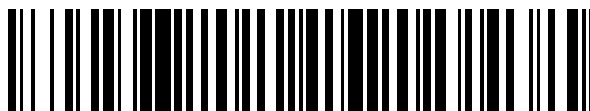


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 532 152**

51 Int. Cl.:

H04S 3/00 (2006.01)

H04S 1/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.09.2009 E 09778738 (6)**

97 Fecha y número de publicación de la concesión europea: **14.01.2015 EP 2335428**

54 Título: **Renderización binaural de una señal de audio multicanal**

30 Prioridad:

07.10.2008 US 103303 P
15.05.2009 EP 09006598

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
24.03.2015

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (33.3%)**
Hansastrasse 27c
80686 München , DE;
KONINKLIJKE PHILIPS N.V. (33.3%) y
DOLBY INTERNATIONAL AB (33.3%)

72 Inventor/es:

KOPPENS, JEROEN;
MUNDT, HARALD;
TERENTIEV, LEONID;
FALCH, CORNELIA;
HILPERT, JOHANNES;
HELLMUTH, OLIVER;
VILLEMOES, LARS;
PLOGSTIES, JAN;
BREEBAART, JEROEN y
ENGDEGARD, JONAS

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 532 152 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

Renderización binaural de una señal de audio multicanal

5 **[0001]** La presente solicitud se relaciona con la renderización binaural de una señal de audio multicanal.

10 **[0002]** Muchos algoritmos de codificación de audio han sido propuestos a fin de codificar con eficacia o comprimir datos de audio de un canal, es decir, señales monoaurales. Usando la psicoacústica, las muestras de audio son apropiadamente modificadas a escala, cuantificadas o conjunto ni siquiera al cero a fin de retirar la irrelevancia de, por ejemplo, el PCM señal de audio codificada. La eliminación de redundancia también es llevada a cabo.

[0003] Como una etapa adicional, las semejanzas entre el canal derecho e izquierdo de señales de audio estéreo han sido explotadas a fin de codificar/comprimir con eficacia señales de audio estéreo.

15 **[0004]** Sin embargo, las aplicaciones próximas plantean demandas adicionales en algoritmos de codificación de audio. Por ejemplo, en teleconferencia, videojuegos, desempeño de música y lo similar, varias señales de audio que son parcialmente o incluso completamente no correlacionadas tienen que transmitirse en paralelo. A fin de mantener la velocidad binaria necesaria para codificar estas señales de audio bastante bajo a fin de ser compatible a aplicaciones de transmisión de tasa con el bajo bit, recientemente, códecs de audio ha sido propuesta que
20 conversión de multicanal a estéreo múltiples señales de audio de entrada en una señal de conversión de multicanal a estéreo, como una señal de conversión monoaural de multicanal a estéreo estéreo o ni siquiera. Por ejemplo, las conversiones de multicanal a estéreo de norma de Sonido envolvente de MPEG los canales de entrada en la señal de conversión de multicanal a estéreo en una manera prescrita por norma. La conversión de multicanal a estéreo se lleva a cabo por el uso de llamados bloques de TTT^{-1} y OTT^{-1} para la conversión de multicanal a estéreo dos señales
25 en una y tres señales en dos, respectivamente. A fin de a la conversión de multicanal a estéreo más de tres señales, una estructura jerárquica de estos bloques se usa. Cada bloque de OTT^{-1} salidas, además de la señal de conversión monoaural de multicanal a estéreo, diferencias de nivel de canal entre los dos canales de entrada, así como coherencia entre canales / parámetros de correlación cruzada que representan la coherencia o la correlación cruzada entre los dos canales de entrada. Los parámetros son enviados junto con la señal de conversión de multicanal a estéreo del codificador de Sonido envolvente de MPEG dentro del tren de datos de Sonido envolvente de MPEG. De forma similar, cada bloque de TTT^{-1} transmite coeficientes de predicción de canal que habilitan recuperar los tres canales de entrada de la señal de conversión estereofónica de multicanal a estéreo resultante. Los coeficientes de predicción de canal también son transmitidos como la información conexa dentro del tren de datos de Sonido envolvente de MPEG. Las conversiones de estéreo a multicanal de Decodificador de sonido envolvente de MPEG la señal de conversión de multicanal a estéreo por el uso de la información conexa transmitida y recuperan, la
35 entrada de canales original en el codificador de Sonido envolvente de MPEG.

[0005] Sin embargo, el Sonido envolvente de MPEG, lamentablemente, no realiza todos los requerimientos planteados por muchas aplicaciones. Por ejemplo, el Decodificador de sonido envolvente de MPEG es dedicado
40 para convertir de estéreo a multicanal la señal de conversión de multicanal a estéreo del codificador de Sonido envolvente de MPEG el que los canales de entrada del codificador de Sonido envolvente de MPEG son recuperados como éstos son. En otras palabras, el tren de datos de Sonido envolvente de MPEG es dedicado para ser puesto por el uso de la configuración de altavoz que se ha usado para la codificación, o por configuraciones comunes como el estéreo.

45 **[0006]** Sin embargo, según algunas aplicaciones, sería favorable si la configuración de altavoz pudiera cambiarse en el lado del decodificador libremente.

[0007] A fin de encargarse a las últimas necesidades, la codificación espacial del objeto de audio (SAOC) la norma es diseñada actualmente. Cada canal se somete a tratamiento como un objeto individual, y todos los objetos se convierten de multicanal a estéreo en una señal de conversión de multicanal a estéreo. Es decir los objetos son manejados como señales de audio que son independientes de entre sí sin adherirse a cualquier configuración de altavoz específica, pero con la capacidad de colocar los altavoces (virtuales) en el lado del decodificador arbitrariamente. Los objetos individuales pueden comprender fuentes de sonido individuales como p.ej documentos
55 o pistas vocales. Diferenciándose del Decodificador de sonido envolvente de MPEG, el decodificador SAOC está libre para convertir de estéreo a multicanal individualmente la señal de conversión de multicanal a estéreo para volver a jugar los objetos individuales en cualquier configuración de altavoz. A fin de habilitar el decodificador SAOC para recuperar los objetos individuales que tienen sido codificado en el tren de datos SAOC, objetan diferencias de nivel y, para objetos que forman conjuntamente un estéreo (o multicanal) señal, los parámetros de correlación cruzados entre objetos se transmiten como la información conexa dentro del tren de bits SAOC. Además de esto, el decodificador/transcodificador SAOC se proporciona con la información que revela como los objetos individuales se han convertido de multicanal a estéreo en la señal de conversión de multicanal a estéreo. Así, en el lado del decodificador, es posible recuperar los canales SAOC individuales y renderizar estas señales en cualquier configuración de altavoz utilizando la información de renderización controlada por el usuario.

65

[0008] Sin embargo, aunque los códecs ya mencionados, es decir Sonido envolvente de MPEG y SAOC, sean capaces de transmitir y renderizar el contenido de audio multicanal en configuraciones de altavoz que tienen a más de dos altavoces, el interés creciente en audífonos ya que el sistema de reproducción de audio requiere esto estos códecs también son capaces de renderizar el contenido de audio en audífonos. En contraste con el pregrabado de altavoz, el contenido de audio estéreo reproducido sobre audífonos es percibido dentro el cabezal. La ausencia del efecto de la vía acústica de fuentes las a ciertas posiciones físicas a los tímpanos hacen que la imagen espacial parezca no natural ya que las señales de aviso que determinan el acimut percibido, elevación y distancia de una fuente de sonido es esencialmente ausente o muy inexacta. Así, para resolver la etapa de sonido no natural causada por señales de aviso de localización de fuente de sonido inexactas o ausentes en audífonos, diversos métodos han sido propuestos para simular una configuración de altavoz virtual. La idea es superimponer señales de aviso de localización de fuente de sonido en cada señal de altavoz. Esto se logra filtrando señales de audio con llamadas funciones de transferencia relacionadas por el cabezal (HRTFs) o respuestas a impulsos de cuarto binaurales (BRIRs) si el cuarto propiedades acústicas se incluye en estos datos de cuantificación. Sin embargo, filtrar cada señal de altavoz con las funciones sólo mencionadas requeriría una cantidad considerablemente más elevada de la energía de cálculo en el lado de decodificador/reproducción. En términos particulares, la renderización la señal de audio multicanal en las ubicaciones de altavoz "virtuales" tendría que llevarse a cabo primero donde, luego, cada señal de altavoz así obtenida, es filtrada con la respectiva función de transferencia o respuesta a impulsos para obtener el canal derecho e izquierdo de la señal de salida binaural. Incluso peor: la señal de salida binaural obtenida tendría una calidad de audio mala debido a que a fin de lograr las señales de altavoz virtuales, un relativamente la gran cantidad de señales de decorrelación sintéticas tendría que mezclarse en las señales convertidas de estéreo a multicanal a fin de compensar la correlación entre señales de entrada de audio al principio no correlacionadas, la correlación que resulta a partir de la conversión de multicanal a estéreo la pluralidad de señales de entrada de audio en la señal de conversión de multicanal a estéreo.

[0009] En la versión actual del códec SAOC, los parámetros SAOC dentro de la información conexa permiten la renderización espacial interactiva por el usuario de los objetos de audio usando cualquier configuración de pregrabado con, en principio, incluyendo audífonos. La renderización binaural a audífonos permite el control espacial de posiciones de objeto virtuales en el espacio en 3D usando la función de transferencia relacionada por el cabezal (HRTF) parámetros. Por ejemplo, la renderización binaural en SAOC podría implementarse restringiendo este caso a la conversión monoaural de multicanal a estéreo caso de SAOC donde las señales de entrada se mezclan en el canal mono igualmente. Lamentablemente, la conversión monoaural de multicanal a estéreo requiere todas las señales de audio para mezclarse en una señal de conversión monoaural de multicanal a estéreo común de modo que las propiedades de correlación originales entre las señales de audio originales sean máximamente perdidas y por lo tanto, la calidad de renderización de la señal de salida de renderización binaural es no óptima.

[00010] En Engdegard et al. : Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Coding, 124ª convención AES, artículo 7377, 17 de mayo de 2008, páginas 1-15, se describe SAOC como una extensión de la norma de sonido envolvente de MPEG. En particular, se describe el modelo de la arquitectura de referencia elegido, la asociación entre los diferentes modos y aplicaciones operativas, y el estado del proceso de normalización de 2008.

[00011] El documento WO 2008/069593 A1 describe un método para procesar una señal de audio que comprende recibir una señal de mezcla descendente, una primera información multicanal y una información de objeto, el procesamiento de la señal de mezcla descendente utilizando la información de objeto y una información de la mezcla, y la transmisión de una de la primera información multicanal y una segunda información de multicanal de acuerdo con la información de la mezcla, en el que la segunda información de canal se genera utilizando la información del objeto y la información de mezcla.

[00012] El documento WO 2007/078254 A2 describe una decodificación personalizada de sonido envolvente multicanal. Un flujo de bits de audio envolvente multicanal paramétrico se recibe en un decodificador multicanal. Los parámetros espaciales recibidos se transforman en un nuevo conjunto de parámetros espaciales que se utilizan con el fin de obtener una decodificación del sonido envolvente multicanal que no es un simple equivalente de la señal envolvente multicanal de entrada original, pero por ejemplo, pueden ser personalizados al hacer la transformación basada en la representación de los filtros relacionados cabeza del usuario. Tales parámetros espaciales personalizados se pueden obtener mediante la combinación de los parámetros espaciales recibidos y la representación del filtro relacionado con el cabezal de usuario con un conjunto de parámetros de representación adicionales que, por ejemplo, son determinados de forma interactiva por el usuario y por lo tanto, son dependientes del tiempo.

[00013] El documento WO 2007/083952 A1 describe un método y un aparato para procesar una señal de medios de comunicación, por el cual la señal de los medios de comunicación se puede convertir en una señal de sonido envolvente mediante el uso de la información espacial de la señal de los medios de comunicación. La información de asignación fuente correspondiente a cada fuente de múltiples fuentes se genera mediante el uso de la información espacial que indica características entre las múltiples fuentes. Al menos una información de renderizado se genera mediante el uso de la información de asignación de fuente y la información de filtro que tiene un efecto de sonido

envolvente. El suavizado se realiza mediante el uso de información de renderizado de vecinos de la al menos una información de renderizado.

[00014] Así, es el objeto de la presente invención para proporcionar esquema de la renderización binaural de una señal de audio multicanal el que el resultado de renderización binaural es mejorado con, simultáneamente, evitando una restricción en la libertad de componer la señal de conversión de multicanal a estéreo de las señales de audio originales.

[00015] Este objeto se logra por un aparato según la reivindicación 1 y un método según la reivindicación 10.

[00016] Una de las ideas básicas que son la base de la presente invención es que la renderización binaural de una señal de audio multicanal inicial de una señal de conversión estereofónica de multicanal a estéreo es ventajosa sobre la renderización binaural inicial de la señal de audio multicanal de una señal de conversión monoaural de multicanal a estéreo de lo mismo en aquella, debido a que pocos objetos se encuentran en los canales individuales de la señal de conversión estereofónica de multicanal a estéreo, la cantidad de la decorrelación entre las señales de audio individuales es mejor conservada, y en aquella la posibilidad de elegir entre los dos canales de la señal de conversión estereofónica de multicanal a estéreo en el lado codificador habilita esto las propiedades de correlación entre señales de audio en diferentes canales de conversión de multicanal a estéreo son parcialmente conservadas. En otras palabras, debido a la conversión de multicanal a estéreo codificador, las coherencias entre objetos son degradadas que tiene para ser representado en el lado decodificador donde la coherencia entre canales de la señal de salida binaural es una medida importante para la percepción de la anchura de fuente de sonido virtual, pero usando la conversión estereofónica de multicanal a estéreo en vez de la conversión monoaural de multicanal a estéreo reduce la cantidad de degradarse de modo que la restauración/generación de la cantidad apropiada de la coherencia entre canales por la renderización binaural la señal de conversión estereofónica de multicanal a estéreo logre la mejor calidad.

[00017] Una idea principal adicional de la presente solicitud consiste en que ICC ya mencionado (ICC = coherencia entre canales) control puede lograrse por medio de una señal no correlacionada que forma un equivalente perceptual con una conversión monoaural de multicanal a estéreo de los canales de conversión de multicanal a estéreo de la anchura de señal de conversión estereofónica de multicanal a estéreo, sin embargo, no correlacionado a la conversión monoaural de multicanal a estéreo. Así, mientras el uso de una señal de conversión estereofónica de multicanal a estéreo en vez de una señal de conversión monoaural de multicanal a estéreo conserva algunas propiedades de correlación de la pluralidad de señales de audio, que se habrían perdido al usar una señal de conversión monoaural de multicanal a estéreo, la renderización binaural puede basarse en una señal no correlacionada que es representativa para ambos, el primero y segundo canal de conversión de multicanal a estéreo, reduciendo así la cantidad de decorrelaciones o procesamiento de señal sintético comparado con decorrelacionar por separado cada canal de conversión estereofónica de multicanal a estéreo.

[00018] Respecto a las figuras, las modalidades preferidas de la presente solicitud se describen más detalladamente. Entre estas figuras, Figura 1 muestra un diagrama de bloques de una configuración de codificador/decodificador SAOC donde las modalidades de la presente invención pueden ponerse en práctica; Figura 2 muestra un diagrama esquemático e ilustrativo de una representación espectral de una señal monoaural; Figura 3 muestra un diagrama de bloques de un decodificador de audio capaz de la renderización binaural según una modalidad de la presente invención; Figura 4 muestra un diagrama de bloques de la conversión de multicanal a estéreo el bloque pre de procesamiento de la Figura 3 según una modalidad de la presente invención; Figura 5 muestra una flujograma de etapas llevadas a cabo por la unidad de procesamiento de datos de parámetro SAOC 42 de la Figura 3 según una primera alternativa; y Figura 6 muestra un gráfico que ilustra los resultados de pruebas de capacidad auditiva.

[00019] Antes de que las modalidades de la presente invención se describan más detalladamente después, el códec SAOC y los parámetros SAOC transmitidos en un tren binario SAOC se presentan a fin de aliviar la comprensión de las modalidades específicas detalladas con mayor detalle después.

[00020] La Figura 1 muestra una configuración general de un codificador SAOC 10 y un decodificador SAOC 12. El codificador SAOC 10 recibe como una entrada N objetos, es decir, señales de audio 14_1 a 14_N . En términos particulares, el codificador 10 comprende un convertidor de multicanal a estéreo 16 que recibe las señales de audio 14_1 a 14_N y conversiones de multicanal a estéreo mismo a una señal de conversión de multicanal a estéreo 18. En la Figura 1, la señal de conversión de multicanal a estéreo es ejemplarmente mostrada como una señal de conversión estereofónica de multicanal a estéreo. Sin embargo, el codificador 10 y decodificador 12 puede ser capaz de funcionar en un modo mono también donde el caso la señal de conversión de multicanal a estéreo sería una señal de conversión monoaural de multicanal a estéreo. La siguiente descripción, sin embargo, se concentra en el caso de conversión estereofónica de multicanal a estéreo. Los canales de la señal 18 de conversión estereofónica de multicanal a estéreo son LO denotado y RO.

[00021] A fin de habilitar el decodificador SAOC 12 para recuperar el paciente se opone 14₁ a 14_N, el convertidor de multicanal a estéreo 16 proporciona el decodificador SAOC 12 por la información conexa que incluye SAOC-parámetros que incluyen diferencias de nivel de objeto parámetros de correlación cruzados (ANTERIORES), entre objetos (IOC), valores de ganancias de conversión de multicanal a estéreo (DMG) y diferencias de nivel de canal de conversión de multicanal a estéreo (DCLD). La información conexa 20 incluyendo los SAOC-parámetros, junto con la señal de conversión de multicanal a estéreo 18, forma la transmisión de datos de salida SAOC 21 recibido por el decodificador SAOC 12.

[00022] El decodificador SAOC 12 comprende una conversión de estéreo a multicanal 22 que recibe la señal de conversión de multicanal a estéreo 18 así como la información conexa 20 a fin de recuperar y renderizar las señales de audio 14₁ y 14_N en cualquier conjunto seleccionado por el usuario de canales 24₁ a 24_{M'}, con la renderización como prescrito por la información de renderización 26 entrada en el decodificador SAOC 12 así como parámetros HRTF 27 el significado de que se describe más detalladamente después. La siguiente descripción se concentra en la renderización binaural, donde M' = 2 y, la señal de salida es sobre todo dedicada para la reproducción de audífonos, aunque la decodificación 12 pueda ser capaz de renderizar en otra configuración de altavoz (no binaural) también, según órdenes dentro de la introducción de datos por el usuario 26.

[00023] Las señales de audio 14₁ a 14_N pueden ser introducidas en el convertidor de multicanal a estéreo 16 en cualquier dominio de codificación, tal como, por ejemplo, a tiempo o dominio espectral. Por si, las señales de audio 14₁ a 14_N se alimenten en el convertidor de multicanal a estéreo 16 en el dominio temporal, como el PCM codificado, convertidor de multicanal a estéreo 16 usos un banco de filtros, como un banco de QMF híbrido, p.ej, un banco del complejo los filtros exponencialmente modulados con una extensión de filtro de Nyquist para las bandas de frecuencia más bajas que aumentan la resolución de la frecuencia en esa parte, a fin de transferir las señales en el dominio espectral donde las señales de audio son representadas en varias subbandas asociadas con diferentes porciones espectrales, a una resolución de banco de filtros específica. Si las señales de audio 14₁ a 14_N están ya en la representación prevista por el convertidor de multicanal a estéreo 16, mismo no tiene que llevar a cabo la decomposición espectral.

[00024] La Figura 2 muestra una señal de audio en el dominio espectral sólo mencionado. Como se puede observar, la señal de audio es representada como una pluralidad de señales de subbanda. Cada señal de subbanda 30₁ a 30_P comprende una secuencia de valores de la subbanda indicados por los pequeños bloques 32. Como se puede observar, los valores de la subbanda 32 de las señales de subbanda 30₁ a 30_P son sincronizadas entre sí a tiempo de modo que para cada uno de intervalos temporales de banco de filtros consecutivos 34, cada subbanda 30₁ a 30_P comprenda exacto un valor de la subbanda 32. Como se ilustra por el eje de frecuencia 35, las señales de subbanda 30₁ a 30_P se asocian con diferentes regiones de frecuencia, y como se ilustra eje 37, los intervalos temporales de banco de filtros 34 son consecutivamente configurados a tiempo.

[00025] Como detallado antes, el convertidor de multicanal a estéreo 16 calcula SAOC-parámetros de las señales de audio de entrada 14₁ a 14_N. El convertidor de multicanal a estéreo 16 lleva a cabo este cálculo en una resolución de tiempo/frecuencia que puede ser disminuida con relación a la resolución de tiempo/frecuencia original como se determina mediante los intervalos temporales de banco de filtros 34 y decomposición de subbanda, por una cierta cantidad, donde esta cierta cantidad puede ser señalada al lado decodificador dentro de la información conexa 20 por respectivos elementos sintácticos bsFrameLength y bsFreqRes. Por ejemplo, los grupos de intervalos temporales de banco de filtros consecutivos 34 pueden formar un cuadro 36, respectivamente. En otras palabras, la señal de audio puede estar dividida en cuadros que se superponen a tiempo o que es inmediatamente adyacente a tiempo, por ejemplo. En este caso, el bsFrameLength puede definir la cantidad de intervalos temporales de parámetro 38 por cuadro, es decir. la unidad de tiempo en donde los parámetros SAOC tal como ANTERIOR e IOC, se calculan en un cuadro SAOC 36 y bsFreqRes puede definir la cantidad de bandas de frecuencia de procesamiento para las cuales los parámetros SAOC se calculan, es decir la cantidad de bandas hacia donde el dominio de frecuencia es subdividido y para que los parámetros SAOC se determinan y transmitido. Por esta medida, cada cuadro está dividido en recuadros de tiempo/frecuencia ejemplificados en la Figura 2 por líneas entrecortadas 39.

[00026] El convertidor de multicanal a estéreo 16 calcula parámetros SAOC según las siguientes fórmulas. En términos particulares, el convertidor de multicanal a estéreo 16 calcula diferencias de nivel de objeto para cada objeto i como:

$$OLD_i = \frac{\sum_n \sum_{k \in m} x_i^{n,k} x_i^{n,k*}}{\max_j \left(\sum_n \sum_{k \in m} x_j^{n,k} x_j^{n,k*} \right)}$$

donde las sumas y los índices n y k, respectivamente, pasan por todos los intervalos temporales de banco de filtros 34, y todas las subbandas de banco de filtros 30 que pertenecen a un cierto recuadro de tiempo/frecuencia 39. Así, las energías de todos los valores de la subbanda x_i de una señal de audio u objeto yo soy resumido y normalizado al valor de energía más elevado de aquel recuadro entre todos los objetos o señales de audio.

[00027] Adicionalmente el convertidor de multicanal a estéreo SAOC 16 es capaz de computar una medida de semejanzas de los recuadros de tiempo/frecuencia correspondientes de pares de la diferente entrada se opone 14₁ a 14_N. Aunque el convertidor de multicanal a estéreo SAOC 16 pueda computar la medida de semejanzas entre todos los pares de la entrada se opone 14₁ a 14_N, el convertidor de multicanal a estéreo 16 también puede suprimir la señalización de las medidas de semejanzas o restringir el cálculo de las medidas de semejanzas a objetos de audio 14₁ a 14_N que forman izquierda o canales derechos de un canal estéreo común. En cualquier caso, la medida de semejanzas se llama el parámetro de correlación cruzado entre objetos IOC_{i,j}. El cálculo es como sigue:

$$IOC_{i,j} = IOC_{j,i} = \text{Re} \left\{ \frac{\sum_n \sum_{k \in m} x_i^{n,k} x_j^{n,k*}}{\sqrt{\sum_n \sum_{k \in m} x_i^{n,k} x_i^{n,k*} \sum_n \sum_{k \in m} x_j^{n,k} x_j^{n,k*}}} \right\}$$

con nuevamente los índices n y k que pasan por todos los valores de la subbanda que pertenecen a un cierto recuadro de tiempo/frecuencia 39, y yo y j denotación de un cierto par de objetos de audio 14₁ a 14_N.

[00028] El convertidor de multicanal a estéreo 16 conversiones de multicanal a estéreo los objetos 14₁ a 14_N por uso de factores de ganancia aplicados a cada objeto 14₁ a 14_N.

[00029] En caso de una señal de conversión estereofónica de multicanal a estéreo, qué caso es ejemplificado en la Figura 1, un factor de ganancia D_{1,i} se aplica al objeto i y luego toda la ganancia los objetos amplificados son sumados a fin de obtener el canal de conversión de multicanal a estéreo de izquierda L0, y factores de ganancia D_{2,i} me aplico para oponerme i y luego los objetos amplificados por la ganancia son sumados a fin de obtener el canal de conversión de multicanal a estéreo derecho R0. Así, factores D_{1,i} y D_{2,i} formo una matriz de conversión de multicanal a estéreo D del tamaño 2xN con:

$$D = \begin{pmatrix} D_{1,1} & \dots & D_{1,N} \\ D_{2,1} & \dots & D_{2,N} \end{pmatrix}_y \begin{pmatrix} LO \\ RO \end{pmatrix} = D \cdot \begin{pmatrix} Obj_1 \\ \vdots \\ Obj_N \end{pmatrix}.$$

[00030] Esta prescripción de conversión de multicanal a estéreo es señalada al lado decodificador por medio de abajo DMG_i de ganancias de mezcla y, en caso de una señal de conversión estereofónica de multicanal a estéreo, diferencias de nivel de canal de conversión de multicanal a estéreo DCLD_i.

[00031] Las ganancias de conversión de multicanal a estéreo se calculan según:

$$DMG_i = 10 \log_{10} (D_{1,i}^2 + D_{2,i}^2 + \epsilon)$$

donde ϵ es un número pequeño tal como 10⁻⁹ o 96dB entrada de señal a continuación máxima.

[00032] Para el DCLD_s la siguiente fórmula aplica:

$$DCLD_1 = 10 \log_{10} \left(\frac{D_{1,i}^2}{D_{2,i}^2} \right).$$

[00033] El convertidor de multicanal a estéreo 16 genera la señal de conversión estereofónica de multicanal a estéreo según:

$$\begin{pmatrix} L0 \\ R0 \end{pmatrix} = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} \cdot \begin{pmatrix} Obj_1 \\ \vdots \\ Obj_N \end{pmatrix}$$

[00034] Así, en las fórmulas anteriormente mencionadas, los parámetros ANTERIORES e IOC son una función de las señales de audio y parámetros DMG y DCLD son una función de D. A propósito, esto se observa que D puede variar a tiempo.

[00035] En caso de la renderización binaural, qué modo de operación del decodificador se describe aquí, la señal de salida naturalmente comprende dos canales, es decir M' = 2. Sin embargo, la información de renderización ya mencionada 26 indica en cuanto a cómo las señales de entrada 14₁ a 14_N deben ser distribuidas en posiciones de altavoz virtuales 1 a M donde M podría ser más elevado que 2. La información de renderización, así, puede comprender **M** de la matriz de renderización de indicación en cuanto a cómo la entrada objeto a que obj_i debieran ser

distribuidos en las posiciones de altavoz virtuales j para obtener señales de altavoz virtuales vs_j con j que está entre 1 y M inclusivamente y y_o que estoy entre 1 y N inclusivamente, con:

$$\begin{pmatrix} vs_1 \\ \vdots \\ vs_M \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} Obj_1 \\ \vdots \\ Obj_N \end{pmatrix}$$

[00036] La información de renderización puede proporcionarse o introducir por el usuario en ninguna forma. Esto puede incluso posible que la información de renderización 26 esté contenida dentro de la información conexas de la transmisión SAOC 21 sí mismo. Por supuesto, la información de renderización puede dejarse hacerse variar a tiempo. Por ejemplo, la resolución de tiempo puede igualar la resolución de cuadro, es decir $\mathbf{M}$ puede definirse por cuadro 36. Incluso una varianza de $\mathbf{M}$ por la frecuencia puede ser posible. Por ejemplo, $\mathbf{M}$ podría definirse para cada

recuadro 39. Después, por ejemplo, $\mathbf{M}_{ren}^{l,m}$ se usará para denotar $\mathbf{M}$, con m de denotación de la banda de frecuencia y l denotación de la fracción de tiempo de parámetro 38.

[00037] Finalmente, en el seguir, el HRTFs 27 se mencionará. Estos HRTFs describen como una señal j de altavoz virtual es volverse en el oído derecho e izquierdo, respectivamente, de modo que las señales de aviso binaurales se conserven. En otras palabras, para cada posición de altavoz virtual j , dos HRTFs existen, a saber un para el oído de izquierda y otro para el oído derecho. COMO se describirá más detalladamente después, es posible que el decodificador se proporcione con parámetros HRTF 27 que comprenden, para cada posición de altavoz virtual j , un

desplazamiento de desplazamiento de fase Φ_j que describe el desplazamiento de desplazamiento de fase entre las señales recibidas por ambos oídos y se origina de la misma fuente j , y dos P_i de ampliaciones/atenuaciones de amplitud, R y $P_{i,L}$ para el oído izquierdo y derecho, respectivamente, describiendo las atenuaciones de ambas señales debido al cabezal del oyente. El parámetro HRTF 27 podría ser constante con el tiempo, pero se define a alguna resolución de la frecuencia que podría ser igual a la resolución de parámetro SAOC, es decir, por banda de

frecuencia. A continuación, los parámetros HRTF se les proporcionan como Φ_j^m , $P_{j,R}^m$ y $P_{j,L}^m$ con m de denotación de la banda de frecuencia.

[00038] La Figura 3 muestra al decodificador SAOC 12 de la Figura 1 más detalladamente. Como se muestra en esa parte, el decodificador 12 comprende una unidad de proceso previo de conversión de multicanal a estéreo 40 y una unidad de procesamiento de datos de parámetro SAOC 42. La unidad de proceso previo de conversión de multicanal a estéreo 40 se configura para recibir la señal 18 de conversión estereofónica de multicanal a estéreo y convertir mismo en la señal de salida binaural 24. La unidad de proceso previo de conversión de multicanal a estéreo 40 lleva a cabo esta conversión en una manera controlada por la unidad de procesamiento de datos de parámetro SAOC 42. En términos particulares, la unidad de procesamiento de datos de parámetro SAOC 42 proporciona la unidad de proceso previo de conversión de multicanal a estéreo 40 por una información de prescripción de renderización 44 que la unidad de procesamiento de datos de parámetro SAOC 42 deriva de la información conexas SAOC 20 e información de renderización 26.

[00039] La Figura 4 muestra la unidad de proceso previo de conversión de multicanal a estéreo 40 de acuerdo con una modalidad de la presente invención más detalladamente. En términos particulares, de acuerdo con la Figura 4, la unidad de proceso previo de conversión de multicanal a estéreo 40 comprende dos trayectos conectados en paralelo entre la entrada en donde la señal 18 de conversión estereofónica de multicanal a estéreo, es decir se recibe $\mathbf{X}^{n,k}$, y una salida de la unidad 40 en donde la señal de salida binaural $\hat{\mathbf{X}}^{n,k}$ es enviada, a saber un trayecto llamado trayecto de renderizado seco 46 hacia donde una unidad de renderización seca es en serie conectada, y un trayecto de renderizado húmedo 48 hacia donde un generador de señal de decorrelación 50 y una unidad de renderización húmeda 52 se conecta en serie, donde unas mezclas de etapa 53 mezcladas las salidas de ambos trayectos de renderizado 46 y 48 para obtener el resultado final, a saber la señal de salida binaural 24.

[00040] Como se describirá más detalladamente después, la unidad de renderización seca 47 se configura para calcular una señal de salida binaural preliminar 54 de la señal 18 de conversión estereofónica de multicanal a estéreo con la señal de salida binaural preliminar 54 que representa la salida del trayecto de renderización seco 46 – también llamado a veces “señal binaural seca” o simplemente “señal seca” a continuación. La unidad de renderización seca 47 lleva a cabo su cálculo basado en una prescripción de renderización seca presentado por la unidad de procesamiento de datos de parámetro SAOC 42. En la modalidad específica descrita después, la prescripción de renderización se define por $\mathbf{G}^{n,k}$ de la matriz de renderización seco,. La condición sólo mencionada se ilustra en la Figura 4 por medio de una flecha rota.

[00041] El generador de señal no correlacionado 50 se configura para generar una señal no correlacionada $\mathbf{X}_d^{n,k}$ de la señal 18 de conversión estereofónica de multicanal a estéreo por la conversión de multicanal a estéreo el que

mismo es un equivalente perceptual con una conversión monoaural de multicanal a estéreo del canal izquierdo y derecho de la señal 18 de conversión estereofónica de multicanal a estéreo con, sin embargo, es decorrelacionado a la conversión monoaural de multicanal a estéreo. como se muestra en la Figura 4, el generador de señal no correlacionado 50 puede comprender un circuito sumador 56 para sumar el canal derecho e izquierdo de la señal 18 de conversión estereofónica de multicanal a estéreo a, por ejemplo, una proporción 1:1 o, por ejemplo, alguna otra proporción fija para obtener la respectiva conversión monoaural de multicanal a estéreo 58, seguido de un

decorrelacionador 60 para generar la señal no correlacionada ya mencionada $X_d^{n,k}$. El decorrelacionador 60 puede

comprender, por ejemplo, uno o más etapas de retraso a fin de formar la señal no correlacionada $X_d^{n,k}$ de la versión retardada o una suma ponderada de las versiones retardadas de la conversión monoaural de multicanal a estéreo 58 o incluso una suma ponderada sobre la conversión monoaural de multicanal a estéreo 58 y la versión (ones) retardada de la conversión monoaural de multicanal a estéreo. Por supuesto, hay muchas alternativas para el decorrelacionador 60. En efecto, la decorrelación llevada a cabo por el decorrelacionador 60 y el generador de señal no correlacionado 50, respectivamente, tiende a bajar la coherencia entre canales entre la señal 62 no correlacionada y la conversión monoaural de multicanal a estéreo 58 cuando medido por la fórmula anteriormente mencionada correspondiente a la correlación cruzada entre objetos, con mantener sustancialmente las diferencias de nivel de objeto de lo mismo cuando medido por la fórmula anteriormente mencionada para diferencias de nivel de objeto.

[00042] La unidad de renderización húmeda 52 se configura para calcular una señal de salida binaural correctiva 64 de la señal 62 no correlacionada, la señal de salida binaural correctiva obtenida 64 representación de la salida del trayecto de renderización húmedo 48 – también llamado a veces “señal binaural húmeda” o simplemente “señal húmeda” a continuación. La unidad de renderización húmeda 52 bases su cálculo en una prescripción de renderización húmeda que, por su parte, depende de la prescripción de renderización seca usada por la unidad de renderización seca 47 como describe después. En consecuencia, la prescripción de renderización húmeda que se indica como $P_2^{n,k}$ en la Figura 4, se obtiene de la unidad de procesamiento de datos de parámetro SAOC 42 como se indica por la flecha rota en la Figura 4.

[00043] Las mezclas de etapa 53 mezcladas ambas señales de salida binaural 54 y 64 de los trayectos de renderización secos y húmedos 46 y 48 para obtener la señal de salida binaural final 24. Como se muestra en la Figura 4, la etapa 53 mezclada se configura para mezclar los canales derechos e izquierdos de las señales de salida binaural 54 y 64 individualmente y puede comprender, en consecuencia, un circuito sumador 66 para sumar los canales de izquierda de lo mismo y un circuito sumador 68 para sumar los canales derechos de lo mismo, respectivamente.

[00044] Después de que tener describió la estructura del decodificador SAOC 12 y la estructura interna de la unidad de proceso previo de conversión de multicanal a estéreo 40, la funcionalidad de lo mismo se describe en el seguir. En términos particulares, las modalidades detalladas descritas después de diferentes alternativas presentes para la unidad de procesamiento de datos de parámetro SAOC 42 para derivar la información de prescripción de renderización 44 así control de la coherencia entre canales de la señal 24 de objeto binaural. En otras palabras, la unidad de procesamiento de datos de parámetro SAOC 42 no sólo calcula la información de prescripción de renderización 44, pero simultáneamente controla la proporción de mezcla por la cual las señales 55 y 64 binaurales preliminares y correctivas se mezclan en la señal de salida binaural final 24.

[00045] De acuerdo con una primera alternativa, la unidad de procesamiento de datos de parámetro SAOC 42 se configura para controlar la proporción de mezcla sólo mencionada como se muestra en la Figura 5. En términos particulares, en una etapa 80, un valor de coherencia entre canales binaural actual de la señal de salida binaural preliminar 54 se determina o estimado por la unidad 42. En una etapa 82, la unidad de procesamiento de datos de parámetro de SAOC 42 determina un valor de coherencia entre canales binaural con especificidad de objetivo. Basado en estos valores de coherencia entre canales determinados, la unidad de procesamiento de datos de parámetro SAOC 42 conjuntos la proporción de mezcla ya mencionada en etapa 84. En términos particulares, la etapa 84 puede comprender la unidad de procesamiento de datos de parámetro SAOC 42 apropiadamente de calcular la prescripción de renderización seca usada por la unidad de renderización seca 42 y la prescripción de renderización húmeda usada por la unidad de renderización húmeda 52, respectivamente, basado en los valores de coherencia entre canales determinados en etapas 80 y 82, respectivamente.

[00046] A continuación, las alternativas ya mencionadas se describirán en una base matemática. Las alternativas se diferencian de entre sí en el modo que la unidad de procesamiento de datos de parámetro SAOC 42 determina la información de prescripción de renderización 44, incluyendo la prescripción de renderización seca y la prescripción de renderización húmeda con controlar intrínsecamente la proporción de mezcla entre trayectos de renderización secos y húmedos 46 y 48. De acuerdo con la primera alternativa representada en la Figura 5, la unidad de procesamiento de datos de parámetro SAOC 42 determina un valor de coherencia entre canales binaural con especificidad de objetivo. Como se describirá más detalladamente después, la unidad 42 puede llevar a cabo esta determinación basada en componentes de una matriz de coherencia con especificidad de objetivo $F = A \cdot E \cdot \text{Un}^*$, con “*” la denotación del conjugado transporta, un que es una matriz de renderización binaural con especificidad de

objetivo relación de los objetos/señales de audio 1 ... N al canal izquierdo y derecho de la señal de salida binaural 24 y señal de salida binaural preliminar 54, respectivamente, y que se deriva de la información de renderización 26 y parámetros HRTF 27, y E que es una matriz los coeficientes de que se derivan del $IOC_{ij}^{l,m}$ y diferencias de nivel de objeto $OLD_i^{l,m}$. El cálculo puede llevarse a cabo en la resolución espacial/temporal de los parámetros SAOC, es decir para cada uno (l, m) . Sin embargo, es adicional posible llevar a cabo el cálculo en una resolución inferior con la interpolación entre los respectivos resultados. La última declaración también es verdadera para los cálculos subsecuentes dispuestos después.

[00047] Como la matriz de renderización binaural con especificidad de objetivo A se relacionan los objetos de entrada 1 ... N a los canales derechos e izquierdos de la señal de salida binaural 24 y la señal de salida binaural preliminar 54, respectivamente, mismo es del tamaño $2 \times N$, es decir,

$$A = \begin{pmatrix} a_{11} & \dots & a_{1N} \\ a_{21} & \dots & a_{2N} \end{pmatrix}$$

[00048] La matriz ya mencionada E es del tamaño $N \times N$ con sus coeficientes que se definen como:

$$e_{ij} = \sqrt{OLD_i \cdot OLD_j \cdot \max(IOC_{ij}, \theta)}$$

[00049] Así, la matriz E con:

$$E = \begin{pmatrix} e_{11} & \dots & e_{1N} \\ \vdots & \ddots & \vdots \\ e_{N1} & \dots & e_{NN} \end{pmatrix}$$

tiene a lo largo de ello la diagonal las diferencias de nivel de objeto, es decir:

$$e_{ii} = OLD_i$$

ya que $IOC_{ij} = 1$ for $i = j$ mientras que la matriz E tiene fuera de sus coeficientes de la matriz diagonales que representan la media geométrica de las diferencias de nivel de objeto de objetos i y j , respectivamente, ponderado

con la medida de correlación cruzada entre objetos IOC_{ij} (se proporcionó mismo es mayor que 0 con los coeficientes que se configuran a 0 otra cosa).

[00050] Comparado al mismo, las segundas y terceras alternativas descritas después, procuran obtener las matrices de renderización encontrando el mejor partido en el sentido menos cuadrado de la ecuación que se corresponde de la señal 18 de conversión estereofónica de multicanal a estéreo en la señal de salida binaural preliminar 54 por medio de la matriz de renderización seca G a la ecuación de renderización con especificidad de objetivo que se corresponde de los objetos de entrada vía lado a lado matriz un en la señal de salida binaural "con especificidad de objetivo" 24 con segundo y tercer diferenciarse alternativo de entre sí en el modo que el mejor partido se forma y el modo que la matriz de renderización húmeda se selecciona.

[00051] A fin de aliviar la comprensión de las siguientes alternativas, la descripción ya mencionada de Figuras 3 y 4 es matemáticamente descrita de nuevo. Como descrito antes, la señal 18 de conversión estereofónica de multicanal a estéreo $X^{n,k}$ alcanza el decodificador SAOC 12 junto con los parámetros SAOC 20 y usuario información de renderización definida 26. Adicionalmente, el decodificador de SAOC 12 y unidad de procesamiento de datos de parámetro SAOC 42, respectivamente, tiene el acceso a una base de datos HRTF como se indica por la flecha 27.

Los parámetros SAOC transmitidos comprenden diferencias de nivel de objeto $OLD_i^{l,m}$, valores de correlación

cruzados entre objetos $IOC_{ij}^{l,m}$, las ganancias de conversión de multicanal a estéreo $DMG_i^{l,m}$ y las diferencias

de nivel de canal de conversión de multicanal a estéreo $DCLD_i^{l,m}$ para todo N se oponen i, j con " l, m " de denotación del respectivo tiempo / recuadro espectral 39 con l especificación del tiempo y m de especificación de la

frecuencia. Se supone, ejemplarmente, que los parámetros HRTF 27 se les proporcionen como $P_{q,L}^m$, $P_{q,R}^m$ y Φ_q^m para todas las posiciones de altavoz virtuales o posición de la fuente de sonido espacial virtual q , para la izquierda (L) y derecha (R) canal binaural y para todas las bandas de frecuencia m .

[00052] La unidad de proceso previo de conversión de multicanal a estéreo 40 se configura para calcular la salida binaural $\hat{\mathbf{X}}^{n,k}$, tan calculada de la conversión estereofónica de multicanal a estéreo $\mathbf{X}^{n,k}$ y señal de conversión monoaural de multicanal a estéreo decorrelacionada $\mathbf{X}_d^{n,k}$ como:

$$\hat{\mathbf{X}}^{n,k} = \mathbf{G}^{n,k} \mathbf{X}^{n,k} + \mathbf{P}_2^{n,k} \mathbf{X}_d^{n,k}$$

[00053] La señal no correlacionada es $\mathbf{X}_d^{n,k}$ perceptivamente equivalente a la suma 58 de los canales de conversión de multicanal a estéreo derechos e izquierdos de la señal 18 de conversión estereofónica de multicanal a estéreo, pero máximamente decorrelacionado a ello según:

$$\mathbf{X}_d^{n,k} = \text{decorrFunction}((1 \quad 1) \mathbf{X}^{n,k})$$

[00054] Respecto a la Figura 4, el generador de señal no correlacionado 50 lleva a cabo la función decorrFunction de la fórmula anteriormente mencionada.

[00055] Adicionalmente, como también descrito antes, la unidad de proceso previo de conversión de multicanal a estéreo 40 comprende dos trayectos de renderizado paralelos 46 y 48. En consecuencia, la ecuación anteriormente mencionada se basa en dos matrices dependientes de tiempo/frecuencia, a saber, $\mathbf{G}^{l,m}$ para el seco y $\mathbf{P}_2^{l,m}$ para el trayecto de renderizado húmedo.

[00056] Como se muestra en la Figura 4, la decorrelación en el trayecto de renderizado húmedo puede ponerse en práctica por la suma del canal de conversión de multicanal a estéreo derecho e izquierdo que se alimenta en un decorrelacionador 60 que genera una señal 62, que es perceptivamente equivalente, pero máximamente decorrelacionada a su entrada 58.

[00057] Los elementos de las matrices sólo mencionadas se calculan por la unidad de proceso previo SAOC 42. Como también denotado antes, los elementos de las matrices sólo mencionadas pueden calcularse en la resolución de tiempo/frecuencia de los parámetros SAOC, es decir para cada intervalo temporal l y cada banda de procesamiento m . Los elementos de la matriz obtenidos pueden ser extendidos sobre la frecuencia e interpolados a tiempo dando como resultado matrices $\mathbf{E}^{n,k}$ y $\mathbf{P}_2^{l,m}$ definido para todos los intervalos temporales de banco de filtros n y subbandas de frecuencia k . Sin embargo, como ya antes, allí también son alternativas. Por ejemplo, la interpolación podría dejarse lejos, de modo que en el antes de la ecuación, los índices n, k pudieran ser con eficacia reemplazados por " l, m ". Más aún, el cálculo de los elementos de las matrices sólo mencionadas podría llevarse a cabo incluso a una resolución de tiempo/frecuencia reducida con la interpolación en la resolución l, m o n, k . Así, nuevamente, aunque en el seguir, los índices l, m indique que los cálculos de la matriz se llevan a cabo para cada recuadro 39, el cálculo puede llevarse a cabo a alguna resolución inferior donde, aplicando las respectivas matrices por la unidad de proceso previo de conversión de multicanal a estéreo 40, las matrices de renderización pueden ser interpoladas hasta una resolución final tal como abajo a la resolución de tiempo/frecuencia QMF de los valores de la subbanda individuales 32.

[00058] Según la primera alternativa anteriormente mencionada, la matriz de renderizado seca $\mathbf{G}^{l,m}$ se calcula para la izquierda y el canal de conversión de multicanal a estéreo derecho por separado el que:

$$\mathbf{G}^{l,m} = \begin{pmatrix} P_L^{l,m,1} \cos(\beta^{l,m} + \alpha^{l,m}) \exp(j \frac{\phi^{l,m,1}}{2}) & P_L^{l,m,2} \cos(\beta^{l,m} + \alpha^{l,m}) \exp(j \frac{\phi^{l,m,2}}{2}) \\ P_R^{l,m,1} \cos(\beta^{l,m} - \alpha^{l,m}) \exp(-j \frac{\phi^{l,m,1}}{2}) & P_R^{l,m,2} \cos(\beta^{l,m} - \alpha^{l,m}) \exp(-j \frac{\phi^{l,m,2}}{2}) \end{pmatrix}$$

[00059] Las ganancias correspondientes $P_L^{l,m,x}$, $P_R^{l,m,x}$ y diferencias de fase $\phi^{l,m,x}$ se define como:

$$P_L^{l,m,x} = \sqrt{\frac{f_{11}^{l,m,x}}{V^{l,m,x}}}, \quad P_R^{l,m,x} = \sqrt{\frac{f_{22}^{l,m,x}}{V^{l,m,x}}},$$

$$\phi^{l,m,x} = \begin{cases} \arg(f_{12}^{l,m,x}) & \text{if } 0 \leq m \leq \text{const}_1 \wedge \frac{|f_{12}^{l,m,x}|}{\sqrt{f_{11}^{l,m,x} f_{22}^{l,m,x}}} \geq \text{const}_2 \\ 0 & \text{else} \end{cases}$$

donde el const_1 puede ser, por ejemplo, 11 y const_2 puede ser 0.6. El índice x denota la izquierda o canal de conversión de multicanal a estéreo derecho y en consecuencia asume o 1 o 2.

[00060] En términos generales, el antes de que la condición se distinga entre un intervalo espectral más elevado y un intervalo espectral inferior y, sobre todo, sólo sea (potencialmente) realizada para el intervalo espectral inferior.

[00061] Además o alternativamente, la condición es dependiente de en cuanto a si uno del valor de coherencia entre canales binaural actual y el valor de coherencia entre canales binaural con especificidad de objetivo tiene una relación predeterminada a un valor umbral de coherencia o no, con la condición que es (potencialmente), realizada sólo si la coherencia excede el valor umbral. Las subcondiciones individuales sólo mencionadas, como se indica antes, pueden combinarse por medio de un y operación.

$\checkmark^{\text{escalar}, m, x}$ se calcula como:

$$V^{l,m,x} = \mathbf{D}^{l,m,x} \mathbf{E}^{l,m} (\mathbf{D}^{l,m,x}) + \varepsilon.$$

[00062] Esto se observa que ε puede ser igual que o diferente al mencionado ε antes con respecto a la definición de las ganancias de conversión de multicanal a estéreo. La matriz $\mathbf{E}$ se ha introducido ya antes. El índice (l, m) simplemente denota la dependencia de tiempo/frecuencia del cálculo de la matriz como ya mencionado antes. Adicionalmente, las matrices $\mathbf{D}^{l,m,x}$ también tiene mencionado antes, con respecto a la definición de las ganancias de conversión de multicanal a estéreo y las diferencias de nivel de canal de conversión de multicanal a estéreo, de modo que $\mathbf{D}^{l,m,1}$ correspondan a $\mathbf{D}_1$ ya mencionado y $\mathbf{D}^{l,m,2}$ corresponde a $\mathbf{D}_2$ ya mencionado.

[00063] Sin embargo, a fin de aliviar la comprensión como la unidad de procesamiento de datos de parámetro SAOC 42 deriva la matriz de renderizado seca $\mathbf{G}^{l,m}$ de los parámetros SAOC recibidos, la correspondencia entre la matriz de conversión de multicanal a estéreo de canal $\mathbf{D}^{l,m,x}$ y la prescripción de conversión de multicanal a estéreo que

comprende las ganancias de conversión de multicanal a estéreo $DMG_i^{l,m}$ y $DCLD_i^{l,m}$ se presenta nuevamente, en la dirección inversa. En términos particulares, los elementos $d_i^{l,m,x}$ de la matriz de conversión de multicanal a estéreo de canal $\mathbf{D}^{l,m,x}$ de tamaño $1 \times N$, es decir, $\mathbf{D}^{l,m,x} = (d_1^{l,m,x}, \dots, d_N^{l,m,x})$ se les proporcionan como:

$$d_i^{l,m,1} = 10 \frac{DMG_i^{l,m}}{20} \sqrt{\frac{\tilde{d}_i^{l,m}}{1+\tilde{d}_i^{l,m}}}, \quad d_i^{l,m,2} = 10 \frac{DMG_i^{l,m}}{20} \sqrt{\frac{1}{1+\tilde{d}_i^{l,m}}}$$

con el elemento $\tilde{d}_i^{l,m}$ que se define como:

$$\tilde{d}_i^{l,m} = 10^{\frac{DCLD_i^{l,m}}{10}}.$$

[00064] En la ecuación superior de $\mathbf{G}^{l,m}$, las ganancias $P_L^{l,m,x}$ y $P_R^{l,m,x}$ y las diferencias de fase $\phi^{l,m,x}$ dependen de coeficientes f_{uv} de una matriz de covarianza de objetivo de paciente de canal-x $\mathbf{F}^{l,m,x}$, que, por su parte, como se configurará más detalladamente después, depende de una matriz $\mathbf{e}^{l,m,x}$ del tamaño $N \times N$ los elementos $e_{ij}^{l,m,x}$ de que se calculan como:

$$e_{ij}^{l,m,x} = e_{ij}^{l,m} \left(\frac{d_i^{l,m,x}}{d_i^{l,m,1} + d_i^{l,m,2}} \right) \left(\frac{d_j^{l,m,x}}{d_j^{l,m,1} + d_j^{l,m,2}} \right)$$

[00065] Los elementos $e_{ij}^{l,m}$ de la matriz $\mathbf{E}^{l,m}$ de tamaño $N \times N$ son, tan declarados antes, proporcionado como: $e_{ij}^{l,m} = \sqrt{OLD_i^{l,m} \cdot OLD_j^{l,m}} \cdot \max(IOC_{ij}^{l,m}, 0)$.

[00066] La matriz de covarianza con especificidad de objetivo sólo mencionada $\mathbf{F}^{l,m,x}$ del tamaño 2×2 con elementos $f_{uv}^{l,m,x}$ es, de manera similar a la matriz de covarianza $\mathbf{F}$ indicada antes, proporcionado como:

$$\mathbf{F}^{l,m,x} = \mathbf{A}^{l,m} \mathbf{E}^{l,m,x} (\mathbf{A}^{l,m})^*$$

donde "*" corresponde al conjugado transportan.

[00067] La matriz de renderización binaural con especificidad de objetivo $\mathbf{A}^{l,m}$ se deriva de los parámetros HRTF Φ_q^m , $P_{q,R}^m$ y $P_{q,L}^m$ para todas las posiciones de altavoz virtuales NHRTF q y la matriz de renderización $\mathbf{M}_{ren}^{l,m}$ es del

tamaño $2 \times N$. Sus elementos $a_{ui}^{l,m}$ definen relación deseada entre todos los objetos i y la señal de salida binaural como:

$$a_{1,i}^{l,m} = \sum_{q=0}^{N_{\text{HRTF}}-1} m_{q,i}^{l,m} P_{q,L}^m \exp\left(j \frac{\phi_q^m}{2}\right), a_{2,i}^{l,m} = \sum_{q=0}^{N_{\text{HRTF}}-1} m_{q,i}^{l,m} P_{q,R}^m \exp\left(-j \frac{\phi_q^m}{2}\right).$$

5 **[00068]** La matriz de renderización $\mathbf{M}_{ren}^{l,m}$ con elementos $m_{qi}^{l,m}$ se relaciona cada objeto de audio i a un altavoz virtual q representado por el HRTF.

[00069] La matriz de renderizado húmeda $\mathbf{P}_2^{l,m}$ se calcula basada en la matriz de renderizado seca $\mathbf{G}^{l,m}$ como:

$$\mathbf{P}_2^{l,m} = \begin{pmatrix} P_L^{l,m} \sin(\beta^{l,m} + \alpha^{l,m}) \exp\left(j \frac{\arg(c_{12}^{l,m})}{2}\right) \\ P_R^{l,m} \sin(\beta^{l,m} - \alpha^{l,m}) \exp\left(-j \frac{\arg(c_{12}^{l,m})}{2}\right) \end{pmatrix}$$

10

[00070] Las ganancias $P_L^{l,m}$ y $P_R^{l,m}$ se definen como:

$$P_L^{l,m} = \sqrt{\frac{c_{11}^{l,m}}{v^{l,m}}}, P_R^{l,m} = \sqrt{\frac{c_{22}^{l,m}}{v^{l,m}}}.$$

[00071] El 2x2 matriz de covariancia $\mathbf{C}^{l,m}$ con elementos $c_{u,v}^{l,m}$ de la señal 54 binaural seca es estimado como:

$$\mathbf{C}^{l,m} = \tilde{\mathbf{G}}^{l,m} \mathbf{D}^{l,m} \mathbf{E}^{l,m} (\mathbf{D}^{l,m})^* (\tilde{\mathbf{G}}^{l,m})^*$$

15

donde

$$\tilde{\mathbf{G}}^{l,m} = \begin{pmatrix} P_L^{l,m,1} \exp\left(j \frac{\phi^{l,m,1}}{2}\right) & P_L^{l,m,2} \exp\left(j \frac{\phi^{l,m,2}}{2}\right) \\ P_R^{l,m,1} \exp\left(-j \frac{\phi^{l,m,1}}{2}\right) & P_R^{l,m,2} \exp\left(-j \frac{\phi^{l,m,2}}{2}\right) \end{pmatrix}$$

[00072] $v^{l,m}$ escalar, m se calcula como:

$$v^{l,m} = \mathbf{W}^{l,m} \mathbf{E}^{l,m} (\mathbf{W}^{l,m})^* + \epsilon.$$

20

[00073] Los elementos $w_i^{l,m}$ de la matriz de conversión monoaural de multicanal a estéreo húmeda $\mathbf{W}^{l,m}$ del tamaño $1 \times N$ se les proporciona como:

$$w_i^{l,m} = d_i^{l,m,1} + d_i^{l,m,2}.$$

25

[00074] Los elementos $d_{x,i}^{l,m}$ de la matriz de conversión estereofónica de multicanal a estéreo $\mathbf{D}^{l,m}$ del tamaño $2 \times N$ se les proporciona como:

$$d_{x,i}^{l,m} = d_i^{l,m,x}.$$

30

[00075] En la ecuación anteriormente mencionada de $\mathbf{G}^{l,m}$, $\alpha^{l,m}$ y $\beta^{l,m}$ representan los ángulos de los rotadores dedicado para el control de ICC. En términos particulares, el ángulo del rotador $\alpha^{l,m}$ controla el mezclado de la señal binaural seca y húmeda a fin de ajustar el ICC de la salida binaural 24 a aquel del objetivo binaural. Al determinar a los ángulos de los rotadores, el ICC de la señal 54 binaural seca debería obtenerse en cuenta que es, según el contenido de audio y la matriz de conversión estereofónica de multicanal a estéreo $\mathbf{D}$, por lo común más pequeño que 1.0 y mayor que ICC con especificidad de objetivo. Esto es en contraste con una conversión monoaural de multicanal a estéreo la renderización binaural basada donde el ICC de la señal binaural seca siempre sería igual a 1.0.

35

[00076] Los ángulos de los rotadores $\alpha^{l,m}$ y $\beta^{l,m}$ controlan el mezclado de la señal binaural seca y húmeda. El ICC $\rho_c^{l,m}$ de la conversión estereofónica de multicanal a estéreo reproducida binaural seca 54 es, en la etapa 80, estimado como:

40

$$\rho_C^{l,m} = \min \left(\frac{|c_{12}^{l,m}|}{\sqrt{c_{11}^{l,m} c_{22}^{l,m}}}, 1 \right)$$

[00077] ICC con especificidad de objetivo binaural total $\rho_C^{l,m}$ es, en la etapa 82, estimado como, o determinado de ser:

$$\rho_T^{l,m} = \min \left(\frac{|f_{12}^{l,m}|}{\sqrt{f_{11}^{l,m} f_{22}^{l,m}}}, 1 \right)$$

[00078] Los ángulos de los rotadores $\alpha^{l,m}$ y $\beta^{l,m}$ para minimizar la energía de la señal húmeda son luego, en la etapa 84, ajustados para ser:

$$\alpha^{l,m} = \frac{1}{2} \left(\arccos(\rho_T^{l,m}) - \arccos(\rho_C^{l,m}) \right),$$

$$\beta^{l,m} = \arctan \left(\tan(\alpha^{l,m}) \frac{P_R^{l,m} - P_L^{l,m}}{P_L^{l,m} + P_R^{l,m}} \right).$$

[00079] Así, según la descripción matemática anteriormente descrita de la funcionalidad del decodificador SAOC 12 para generar la señal de salida binaural 24, la unidad de procesamiento de datos de parámetro SAOC 42 computa,

en determinar de ICC binaural actual, $\rho_C^{l,m}$ por el uso de las ecuaciones anteriormente presentadas para $\rho_C^{l,m}$ y las ecuaciones subsidiarias también presentadas antes. De forma similar, la unidad de procesamiento de datos de parámetro de SAOC 42 computa, en determinar de ICC binaural con especificidad de objetivo en la etapa 82, el

parámetro $\rho_C^{l,m}$ por la ecuación anteriormente indicada y las ecuaciones subsidiarias. En la base de lo mismo, la unidad de procesamiento de datos de parámetro SAOC 42 determina en la etapa 84 los ángulos de los rotadores, por tanto, estableciendo la proporción de mezcla entre el trayecto de renderización seco y húmedo. Con ángulos de los rotadores, la unidad de procesamiento de datos de parámetro de SAOC 42 construye las matrices de

renderización secas y húmedas o parámetros de conversión de estéreo a multicanal $\mathbf{G}^{l,m}$ y $\mathbf{P}_2^{l,m}$ que, por su parte, se usan por la unidad de proceso previo de conversión de multicanal a estéreo 40 - a la resolución n, k - a fin de derivar la señal de salida binaural 24 de la conversión estereofónica de multicanal a estéreo 18.

[00080] Deberse observar que la primera alternativa ya mencionada puede hacerse variar de algún modo. Por ejemplo, la ecuación anteriormente presentada para la diferencia de fase de intercanal $\Phi_C^{l,m}$ podría cambiarse hasta el punto de que la segunda subcondición podría comparar ICC actual de la conversión estereofónica de multicanal a estéreo reproducida binaural seca a const_2 en vez del ICC determinado de la matriz de covariancia de paciente de

canal $\mathbf{F}^{l,m,x}$ de modo que en aquella la ecuación la porción $\frac{|f_{12}^{l,m,x}|}{\sqrt{f_{11}^{l,m,x} f_{22}^{l,m,x}}}$ fuera reemplazada por el término $\frac{|c_{12}^{l,m}|}{\sqrt{c_{11}^{l,m} c_{22}^{l,m}}}$.

[00081] Adicionalmente, débese observar que, de acuerdo con la notación seleccionada, en algunas ecuaciones superiores, una matriz de todos los se ha dejado lejos cuando una constante escalar, como el e se agrega a una matriz de modo que esta constante se agregue a cada coeficiente de la respectiva matriz.

[00082] Una generación alternativa de la matriz de renderización seca con el potencial más elevado de la extracción de objeto se basa en un tratamiento conjunto de los canales de conversión de multicanal a estéreo derechos e izquierdos. Omitiendo al par de índice de subbanda para la claridad, el principio debe apuntar en el mejor partido en la menor parte de sentido de cuadrados de:

$$\hat{\mathbf{X}} = \mathbf{G}\mathbf{X}$$

a la renderización con especificidad de objetivo:

$$\mathbf{Y} = \mathbf{A}\mathbf{S}$$

[00083] Esto produce la matriz de covariancia con especificidad de objetivo:

$$\mathbf{Y}\mathbf{Y}^* = \mathbf{A}\mathbf{S}\mathbf{S}^*\mathbf{A}^*$$

donde el complejo que proporcionan a la matriz de renderización binaural con especificidad de objetivo valorada $\mathbf{A}$ en una fórmula anterior y la matriz $\mathbf{S}$ contiene las señales de subbanda de objetos originales como hileras.

- 5 **[00084]** La menor parte de partido de cuadrados se calcula de la segunda información de orden derivada del objeto transportado y datos de conversión de multicanal a estéreo. Es decir las siguientes substituciones se llevan a cabo:

$$\begin{aligned} \mathbf{XX}^* &\leftrightarrow \mathbf{DED}^*, \\ \mathbf{YX}^* &\leftrightarrow \mathbf{AED}^*, \\ \mathbf{YY}^* &\leftrightarrow \mathbf{AEA}^*. \end{aligned}$$

- 10 Para motivar las substituciones, recuerde que SAOC objetan parámetros información por lo común transportable sobre las energías de objeto correlaciones cruzadas entre objetos (ANTERIORES) y (seleccionadas) (IOC). De estos parámetros, la matriz de covariancia de objeto de $N \times N$ $\mathbf{E}$ se deriva, que representa una aproximación a $\mathbf{SS}^*$, es decir. $\mathbf{E} \approx \mathbf{SS}^*$, produciendo $\mathbf{YY}^* = \mathbf{AEA}^*$.

- 15 **[00085]** Adicionalmente, $\mathbf{X} = \mathbf{DS}$ y la matriz de covariancia de conversión de multicanal a estéreo se convierten:

$$\mathbf{XX}^* = \mathbf{DSS}^* \mathbf{D}^*,$$
que nuevamente puede derivarse de $\mathbf{E}$ por $\mathbf{XX}^* = \mathbf{DED}^*$.

- 20 **[00086]** La matriz de renderización seca $\mathbf{G}$ se obtiene solucionando la menor parte de problema de cuadrados:

$$\min \{ \text{norma} \{ \mathbf{Y} - \mathbf{X} \} \}.$$

$$\mathbf{G} = \mathbf{G}_0 = \mathbf{YX}^* (\mathbf{XX}^*)^{-1}$$

donde $\mathbf{YX}^*$ se calcula como $\mathbf{YX}^* = \mathbf{AED}^*$.

- 25 **[00087]** Así, la unidad de renderización seca 42 determina que la señal de salida binaural $\hat{\mathbf{X}}$ forma la señal de conversión de multicanal a estéreo $\mathbf{X}$ por el uso de la matriz de renderizado seca 2×2 $\mathbf{G}$, por $\hat{\mathbf{X}} = \mathbf{GX}$, y la unidad de procesamiento de datos de parámetro SAOC determina $\mathbf{G}$ por el uso de las fórmulas superiores de ser:

$$\mathbf{G} = \mathbf{AED}^* (\mathbf{DED}^*)^{-1},$$

- 30 **[00088]** Considerando este complejo matriz de renderización seca valorada, el complejo la matriz de renderización húmeda valorada $\mathbf{P}$ - antes denotó que $\mathbf{P}_2$ - se calcula en la unidad de procesamiento de datos de parámetro SAOC 42 tomando en cuenta la matriz de error de covariancia ausente:

$$\Delta \mathbf{R} = \mathbf{YY}^* - \mathbf{G}_0 \mathbf{XX}^* \mathbf{G}_0^*.$$

- 35 **[00089]** Esto puede mostrarse que esta matriz tiene resultado positivo y una opción preferida de $\mathbf{P}$ se les proporciona seleccionando un vector propio de norma de unidad $\mathbf{u}$ correspondiente al valor propio más grande λ de $\Delta \mathbf{R}$ y modificándolo a escala según:

$$\mathbf{P} = \sqrt{\frac{\lambda}{V}} \mathbf{u},$$

donde el escalar V se calcula como observado antes, es decir. $V = \mathbf{WE}(\mathbf{W})^* + \epsilon$

- 40 **[00090]** En otras palabras, ya que el trayecto de renderizado húmedo se instala para corregir la correlación de la solución seca obtenida, $\Delta \mathbf{R} = \mathbf{AEA}^* - \mathbf{G}_0 \mathbf{DED}^* \mathbf{G}_0^*$ representa la matriz de error de covariancia ausente, es decir. $\mathbf{YY}^* = \hat{\mathbf{X}} \hat{\mathbf{X}}^* + \Delta \mathbf{R}$ o, respectivamente, $\Delta \mathbf{R} = \mathbf{YY}^* - \hat{\mathbf{X}} \hat{\mathbf{X}}^*$, y, por lo tanto, la unidad de procesamiento de datos de parámetro SAOC 42 ajusta $\mathbf{P}$ tal que $\mathbf{PPS}^* = \Delta \mathbf{R}$, una solución para la cual se les proporciona seleccionando el vector propio de norma de unidad anteriormente mencionado $\mathbf{u}$.

- 45 **[00091]** Un tercer método para generar matrices de renderización secas y húmedas representa una valoración de los parámetros de renderización basados en la señal de aviso predicción compleja reprimida y combina la ventaja de rehabilitar la estructura de covariancia compleja correcta con las ventajas del tratamiento conjunto de canales de conversión de multicanal a estéreo para la extracción de objeto mejorada. Una oportunidad adicional ofrecida por este método es ser capaz de omitir la conversión de estéreo a multicanal húmeda totalmente en muchos casos, que así preparan el terreno para una versión de la renderización binaural con la complejidad computacional inferior.

[00092] Como con la segunda alternativa, la tercera alternativa presentada después se basa en un tratamiento conjunto de los canales de conversión de multicanal a estéreo derechos e izquierdos.

El principio debe apuntar en el mejor partido en la menor parte de sentido de cuadrados de:

$$\hat{\mathbf{X}} = \mathbf{GX}$$

5 a la renderización con especificidad de objetivo $\mathbf{Y} = \mathbf{COMO}$ bajo la coacción de covariancia compleja correcta:

$$\mathbf{GXX}^* \mathbf{G}^* + \mathbf{VPP}^* = \hat{\mathbf{Y}}\hat{\mathbf{Y}}^*.$$

[00093] Así, es el objetivo de encontrar una solución para $\mathbf{G}$ y $\mathbf{P}$, este

1) $\hat{\mathbf{Y}}\hat{\mathbf{Y}}^* = \mathbf{YY}^*$ (que es la coacción a la formulación en 2); y

10 2) el minuto {norma $\{\mathbf{Y} - \hat{\mathbf{Y}}\}$ }, ya que es solicitado dentro de la segunda alternativa.

[00094] De la teoría de multiplicadores de Lagrange, resulta que allí existe una matriz autoadjunta $\mathbf{M} = \mathbf{M}^*$, el que:

$$\mathbf{MP} = \mathbf{0}, \text{ y}$$

$$15 \quad \mathbf{MGXX}^* = \mathbf{YX}^*$$

[00095] En el caso genérico donde tanto $\mathbf{YX}^*$ como $\mathbf{XX}^*$ son no singulares esto sigue de la segunda ecuación que $\mathbf{M}$ es no singular, y por lo tanto $\mathbf{P} = \mathbf{0}$ es la única solución de la primera ecuación. Esto es una solución sin la renderización húmeda. La configuración $\mathbf{K} = \mathbf{m}^{-1}$ esto puede observarse por que la conversión de estéreo a multicanal seca correspondiente se les proporciona:

$$\mathbf{G} = \mathbf{KG}_0$$

donde $\mathbf{G}_0$ es la solución profética derivada antes con respecto a la segunda alternativa, y la matriz autoadjunta $\mathbf{K}$ soluciona:

$$25 \quad \mathbf{KG}_0\mathbf{XX}^*\mathbf{G}_0^*\mathbf{K}^* = \mathbf{YY}^*.$$

[00096] Si la positiva única y por lo tanto autoadjunta matriz raíz cuadrada de la matriz $\mathbf{G}_0\mathbf{XX}^*\mathbf{G}_0^*$ se denotada por $\mathbf{Q}$, entonces la solución puede ser escrita como:

$$\mathbf{K} = \mathbf{Q}^{-1}(\mathbf{QYY}^*\mathbf{Q})^{1/2}\mathbf{Q}^{-1}.$$

30 [00097] Así, la unidad de procesamiento de datos de parámetro SAOC 42 determina $\mathbf{G}$ para ser: $\mathbf{KG}_0 = \mathbf{q}^{-1}(\mathbf{QYY}^*\mathbf{Q})^{1/2}\mathbf{Q}^{-1}\mathbf{G}_0 = (\mathbf{G}_0\mathbf{DED}^*\mathbf{G}_0^*)^{-1}(\mathbf{G}_0\mathbf{DED}^*\mathbf{G}_0^*\mathbf{AEA}^*\mathbf{G}_0\mathbf{DED}^*\mathbf{G}_0^*)^{1/2}(\mathbf{G}_0\mathbf{DED}^*\mathbf{G}_0^*)^{-1}\mathbf{G}_0$ con $\mathbf{G}_0 = \mathbf{AED}^*(\mathbf{DED}^*)^{-1}$.

[00098] Para las raíces cuadradas internas habrá en general cuatro soluciones autoadjuntas, y la solución que da como resultado al mejor partido de $\hat{\mathbf{X}}$ a $\mathbf{Y}$ se selecciona.

35 [00099] En la práctica, uno tiene que limitar la matriz de renderización seca $\mathbf{G} = \mathbf{KG}_0$ a un tamaño máximo, por ejemplo limitando la condición en la suma de cuadrados de valores absolutos de todos los coeficientes de la matriz de renderización secos, que pueden expresarse como:

$$40 \quad \text{traza}(\mathbf{GG}^*) \leq g_{\max}.$$

[000100] Si la solución viola esta condición que limita, una solución que apoya en el límite se encuentra en cambio. Esto se logra agregando la coacción:

$$\text{traza}(\mathbf{GG}^*) = g_{\max}$$

a las coacciones anteriores y derivar de nuevo las ecuaciones de Lagrange. Resulta que la ecuación anterior:

$$45 \quad \mathbf{MGXX}^* = \mathbf{YX}^*$$

tiene que ser reemplazado por:

$$\mathbf{MGXX}^* + \mu \mathbf{I} = \mathbf{YX}^*$$

donde μ es un parámetro complejo intermedio adicional e $\mathbf{I}$ es la matriz de identidad 2x2. Una solución con la renderización húmeda distinta a cero $\mathbf{P}$ resultará. En términos particulares, una solución para la matriz de renderizado húmeda puede encontrarse por $\mathbf{PPS}^* = (\mathbf{YY}^* - \mathbf{GXX}^*\mathbf{G}^*) / V = (\mathbf{AEA}^* - \mathbf{GDED}^*\mathbf{G}^*) / V$, donde la opción de $\mathbf{P}$ es preferentemente basada en la consideración de valor propio ya indicada antes de que con respecto a la segunda alternativa, y V sea $\mathbf{WEW}^* + e$. La última determinación de $\mathbf{P}$ también es hecha por la unidad de procesamiento de datos de parámetro SAOC 42.

55 [000101] Las matrices determinadas $\mathbf{G}$ y $\mathbf{P}$ son usados luego por las unidades de renderización húmedas y secas como descritas antes.

[000102] Si se requiere una baja versión de complejidad, la siguiente etapa debe reemplazar incluso esta solución con una solución sin la renderización húmeda. Un método preferido de lograr esto es reducir los requerimientos en la

covariancia compleja para emparejar sólo en la diagonal, el que las energías de señal correctas todavía son logradas en los canales izquierdos y derechos, pero la covariancia cruzada se deja abierta.

[000103] En cuanto a las primeras pruebas de capacidad auditiva alternativas, subjetivas se llevan a cabo en un cuarto de capacidad auditiva acústicamente aislado que es diseñado para permitir la escucha de alta calidad. El resultado es detallado después.

[000104] El pregrabado se lleva a cabo usando audífonos (STAX SR Lambda Pro con la gente de la laca Convertidor de D/A y SRM-Monitor de STAX). El método de prueba procedimientos de norma seguidos usados en las pruebas de verificación de audio espaciales, basadas en "Múltiple Estímulo con Referencia Escondida y Anclas" (MUSHRA) método para el asesoramiento subjetivo de audio de calidad intermedio.

[000105] Un total de 5 oyentes participó en cada una de las pruebas llevadas a cabo. Todos los sujetos pueden considerarse como oyentes experimentados. De acuerdo con la metodología MUSHRA, los oyentes son con instrucciones para comparar todas las condiciones de prueba con la referencia. Las condiciones de prueba son aleatorizadas automáticamente para cada artículo de prueba y para cada oyente. Respuestas subjetivas son registradas por un programa de MUSHRA asistido por ordenador en una escala en los límites 0 a 100. Una conmutación instantánea entre los artículos bajo la prueba es permitida. Las pruebas de MUSHRA se han llevado a cabo para evaluar el desempeño perceptual del procesamiento de estéreo a binaural descrito del MPEG sistema de SAOC.

[000106] A fin de evaluar una ganancia de calidad perceptual del sistema descrito comparado con el mono al desempeño binaural, los artículos procesados por el mono al sistema binaural también son incluidos en la prueba. La correspondencia mono y señales de conversión de multicanal a estéreo y mono correspondientes son codificadas por AAC a 80 kbits por segundo y por canal.

[000107] Ya que la base de datos de HRTF "KEMAR_MIT_COMPACT" se usa. La condición de referencia se ha generado por filtrar binaural de objetos con las respuestas a impulsos HRTF apropiadamente ponderadas que tienen en cuenta la renderización deseada. La condición de ancla es la baja pasada condición de referencia filtrada (a 3.5 kilohercios).

[000108] La Tabla 1 contiene la lista de los artículos de audio analizados.

Tabla 1 - Artículos de audio de las pruebas de capacidad auditiva

Artículos para capacidad auditiva	Objetos Nr. mono/estéreo	Ángulos de objeto / Ganancias de objeto (dB)
disco1 disco2	10/0	[-30, 0, -20, 40, 5, -5, 120, 0, -20, -40] [-3, -3, -3, -3, -3, -3, -3, -3, -3, -3] [-30, 0, -20, 40, 5, -5, 120, 0, -20, -40] [-12, -12, 3, 3, -12, -12, 3, -12, 3, -12]
coffee1 coffee2	6/0	[10, -20, 25, -35, 0, 120] [0, -3, 0, 0, 0, 0] [10, -20, 25, -35, 0, 120] [3, -20, -15, -15, 3, 3]
pop2	1/5	[0, 30, -30, -90, 90, 0, 0, -120, 120, -45, 45] [4, -6, -6, 4, 4, -6, -6, -6, -6, -16, -16]

[000109] Cinco diferentes escenas se han analizado, que son el resultado de la renderización (mono o estéreo) objetos de 3 diferentes grupos de la fuente de objeto. Tres diferentes matrices de conversión de multicanal a estéreo se han aplicado en el codificador SAOC, ver la Tabla. 2.

Tabla 2 - Tipos de conversión de multicanal a estéreo

Tipo de conversión de multicanal a estéreo	Mono	Stereo	Dual mono
Notación Matlab	dmx1=unos(1,N);	dmx2=ceros(2,N); dmx2(1,1:2:N)=1; smx2(2,2:2:N)=1;	dmx3=unos(2,N):

[000110] Las pruebas de evaluación de calidad de presentación de conversión de estéreo a multicanal se han definido como enumerado en la Tabla 3.

Tabla 3 - Condiciones de prueba para capacidad auditiva

Condición de texto	Tipo de conversión de multicanal a estéreo	Codificador central
x-1-b	Mono	AAC@80kbps
x-2-b	Estéreo	AAC@160kbps
x-2-b Dual/Mono	Dual Mono	AAC@160kbps
5222	Estéreo	AAC@160kbps
5222 DualMono	Dual Mono	AAC@160kbps

[000111] El sistema "5222" usa el preprocesador de conversión estereofónica de multicanal a estéreo como descrito en ISO/IEC JTC 1/Consejero principal 29/WG 11 (MPEG), el Documento N10045, "CD ISO/IEC Codificación espacial del objeto de audio 23003-2:200x (SAOC)", 85^{to} MPEG Hacer contacto, julio de 2008, Hannover, Alemania, con el complejo matriz de renderización con especificidad de objetivo binaural valorada $A^{l,m}$ como una entrada. Es decir ningún control de ICC se lleva a cabo. La prueba de capacidad auditiva informal ha mostrado que considerando la magnitud de $A^{l,m}$ para bandas superiores en vez de mantener ella el complejo valorado por todas las bandas mejora el desempeño. El mejorado "5222" sistema se ha usado en la prueba.

[000112] Una descripción corta en términos de diagramas que demuestran los resultados de pruebas de capacidad auditiva obtenidos puede encontrarse en la Figura 6. Estos complots muestran MUSHRA promedio que clasifica por artículo sobre todos los oyentes y el valor promedio estadístico sobre todos los artículos evaluados conjuntamente con los intervalos de confianza del 95 % asociados. Habría que observar que los datos para la referencia escondida se omiten en los complots de MUSHRA porque todos los sujetos lo han identificado correctamente.

[000113] Las siguientes observaciones pueden hacerse basadas mediante los resultados de las pruebas de capacidad auditiva:

- "x-2-b_DualMono" lleva a cabo comparable a "5222".
- "x-2-b_DualMono" lleva a cabo claramente mejor que "5222_DualMono".
- "x-2-b_DualMono" lleva a cabo comparable a "x-1-b"
- "x-2-b" puesto en práctica según la primera alternativa superior, lleva a cabo ligeramente mejor que todas otras condiciones.
- el artículo "disco1" no muestra mucha variación en los resultados y puede no ser adecuado.

[000114] Así, un concepto para la renderización binaural de señales de conversión estereofónica de multicanal a estéreo en SAOC se ha descrito antes, que cumple requisitos para de diferentes matrices de conversión de multicanal a estéreo. Particularmente la calidad para el dual mono como conversiones de multicanal a estéreo es igual que para conversiones monoaurales de multicanal a estéreo verdaderas que ha sido verificado en una prueba de capacidad auditiva. La mejora de la calidad que puede obtenerse de conversiones estereofónicas de multicanal a estéreo comparado con conversiones monoaurales de multicanal a estéreo también puede observarse de la prueba de capacidad auditiva. Los bloques de procesamiento básicos de las modalidades anteriores son la renderización binaural seca de la conversión estereofónica de multicanal a estéreo y el mezclado con una señal binaural húmeda con una combinación apropiada de ambos bloques.

●En términos particulares, la señal binaural húmeda se calcula usando un decorrelacionador con la entrada de conversión monoaural de multicanal a estéreo de modo que las energías derechas e izquierdas y el IPD son iguales como en la señal binaural seca.

●El mezclado de las señales binaurales húmedas y secas se controla por ICC con especificidad de objetivo y el ICC de la señal binaural seca de modo que la decorrelación por lo común menos sea requerida que para la conversión monoaural de multicanal a estéreo renderización binaural basada calidad de sonido total dando como resultado más elevada.

●Adicionalmente, las modalidades anteriores, puede ser fácilmente modificado para cualquier combinación de entrada mono / entrada de conversión estereofónica de multicanal a estéreo y salida mono/estéreo/binaural en una manera estable.

[000115] En otras palabras, las modalidades que proporcionan una estructura de procesamiento de señal y método para decodificación y renderización binaural de la conversión estereofónica de multicanal a estéreo trenes de bits SAOC basados con el control de coherencia entre canales se describen antes. Todas las combinaciones de mono o entrada de conversión estereofónica de multicanal a estéreo y salida mono, estéreo o binaural pueden ser manejadas como casos especiales de la conversión estereofónica de multicanal a estéreo descrita concepto basado. La calidad de la conversión estereofónica de multicanal a estéreo el concepto basado resultó ser por lo común mejor que la conversión monoaural de multicanal a estéreo el concepto basado que es verificado en MUSHRA anteriormente descrito que escucha la prueba.

[000116] En la Codificación espacial del objeto de audio (SAOC) ISO/IEC JTC 1/Consejero principal 29/WG 11 (MPEG), el Documento N10045, "el CD ISO/IEC Codificación espacial del objeto de audio 23003-2:200x (SAOC)", el

85 Congreso MPEG, julio de 2008, Hannover, Alemania, se convierten múltiples objetos de audio de multicanal a una señal mono o estéreo. Esta señal es codificada y transmitida conjuntamente con la información conexas (parámetros de SAOC) al decodificador SAOC. Las modalidades anteriores habilitan la coherencia entre canales (ICC) de la señal de salida binaural que es una medida importante para la percepción de la anchura de fuente de sonido virtual, y que es, debido a la conversión de multicanal a estéreo codificador, degradada o incluso destruyó, (casi) completamente para corregirse.

[000117] Las entradas al sistema son la conversión estereofónica de multicanal a estéreo, parámetros de SAOC, información de renderización espacial y una base de datos HRTF. La salida es la señal binaural. Ambas entrada y salida se les proporciona en el dominio de transformación de decodificador por lo común por medio de un complejo sobremuestreado el banco de filtros de análisis modulado, como el banco de filtros QMF de híbrido de Sonido envolvente de MPEG, ISO/IEC 23003-1:2007, Tecnología de la información - tecnologías de audio de MPEG - la Parte 1: Sonido envolvente de MPEG con suficientemente baja distorsión por plegado del espectro intrabanda. La señal de salida binaural se convierte nuevamente al dominio temporal de PCM por medio del banco de filtros de síntesis. El sistema es así, en otras palabras, una extensión de una conversión monoaural de multicanal a estéreo potencial renderización binaural basada hacia señales de conversión estereofónica de multicanal a estéreo. Ya que la conversión monoaural de multicanal a estéreo dual señala que la salida del sistema es igual que para la conversión monoaural de multicanal a estéreo el sistema basado. Por lo tanto el sistema puede manejar cualquier combinación de entrada mono / entrada de conversión estereofónica de multicanal a estéreo y salida mono/estéreo/binaural fijando los parámetros de renderización apropiadamente en una manera estable.

[000118] En incluso otras palabras, las modalidades anteriores llevan a cabo la renderización binaural y la decodificación de la conversión estereofónica de multicanal a estéreo trenes binarios SAOC basados con el control de ICC. Comparado con una conversión monoaural de multicanal a estéreo renderización binaural basada, las modalidades pueden aprovechar la conversión estereofónica de multicanal a estéreo de dos modos:

- Las propiedades de correlación entre objetos en diferentes canales de conversión de multicanal a estéreo son en parte conservadas
- Objetar a que la extracción sea mejorada ya que pocos objetos se encuentran en un canal de conversión de multicanal a estéreo

[000119] Así, un concepto para la renderización binaural de señales de conversión estereofónica de multicanal a estéreo en SAOC se ha descrito antes de que esto cumpla requisitos para de diferentes matrices de conversión de multicanal a estéreo. En términos particulares, la calidad para el dual mono como conversiones de multicanal a estéreo es igual que para conversiones monoaurales de multicanal a estéreo verdaderas que ha sido verificado en una prueba de capacidad auditiva. La mejora de la calidad que puede obtenerse de conversiones estereofónicas de multicanal a estéreo comparado con conversiones monoaurales de multicanal a estéreo también puede observarse de la prueba de capacidad auditiva. Los bloques de procesamiento básicos de las modalidades anteriores son la renderización binaural seca de la conversión estereofónica de multicanal a estéreo y el mezclado con una señal binaural húmeda con una combinación apropiada de ambos bloques. En términos particulares, la señal binaural húmeda se calcula usando un decorrelacionador con la entrada de conversión monoaural de multicanal a estéreo de modo que las energías derechas e izquierdas y el IPD son iguales como en la señal binaural seca. El mezclado de las señales binaurales húmedas y secas se controla por ICC con especificidad de objetivo y la conversión monoaural de multicanal a estéreo la renderización binaural basada calidad de sonido total dando como resultado más elevada. Adicionalmente, las modalidades anteriores pueden ser fácilmente modificadas para cualquier combinación de entrada mono / entrada de conversión estereofónica de multicanal a estéreo y salida mono/estéreo/binaural en una manera estable. De acuerdo con las modalidades, la señal de conversión estereofónica de multicanal a estéreo $X^{n,k}$ se obtiene conjuntamente con los parámetros SAOC, usuario información de renderización definida y una base de datos HRTF como entradas. Los parámetros SAOC transmitidos son $OLD_i^{l,m}$ (objeto diferencias de nivel), $IOC_{ij}^{l,m}$ (correlación cruzada entre objetos), $DMG_i^{l,m}$ (ganancias de conversión de multicanal a estéreo) y $DCLD_i^{l,m}$ (diferencias de nivel de canal de conversión de multicanal a estéreo) para todo N se opone i, j . Los parámetros

HRTF se les proporcionan como $P_{q,L}^m, P_{q,R}^m$ y ϕ_q^m para todo el índice q de base de datos HRTF, que se asocia con una cierta posición de la fuente de sonido espacial.

[000120] Finalmente, esto se observa que aunque dentro de la descripción anterior, los términos "entre canales de coherencia" y "correlación cruzada entre objetos" se hayan construido diferentemente en aquella "la coherencia" se usa en un término y "se cruzan la correlación" se usa en el otro, los últimos términos se pueden usar de modo indistinto como una medida para semejanzas entre canales y objetos, respectivamente.

[000121] Según una realización actual, el concepto de renderización binaural inventivo puede ponerse en práctica en el hardware o en el software. Por lo tanto, la presente invención también se relaciona con un programa informático, que puede almacenarse en un medio legible por computadora, como un CD, un disco, DVD, una barra de memoria, una tarjeta de memoria o un microcircuito de memoria. La presente invención es, por lo tanto, también un programa informático que tiene un código de programación que, cuando ejecutado en una computadora, lleva a cabo el método inventivo de codificar, convirtiendo o decodificando descrito en relación con las figuras anteriores. Más aún, esto se observa que todas las etapas indicadas en los organigramas se ponen en práctica por el

respectivo medio en el decodificador, respectivamente, un que las realizaciones pueden comprender subrutinas que funcionan en una unidad central de proceso, las partes de circuito de un ASIC o lo similar. Una declaración similar es verdad para las funciones de los bloques en los diagramas de bloques

- 5 **[000122]** En otras palabras, según una modalidad un aparato para la renderización binaural de una señal de audio multicanal (21) en una señal de salida binaural (24) se proporciona, la señal de audio multicanal (21) que comprende una señal (18) de conversión estereofónica de multicanal a estéreo hacia donde una pluralidad de señales de audio (14₁-14_N) se convierte de multicanal a estéreo, e información conexa (20) que comprende una información de conversión de multicanal a estéreo (DMG, DCLD) indicación, para cada señal de audio, hasta qué punto la
- 10 respectiva señal de audio se ha mezclado en un primer canal (L0) y un segundo canal (R0) de la señal (18) de conversión estereofónica de multicanal a estéreo, respectivamente, así como información sobre el nivel de objeto (ANTERIOR) de la pluralidad de señales de audio e información de correlación cruzada entre objetos (IOC) descripción de semejanzas entre pares de señales de audio de la pluralidad de señales de audio, el aparato que comprende el medio (47) para la informática, basada en una primera prescripción de renderización ($G^{l,m}$) según la
- 15 información de correlación cruzada entre objetos, la información sobre el nivel de objeto, la información de conversión de multicanal a estéreo, información de renderización que relaciona cada señal de audio con una posición de altavoz virtual y parámetros HRTF, una señal (54) de salida binaural preliminar de los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo; el medio (50) para generar una señal no correlacionada ($X_d^{n,k}$) como un equivalente perceptual con una conversión monoaural de multicanal a
- 20 estéreo (58) de los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo que es, sin embargo, decorrelacionada a la conversión monoaural de multicanal a estéreo (58); medio (52) para informática, según una segunda prescripción de renderización ($P_2^{l,m}$) según la información de correlación cruzada entre objetos, la información sobre el nivel de objeto, la información de conversión de multicanal a estéreo, la información de renderización y los parámetros HRTF, una señal (64) de salida binaural correctiva de la señal (62)
- 25 no correlacionada; y medio (53) para mezclado la señal (54) de salida binaural preliminar con la señal (64) de salida binaural correctiva de obtener la señal de salida binaural (24).

Referencias

- 30 **[000123]**
- ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Document N10045, "ISO/IEC CD 23003-2:200x Spatial Audio Object Coding (SAOC)", 85th MPEG Meeting, July 2008, Hannover, Alemania.
- 35 EBU Technical recommendation: "MUSHRA-EBU Method for Subjective Listening Tests of Intermediate Audio Quality", Doc. B/AIM022, October 1999.
- ISO/IEC 23003-1:2007, Information technology - MPEG audio technologies - Part 1: MPEG Surround.
- 40 ISO/IEC JTC1/SC29/WG11 (MPEG), Document N9099: "Final Spatial Audio Object Coding Evaluation Procedures and Criterion". Abril de 2007, San Jose, EUA.
- Jeroen, Breebaart, Christof Faller: Spatial Audio Processing. MPEG Surround and Other Applications. Wiley & Sons, 2007.
- 45 Jeroen, Breebaart et al.: Multi-Channel goes Mobile : MPEG Surround Binaural Rendering. AES 29th International Conference, Seúl, Corea, 2006.

REIVINDICACIONES

1. Un aparato para la renderización binaural de una señal de audio multicanal (21) en una señal de salida binaural (24), la señal de audio multicanal (21) comprende una señal (18) de conversión estereofónica de multicanal a estéreo hacia donde una pluralidad de señales de audio (14₁-14_N) se convierte de multicanal a estéreo, e información conexa (20) que comprende una información de conversión de multicanal a estéreo (DMG, DCLD) que indica, para cada señal de audio, hasta qué punto la respectiva señal de audio se ha mezclado en un primer canal (L0) y un segundo canal (R0) de la señal (18) de conversión estereofónica de multicanal a estéreo, respectivamente, así como información sobre el nivel de objeto (ANTERIOR) de la pluralidad de señales de audio e información de correlación cruzada entre objetos (IOC) que describe las semejanzas entre pares de señales de audio de la pluralidad de señales de audio, el aparato estando configurado para:

computar (47), basado en una primera prescripción de renderización ($G^{l,m}$) según la información de correlación cruzada entre objetos, la información sobre el nivel de objeto, la información de conversión de multicanal a estéreo, información de renderización que relaciona cada señal de audio con una posición de altavoz virtual y parámetros HRTF, una señal (54) de salida binaural preliminar de los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo;

generar (50), a partir de la señal (18) de conversión estereofónica de multicanal a estéreo, una señal no correlacionada ($X_d^{n,k}$) como un equivalente perceptual con una conversión monoaural de multicanal a estéreo (58) de los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo que es, sin embargo, decorrelacionada a la conversión monoaural de multicanal a estéreo (58);

computar (52), según una segunda prescripción de renderización ($P_2^{l,m}$) según la información de correlación cruzada entre objetos, la información sobre el nivel de objeto, la información de conversión de multicanal a estéreo, la información de renderización y los parámetros HRTF, una señal (64) de salida binaural correctiva de la señal (62) no correlacionada; y

mezclar (53) la señal (54) de salida binaural preliminar con la señal (64) de salida binaural correctiva de obtener la señal de salida binaural (24).

2. Aparato según la reivindicación 1, en donde el aparato está configurado adicionalmente para, en la generación de la señal no correlacionada ($X_d^{n,k}$), sumar el primer y segundo canal de la señal (18) de conversión estereofónica de multicanal a estéreo y decorrelacionar la suma para obtener la señal (62) no correlacionada.

3. Aparato según la reivindicación 1 o 2, adicionalmente configurado para:

estimar (80) un valor de coherencia entre canales binaural actual de la señal (54) de salida binaural preliminar;

determinar (82) un valor de coherencia entre canales binaural con especificidad de objetivo; y

conformar (84) una proporción de mezcla determinante a cual grado la señal de salida binaural (24) es bajo la influencia de los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo como procesado por el cálculo (47) de la señal (54) de salida binaural preliminar y los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo como procesado por la generación (50) de una señal no correlacionada y el cálculo (52) de la señal (64) de salida binaural correctiva, respectivamente, basada en el valor de coherencia entre canales binaural actual y el valor de coherencia entre canales binaural con especificidad de objetivo.

4. Aparato según la reivindicación 3, en donde el aparato está configurado adicionalmente para, en la configuración de la proporción de mezcla, determinar la proporción de mezcla determinando la primera prescripción de renderización ($G^{l,m}$) y la segunda prescripción de renderización ($P_2^{l,m}$) basado en el valor de coherencia entre canales binaural actual y el valor de coherencia entre canales binaural con especificidad de objetivo.

5. Aparato según la reivindicación 3 o 4, en donde el aparato está configurado adicionalmente para, al determinar el valor de coherencia entre canales binaural con especificidad de objetivo, llevar a cabo la determinación basada en componentes de una matriz de covarianza con especificidad de objetivo $F = A E A^*$, con "*" denotando transpuesto conjugado, A siendo una matriz de renderización binaural con especificidad de objetivo que relaciona las señales de audio con los primeros y segundos canales de la señal de salida binaural, respectivamente, y estando exclusivamente determinada por la información de renderización y los parámetros HRTF, y E siendo una matriz que es exclusivamente determinada por la información de correlación cruzada entre objetos y la información sobre el nivel de objeto.

6. Aparato según la reivindicación 5, en donde el aparato está configurado adicionalmente para, al computar la señal (54) de salida binaural preliminar, llevar a cabo el cálculo de modo que:

$$\hat{X}_1 = G \cdot X$$

en donde X es un vector 2x1 cuyos componentes corresponden a los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo, $\hat{X}_1$ es un vector 2x1 cuyos componentes corresponden a los

primeros y segundos canales de la señal (54) de salida binaural preliminar, $\mathbf{G}$ es una primera matriz de renderización que representa la primera prescripción de renderización y que tiene un tamaño de 2x2 con:

$$\mathbf{G} = \begin{pmatrix} P_L^1 \cos(\beta + \alpha) \exp(j \frac{\phi^1}{2}) & P_L^2 \cos(\beta + \alpha) \exp(j \frac{\phi^2}{2}) \\ P_R^2 \cos(\beta - \alpha) \exp(-j \frac{\phi^1}{2}) & P_R^2 \cos(\beta - \alpha) \exp(-j \frac{\phi^2}{2}) \end{pmatrix}$$

donde, con $x \in \{1,2\}$:

$$P_L^x = \sqrt{\frac{f_{11}^x}{V^x}},$$

$$\phi^x = \begin{cases} \arg(f_{12}^x) & \text{Si aplica una condición} \\ 0 & \text{En caso contrario} \end{cases}$$

en donde f_{11}^x, f_{12}^x y f_{22}^x son coeficientes de matrices de covariancia con especificidad de objetivo $\mathbf{F}^x$ del tamaño 2x2 con: $\mathbf{F}^x = \mathbf{A} \mathbf{E}^x \mathbf{A}^*$,

en donde: $e_{ij}^x = e_{ij} \left(\frac{d_i^x}{d_i^1 + d_i^2} \right) \left(\frac{d_j^x}{d_j^1 + d_j^2} \right)$ son coeficientes de la matriz de NxN $\mathbf{E}^x$, N siendo la cantidad de señales de

15 audio, e_{ij} son coeficientes de la matriz $\mathbf{E}$ siendo de tamaño NxN, y d_i^x son exclusivamente determinados por la información de conversión de multicanal a estéreo, donde d_i^1 indica el grado al cual la señal de audio i ha sido mezclada dentro del primer canal de la señal (18) de conversión estereofónica de multicanal a estéreo y d_i^2 define hasta qué punto la señal de audio i ha sido mezclada dentro del segundo canal de la señal de salida estéreo (18),

20 en donde V^x es un escalar con $V^x = \mathbf{D}^x \mathbf{E} (\mathbf{D}^x)^* + \epsilon$ y $\mathbf{D}^x$ es una matriz 1xN cuyos coeficientes son d_i^x , en donde el aparato está configurado adicionalmente para, al computar una señal de salida binaural correctiva (64), llevar a cabo el cálculo de modo que:

$$\hat{X}_2 = \mathbf{P}_2 \cdot X_d$$

25 en donde X_d es la señal no correlacionada, $\hat{X}_2$ es un vector 2x1 cuyos componentes corresponden a primeros y segundos canales de la señal (64) de salida binaural correctiva, y $\mathbf{P}_2$ es una segunda matriz de renderización que representa la segunda prescripción de renderización y que tiene un tamaño 2x2 con:

$$\mathbf{P}_2 = \begin{pmatrix} P_L \sin(\beta + \alpha) \exp(j \frac{\arg(c_{12})}{2}) \\ P_R \sin(\beta - \alpha) \exp(-j \frac{\arg(c_{12})}{2}) \end{pmatrix}$$

en donde P_L y P_R de ganancias se definen como:

$$P_L = \sqrt{\frac{c_{11}}{V}}, P_R = \sqrt{\frac{c_{22}}{V}}$$

30 en donde c_{11} y c_{22} son coeficientes de una matriz 2x2 $\mathbf{C}$ de covariancia de la señal (54) de salida binaural preliminar con:

$$\mathbf{C} = \tilde{\mathbf{G}} \mathbf{D} \mathbf{E} \mathbf{D}^* \tilde{\mathbf{G}}^*$$

en donde V es un escalar con $V = \mathbf{W} \mathbf{E} \mathbf{W}^* + \epsilon$, $\mathbf{W}$ es una matriz de conversión monoaural de multicanal a

estéreo del tamaño 1xN cuyos coeficientes son exclusivamente determinados por: $d_i^x \mathbf{D} = \begin{pmatrix} d_i^1 \\ d_i^2 \end{pmatrix}_y \tilde{\mathbf{G}}$

$$\tilde{\mathbf{G}}^{l,m} = \begin{pmatrix} P_L^1 \exp(j \frac{\phi^1}{2}) & P_L^{l,m,2} \exp(j \frac{\phi^2}{2}) \\ P_R^1 \exp(-j \frac{\phi^1}{2}) & P_R^2 \exp(-j \frac{\phi^2}{2}) \end{pmatrix}$$

es

35 en donde el aparato está configurado adicionalmente para, al estimar el valor de coherencia entre canales binaural actual, determinar el valor de coherencia entre canales binaural actual como:

$$\rho_C = \min \left(\frac{|c_{12}|}{\sqrt{c_{11}c_{22}}}, 1 \right)$$

en donde el aparato está configurado adicionalmente para, al determinar el valor de coherencia entre canales binaural con especificidad de objetivo, determinar el valor de coherencia entre canales binaural con especificidad de objetivo como:

$$\rho_T = \min\left(\frac{|f_{12}|}{\sqrt{f_{11}f_{22}}}, 1\right), y$$

- 5 en donde el aparato está configurado adicionalmente para, en la configuración de la proporción de mezcla, determinar los ángulos de los rotadores α y β según:

$$\alpha = \frac{1}{2}(\arccos(\rho_T) - \arccos(\rho_C)),$$

$$\beta = \arctan\left(\tan(\alpha) \frac{P_R - P_L}{P_L + P_R}\right),$$

con e denotando una constante pequeña para evitar divisiones por cero, respectivamente.

- 10 7. Aparato según la reivindicación 1, donde el aparato está configurado adicionalmente para, al computar la señal (54) de salida binaural preliminar, llevar a cabo el cálculo de modo que:

$$\hat{X}_1 = G \cdot X$$

en donde X es un vector 2x1 cuyos componentes corresponden a los primeros y segundos canales de la señal (18)

- 15 de conversión estereofónica de multicanal a estéreo, $\hat{X}_1$ es un vector 2x1 cuyos componentes corresponden a los primeros y segundos canales de la señal (54) de salida binaural preliminar, G es una primera matriz de renderización que representa la primera prescripción de renderización y que tiene un tamaño de 2x2 con:

$$G = AED^*(DED^*)^{-1},$$

- 20 en donde E es una matriz siendo exclusivamente determinada por la información de correlación cruzada entre objetos y la información sobre el nivel de objeto;

D es una matriz 2xN cuyos coeficientes d_{ij} están exclusivamente determinados por la información de conversión de multicanal a estéreo, en donde d_{1j} indica el grado al cual la señal de audio j se ha mezclado dentro del primer canal de la señal (18) de conversión estereofónica de multicanal a estéreo y d_{2j} define hasta qué punto la señal de audio j se ha mezclado dentro del segundo canal de la señal de salida estéreo (18);

- 25 A es una matriz de renderización binaural con especificidad de objetivo relación de las señales de audio a los primeros y segundos canales de la señal de salida binaural, respectivamente, y es exclusivamente determinada por la información de renderización y los parámetros HRTF,

- 30 en donde el aparato está configurado adicionalmente para, al computar una señal de salida binaural correctiva (64), llevar a cabo el cálculo de modo que:

$$\hat{X}_2 = P \cdot X_d$$

en donde X_d es la señal no correlacionada, $\hat{X}_2$ es un vector 2x1 cuyos componentes corresponden a primeros y segundos canales de la señal (64) de salida binaural correctiva, y P es una segunda matriz de renderización que representa la segunda prescripción de renderización y que tiene un tamaño 2x2 y se determina de modo que: PPS *

- 35 $= \Delta R$, con $\Delta R = AEA^* - G_0 DED^* G_0^*$ con $G_0 = G$.

8. Aparato según la reivindicación 1, en donde el aparato está configurado adicionalmente para, al calcular la señal (54) de salida binaural preliminar, llevar a cabo el cálculo de modo que:

$$\hat{X}_1 = G \cdot X$$

- 40 en donde X es un vector 2x1 cuyos componentes corresponden a los primeros y segundos canales de la señal (18)

de conversión estereofónica de multicanal a estéreo, $\hat{X}_1$ es un vector 2x1 cuyos componentes corresponden a los primeros y segundos canales de la señal (54) de salida binaural preliminar, G es una primera matriz de renderización que representa la primera prescripción de renderización y que tiene un tamaño de 2x2 con:

$$G = (G_0 DED^* G_0^*)^{-1} (G_0 DED^* G_0^* AEA^* G_0 DED^* G_0^*)^{1/2} (G_0 DED^* G_0^*)^{-1} G_0 \text{ con } G_0 = AED^* (DED^*)^{-1}$$

- 45 en donde E es una matriz que es exclusivamente determinada por la información de correlación cruzada entre objetos y la información sobre el nivel de objeto;

D es una matriz $2 \times N$ cuyos coeficientes d_{ij} son exclusivamente determinados por la información de conversión de multicanal a estéreo, en donde d_{1j} indica el grado al cual la señal de audio j se ha mezclado dentro del primer canal de la señal (18) de conversión estereofónica de multicanal a estéreo y d_{2j} define hasta qué punto la señal de audio j se ha mezclado dentro del segundo canal de la señal de salida estéreo (18);

5 A es una matriz de renderización binaural con especificidad de objetivo relación de las señales de audio a los primeros y segundos canales de la señal de salida binaural, respectivamente, y es exclusivamente determinado por la información de renderización y los parámetros HRTF, en donde el aparato está configurado adicionalmente para, al computar una señal de salida binaural correctiva (64), lleve a cabo el cálculo de modo que:

$$\hat{X}_2 = P \cdot X_d$$

10 en donde X_d es la señal no correlacionada, $\hat{X}_2$ es un vector 2×1 cuyos componentes corresponden a primeros y segundos canales de la señal (64) de salida binaural correctiva, y P es una segunda matriz de renderización que representa la segunda prescripción de renderización y que tiene un tamaño 2×2 y se determina el que $PPS^* = (AEA^* - GDED^*G^*) / V$ con V que es un escalar.

15 9. Aparato según cualquiera de las reivindicaciones anteriores, en donde la información de conversión de multicanal a estéreo (DMG, DCLD) es dependiente del tiempo, y la información sobre el nivel de objeto (ANTERIOR) y la información de correlación cruzada entre objetos (IOC) es dependiente de tiempo y frecuencia.

20 10. Método para la renderización binaural de una señal de audio multicanal (21) en una señal de salida binaural (24), la señal de audio multicanal (21) que comprende una señal (18) de conversión estereofónica de multicanal a estéreo hacia donde una pluralidad de señales de audio (14₁-14_N) se convierte de multicanal a estéreo, e información conexa (20) que comprende una información de conversión de multicanal a estéreo (DMG, DCLD) indicando, para cada señal de audio, hasta qué punto la respectiva señal de audio se ha mezclado dentro de un primer canal (L0) y un segundo canal (R0) de la señal (18) de conversión estereofónica de multicanal a estéreo, respectivamente, así como información sobre el nivel de objeto (ANTERIOR) de la pluralidad de señales de audio e información de correlación cruzada entre objetos (IOC) que describe las semejanzas entre pares de señales de audio de la pluralidad de señales de audio, en donde el método comprende:

30 computar, basado en una primera prescripción de renderización ($G^{(m)}$) según la información de correlación cruzada entre objetos, la información sobre el nivel de objeto, la información de conversión de multicanal a estéreo, información de renderización que relaciona cada señal de audio con una posición de altavoz virtual y parámetros HRTF, una señal (54) de salida binaural preliminar de los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo;

35 generar, a partir de la señal (18) de conversión estereofónica de multicanal a estéreo, una señal no correlacionada ($X_d^{n,k}$) como un equivalente perceptual con una conversión monoaural de multicanal a estéreo (58) de los primeros y segundos canales de la señal (18) de conversión estereofónica de multicanal a estéreo que es, sin embargo, decorrelacionada a la conversión monoaural de multicanal a estéreo (58);

40 computar, según una segunda prescripción de renderización ($P_2^{(m)}$) según la información de correlación cruzada entre objetos, la información sobre el nivel de objeto, la información de conversión de multicanal a estéreo, la información de renderización y los parámetros HRTF, una señal (64) de salida binaural correctiva de la señal (62) no correlacionada; y mezclar la señal (54) de salida binaural preliminar con la señal (64) de salida binaural correctiva de obtener la señal de salida binaural (24).

45 11. Programa de computadora que tiene instrucciones para llevar a cabo, cuando se ejecuta en una computadora, el método según la reivindicación 10.

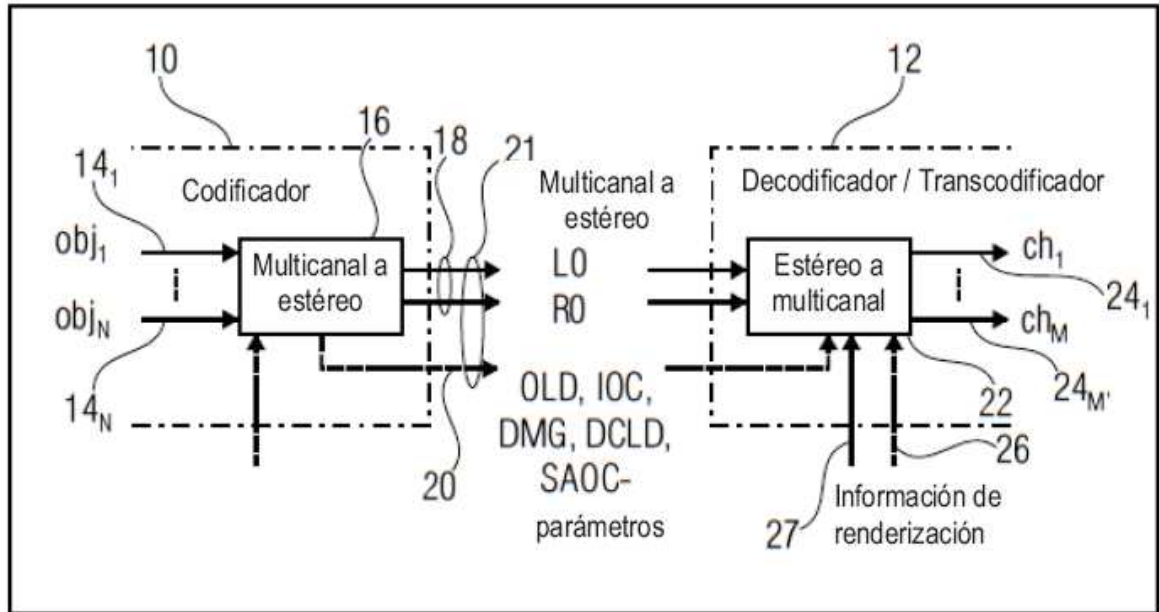


FIG 1

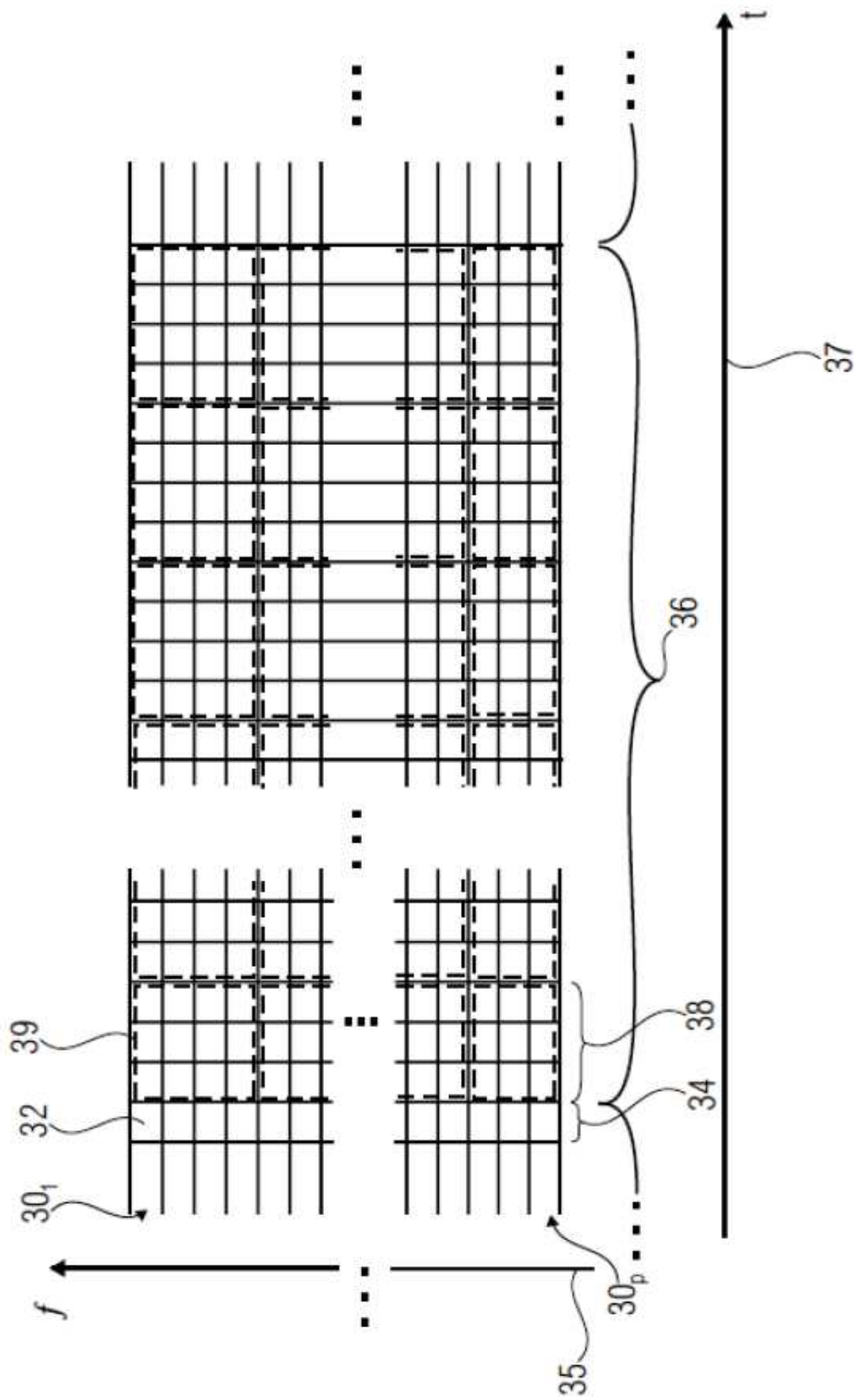


FIG 2

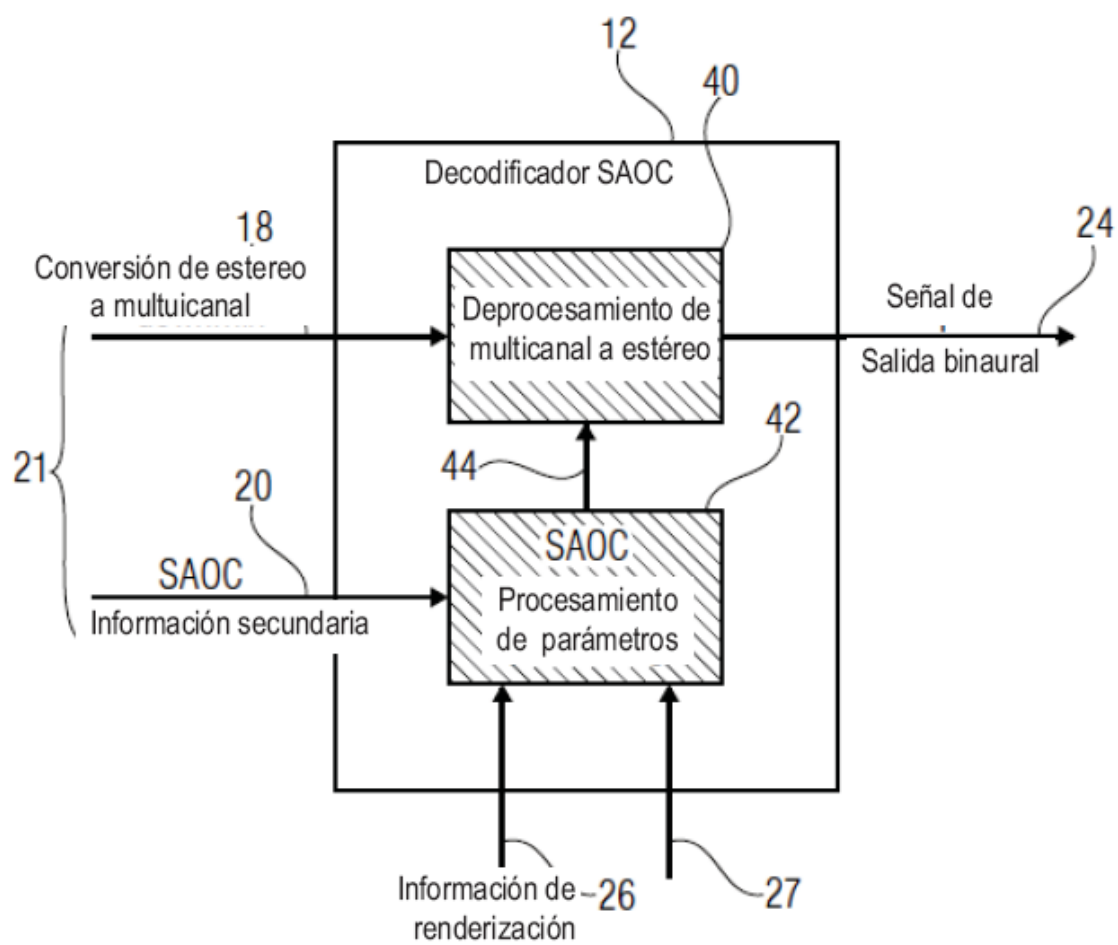


FIG 3

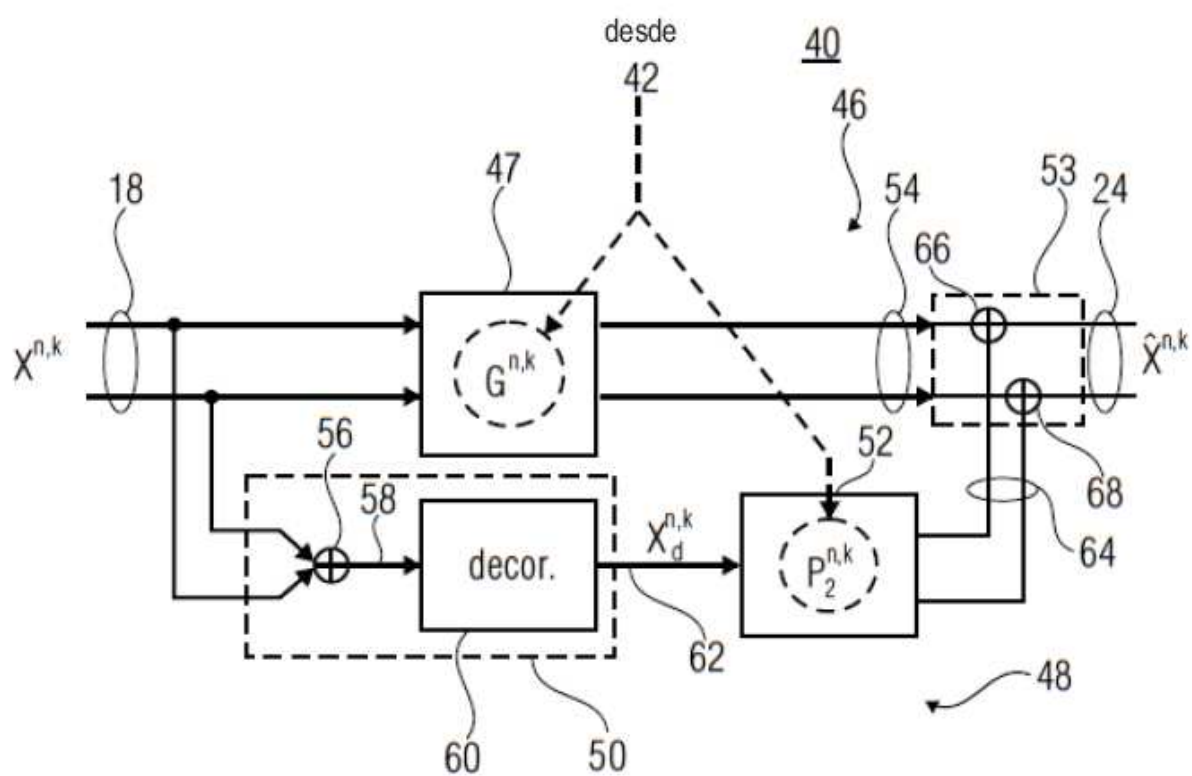


FIG 4

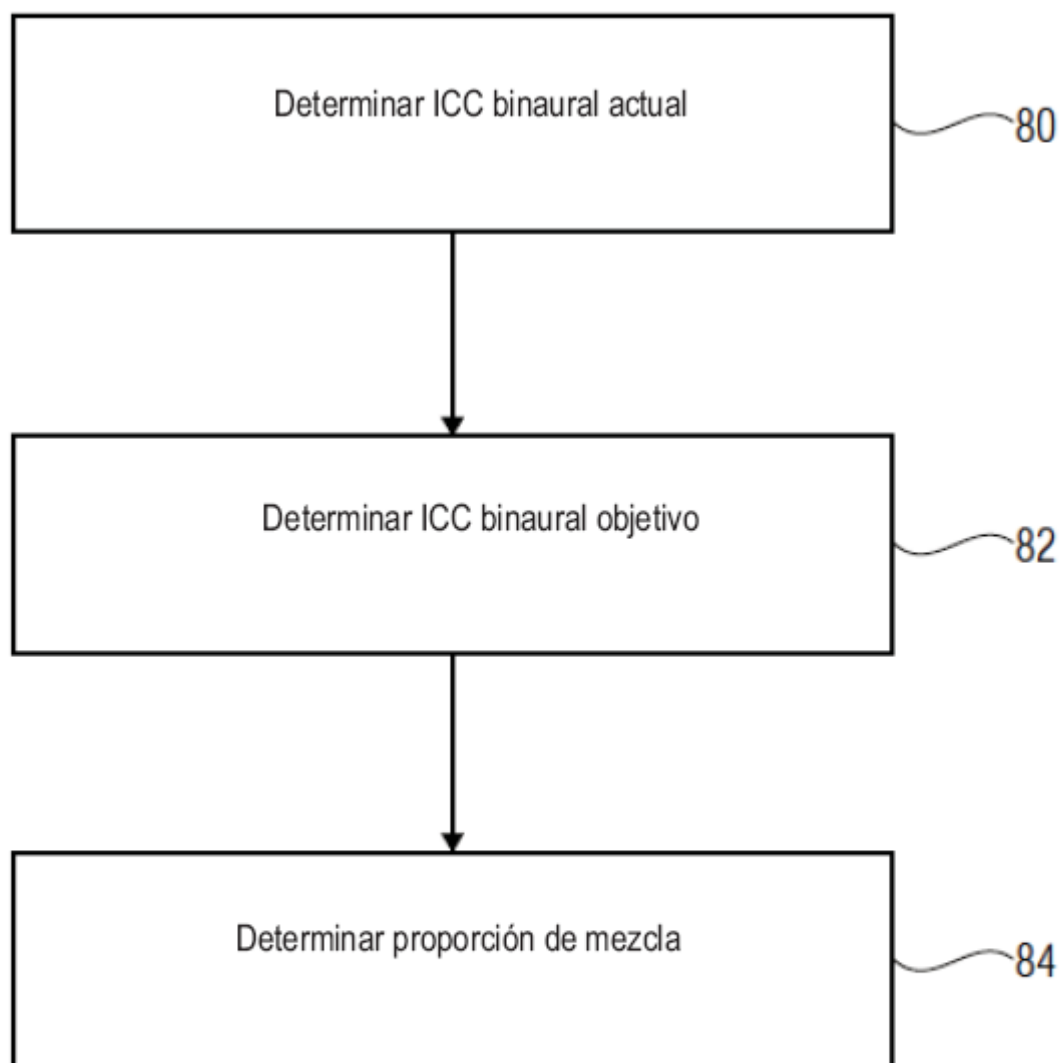


FIG 5

FIG 6

