono reduciendo así la complejidad, en comparación, por ejemplo, con procedimientos que requieren cuatro filtros.

15 **[0080]** Los filtros 415, 417 son filtros de sub-banda, en donde cada sub-banda se filtra individualmente. Específicamente, cada uno de los filtros puede ser un filtro de respuesta finita al impulso (FIR) en cada sub-banda, que lleva a cabo un filtrado proporcionado sustancialmente por:

20

$$z^{n,k} = \sum_{i=0}^{N_i-1} c_i^k \cdot y_0^{n-i,k}$$

25 en donde  $y$  representa las muestras de sub-banda recibidas del procesador de matriz 409,  $c$  son los coeficientes de filtro,  $n$  es el número de muestra (correspondiente al número de intervalo de transformación),  $k$  es la sub-banda y  $N$  es la longitud de la respuesta al impulso del filtro. Por tanto, en cada sub-banda individual, se lleva a cabo un filtrado de "dominio de tiempo" extendiendo así el procesamiento de encontrarse en un solo intervalo de transformación a tomar en cuenta las muestras de sub-banda de una pluralidad de intervalos de transformación.

30 **[0081]** Las modificaciones de señal del sonido envolvente MPEG se llevan a cabo en el dominio de un complejo banco de filtro modulado, el QMF, que no se muestrea críticamente. Su particular diseño permite implementar un filtro de dominio de tiempo dado a una alta precisión filtrando cada señal de sub-banda en la dirección de tiempo con un filtro separado. La SNR total resultante para la implementación del filtro se encuentra en el rango de 50 dB con la parte alienada del error significativamente más pequeña. Además, estos filtros de dominio de sub-banda pueden  
35 suministrarse directamente desde el filtro de dominio de tiempo dado. Un procedimiento particularmente atractivo para computar el filtro de dominio de sub-banda correspondiente al filtro de dominio de tiempo  $h(\nu)$ , es utilizar un segundo banco complejo de filtro de análisis modulado con un filtro prototipo FIR  $q(\nu)$  derivado del filtro prototipo del banco de filtro QMF. Específicamente.

40

$$c_i^k = \sum_{\nu} h(\nu + iL)q(\nu) \exp\left(-j\frac{\pi}{L}\left(k + \frac{1}{2}\right)\nu\right)$$

45 en donde  $L = 64$ . Para el banco QMF de sonido envolvente MPEG, el filtro prototipo convertidor de filtro  $q(\nu)$  tiene 192 derivaciones. Como ejemplo, un filtro de dominio de tiempo con 1024 derivaciones se convertirá en un conjunto de 64 filtros de sub-banda teniendo todos 18 derivaciones en la dirección de tiempo.

50 **[0082]** Las características del filtro, en el ejemplo generado, son para reflejar ambos aspectos de los parámetros espaciales así como los aspectos de los HRTFs deseados. Específicamente, se determinan los coeficientes de filtro en respuesta a las respuestas al impulso de HRTF y a las claves de localización espacial, de tal manera que las características de reverberación y localización de la señal binaural generada se introducen y se controlan por medio de los filtros. La correlación o coherencia de la parte directa de las señales binaurales no se afecta por el filtrado  
55 asumiendo que la parte directa de los filtros es (casi) coherente y, por tanto, la coherencia del sonido directo de la salida binaural se define totalmente mediante la operación de matriz precedente. La parte de reverberación tardía de los filtros, por otra parte, se asume no correlacionada entre los filtros del oído izquierdo y derecho y, por tanto, la salida de esa parte específica será siempre no correlacionada, independientemente de la coherencia de la señal alimentada en estos filtros. De aquí que no se requiere ninguna modificación para los filtros en respuesta a la coherencia deseada. Por tanto, la operación de matriz que procede a los filtros determina la coherencia deseada de la parte directa, mientras que  
60 la parte de reverberación restante tendrá automáticamente la correlación correcta (baja), independientemente de los valor de matriz reales. Por tanto, el filtrado mantiene la coherencia deseada introducida por el procesador de matriz 409.

**[0083]** Por tanto, en el dispositivo de la Figura 4, los parámetros binaurales (en forma de los coeficientes de matriz) utilizados por el procesador de matriz 409 son parámetros de coherencia indicativos de una correlación entre los canales de la señal de audio binaural. Sin embargo, estos parámetros no comprenden parámetros de localización indicativos de la localización de cualquier fuente de sonido de la señal de audio binaural o parámetros de reverberación indicativos de la reverberación de cualquier componente de sonido de la señal de audio binaural. Por el contrario, estos parámetros/características se introducen mediante el subsecuente filtrado de sub-banda determinando los coeficientes de filtro, de tal manera que reflejan las claves de localización y las claves de reverberación para la señal de audio binaural.

**[0084]** Específicamente, los filtros se encuentran acoplados a un procesador de coeficiente 419 que se encuentra acoplado además al desmultiplexor 401 y al almacén de HRTF 413. El procesador de coeficiente 419 determina los coeficientes de filtro para el filtro de estéreo 415, 417, en respuesta a la(s) función(es) de transferencia perceptual binaural. Además, el procesador de coeficiente 419 recibe los datos espaciales del desmultiplexor 401 y utiliza esto para determinar los coeficientes de filtro.

**[0085]** Específicamente, las respuestas al impulso HRTF se convierten en el dominio de sub-banda y, a medida que la respuesta al impulso excede un solo intervalo de transformación, esto da como resultado una respuesta al impulso para cada canal en cada sub-banda más que en un solo coeficiente de sub-banda. Las respuestas al impulso para cada filtro HRTF que corresponden a cada uno de los canales N se suman entonces en una suma calculada. Los pesos que se aplican a cada una de las respuestas al impulso del filtro N HRTF se determinan en respuesta a los datos espaciales y se determinan específicamente para dar como resultado la distribución de la potencia apropiada entre los diferentes canales. Los detalles específicos de cómo pueden generarse los coeficientes de filtro se describirán más adelante.

**[0086]** La salida de los filtros 415, 417, es, por tanto, una representación de la sub-banda estéreo de una señal de audio binaural que emula efectivamente una señal envolvente total cuando se presenta en los audífonos. Los filtros 415, 417, se encuentran acoplados a un procesador de transformación inversa 421 que lleva a cabo una transformación inversa para convertir la señal de sub-banda en el dominio de tiempo. Específicamente, el procesador de transformación inversa 421 puede llevar a cabo una transformación QMF inversa.

**[0087]** Por tanto, la salida del procesador de transformación inversa 421 es una señal binaural que puede proporcionar una experiencia de sonido envolvente desde un juego de audífonos. La señal, por ejemplo, puede codificarse utilizando un codificador estéreo convencional y/o puede convertirse en el dominio análogo en un convertidor análogo o digital para proporcionar una señal que pueda alimentarse directamente a los audífonos.

**[0088]** Por tanto, el dispositivo de la Figura 4 combina el procesamiento de matriz HRTF paramétrico y el filtrado de sub-banda para proporcionar una señal binaural. La separación de una multiplicación de matriz de correlación/coherencia y un filtro, en base al filtrado de localización y reverberación, proporciona un sistema en donde pueden computarse fácilmente los parámetros requeridos, por ejemplo, para una señal mono. Específicamente, en contraste con un procedimiento de filtrado puro, en donde es difícil o imposible determinar e implementar el parámetro de coherencia, la combinación de diferentes tipos de procesamiento permite controlar eficientemente la coherencia incluso para aplicaciones basadas en una señal mono de mezcla reductora.

**[0089]** Por tanto, el procedimiento descrito tiene la ventaja de que la síntesis de la coherencia correcta (por medio de la multiplicación de matriz) y la generación de las claves de localización y reverberación (por medio de los filtros) se separan y se controlan completamente de manera independiente. Además, el número de filtros se limita a dos, dado que no se requiere un filtrado de canal cruzado. A medida que los filtros son típicamente más complejos que la simple multiplicación de matriz, la complejidad se reduce.

**[0090]** En lo siguiente, se describirá un ejemplo específico de cómo pueden calcularse los parámetros binaurales de matriz y los coeficientes de filtro requeridos. En el ejemplo, la señal recibida es una corriente de bits envolvente de MPEG codificada utilizando una estructura de árbol '5151'.

**[0091]** En la descripción se utilizarán los siguientes acrónimos:

- l o L: canal izquierdo
- r o R: canal derecho
- f: canal(es) frontal(es)
- s: canal(es) de sonido envolvente
- c: canal central
- ls: sonido envolvente izquierdo

rs: sonido envolvente derecho  
 lf: frente izquierdo  
 lr: derecha izquierda

5 Los datos espaciales comprendidos en la corriente de datos MPEG incluyen los siguientes parámetros:

| Parámetro             | Descripción  |
|-----------------------|--|
| 10 CLD <sub>fs</sub>  | diferencia de nivel frontal vs sonido envolvente                                   |
| CLD <sub>fc</sub>     | diferencia de nivel frontal vs centro  |
| CLD <sub>f</sub>      | diferencia de nivel izquierda frontal vs derecha frontal                           |
| 15 CLD <sub>s</sub>   | diferencia de nivel izquierda de sonido envolvente vs derecha de sonido envolvente |
| ICC <sub>fs</sub>     | correlación frontal vs sonido envolvente   |
| 20 ICC <sub>fc</sub>  | correlación frontal vs central   |
| ICC <sub>f</sub>      | correlación izquierda frontal vs derecha frontal                                   |
| ICC <sub>s</sub>      | correlación izquierda de sonido envolvente vs derecha de sonido envolvente         |
| 25 CLD <sub>ife</sub> | diferencia de nivel central vs LFE   |

**[0092]** Primeramente, se describirá la generación de los parámetros binaurales utilizados para la multiplicación de matriz por el procesador de matriz 409.

30 **[0093]** El procesador de conversión 411 calcula primero un estimado de la coherencia binaural que es un parámetro que refleja la coherencia deseada entre los canales de la señal de salida binaural. La estimación utiliza los parámetros espaciales así como los parámetros HRTF determinados para las funciones HRTF.

35 **[0094]** Específicamente, se utilizan los siguientes parámetros HRTF:

$P_l$ , que es la potencia rms dentro de cierta banda de frecuencia de un HRTF correspondiente al oído izquierdo.

$P_r$ , que es la potencia rms dentro de cierta banda de frecuencia de un HRTF correspondiente al oído derecho.

40  $\rho$ , que es la coherencia dentro de cierta banda de frecuencia entre el HRTF del oído izquierdo y derecho para cierta posición virtual de la fuente de la fuente de sonido.

$\phi$ , que es la diferencia porcentual de fase dentro de cierta banda de frecuencia entre el HRTF del oído izquierdo y derecho para cierta posición virtual de la fuente de sonido.

45 **[0095]** Asumiendo que la representación HRTF del dominio de frecuencia es  $H_l(f)$ ,  $H_r(f)$ , para los oídos izquierdo y derecho, respectivamente y  $f$  el índice de frecuencia, estos parámetros pueden calcularse de acuerdo con:

5

$$P_l = \sqrt{\sum_{f=f(b)}^{f=f(b+1)-1} H_l(f)H_l^*(f)}$$

10

$$P_r = \sqrt{\sum_{f=f(b)}^{f=f(b+1)-1} H_r(f)H_r^*(f)}$$

15

$$\varphi = \arg\left(\sum_{f=f(b)}^{f=f(b+1)-1} H_l(f)H_r^*(f)\right)$$

20

$$\rho = \frac{\left|\sum_{f=f(b)}^{f=f(b+1)-1} H_l(f)H_r^*(f)\right|}{P_l P_r}$$

25

**[0096]** Cuando se lleva a cabo la suma a través de  $f$  para cada banda de parámetro, da como resultado un conjunto de parámetros para cada banda de parámetro  $b$ . Puede obtenerse más información acerca de este proceso de parametrización HRTF de Breebaart J., "Analysis and synthesis of binaural parameters for efficient 3D audio rendering in MPEG Surround" (Análisis y síntesis de parámetros binaurales para la emisión eficiente de audio 3D en sonido envolvente MPEG), Proc. ICME, Beijing, China (2007) y Breebaart J., Faller C., "Spatial audio processing: MPEG Surround and other applications" (Procesamiento de audio espacial: sonido envolvente MPEG y otras aplicaciones), Wiley & Sons, New York (2007).

30

**[0097]** El proceso de parametrización anterior se lleva a cabo independientemente para cada banda de parámetro y cada posición virtual de altavoz. En lo siguiente, la posición del altavoz se denota mediante  $P_i(X)$ , siendo  $X$  el identificador del altavoz (lf, rf, c, ls, o rs).

35

**[0098]** Como primera etapa, las potencias relativas (con respecto a la potencia de la señal de entrada mono) de la señal del canal 5.1 se computan utilizando los parámetros CLD transmitidos. La potencia relativa del canal izquierdo-frontal se proporciona mediante:

$$\sigma_{lf}^2 = r_1(\text{CLD}_{fs})r_1(\text{CLD}_{fc})r_1(\text{CLD}_f)$$

con

$$r_1(\text{CLD}) = \frac{10^{\text{CLD}/10}}{1 + 10^{\text{CLD}/10}}$$

40

y

$$r_2(\text{CLD}) = \frac{1}{1 + 10^{\text{CLD}/10}}$$

45

**[0099]** De manera similar, las potencias relativas de los otros canales se proporciona mediante:

$$\sigma_{rf}^2 = r_1(\text{CLD}_{fs})r_1(\text{CLD}_{fc})r_2(\text{CLD}_f)$$

50

$$\sigma_c^2 = r_1(\text{CLD}_{fs})r_2(\text{CLD}_{fc})$$

$$\sigma_{ls}^2 = r_2(\text{CLD}_{fs})r_1(\text{CLD}_s)$$

55

$$\sigma_{rs}^2 = r_2(\text{CLD}_{fs})r_2(\text{CLD}_s)$$

60

**[0100]** Dadas las potencias  $\sigma$  de cada altavoz virtual, los parámetros ICC que representan los valores de coherencia entre ciertos pares de altavoces, y los parámetros HRTF  $P_l$ ,  $P_r$ ,  $\sigma$  y  $\varphi$ , para cada altavoz virtual, pueden estimarse los atributos estadísticos de la señal binaural resultante. Esto se logra agregando la contribución en términos de potencia  $\sigma$  para cada altavoz virtual, multiplicada por la potencia del HRTF  $P_l$ ,  $P_r$  para cada oído individualmente, para reflejar el cambio en la potencia introducido por el HRTF. Se requieren términos adicionales para incorporar el efecto de las correlaciones mutuas entre las señales virtuales de altavoz (ICC) y las diferencias en la longitud de trayectoria del HRTF

(representado por el parámetro  $\phi$ ) (ref., por ejemplo, Breebaart J., Faller C., "Spatial audio processing: MPEG Surround and other applications" (Procesamiento de audio espacial: sonido envolvente MPEG y otras aplicaciones), Wiley & Sons, New York (2007)).

5 **[0101]** El valor esperado de la potencia relativa del canal izquierdo de salida binaural  $\sigma_L^2$  (con respecto al canal de entrada mono), se proporciona mediante:

$$10 \quad \sigma_L^2 = P_l^2(C)\sigma_c^2 + P_l^2(Lf)\sigma_{lf}^2 + P_l^2(Ls)\sigma_{ls}^2 + P_l^2(Rf)\sigma_{rf}^2 + P_l^2(Rs)\sigma_{rs}^2 + \dots \\ 2P_l(Lf)P_l(Rf)\rho(Rf)\sigma_{lf}\sigma_{rf}ICC_f \cos(\phi(Rf)) + \dots \\ 2P_l(Ls)P_l(Rs)\rho(Rs)\sigma_{ls}\sigma_{rs}ICC_s \cos(\phi(Rs))$$

De manera similar, la potencia (relativa) para el canal derecho se proporciona mediante:

$$15 \quad \sigma_R^2 = P_r^2(C)\sigma_c^2 + P_r^2(Lf)\sigma_{lf}^2 + P_r^2(Ls)\sigma_{ls}^2 + P_r^2(Rf)\sigma_{rf}^2 + P_r^2(Rs)\sigma_{rs}^2 + \dots \\ 2P_r(Lf)P_r(Rf)\rho(Lf)\sigma_{lf}\sigma_{rf}ICC_f \cos(\phi(Lf)) + \dots \\ 20 \quad 2P_r(Ls)P_r(Rs)\rho(Ls)\sigma_{ls}\sigma_{rs}ICC_s \cos(\phi(Ls))$$

**[0102]** En base a presunciones similares y utilizando técnicas similares, el valor esperado para el producto cruzado  $L_B R_B^*$  del par de señales binaurales, puede calcularse a partir de

$$25 \quad \langle L_B R_B^* \rangle = \sigma_c^2 P_l(C)P_r(C)\rho(C)\exp(j\phi(C)) + \dots \\ \sigma_{lf}^2 P_l(Lf)P_r(Lf)\rho(Lf)\exp(j\phi(Lf)) + \dots \\ 30 \quad \sigma_{rf}^2 P_l(Rf)P_r(Rf)\rho(Rf)\exp(j\phi(Rf)) + \dots \\ \sigma_{ls}^2 P_l(Ls)P_r(Ls)\rho(Ls)\exp(j\phi(Ls)) + \dots \\ \sigma_{rs}^2 P_l(Rs)P_r(Rs)\rho(Rs)\exp(j\phi(Rs)) + \dots \\ 35 \quad P_l(Lf)P_r(Rf)\sigma_{lf}\sigma_{rf}ICC_f + \dots \\ P_l(Ls)P_r(Rs)\sigma_{ls}\sigma_{rs}ICC_s + \dots \\ P_l(Rs)P_r(Ls)\sigma_{ls}\sigma_{rs}ICC_s\rho(Ls)\rho(Rs)\exp(j(\phi(Rs) + \phi(Ls))) + \dots \\ 40 \quad P_l(Rf)P_r(Lf)\sigma_{lf}\sigma_{rf}ICC_f\rho(Lf)\rho(Rf)\exp(j(\phi(Rf) + \phi(Lf)))$$

La coherencia de la salida binaural ( $ICC_B$ ) se proporciona entonces mediante:

$$45 \quad ICC_B = \frac{|\langle L_B R_B^* \rangle|}{\sigma_L \sigma_R}$$

50 **[0103]** En base a la coherencia determinada de la señal de salida binaural  $ICC_B$  (e ignorando las claves de localización y las características de reverberación), pueden calcularse los coeficientes de matriz requeridos para reinstalar los parámetros  $ICC_B$  utilizando procedimientos convencionales como se especifica en Breebaart J., van de Par S., Koohrausch A., Schuijers E., "Parametric coding of stereo audio" (Codificación paramétrica de audio en estéreo), EURASIP J. Applied Signal Proc. 9, p. 1305-1322 (2005):

$$55 \quad h_{12} = \text{sen}(\alpha + \beta) \\ 60 \quad \bar{h}_{22} = \text{sen}(-\alpha + \beta)$$

con  $\alpha = 0.5 \arccos(\text{ICC}_B)$

$$\beta = \arctan\left(\frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L} \tan(\alpha)\right)$$

[0104] En lo siguiente se describirá la generación de los coeficientes de filtro mediante el procesador de coeficiente 419.

[0105] Primeramente, se generan las representaciones de sub-banda de las respuestas al impulso de la función de transferencia perceptual binaural correspondiente a diferentes fuentes de sonido en la señal de audio binaural.

[0106] Específicamente, los HRTFs (o BRIRs) se convierten en el dominio QMF dando como resultado las representaciones de QMF-dominio  $H_{L,X}^{n,k}$ ,  $H_{R,X}^{n,k}$  para las respuestas al impulso del oído izquierdo y del oído derecho, respectivamente, utilizando el procedimiento de convertidor de filtro señalado anteriormente en la descripción de la Figura 4. En la representación, X denota el canal de la fuente (X = Lf, Rf, C, Ls, Rs), R y L denotan los canales binaurales izquierdo y derecho, respectivamente, n es el número de bloques de transformación y k denota la sub-banda.

[0107] El procesador de coeficiente 419 procede entonces para determinar los coeficientes del filtro como una combinación ponderada de los coeficientes correspondientes de las representaciones de sub-banda  $H_{L,X}^{n,k}$ ,  $H_{R,X}^{n,k}$ . Específicamente, los coeficientes de filtro para los filtros FIR 415, 417, se proporcionan mediante:

$$H_{L,M}^{n,k} = g_L^k \cdot (t_{Lf}^k H_{L,Lf}^{n,k} + t_{Ls}^k H_{L,Ls}^{n,k} + t_{Rf}^k H_{L,Rf}^{n,k} + t_{Rs}^k H_{L,Rs}^{n,k} + t_C^k H_{L,C}^{n,k})$$

$$H_{R,M}^{n,k} = g_R^k \cdot (s_{Lf}^k H_{R,Lf}^{n,k} + s_{Ls}^k H_{R,Ls}^{n,k} + s_{Rf}^k H_{R,Rf}^{n,k} + s_{Rs}^k H_{R,Rs}^{n,k} + s_C^k H_{R,C}^{n,k})$$

[0108] El procesador de coeficiente 419 calcula los pesos  $t^k$  y  $s^k$  como se describe a continuación.

[0109] Primeramente, el módulo de los pesos de combinación lineal se seleccionan de tal manera que:

$$|t_X^k| = \sigma_X^k, \quad |s_X^k| = \sigma_X^k$$

[0110] Por tanto, el peso para un HRTF dado correspondiente a un canal espacial dado, se selecciona para corresponder al nivel de potencia de ese canal.

[0111] Segundo, se computan las ganancias de escalado  $g_Y^k$ , como sigue.

Se hace notar la potencia de salida binaural objetivo normalizada para la banda híbrida k por  $(\sigma_Y^k)^2$  para el canal de salida  $Y = L, R$ , y se hace notar la ganancia de potencia del filtro  $H_{Y,M}^{n,k}$ , por  $(\sigma_{Y,M}^k)^2$ , después las ganancias de escalado  $g_Y^k$  se ajustan a fin de lograr

$$\sigma_{Y,M}^k = \sigma_Y^k$$

[0112] Nótese aquí que si esto puede lograrse aproximadamente con ganancias de escalado que son constantes en cada banda de parámetro, entonces puede omitirse el escalado de la morfología y llevarse a cabo modificando los elementos de matriz de la sección previa a

$$h_{11} = g_L \cos(\alpha + \beta)$$

$$h_{12} = g_L \sin(\alpha + \beta)$$

$$h_{22} = g_R \sin(-\alpha + \beta)$$

$$h_{21} = g_R \cos(-\alpha + \beta)$$



[0113] Para que esto sea verdadero, se requiere que la combinación ponderada no escalada

$$\begin{aligned}
 & t_{L_f}^k H_{L,L_f}^{n,k} + t_{L_s}^k H_{L,L_s}^{n,k} + t_{R_f}^k H_{L,R_f}^{n,k} + t_{R_s}^k H_{L,R_s}^{n,k} + t_C^k H_{L,C}^{n,k} \\
 & s_{L_f}^k H_{R,L_f}^{n,k} + s_{L_s}^k H_{R,L_s}^{n,k} + s_{R_f}^k H_{R,R_f}^{n,k} + s_{R_s}^k H_{R,R_s}^{n,k} + s_C^k H_{R,C}^{n,k}
 \end{aligned}$$

tenga ganancias de potencia que no varíen demasiado dentro de las bandas de parámetro. Típicamente, surge una contribución principal a tales variaciones a partir de las diferencias de retraso principales entre las respuestas HRTF. En algunas modalidades de la presente invención, se lleva a cabo una pre-alineación en el dominio de tiempo para los filtros HRTF dominantes y pueden aplicarse los pesos simples de la combinación valuados reales:

$$t_X^k = s_X^k = \sigma_X^k$$

[0114] En otras modalidades de la presente invención, esas diferencias de retraso se contraatacan de manera adaptada en los pares HRTF dominantes por medio de la introducción de pesos valuados complejos. En el caso de pares frontales/posteriores, esto asciende al uso de los siguientes pesos:

$$t_{L_f}^k = \sigma_{L_f}^k \exp \left[ -j\phi_{L_f,L_s}^{L,k} \frac{(\sigma_{L_s}^k)^2}{(\sigma_{L_f}^k)^2 + (\sigma_{L_s}^k)^2} \right]$$

$$t_{L_s}^k = \sigma_{L_s}^k \exp \left[ j\phi_{L_f,L_s}^{L,k} \frac{(\sigma_{L_f}^k)^2}{(\sigma_{L_f}^k)^2 + (\sigma_{L_s}^k)^2} \right]$$

y  $t_X^k = \sigma_X^k$  por  $X = C, R_f, R_s$ .

$$s_{R_f}^k = \sigma_{R_f}^k \exp \left[ -j\phi_{R_f,R_s}^{R,k} \frac{(\sigma_{R_s}^k)^2}{(\sigma_{R_f}^k)^2 + (\sigma_{R_s}^k)^2} \right]$$

$$s_{R_s}^k = \sigma_{R_s}^k \exp \left[ j\phi_{R_f,R_s}^{R,k} \frac{(\sigma_{R_f}^k)^2}{(\sigma_{R_f}^k)^2 + (\sigma_{R_s}^k)^2} \right]$$

y  $s_X^k = \sigma_X^k$  por  $X = C, L_f, L_s$ .

[0115] Aquí,  $\phi_{X_f,X_s}^{X,k}$  es el ángulo de fase no envuelta de la compleja correlación cruzada entre los filtros de sub-banda  $H_{X,X_f}^{n,k}$  y  $H_{X,X_s}^{n,k}$ . Esta correlación cruzada se define mediante

$$(CIC)_k = \frac{\sum_n (H_{X,X_f}^{n,k})(H_{X,X_s}^{n,k})^*}{\left( \sum_n |H_{X,X_f}^{n,k}|^2 \right)^{1/2} \left( \sum_n |H_{X,X_s}^{n,k}|^2 \right)^{1/2}}$$

en donde el asterisco denota una conjugación compleja.

[0116] El propósito del desenvolvimiento de fase es utilizar la libertad de selección de un ángulo de fase hasta múltiplos de  $2\pi$  a fin de obtener una curva de fase que varía tan lentamente como es posible como una función del índice de sub-banda  $k$ .

5 **[0117]** El papel de los parámetros de ángulo de fase en la combinación de las fórmulas anteriores es en dos sentidos. Primero, realiza una compensación de retraso de los filtros frontal/posterior previo a su superposición, lo cual conduce a una respuesta combinada que modela un tiempo de retraso correspondiente a una posición de la fuente entre los altavoces frontal y posterior. Segundo, reduce la variabilidad de las ganancias de potencia de los filtros no escalados.

**[0118]** Si la coherencia  $ICC_M$  de los filtros combinados  $H_{L,M}$ ,  $H_{R,M}$  en una banda de parámetro o en una banda híbrida es menor que uno, la salida binaural puede volverse menos coherente que lo pretendido, como sigue a partir de la relación

10  $ICC_{B-fuera} = ICC_M - ICC_B$

**[0119]** La solución a este problema, de acuerdo con algunas modalidades de la presente invención, es utilizar un valor  $ICC_B$  modificado para la definición del elemento de matriz definido mediante

15 
$$ICC'_B = \min\left\{1, \frac{ICC_B}{ICC_M}\right\}$$

20 **[0120]** La Figura 5 ilustra un diagrama de flujo de un ejemplo de un procedimiento para generar una señal de audio binaural de acuerdo con algunas modalidades de la invención.

25 **[0121]** El procedimiento comienza en la etapa 501, en donde se reciben los datos de audio que comprenden una señal de audio de M canales que es una mezcla reductora de una señal de audio de N canales y los datos de parámetros espaciales para el mezclado ascendente de la señal de audio de M canales a la señal de audio de N canales.

**[0122]** A la etapa 501 le sigue la etapa 503, en donde los parámetros espaciales de los datos de parámetros espaciales se convierten en los primeros parámetros binaurales en respuesta a una función de transferencia perceptual binaural.

30 **[0123]** A la etapa 503 le sigue la etapa 505, en donde la señal de audio de M canales se convierte en una primera señal estéreo en respuesta a los primeros parámetros binaurales.

**[0124]** A la etapa 505 le sigue la etapa 507, en donde se determinan los coeficientes de filtro para un filtro estéreo en respuesta a la función de transferencia perceptual binaural.

35 **[0125]** A la etapa 507 le sigue la etapa 509, en donde la señal de audio binaural se genera filtrando la primera señal estéreo en el filtro de estéreo.

40 **[0126]** El aparato de la Figura 4, por ejemplo, puede utilizarse en un sistema de transmisión. La Figura 6 ilustra un ejemplo de un sistema de transmisión para la comunicación de una señal de audio de acuerdo con algunas modalidades de la invención. El sistema de transmisión comprende un transmisor 601 que se encuentra acoplado a un receptor 603 a través de una red 605 que puede ser, específicamente, la Internet.

45 **[0127]** En el ejemplo específico, el transmisor 601 es un dispositivo de grabación de señal y el receptor 603 es un dispositivo reproductor de señal, pero se apreciará que, en otras modalidades, pueden utilizarse un transmisor y un receptor en otras aplicaciones y para otros propósitos. Por ejemplo, el transmisor 601 y/o el receptor 603, pueden ser parte de una funcionalidad de transcodificación y por ejemplo, pueden proporcionar interfaz a otras fuentes o destinos de la señal. Específicamente, el receptor 603 puede recibir una señal de sonido envolvente codificada y generar una señal binaural codificada que emula la señal de sonido envolvente. La señal binaural codificada puede distribuirse entonces a otras fuentes.

50 **[0128]** En el ejemplo específico en el que se soporta una función de grabación de señal, el transmisor 601 comprende un digitalizador 607 que recibe una señal multicanal (envolvente) analógica que se convierte en una señal PCM digital (modulada por códigos de impulso) muestreando una conversión de analógica-a-digital.

55 **[0129]** El digitalizador 607 se encuentra acoplado al codificador 609 de la Figura 1, que codifica la señal multicanal PCM de acuerdo con un algoritmo de codificación. En el ejemplo específico, el codificador 609 codifica la señal como una señal MPEG de sonido envolvente codificada. El codificador 609 se encuentra acoplado a un transmisor de red 611 que recibe la señal codificada y se interfaza a la Internet 605. El transmisor de red puede transmitir la señal codificada al receptor 603 a través de la Internet 605.

[0130] El receptor 603 comprende un receptor de red 613 que se interfasa a la Internet 605 y que se encuentra dispuesto para recibir la señal codificada desde el transmisor 601.

5 [0131] El receptor de red 613 se encuentra acoplado a un decodificador binaural 615 que, en el ejemplo, es el dispositivo de la Figura 4.

10 [0132] En el ejemplo específico en el cual se soporta una función de reproducción de señal, el receptor 603 comprende además un reproductor de señal 1617 que recibe la señal de audio binaural desde el decodificador binaural 615 y la presenta al usuario. Específicamente, el reproductor de señal 117 puede comprender un convertidor de digital-a-análogo, amplificadores y altavoces para dar salida a la señal de audio binaural hacia un juego de audífonos.

15 [0133] Se apreciará que la descripción anterior, por claridad, ha descrito las modalidades de la invención con referencia a diferentes unidades y procesadores funcionales. Sin embargo, será aparente que puede utilizarse cualquier distribución de funcionalidad adecuada entre las diferentes unidades o procesadores funcionales sin detracción de la invención. Por ejemplo, la funcionalidad ilustrada para llevarse a cabo mediante procesadores o controladores separados, puede efectuarse por el mismo procesador o controladores. De aquí que, las referencias a las unidades funcionales específicas se contemplan solamente como referencias a los medios adecuados para proporcionar la funcionalidad descrita más que indicativas de una estructura u organización lógica o física estricta.

20 [0134] La invención puede implementarse en cualquier forma adecuada incluyendo hardware, software, firmware o cualquier combinación de éstos. La invención puede implementarse opcionalmente, al menos parcialmente, como un software de computadora operando en uno o más procesadores de datos y/o procesadores de señal digital. Los elementos y componentes de una modalidad de la invención pueden implementarse físicamente, funcionalmente y lógicamente en cualquier manera adecuada. De hecho, la funcionalidad puede implementarse en una sola unidad, en 25 una pluralidad de unidades o como parte de otras unidades funcionales. Como tal, la invención puede implementarse en una sola unidad o puede distribuirse físicamente y funcionalmente entre diferentes unidades y procesadores.

30 [0135] Aunque la presente invención se ha descrito en conexión con algunas modalidades, ésta no pretende limitarse a la forma específica expuestas en la presente. Por el contrario, el alcance de la presente invención se limita solamente por las reivindicaciones anexas. Adicionalmente, aunque puede parecer que una característica se describe en conexión con las modalidades particulares, el experto en la técnica reconocería que varias características de las modalidades descritas pueden combinarse de acuerdo con la invención. En las reivindicaciones, el término, que comprende, no excluye la presencia de otros elementos o etapas.

35 [0136] Además, aunque se listan individualmente, pueden implementarse una pluralidad de medios, elementos o etapas del procedimiento, por ejemplo, por medio de una sola unidad o procesador. Adicionalmente, aunque pueden incluirse características individuales en las diferentes reivindicaciones, ventajosamente éstas posiblemente pueden combinarse y la inclusión en las diferentes reivindicaciones no implica que no sea posible y/o ventajosa una combinación de características. También, la inclusión de una característica en una categoría de reivindicaciones no implica una limitación a esta categoría, sino por el contrario, indica que la característica es igualmente aplicable a otras categorías de 40 reivindicación, según sea apropiado. Además, el orden de las características en las reivindicaciones no implica ningún orden específico en el cual deban operarse las características y, en particular, el orden de las etapas individuales en una reivindicación del procedimiento no implica que las etapas deban llevarse a cabo en este orden. Por el contrario, las etapas pueden llevarse a cabo en cualquier orden adecuado. Además, las referencias en singular no excluyen la pluralidad. Por tanto, las referencias a "un" "una", "primero(a)", "segundo(a)", etc., no excluyen la pluralidad. Los signos de referencia en las reivindicaciones se proporcionan meramente como un ejemplo de aclaración y no deben interpretarse como limitantes del alcance de las reivindicaciones en modo alguno.

45

**REIVINDICACIONES**

1. Aparato para generar una señal de audio binaural, comprendiendo el aparato:

- 5 - un medio (401, 403) destinado a recibir datos de audio que comprenden una señal de audio con M canales que es una mezcla descendente de una señal de audio de N canales y datos de parámetros espaciales para la mezcla ascendente de la señal de audio con M canales, para obtener la señal de audio de N canales;
- un medio de datos de parámetros (411) destinado a convertir parámetros espaciales de los datos de parámetros espaciales, para obtener unos primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptiva binaural;
- 10 - un medio de conversión (409) destinado a convertir la señal de audio con M canales en una primera señal estéreo en respuesta a los primeros parámetros binaurales;

**caracterizado por**

- 15 - un filtro estéreo (415, 417) destinado a generar la señal de audio binaural por filtrado de la primera señal estéreo, y un medio de coeficientes (419) destinado a determinar coeficientes de filtro para el filtro estéreo en respuesta a la función de transferencia perceptiva binaural.

20 2. Aparato según la reivindicación 1, que comprende por otro lado:

- un medio de transformación (405) destinado a transformar la señal de audio con M canales de un ámbito temporal a un ámbito de sub-banda y donde el medio de conversión y el filtro estéreo están dispuestos para tratar individualmente cada sub-banda del ámbito de sub-banda.

25 3. Aparato según la reivindicación 2, en el cual una duración de una respuesta impulsional de la función de transferencia perceptiva binaural excede un intervalo de actualización de transformación.

30 4. Aparato según la reivindicación 2, en el cual el medio de conversión (409) está dispuesto para generar, para cada sub-banda, muestras de salida estéreo sensiblemente como:

$$\begin{bmatrix} L_o \\ R_o \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} L_l \\ R_l \end{bmatrix}$$

35 donde al menos una de entre  $L_l$  y  $R_l$  es una muestra de un canal audio de la señal de audio con M canales en la sub-banda y el medio de conversión está dispuesto para determinar coeficientes de matriz  $h_{xy}$  en respuesta tanto a los datos de parámetros espaciales como a la al menos una función de transferencia perceptiva binaural.

40 5. Aparato según la reivindicación 2, en el cual el medio de coeficientes (419) comprende:

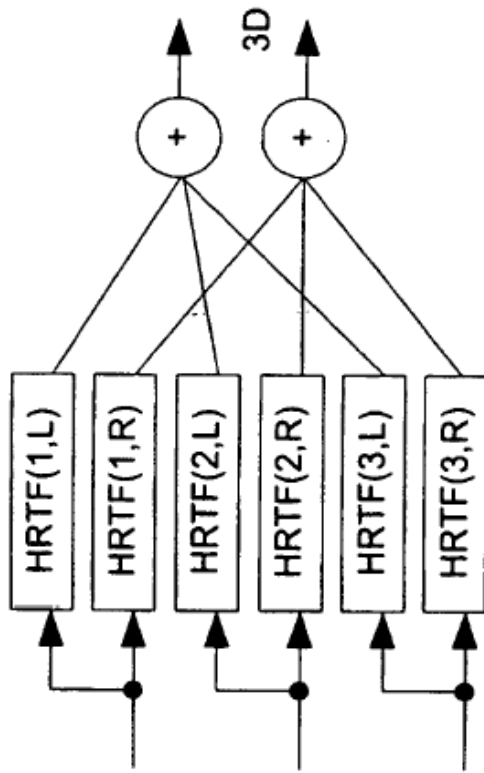
- un medio destinado a suministrar representaciones de sub-banda de respuestas impulsionales de una pluralidad de funciones de transferencia perceptiva binaural correspondientes a unas fuentes sonoras diferentes en la señal de N canales;
- 45 - un medio destinado a determinar los coeficientes de filtro por una combinación ponderada de coeficientes correspondientes de las representaciones de sub-banda; y
- un medio destinado a determinar los pesos de las representaciones de sub-banda para la combinación ponderada en respuesta a los datos de parámetros espaciales.

50 6. Aparato según la reivindicación 1, en el cual los primeros parámetros binaurales comprenden unos parámetros de coherencia que indican una correlación entre los canales de la señal de audio binaural.

55 7. Aparato según la reivindicación 1, en el cual los primeros parámetros binaurales no comprenden al menos uno de entre parámetros de localización que indican un emplazamiento de una fuente sonora cualquiera de la señal de N canales y de los parámetros de reverberación que indican una reverberación de una componente de sonido cualquiera de la señal de audio binaural.

60 8. Aparato según la reivindicación 1, en el cual el medio de coeficientes (419) está dispuesto para determinar los coeficientes de filtro para reflejar al menos una de entre las referencias de localización y las referencias de reverberación para la señal de audio binaural.

- 5 9. Aparato según la reivindicación 1, en el cual la señal de audio con M canales audio es una señal de audio mono y el medio de conversión (407, 409) está dispuesto para generar una señal descorrelacionada a partir de la señal de audio mono y para generar la primera señal estéreo por una multiplicación matricial aplicada a unas muestras de una señal estéreo que comprende la señal descorrelacionada y la señal de audio mono.
10. Procedimiento de generación de una señal de audio binaural, comprendiendo el procedimiento que comprende
- recibir (501) datos de audio que comprenden una señal de audio con M canales que es una mezcla descendente de una señal de audio de N canales y datos de parámetros espaciales para la mezcla ascendente de la señal de audio con
  - 10 M canales, para obtener la señal de audio de N canales;
  - convertir (503) los parámetros espaciales de los datos de parámetros espaciales, para obtener los primeros parámetros binaurales en respuesta a al menos una función de transferencia perceptiva binaural;
  - convertir (505) la señal de audio con M canales en una primera señal estéreo en respuesta a los primeros parámetros binaurales;
  - 15 caracterizado por el hecho de
  - generar (509) la señal de audio binaural por filtrado de la primera señal estéreo; y
  - determinar (507) los coeficientes de filtro para el filtro estéreo en respuesta a la al menos una función de transferencia perceptiva binaural.
  - 20
11. Emisor para transmitir una señal de audio binaural, comprendiendo el emisor un aparato para generar una señal de audio binaural según la reivindicación 1.
- 25 12. Sistema de transmisión para transmitir una señal de audio, comprendiendo el sistema de transmisión
- un emisor que comprende un emisor para transmitir la señal de audio binaural según la reivindicación 11, y
  - un receptor para recibir la señal de audio binaural.
- 30 13. dispositivo de grabación audio para grabar una señal de audio binaural, comprendiendo el dispositivo de grabación audio un aparato para generar una señal de audio binaural según la reivindicación 1.
14. Procedimiento de transmisión de una señal de audio binaural, comprendiendo el procedimiento un procedimiento de generación de una señal de audio binaural según la reivindicación 10.
- 35 15. Procedimiento de emisión y de recepción de una señal de audio binaural, comprendiendo el procedimiento: un emisor que realiza un procedimiento de transmisión de señal de audio binaural según la reivindicación 10; y un receptor que realiza la etapa consistente en recibir la señal de audio binaural.
- 40 16. Producto de programa de ordenador para realizar el procedimiento según cualquiera de las reivindicaciones 14 y 15.



Estado de la técnica

FIG. 1

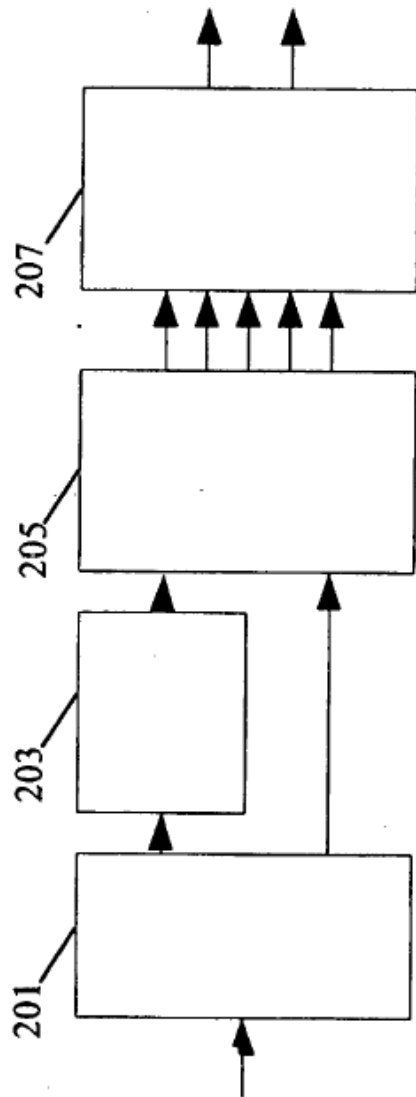


FIG. 2

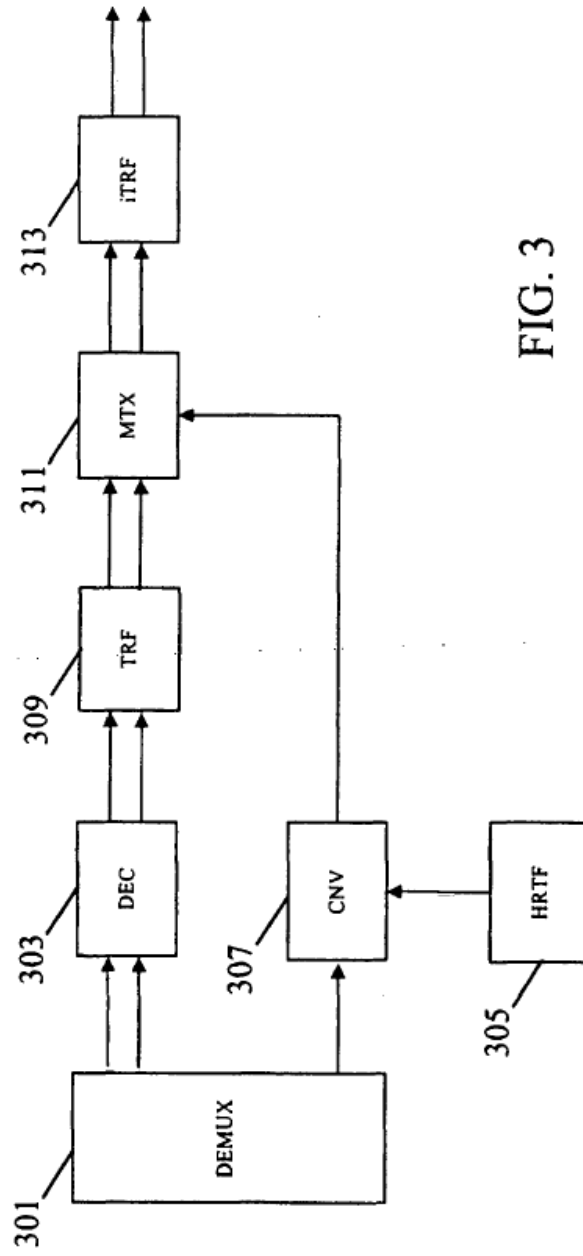


FIG. 3



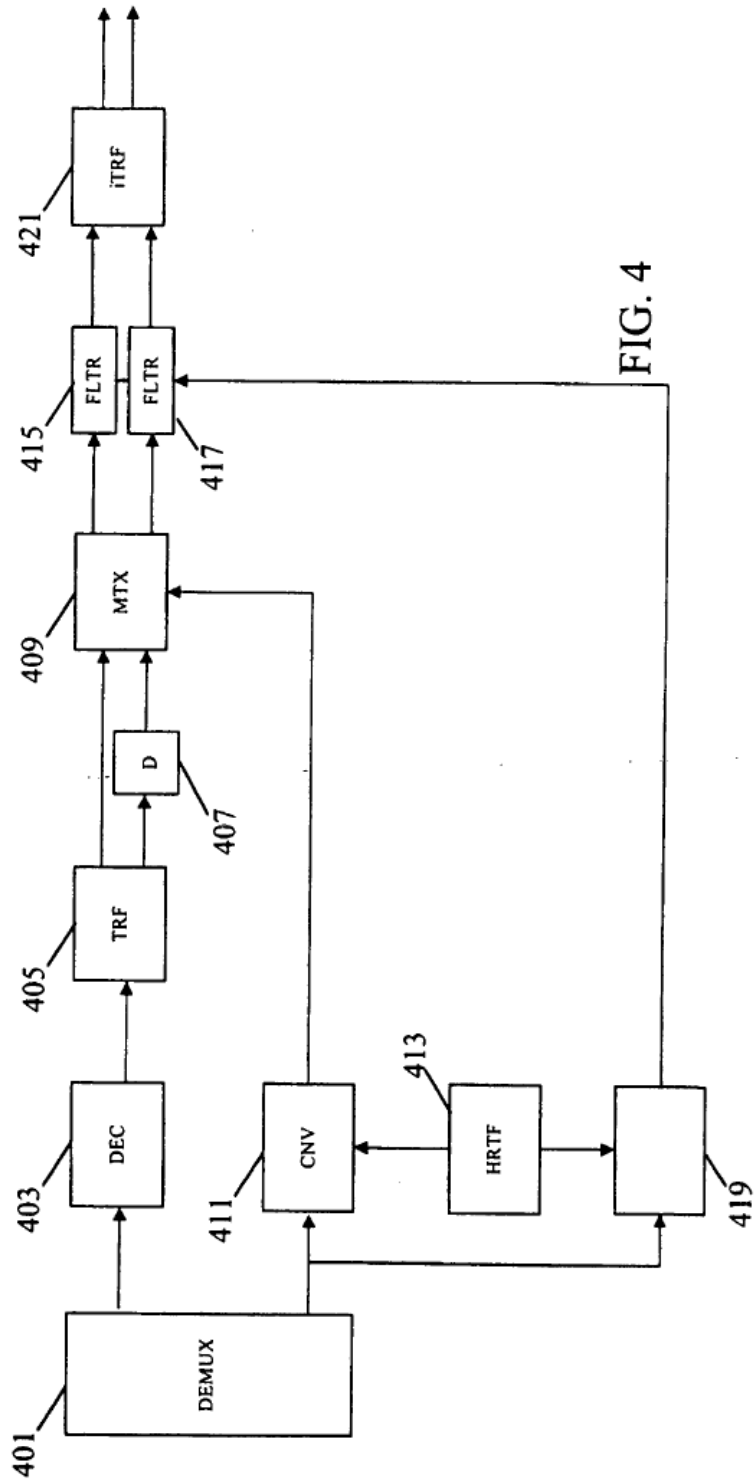


FIG. 4

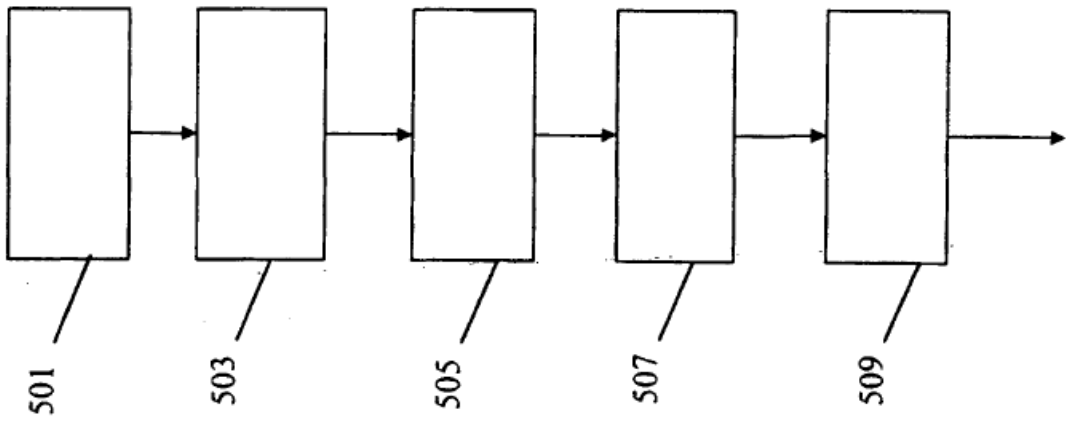


FIG. 5

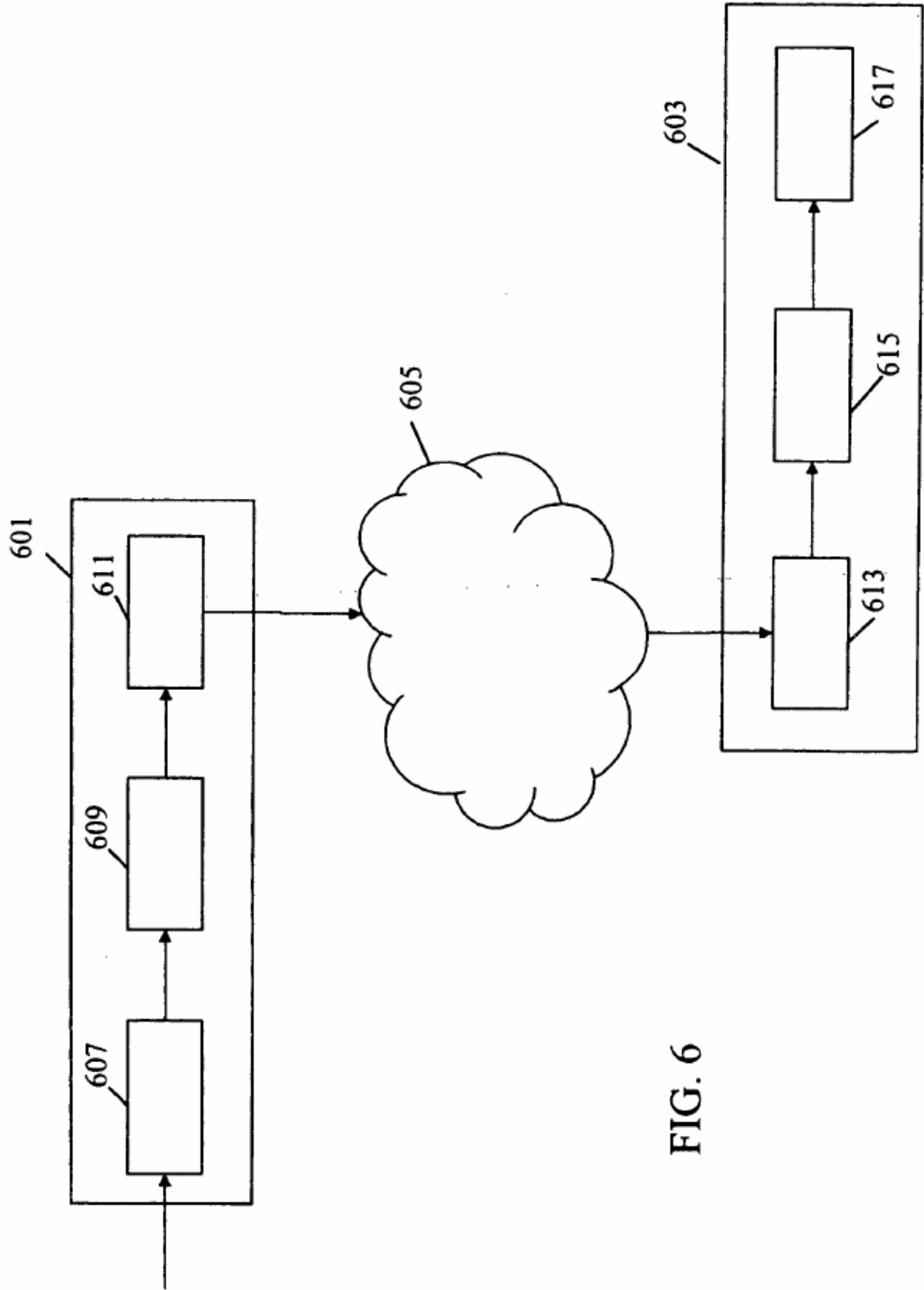


FIG. 6