

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 986 617**

51 Int. Cl.:

**H04S 7/00** (2006.01)

**G10L 19/008** (2013.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.12.2018** **PCT/FR2018/053161**

87 Fecha y número de publicación internacional: **27.06.2019** **WO19122580**

96 Fecha de presentación y número de la solicitud europea: **07.12.2018** **E 18833274 (6)**

97 Fecha y número de publicación de la concesión europea: **26.06.2024** **EP 3729832**

54 Título: **Procesamiento de una señal monofónica en un decodificador de audio 3d que reproduce contenido binaural**

30 Prioridad:

**19.12.2017 FR 1762478**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**12.11.2024**

73 Titular/es:

**ORANGE (100.0%)**  
**111, quai du Président Roosevelt**  
**92130 Issy-les-Moulineaux, FR**

72 Inventor/es:

**PALLONE, GRÉGORY**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

ES 2 986 617 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

Procesamiento de una señal monofónica en un decodificador de audio 3d que reproduce contenido binaural

- 5 La presente invención hace referencia al procesamiento de una señal de audio en un sistema de decodificación de audio 3D de tipo codificada normalizada MPEG-H 3D audio. La invención hace referencia más particularmente al procesamiento de una señal monofónica para ser reproducida en un auricular que también recibe señales de audio binaurales.
- 10 El término binaural hace referencia a la reproducción de una señal de sonido en cascos o un par de auriculares, pero no obstante con efectos espaciales. El procesamiento binaural de señales de audio, denominado en lo sucesivo binauralización o procesamiento de binauralización, utiliza filtros HRTF (de "Head Related Transfert Function" en inglés) en el dominio de la frecuencia o filtros HRIR, BRIR (de "Head Related Transfert Function, Binaural Room Impulse Response" en inglés) en el dominio del tiempo para reproducir las funciones de transferencia acústica entre
- 15 las fuentes de sonido y los oídos del oyente. Estos filtros se utilizan para simular los índices de localización auditiva que permiten al oyente localizar las fuentes de sonido como en una situación de escucha real.
- La señal del oído derecho se obtiene filtrando una señal monofónica utilizando la función de transferencia (HRTF) del oído derecho y la señal del oído izquierdo se obtiene filtrando la misma señal monofónica utilizando la función de
- 20 transferencia del oído izquierdo.
- En los códecs de tipo NGA (de "Next Generation Audio" en inglés), como por ejemplo, el MPEG-H 3D audio descrito en el documento indicado ISO/IEC 23008-3: "High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio" publicado el 25/07/2014 o también AC4 descrito en el documento indicado ETSI TS 103 190: "Digital Audio Compression Standard" publicado en abril de 2014, las señales recibidas en el decodificador se decodifican primero y, a continuación, se someten a un procesamiento de binauralización como el descrito anteriormente antes de ser reproducidas en unos cascos. En este caso, nos interesa el caso de la reproducción en cascos, con sonido espacializado, es decir, una señal binauralizada.
- 25
- 30 Por lo tanto, los códecs mencionados prevén la posibilidad de una reproducción en varios altavoces virtuales escuchando una señal binauralizada en cascos, pero también prevén la posibilidad de reproducir en varios altavoces reales, un sonido espacializado.
- En algunos casos, el procesamiento de binauralización se combina con una función de seguimiento de la cabeza del oyente ("Head tracking" en inglés) que se denomina renderizado dinámico, en contraposición al renderizado estático. Este procesamiento tiene en cuenta el movimiento de la cabeza del oyente para modificar la reproducción del sonido en cada oído con el fin de mantener estable la reproducción del escenario de sonido. En otras palabras, el oyente percibirá las fuentes de sonido en el mismo lugar del espacio físico mueva o no la cabeza.
- 35
- 40 Esto puede ser importante para ver y escuchar contenidos de vídeo en 360°.
- Sin embargo, para algunos contenidos no es deseable utilizar este tipo de procesamiento. De hecho, en algunos casos, cuando el contenido ha sido creado específicamente para la reproducción binaural, por ejemplo, si las señales han sido grabadas directamente por una cabeza artificial o ya han sido procesadas mediante binauralización, entonces se deben reproducir directamente en los auriculares de los cascos. Estas señales no requieren ningún procesamiento de binauralización adicional.
- 45
- Del mismo modo, un productor de contenidos puede querer que una señal sonora se reproduzca de forma independiente de la escena de sonido, es decir, que se perciba como un sonido independiente de la escena de sonido, por ejemplo, como en el caso de una voz en "OFF".
- 50
- Este tipo de reproducción puede permitir, por ejemplo, dar explicaciones de una escena de sonido reproducida de otro modo. Por ejemplo, el productor de contenidos puede querer que el sonido se reproduzca sólo en un oído para obtener un efecto deliberado de "auricular", es decir, que el sonido sólo se escuche en un oído. También puede ser deseable que el sonido permanezca permanentemente sólo en ese oído, aunque el oyente mueva la cabeza, como ocurre en el ejemplo anterior. El productor de contenidos también puede querer que este sonido se reproduzca en una posición precisa en el espacio de sonido, en relación con uno de los oídos del oyente (y no sólo en un único oído), aunque éste mueva la cabeza.
- 55
- Una señal monofónica de este tipo, decodificada e introducida en un sistema de reproducción mediante un códec de tipo MPEG-H 3D audio o AC4, se binauralizará. El sonido se distribuirá entonces entre los dos oídos (aunque será menos fuerte en el oído opuesto) y si el oyente mueve la cabeza, no percibirá el sonido de la misma forma en su oído, ya que el procesamiento de seguimiento de la cabeza, si está implementado, garantizará que la posición de la fuente de sonido siga siendo la misma que en la escena de sonido inicial: dependiendo de la posición de la cabeza, el sonido aparecerá por tanto más fuerte en uno u otro de los oídos.
- 60
- 65

En una propuesta de modificación del códec MPEG-H 3D audio, una contribución referenciada "ISO/IEC JTC1/SC29/WG11 MPEG2015/M37265" de octubre de 2015 propone identificar los contenidos que no deben ser alterados por la binauralización.

- 5 De este modo, se asocia una identificación "dicótica" a los contenidos que no deben ser procesados por la binauralización.

Todos los elementos de audio estarán entonces binauralizados excepto los referenciados como "dicóticos". "Dicótico" significa que tienes una señal diferente en cada oído.

- 10 De la misma forma, en la norma AC4, un bit de información indica que una señal ya ha sido virtualizada. Este bit desactiva el postprocesamiento. Los contenidos identificados de este modo ya están formateados para cascos, es decir, en binaural. Constan de dos canales.

- 15 Estos métodos no tratan el caso de una señal monofónica para la que el productor de la escena de sonido no desea la binauralización.

- 20 Esto significa que no es posible reproducir una señal monofónica de forma independiente de la escena de sonido, en una posición precisa con respecto al oído de un oyente, lo que denominaremos en modo "auricular". Utilizando las técnicas bicanales del estado de la técnica, una solución sería crear un contenido bicanal compuesto por una señal en uno de los canales y silencio en el otro canal para una reproducción deseada en un solo oído, o crear un contenido estereofónico teniendo en cuenta la posición espacial deseada e identificar este contenido como ya espacializado antes de transmitirlo.

- 25 Sin embargo, este tipo de procesamiento crea complejidad al crear contenido estereofónico y requiere una velocidad de transmisión de datos adicional para este contenido estereofónico.

- 30 El documento US2016/0266865 describe un método para reproducir una señal de audio en cascos utilizando parámetros de binauralización elegidos por medio de una interfaz de usuario para optimizar la reproducción en cascos.

Por lo tanto, es necesario ofrecer una solución que permita transmitir una señal que se reproducirá en una posición precisa con respecto a un oído de un usuario de cascos, de forma independiente de una escena de sonido reproducida por este mismo auricular, optimizando al mismo tiempo la velocidad de transmisión de datos de codificación utilizada.

- 35 La presente invención mejora la situación.

- 40 Para ello, propone un método de procesamiento de una señal de audio monofónica en un decodificador de audio 3D que comprende una etapa de procesamiento de la binauralización de las señales decodificadas destinadas a ser reproducidas espacialmente por cascos. El método es tal que, al detectar, en un flujo de datos representativo de la señal monofónica, una indicación de que no se está llevando a cabo el procesamiento de binauralización asociado a la información de posición espacial de reproducción, siendo la información de posición espacial de reproducción un dato binario que indica un único canal de los cascos de reproducción, la señal monofónica decodificada se envía a un módulo de mezcla que comprende un motor de renderizado estereofónico que tiene en cuenta la información de posición para construir dos canales de reproducción procesados por una etapa de mezcla directa que suma estos dos canales con una señal binauralizada resultante del procesamiento de binauralización, para ser reproducirla en los cascos.

- 45 De este modo, es posible especificar que el contenido monofónico se debe reproducir en una posición espacial precisa en relación con el oído de un oyente y que no se debe someter a un procesamiento de binauralización para que esta señal reproducida pueda tener un efecto "auricular", es decir, que sea escuchada por el oyente en una posición específica en relación con un oído, en el interior de la cabeza de la misma forma que una señal estereofónica, aunque la cabeza del oyente se mueva.

- 50 De hecho, las señales estereofónicas se caracterizan porque cada fuente de sonido está presente en cada uno de los 2 canales de salida (izquierdo y derecho) con una diferencia de intensidad (o ILD por "Interaural Level Difference") y a veces de tiempo (o ITD por "Interaural Time Difference") entre los canales. Al escuchar una señal estereofónica a través de unos cascos, las fuentes se perciben en el interior de la cabeza, en un punto situado entre el oído izquierdo y el derecho, en función de la ILD y/o la ITD. Las señales binaurales difieren de las estereofónicas en que se aplica un filtro a las fuentes para reproducir el trayecto acústico desde la fuente hasta el oído del oyente. Al escuchar una señal binaural a través de unos cascos, las fuentes se perciben fuera de la cabeza, en un punto de una esfera, dependiendo del filtro utilizado.

- 55 Las señales estereofónicas y binaurales se parecen en que están formadas por 2 canales izquierdo y derecho, y se distinguen por el contenido de estos 2 canales.

- 65

Esta señal mono (de monofónica) reproducida se superpone entonces a las demás señales reproducidas para formar una escena de sonido en 3D.

5 La velocidad de transmisión de datos necesaria para indicar este tipo de contenido se optimiza porque sólo es necesario codificar una indicación de posición en la escena de sonido además de la indicación de no binauralización para informar al decodificador del procesamiento que se debe llevar a cabo, a diferencia de un método que exigiría codificar, transmitir y, a continuación, decodificar una señal estereofónica que tenga en cuenta esta posición espacial.

10 La información de posición espacial de reproducción son datos binarios que indican un único canal de los cascos de reproducción.

Esta información sólo requiere un bit de codificación, lo que reduce aún más la velocidad de transmisión de datos necesaria.

15 Los diferentes formas de realización particulares que se mencionan a continuación se pueden añadir, independientemente o combinadas entre sí, a las etapas del método de procesamiento definido anteriormente.

20 En esta forma de realización, sólo el canal de reproducción correspondiente al canal indicado por los datos binarios se suma con el canal correspondiente de la señal binauralizada en la etapa de mezcla directa, teniendo el otro canal de reproducción un valor de cero.

La suma llevada a cabo de este modo es sencilla y proporciona el efecto "auricular" deseado, superponiendo la señal mono a la escena de sonido reproducida.

25 En una forma de realización particular, la señal monofónica es una señal de tipo canal dirigida al motor de renderizado estereofónico con la información de posición espacial de reproducción.

30 De este modo, la señal monofónica no se somete a una etapa de procesamiento de binauralización y no se procesa del mismo modo que las señales de tipo canal procesadas habitualmente por los métodos del estado de la técnica. Esta señal es procesada por un motor de renderizado estereofónico diferente al existente para las señales de tipo canal. Este motor de renderizado consiste en duplicar la señal monofónica en los 2 canales, aplicando factores que son funciones de la información sobre la posición espacial de la reproducción en los dos canales.

35 Este motor de renderizado estereofónico también se puede integrar en el motor de renderizado de canales con procesamiento diferenciado según la detección realizada para la señal a la entrada de este motor de renderizado o en el módulo de mezcla directa sumando los canales de este motor de renderizado estereofónico a la señal binauralizada procedente del módulo de procesamiento de binauralización.

40 En otra forma de realización, la señal monofónica es una señal de tipo objeto asociada a un conjunto de parámetros de reproducción que comprenden la indicación de no binauralización y la información de posición de reproducción, dirigiéndose la señal hacia el motor de renderizado estereofónico con la información de posición espacial de reproducción.

45 De este modo, la señal monofónica no se somete a una etapa de procesamiento de binauralización y no se procesa del mismo modo que las señales de tipo canal procesadas habitualmente por los métodos del estado de la técnica. Esta señal es procesada por un motor de renderizado estereofónico diferente al existente para las señales de tipo objeto. La indicación de no procesamiento de binauralización y la información de posición de reproducción se incluyen en los parámetros de reproducción (Metadatos) asociados a la señal tipo objeto. Este motor de renderizado también se puede integrar en el motor de renderizado de objetos o en el módulo de mezcla directa sumando los canales de este motor de renderizado estereofónico a la señal binauralizada procedente del módulo de procesamiento de binauralización.

50 La presente invención también hace referencia a un dispositivo de procesamiento de una señal de audio monofónica que comprende un módulo de procesamiento de binauralización de señales decodificadas destinadas a ser reproducidas espacialmente por unos cascos. Este dispositivo es tal que comprende:

60 • un módulo de detección capaz de detectar, en un flujo de datos representativo de la señal monofónica, una indicación de no procesamiento de binauralización asociada a una información de posición espacial de reproducción, siendo la información de posición espacial de reproducción datos binarios que indican un único canal de los cascos de reproducción;

• un módulo de redirección, en caso de detección positiva por el módulo de detección, capaz de dirigir la señal monofónica hacia un motor de un módulo de mezcla;

- un módulo de mezcla que comprende un motor de renderizado estereofónico capaz de tener en cuenta la información de posición para construir dos canales de reproducción, siendo el módulo de mezcla directa capaz de procesar directamente los dos canales de reproducción sumándolos con una señal binauralizada procedente del módulo de procesamiento de binauralización, para ser reproducida en los cascos.

Este dispositivo tiene las mismas ventajas que el método descrito anteriormente, que pone en práctica.

De este modo, los canales de reproducción sólo se construyen en el módulo de mezcla directa, transmitiéndose entonces únicamente la información de posición con la señal mono hasta el módulo de mezcla directa. Esta señal puede ser de tipo canal u objeto.

En una forma de realización, la señal monofónica es una señal de tipo canal y el motor de renderizado estereofónico se integra con un motor de renderizado de canales que también construye canales de reproducción para señales multicanal.

En otra forma de realización, la señal monofónica es una señal de tipo objeto y el motor de renderizado estereofónico se integra con un motor de renderizado de objetos que también construye canales de reproducción para señales monofónicas asociadas con conjuntos de parámetros de reproducción.

La presente invención hace referencia a un decodificador de audio que comprende un dispositivo de procesamiento tal como el descrito, así como un programa informático que comprende instrucciones de código para implementar las etapas del método de procesamiento como el descrito, cuando estas instrucciones son ejecutadas por un procesador.

Por último, la invención hace referencia a un medio de almacenamiento, legible por un procesador, que puede o no estar integrado en el dispositivo de procesamiento, y que puede o no ser extraíble, que almacena un programa informático que comprende instrucciones para la ejecución del método de procesamiento tal como se ha descrito anteriormente.

Otras características y ventajas de la invención quedarán más claras con la lectura de la siguiente descripción, dada únicamente a modo de ejemplo no restrictivo, y hecha con referencia a los dibujos adjuntos, en los que:

- la figura 1 ilustra un decodificador del tipo MPEG-H 3D audio tal y como existe en el estado de la técnica;
- la figura 2 ilustra las etapas de un método de procesamiento de acuerdo con una forma de realización de la invención;
- la figura 3 ilustra un decodificador que comprende un dispositivo de procesamiento de acuerdo con una primera forma de realización útil para comprender la invención;
- la figura 4 ilustra un decodificador que comprende un dispositivo de procesamiento de acuerdo con una segunda forma de realización útil para comprender la invención; y
- la figura 5 ilustra una representación hardware de un dispositivo de procesamiento de acuerdo con una forma de realización de la invención.

La **figura 1** ilustra esquemáticamente un decodificador como el normalizado en la norma MPEG-H 3D audio de acuerdo con el documento citado anteriormente. El bloque 101 es un módulo de decodificación central que descodifica señales de audio multicanal (Ch.) de tipo "canal", señales de audio monofónicas de tipo "objeto" (Obj.) asociadas a parámetros de espacialización ("Metadatos") (Obj.MeDa.) y señales de audio en formato de audio ambifónico de orden superior (HOA) (HOA de "Higher Order Ambisonic" en inglés).

Una señal de tipo canal es descodificada y procesada por un motor de renderizado de canales 102 ("Channel renderer" en inglés, también denominado "Format Converter" Conversor de Formato en MPEG-H 3D Audio) con el fin de adaptar esta señal de canal al sistema de reproducción de audio. El motor de renderizado de canales conoce las características del sistema de reproducción y, por lo tanto, proporciona una señal por canal de reproducción (Rdr.Ch.) para alimentar altavoces reales o altavoces virtuales (que luego se binauralizarán para la reproducción a través de unos cascos).

Estos canales de reproducción son mezclados por el módulo de mezcla 110 con otros canales de reproducción procedentes de los motores de renderizado de objetos 103 y HOA 105 descritos más adelante.

Las señales de tipo objeto (Obj.) son señales monofónicas asociadas a datos ("Metadatos") tales como parámetros de espacialización (ángulos acimutales, elevación) que permiten posicionar la señal monofónica en la escena de sonido espacializada, parámetros de prioridad o parámetros de volumen de sonido. Estas señales objeto son

descodificadas, junto con los parámetros asociados, por el módulo de descodificación 101 y son procesadas por un motor de renderizado de objetos 103 ("Object Renderer" en inglés) que, conociendo las características del sistema de reproducción, adapta estas señales monofónicas a dichas características. Los diferentes canales de reproducción (Rdr.Obj.) creados de este modo se mezclan con los demás canales de reproducción procedentes de los motores de renderizado de canales y HOA mediante el módulo de mezcla 110.

De la misma forma, se descodifican las señales ambifónicas de orden superior (HOA de "Higher Order Ambisonic" en inglés) y los componentes ambifónicos descodificados se introducen en un motor de renderizado ambifónico HOA 105 ("HOA renderer" en inglés) para adaptar estos componentes al sistema de reproducción de sonido.

Los canales de reproducción (Rdr.HOA) creados por este motor de renderizado HOA se mezclan en 110 con los canales de reproducción creados por los otros motores de renderizado 102 y 103.

Las señales a la salida del módulo de mezcla 110 pueden ser reproducidas por altavoces reales HP situados en una sala de reproducción. En este caso, las señales a la salida del módulo de mezcla se pueden alimentar directamente a estos altavoces reales, correspondiendo un canal a un altavoz.

Si las señales emitidas por el módulo mezclador van a ser reproducidas en cascos CA, entonces estas señales son procesadas por un módulo de procesamiento de binauralización 120 de acuerdo con las técnicas de binauralización descritas, por ejemplo, en el documento citado para la norma MPEG-H 3D audio.

De este modo, todas las señales destinadas a la reproducción en cascos son procesadas por el módulo de procesamiento de binauralización 120.

La **figura 2** describe ahora las etapas de un método de procesamiento de acuerdo con una forma de realización de la invención.

Este método hace referencia al procesamiento de una señal monofónica en un descodificador de audio 3D. Una etapa E200 detecta si el flujo de datos (SMo) que representa la señal monofónica (por ejemplo, el flujo de bits a la entrada del descodificador de audio) incluye una indicación de que no se está procesando la binauralización, asociada a información sobre la posición espacial de la reproducción. En caso contrario (N en la etapa E200), la señal debe ser binauralizada. Se procesa mediante un procesamiento de binauralización en la etapa E210, antes de ser reproducida en los cascos de reproducción en la etapa E240. Esta señal binauralizada se puede mezclar con otras señales estereofónicas procedentes de la etapa E220 descrita a continuación.

En el caso de que el flujo de datos que representa la señal monofónica incluya tanto una indicación de no binauralización (Di.) como información de posición espacial de reproducción (Pos.) (O en la etapa E200), la señal monofónica descodificada se envía a un motor de renderizado estereofónico para ser procesada por una etapa E220.

Esta indicación de no binauralización puede ser, por ejemplo, como en el estado de la técnica, una identificación "Dicótico" dada a la señal monofónica u otra identificación entendida como una instrucción de no procesar la señal mediante un procesamiento de binauralización. En formas de realización útiles para la comprensión de la invención, la información sobre la posición espacial de reproducción puede ser, por ejemplo, un ángulo acimutal que indique la posición de reproducción del sonido con respecto a un oído, derecho o izquierdo, o incluso una indicación de la diferencia de nivel entre los canales izquierdo y derecho, como una información ILD que permita distribuir la energía de la señal monofónica entre los canales izquierdo y derecho. De acuerdo con la invención, la información de posición espacial consiste en la indicación de un único canal de reproducción, correspondiente al oído derecho o izquierdo. En este último caso, esta información es binaria y requiere muy poca velocidad de transmisión de datos (sólo 1 bit de información).

En la etapa E220, la información de posición se tiene en cuenta para construir dos canales de reproducción para los dos auriculares de los cascos. Estos dos canales de reproducción contruidos de este modo se procesan directamente mediante una etapa de mezcla directa E230 que suma estos dos canales estereofónicos con los dos canales de la señal binauralizada resultante del procesamiento de binauralización E210.

A continuación, cada uno de los canales de reproducción estereofónico se suma con el canal correspondiente de la señal binauralizada.

Tras esta etapa de mezcla directa, los dos canales de reproducción procedentes de la etapa de mezcla E230 se reproducen en E240 en los cascos CA.

En una forma de realización de acuerdo con la invención, donde la información de posición espacial de reproducción es un dato binario que indica un único canal de los cascos de reproducción, esto significa que la señal monofónica debe reproducirse sólo en un auricular de estos auriculares. Los dos canales de reproducción contruidos en la etapa E220 por el motor de renderizado estereofónico consisten en un canal que contiene la señal monofónica, siendo el otro canal cero, y por lo tanto posiblemente ausente.

En la etapa de mezcla directa E230, un solo canal se suma con el canal correspondiente de la señal binauralizada, siendo el otro canal cero. Por lo tanto, esta etapa de mezcla se simplifica.

De este modo, el oyente que lleva los cascos escucha por una parte una escena de sonido espacializada procedente de la señal binauralizada en la misma ubicación física, aunque mueva la cabeza en el caso del renderizado dinámico, y por otra parte un sonido situado en el interior de la cabeza, entre un oído y el centro de la cabeza, que se superpone a la escena de sonido de forma independiente, es decir, si el oyente mueve la cabeza, este sonido se escuchará en la misma posición en relación con un oído.

Por lo tanto, este sonido se percibe como una superposición de los demás sonidos binaurales de la escena de sonido y actuará, por ejemplo, como una voz "OFF" en esta escena de sonido.

Así se consigue el efecto "auricular".

La **figura 3** ilustra una primera forma de realización de un decodificador que comprende un dispositivo de procesamiento que utiliza el método de procesamiento descrito con referencia a la figura 2. En este ejemplo de forma de realización, la señal monofónica procesada por el método utilizado es una señal de tipo canal (Ch.).

Las señales de tipo objeto (obj.) y tipo HOA (HOA) son procesadas de la misma forma por los respectivos bloques 303, 304 y 305 que los bloques 103, 104 y 105 descritos con referencia a la figura 1. De la misma forma, el bloque de mezcla 310 lleva a cabo una mezcla tal como se describe para el bloque 110 de la figura 1.

El bloque 330 que recibe las señales de tipo canal procesa una señal monofónica que comprende una indicación de no binauralización (Di.) asociada a una información de posición espacial de reproducción (Pos.) de forma diferente a otra señal que no contiene esta información, en particular una señal multicanal. En el caso de estas señales que no comprenden esta información, son procesadas por el bloque 302 de la misma forma que el bloque 102 descrito con referencia a la figura 1.

Para una señal monofónica que comprende la indicación de no binauralización asociada a una información de posición espacial de reproducción, el bloque 330 actúa como un enrutador o conmutador y dirige la señal monofónica descodificada (Mo.) a un motor de renderizado estereofónico 331. Este motor de renderizado estereofónico también recibe información sobre la posición espacial de reproducción (Pos.) del módulo de decodificación. Utilizando esta información, construye dos canales de reproducción (2 Vo.), correspondientes a los canales izquierdo y derecho de los cascos de reproducción, para que estos canales se puedan reproducir en los cascos CA.

En una variante no contemplada en el texto de las reivindicaciones, la información sobre la posición espacial de reproducción es una información sobre la diferencia interaural de nivel de sonido entre los canales izquierdo y derecho. Esta información permite definir un factor que se debe aplicar a cada uno de los canales de reproducción para respetar esta posición espacial de reproducción.

La definición de estos factores se puede llevar a cabo como en el documento denominado MPEG-2 AAC: ISO/IEC 13818-4:2004/DCOR 2, AAC en la sección 7.2 que describe la intensidad estéreo.

Antes de ser reproducidos en los cascos, estos canales de reproducción se añaden a los canales de una señal binauralizada procedente del módulo de binauralización 320, que lleva a cabo el procesamiento de binauralización de la misma forma que el bloque 120 de la figura 1.

Esta etapa de suma de canales se lleva a cabo mediante el módulo de mezcla directa 340, que suma el canal izquierdo procedente del motor de renderizado estereofónico 331 con el canal izquierdo de la señal binauralizada procedente del módulo de procesamiento binauralizado 320 y el canal derecho procedente del motor de renderizado estereofónico 331 con el canal derecho de la señal binauralizada procedente del módulo de procesamiento de binauralización 320, antes de la reproducción en los auriculares CA.

De este modo, la señal monofónica no pasa por el módulo de procesamiento de binauralización 320, sino que se transmite directamente al motor de renderizado estereofónico 331 antes de mezclarse directamente con una señal binauralizada.

Por lo tanto, esta señal tampoco sufrirá ningún procesamiento de seguimiento de la cabeza. Por tanto, el sonido reproducido estará en una posición de reproducción con respecto a uno de los oídos del oyente y permanecerá en esa posición, aunque el oyente mueva la cabeza.

En esta forma de realización, el motor de renderizado estereofónico 331 se puede integrar con el motor de renderizado de canales 302. En este caso, este motor de renderizado de canales utiliza a la vez la adaptación de las señales convencionales de tipo canal, como se describe en la figura 1, y la construcción de dos canales de reproducción del motor de renderizado 331 según se ha explicado anteriormente recibiendo la información de posición espacial de

reproducción (Pos.). A continuación, sólo los dos canales de reproducción se redirigen hacia el módulo de mezcla directa 340 antes de la reproducción en los cascos CA.

De acuerdo con la invención, el motor de renderizado estereofónico 331 se integra en el módulo de mezcla directa 340. En este caso, el módulo de enrutamiento 330 dirige la señal monofónica descodificada (para la que se han detectado la indicación de no binauralización y la información de posición espacial de reproducción) al módulo de mezcla directa 340. Por otra parte, la información de posición espacial de reproducción descodificada (Pos.) también se transmite al módulo de mezcla directa 340. Este módulo de mezcla directa, que comprende entonces el motor de renderizado estereofónico, utiliza la construcción de los dos canales de reproducción teniendo en cuenta la información de posición espacial de reproducción, así como la mezcla de estos dos canales de reproducción con los canales de reproducción de una señal binauralizada procedente del módulo de procesamiento de binauralización 320.

La **figura 4** ilustra una segunda forma de realización de un descodificador que comprende un dispositivo de procesamiento que utiliza el método de procesamiento descrito con referencia a la figura 2. En este ejemplo de forma de realización, la señal monofónica procesada por el método utilizado es una señal de tipo objeto (Obj.).

Las señales de tipo canal (Ch.) y de tipo HOA (HOA) son procesadas de la misma forma por los respectivos bloques 402 y 405 que los bloques 102 y 105 descritos con referencia a la figura 1. De la misma forma, el bloque de mezcla 410 lleva a cabo una mezcla tal como se describe para el bloque 110 de la figura 1.

El bloque 430, que recibe señales de tipo objeto (Obj.), trata una señal monofónica para la que se ha detectado una indicación de no binauralización (Di.) asociada a información de posición espacial de reproducción (Pos.) de forma diferente a otra señal monofónica para la que no se ha detectado esta información.

Para aquellas señales monofónicas para las que no se ha detectado esta información, son procesadas por el bloque 403 de la misma forma que el bloque 103 descrito con referencia a la figura 1, utilizando los parámetros descodificados del bloque 404 descodificando los Metadatos de la misma forma que el bloque 104 de la figura 1.

Para una señal monofónica de tipo objeto para la que se ha detectado la indicación de no binauralización asociada a la información de posición espacial de reproducción, el bloque 430 actúa como un enrutador o conmutador y dirige la señal monofónica descodificada (Mo.) a un motor de renderizado estereofónico 431.

La indicación de no binauralización (Di.) y la información de posición espacial de reproducción (Pos.) se descodifican mediante el bloque de descodificación 404 a partir de los metadatos o parámetros asociados con las señales de tipo objeto. La indicación de no binauralización (Di.) se transmite al bloque de enrutamiento 430 y la información de posición espacial de reproducción se transmite al motor de renderizado estereofónico 431.

Este motor de renderizado estereofónico recibe de este modo información sobre la posición espacial de la reproducción (Pos.) y construye dos canales de reproducción, correspondientes a los canales izquierdo y derecho de los cascos de reproducción, para que estos canales puedan reproducirse en los cascos CA.

En una variante no contemplada en el texto de las reivindicaciones, la información de posición espacial de reproducción es una información de ángulo acimutal que define un ángulo entre la posición de reproducción deseada y el centro de la cabeza del oyente.

Esta información permite definir un factor que se debe aplicar a cada uno de los canales de reproducción para respetar esta posición espacial de reproducción.

Los factores de ganancia para los canales izquierdo y derecho se pueden calcular como se presenta en el artículo titulado "Virtual Sound Source Positioning Using Vector Base Amplitude Panning" de Ville Pulkki en J. Audio Eng. Soc, Vol.45, No.6, Junio 1997.

Por ejemplo, los factores de ganancia para el motor de renderizado estereofónico pueden venir dados por:  

$$g1 = (\cos O \cdot \sin H + \sin O \cdot \cos H) / (2 \cdot \cos H \cdot \sin H) \quad g2 = (\cos O \cdot \sin H - \sin O \cdot \cos H) / (2 \cdot \cos H \cdot \sin H)$$

Donde g1 y g2 corresponden a los factores para las señales de los canales izquierdo y derecho, O es el ángulo entre la dirección frontal y el objeto (llamado acimut), y H es el ángulo entre la dirección frontal y la posición del altavoz virtual (correspondiente al semiángulo entre los altavoces), fijado por ejemplo en 45°.

Antes de ser reproducidos en los cascos, estos canales de reproducción se añaden a los canales de una señal binauralizada procedente del módulo de binauralización 420, que lleva a cabo el procesamiento de binauralización de la misma forma que el bloque 120 de la figura 1.

Esta etapa de suma de canales se lleva a cabo mediante el módulo de mezcla directa 440, que suma el canal izquierdo procedente del motor de renderizado estereofónico 431 con el canal izquierdo de la señal binauralizada procedente



del módulo de procesamiento binauralizado 420 y el canal derecho procedente del motor de renderizado estereofónico 431 con el canal derecho de la señal binauralizada procedente del módulo de procesamiento de binauralización 420, antes de la reproducción en los auriculares CA.

5 De este modo, la señal monofónica no pasa por el módulo de procesamiento de binauralización 420, sino que se transmite directamente al motor de renderizado estereofónico 431 antes de mezclarse directamente con una señal binauralizada.

10 Por lo tanto, esta señal tampoco sufrirá ningún procesamiento de seguimiento de la cabeza. Por tanto, el sonido reproducido estará en una posición de reproducción con respecto a uno de los oídos del oyente y permanecerá en esa posición, aunque el oyente mueva la cabeza.

15 En esta forma de realización no cubierta por el texto de las reivindicaciones, el motor de renderizado estereofónico 431 se puede integrar con el motor de renderizado de objetos 403. En este caso, este motor de renderizado de objetos utiliza tanto la adaptación de las señales convencionales de tipo objeto, como se describe en la figura 1, como la construcción de los dos canales de reproducción del motor de renderizado 431, como se ha explicado anteriormente, recibiendo la información de posición espacial de reproducción (Pos.) del módulo de descodificación de parámetros 404. Sólo los dos canales de reproducción (2Vo.) son entonces redirigidos al módulo de mezcla directa 440 antes de la reproducción en los cascos CA.

20 De acuerdo con la invención, el motor de renderizado estereofónico 431 se integra en el módulo de mezcla directa 440. En este caso, el módulo de enrutamiento 430 dirige la señal monofónica descodificada (Mo.) (para la que se han detectado la indicación de no binauralización y la información de posición espacial de reproducción) hacia el módulo de mezcla directa 440. Por otra parte, la información de posición espacial de reproducción descodificada (Pos.) también se transmite al módulo de mezcla directa 440 mediante el módulo de descodificación de parámetros 404. Este módulo de mezcla directa, que comprende entonces el motor de renderizado estereofónico, utiliza la construcción de los dos canales de reproducción teniendo en cuenta la información de posición espacial de reproducción, así como la mezcla de estos dos canales de reproducción con los canales de reproducción de una señal binauralizada procedente del módulo de procesamiento de binauralización 420.

30 La **figura 5** ilustra ahora un ejemplo de forma de realización del hardware de un dispositivo de procesamiento capaz de utilizar el método de procesamiento de acuerdo con la invención.

35 El dispositivo DIS comprende un espacio de almacenamiento 530, por ejemplo una memoria MEM, una unidad de procesamiento 520 que comprende un procesador PROC, controlado por un programa informático Pg, almacenado en la memoria 530 y que utiliza el método de procesamiento de acuerdo con la invención.

40 El programa de ordenador Pg incluye instrucciones de código para la implementación de las etapas del método de procesamiento en el sentido de la invención, cuando estas instrucciones son ejecutadas por el procesador PROC, y en particular, en la detección, en un flujo de datos representativo de la señal monofónica, de una indicación de no procesamiento de binauralización asociada a una información de posición espacial de reproducción, una etapa consistente en dirigir la señal monofónica descodificada hacia un motor de renderizado estereofónico que tenga en cuenta la información de posición, para construir dos canales de reproducción procesados directamente por una etapa de mezcla directa que sume estos dos canales con una señal binauralizada resultante del procesamiento de binauralización, con el fin de ser reproducida en los cascos.

Normalmente, la descripción de la figura 2 retoma las etapas de un algoritmo para un programa informático de este tipo.

50 En la inicialización, las instrucciones de código de programa Pg se cargan, por ejemplo, en una memoria RAM (no mostrada) antes de ser ejecutadas por el procesador PROC de la unidad de procesamiento 520. Las instrucciones del programa se pueden memorizar en un medio de almacenamiento como por ejemplo una memoria flash, un disco duro o cualquier otro medio de almacenamiento no transitorio.

55 El dispositivo DIS incluye un módulo de recepción 510 capaz de recibir un flujo de datos SMO representativo en particular de una señal monofónica. Comprende un módulo de detección 540 capaz de detectar, en este flujo de datos, una indicación de que no se está procesando la binauralización, asociada a una información sobre la posición espacial de la reproducción. Comprende un módulo de dirección 550, en caso de detección positiva por el módulo de detección 540, de la señal monofónica descodificada hacia un motor de renderizado estereofónico 560, siendo el motor de renderizado estereofónico 560 capaz de tener en cuenta la información de posición para construir dos canales de reproducción.

60 El dispositivo DIS también incluye un módulo de mezcla directa 570 capaz de procesar directamente los dos canales de reproducción sumándolos con los dos canales de una señal binauralizada procedente de un módulo de procesamiento de binauralización. Los canales de reproducción obtenidos de este modo se transmiten a unos cascos CA por medio de un módulo de salida 560, para su reproducción.

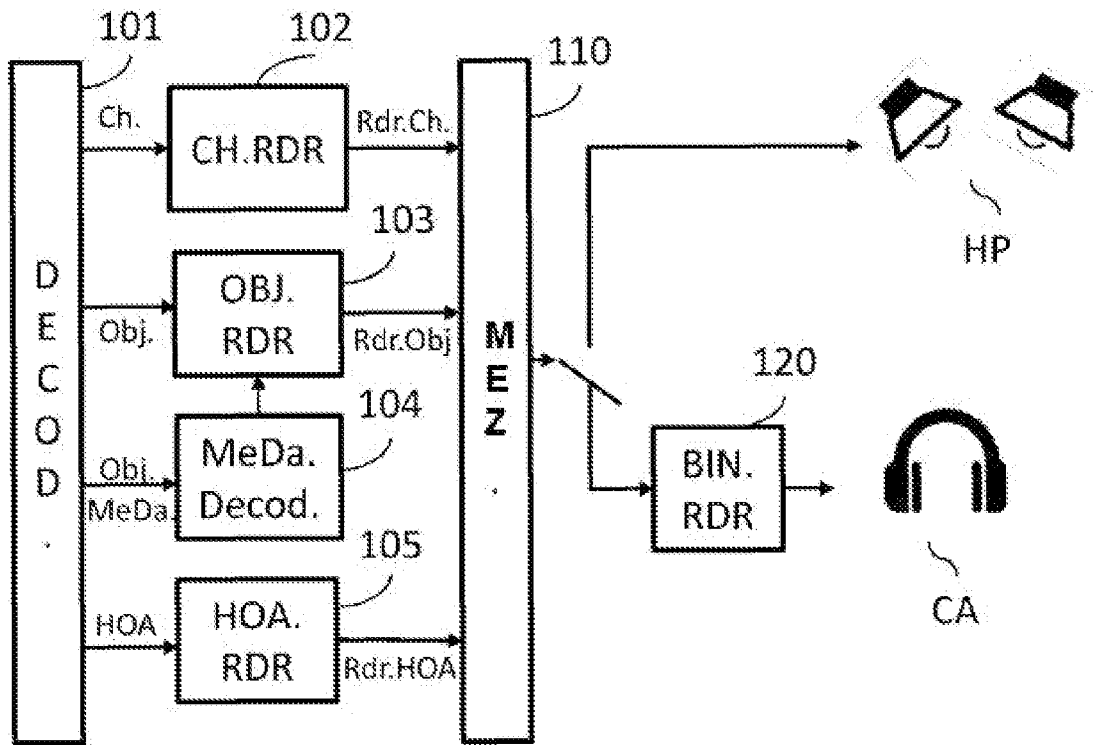
Estos diferentes módulos son tal como se describen con referencia a las figuras 3 y 4 de acuerdo con las formas de realización.

- 5 El término módulo puede corresponder tanto a un componente de software como a un componente de hardware o a un conjunto de componentes de hardware y software, correspondiendo un componente de software propiamente dicho a uno o más programas o subprogramas informáticos o, de forma más general, a cualquier elemento de un programa capaz de llevar a cabo una función o un conjunto de funciones tales como las descritas para los módulos en cuestión. Del mismo modo, un componente hardware corresponde a cualquier elemento de un conjunto hardware capaz de
- 10 llevar a cabo una función o un conjunto de funciones para el módulo en cuestión (circuito integrado, tarjeta inteligente, tarjeta de memoria, etc.).

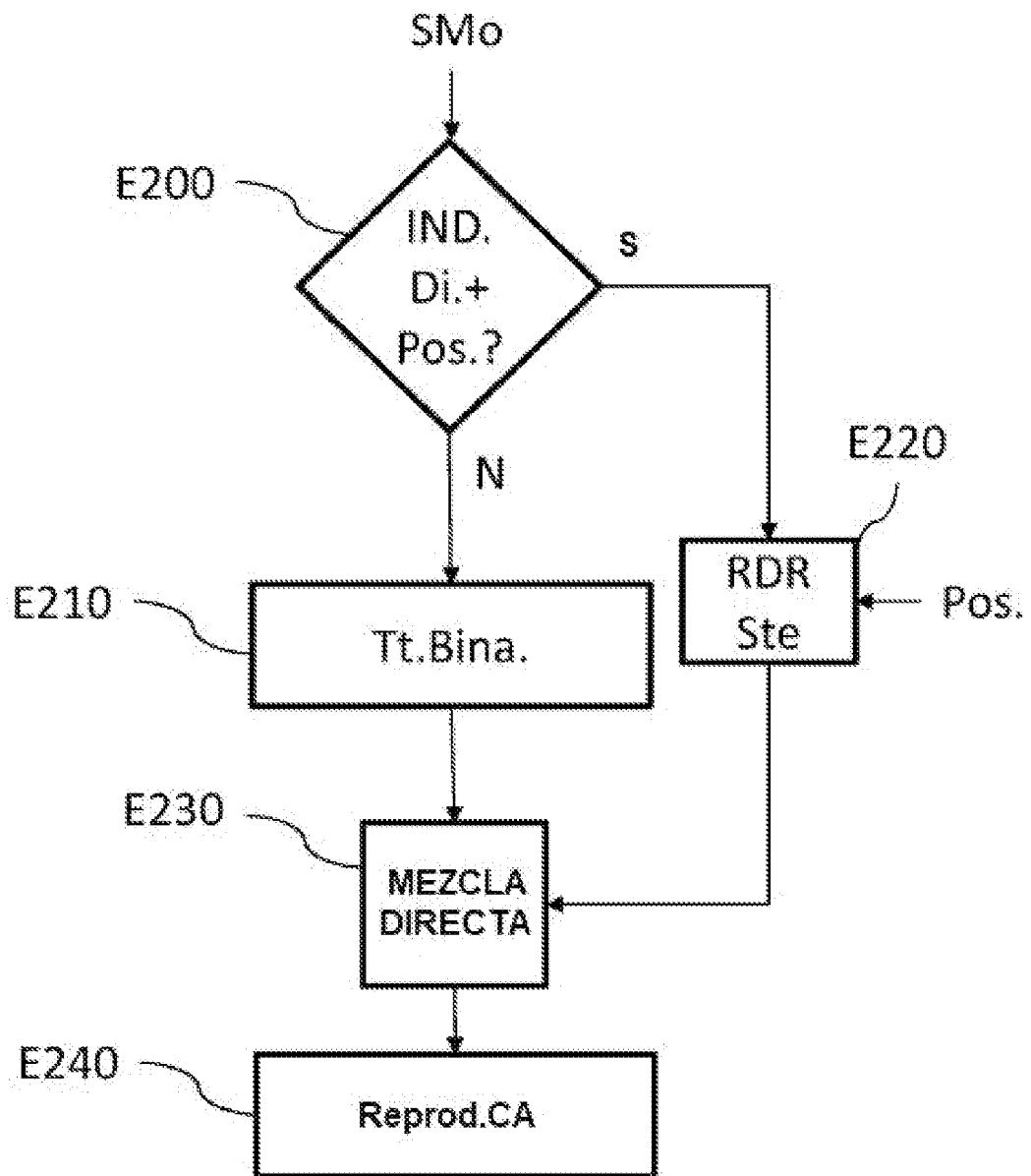
- El dispositivo se puede integrar en un decodificador de audio tal como el descrito en las figuras 3 o 4 y se puede integrar, por ejemplo, en equipos multimedia de tipo decodificadores de salón, decodificadores o reproductores de
- 15 contenidos de audio o vídeo. También se puede integrar en equipos de comunicación de tipo teléfonos móviles o pasarelas de comunicación.

## REIVINDICACIONES

1. Método de procesamiento de una señal de audio monofónica en un decodificador de audio 3D que comprende una etapa de procesamiento de binauralización de las señales descodificadas destinadas a ser reproducidas espacialmente por unos cascos, en la que, al detectar (E200), en un flujo de datos representativo de la señal monofónica, una indicación de no procesamiento de binauralización asociada a una información de posición espacial de reproducción, la señal monofónica descodificada se envía (O-E200) hacia un módulo de mezcla que comprende un motor de renderizado estereofónico que tiene en cuenta la información de posición para construir dos canales de reproducción (E220), procesados directamente por una etapa de mezcla directa (E230) que suma estos dos canales con una señal binauralizada procedente del procesamiento de binauralización, para ser reproducida (E240) en los cascos, **caracterizado por que** la información de posición espacial de reproducción es un dato binario que indica un único canal de reproducción de los cascos de reproducción.
2. Método de acuerdo con la reivindicación 1, en el que sólo el canal de reproducción correspondiente al canal indicado por los datos binarios se suma al canal correspondiente de la señal binauralizada en la etapa de mezcla directa, siendo el otro canal de reproducción de valor cero.
3. Método de acuerdo con la reivindicación 1, en el que la señal monofónica es una señal de tipo canal dirigida hacia el módulo de mezcla, con la información de posición espacial de reproducción.
4. Método de acuerdo con la reivindicación 1, en el que la señal monofónica es una señal de tipo objeto asociada a un conjunto de parámetros de reproducción que comprenden la indicación de no binauralización y la información de posición de reproducción, dirigiéndose la señal al módulo de mezclado con la información de posición de reproducción.
5. Dispositivo para procesar una señal de audio monofónica que comprende un módulo de procesamiento de la binauralización de señales descodificadas destinadas a ser reproducidas espacialmente por unos cascos, que comprende:
  - un módulo de detección (330; 430) capaz de detectar, en un flujo de datos representativo de la señal monofónica, una indicación de no procesamiento de la binauralización asociada a la información sobre la posición espacial de la reproducción;
  - un módulo de redirección (330, 430), en caso de detección positiva por el módulo de detección, capaz de dirigir la señal monofónica descodificada hacia un módulo de mezcla;
  - un módulo de mezcla (340; 440) que comprende un motor de renderizado estereofónico (331; 431) capaz de tener en cuenta la información de posición para construir dos canales de reproducción, pudiendo el módulo de mezcla procesar directamente los dos canales de reproducción sumándolos con una señal binauralizada procedente del módulo de procesamiento de binauralización (320; 420), para ser reproducida en los cascos;**caracterizado por que** la información de posición espacial de reproducción son datos binarios que indican un único canal de los cascos de reproducción.
6. Dispositivo de acuerdo con la reivindicación 5, en el que la señal monofónica es una señal de tipo canal y en el que el motor de renderizado estereofónico se integra con un motor de renderizado de canales que construye de otro modo canales de reproducción para señales multicanal.
7. Dispositivo de acuerdo con la reivindicación 5, en el que la señal monofónica es una señal de tipo objeto y en el que el motor de renderizado estereofónico se integra con un motor de renderizado de objetos construyendo de otro modo canales de reproducción para señales monofónicas asociadas con conjuntos de parámetros de reproducción.
8. Decodificador de audio que comprende un dispositivo de procesamiento de acuerdo con una de las reivindicaciones 5 a 7.
9. Programa de ordenador que comprende instrucciones de código para la implementación de las etapas del método de procesamiento de acuerdo con una de las reivindicaciones 1 a 4, cuando estas instrucciones son ejecutadas por un procesador.
10. Medio de almacenamiento, legible por un procesador, que almacena un programa informático que comprende instrucciones para la ejecución del método de procesamiento de acuerdo con una de las reivindicaciones 1 a 4.



**FIG. 1** (Estado de la técnica)



**FIG. 2**

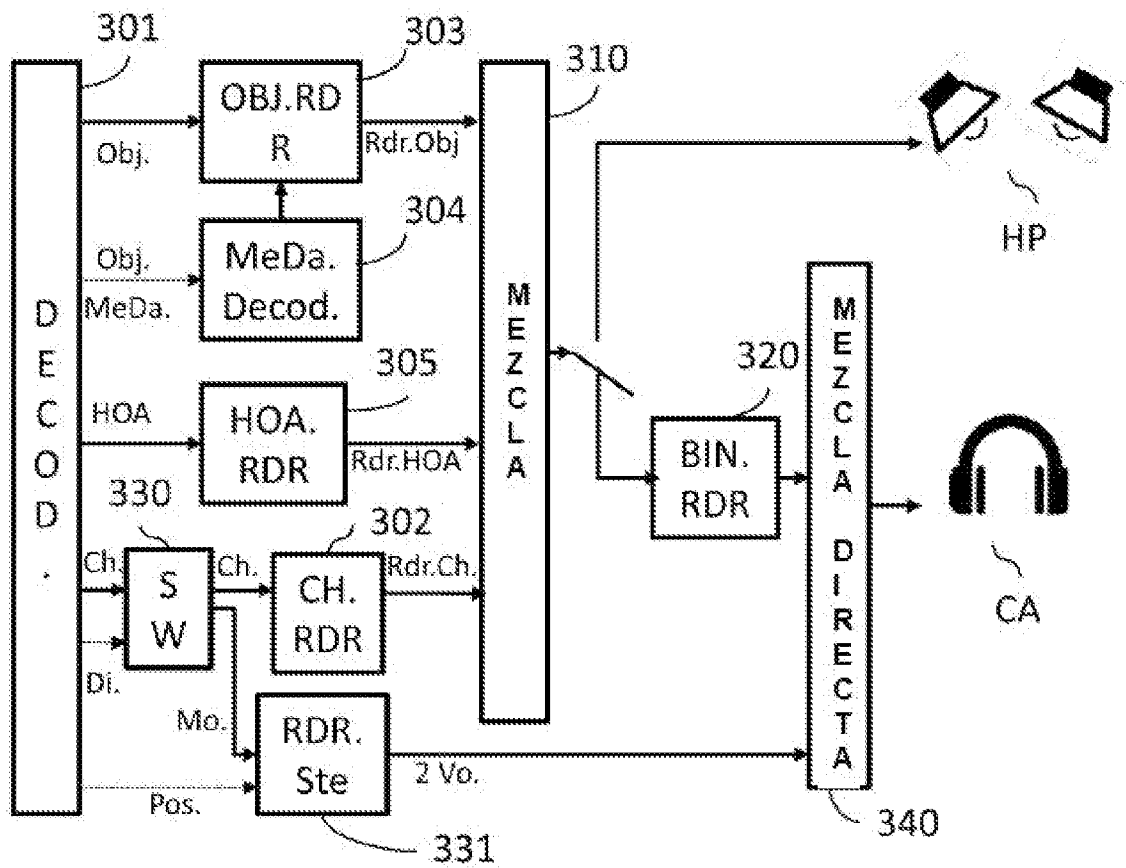


FIG. 3

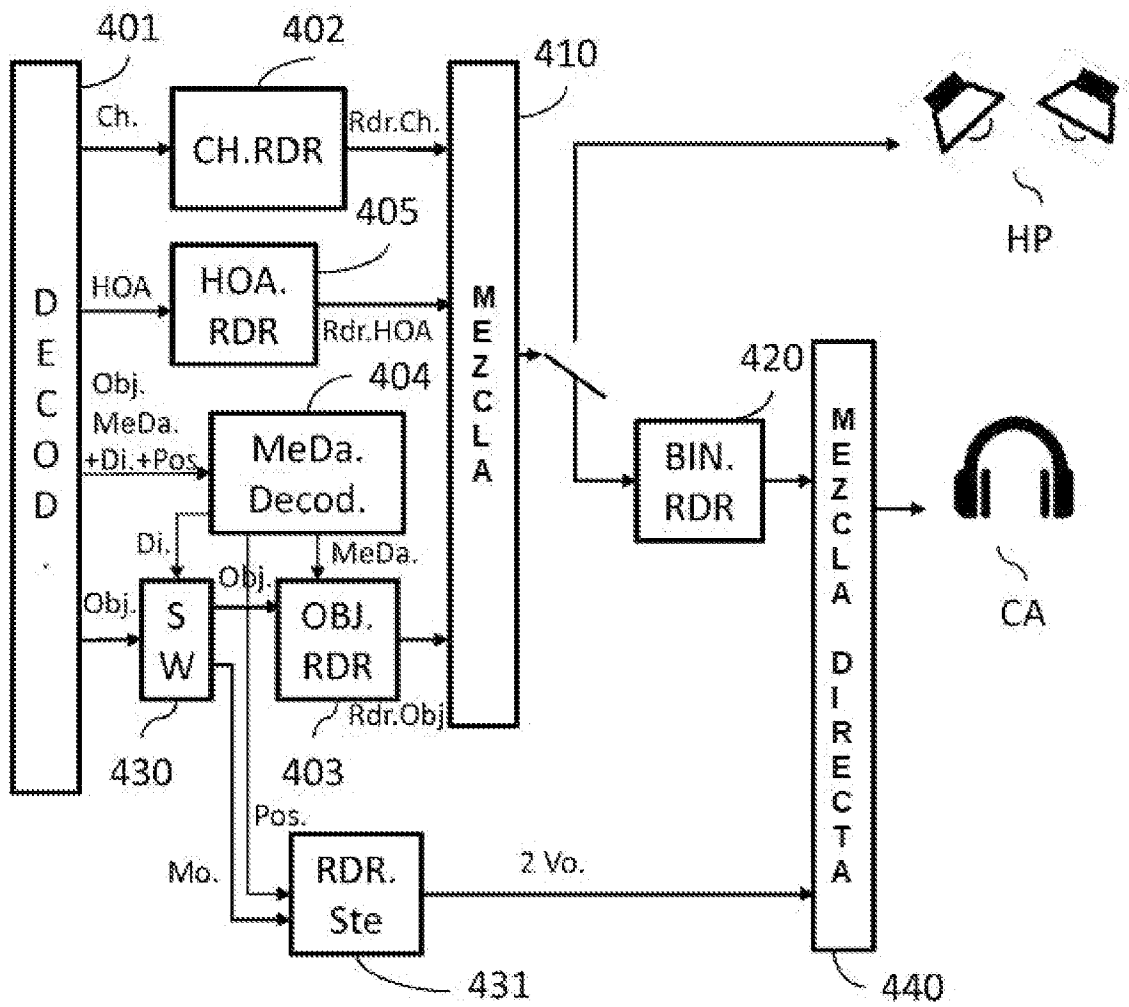


FIG. 4

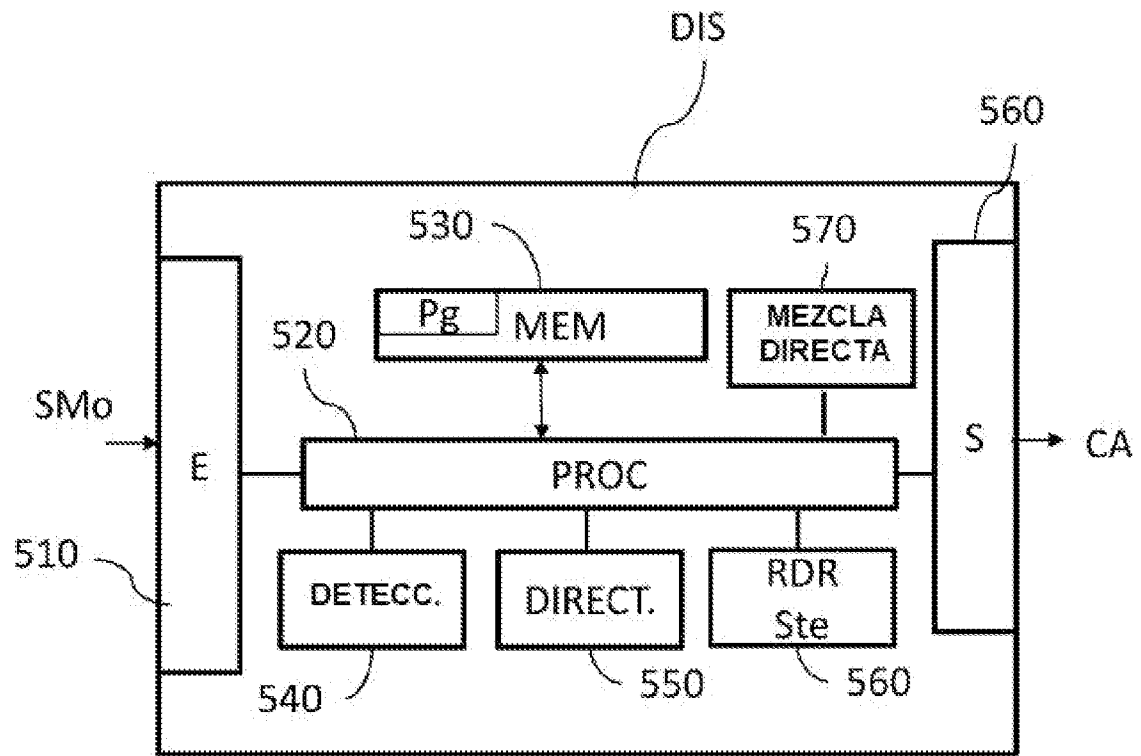


FIG. 5