

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 755 349**

51 Int. Cl.:

H04S 7/00 (2006.01)

G10L 19/20 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.10.2014 PCT/US2014/062705**

87 Fecha y número de publicación internacional: **07.05.2015 WO15066062**

96 Fecha de presentación y número de la solicitud europea: **28.10.2014 E 14795767 (4)**

97 Fecha y número de publicación de la concesión europea: **16.10.2019 EP 3063955**

54 Título: **Renderización binaural para auriculares utilizando procesamiento de metadatos**

30 Prioridad:

31.10.2013 US 201361898365 P

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

22.04.2020

73 Titular/es:

**DOLBY LABORATORIES LICENSING CORPORATION (100.0%)
1275 Market Street
San Francisco, CA 94103, US**

72 Inventor/es:

**TSINGOS, NICOLAS R.;
WILSON, RHONDA;
BHARITKAR, SUNIL;
BROWN, C. PHILLIP;
SEEFELDT, ALAN J. y
AUDFRAY, REMI**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 755 349 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 para auriculares utilizando procesamiento de metadatos

Campo de la invención

5 Una o más implementaciones se refieren generalmente al procesamiento de señales de audio y, más específicamente, a la renderización binaural de audio basado en canales y objetos para la reproducción en auriculares.

Antecedentes

10 La renderización virtual de audio espacial en un par de altavoces implica comúnmente la creación de una señal binaural estéreo que representa el sonido deseado que llega a los oídos izquierdo y derecho del oyente y se sintetiza para simular una escena de audio particular en un espacio tridimensional (3D), que contiene posiblemente una multitud de fuentes en diferentes ubicaciones. Para la reproducción a través de auriculares en lugar de altavoces, el procesamiento o renderización binaural se puede definir como un conjunto de operaciones de procesamiento de señales destinadas a reproducir la ubicación 3D deseada de una fuente de sonido a través de auriculares emulando las señas de escucha espaciales naturales de los seres humanos. Los componentes centrales típicos de un renderizador binaural son el filtrado relacionado con la cabeza para reproducir señas dependientes de la dirección, así como el procesamiento de señas de distancia, lo que puede implicar modelar la influencia de una sala o entorno de escucha real o virtual. Un ejemplo de una renderización binaural actual procesa cada uno de los 5 o 7 canales de un sonido envolvente 5.1 o 7.1 en una presentación de audio basada en canales a fuentes de sonido virtual 5/7 en un espacio 2D alrededor del oyente. La renderización binaural también se encuentra comúnmente en juegos o hardware de audio para juegos, en cuyo caso el procesamiento se puede aplicar a objetos de audio individuales en el juego basándose en su posición 3D individual.

25 Tradicionalmente, la renderización binaural es una forma de postprocesamiento ciego aplicado al contenido de audio basado en objetos multicanal. Parte del procesamiento involucrado en la renderización binaural puede tener efectos negativos y no deseados en el timbre del contenido, como el suavizado de transitorios o la reverberación excesiva agregada al diálogo o algunos efectos y elementos musicales. Con la creciente importancia de la audición en auriculares y la flexibilidad adicional que brinda el contenido basado en objetos (como el sistema Dolby® Atmos™), existe una mayor oportunidad y necesidad de que los mezcladores creen y codifiquen metadatos específicos de renderización binaural en el momento de la creación de contenido, por ejemplo, dando instrucciones al renderizador para que procese partes del contenido con diferentes algoritmos o con diferentes configuraciones. Los sistemas actuales no cuentan con esta capacidad, ni permiten que tales metadatos se transporten como parte de una carga útil de auriculares adicional específica en los códecs.

35 Los sistemas actuales tampoco están optimizados en el extremo de reproducción de la cadena de transmisión, en la medida en que el contenido no está configurado para ser recibido en un dispositivo con metadatos adicionales que pueden proporcionarse en vivo al renderizador binaural. Si bien el seguimiento de la cabeza en tiempo real se ha implementado previamente y se ha demostrado que mejora la renderización binaural, esto generalmente bloquea otras características como la detección continua y automática del tamaño de la cabeza y la detección de la sala, y otras características de personalización que mejoran la calidad de la renderización binaural para que sea efectiva y se implemente eficientemente en sistemas de reproducción basados en auriculares.

40 Aki Härmä y otros: "Techniques and applications of wereable augmented reality audio", documento de la Engineering Society Convention 5768, de 22 de marzo de 2003, describe un método para renderizar audio para su reproducción a través de auriculares que ajusta la renderización de la señal de entrada recibida en función de datos de ubicación y orientación obtenidos de los seguidores de cabeza. Este documento también sugiere recibir contenido dependiente de metadatos del contenido generados por una herramienta de autoría que procesa el contenido de audio digital recibido. El documento WO 2013/006338 A2 revela un método para renderizar audio para reproducción a través de auriculares que comprende: recibir contenido de audio digital; recibir metadatos dependientes de contenido generados por una herramienta de autoría que procesa el contenido de audio digital recibido, en el que los metadatos dependientes de contenido controlan una pluralidad de características de canal y objeto e incluyen posiciones de objetos de audio en el audio basado en objetos.

50 MV LAITINEN ET AL: "Reproduction binaural for directional audio coding", WASPAA '09, APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS", 2009, IEEE, PISCATAWAY, NJ, EE.UU., 18 de octubre de 2009, páginas 337-340, describe un sistema de renderización binaural que incluye el seguimiento de la cabeza para renderizar contenido de audio difuso y no difuso.

55 Lo que se necesita, por lo tanto, es una renderización binaural que se ejecute en el dispositivo de reproducción que combine metadatos de autoría con metadatos generados localmente en tiempo real para proporcionar la mejor experiencia posible al usuario final cuando escuche audio basado en canales y objetos a través de auriculares. Además, para el contenido basado en canales, generalmente se requiere que se conserve la intención artística incorporando análisis de segmentación de audio.

No debe suponerse que la materia discutida en la sección de antecedentes es técnica anterior simplemente como resultado de su mención en la sección de antecedentes. Del mismo modo, no debe suponerse que un problema mencionado en la sección de antecedentes o asociado con la materia de la sección de antecedentes haya sido previamente reconocido en la técnica anterior. La materia de la sección de antecedentes representa simplemente diferentes enfoques, que en sí mismos también pueden ser invenciones.

Breve compendio de las realizaciones

Se describen realizaciones para sistemas y métodos de renderización virtual de contenido de audio basado en objetos y ecualización mejorada en sistemas de reproducción basados en auriculares. Las realizaciones incluyen un método según la reivindicación 1 adjunta.

El método comprende además separar el contenido de audio digital en uno o más componentes en función del tipo de contenido, y en el que el tipo de contenido se selecciona del grupo que consiste en: diálogo, música, efectos de audio, señales transitorias y señales ambientales. Los metadatos de renderización binaural controlan una pluralidad de características de canal y objeto que incluyen: posición, tamaño, ajuste de ganancia y configuraciones dependientes del contenido o preconfiguraciones de procesamiento; y los metadatos de reproducción controlan una pluralidad de características específicas del oyente, que incluyen la posición de la cabeza, la orientación de la cabeza, el tamaño de la cabeza, los niveles de ruido de la sala de escucha, las propiedades de la sala de escucha y la posición del dispositivo o la pantalla de reproducción con relación al oyente. El método puede incluir además recibir uno o más comandos de entrada del usuario que modifiquen los metadatos de renderización binaural, los comandos de entrada del usuario controlan una o más características que incluyen: énfasis de elevación donde los objetos y canales elevados podrían recibir un aumento de ganancia, radio de sonido 1D (unidimensional) preferido factores de escala 3D para el posicionamiento de objetos o canales, y la habilitación del modo de procesamiento (por ejemplo, para alternar entre estéreo tradicional o procesamiento completo de contenido). Los metadatos de reproducción pueden generarse en respuesta a datos de sensor proporcionados por un auricular habilitado que alberga una pluralidad de sensores, comprendiendo el auricular habilitado parte del dispositivo de reproducción. El método puede comprender además separar el audio de entrada en subseñales separadas, por ejemplo, por tipo de contenido o sin mezclar el audio de entrada (basado en canales y en objetos), en contenido directo constituyente y contenido difuso, en donde el contenido difuso comprende elementos de sonido reverberados o reflejados, y realizando una renderización binaural en las subseñales separadas de forma independiente.

El método puede comprender además realizar una o más funciones de postprocesamiento en el contenido de audio fuente antes de la reproducción a través de auriculares; en donde las funciones de postprocesamiento comprenden al menos uno de: mezcla descendente de una pluralidad de canales de sonido envolvente a una mezcla binaural o una mezcla estéreo, gestión de nivel, ecualización, corrección de timbre y cancelación de ruido.

Las realizaciones se dirigen además a sistemas y artículos de fabricación que realizan o incorporan comandos de procesamiento que realizan o implementan los actos del método descritos anteriormente.

Breve descripción de los dibujos

En los siguientes dibujos, se usan números de referencia similares para referirse a elementos similares. Aunque las siguientes figuras representan diversos ejemplos, la una o más implementaciones no se limitan a los ejemplos renderizados en las figuras.

La figura 1 ilustra un sistema general que incorpora realizaciones de un sistema de creación, renderización y reproducción de contenido, en algunas realizaciones.

La figura 2A es un diagrama de bloques de una herramienta de autoría utilizada en un sistema de renderización en auriculares basado en objetos, en una realización.

La figura 2B es un diagrama de bloques de una herramienta de autoría utilizada en un sistema de renderización en auriculares basado en objetos, en una realización alternativa

La figura 3A es un diagrama de bloques de un componente de renderización utilizado en un sistema de renderización en auriculares basado en objetos, en una realización.

La figura 3B es un diagrama de bloques de un componente de renderización utilizado en un sistema de renderización en auriculares basado en objetos, en una realización alternativa.

La figura 4 es un diagrama de bloques que proporciona una visión general del sistema de renderización binaural de doble extremo, en una realización.

La figura 5 ilustra una GUI de herramienta de autoría que puede usarse con realizaciones de un sistema de renderización en auriculares, en una realización.

La figura 6 ilustra un auricular habilitado que comprende uno o más sensores que detectan las condiciones de reproducción para codificarlos como metadatos utilizados en un sistema de renderización en auriculares, en una realización.

5 La figura 7 ilustra la conexión entre un auricular y un dispositivo que incluye un procesador de sensor de auriculares, en una realización.

La figura 8 es un diagrama de bloques que ilustra los diferentes componentes de metadatos que pueden usarse en un sistema de renderización en auriculares, en una realización.

La figura 9 ilustra componentes funcionales de un componente de renderización binaural para el procesamiento de auriculares, en una realización.

10 La figura 10 ilustra un sistema de renderización binaural para renderizar objetos de audio en un sistema de renderización en auriculares, en una realización.

La figura 11 ilustra una representación más detallada del sistema de renderización binaural de la figura 10, en una realización.

15 La figura 12 es un diagrama de sistema que muestra las diferentes herramientas utilizadas en un sistema de modelado HRTF utilizado en un sistema de renderización en auriculares, en una realización.

La figura 13 ilustra una estructura de datos que permite la entrega de metadatos para un sistema de renderización en auriculares, en una realización.

La figura 14 ilustra un caso de ejemplo de tres mediciones de respuesta de impulso para cada oído, en una realización de un proceso de ecualización de auriculares.

20 La figura 15A ilustra un circuito para calcular la transmisión de sonido de campo libre, en una realización.

La figura 15B ilustra un circuito para calcular la transmisión de sonido en auriculares, en una realización.

Descripción detallada

25 Se describen sistemas y métodos para la renderización virtual de contenido basado en objetos a través de auriculares, y un sistema de entrega y procesamiento de metadatos para la renderización virtual, aunque las aplicaciones no son tan limitadas. Los aspectos de una o más realizaciones descritas en la presente memoria pueden implementarse en un sistema de audio o audiovisual que procesa información de audio fuente en un sistema de mezcla, renderización y reproducción que incluye uno o más ordenadores o dispositivos de procesamiento que ejecutan instrucciones de software. Cualquiera de las realizaciones descritas puede usarse sola o junto con otras en cualquier combinación. Aunque diversas realizaciones pueden haber sido motivadas por diversas deficiencias de la técnica anterior, que pueden discutirse o aludirse en uno o más lugares de la memoria, las realizaciones no abordan necesariamente ninguna de estas deficiencias. En otras palabras, diferentes realizaciones pueden abordar diferentes deficiencias que pueden discutirse en la memoria. Algunas realizaciones sólo pueden abordar parcialmente algunas deficiencias o solo una deficiencia que puede discutirse en la memoria, y algunas realizaciones pueden no abordar ninguna de estas deficiencias.

35 Las realizaciones están dirigidas a un sistema de producción y reproducción de contenido de audio que optimiza la renderización y reproducción de audio basado en objetos y/o canales a través de auriculares. La figura 1 ilustra un sistema general que incorpora realizaciones de un sistema de creación, renderización y reproducción de contenido, en algunas realizaciones. Como se muestra en el sistema 100, un creador utiliza una herramienta de autoría 102 para generar contenido de audio para su reproducción a través de uno o más dispositivos 104 para que un usuario lo escuche a través de unos auriculares 116 o 118. El dispositivo 104 es generalmente un reproductor de música o audio portátil o un ordenador pequeño o un dispositivo de telecomunicaciones móvil que ejecuta aplicaciones que permiten la reproducción de contenido de audio. Tal dispositivo puede ser un teléfono móvil o reproductor de audio (por ejemplo, MP3) 106, una tableta (por ejemplo, iPad de Apple o dispositivo similar) 108, una consola de música 110, un ordenador portátil 111 o cualquier dispositivo de reproducción de audio similar. El audio puede comprender música, diálogo, efectos o cualquier audio digital que se desee escuchar a través de auriculares, y tal audio puede transmitirse de forma inalámbrica desde una fuente de contenido, reproducirse localmente desde un medio de almacenamiento (por ejemplo, disco, unidad flash, etc.), o generarse localmente. En la siguiente descripción, el término "auriculares" se refiere específicamente en general a un dispositivo de reproducción de acoplamiento cerrado que usa el usuario directamente sobre sus oídos o a dispositivos de escucha en el oído; también puede referirse en general a al menos parte del procesamiento realizado para renderizar señales destinadas a la reproducción en auriculares como una alternativa a los términos "procesamiento de auriculares" o "reproducción en auriculares".

El audio procesado por el sistema comprende audio basado en objetos y canales (por ejemplo, audio híbrido o adaptativo). El audio comprende o está asociado con metadatos que dictan cómo se renderiza el audio para su

reproducción en dispositivos de punto final y entornos de escucha específicos. El audio basado en canales generalmente se refiere a una señal de audio más metadatos en los que la posición se codifica como un identificador de canal, donde el audio está formateado para su reproducción a través de un conjunto predefinido de zonas de altavoces con ubicaciones de sonido envolvente nominales asociadas, por ejemplo, 5.1, 7.1, y así sucesivamente; y basado en objetos significa uno o más canales de audio con una descripción de fuente paramétrica, como la posición de fuente aparente (por ejemplo, coordenadas 3D), ancho de fuente aparente, etc. El término "audio adaptativo" puede usarse para referirse a señales de audio basadas en objetos y/o canales más metadatos que renderizan las señales de audio basadas en el entorno de reproducción utilizando una secuencia de audio más metadatos en los que la posición se codifica como una posición 3D en el espacio. En general, el entorno de escucha puede ser abierto, parcialmente cerrado, o un área completamente cerrada, tal como una sala, pero las realizaciones descritas en la presente memoria generalmente están dirigidas a la reproducción a través de auriculares u otros dispositivos de punto final de proximidad cercana. Los objetos de audio pueden considerarse como grupos de elementos de sonido que pueden percibirse que emanan de una ubicación física particular o ubicaciones en el entorno, y tales objetos pueden ser estáticos o dinámicos. Los objetos de audio están controlados por metadatos que, entre otras cosas, detallan la posición del sonido en un punto dado en el tiempo, y al reproducirlos se renderizan según los metadatos posicionales. En un sistema de audio híbrido, el contenido basado en canales (por ejemplo, "camas") puede procesarse además de los objetos de audio, donde las camas son submezclas o troncos basados en canales. Estos se pueden entregar para la reproducción final (renderización) y se pueden crear en diferentes configuraciones basadas en canales tales como 5.1, 7.1.

Los auriculares utilizados por el usuario son unos auriculares habilitados 116, como se muestra en la figura 1, que incluye sensores y otros componentes (con o sin alimentación) que proporcionan ciertos parámetros operativos al renderizador para un procesamiento y optimización adicionales del contenido de audio. La figura 1 también ilustra un auricular 118 pasivo o legado que solo incluye transductores sin alimentación que simplemente recrean la señal de audio. Sin embargo, en el resto de la descripción. Los ejemplos descritos en combinación con tales auriculares 118 no están comprendidos en el alcance de las reivindicaciones adjuntas. Los auriculares 116 o 118 pueden incorporarse en cualquier dispositivo de oído cerrado apropiado, tales como auriculares abiertos o cerrados, auriculares sobre el oído o dentro del oído, auriculares individuales, botones auditivos, almohadillas auditivas, de cancelación de ruido, de aislamiento u otro tipo de dispositivo de auriculares. Tales auriculares pueden ser cableados o inalámbricos con respecto a su conexión a la fuente de sonido o dispositivo 104.

En una realización, el contenido de audio de la herramienta de autoría 102 incluye audio estéreo o basado en canales (por ejemplo, sonido envolvente 5.1 o 7.1) además del audio basado en objetos. Para la realización de la figura 1, un renderizador 112 recibe el contenido de audio de la herramienta de autoría y proporciona ciertas funciones que optimizan el contenido de audio para la reproducción a través del dispositivo 104 y los auriculares 116 o 118. En una realización, el renderizador 112 incluye una etapa de preprocesamiento 113, una etapa de renderización binaural 114, y una etapa de postprocesamiento 115. La etapa de preprocesamiento 113 generalmente realiza ciertas operaciones de segmentación en el audio de entrada, tales como segmentar el audio en función de su tipo de contenido, entre otras funciones; la etapa de renderización binaural 114 generalmente combina y procesa los metadatos asociados con los componentes de canal y objeto del audio y genera una salida de audio binaural estéreo o multicanal con estéreo binaural y salidas de baja frecuencia adicionales; y el componente de postprocesamiento 115 generalmente realiza mezcla descendente, ecualización, control de intervalo de ganancia/volumen/ dinámico y otras funciones antes de la transmisión de la señal de audio al dispositivo 104. Cabe señalar que, si bien el renderizador probablemente generará señales de dos canales en la mayoría de los casos, podría configurarse para proporcionar más de dos canales de entrada a auriculares habilitados específicos, por ejemplo, para entregar canales de bajos separados (similar al canal LFE .1 en sonido envolvente tradicional). Los auriculares habilitados pueden tener conjuntos específicos de controladores para reproducir componentes de bajos por separado a partir del sonido de frecuencia media a más alta.

Debe observarse que los componentes de la figura 1 generalmente representan los principales bloques funcionales de los sistemas de generación, renderización y reproducción de audio, y que ciertas funciones pueden incorporarse como parte de uno o más componentes. Por ejemplo, una o más partes del renderizador 112 pueden incorporarse parcial o totalmente en el dispositivo 104. En este caso, el reproductor de audio o tableta (u otro dispositivo) puede incluir un componente de renderización integrado dentro del dispositivo. De manera similar, los auriculares habilitados 116 pueden incluir al menos algunas funciones asociadas con el dispositivo de reproducción y/o renderización. En tal caso, un auricular totalmente integrado puede incluir un dispositivo de reproducción integrado (por ejemplo, un decodificador de contenido incorporado, por ejemplo, un reproductor MP3), así como un componente de renderización integrado. Además, uno o más componentes del renderizador 112, tal como el componente de preprocesamiento 113, pueden implementarse al menos en parte en la herramienta de autoría, o como parte de un componente de preprocesamiento separado.

La figura 2A es un diagrama de bloques de una herramienta de autoría utilizada en un sistema de renderización en auriculares basado en objetos, en una realización. Como se muestra en la figura 2A, el audio de entrada 202 procedente de una fuente de audio (por ejemplo, fuente en vivo, grabación, etc.) se introduce en una estación de trabajo de audio digital (DAW) 204 para que la procese un ingeniero de sonido. El audio de entrada 201 está típicamente en forma digital, y si se usa audio analógico, se requiere un paso de conversión A/D (analógico a digital) (no mostrado). Este audio generalmente comprende contenido basado en objetos y canales, tal como puede usarse

en un sistema de audio adaptativo (por ejemplo, Dolby Atmos) y a menudo incluye varios tipos diferentes de contenido. El audio de entrada puede segmentarse mediante un preproceso de segmentación de audio (opcional) 204 que separa (o segmenta) el audio en función de su tipo de contenido para que los diferentes tipos de audio se puedan renderizar de manera diferente. Por ejemplo, el diálogo se puede renderizar de manera diferente que las señales transitorias o las señales ambientales. La DAW 204 puede implementarse como una estación de trabajo para editar y procesar el audio digital 202 segmentado o no segmentado, y puede incluir una consola de mezcla, superficie de control, conversor de audio, almacenamiento de datos y otros elementos apropiados. En una realización, la DAW es una plataforma de procesamiento que ejecuta software de audio digital que proporciona una funcionalidad de edición integral, así como una interfaz para uno o más programas de complemento, tal como un complemento de panoramizador, entre otras funciones, tales como ecualizadores, sintetizadores, efectos, etc. El complemento de panoramizador en la DAW 204 realiza una función de panoramizador configurada para distribuir cada señal de objeto a pares o ubicaciones de altavoces específicos en el espacio 2D/3D de manera que transmite al oyente la posición deseada de cada señal de objeto respectiva.

En la herramienta de autoría 102a, el audio procesado desde la DAW 204 se introduce en un componente de renderización binaural 206. Este componente incluye una función de procesamiento de audio que produce una salida de audio binaural 210, así como metadatos de renderización binaural 208 y metadatos de tipo de medios espaciales 212. El audio 210 y los componentes de metadatos 208 y 212 forman un flujo de bits de audio codificado con una carga útil de metadatos binaurales 214. En general, el componente de audio 210 comprende audio basado en canales y objetos que se pasa al flujo de bits 214 con los componentes de metadatos 208 y 212; sin embargo, debe tenerse en cuenta que el componente de audio 210 puede ser audio multicanal estándar, audio renderizado binauralmente o una combinación de estos dos tipos de audio. Un componente de renderización binaural 206 también incluye una función de entrada de metadatos binaurales que produce directamente una salida de auriculares 216 para la conexión directa a los auriculares. Para la realización de la figura 2A, los metadatos para la renderización binaural se generan en el momento de la mezcla dentro de la herramienta de autoría 102a. En una realización alternativa, los metadatos pueden generarse en el tiempo de codificación, como se muestra con referencia a la figura 2B. Como se muestra en la figura 2A, un mezclador 203 usa una aplicación o herramienta para crear datos de audio y los metadatos binaurales y espaciales. El mezclador 203 proporciona entradas a la DAW 204. Alternativamente, también puede proporcionar entradas directamente al proceso de renderización binaural 206. En una realización, el mezclador recibe la salida de audio de auriculares 216 para que el mezclador pueda controlar la entrada del efecto del audio y los metadatos. Esto constituye efectivamente un ciclo de realimentación en el que el mezclador recibe la salida de audio renderizada de los auriculares a través de los auriculares para determinar si se necesitan cambios de entrada. El mezclador 203 puede ser una persona que opera un equipo, tal como una consola o un ordenador de mezclas, o puede ser un proceso automatizado que se controla de forma remota o se preprograma.

La figura 2B es un diagrama de bloques de una herramienta de autoría utilizada en un sistema de renderización en auriculares basado en objetos, en una realización alternativa. En esta realización, los metadatos para la renderización binaural se generan en el tiempo de codificación, y el codificador ejecuta un clasificador de contenido y un generador de metadatos para generar metadatos adicionales a partir de contenido basado en canales legado. Para la herramienta de autoría 102b de la figura 2B, el contenido multicanal legado 220, que no incluye ningún objeto de audio, sino el audio basado en canales sólo se introduce en una herramienta de codificación y un componente de emulación de auriculares de renderización 226. El contenido basado en objetos 222 también se introduce por separado en este componente. El contenido legado basado en canales 220 se puede introducir primero en un preprocesador de segmentación de audio opcional 224 para la separación de diferentes tipos de contenido para su renderización individual. En la herramienta de autoría 102b, el componente de renderización binaural 226 incluye una función de emulación de auriculares que produce una salida de audio binaural 230, así como metadatos de renderización binaural 228 y metadatos de tipo de medios espaciales 232. Los componentes de audio 230 y de metadatos 228 y 232 forman un flujo de bits de audio codificado con una carga útil de metadatos binaurales 236. Como se indicó anteriormente, el componente de audio 230 comprende audio basado en canales y objetos que se pasa al flujo de bits 236 con los componentes de metadatos 228 y 232.

Cuando se introduce contenido legado, el flujo de bits de audio codificado de salida podría contener datos o metadatos de audio de subcomponentes separados explícitamente que describen implícitamente el tipo de contenido que permite que el punto final receptor realice la segmentación y procese de manera apropiada cada subcomponente. El componente de renderización binaural 226 también incluye una función de entrada de metadatos binaurales que produce directamente una salida de auriculares 234 para conexión directa a los auriculares. Como se muestra en la figura 2B, se puede incluir un mezclador opcional (persona o proceso) 223 para supervisar la salida de auriculares 234 con el fin de introducir y modificar datos de audio y entradas de metadatos que pueden proporcionarse directamente al proceso de renderización 226.

Con respecto al tipo de contenido y al funcionamiento del clasificador de contenido, el audio generalmente se clasifica como uno de varios tipos de contenido definidos, tal como diálogo, música, ambiente, efectos especiales, etc. Un objeto puede cambiar el tipo de contenido a lo largo de su duración, pero en cualquier momento específico es generalmente solo un tipo de contenido. En una realización, el tipo de contenido se expresa como una probabilidad de que el objeto sea un tipo particular de contenido en cualquier momento. Así, por ejemplo, un objeto de diálogo constante se expresaría como un objeto de diálogo con una probabilidad del cien por cien, mientras que

un objeto que se transforma de diálogo a música puede expresarse como cincuenta por ciento diálogo/cincuenta por ciento música. El procesamiento de objetos que tienen diferentes tipos de contenido podría realizarse promediando sus respectivas probabilidades para cada tipo de contenido, seleccionando las probabilidades de tipo de contenido para el objeto más dominante dentro de un grupo de objetos, o un solo objeto a lo largo del tiempo, o alguna otra combinación lógica de medidas de tipo de contenido. El tipo de contenido también puede expresarse como un vector n-dimensional (donde n es el número total de diferentes tipos de contenido, por ejemplo, cuatro, en el caso de diálogo/música/ambiente/efectos). Los metadatos de tipo de contenido pueden incorporarse como una definición de metadatos de tipo de contenido combinada, donde una combinación de tipos de contenido refleja las distribuciones de probabilidad que se combinan (por ejemplo, un vector de probabilidades de música, habla, etc.).

Con respecto a la clasificación de audio, en una realización, el proceso opera en una base de cuadro por tiempo para analizar la señal, identificar características de la señal y comparar las características identificadas con las características de clases conocidas para determinar cómo de bien coinciden las características del objeto con las características de una clase particular. Basándose como de bien coinciden las características con una clase particular, el clasificador puede identificar la probabilidad de que un objeto pertenezca a una clase particular. Por ejemplo, si en el tiempo $t = T$ las características de un objeto coinciden muy bien con las características de diálogo, entonces el objeto se clasificaría como diálogo con una alta probabilidad. Si, en el tiempo $t = T + N$, las características de un objeto coinciden muy bien con las características de música, el objeto se clasificaría como música con una alta probabilidad. Finalmente, si en el momento $T = T + 2N$ las características de un objeto no coinciden particularmente bien con el diálogo o la música, el objeto podría clasificarse como 50% música y 50% diálogo. Por lo tanto, en una realización, basada en las probabilidades de tipo de contenido, el contenido de audio se puede separar en diferentes subseñales correspondientes a los diferentes tipos de contenido. Esto se logra, por ejemplo, enviando un porcentaje de la señal original a cada subseñal (ya sea en una banda ancha o en una subbanda de frecuencia), en una proporción motivada por las probabilidades de tipo de medios calculadas.

Con referencia a la figura 1, la salida de la herramienta de autoría 102 se introduce en el renderizador 112 para renderizar como salida de audio para su reproducción a través de auriculares u otros dispositivos de punto final. La figura 3A es un diagrama de bloques de un componente de renderización 112a utilizado en un sistema de renderización en auriculares basado en objetos, en una realización. La figura 3A ilustra el procesamiento 113, la renderización binaural 114 y el postprocesamiento 115 de subcomponentes del renderizador 112 con mayor detalle. Desde la herramienta de autoría 102, los metadatos y el audio se introducen en los componentes de procesamiento o preprocesamiento con la forma de un flujo de bits de audio codificado 301. Los metadatos 302 se introducen en un componente de procesamiento de metadatos 306, y el audio 304 se introduce en un preprocesador de segmentación de audio opcional 308. Como se muestra con referencia a las figuras 2A y 2B, la herramienta de autoría puede realizar la segmentación de audio mediante unos preprocesadores 202 o 224. Si la herramienta de autoría no realiza tal segmentación de audio, el renderizador puede realice esta tarea mediante el preprocesador 308. Los metadatos procesados y el audio segmentado se introducen entonces en un componente de renderización binaural 310. Este componente realiza ciertas funciones de renderización específicas de auriculares, tales como posicionamiento 3D, control de distancia, procesamiento del tamaño de la cabeza, etc. El audio renderizado binaural se introduce luego en el postprocesador de audio 314, que aplica ciertas operaciones de audio, tales como gestión de nivel, ecualización, compensación o cancelación de ruido, etc. El audio postprocesado se emite 312 para reproducirlo a través de los auriculares 116 o 118. Para una realización en la que los auriculares o el dispositivo de reproducción 104 están equipados con sensores y/o micrófonos para retroalimentar al renderizador, se introducen los datos de micrófono y sensor 316 en al menos uno de los componentes de procesamiento de metadatos 306, el componente de renderización binaural 310 o el componente de postprocesamiento de audio 314. Para auriculares estándar que no están equipados con sensores, el seguimiento de la cabeza podría reemplazarse por una "oscilación" de la cabeza generada de forma pseudoaleatoria más sencilla que imita continuamente el cambio de pequeños movimientos de la cabeza. Esto permite que cualesquiera datos ambientales u operativos relevante en el punto de reproducción se utilicen por el sistema de renderización para modificar aún más el audio para contrarrestar o mejorar ciertas condiciones de reproducción.

Como se mencionó anteriormente, la segmentación del audio puede ser realizada por la herramienta de autoría o el renderizador. Para la realización en la que el audio está previamente segmentado, el renderizador procesa este audio directamente. La figura 3B es un diagrama de bloques de un componente de renderización utilizado en un sistema de renderización en auriculares basado en objetos, en esta realización alternativa. Como se muestra para el renderizador 112b, el flujo de bits de audio codificado 321 de la herramienta de autoría se proporciona en sus partes constituyentes de entrada de metadatos 322 al componente de procesamiento de metadatos 326, y el audio 324 al componente de renderización binaural 330. Para la realización de la figura 3B, el audio se segmenta previamente mediante un proceso de presegmentación de audio 202 o 224 en la herramienta de autoría apropiada. El componente de renderización binaural 330 realiza ciertas funciones de renderización específicas de auriculares, tales como posicionamiento 3D, control de distancia, procesamiento del tamaño de la cabeza, etc. El audio renderizado binaural se introduce luego en el postprocesador de audio 334, que aplica ciertas operaciones de audio, tales como gestión de nivel, ecualización, compensación o cancelación de ruido, etc. El audio postprocesado se emite 332 para reproducirlo a través de los auriculares 116 o 118. Para una realización en la que los auriculares o el dispositivo de reproducción 104 están equipados con sensores y/o micrófonos para retroalimentar al procesador, el micrófono y los datos de sensor 336 se devuelven a al menos uno del componente de procesamiento de metadatos

326, el componente de renderización binaural 330 o el componente de postprocesamiento de audio 334. Los sistemas de autoría y renderización de las figuras 2A, 2B, 3A y 3B permiten a los autores de contenido crear y codificar metadatos de renderización binaural específicos en el momento de la creación de contenido utilizando la herramienta de autoría 102. Esto permite que los datos de audio se usen para instruir al renderizador para que procese partes del contenido de audio con diferentes algoritmos o con diferentes configuraciones. En una realización, la herramienta de autoría 102 representa una estación de trabajo o aplicación informática que permite a un creador de contenido (autor) seleccionar o crear contenido de audio para su reproducción y definir ciertas características para cada uno de los canales y/u objetos que componen el contenido de audio. La herramienta de autoría puede incluir una interfaz de consola de tipo mezclador o una renderización de interfaz gráfica de usuario (GUI) de una consola de mezcla. La figura 5 ilustra una GUI de herramienta de autoría que puede usarse con realizaciones de un sistema de renderización en auriculares, en una realización. Como se puede ver en la pantalla de GUI 500, el autor puede establecer una serie de características diferentes, como niveles de ganancia, características de baja frecuencia, ecualización, panning, posición y densidad del objeto, retrasos, desvanecimientos, etc. Para la realización mostrada, la entrada del usuario se facilita mediante el uso de controles deslizantes virtuales para que el autor especifique valores de configuración, aunque también son posibles otros medios de entrada virtualizados o directos, como entrada directa de texto, configuraciones de potenciómetros, diales giratorios, etc. Al menos algunas de las configuraciones de parámetros introducidas por el usuario están codificadas como metadatos que están asociados con los canales u objetos de audio relevantes para su transporte con el contenido de audio. En una realización, los metadatos se pueden empaquetar como parte de una carga útil adicional de auriculares en los circuitos de códec (codificador/decodificador) en el sistema de audio. Al usar dispositivos habilitados, los metadatos en tiempo real que codifican ciertas condiciones operativas y ambientales (por ejemplo, seguimiento de la cabeza, detección del tamaño de la cabeza, detección de la sala, condiciones ambientales, niveles de ruido, etc.) pueden proporcionarse en vivo a la renderización binaural. La renderización binaural combina el contenido de metadatos creado y los metadatos generados localmente en tiempo real para proporcionar al usuario una experiencia de escucha optimizada. En general, los controles de objeto proporcionados por las herramientas de autoría y las interfaces de entrada de usuario permiten al usuario controlar ciertos parámetros específicos de auriculares importantes, tales como modos de renderización dinámica binaural y estéreo-bypass, ganancia LFE (elemento de baja frecuencia) y ganancias de objeto, inteligencia de medios y controles dependientes del contenido. Más específicamente, el modo de renderización podría seleccionarse en función del tipo de contenido u objeto entre estéreo (Lo/Ro), estéreo matricial (Lt/Rt), utilizando una combinación de retardos de tiempo interaural y amplitud estéreo o panning de intensidad, o renderización binaural completa (es decir, combinación de retrasos y niveles temporales interaural, así como señales espectrales dependientes de la frecuencia). Además, se puede especificar un punto de cruce de frecuencia para volver al procesamiento estéreo por debajo de una frecuencia dada. Las ganancias de baja frecuencia también se pueden especificar para atenuar componentes de baja frecuencia o contenido LFE. El contenido de baja frecuencia también podría transportarse por separado a auriculares habilitados, como se describe con mayor detalle a continuación. Se pueden especificar otros metadatos por tipo de contenido o por canal/objeto, tal como el modelo de sala generalmente descrito por una ganancia directa/reverberante y un tiempo de reverberación dependiente de la frecuencia y la correlación cruzada de objetivo interaural. También podría incluir otros modelos más detallados de la sala (por ejemplo, posiciones de reflexiones tempranas, ganancias y ganancias de reverberación tardía). También podría incluir filtros directamente especificados que modelen una respuesta de sala particular. La estructura y la lista de programas de ejemplo que se muestran en el Apéndice 1 adjunto a esta memoria dan un ejemplo de tales metadatos y cómo se pueden señalar según algunas realizaciones. Otros metadatos incluyen banderas de adaptación al alabeo según la pantalla (que controlan cómo los objetos se reasignan para ajustarse a la relación de aspecto de la pantalla y al ángulo de visión en función de la distancia). Finalmente, un indicador relativo al oyente (es decir, para aplicar o no información de seguimiento de la cabeza), escalado preferido (especifíquese un relación de tamaño/aspecto predeterminado de una "sala virtual" para renderizar el contenido utilizado para escalar las posiciones de los objetos, así como reasignar a la pantalla (en función del tamaño de la pantalla del dispositivo y la distancia al dispositivo)) así como el exponente del modelo de distancia que controla la ley de atenuación de la distancia (por ejemplo, $1/(1+r^a)$) también es posible señalar grupos de parámetros o "preconfiguraciones" que pueden aplicarse a diferentes canales/objetos o dependiendo del tipo de contenido.

Como se muestra con respecto a los componentes de presegmentación de la herramienta de autoría y/o renderización, los diferentes tipos de contenido (por ejemplo, diálogo, música, efectos, etc.) pueden procesarse de manera diferente en función de la intención del autor y la configuración de renderización óptima. La separación del contenido en función del tipo u otra característica destacada se puede lograr a priori durante la autoría, por ejemplo, manteniendo manualmente el diálogo separado en su propio conjunto de pistas u objetos, o a posteriori en vivo antes de su renderizado en el dispositivo receptor. Se pueden usar herramientas de inteligencia de medios adicionales durante la autoría para clasificar el contenido según diferentes características y generar canales u objetos adicionales que pueden llevar diferentes conjuntos de metadatos de renderización. Por ejemplo, teniendo conocimiento de los troncos (música, diálogo, Foley, efectos, etc.) y una mezcla envolvente asociada (por ejemplo, 5.1), los clasificadores de medios podrían ser entrenados para el proceso de creación de contenido con el fin de desarrollar un modelo para identificar diferentes proporciones de mezclas tronco. Se podría emplear una técnica de separación de fuente asociada para extraer los troncos aproximados utilizando funciones de ponderación derivadas del clasificador de medios. A partir de los troncos extraídos, los parámetros binaurales, que se codificarían como metadatos, se pueden aplicar durante la autoría. En una realización, se aplica un proceso reflejado en el dispositivo

del usuario final mediante el cual el uso de los parámetros de metadatos decodificados crearía una experiencia sustancialmente similar a la de la creación de contenido.

En una realización, las extensiones de las herramientas de autoría de estudio existentes incluyen supervisión binaural y grabación de metadatos. Los metadatos típicos capturados en tiempo de autoría incluyen: información de posición/tamaño de canal/objeto para cada canal y objeto de audio, ajuste de ganancia de canal/objeto, metadatos dependientes del contenido (pueden variar según el tipo de contenido), omitir banderas para indicar que deben usarse configuraciones, tales como renderización estéreo/izquierda/derecha, en lugar de binaural, puntos de cruce y niveles que indican que la frecuencia de bajos por debajo del punto de cruce se debe omitir y/o atenuar, y la información del modelo de sala para describir una ganancia directa/reverberante y un tiempo de reverberación dependiente de la frecuencia u otras características, tales como reflexiones tempranas y ganancia de reverberación tardía. Otros metadatos dependientes del contenido podrían proporcionar una funcionalidad de deformación a pantalla que reasigna las imágenes para ajustarse a la relación de aspecto de la pantalla o cambiar el ángulo de visualización en función de la distancia. La información de seguimiento de la cabeza se puede aplicar para proporcionar una experiencia relativa al oyente. También se podrían usar metadatos que implementen un exponente de modelo de distancia que controle la distancia en función de la ley de atenuación (por ejemplo, $1/(1+r^a)$). Estas representan solo ciertas características que pueden ser codificadas por los metadatos, y otras características también pueden ser codificadas.

La figura 4 es un diagrama de bloques que proporciona una visión general del sistema de renderización binaural de doble extremo, en una realización. En una realización, el sistema 400 proporciona metadatos dependientes del contenido y configuraciones de renderización que afectan la forma en cómo se renderizan los diferentes tipos de contenido de audio. Por ejemplo, el contenido de audio original puede comprender diferentes elementos de audio, tales como diálogo, música, efectos, sonidos ambientales, transitorios, etc. Cada uno de estos elementos puede renderizarse de manera óptima de diferentes maneras, en lugar de limitarlos a que se rendericen de una sola manera. Para la realización del sistema 400, la entrada de audio 401 comprende audio híbrido de canal más objetos. El audio se introduce en un codificador 402 que agrega o modifica metadatos asociados con los objetos y canales de audio. Como se muestra en el sistema 400, el audio se introduce en un componente de supervisión de auriculares 410 que aplica herramientas paramétricas ajustables por el usuario para controlar el procesamiento, ecualización, mezcla descendente y otras características apropiadas para la reproducción en auriculares. El codificador 402 incrusta el conjunto de parámetros optimizado por el usuario (M) como metadatos o metadatos adicionales para formar un flujo de bits que se transmite al decodificador 404. El decodificador 404 decodifica los metadatos y el conjunto de parámetros M del objeto y el audio basado en canales para controlar el procesamiento de auriculares y componente de mezcla descendente 406, que produce una salida de audio 408 optimizada y mezclada de auriculares (por ejemplo, 5.1 a estéreo) a los auriculares. Aunque se ha implementado cierto procesamiento dependiente del contenido en los sistemas actuales y las cadenas de postprocesamientos, generalmente no se ha aplicado a la renderización binaural, como se ilustra en el sistema 400 de la figura 4.

Como se muestra en la figura 4, pueden proporcionarse ciertos metadatos por un componente de supervisión de auriculares 410 que proporciona herramientas paramétricas ajustables específicas del usuario para controlar la reproducción específica de los auriculares. Tal componente puede configurarse para proporcionar al usuario cierto grado de control sobre la renderización en auriculares para auriculares tradicionales 118 que reproducen pasivamente contenido de audio transmitido. En las realizaciones, el dispositivo de punto final es un auricular habilitado 116 que incluye sensores y, opcionalmente, cierto grado de capacidad de procesamiento para generar metadatos o datos de señal que pueden codificarse como metadatos compatibles para modificar aún más los metadatos creados con el fin de optimizar el contenido de audio se renderización a través de auriculares. Por lo tanto, en el extremo receptor del contenido. La renderización se realiza en vivo y puede dar cuenta de los datos de matriz de sensores generados localmente que pueden generarse mediante un auricular o un dispositivo móvil real 104 al que están conectados los auriculares, y tales metadatos generados por hardware pueden combinarse además con los metadatos creados por el creador de contenido en el momento de la autoría para mejorar la experiencia de renderización binaural.

Como se indicó anteriormente, en algunas realizaciones, el contenido de baja frecuencia puede transportarse por separado a unos auriculares habilitados que permiten más de una entrada estéreo (típicamente 3 o 4 entradas de audio), o codificarse y modularse en las frecuencias más altas de las principales formas de onda estéreo transportadas a un auricular con una sola entrada estéreo. Esto permitiría que tenga lugar un mayor procesamiento de baja frecuencia en los auriculares (por ejemplo, enrutamiento a controladores específicos optimizados para bajas frecuencias). Tales auriculares pueden incluir controladores y/o filtros específicos de baja frecuencia más circuitos cruzados y de amplificación para optimizar la reproducción de señales de baja frecuencia.

En una realización, se proporciona un enlace desde los auriculares al componente de procesamiento de auriculares en el lado de reproducción para permitir la identificación manual de los auriculares para la carga automática de preajustes de auriculares u otra configuración de los auriculares. Tal enlace puede implementarse como un enlace inalámbrico o por cable desde los auriculares al proceso de auriculares 406 de la figura 4, por ejemplo. La identificación se puede usar para configurar los auriculares objetivo o con el fin de enviar contenido específico o contenido específicamente renderizado a un conjunto específico de auriculares si se utilizan múltiples auriculares objetivo. El identificador de auriculares puede estar incorporado en cualquier código alfanumérico o binario

apropiado que sea procesado por el proceso de renderización como parte de los metadatos o de una operación de procesamiento de datos separada.

La figura 6 ilustra un auricular habilitado que comprende uno o más sensores que detectan las condiciones de reproducción para codificarlas como metadatos utilizados en un sistema de renderización en auriculares, en una realización. Los diversos sensores pueden estar dispuestos en una matriz de sensores que puede usarse para proporcionar metadatos en vivo a la renderización en el momento de renderización. Para el ejemplo de unos auriculares 600 de la fig. 6, los sensores incluyen un sensor de alcance (tal como un infrarrojo IR o una cámara de tiempo de vuelo TOF) 602, un sensor de tensión/tamaño de cabeza 604, un sensor giroscópico 606, un micrófono externo (o par) 610, un procesador de cancelación de ruido ambiental 608, un micrófono interno (o par) 612, entre otros sensores apropiados. Como se muestra en la figura 6, la matriz de sensores puede comprender tanto sensores de audio (es decir, micrófonos) como sensores de datos (por ejemplo, sensores de orientación, tamaño, tensión/tensión y alcance). Específicamente para usar con auriculares, los datos de orientación se pueden usar para "bloquear" o rotar el objeto de audio espacial según el movimiento de la cabeza del oyente, los sensores de tensión o los micrófonos externos se pueden usar para inferir el tamaño de la cabeza del oyente (por ejemplo, al supervisar la correlación cruzada de audio en dos micrófonos externos ubicados en los audífonos) y ajustar los parámetros de renderización binaurales relevantes (por ejemplo, retrasos temporales interaurales, tiempo de reflexión del hombro, etc.). Los sensores de alcance 602 se pueden usar para evaluar la distancia a la pantalla en caso de una reproducción móvil de A/V y corregir la ubicación de los objetos en pantalla para tener en cuenta el ángulo de visión dependiente de la distancia (es decir, ampliar los objetos a medida que se acerca la pantalla al oyente) o ajustar la ganancia global y el modelo de sala para transmitir la renderización de distancia adecuada. Tal función de sensor es útil si el contenido de audio es parte de un contenido de A/V que se reproduce en dispositivos que pueden variar desde teléfonos móviles pequeños (por ejemplo, tamaño de pantalla de 2 a 4" pulgadas) a tabletas (por ejemplo, tamaño de pantalla de 7 a 10" pulgadas) a ordenadores portátiles (por ejemplo, tamaño de pantalla de 15-17"). Además, los sensores también se pueden usar para detectar y configurar automáticamente el enrutamiento de las salidas de audio izquierda y derecha a los transductores correctos, sin requerir una orientación específica a priori o marcas explícitas de "Izquierda/Derecha" en los auriculares.

Como se muestra en la figura 1, el contenido de audio o A/V transmitido a los auriculares 116 o 118 puede proporcionarse a través de un dispositivo de mano o portátil 104. En una realización, el dispositivo 104 puede incluir en sí mismo uno o más sensores. Por ejemplo, si el dispositivo es una consola de juegos portátil o un controlador de juegos, se pueden proporcionar ciertos sensores giroscópicos y acelerómetros para seguir el movimiento y la posición de los objetos. Para esta realización, el dispositivo 104 al que está conectado el auricular también puede proporcionar datos de sensor adicionales, tales como orientación, tamaño de la cabeza, cámara, etc., como metadatos del dispositivo.

Para esta realización, se implementan ciertos medios de comunicación de auriculares a dispositivo. Por ejemplo, los auriculares se pueden conectar al dispositivo a través de un enlace digital con cable o inalámbrico, o a un enlace de audio analógico (entrada de micrófono), en cuyo caso los metadatos serán modulados en frecuencia y se agregarán a la entrada de micrófono analógico. La figura 7 ilustra la conexión entre un auricular y un dispositivo 104 que incluye un procesador de sensor de auriculares 702, en una realización. Como se muestra en el sistema 700, los auriculares 600 transmiten ciertos datos de sensor, audio y micrófono 701 a través de un enlace cableado o inalámbrico al procesador de sensor de auriculares 702. Los datos procesados del procesador 702 pueden comprender audio analógico con metadatos 704 o una salida de audio espacial 706. Como se muestra en la figura 7, cada una de las conexiones comprende un enlace bidireccional entre los auriculares, el procesador y las salidas. Esto permite que los datos de sensor y el micrófono se transmitan hacia y desde los auriculares y el dispositivo para la creación o modificación de metadatos apropiados. Además de los metadatos generados por hardware, también se pueden proporcionar controles de usuario para complementar o generar metadatos apropiados si no están disponibles a través de matrices de sensores de hardware. Los controles de usuario de ejemplo pueden incluir: énfasis de elevación, interruptor de encendido/apagado binaural, radio o tamaño de sonido preferido y otras características similares. Tales controles de usuario pueden proporcionarse a través de elementos de interfaz de hardware o software asociados con el componente de procesador de auriculares, dispositivo de reproducción y/o auriculares.

La figura 8 es un diagrama de bloques que ilustra los diferentes componentes de metadatos que pueden usarse en un sistema de renderización en auriculares, en una realización. Como se muestra en el diagrama 800, los metadatos procesados por el procesador de auriculares 806 comprenden metadatos creados, tales como los producidos por la herramienta de autoría 102 y la consola de mezclas 500, y los metadatos generados por hardware 804. Los metadatos generados por hardware 804 incluyen metadatos del lado de los auriculares proporcionados o generados a partir de datos enviados desde el auricular 810, y opcionalmente metadatos de entrada del usuario, y/o metadatos del lado del dispositivo proporcionados o generados a partir de datos enviados desde el dispositivo 808.

En una realización, se procesan los metadatos generados por hardware 804 y/o creados 802 en un componente de renderización binaural 114 del renderizador 112. Los metadatos proporcionan control sobre canales u objetos de audio específicos para optimizar la reproducción en los auriculares 116 o 118. La figura 9 ilustra unos componentes funcionales de un componente de renderización binaural para el procesamiento de auriculares, en una realización. Como se muestra en el sistema 900, el decodificador 902 emite la señal multicanal o las pistas de canal más objeto junto con el conjunto de parámetros decodificados, M, para controlar el procesamiento de auriculares realizado por

el procesador de auriculares 904. El procesador de auriculares 904 también recibe ciertas actualizaciones de parámetros espaciales 906 de un dispositivo de seguimiento basado en cámara o basado en sensor 910. El dispositivo de seguimiento 910 es un dispositivo de seguimiento facial o de seguimiento de la cabeza que mide ciertos parámetros angulares y posicionales (r , θ , ϕ) asociados con la cabeza del usuario. Los parámetros espaciales pueden corresponder a la distancia y ciertos ángulos de orientación, como guiñada, cabeceo y balanceo. Se puede actualizar un conjunto original de parámetros espaciales, x , a medida que se procesan los datos de sensor 910. Estas actualizaciones de parámetros espaciales Y se pasan luego al procesador de auriculares 904 para una modificación adicional del conjunto de parámetros M . Los datos de audio procesados se transmiten entonces a una etapa de postprocesamiento 908 que realiza cierto procesamiento de audio, tal como corrección de timbre, filtrado, mezcla descendente, y otros procesos relevantes. Luego, el ecualizador 912 ecualiza el audio y lo transmite a los auriculares. En una realización, el ecualizador 912 puede realizar la ecualización con o sin usar una transformada de la relación de presión-división (PDR), como se describe con más detalle en la descripción que sigue.

La figura 10 ilustra un sistema de renderización binaural para renderizar objetos de audio en un sistema de renderización en auriculares, en una realización. La figura 10 ilustra algunos de los componentes de señal a medida que se procesan mediante un procesador de auriculares binaural. Como se muestra en el diagrama 1000, los componentes de audio del objeto se introducen en un desmezclador 1002 que separa los componentes directos y difusos (por ejemplo, directo de la ruta de reverberación) del audio. El componente directo se introduce en un componente de mezcla descendente 1006 que mezcla canales envolventes (por ejemplo, envolvente 5.1) a estéreo con información de desplazamiento de fase. El componente directo también se introduce en una renderización binaural de contenido directo 1008. Ambos componentes de dos canales se introducen en un ecualizador de timbre dinámico 1012. Para la entrada de audio basada en objetos, la posición del objeto y las señales de control del usuario se introducen en un componente piloto de virtualizador 1004. Esto genera una posición de objeto escalada que se introduce en el renderizador binaural 1008 junto con el componente directo. El componente difuso del audio se introduce en una renderización binaural separada 1010, y se combina con el contenido directo renderizado por un circuito sumador antes de su salida como audio de salida de dos canales.

La figura 11 ilustra una representación más detallada del sistema de renderización binaural de la figura 10, en una realización. Como se muestra en el diagrama 1100 de la figura 11, el audio multicanal y basado en objetos se introduce en el desmezclador 1102 para su separación en componentes directos y difusos. El contenido directo es procesado por el renderizador binaural directo 1118, y el contenido difuso es procesado por el renderizador binaural difuso 1120. Después de la mezcla descendente 1116 y la ecualización de timbre 1124 del contenido directo, los componentes de audio directo y difuso se combinan mediante un circuito sumador para su postprocesamiento, tal como por el ecualizador de auriculares 1122, y otros circuitos posibles. Como se muestra en la figura 11, ciertos datos de entrada y realimentación del usuario se utilizan para modificar la renderización binaural del contenido difuso en el renderizador binaural difuso 1120. Para la realización del sistema 1100, el sensor de entorno de reproducción 1106 proporciona datos sobre las propiedades de la sala de audición y la estimación del ruido (niveles de sonido ambiental), el sensor de seguimiento de cabeza/cara 1108 proporciona datos de posición, orientación y tamaño de cabeza, el sensor de seguimiento de dispositivo 1110 proporciona datos de posición del dispositivo y la entrada del usuario 1112 proporciona datos de radio de reproducción. Estos datos pueden ser proporcionados por sensores ubicados en los auriculares 116 y/o el dispositivo 104. Los diversos datos de sensor y los datos de entrada del usuario se combinan con metadatos de contenido, que proporcionan información sobre la posición del objeto y los parámetros de la sala en un componente piloto de virtualizador 1104. Este componente también recibe información de energía directa y difusa del desmezclador 1102. El piloto de virtualizador 1104 emite datos, incluida posición del objeto, posición/orientación/tamaño de la cabeza, parámetros de la sala y otra información relevante para el renderizador binaural de contenido difuso 1120. De esta manera, el contenido difuso del audio de entrada se ajusta para acomodar los datos de entrada del sensor y del usuario.

Si bien se logra un rendimiento óptimo del piloto de virtualizador cuando se reciben datos de sensor, datos de entrada del usuario y metadatos de contenido, es posible lograr un rendimiento beneficioso del piloto de virtualizador incluso en ausencia de una o más de estas entradas, en ejemplos no cubiertos por las reivindicaciones adjuntas. Por ejemplo, cuando se procesa contenido legado (por ejemplo, flujos de bits codificados que no contienen metadatos de renderización binaural) para la reproducción en auriculares convencionales (por ejemplo, auriculares que no incluyen diversos sensores, micrófonos, etc.), aún se puede obtener un resultado beneficioso al proporcionar la energía directa y las salidas de energía difusa del desmezclador 1102 al piloto de virtualizador 1104 para generar información de control para el renderizador de contenido binaural difuso 1120, incluso en ausencia de una o más entradas adicionales al piloto de virtualizador.

En una realización, el sistema de renderización 1100 de la figura 11 permite que el renderizador de auriculares binaural proporcione de manera eficiente una individualización basada en la diferencia de tiempo interaural (ITD) y la diferencia de nivel interaural (ILD) y la detección del tamaño de la cabeza. La ILD y la ITD son señales importantes para el acimut, que es el ángulo de una señal de audio con relación a la cabeza cuando se produce en el plano horizontal. La ITD se define como la diferencia en el tiempo de llegada de un sonido entre dos oídos, y el efecto ILD usa diferencias en el nivel de sonido que entra en los oídos para proporcionar señales de localización. En general, se acepta que las ITD se usan para localizar sonidos de baja frecuencia y las ILD se usan para localizar sonidos de alta frecuencia, mientras que ambos se usan para contenido que contiene frecuencias altas y bajas.

El sistema de renderización 1100 también permite acomodar el control de distancia de origen y el modelo de sala. Además, permite la extracción y el procesamiento de contenido directo versus difuso/reverberación (seco/húmedo), la optimización de los reflejos de la sala y la coincidencia tímbrica.

Modelo HRTF

5 En la reproducción de audio espacial, se virtualizan ciertas señales de la fuente de sonido. Por ejemplo, los sonidos destinados a ser escuchados por detrás de los oyentes pueden ser generados por altavoces ubicados físicamente detrás de ellos y, en consecuencia, todos los oyentes perciben estos sonidos como provenientes de atrás. Con la renderización espacial virtual en auriculares, por otro lado, la percepción del audio desde atrás se controla mediante las funciones de transferencia relacionadas con la cabeza (HRTF) que se utilizan para generar la señal binaural. En una realización, el sistema de procesamiento de auriculares basado en metadatos 100 puede incluir ciertos mecanismos de modelado HRTF. La base de tal sistema se basa generalmente en el modelo estructural de la cabeza y el torso. Este enfoque permite construir algoritmos sobre el modelo central en un enfoque modular. En este algoritmo, los algoritmos modulares se denominan "herramientas". Además de proporcionar señales ITD e ILD, el enfoque del modelo proporciona un punto de referencia con respecto a la posición de los oídos en la cabeza y, más ampliamente, a las herramientas que están construidas sobre el modelo. El sistema podría ajustarse o modificarse según las características antropométricas del usuario. Otros beneficios del enfoque modular permiten acentuar ciertas características para amplificar señales espaciales específicas. Por ejemplo, ciertas señales podrían exagerarse más allá de lo que un filtro binaural acústico impartiría a un individuo. La figura 12 es un diagrama del sistema que muestra las diferentes herramientas utilizadas en un sistema de modelado HRTF utilizado en un sistema de renderización en auriculares, en una realización. Como se muestra en la figura 12, ciertas entradas que incluyen azimut, elevación, fs y alcance se introducen en la etapa de modelado 1204, después de que se filtren al menos algunos componentes de entrada 1202. En una realización, la etapa de filtro 1202 puede comprender un modelo de filtro de muñeco de nieve que consiste en una cabeza esférica en la parte superior de un cuerpo esférico y tiene en cuenta las contribuciones del torso, así como de la cabeza, al HRTF. La etapa de modelado 1204 calcula los modelos de pabellón auricular y torso y los componentes izquierdo y derecho (i, d) se procesan posteriormente 1206 para la salida final 1208.

Estructura de metadatos

Como se describió anteriormente, el contenido de audio procesado por el sistema de reproducción en auriculares comprende canales, objetos y metadatos asociados que proporcionan las señales espaciales y de procesamiento necesarias para optimizar la renderización del audio a través de auriculares. Tales metadatos se pueden generar como metadatos creados a partir de herramientas de autoría, así como metadatos generados por hardware desde uno o más dispositivos de punto final. La figura 13 ilustra una estructura de datos que permite la entrega de metadatos para un sistema de renderización en auriculares, en una realización. En una realización, la estructura de metadatos de la figura 13 está configurada para complementar los metadatos entregados en otras partes de un flujo de bits que pueden empaquetarse según un formato de audio basado en canales conocido, como Dolby digital AC-3 o sintaxis de flujo de bits AC-3 Mejorado. Como se muestra en la figura 13, la estructura de datos consta de un contenedor 1300 que contiene una o más cargas útiles de datos 1304. Cada carga útil se identifica en el contenedor utilizando un valor único de identificador de carga útil para proporcionar una indicación inequívoca del tipo de datos presentes en la carga útil. El orden de las cargas útiles dentro del contenedor no está definido. Las cargas útiles se pueden almacenar en cualquier orden, y un analizador debe poder analizar todo el contenedor para extraer cargas útiles relevantes e ignorar cargas útiles que no son relevantes o no son compatibles. Los datos de protección 1306 siguen a la carga útil final en el contenedor que puede ser utilizada por un dispositivo de decodificación para verificar que el contenedor y las cargas útiles dentro del contenedor están libres de errores. Una parte preliminar 1302 que contiene información de sincronización, versión e ID de clave precede a la primera carga útil en el contenedor.

45 La estructura de datos soporta la extensibilidad mediante el uso de versionado e identificadores para tipos de carga útil específicos. Las cargas útiles de metadatos pueden usarse para describir la naturaleza o la configuración del programa de audio que se entrega en un flujo de bits AC-3 o AC-3 Mejorado (u otro tipo), o pueden usarse para controlar algoritmos de procesamiento de audio diseñados para procesar aún más la salida del proceso de decodificación.

50 Los contenedores pueden definirse usando diferentes estructuras de programación, en base a las preferencias de implementación. La siguiente tabla ilustra un ejemplo de sintaxis de un contenedor, en una realización.

```

container()
{
  Version ..... 2
  if (version == 3)
  {
    version += variable_bits(2) ..... variable_bits(2)
  }
  key_id ..... 3
  if (key_id == 7)
  {
    key_id += variable_bits(3) ..... variable_bits(3)
  }
  payload_id ..... 5
  while (payload_id != 0x0)
  {
    if (payload_id == 0x1F)
    {
      payload_id += variable_bits(5) ..... variable_bits(5)
    }
    payload_config()
    payload_size ..... variable_bits(8)
    for (i = 0; i < payload_size; i++)
    {
      payload_bytes ..... 8
    }
  }
  protection()
}

```

Un ejemplo de posible sintaxis de los bits variables para la sintaxis de contenedor de ejemplo proporcionada anteriormente se muestra en la siguiente tabla:

Sintaxis
<pre> variable_bits (n_bits) { value = 0; Do { value += read n_bits read_more 1 if (read_more) { value <<= n_bits; value += (1<<n_bits); } } while (read_more); return value; } </pre>

- 5 Un ejemplo de una posible sintaxis de la configuración de carga útil para la sintaxis de contenedor de ejemplo proporcionada anteriormente se muestra en la siguiente tabla:

Sintaxis	No. of bits
<code>payload_config()</code>	
{	
Smploffste	1
if(smploffste) { smploffst }.....	11
Reserved	1
Duratione	1
if(duratione) { duration }.....	variable_bits(11)
Groupide	1
if (groupide) { groupid }.....	variable_bits(2)
Codecdatae	1
if (codecdatae) { reserved }.....	8
discard_unknown_payload	1
if(discard_unknown_payload == 0)	
{	
if (smploffste == 0)	
{	
payload_frame_aligned	1
if (payload_frame_aligned)	
{	
create_duplicate	1
remove_duplicate	1
}	
}	
if (smploffste == 1 payload_frame_aligned == 1)	
{	
Priority	5
proc_allowed	2
}	
}	
}	

Se proporcionan las definiciones de sintaxis anteriores como implementaciones de ejemplo, y no están destinadas a ser limitantes, ya que se pueden usar muchas otras estructuras de programa diferentes. En una realización, se codifican varios campos dentro de la estructura del contenedor y los datos de la carga útil utilizando un método conocido como bits variables. Este método permite la codificación eficiente de valores de campo pequeños con extensibilidad para poder expresar valores de campo arbitrariamente grandes. Cuando se utiliza la codificación variable_bit, el campo consta de uno o más grupos de n bits, con cada grupo seguido de un campo read_more de 1 bit. Como mínimo, la codificación de n bits requiere que se transmitan n + 1 bits. Todos los campos codificados con variable_bits se interpretan como enteros sin signo. Se pueden implementar otros aspectos de codificación diferentes según prácticas y métodos conocidos por los expertos en la materia. Las tablas anteriores y la figura 13 ilustran un ejemplo de estructura de metadatos, formato y contenido del programa que se proporciona con mayor detalle en el Apéndice 1. Cabe señalar que estos están destinados a representar una realización de ejemplo de una renderización de metadatos, y también son posibles otras definiciones y contenido de metadatos.

Ecuación y corrección de auriculares

- 15 Como se ilustra en la figura 1, el renderizador 112 puede realizar ciertas funciones de postprocesamiento 115. Una de tales funciones de postprocesamiento comprende la ecualización de auriculares, como se muestra en un elemento 912 de la figura 9. En una realización, la ecualización puede realizarse obteniendo medidas de respuesta al impulso del canal auditivo bloqueado para diferentes ubicaciones de auriculares para cada oído. La figura 14 ilustra un caso de ejemplo de tres mediciones de respuesta de impulso para cada oído, en una realización de un proceso de ecualización de auriculares. El proceso posterior de ecualización calcula la Transformada rápida de Fourier (FFT) de cada respuesta y realiza un promedio RMS (raíz cuadrática media) de la respuesta derivada. Las respuestas pueden ser variables, octava suavizada, ERB suavizada, etc. El proceso luego calcula la inversión, $|F(\omega)|$, del promedio RMS con restricciones en los límites (+/- x dB) de la respuesta de magnitud de inversión a frecuencias medias y altas. El proceso determina entonces el filtro de dominio tiempo.
- 25 El postproceso también puede incluir una función de transformada de cerrado a abierto. Este método de relación de presión-división (PDR) consiste en diseñar una transformada para que coincida con la impedancia acústica entre el tímpano y el campo libre para auriculares cerrados con modificaciones en términos de cómo se obtienen las mediciones para la transmisión de sonido en campo libre en función de la dirección de llegada del primer sonido. Esto permite indirectamente hacer coincidir las señales de presión del tímpano entre auriculares cerrados y condiciones equivalentes de campo libre sin requerir mediciones complicadas del tímpano.
- 30

La figura 15A ilustra un circuito para calcular la transmisión de sonido de campo libre, en una realización. El circuito 1500 se basa en un modelo de impedancia acústica de campo libre. En este modelo, $P_1(\omega)$ es la presión de Thevenin medida en la entrada del canal auditivo bloqueado con un altavoz a θ grados sobre el plano medio (por ejemplo, aproximadamente 30 grados a la izquierda y en frente del oyente) que implica la extracción de sonido directo de la respuesta de impulso medida. La medición $P_1(\omega)$ se puede hacer en la entrada del canal auditivo o a

una cierta distancia dentro (x mm) dentro del canal auditivo (o en el tímpano) desde la abertura para el mismo altavoz en la misma ubicación para medir $P_1(\omega)$ lo que implica la extracción de sonido directo de la respuesta de impulso medida.

Para este modelo, la relación de $P_2(\omega)/P_1(\omega)$ se calcula de la siguiente manera:

$$\frac{P_2(\omega)}{P_1(\omega)} = \frac{Z_{\text{tímpano}}(\omega)}{Z_{\text{tímpano}}(\omega) + Z_{\text{radiación}}(\omega)}$$

5

La figura 15B ilustra un circuito para calcular la transmisión de sonido de auriculares, en una realización. El circuito 1510 se basa en un modelo analógico de impedancia acústica de auriculares. En este modelo, P_4 se mide en la entrada del canal auditivo bloqueado con la medición de estado estable de los auriculares (promedio RMS), y la medida $P_5(\omega)$ se realiza en la entrada del canal auditivo o a una distancia dentro del canal auditivo (o en el tímpano) desde la abertura para la misma colocación de auriculares utilizada para medir $P_4(\omega)$.

10

Para este modelo, la relación $P_5(\omega)/P_4(\omega)$ se calcula de la siguiente manera:

$$\frac{P_5(\omega)}{P_4(\omega)} = \frac{Z_{\text{tímpano}}(\omega)}{Z_{\text{tímpano}}(\omega) + Z_{\text{auriculares}}(\omega)}$$

La relación de presión-división (PDR) se puede calcular utilizando la siguiente fórmula:

$$PDR(\omega) = \frac{P_2(\omega)}{P_1(\omega)} \div \frac{P_5(\omega)}{P_4(\omega)}$$

15 Los aspectos de los métodos y sistemas descritos en la presente memoria pueden implementarse en un entorno de red de procesamiento de sonido basado en ordenador apropiado para procesar archivos de audio digital o digitalizado. Algunas partes del sistema de audio adaptativo pueden incluir una o más redes que comprenden cualquier número deseado de máquinas individuales, incluido uno o más enrutadores (no mostrados) que sirven para almacenar y enrutar los datos transmitidos entre los ordenadores. Tal red se puede construir en diversos protocolos de red diferentes, y puede ser Internet, una red de área amplia (WAN), una red de área local (LAN) o cualquier combinación de las mismas. En una realización en la que la red comprende Internet, una o más máquinas pueden configurarse para acceder a Internet a través de programas de navegador web.

20

Uno o más de los componentes, bloques, procesos u otros componentes funcionales pueden implementarse mediante un programa informático que controla la ejecución de un dispositivo informático del sistema basado en procesador. También debe tenerse en cuenta que las diversas funciones descritas en la presente memoria pueden describirse utilizando cualquier número de combinaciones de hardware, firmware y/o como datos y/o instrucciones incorporados en diversos medios legibles por máquina o legibles por ordenador, en términos de su comportamiento, transferencia de registro, componente lógico y/u otras características. Los medios legibles por ordenador en los que se pueden incorporar tales datos y/o instrucciones formateados incluyen, sin limitación, medios de almacenamiento físicos (no transitorios) y no volátiles en diversas formas, tales como medios de almacenamiento ópticos, magnéticos o semiconductores.

25

30

A menos que el contexto claramente requiera lo contrario, a lo largo de la descripción y las reivindicaciones, las palabras "comprende", "que comprende", y similares deben interpretarse en un sentido inclusivo en lugar de en un sentido exclusivo o exhaustivo; es decir, en un sentido de "incluyendo, pero no limitado a". Las palabras que usan el número singular o plural también incluyen el número plural o singular respectivamente. Además, las palabras "aquí", "a continuación" "anteriormente", "más adelante", y las palabras de importancia similar se refieren a esta solicitud en su conjunto y no a partes particulares de esta aplicación. Cuando la palabra "o" se usa en referencia a una lista de dos o más elementos, esa palabra cubre todas las siguientes interpretaciones de la palabra: cualquiera de los elementos de la lista, todos los elementos de la lista y cualquier combinación de los elementos de la lista.

35

Aunque se han descrito una o más implementaciones a modo de ejemplo y en términos de las realizaciones específicas, debe entenderse que una o más implementaciones no están limitadas a las realizaciones descritas. Por el contrario, se pretende cubrir diversas modificaciones y disposiciones similares como las que resultarían evidentes para los expertos en la materia.

40

REIVINDICACIONES

1. Un método para renderizar audio para su reproducción a través de unos auriculares (116) utilizados por un oyente, que comprende:

recibir contenido de audio digital que incluye audio basado en canales y en objetos;

5 recibir metadatos dependientes del contenido generados por una herramienta de autoría (102) que procesa el contenido de audio digital recibido, en el que los metadatos dependientes del contenido controlan una pluralidad de características de canal y objeto e incluyen posiciones de objetos de audio en el audio basado en objetos;

caracterizado por que el método comprende además

10 recibir datos de sensor (316; 336) proporcionados por una pluralidad de sensores instalados en los auriculares (116), incluyendo dichos datos de sensor parámetros angulares y posicionales (r , θ , ϕ) asociados con la cabeza del oyente; y

renderizar el contenido de audio digital en respuesta a una combinación de los metadatos dependientes del contenido y los datos de sensor;

15 en donde los metadatos dependientes del contenido comprenden además una bandera que indica que se aplique seguimiento de la cabeza a un objeto, y

en donde el paso de renderizar el contenido de audio digital incluye, para dicho objeto al cual se indica que se ha de aplicar seguimiento de la cabeza, la modificación de una posición de dicho objeto en respuesta a los parámetros angulares y posicionales recibidos (r , θ , ϕ).

20 2. El método de la reivindicación 1, que además comprende separar el contenido de audio digital en uno o más componentes basados en el tipo de contenido, y en el que el tipo de contenido se selecciona del grupo que consiste en: diálogo, música, efectos de audio, señales transitorias y señales ambientales.

3. El método de la reivindicación 1, que comprende además desmezclar al menos uno del audio basado en canales y en objetos en contenido directo constituyente y en contenido difuso, en el que el contenido difuso comprende elementos de sonido reverberados o reflejados.

25 4. El método de la reivindicación 3, que comprende además realizar la renderización binaural sobre el contenido directo por separado de la renderización binaural sobre el contenido difuso.

5. Un sistema que comprende:

30 una herramienta de autoría (102) que genera metadatos dependientes del contenido para renderizar contenido de audio digital a través de unos auriculares (116) utilizados por un oyente, en el que el contenido de audio digital comprende audio basado en canales y objetos, y en el que los metadatos dependientes del contenido controlan una pluralidad de características de canal y objeto e incluyen posiciones de objetos de audio en el audio basado en objetos;

caracterizado por

35 una interfaz de realimentación que incluye una pluralidad de sensores instalados en los auriculares (116) y dispuestos para proporcionar datos de sensor (316; 336) que incluyen parámetros angulares y posicionales (r , θ , ϕ) asociados con la cabeza del oyente; y

40 un componente de renderización (112) acoplado funcionalmente para recibir el contenido de audio digital y los metadatos dependientes del contenido de la herramienta de autoría, y para recibir los datos de sensor de los sensores, y configurado para realizar la renderización del contenido de audio digital en respuesta a una combinación de los metadatos dependientes del contenido y los datos de sensor;

en donde los metadatos dependientes del contenido comprenden además una bandera que indica que se ha de aplicar seguimiento de cabeza a un objeto, y

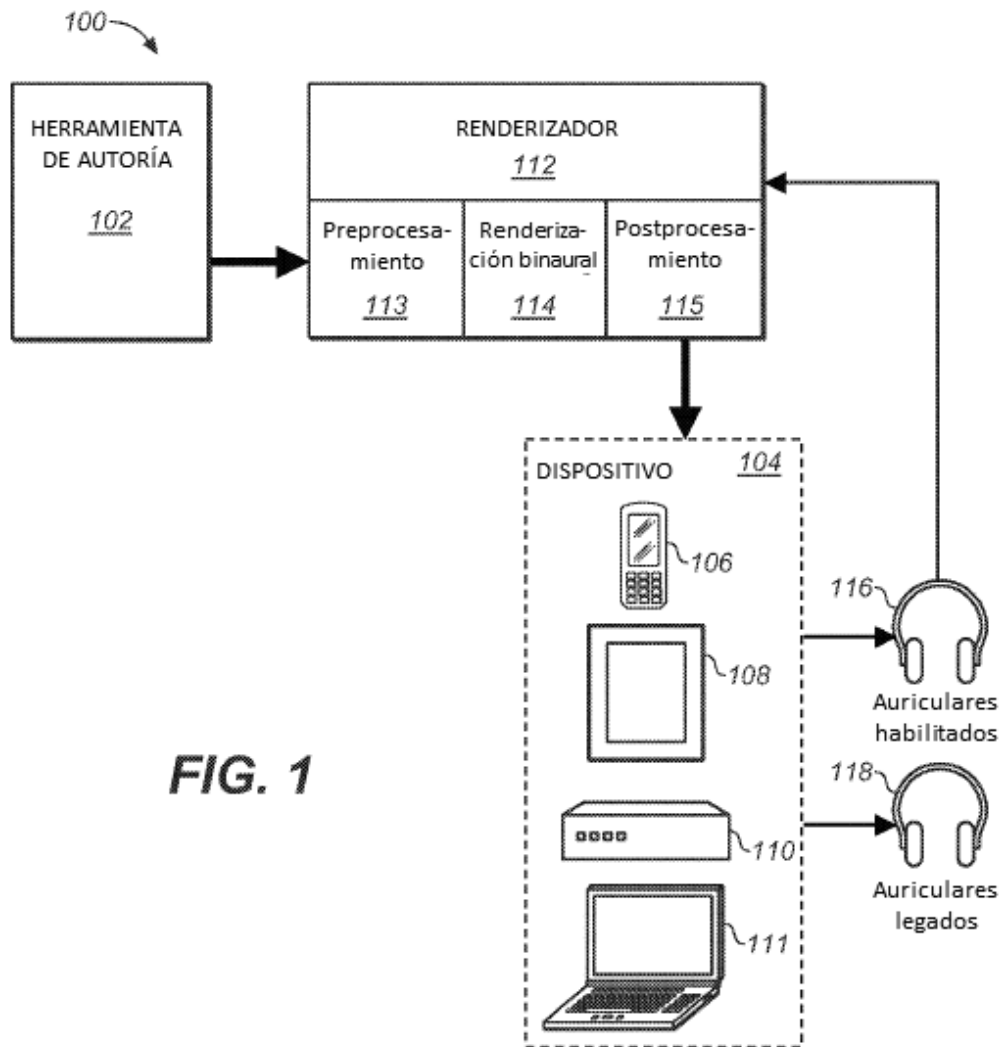
45 en donde el componente de renderización, para dicho objeto para el cual se indica que se ha de aplicar seguimiento de cabeza, está configurado adicionalmente para modificar una posición de dicho objeto en respuesta a los parámetros angulares y posicionales recibidos (r , θ , ϕ).

6. El sistema de la reivindicación 5 en el que el contenido de audio digital se transmite para su reproducción a través de los auriculares mediante un dispositivo de fuente de audio portátil.

50 7. El sistema de cualquiera de las reivindicaciones 5 o 6 que comprende además un componente de formateo que formatea los datos de sensor en un formato de metadatos compatible con los metadatos dependientes del contenido para producir metadatos de reproducción.

8. El sistema de la reivindicación 7, en el que los auriculares están acoplados con el componente de formateado a través de uno de un enlace inalámbrico y un enlace por cable.

5 9. El sistema de cualquiera de las reivindicaciones 5-8, en el que el componente de renderización comprende una etapa de postprocesamiento configurada para realizar funciones que comprenden al menos una de: mezcla descendente a partir de una pluralidad de canales de sonido envolvente hasta uno de una mezcla binaural, o una mezcla estéreo, gestión de nivel, ecualización, corrección de timbre y cancelación de ruido.



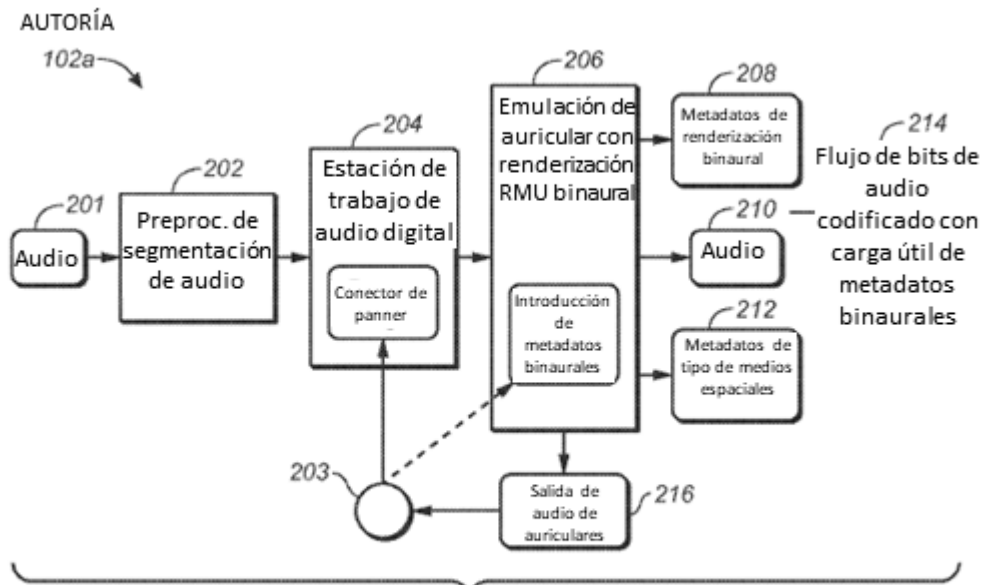


FIG. 2A

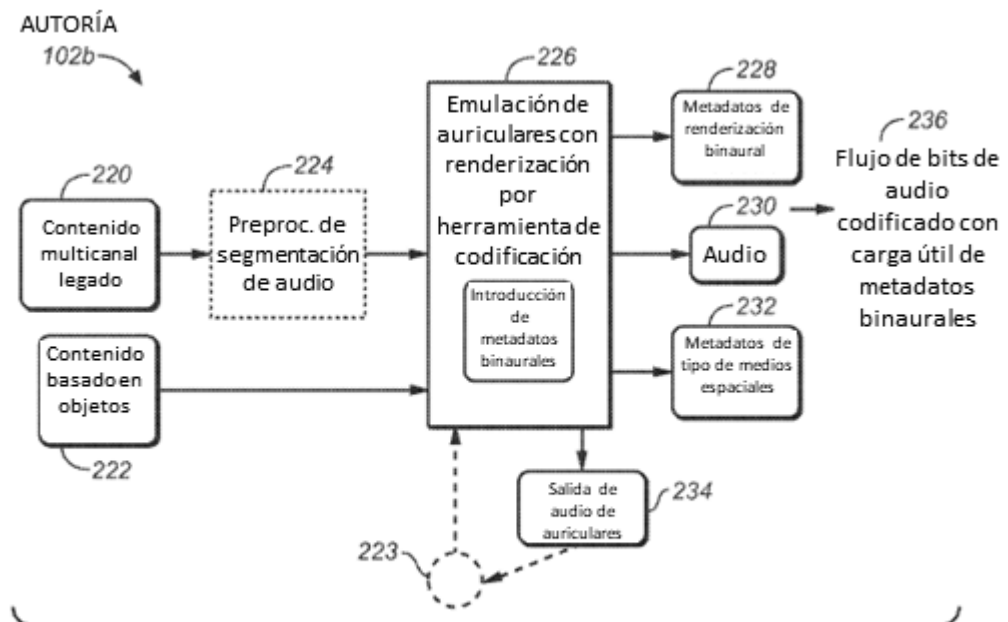


FIG. 2B

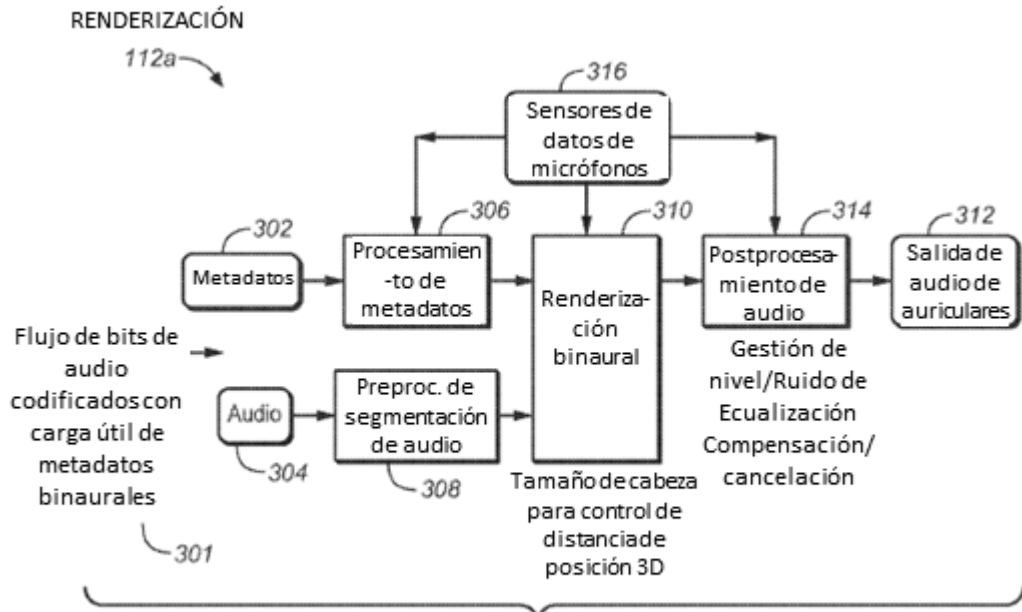


FIG. 3A

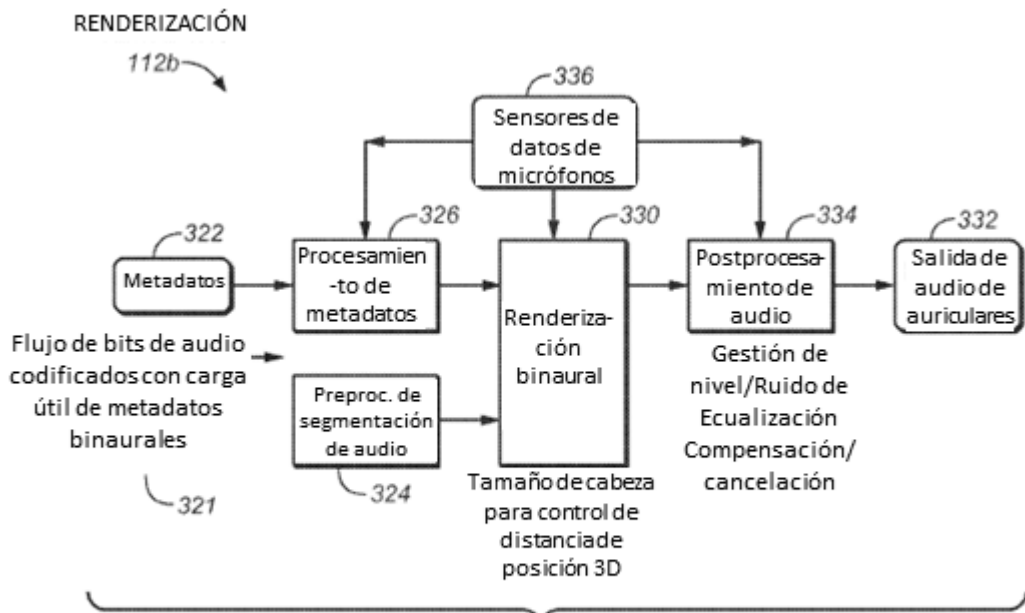


FIG. 3B

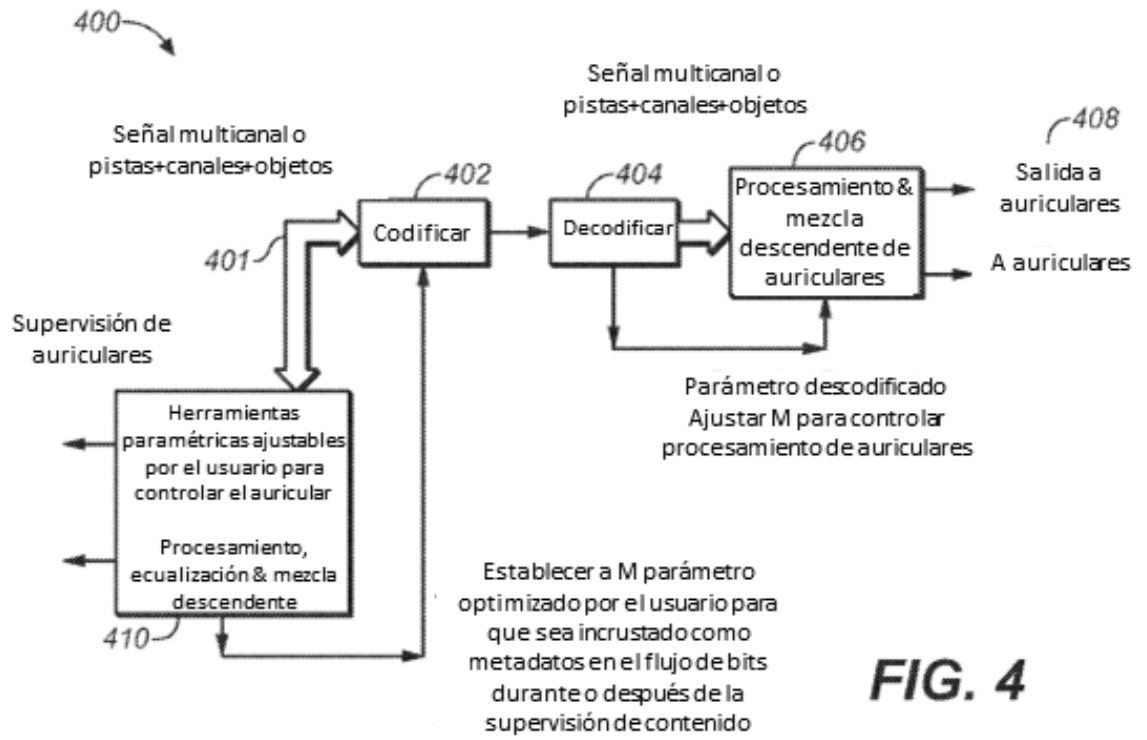


FIG. 4

500

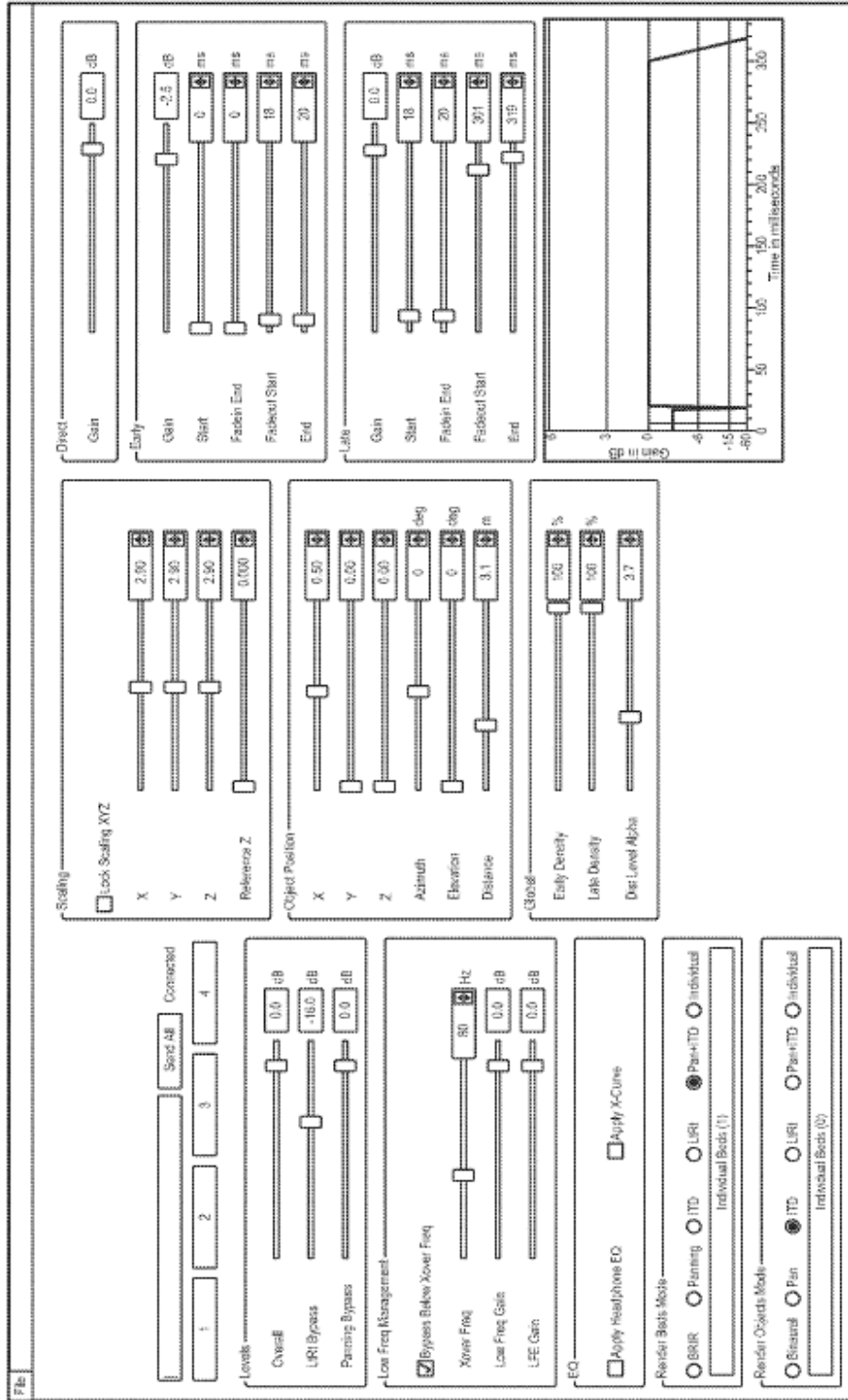


FIG. 5

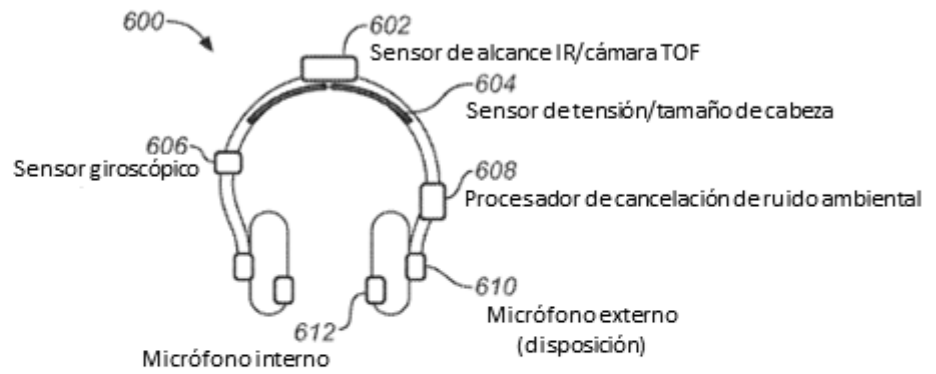


FIG. 6

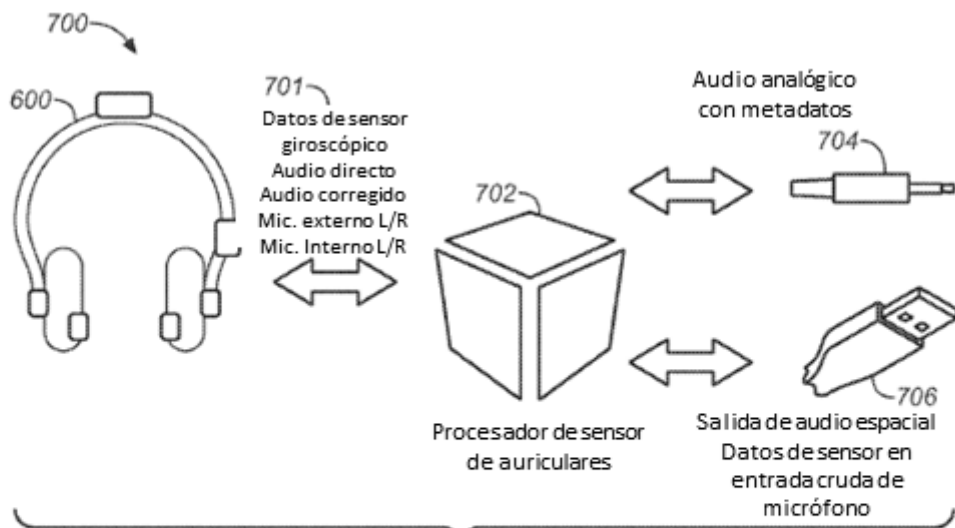


FIG. 7

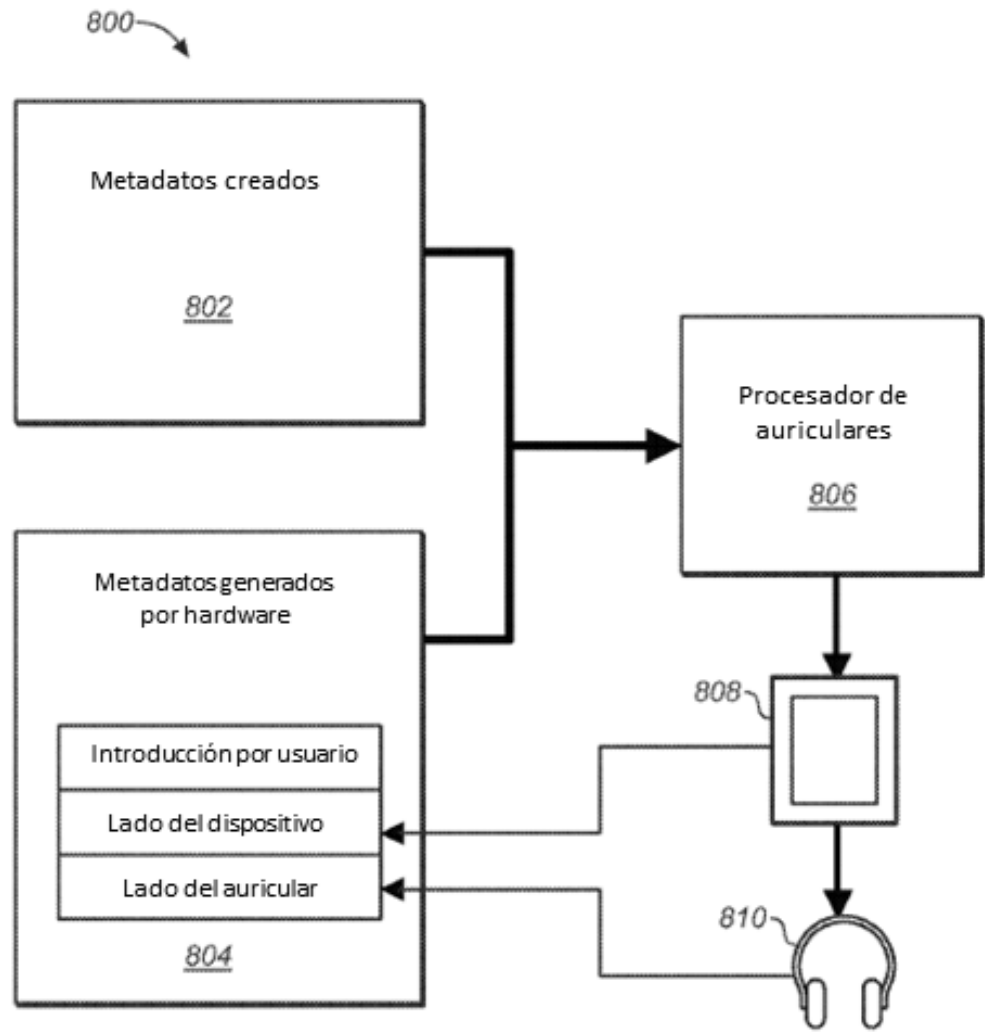


FIG. 8

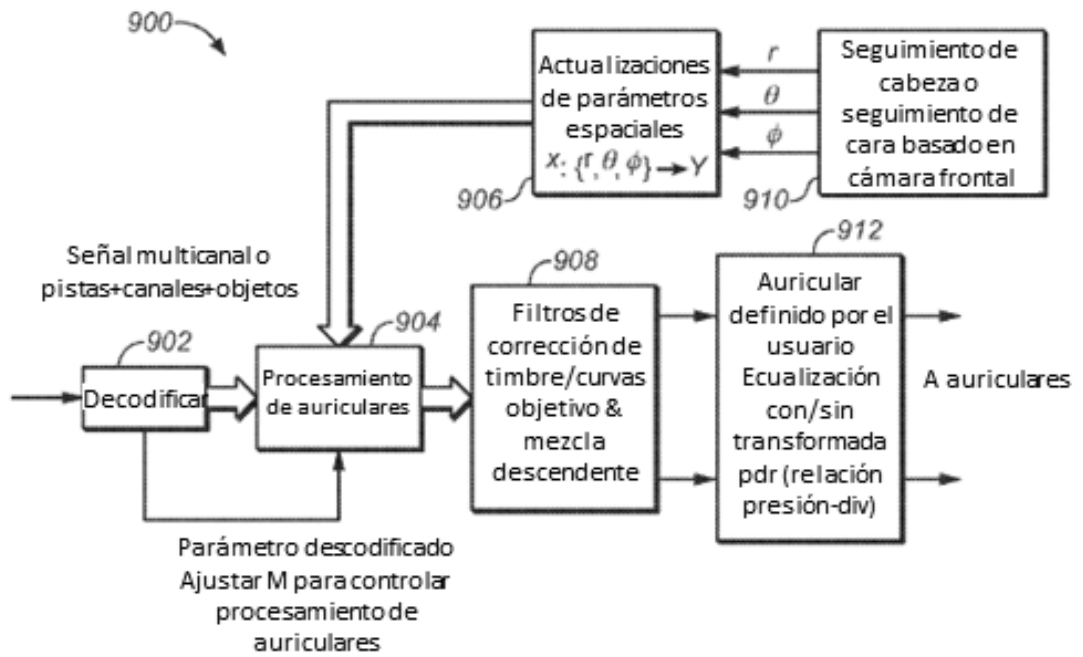


FIG. 9

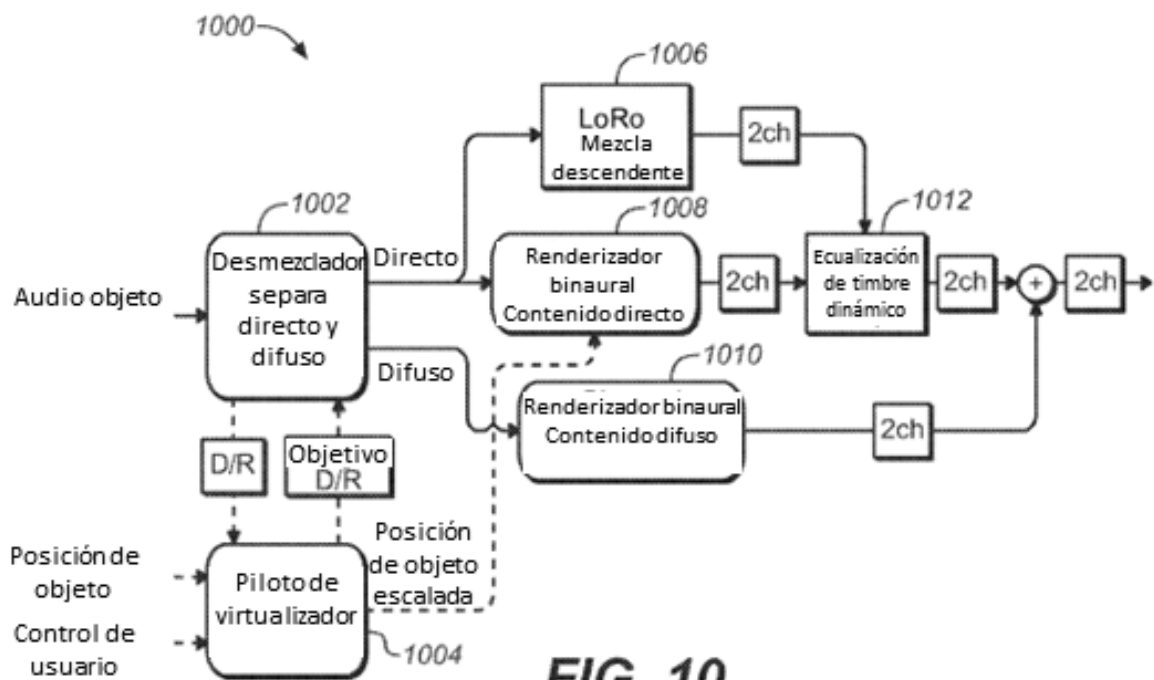


FIG. 10

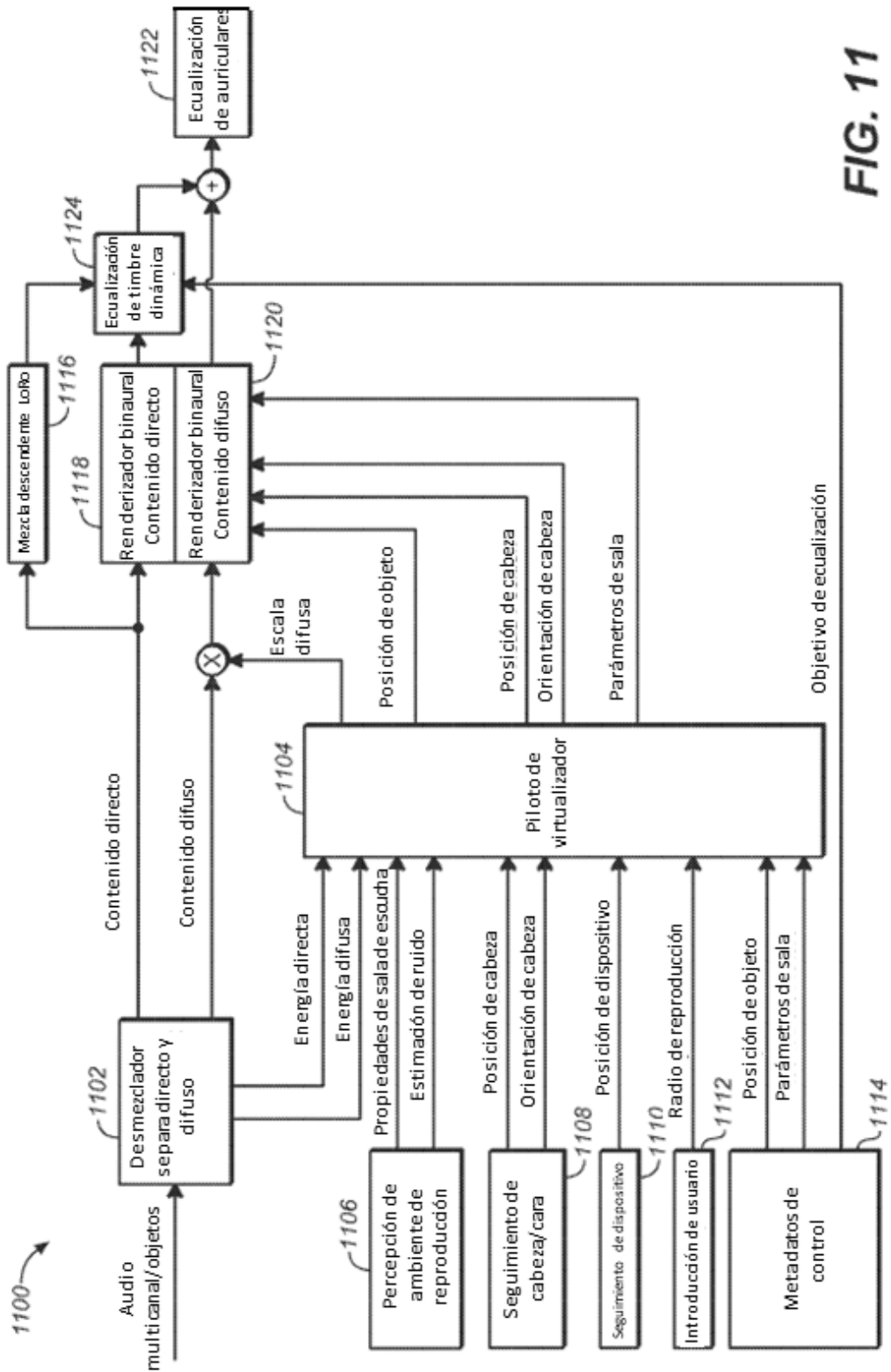


FIG. 11

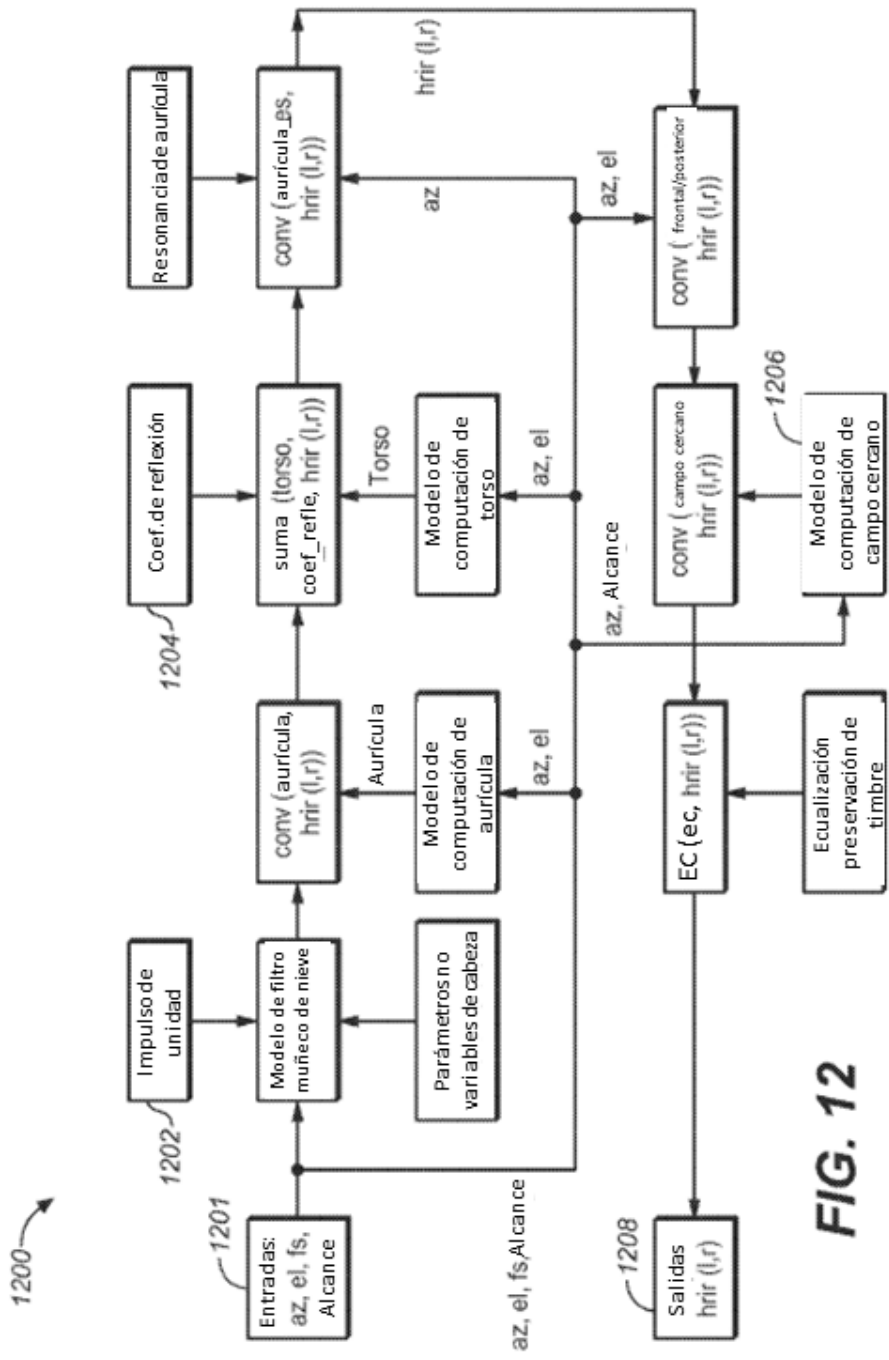


FIG. 12

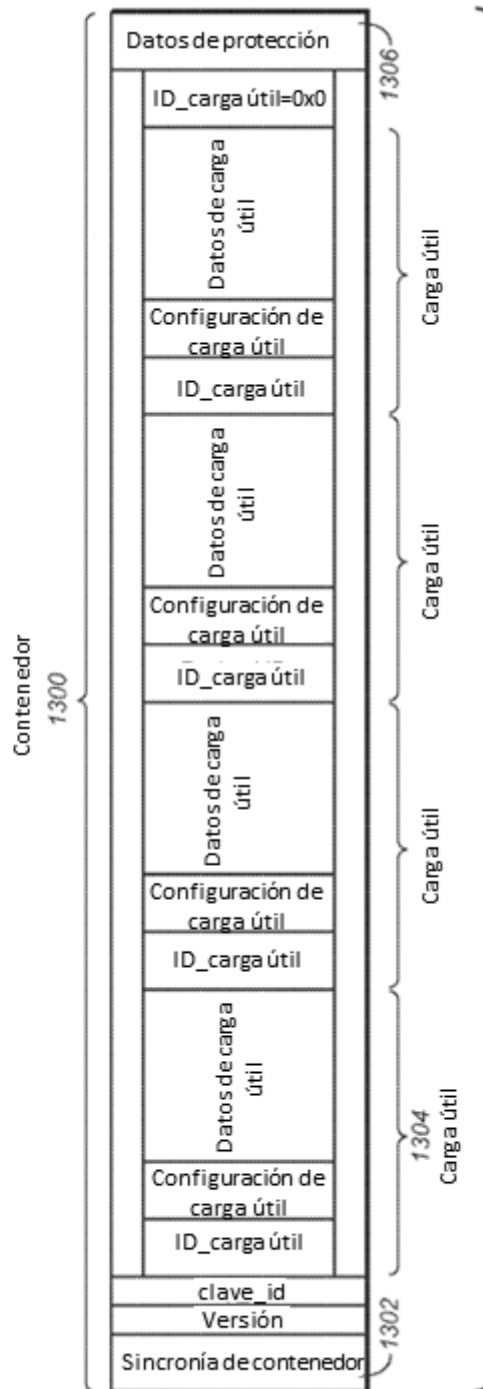


FIG. 13

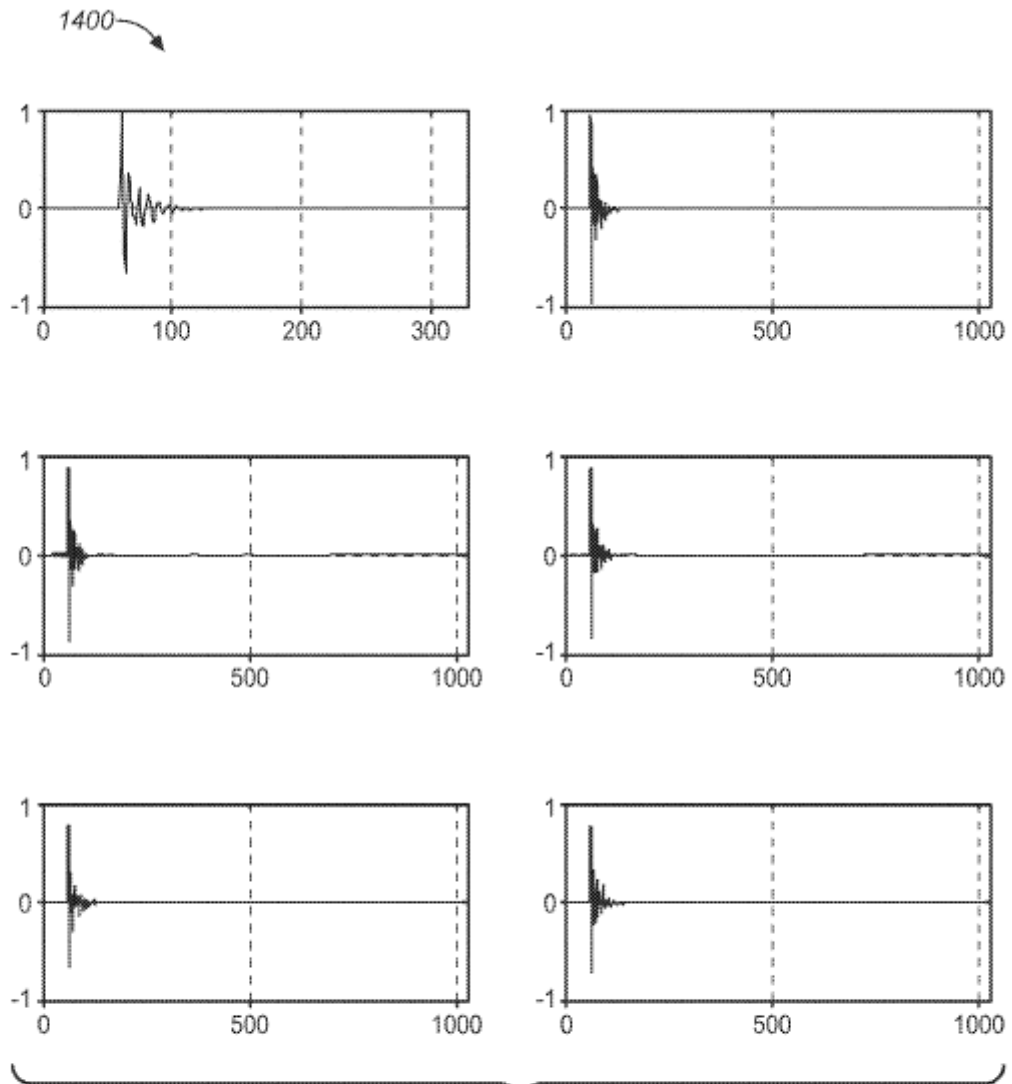


FIG. 14

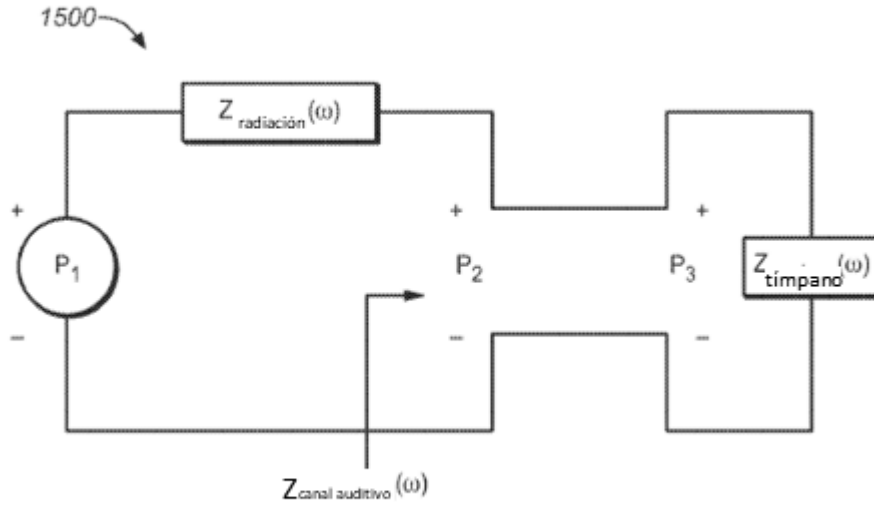


FIG. 15A

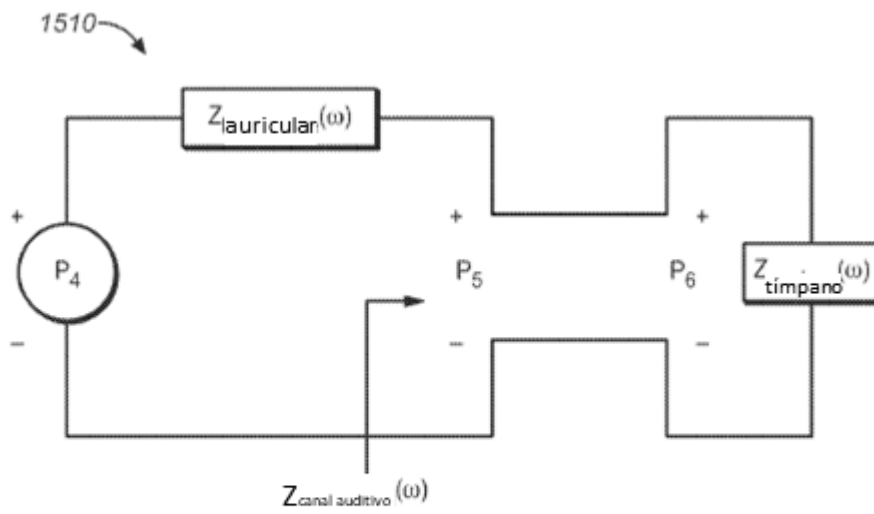


FIG. 15B