

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 606 678**

51 Int. Cl.:

**H04R 5/02** (2006.01)

**H04S 7/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.08.2013 PCT/US2013/056989**

87 Fecha y número de publicación internacional: **06.03.2014 WO14036085**

96 Fecha de presentación y número de la solicitud europea: **28.08.2013 E 13759397 (6)**

97 Fecha y número de publicación de la concesión europea: **05.10.2016 EP 2891337**

54 Título: **Presentación de sonido reflejado para audio con base de objeto**

30 Prioridad:

**31.08.2012 US 201261695893 P**

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

**27.03.2017**

73 Titular/es:

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

72 Inventor/es:

**CROCKETT, BRETT, G.;  
HOOKS, SPENCER;  
SEEFELDT, ALAN;  
LANDO, JOSHUA, B.;  
BROWN, C., PHILLIP;  
MEHTA, SRIPAL, S. y  
MURRIE, STEWART**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

ES 2 606 678 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

Presentación de sonido reflejado para audio con base de objeto.

## 5 CAMPO DE LA INVENCION

Una o más realizaciones se refieren, en general, al procesamiento de señales sonoras y más en particular, a la presentación de un contenido de audio adaptativo a través de controladores directos y reflejados en algunos entornos de escucha.

10

## ANTECEDENTES DE LA INVENCION

El contenido en la sección de antecedentes de la invención no debe suponerse que es la técnica anterior sino simplemente un resultado de su mención en la sección de antecedentes. De modo similar, un problema mencionado en la sección de antecedentes de la invención o asociado con el contenido de dicha sección no debe suponerse que ha sido anteriormente reconocido en la técnica anterior.

15

Las pistas de sonido cinematográfico suelen comprender numerosos elementos sonoros diferentes que corresponden a imágenes en la pantalla, ventanas de diálogo, ruidos y efectos sonoros que emanan de diferentes lugares en la pantalla y se combinan con la música de fondo y los efectos ambientales para crear la experiencia de audiencia global. Una reproducción exacta requiere que los sonidos se reproduzcan en una manera que corresponda lo más estrechamente posible a lo que aparece en la pantalla con respecto a la posición de la fuente del sonido, intensidad, movimiento y profundidad. Los sistemas de audio basados en canales tradicionales envían el contenido de audio en la forma de señales de alimentación de altavoces para altavoces individuales en un entorno de reproducción. La introducción del sistema cinematográfico digital ha creado nuevas normas para el sonido cinematográfico, tal como la incorporación de múltiples canales de audio para permitir una mayor creatividad para creadores de contenidos y una experiencia de audición más envolvente y realista para las audiencias. La expansión más allá de las alimentaciones de altavoces tradicionales y de audio basado en canales como un medio para distribuir la señal de audio espacial es crítica y existe un interés considerable en una descripción de audio basada en un modelo que permita a quien escucha seleccionar una configuración de reproducción deseada con la señal de audio presentada específicamente para su configuración elegida. Para mejorar todavía más la experiencia del oyente, la reproducción del sonido en entornos verdaderamente tridimensionales (3D) o 3D virtuales se ha convertido en un área de investigación y desarrollo cada vez mayor. La presentación espacial del sonido utiliza objetos de audio, que son señales de audio con descripciones de fuentes paramétricas asociadas de la posición de la fuente aparente (p.ej., coordenadas de 3D), anchura de fuente aparente y otros parámetros. La señal de audio basada en el objeto puede utilizarse para numerosas aplicaciones multimedia, tales como películas cinematográficas digitales, juegos de vídeo, simuladores y siendo de particular importancia en un entorno residencial en donde el número de altavoces y su emplazamiento suelen estar limitado o restringido por los confines de un entorno de escucha relativamente pequeño.

20

25

30

35

40

Varias tecnologías se han desarrollado para mejorar los sistemas acústicos en entornos cinematográficos y para capturar y reproducir, con mayor exactitud, la intención artística del creador para una pista sonora de imágenes en movimiento. A modo de ejemplo, se ha desarrollado un formato de audio espacial de la siguiente generación (también referido como "audio adaptativo"), que comprende una mezcla de objetos de audio y altavoces basados en canales tradicionales que se alimentan junto con metadatos posicionales para los objetos de audio. En un decodificador de audio espacial, los canales se envían directamente a sus altavoces asociados (si existen altavoces adecuados) o son un objeto de una mezcla descendente hacia un conjunto de altavoces existente y los objetos de audio son presentados por el decodificador en una manera flexible. La descripción de fuente paramétrica asociada con cada objeto, tal como una trayectoria posicional en un espacio en 3D se toma como una entrada junto con el número y la posición de altavoces conectados al decodificador. El dispositivo de presentación utiliza entonces algunos algoritmos, tal como una ley para obtener panorámicas, para distribuir el audio asociado con cada objeto a través del conjunto adjunto de altavoces. De este modo, la intención espacial de autor de cada objeto se presenta de forma óptima a través de la configuración de altavoz específica que está presente en el entorno de escucha.

45

50

55

60

65

Los sistemas de audio espacial actuales han sido generalmente desarrollados para uso cinematográfico y por ello, implican su despliegue en grandes recintos y el uso de equipos de coste relativamente alto, incluyendo redes de múltiples altavoces distribuidos alrededor del entorno de escucha. Una cantidad creciente de contenidos cinematográfico que actualmente se produce se está haciendo disponible para su reproducción en el entorno residencial por intermedio de tecnología de flujos continuos y tecnología de soportes avanzados, tal como un así denominado *Blu-ray*, etc. Además, tecnologías emergentes tales como televisión en 3D y juegos informáticos avanzados y simulares están estimulando el uso de equipos relativamente sofisticados, tales como monitores de pantallas de grandes dimensiones, reflectores de sonido envolvente y redes de altavoces en el entorno residencial y otros entornos de escucha (no de cine/teatro). Sin embargo, el coste del equipo, la complejidad de la información y las dimensiones de la sala son limitaciones realistas que impiden la plena explotación de las señales de audio espaciales en la mayoría de los entornos residenciales. A modo de ejemplo, sistemas de audio basados en objetos avanzados suelen emplear altavoces de altura o aéreos para reproducir el sonido que está provisto que se origine

por encima de la cabeza de la persona que los escucha. En numerosos casos, y en particular, en el entorno residencial, dichos altavoces de altura pueden no estar disponibles. En este caso, la información de altura se pierde si dichos objetos de sonido se reproducen solamente a través de altavoces montados en la pared o en el suelo.

5 Por lo tanto lo que se necesita es un sistema que permita que una información espacial completa de un sistema de audio adaptativo se reproduzca en un entorno de escucha que puede incluir solamente una parte de la red de altavoces completa prevista para la reproducción, tal como altavoces limitados o sin sobrecarga, y que se pueda utilizar altavoces reflejados para emanar sonido desde lugares en donde no pueden existir altavoces directos.

10 El denominado Informe de Investigación Internacional emitido en relación con el presente documento citaba *inter alia* los siguientes documentos:

15 La publicación de solicitud de patente de Estados Unidos número US 2007/0263890 A1 da a conocer un receptor de sonido envolvente (AVR) de audio-vídeo reconfigurable. Un circuito de procesamiento dentro del receptor AVR genera señales para altavoces de canal principal y envolventes que proporcionan una operación seleccionable entre el emplazamiento del altavoz en instalación sonora envolvente ordinaria o en una instalación sonora envolvente simulada con altavoces situados en un extremo de una sala de escucha.

20 La publicación de solicitud de patente europea nº EP 1 971 187 A2 da a conocer un aparato que incluye una red de altavoces en donde las unidades de altavoces plurales están dispuestas en un cuerpo único, una unidad de adición de localización de fuente sonora que genera un señal de audio izquierda y derecha realizando un procesamiento de localización para añadir características del sonido a señales de audio de un canal frontal-izquierdo y un canal frontal-derecho sobre la base de funciones de transferencia de cabezas y una unidad de control de dirección de emisión sonora que distribuye las señales de audio izquierda y derecha a una o varias unidades de altavoces de la red de altavoces.

25 La publicación de solicitud de patente europea nº EP 1 416 769 A1 da a conocer una unidad de edición/producción de audio que separa las fuentes sonoras aplicadas por intermedio de la unidad de entrada de audio en sonidos objeto y sonidos de fondo en función de la selección del oyente y los convierte en información de escena de audio tridimensional (3-D). Una unidad de codificación de audio codifica información en 3-D y señales objeto de información de escena de audio de 3-D, con el fin de transmitir por intermedio del soporte.

30 La publicación de solicitud de patente de Estados Unidos nº US 2006/0109992 A1 da a conocer, para una corrección de nivel en un sistema de síntesis de campo de ondas que tiene un módulo de síntesis de campo de ondas y una red de altavoces para proporcionar sonido a una zona de presentación, un valor de corrección que está basado en un estado de amplitud establecido en una zona de presentación es objeto de determinación. El estado de amplitud establecido depende de una posición de la fuente virtual o un tipo de la fuente virtual, y el estado de amplitud real en la zona de presentación depende de las señales componentes para los altavoces debido a la fuente virtual. El valor de corrección se alimenta a un manipulador que manipula la señal de audio asociada a la fuente virtual antes de la alimentación al módulo de síntesis de campo de ondas o las señales componentes para los altavoces individuales debido a la fuente virtual se manipulan para reducir una desviación entre un estado de amplitud establecido y un estado de amplitud real en un punto o varios puntos en la zona de presentación.

#### 45 SUMARIO DE LAS FORMAS DE REALIZACIÓN

En conformidad con la presente invención, el objetivo anterior se consigue mediante un sistema para presentar sonido utilizando elementos sonoros reflejados según se define en la reivindicación independiente 1 y mediante un método correspondiente según se define en la reivindicación independiente 11.

50 Además, se describen sistemas y métodos para un formato de audio y un sistema que incluye herramientas de creación de contenidos actualizadas, métodos de distribución y una mejor experiencia del oyente basada en un sistema de audio adaptativo que incluye nuevas configuraciones de altavoces y canales, así como un nuevo formato de descripción espacial hecho posible por un conjunto de herramientas de creación de contenidos avanzadas creadas para mezcladores de sonido cinematográficos. Las formas de realización incluyen un sistema que expande el concepto de audio adaptativo basado en cinematografía para un ecosistema de reproducción de audio particular que incluye un entorno residencial (p.ej., receptor A/V, barra de sonido y reproductor blu-ray), soportes electrónicos (p.ej., ordenador personal PC, tableta electrónica, dispositivo móvil y reproducción de auriculares), difusión (p.ej., TV y decodificador), música, juegos, sonido en vivo, contenido generado por el oyente ("UGC"),etc. El sistema de entorno residencial incluye componentes que proporcionan compatibilidad con el contenido de la sala de audición e incluye definiciones de metadatos que comprende información de creación de contenidos para transmitir la idea creativa, información de inteligencia de soporte con respecto a objetos de audio, alimentaciones de altavoces, información de presentación espacial y metadatos dependientes de los contenidos que indican el tipo de contenido tal como diálogo, música, ambiente, etc. Las definiciones de audio adaptativo pueden incluir un altavoz estándar que se alimenta por intermedio de canales de audio más objetos de audio con información de presentación espacial asociada (tal como tamaño, velocidad y emplazamiento en un espacio tridimensional). Una nueva disposición de altavoces (o configuración de canales) y un nuevo formato de descripción espacial adjunto que soportará múltiples

5 tecnologías de presentación se describen también a este respecto. Flujos de audio (que incluyen, en general, canales y objetos) se transmiten junto con los metadatos que describen la intención del creador de contenidos o del mezclador de sonidos, incluyendo la posición deseada del flujo de audio. La posición puede expresarse como un canal nombrado (desde dentro de la configuración de canal predefinida) o una información de posición espacial en 3D. Este canal junto con el formato del objeto proporciona los mejores métodos de descripción de escenas de audio basadas en canales y basadas en modelos.

10 Las formas de realización están concretamente dirigidas a un sistema para presentar sonido utilizando elementos sonoros reflejados, que comprende una red de controladores de audio para distribución alrededor de un entorno de escucha, en donde algunos de los excitaciones son excitaciones directos y otros son excitaciones reflejados que están configurados para proyectar ondas sonoras hacia una o más superficies del entorno de escucha para reflexión a una zona de escucha específica; un dispositivo de presentación para procesar flujos de audio y uno o más conjuntos de metadatos que están asociados con cada flujo de audio y que especifican un emplazamiento de reproducción en el entorno de escucha de un flujo de audio respectivo, en donde los flujos de audio comprenden uno o más flujos de audio reflejados y uno o más flujos de audio directos; y un sistema de reproducción para presentar los flujos de audio a la red de controladores de audio en conformidad con los uno o más conjuntos de metadatos, y en donde los uno o más flujos de audio reflejados se transmiten a los controladores de audio reflejados.

20 BREVE DESCRIPCIÓN DE LOS DIBUJOS

En los dibujos siguientes se utilizan referencias numéricas similares para referirse a elementos similares. Aunque las siguientes Figuras ilustran varios ejemplos, las una o más puestas en práctica no están limitadas a los ejemplos ilustrados en las Figuras.

25 La Figura 1 ilustra un emplazamiento de altavoz ejemplo en un sistema envolvente (p.ej., 9.1 envolvente) que proporciona altavoces de altura para reproducción de canales de altura.

30 La Figura 2 ilustra la combinación de datos basados en objetos y en canales para obtener una mezcla de audio adaptativo, en conformidad con una forma de realización.

La Figura 3 es un diagrama de bloques de una arquitectura de reproducción para uso en un sistema de audio adaptativo, en conformidad con una forma de realización.

35 La Figura 4A es un diagrama de bloques que ilustra los componentes funcionales para adaptar el contenido de audio con base cinematográfica para uso en un entorno de escucha en conformidad con un forma de realización.

La Figura 4B es un diagrama de bloques detallado de los componentes representados en la Figura 3A, en conformidad con una forma de realización.

40 La Figura 4C es un diagrama de bloques de los componentes funcionales de un entorno de audio adaptativo, en conformidad con una forma de realización.

45 La Figura 5 ilustra el despliegue de un sistema de audio adaptativo en entorno de auditorio residencial a modo de ejemplo.

La Figura 6 ilustra el uso de un controlador de proyección hacia arriba que utiliza el sonido reflejado para simular un altavoz aéreo en un entorno de escucha.

50 La Figura 7A ilustra un altavoz que tiene una pluralidad de excitaciones en una primera configuración para uso en un sistema de audio adaptativo que tiene un dispositivo de presentación de sonido reflejado, en conformidad con una forma de realización.

55 La Figura 7B ilustra un sistema de altavoces que tiene excitaciones distribuidos en múltiples recintos para uso en un sistema de audio adaptativo que tiene un dispositivo de presentación de sonido reflejado, en conformidad con una forma de realización.

La Figura 7C ilustra una configuración ejemplo para una barra de sonido utiliza en un sistema de audio adaptativo que usa un dispositivo de presentación de sonido reflejado, en conformidad con una forma de realización.

60 La Figura 8 ilustra un emplazamiento, a modo de ejemplo, de altavoces que tienen excitaciones direccionables individualmente que incluyen excitaciones de proyección hacia arriba situados dentro de un entorno de escucha.

65 La Figura 9A ilustra una configuración de altavoces para un sistema de audio adaptativo 5.1 que utiliza múltiples excitaciones direccionables para audio reflejado, en conformidad con una forma de realización.

La Figura 9B ilustra una configuración de altavoces para un sistema de audio adaptativo 7.1 que utiliza múltiples

excitaciones direccionables para audio reflejado, en conformidad con una forma de realización.

La Figura 10 es un diagrama que ilustra la composición de una interconexión bidireccional, en conformidad con una forma de realización.

5 La Figura 11 ilustra una configuración automática y un proceso de calibración del sistema para uso en un sistema de audio adaptativo, en conformidad con una forma de realización.

10 La Figura 12 es un diagrama de flujo que ilustra las etapas de proceso para un método de calibración utilizado en un sistema de audio adaptativo, en conformidad con una forma de realización.

La Figura 13 ilustra el uso de un sistema de audio adaptativo en un caso de uso, a modo de ejemplo, de televisión y barra de sonido.

15 La Figura 14 ilustra una representación simplificada de una virtualización de auriculares binaurales tridimensionales en un sistema de audio adaptativo, en conformidad con una forma de realización.

20 La Figura 15 es una tabla que ilustra algunas definiciones de metadatos para uso en un sistema de audio adaptativo que utiliza un dispositivo de presentación de sonido reflejado para entornos de escucha, en conformidad con una forma de realización.

La Figura 16 es un gráfico que ilustra la respuesta de frecuencia para un filtro combinado, en conformidad con una forma de realización.

## 25 DESCRIPCIÓN DETALLADA DE LA INVENCION

Se describen sistemas y métodos para un sistema de audio adaptativo que presenta sonido reflejado para sistemas de audio adaptativos que carecen de altavoces aéreos. Los aspectos de las una o más formas de realización aquí descritas pueden ponerse en práctica en un sistema audiovisual o de audio que procesa información de audio origen en un sistema de mezcla, presentación y reproducción que incluye uno o más ordenadores o dispositivos de procesamiento que ejecutan instrucciones de software. Cualquiera de las formas de realización descritas puede utilizarse sola o junto con otra en cualquier combinación. Aunque varias formas de realización pueden haber sido motivadas por diversas deficiencias de la técnica anterior, que pueden describirse o referirse en uno o más lugares en esta especificación, las formas de realización no eliminan necesariamente cualquiera de estas deficiencias. Dicho de otro modo, diferentes formas de realización pueden resolver diferentes deficiencias que pueden examinarse en la especificación. Algunas formas de realización solo pueden resolver parcialmente algunas deficiencias o solamente una deficiencia que puede describirse en la especificación, y algunas formas de realización no pueden resolver ninguna de estas deficiencias.

40 Para los fines de la presente descripción, los términos siguientes tienen los significados asociados: el término "canal" significa una señal de audio más metadatos en donde la posición está codificada como un identificador de canal, p.ej., envolvente de parte frontal izquierda o parte superior derecha; "audio basado en canal" es un audio formateado para la reproducción por intermedio de un conjunto predefinido de zonas de altavoces con localizaciones nominales asociadas, p.ej., 5.1, 7.1, y así sucesivamente; el término "objeto" u "objeto basado en objeto" significa uno o más canales de audio con una descripción de origen paramétrico, tal como una posición de fuente aparente (p.ej., coordenadas de 3D), anchura de fuente aparente, etc.; y el término "audio adaptativo" significa señales de audio basadas en canal y/o basadas en objeto más metadatos que presentan las señales de audio sobre la base del entorno de reproducción utilizando un flujo de audio más metadatos en los que la posición se codifica como una posición 3D en el espacio; y el término "entorno de escucha" significa cualquier zona abierta, parcialmente cerrada o completamente cerrada, tal como una sala que puede utilizarse para la reproducción de contenido de audio solo o con vídeo u otro contenido, y puede modificarse en un entorno residencial, sala cinematográfica, teatro, auditorio, estudio, consola de juegos y soportes similares. Dicha zona puede tener una o más superficies dispuestas en ella, tales como paredes o deflectores que pueden reflejar, de forma directa o difusa, ondas sonoras.

## 55 Formato y sistema de audio adaptativo

Las formas de realización están destinadas a un sistema de presentación de sonido reflejado que está configurado para funcionar con un formato de sonido y un sistema de procesamiento que puede referirse como un "sistema de audio espacial" o "sistema de audio adaptativo" que está basado en un formato de audio y tecnología de presentación para permitir la mejora de la inmersión de la audiencia, mayor control artístico y mejor flexibilidad y escalabilidad del sistema. Un sistema de audio adaptativo global suele comprender un sistema de codificación, distribución y decodificación de audio configurado para generar uno o más flujos de bits que contienen ambos elementos de audio basados en canales convencionales y elementos de codificación de objetos de audio. Dicho método combinado proporciona mayor eficiencia de codificación y flexibilidad de presentación en comparación con los métodos basados en el objeto o basados en el canal que se consideran por separado. Un ejemplo de un sistema de audio adaptativo que puede utilizarse en conjunción con las presentes formas de realización se describe en la

solicitud de patente provisional de Estados Unidos pendiente 61/636,429, presentada con fecha 20 de abril de 2012 y titulada "Sistema y método para la generación, codificación y presentación de señales de audio adaptativo".

Una puesta en práctica ejemplo de un sistema de audio adaptativo y formato de audio asociado es la plataforma de Dolby® Atmos™. Dicho sistema incorpora una dimensión de altura (arriba/abajo) que puede ponerse en práctica como un sistema envolvente 9.1 o una configuración sonora envolvente similar. La Figura 1 ilustra el emplazamiento de los altavoces en un presente sistema envolvente (p.ej., envolvente 9.1) que proporciona altavoces de altura para la reproducción de canales de altura. La configuración de altavoces del sistema 9.1 100 está constituido por cinco altavoces 102 en el plano del suelo y cuatro altavoces 104 en el plano de altura. En general, estos altavoces pueden utilizarse para producir sonido que esté diseñado para emanar desde cualquier posición, con mayor o menor exactitud dentro del entorno de escucha. Configuraciones de altavoces predefinidas, tales como las ilustradas en la Figura 1, pueden limitar naturalmente la capacidad para representar, con exactitud, la posición de una fuente sonora dada. A modo de ejemplo, una fuente sonora no puede ser objeto de panorámicas más a la izquierda que el propio altavoz izquierdo. Esto se aplica a cada altavoz, constituyendo, por lo tanto, una forma geométrica unidimensional (p.ej., izquierda-derecha), bidimensional (p.ej., frontal-posterior) o tridimensional (p.ej., izquierda-derecha, frontal-posterior, arriba-abajo), en donde está limitada la mezcla. Varias diferentes configuraciones y tipos de altavoces pueden utilizarse para dicha configuración de altavoces. A modo de ejemplo, algunos sistemas de audio mejorados pueden utilizarse altavoces en una configuración 9.1, 11.1, 13.1, 19.4, u otra configuración. Los tipos de altavoces pueden incluir una gama completa de altavoces directos, redes de altavoces, altavoces envolventes, altavoces de graves, altavoces de agudos y otros tipos de altavoces.

Los objetos de audio pueden considerarse como grupos de elementos sonoros que pueden percibirse para emanar desde un lugar físico particular o ubicaciones en el entorno de escucha. Dichos objetos pueden ser estáticos (es decir, estacionarios) o dinámicos (es decir, móviles). Los objetos de audio están controlados por metadatos que definen la posición del sonido en un punto dado en el tiempo, junto con otras funciones. Cuando se reproducen objetos, se presentan en función de los metadatos posicionales que utilizan los altavoces que están presentes, y no necesariamente siendo salida a un canal físico predefinido. Una pista en una sesión puede ser un objeto de audio, y datos de obtener panorámicas estándar que sean análogas a los metadatos posicionales. De este modo, el contenido colocado en la pantalla podría obtener panorámicas efectivamente de la misma manera que con un contenido basado en canal, pero el contenido colocado en las envolventes puede presentarse a un altavoz individual si así se desea. Aunque el uso de objetos de audio proporciona el control deseado para efectos discretos, otros aspectos de una pista sonora pueden funcionar efectivamente en un entorno basado en canal. A modo de ejemplo numerosos efectos ambientales o reverberación se benefician actualmente de ser alimentados a redes de altavoces. Aunque podrían tratarse con objetos con anchura suficiente para rellenar un conjunto matricial, es conveniente retener alguna funcionalidad basada en canal.

El sistema de audio adaptativo está configurado para soportar "lechos" además de los objetos de audio, en donde los lechos son efectivamente mezclas o derivaciones basadas en canal. Se pueden proporcionar para la reproducción final (presentación) bien sea individualmente, bien sea en combinación en un lecho auditivo único, dependiendo de la intención del creador de contenidos. Estos lechos pueden crearse en diferentes configuraciones basadas en canal tales como 5.1, 7.1 y 9.1, y redes matriciales que incluyen altavoces aéreos, tales como los ilustrados en la Figura 1. La Figura 2 ilustra la combinación de datos basados en canal y datos basados en objeto para proporcionar una mezcla de audio adaptativa, en conformidad con una forma de realización. Según se ilustra en el proceso 200, los datos basados en canal 202 que, a modo de ejemplo, pueden ser datos sonoros envolventes 5.1 o 7.1 proporcionados en la forma de datos modulados por códigos de pulsos (PCM) se combinan con los datos de objetos de audio 204 para obtener una mezcla de audio adaptativo 208. Los datos de objeto de audio 204 se obtienen combinando los elementos de los datos basados en canal originales con metadatos asociados que especifican algunos parámetros que pertenecen a la localización de los objetos de audio. Según se ilustra conceptualmente en la Figura 2, las herramientas de autoría de edición proporcionan la capacidad para crear programas de audio que contengan una combinación de grupos de canales de altavoces y canales de objetos, simultáneamente. A modo de ejemplo, un programa de audio podría contener uno o más canales de altavoces opcionalmente organizados en grupos (o pistas, p.ej., una pista estéreo o 5.1), metadatos descriptivos para uno o más canales de altavoces, uno o más canales de objetos y metadatos descriptivos para uno o más canales de objetos.

Un sistema de audio adaptativo se mueve efectivamente más allá de simples "alimentaciones de altavoces" como un medio para distribuir audio espacial, y se han desarrollado descripciones de audio basadas en modelos de carácter avanzado que permiten al oyente la libertad para seleccionar una configuración de reproducción que se adapte mejor a sus necesidades individuales o presupuesto y tienen el audio presentado específicamente para su configuración individualmente elegida. A un nivel alto, existen cuatro formatos de descripción de audio espaciales principales: (1) alimentación de altavoz, en donde el contenido de audio se describe como señales previstas para altavoces situados en posiciones nominales; (2) alimentación de micrófono, en donde el contenido de audio se describe como señales capturadas por micrófonos reales o virtuales en una configuración predefinida (el número de micrófonos y su posición relativa); (3) descripción basada en el modelo, en donde el contenido de audio se describe en términos de una secuencia de eventos de audio en tiempos y posiciones que se describen y (4) binaurales, en donde el contenido de audio se describe por las señales que llegan a los dos oídos de la persona que escucha.

Los cuatro formatos de descripción se suelen asociar con las tecnologías de presentación comunes siguientes, en donde el término “presentación” significa la conversión a señales eléctricas utilizadas como alimentaciones de altavoces: (1) obtener panorámicas en donde el flujo de audio se convierte a alimentaciones de altavoces que utilizan un conjunto de leyes para obtener panorámicas y posiciones de altavoces conocidas o asumidas (normalmente presentadas antes de la distribución); (2) la así denominada ambisónica, en donde las señales microfónicas se convierten a alimentaciones para una red escalable de altavoces (normalmente presentados después de la distribución); (3) Síntesis de Campo de Ondas (WFS), en donde los eventos sonoros se convierten en las señales de altavoces adecuadas para sintetizar un campo sonoro (normalmente presentado después de la distribución); y (4) binaurales, en donde las señales binaurales de izquierda/derecha se entregan al oído izquierdo/derecho normalmente a través de auriculares, pero también por intermedio de altavoces en conjunción con una cancelación de la diafonía.

En general, cualquier formato puede convertirse a otro formato (aunque esta operación pueda requerir una separación de orígenes a ciegas o una tecnología similar) y presentado utilizando cualquiera de las tecnologías antes citadas; sin embargo, no todas las transformaciones proporcionan resultados adecuados en la práctica. El formato de alimentación de altavoz es el más común porque resulta simple y eficaz. Los mejores resultados sónicos (es decir, los más exactos y fiables) se consiguen mezclando/supervisando y distribuyendo luego las alimentaciones de altavoces directamente puesto que no existe ningún procesamiento requerido entre el creador de contenidos y el oyente. Si el sistema de reproducción se conoce por anticipado, una descripción de alimentación de altavoz proporciona la más alta fidelidad, sin embargo, el sistema de reproducción y su configuración no suele ser conocidos de antemano. Por el contrario, la descripción basada en el modelo es la más adaptable puesto que no establece ningún supuesto sobre el sistema de reproducción y por lo tanto, es de más fácil aplicación a múltiples tecnologías de presentación. La descripción basada en el modelo puede capturar eficientemente información espacial, pero se hace muy ineficiente cuando aumenta el número de fuentes de audio.

El sistema de audio adaptativo combina los beneficios de ambos sistemas basados en canal y basados en el modelo, con beneficios específicos que incluyen alta calidad del timbre, reproducción óptima de la intención artística cuando se mezclan y presentan utilizando la misma configuración de canal, inventario único con adaptación “descendente” para la configuración de presentación, impacto relativamente bajo sobre la conducción acústica del sistema y mayor inmersión mediante una resolución espacial del altavoz horizontal más fina y nuevos canales de altura. El sistema de audio adaptativo proporciona varias nuevas características que incluyen: un inventario único con adaptación en sentido descendente y ascendente a una configuración de presentación cinematográfica específica, esto es, presentación de retardo y uso óptimo de altavoces disponibles en un entorno de reproducción; envolvente aumentada, incluyendo la mezcla descendente optimizada para evitar artefactos de correlación intercanales (ICC); mayor resolución espacial mediante redes denominadas *steer-thru* (p.ej., que permiten a un objeto de audio asignarse dinámicamente a uno o más altavoces dentro de una red de sonido); y aumento de la resolución del canal frontal mediante una configuración de altavoces similares o centrales de alta resolución.

Los efectos espaciales de señales de audio son críticas para proporcionar una experiencia inmersiva para el oyente. Los sonidos que se conoce que emanan desde una zona específica de una pantalla de visión o de un entorno de escucha deben reproducirse a través de altavoces situados en esa misma localización relativa. De este modo, el metadatos de audio primario de un evento sonoro en una descripción basada en el modelo es la posición, aunque otros parámetros tales como tamaño, orientación, velocidad y dispersión acústica pueden describirse también. Para transmitir la posición, una descripción espacial de audio en 3D, basada en el modelo, requiere un sistema de coordenadas de 3D. El sistema de coordenadas utilizado para la transmisión (euclidiana, esférica, cilíndrica) suele elegirse por conveniencia o por naturaleza compacta; sin embargo, otros sistemas de coordenadas pueden utilizarse para el procesamiento de presentación. Además de un sistema de coordenadas, se requiere una trama de referencia para representar las localizaciones de los objetos en el espacio. Para sistemas para reproducir, con exactitud, el sonido basado en la posición en una diversidad de entornos diferentes, la selección de la trama de referencia adecuada puede resultar crítica. Con una trama de referencia alocéntrica, una posición de fuente de audio se define en relación con características dentro del entorno de presentación tales como paredes y esquinas de la sala, localizaciones de altavoces estándar y localización de la pantalla. En una trama de referencia egocéntrica, las localizaciones se representan con respecto a la perspectiva del oyente, tal como “en frente de mí”, “ligeramente a la izquierda”, etc. Estudios científicos de la percepción espacial (audio y de otro modo) han demostrado que la perspectiva egocéntrica se utiliza casi universalmente. Para el evento cinematográfico, sin embargo, la trama de referencia alocéntrica suele ser la más adecuada. A modo de ejemplo, la localización precisa de un objeto de audio es más importante cuando existe un objeto asociado en la pantalla. Cuando se utiliza una referencia alocéntrica, para cada posición de escucha y para cualquier tamaño de la pantalla, el sonido se localizará en la misma posición relativa en la pantalla, p.ej., “un tercio a la izquierda del centro de la pantalla”. Otra razón es que los mezcladores tienden a considerar y mezclar en términos alocéntricos, y las herramientas para obtener panorámicas se depositan con una trama alocéntrica (es decir, las paredes de la sala) y los mezcladores esperan que se presenten de esa manera, p.ej., “este sonido debe estar en la pantalla”, “este sonido debe estar fuera de la pantalla” o “desde la pared izquierda”, etc.

A pesar del uso de la trama alocéntrica de referencia en el entorno cinematográfico, existen algunos casos en donde

una trama egocéntrica de referencia puede ser de utilidad y más adecuado. Estos casos incluyen sonidos no diegéticos, esto es, los que no están presentes en el “espacio de historia”, p.ej., música de fondo de ambiente, para los cuales puede ser deseable una presentación egocéntricamente uniforme. Otro caso es el de los efectos de campos cercanos (p.ej., un zumbido de mosquito en el oído izquierdo del oyente) que requieren una representación egocéntrica. Además, fuentes sonoras infinitamente lejanas (y las ondas planas resultantes) pueden aparecer desde una posición egocéntrica constante (p.ej., 30 grados a la izquierda), y dichos sonidos son más fáciles de describir en términos egocéntricos que en términos alocéntricos. En algunos casos, es posible utilizar una trama de referencia alocéntrica en tanto que se defina una posición de escucha nominal, mientras que algunos ejemplos requieren una representación egocéntrica que no es todavía posible presentar. Aunque una referencia alocéntrica puede ser más útil y adecuada, la representación de audio debe ser extensible, puesto que numerosas nuevas características, incluyendo la presentación egocéntrica, pueden ser más deseables para determinadas aplicaciones y entornos de escucha.

Las formas de realización del sistema de audio adaptativo incluyen un método de descripción espacial híbrido que comprende una configuración de canal recomendada para una fidelidad óptima y para la presentación de fuentes difusas o complejas, de tipo multipunto (p.ej., congestión de estadio, ambiente) utilizando una referencia egocéntrica, junto con una descripción del sonido basada en el modelo alocéntrica para permitir un aumento eficiente de la resolución espacial y de la escalabilidad. La Figura 3 es un diagrama de bloques de una arquitectura de reproducción para uso en un sistema de audio adaptativo, en conformidad con una forma de realización. El sistema ilustrado en la Figura 3 incluye el procesamiento de bloques que realizan una decodificación de audio de canal y de objeto, como legado, presentación de objetos, remapeado de canales y procesamiento de señales antes de que la señal de audio se envíe para las etapas de post-procesamiento y/o amplificación y altavoces.

El sistema de reproducción 300 está configurado para presentar y reproducir un contenido de audio que se genera mediante uno o más componentes de captura, pre-procesamiento, autoría de edición y codificación. Un preprocesador de audio adaptativo puede incluir una funcionalidad de detección del tipo de contenidos y de separación de fuentes que genera automáticamente metadatos adecuados mediante análisis de audio de salida. A modo de ejemplo, metadatos posicionales pueden derivarse desde un registro multicanal mediante un análisis de los niveles relativos de entrada correlacionada entre pares de canales. La detección del tipo de contenido, tal como “voz” o “música”, puede conseguirse, a modo de ejemplo, mediante extracción y clasificación de características. Algunas herramientas de autoría de edición permiten determinar la autoría del programa de audio optimizando la entrada y la codificación de la intención creativa del ingeniero de sonido que le permite crear la mezcla de audio final una vez que esté optimizada para su reproducción en prácticamente cualquier entorno de reproducción. Lo que antecede puede conseguirse mediante el uso de objetos de audio y datos posicionales que están asociados y codificados con el contenido de audio original. Con el fin de situar, con exactitud los sonidos en torno a un auditorio, el ingeniero de sonido necesita controlar cómo el sonido se presentará, en última instancia, sobre la base de las limitaciones reales y las características del entorno de reproducción. El sistema de audio adaptativo proporciona este control permitiendo al ingeniero de sonido cambiar cómo el contenido de audio está diseñado y mezclado mediante el uso de objetos de audio y datos posicionales. Una vez que el contenido de audio adaptativo haya sido objeto de determinación de autoría y codificado en los dispositivos códec adecuados, se decodifica y presenta en los diversos componentes del sistema de reproducción 300.

Según se ilustra en la Figura 3, (1) el audio de sonido envolvente de legado 302, (2) el audio de objeto incluyendo metadatos de objeto 304, y (3) audio de canal incluyendo metadatos de canales 306 son objeto de entrada a los datos de decodificador 308, 309 dentro del bloque de procesamiento 310. Los metadatos de objetos se presentan en dispositivos de presentación de objetos 312, mientras que los metadatos de canales pueden ser objeto de un nuevo mapeado de correspondencia si es necesario. La información de configuración del entorno de escucha 307 se proporciona al dispositivo presentador de objetos y al componente de remapeado de canales. Los datos de audio híbridos se procesan luego por intermedio de una o más etapas de procesamiento de señales, tales como ecualizadores y limitadores 314 antes de la salida a la etapa de procesamiento de cadena B 316 y la reproducción por intermedio de altavoces 318. El sistema 300 representa un ejemplo de un sistema de reproducción para audio adaptativo, y son posibles también otras configuraciones, componentes e interconexiones.

El sistema de la Figura 3 ilustra una forma de realización en la que el dispositivo de presentación comprende un componente que aplica metadatos de objetos a los canales de audio de entrada para procesar el contenido de audio basado en el objeto en conjunción con el contenido de audio basado en el canal opcional. Las formas de realización pueden destinarse también a un caso en el que los canales de audio de entrada comprenden un contenido basado en canal de legado solamente, y el dispositivo de presentación comprende un componente que genera alimentaciones de altavoces para la transmisión a un conjunto de excitaciones en una configuración de sonido envolvente. En este caso, la entrada no es necesariamente un contenido con base en el objeto, sino un contenido de legado 5.1 o 7.1 (u registro contenido basado no basado en el objeto), tal como se proporciona en el sistema de Dolby Digital o Dolby Digital Pluralidades o sistemas similares.

Aplicaciones de reproducción

Según se indicó con anterioridad, una puesta en práctica inicial del formato de audio adaptativo y del sistema



correspondiente está en el contexto cinematográfico digital (D-cinema) que incluye la captura del contenido (objetos y canales) cuya autoría se consigue utilizando nuevas herramientas de autoría de edición, empaquetados utilizados un codificador cinematográfico de audio adaptativo, y distribuido utilizando PCM sin pérdidas patentado que utiliza el mecanismo de distribución de Iniciativa Cinematográfica Digital (DCI). En este caso, el contenido de audio está previsto para decodificarse y presentarse en una sala cinematográfica digital para crear una experiencia cinematográfica de audio espacial inmersiva. Sin embargo, como con las anteriores mejoras cinematográficas, tales como sonido envolvente analógico, audio multicanal digital, etc., existe un imperativo para proporcionar una experiencia del oyente mejorada, proporcionada por el formato de audio adaptativo directamente a los oyentes en su entorno residencial. Lo que antecede requiere que algunas características del formato y del sistema sean adaptadas para uso en entornos de escucha más limitados. A modo de ejemplo, domicilios propios, salas, pequeños auditorios o lugares similares pueden tener un espacio reducido, con propiedades acústicas y capacidades de equipo comparables a las de un entorno de sala cinematográfica o teatro. Para los fines de la descripción, el término “entorno basado en el consumo” está previsto que se incluya cualquier entorno no cinematográfico que comprenda un entorno de escucha para uso por consumidores o profesionales normales, tal como una vivienda, un estudio, una sala, una zona de consola, un auditorio y recinto similar. El contenido de audio puede ser captado en origen y presentado solo o puede asociarse con contenido de gráficos, p.ej., imágenes fijas, visualización luminosa, vídeo, etc.

La Figura 4A es un diagrama de bloques que ilustra los componentes funcionales para adaptar el contenido de audio basado en cinematografía para uso en un entorno de escucha en conformidad con una forma de realización. Según se ilustra en la Figura 4A, el contenido cinematográfico suele comprender una pista sonora de imágenes animadas que se captura y/o es objeto de autoría de edición utilizando un equipo adecuado y herramientas contenidas en el bloque 402. En un sistema de audio adaptativo, este contenido se procesa mediante la codificación/decodificación y presentación de componentes e interfaces en el bloque 404. El objeto resultante y las alimentaciones de audio de canales se envían luego a los altavoces adecuados en la sala cinematográfica o teatro 406. En el sistema 400, el contenido del sistema se procesa también para su reproducción en un entorno de escucha, tal como un sistema de entorno residencial 416. Se presupone que el entorno de escucha no es global o capaz de reproducir todo el contenido sonoro que está previsto por el creador de contenidos debido a la limitación de espacio, número reducido de altavoces, etc. Sin embargo, las formas de realización están destinadas a sistemas y métodos que permitan que el contenido de audio original sea presentado en una manera que reduzca al mínimo las restricciones impuestas por la capacidad reducida del entorno de escucha y permitan el procesamiento de las pistas posicionales en una manera que haga máximo el rendimiento del equipo disponible. Según se ilustra en la Figura 4A, el contenido de audio cinematográfico se procesa mediante un componente de conversión de ambiente cinematográfico a de consumo 408, en donde se procesa en la cadena de codificación y presentación de contenidos de consumidores 414. Esta cadena procesa también el contenido de audio que se captura y/o es objeto de autoría de edición en el bloque 412. El contenido original y/o el contenido cinematográfico convertido se reproducen luego en el entorno de escucha 416. De este modo, la información espacial pertinente que se codifica en el contenido de audio puede utilizarse para presentar el sonido en una manera más inmersiva, incluso utilizando la configuración de altavoces posiblemente limitada del entorno de escucha o residencial 416.

La Figura 4B ilustra los componentes de la Figura 4A con mayor detalle. La Figura 4B ilustra un mecanismo de distribución ejemplo para contenido cinematográfico de audio adaptativo por intermedio de un ecosistema de reproducción de audio. Según se ilustra en la diagrama 420, el contenido de TV y cinematográfico original se captura 422 y es objeto de autoría de edición 423 para su reproducción en una diversidad de entornos diferentes para proporcionar una experiencia cinematográfica 427 o experiencias de entornos de consumidores 434. Análogamente, algún contenido generado por el usuario (UGC) o contenido de consumidor se captura 423 y es objeto de autoría de edición 425 para su reproducción en el entorno de escucha 434. El contenido cinematográfico para reproducción en el entorno cinematográfico 427 se procesa mediante procesos cinematográficos conocidos 426. Sin embargo, en el sistema 420, la salida de la caja de herramientas de autoría cinematográfica 423 consiste también en objetos de audio, canales de audio y metadatos que transmiten la intención artística del mezclador de sonidos. Lo que antecede puede conseguirse como un paquete de audio de estilo intermedio que puede utilizarse para crear múltiples versiones del contenido cinematográfico para su reproducción. En una forma de realización, esta funcionalidad se proporciona por un convertidor de audio adaptativo de entorno cinematográfico a entorno del consumidor 430. Este convertidor tiene una entrada para el contenido de audio adaptativo y obtiene a partir de dicho contenido el contenido de audio adecuado y metadatos para los puntos finales del consumidor deseados 434. El convertidor crea salidas de audio y metadatos separadas, y posiblemente diferentes, dependiendo del mecanismo de distribución y del punto final.

Según se ilustra en el ejemplo del sistema 420, el convertidor de entorno cinematográfico al del consumidor 430 alimenta el sonido para la imagen (difusión, disco, OTT, etc.) y módulos de creación de flujos de bits de audio de juegos 428. Estos dos módulos que son adecuados para proporcionar un contenido cinematográfico, pueden alimentarse en múltiples conductos acústicos de distribución 432, todos los cuales se pueden proporcionar a los puntos finales de consumidores. A modo de ejemplo, el contenido cinematográfico de audio adaptativo puede codificarse utilizando un códec adecuado para fines de difusión tal como un Dolby Digital Plus, que puede modificarse para transmitir canales, objetos y metadatos asociados, y se transmite por intermedio de la cadena de difusión mediante cable o satélite y luego, se decodifica y presenta en un entorno residencial o reproducción

televisiva. De modo similar, el mismo contenido podría codificarse utilizando un códec adecuado para distribución en línea en donde esté limitado el ancho de banda, en donde se transmite luego a través de una red móvil 3G o 4G y luego, se decodifica y presenta para su reproducción por intermedio de un dispositivo móvil con el uso de auriculares. Otras fuentes de contenido tales como TV, difusión en directo, juegos y música pueden utilizar también el formato de audio adaptativo para crear y proporcionar un contenido para un formato de audio de la siguiente generación.

El sistema de la Figura 4B proporciona experiencia del oyente mejorada por intermedio del ecosistema de audio del consumidor global que puede incluir un entorno residencial (receptor A/V, barra de sonido y BluRay), soportes electrónicos (ordenador personal PC, tableta electrónica, sistema móvil incluyendo reproducción por auriculares), difusión (TV y decodificador, música, juegos, sonido en directo, contenido generado por el usuario ("UGC")). Dicho sistema proporciona: una inmersión mejorada para la audiencia para todos los dispositivos de punto final, mayor control artístico para los creadores de contenido de audio, mejores metadatos dependientes del contenido (descriptivos) para una presentación mejorada, mayor flexibilidad y escalabilidad para los sistemas de reproducción, preservación y coincidencia del timbre y la oportunidad para la presentación dinámica del contenido sobre la base de la posición del oyente y de la interacción. El sistema incluye varios componentes que comprenden nuevas herramientas mezcladoras para creadores de contenidos, herramientas de empaquetado y codificación nueva y actualizada para distribución y reproducción, mezcla y presentación dinámica en el entorno residencial (adecuado para diferentes configuraciones), localizaciones y diseños de altavoces adicionales.

El sistema de audio adaptativo está configurado para ser un sistema de audio global, de extremo a extremo y de la siguiente generación con el uso del formato de audio adaptativo que incluye creación de contenidos, empaquetado, distribución y reproducción/presentación a través de un número amplio de dispositivos de punto final y casos de uso. Según se ilustra en la Figura 4B, el sistema tiene su origen con el contenido capturado desde y para varios diferentes casos de uso 422 y 424. Estos puntos de captura incluyen todos los formatos de contenido pertinentes incluyendo el formato cinematográfico, TV, difusión en directo (y sonido), UGC, juegos y música. El contenido cuando pasa a través del ecosistema, pasa a través de varias fases claves, tales como preprocesamiento y herramientas de autoría de audición, herramientas de conversión (p.ej., conversión de contenido de audio adaptativo para aplicaciones de distribución de contenidos cinematográficos o de consumidores), empaquetado de audio adaptativo específico/codificación de flujo de bits (que captura datos esenciales de audio así como metadatos adicionales y la información de reproducción de audio), codificación de distribución utilizando *codecs* nuevos o existentes (p.ej., DD+, TrueHD, Dolby Pulse) para una distribución eficiente a través de varios canales de audio, una transmisión a través de los canales de distribución pertinentes (difusión, disco, móvil, Internet, etc.) y por último, presentación dinámica dependiendo del punto final para reproducir y transmitir la experiencia del oyente de audio adaptativo definida por el creador de contenidos que proporciona las ventajas de la experiencia de audio espacial. El sistema de audio adaptativo puede utilizarse durante la presentación para un número ampliamente variable de puntos finales del consumidor, y la técnica de presentación que se aplica puede optimizarse dependiendo del dispositivo del punto final. A modo de ejemplo, los sistemas de entorno residencial y barras de sonido pueden tener 2, 3, 5, 7 o incluso 9 altavoces separados en varias localizaciones. Muchos otros tipos de dispositivos tienen solamente dos altavoces (TV, ordenador portátil, caja de música) y casi todos los dispositivos de uso común tienen una salida de auriculares (ordenador personal PC, ordenador portátil, tableta electrónica, teléfono móvil, reproductor de música, etc.).

Los sistemas de autoría de edición y de distribución actuales para crear un audio sonoro envolvente y proporcionar audio que esté previsto para reproducción para localizaciones de altavoces fijas y predefinidas con conocimiento limitado del tipo de contenido transmitido en la esencia de audio (esto es, el audio real que se reproduce por el sistema de reproducción). El sistema de audio adaptativo, sin embargo, proporciona un nuevo método híbrido para creación de audio que incluye la opción para a la vez, audio específico de localización de altavoz fija (canal izquierdo, canal derecho, etc.) y elementos de audio basados en el objeto que tienen información espacial en 3D generalizada que incluye la posición, magnitud y velocidad. Este método híbrido proporciona un enfoque equilibrado para la fidelidad (proporcionada por localizaciones de altavoces fijas) y flexibilidad en la presentación (objetos de audio generalizados). Este sistema proporciona también información de utilidad adicional sobre el contenido de audio por intermedio de nuevos metadatos que están emparejados con la esencia de audio por el creador de contenidos en el momento de la creación/autoría de edición de contenidos. Esta información proporciona detalles sobre los atributos de la señal de audio que pueden utilizarse durante la presentación. Dichos atributos pueden incluir el tipo de contenido (diálogo, música, efecto, Foley, fondo/ambiente, etc.) así como información de objeto de audio tal como atributos espaciales (posición 3D, magnitud de objeto, velocidad, etc.) e información de presentación de utilidad (permitir localización de altavoces, pesos de canales, ganancia, información de gestión de bajos, etc.). El contenido de audio y los metadatos de intención de reproducción pueden crearse bien sea manualmente por el creador de contenidos o bien, crearse mediante el uso de algoritmos de inteligencia multimedia automáticos que pueden ejecutarse en la estructura de fondo durante el proceso de autoría de edición y revisarse por el creador de contenidos durante una fase de control de calidad final, si así se desea.

La Figura 4C es un diagrama de bloques de los componentes funcionales de un entorno de audio adaptativo en conformidad con una forma de realización. Según se ilustra en el diagrama 450, el sistema procesa un flujo de bits codificado 452 que soporta un flujo de audio basado en canal y basado en objeto híbrido. El flujo de bits se procesa

mediante el bloque de procesamiento de señal/presentación 454. En una forma de realización, al menos partes de este bloque funcional pueden ponerse en práctica en el bloque de presentación 312 ilustrado en la Figura 3. La función de presentación 454 pone en práctica varios algoritmos de presentación para audio adaptativo, así como algunos algoritmos post-procesamiento tales como mezcla ascendente, procesamiento de sonido directo en comparación con el sonido reflejado, y similares. La salida desde el dispositivo de presentación se proporciona a los altavoces 458 mediante interconexiones bidireccionales 456. En una forma de realización, los altavoces 458 comprenden varios controladores individuales que pueden disponerse en una configuración de sonido envolvente o similar. Los controladores son individualmente direccionables y pueden materializarse en recintos individuales o armarios o redes de múltiples controladores. El sistema 450 puede incluir también micrófonos 460 que proporcionan mediciones del entorno de escucha o características de la sala que puede utilizarse para calibrar el proceso de presentación. Las funciones de calidad y configuración del sistema se proporcionan en el bloque 462. Estas funciones pueden incluirse como parte de los componentes de presentación, o pueden ponerse en práctica como componentes separados que están funcionalmente acoplados al dispositivo de presentación. Las interconexiones bidireccionales 456 proporcionan la ruta de señal de realimentación acústica desde los altavoces en el entorno de escucha de nuevo hacia el componente de calibración 462.

#### Entornos de escucha

Puestas en práctica del sistema de audio adaptativo pueden desplegarse en una diversidad de entornos de escucha diferentes. Estos últimos incluyen tres áreas primarias de aplicaciones de reproducción de audio: sistemas de entorno residencial, televisiones y barras acústicas así como auriculares. La Figura 5 ilustra el despliegue de un sistema de audio adaptativo en un entorno residencial a modo de ejemplo. El sistema ilustrado en la Figura 5 es un superconjunto de componentes y funciones que puede proporcionarse por un sistema de audio adaptativo y algunos aspectos pueden reducirse o eliminarse sobre la base de las necesidades del oyente, al mismo tiempo que se proporciona una experiencia mejorada. El sistema 500 incluye varios altavoces diferentes y controladores en una diversidad de armarios o redes diferentes 504. Los altavoces incluyen controladores individuales que proporcionan opciones de proyección frontal, lateral y hacia arriba así como una virtualización dinámica de audio utilizando algunas técnicas de procesamiento de audio. El diagrama 500 ilustra varios altavoces desplegados en una configuración de altavoces 9.1 estándar. Estos altavoces de altura izquierdo y derecho (LH, RH), altavoces izquierdos y derechos (L, R), un altavoz central (ilustrado como un altavoz central modificado) y altavoces envolventes izquierdo y derecho y posteriores (LS, RS, LB y RB, no siendo ilustrado el elemento de baja frecuencia LFE).

La Figura 5 ilustra el uso de un altavoz de canal central 510 utilizado en una localización central del entorno de escucha. En una forma de realización, este altavoz se pone en práctica utilizando un canal central modificado o un canal central de alta resolución 510. Dicho altavoz puede ser una red de canal central de proyección frontal con altavoces individualmente direccionables que permiten panorámicas discretas de objetos de audio mediante la red que hace coincidir el movimiento de los objetos de vídeo en la pantalla. Puede materializarse como un altavoz de canal central de alta resolución (HRC), tal como se describe en la solicitud internacional n° PCT/US2011/028783. El altavoz HRC 510 puede incluir también altavoces de proyección lateral, según se ilustra. Estos últimos podrían activarse y utilizarse si el altavoz HRC se utiliza no solamente como un altavoz central sino también como un altavoz con capacidades de barra de sonido. El altavoz HRC puede incorporarse también por encima y/o las partes laterales de la pantalla 502 para proporcionar una opción de obtener panorámicas de alta resolución bidimensional para objetos de audio. El altavoz central 510 podría incluir también controladores adicionales y poner en práctica un haz sonoro direccionable con zonas sonoras controladas por separado.

El sistema 500 incluye también un altavoz de efecto de campo cercano (NFE) 512 que puede localizarse en la parte frontal derecha o próximo en frente del oyente que escucha, tal como sobre la base en frente de un lugar de asiento. Con el audio adaptativo es posible llevar objetos de audio a la sala y no solamente enclavarse para el perímetro de la sala. Por lo tanto, tener objetos en sentido transversal a través del espacio tridimensional es una opción. Una realización ejemplo es donde un objeto puede originarse en el altavoz L, desplazarse a través del entorno de escucha hacia el altavoz NFE y terminar en el altavoz RS. Varios altavoces diferentes pueden ser adecuados para uso como un altavoz NFE, tal como un altavoz inalámbrico alimentado por baterías.

La Figura 5 ilustra el uso de virtualización de altavoz dinámica para proporcionar una experiencia del oyente inmersiva en el entorno de escucha residencial. La virtualización de altavoces de tipo dinámico se permite mediante el control dinámico de los parámetros de algoritmos de virtualización de altavoces sobre la base de la información espacial de objetos que se proporciona por el contenido de audio adaptativo. Esta virtualización dinámica se ilustra en la Figura 5 para los altavoces L y R en donde es natural considerarla para crear la percepción de objeto que se desplazan a lo largo de las partes laterales del entorno de escucha. Un virtualizador separado puede utilizarse para cada objeto pertinente y la señal combinada puede enviarse a los altavoces L y R para crear un efecto de virtualización de objeto múltiple. Los efectos de virtualización de tipo dinámico se ilustran para los altavoces L y R, así como para el altavoz NFE, que está previsto para ser un altavoz estéreo (con dos entradas independientes). Este altavoz, junto con la información de magnitud y posición de audio, podría utilizarse para crear una experiencia de audio de campos cercanos de origen puntual o difuso. Efectos de virtualización similares pueden aplicarse también a cualquiera o la totalidad de los demás altavoces en el sistema. En una forma de realización, una cámara puede

proporcionar información de posición y de identidad adicional del oyente que podría utilizarse por el dispositivo de presentación de audio adaptativo para proporcionar una experiencia más impresionante y más verdadera para la intención artística del mezclador.

5 El dispositivo de presentación de audio adaptativo entiende la relación espacial entre el sistema de mezcla y de reproducción. En algunas instancias operativas de un entorno de reproducción, altavoces discretos pueden estar disponibles en todas las zonas pertinentes del entorno de escucha, incluyendo posiciones de auriculares según se ilustra en la Figura 1. En estos casos en donde altavoces discretos están disponibles en algunas localizaciones, el dispositivo de presentación puede configurarse para fijar objetos para los altavoces más próximos en lugar de crear una imagen fantasma entre dos o más altavoces mediante la función de obtener panorámicas o el uso de algoritmos de virtualización de altavoces. Aunque distorsiona ligeramente la representación espacial de la mezcla, permite también al dispositivo de presentación evitar imágenes fantasmas imprevistas. A modo de ejemplo, si la posición angular de altavoz izquierdo de la etapa de mezcla no corresponde a la posición angular del altavoz izquierdo del sistema de reproducción, lo que permite a esta función evitar tener una imagen fantasma constante del canal izquierdo inicial.

En numerosos casos, sin embargo, y en particular en un entorno residencial, algunos altavoces, tales como los altavoces aéreos montados en el techo no están disponibles. En este caso, algunas técnicas de virtualización son puestas en práctica por el dispositivo de presentación para reproducir un contenido de audio aéreo a través de los altavoces montados en las paredes o suelo existentes. En una forma de realización, el sistema de audio adaptativo incluye una modificación para la configuración estándar mediante la inclusión de, a la vez, una capacidad de proyección frontal y una capacidad de proyección superior (o "hacia arriba") para cada altavoz. En las aplicaciones domésticas tradicionales, los fabricantes de altavoces han intentado introducir nuevas configuraciones controladoras distintas a los transductores de proyección frontal y se han enfrentado con el problema de intentar identificar cuáles de las señales de audio originales (o modificaciones de ellas) deben enviarse a estos nuevos controladores. Con el sistema de audio adaptativo, existen una información muy específica con respecto a qué objeto de audio debe presentarse por encima del plano horizontal estándar. En una forma de realización, se presenta la información de altura presente en el sistema de audio adaptativo utilizando los controladores de proyección hacia arriba. Análogamente, los altavoces de proyección lateral pueden utilizarse para presentar algún otro contenido, tal como efectos ambientales.

Una ventaja de los controladores de proyección hacia arriba es que pueden utilizarse para reflejar sonido fuera de una superficie de techo para simular la presencia de altavoces aéreos/de altura situados en el techo. Un atributo importante del contenido de audio adaptativo es que el audio espacialmente diverso se reproduce utilizando una red de altavoces aéreos. Según se indicó con anterioridad, sin embargo, en numerosos casos, la instalación de altavoces aéreos es demasiado costosa o no resulta práctica en un entorno residencial. Simulando altavoces de altura que utilizan altavoces normalmente situados en el plano horizontal, una experiencia en 3D impresionante puede crearse con facilidad para situar los altavoces. En este caso, el sistema de audio adaptativo está utilizando los controladores de simulación de altura/proyección hacia arriba en una nueva forma en la que los objetos de audio y su información de reproducción espacial se está utilizando para crear el audio que se reproduce por los controladores de proyección hacia arriba.

La Figura 6 ilustra el uso de un controlador de proyección hacia arriba que utiliza sonido reflejado para simular un altavoz aéreo único en un entorno residencial. Conviene señalar que cualquier número de controladores de proyección hacia arriba podría utilizarse en combinación para crear múltiples altavoces de altura simulados. Como alternativa, varios controladores de proyección hacia arriba pueden configurarse para transmitir sonido a prácticamente el mismo lugar en el techo para conseguir un determinado efecto o intensidad sonora. El diagrama 600 ilustra un ejemplo en el que la posición de escucha usual 602 está situada en un lugar particular dentro de un entorno de escucha. El sistema no incluye cualesquiera altavoces de altura para transmitir un contenido de audio que contenga pistas de altura. En cambio, el armario de altavoces o la red de altavoces 604 incluye un controlador de proyección hacia arriba junto con los controladores de proyección frontal. El controlador de proyección hacia arriba está configurado (con respecto al ángulo de localización e inclinación) para enviar la onda sonora 606 hasta un punto particular en el techo 608 en donde se reflejará de nuevo hacia la posición de escucha 602. Se supone que el techo está constituido por un material y composición adecuados para reflejar adecuadamente el sonido hacia abajo en el entorno de escucha. Las características pertinentes del controlador de proyección hacia arriba (p.ej., tamaño, potencia, localización, etc.) pueden seleccionarse sobre la base de la composición del techo, magnitud de la sala y otras características pertinentes del entorno de escucha. Aunque solamente el controlador de proyección hacia arriba se ilustra en la Figura 6, múltiples controladores de proyección hacia arriba pueden incorporarse en un sistema de reproducción en algunas formas de realización.

En conformidad con la invención, el sistema de audio adaptativo utiliza controladores de proyección hacia arriba para proporcionar el elemento de altura. En conformidad con la invención, la incorporación de procesamiento de señal para introducir pistas de altura de percepción en la señal de audio que se alimenta a los controladores de proyección hacia arriba mejora el posicionamiento y la calidad percibida de la señal de altura virtual. A modo de ejemplo, un modelo de audición binaural perceptual paramétrica se ha desarrollado para crear un filtro de pista de altura, que cuando se utiliza para procesar las señales de audio que se reproducen por un controlador de proyección

hacia arriba mejora la calidad percibida de la reproducción. En una forma de realización, el filtro de pista de altura se deriva de la localización del altavoz físico (aproximadamente a nivel con el oyente y la localización del altavoz reflejado (por encima del oyente). Para la localización del altavoz físico, se determina un filtro direccional sobre la base de un modelo del oído exterior (o pabellón auricular). Una forma inversa de este filtro se determina a continuación y se utiliza para eliminar las pistas de altura desde el altavoz físico. A continuación, para la localización del altavoz reflejado, se determina un segundo filtro direccional utilizando el mismo modelo del oído externo. Este filtro se aplica directamente, con la reproducción esencial de las pistas que el oído recibiría si el sonido estuviera por encima del oyente. En la práctica, estos filtros pueden combinarse en una manera que permita un filtro único que, a la vez, (1) elimina la pista de altura desde la localización del altavoz físico y (2) inserte la pista de altura desde la localización del altavoz reflejado. La Figura 16 es un gráfico que ilustra la respuesta de frecuencia para dicho filtro combinado. El filtro combinado puede utilizarse de tal manera que permita alguna ajustabilidad con respecto a la 'agresividad' o la magnitud del filtrado que se aplica. A modo de ejemplo, en algunos casos, puede ser ventajoso no eliminar completamente la pista de altura del altavoz físico, o aplicar totalmente la pista de altura del altavoz reflejado puesto que solamente parte del sonido del altavoz físico llega directamente al oyente (con el resto siendo reflejado desde el techo).

#### Configuración de altavoces

Una consideración principal del sistema de audio adaptativo es la configuración de altavoces. El sistema utiliza controladores individualmente direccionables y una red de dichos controladores está configurada para proporcionar una combinación de fuentes sonoras directas y reflejadas. Un enlace bidireccional al controlador del sistema (p.ej., receptor A/V, decodificador), permite que los datos de audio y de configuración sean enviados al altavoz y la información del altavoz y del sensor sea enviada de nuevo al controlador, creando un sistema en bucle cerrado activo.

Para los fines de la descripción, el término "controlador" significa un transductor electroacústico único que produce sonidos en respuesta a una señal de entrada de audio eléctrica. Un controlador puede ponerse en práctica en cualquier tipo, geometría y tamaño adecuados y puede incluir bocinas, conos, transductores de cinta y componentes similares. El término "altavoz" significa uno o más controladores en un recinto unitario. La Figura 7A ilustra un altavoz que tiene una pluralidad de controladores en una primera configuración, en conformidad con una forma de realización. Según se ilustra en la Figura 7A, un recinto de altavoces 700 tiene varios controladores individuales montados dentro del recinto. En condiciones normales, el recinto incluirá uno o más controladores de proyección frontal 702, tales como altavoces de graves, altavoces de gama media o altavoces de agudos o cualquiera de sus combinaciones. Uno o más controladores de proyección lateral 704 pueden incluirse también a este respecto. Los controladores de proyección frontal y lateral suelen estar montados a tope contra la parte lateral del recinto, de modo que proyecten el sonido perpendicularmente hacia fuera desde el plano vertical definido por el altavoz y estos controladores suelen estar permanentemente fijos dentro del armario 700. Para el sistema de audio adaptativo que tiene la funcionalidad de presentación del sonido reflejado se proporcionan también uno o más controladores inclinados hacia arriba 706. Estos controladores están situados de modo que proyecten sonido en un ángulo hacia arriba al techo en donde pueden rebotar luego hacia un oyente, según se ilustra en la Figura 6. El lado de inclinación puede establecerse dependiendo de las características del entorno de escucha y de los requisitos del sistema. A modo de ejemplo, el controlador de proyección hacia arriba 706 puede incluirse entre 30 y 60 grados y puede situarse por encima del controlador de proyección frontal 702 en el recinto de altavoces 700 con el fin de minimizar la interferencia con las ondas sonoras producidas desde el controlador de proyección frontal 702. El controlador de proyección hacia arriba 706 puede instalarse en un ángulo fijo, o puede instalarse de modo que el ángulo de inclinación se pueda ajustar manualmente. Como alternativa, un servomecanismo puede utilizarse para permitir el control automático o eléctrico del ángulo de inclinación y la dirección de proyección del controlador de proyección hacia arriba. Para algunos sonidos, tal como un sonido ambiente, el controlador de proyección hacia arriba puede apuntarse directamente desde una superficie superior del recinto de altavoces 700 para crear lo que podría referirse como un controlador de "proyección superior". En este caso, una gran componente del sonido puede reflejarse de nuevo hacia el altavoz, dependiendo de las características acústicas del techo. En la mayor parte de los casos, sin embargo, algún ángulo de inclinación se suele utilizar para ayudar a proyectar el sonido mediante la recepción desde el techo a una posición diferente o más central dentro del entorno de escucha, según se ilustra en la Figura 6.

La Figura 7A está prevista para ilustrar un ejemplo de una configuración de altavoces y controladores, y son posibles muchas otras configuraciones. A modo de ejemplo, el controlador de proyección hacia arriba puede proporcionarse en su propio recinto para permitir el uso de los altavoces existentes. La Figura 7B ilustra un sistema de altavoces que tienen controladores distribuidos en múltiples recintos, en conformidad con una forma de realización. Según se ilustra en la Figura 7B, el controlador de proyección hacia arriba 712 está provisto en un recinto separado 710, que puede colocarse luego próximo a o por encima de un recinto 714 que tiene controladores de proyección frontal y/o lateral 716 y 718. Los controladores pueden encerrarse también dentro de una barra de sonido de altavoz, tal como se utiliza en numerosos entornos de salas residenciales, en donde un número de controladores de tamaño pequeño o medio están dispuestos a lo largo de un eje dentro de un recinto vertical u horizontal único. La Figura 7C ilustra la colocación de controladores dentro de una barra de sonido, en conformidad con una forma de realización. En este ejemplo, el recinto de barra de sonido 730 es una barra de sonido horizontal que incluye controladores de proyección lateral 734, controladores de proyección hacia arriba 736 y controladores de proyección frontal 732. La

Figura 7C está prevista para ser una configuración ejemplo solamente, y cualquier número práctico de controladores para cada una de las funciones – proyección frontal, lateral y hacia arriba – puede utilizarse a este respecto.

5 Para la forma de realización ilustrada en las Figuras 7A-C, debe indicarse que los controladores pueden ser de cualquier forma, tamaño y tipo adecuados, dependiendo de las características de respuestas de frecuencias requeridas, así como de cualesquiera otras limitaciones pertinentes, tales como magnitud, potencia nominal, coste de componentes, etc.

10 En un entorno de audio adaptativo típico, varios recintos de altavoces estarán contenidos dentro del entorno de escucha. La Figura 8 ilustra una colocación, a modo de ejemplo, de altavoces que tienen controladores individualmente direccionables incluyendo controladores de proyección hacia arriba situados dentro de un entorno de escucha. Según se ilustra en la Figura 8, el entorno de escucha 800 incluye cuatro altavoces individuales 806  
15 teniendo cada uno de ellos al menos un controlador de proyección frontal, proyección lateral y proyección hacia arriba. El entorno de escucha puede contener también controladores fijos utilizados para aplicaciones de sonido envolvente, tales como un altavoz central 802 y altavoz de graves o LFE 804. Como puede observarse en la Figura 8, dependiendo de la magnitud del entorno de escucha y de las respectivas unidades de altavoces, la colocación adecuada de altavoces 806 dentro del entorno de escucha puede proporcionar un entorno de audio funcionalmente rico que resulte de la recepción de sonidos desde el techo procedente de varios controladores de proyección hacia arriba. Los altavoces pueden tener como objetivo proporcionar una reflexión de uno o más puntos en el plano del  
20 techo dependiendo del contenido, de la magnitud del entorno de escucha, de la posición del oyente, de las características acústicas y de otros parámetros pertinentes.

25 Los altavoces utilizados en un sistema de audio adaptativo para un entorno residencial o entorno de escucha similar puede utilizar una configuración que esté basada en configuraciones de sonido envolvente ya existentes (p.ej., 5.1, 7.1, 9.1, etc.). En este caso, varios controladores se proporcionan y definen según el convenio de sonido envolvente conocido, con controladores adicionales y definiciones establecidas para los componentes de sonido de proyección hacia arriba.

30 La Figura 9A ilustra una configuración de altavoz para un sistema de audio adaptativo 5.1 que utiliza múltiples controladores direccionables para audio reflejado, en conformidad con una forma de realización. En la configuración 900, una huella de altavoz 5.1 estándar que comprende LFE 901, un altavoz central 902, altavoces frontales L/R 904/906 y altavoces posteriores L/R 908/910 están provistos de ocho controladores adicionales que proporcionan un total de 14 controladores direccionables. Estos ocho controladores adicionales se indican como "hacia arriba" y "hacia parte lateral" además de "hacia delante" (o "frontal") en cada unidad de altavoces 902-910. Los controladores  
35 hacia delante directos serían controlados por subcanales que contienen objetos de audio adaptativo y cualesquiera otros componentes que designen para tener un alto grado de direccionalidad. Los controladores con proyección hacia arriba (reflejados) podrían contener un contenido de subcanales que sea más omnidireccional o sin dirección, pero no está así limitado. Ejemplos incluirían la música de fondo, o sonidos medioambientales. Si la entrada al sistema comprende un contenido de sonido envolvente de legado, en tal caso, este contenido podría ser  
40 inteligentemente plasmado en subcanales directos y reflejados y alimentados a los controladores adecuados.

45 Para los subcanales directos, el recinto de altavoces contendría controladores en los que el eje medio del controlador bisecta el "punto ideal", o el centro acústico del entorno de escucha. Los controladores de proyección hacia arriba estarían situados de modo que el ángulo entre el plano medio del controlador y el centro acústico sería a un ángulo en el margen de 45 a 180 grados. En el caso de posicionamiento del controlador a 180 grados, el controlador de proyección hacia atrás podría proporcionar difusión acústica mediante la reflexión desde una pared posterior. Esta configuración utiliza el principal acústico que después de la alineación temporal de los controladores con proyección hacia arriba con los controladores directos, las componentes de señal de llegada pronta sería coherente mientras que los componentes de llegada tardía se beneficiarían de la difusión natural proporcionada por  
50 el entorno de escucha.

55 Con el fin de conseguir las pistas de altura proporcionadas por el sistema de audio adaptativo, los controladores de proyección hacia arriba podrían angularse hacia arriba desde el plano horizontal y en el extremo, podrían posicionarse para radiar en sentido directo y reflejar desde una o más superficies reflectoras tales como un falso techo, o un difusor acústico situado inmediatamente por encima del recinto. Para proporcionar una direccionalidad adicional, el altavoz central podría utilizar una configuración de barra de sonido (tal como la ilustrada en la Figura 7C) con la capacidad para dirigir el sonido a través de la pantalla para proporcionar un canal central de alta resolución.

60 La configuración 5.1 de la Figura 9A podría expandirse añadiendo dos recintos posteriores adicionales similares a una configuración estándar 7.1. La Figura 9B ilustra una configuración de altavoces para un sistema de audio adaptativo 7.1 que utiliza múltiples controladores direccionables para el audio reflejado, en conformidad con dicha forma de realización. Según se ilustra en la configuración 920, los dos recintos adicionales 922 y 924 están colocados en las posiciones de "envolvente del lado izquierdo" y "envolvente del lado derecho" con los altavoces laterales apuntando hacia las paredes laterales en forma similar a los recintos frontales y los controladores de proyección hacia arriba establecidos para el rebote desde el techo a medio camino entre los pares frontales y  
65

posteriores existentes. Dichas adiciones incrementales pueden realizarse tantas veces como sea deseable, con los pares adicionales relleno los espacios vacíos a lo largo de las paredes laterales o posteriores. Las Figuras 9A y 9B ilustran solamente algunos ejemplos de posibles configuraciones de disposiciones de altavoces de sonido envolvente extendidas que pueden utilizarse en conjunción con altavoces de proyección hacia arriba o lateral en un sistema de audio adaptativo para entornos de escucha, y muchas otras configuraciones son también posibles.

Como una alternativa a las n.1 configuraciones anteriormente descritas, se puede utilizar un sistema basado en un soporte más flexible en donde cada controlador está contenido dentro de su propio recinto, que podría montarse luego en cualquier posición conveniente. Esta disposición utilizaría una configuración de controladores tal como se ilustra en la Figura 7B. Estas unidades individuales pueden agruparse luego en una manera similar a las n.1 configuraciones, o podrían dispersarse individualmente alrededor del entorno de escucha. Los soportes no están necesariamente restringidos a colocarse en los bordes del entorno de escucha, sino que podrían colocarse también en cualquier superficie en su interior (p.ej., mesa de café, estante de librería, etc.). Dicho sistema sería fácil de expandir, permitiendo al usuario añadir más altavoces en el transcurso del tiempo para crear una experiencia más inmersiva. Si los altavoces son inalámbricos, entonces el sistema de soporte podría incluir la capacidad para establecer altavoces para fines de recarga. En este diseño, los soportes podrían establecerse juntos de modo que actuaran como un altavoz único mientras se recargan, quizás para la escucha de una música estéreo, y luego, desbloqueados y situados alrededor del entorno de escucha para contenido de audio adaptativo.

Con el fin de mejorar la configurabilidad y exactitud del sistema de audio adaptativo utilizando controladores direccionables de proyección hacia arriba, varios sensores y dispositivos de realimentación podrían añadirse a los recintos para informar al presentador sobre las características que podrían utilizarse en el algoritmo de presentación. A modo de ejemplo, un micrófono instalado en el recinto permitiría al sistema medir las características de fase, frecuencia y reverberación del entorno de escucha, junto con la posición de los altavoces en relación mutua con cada uno utilizando una triangulación y las funciones de HRTF (Transferencia relacionada con la cabeza) de los propios recintos. Sensores inerciales (p.ej., giroscopios, brújulas, etc.) podrían utilizarse para detectar la dirección y el ángulo de los recintos; y sensores ópticos y visuales (p.ej., utilizando un goniómetro de infrarrojos basado en láser) podrían utilizarse para proporcionar información posicional relativa al propio entorno de escucha. Lo que antecede representa solamente unas pocas posibilidades de sensores adicionales que podrían utilizarse en el sistema, siendo también posible la incorporación de otros.

Dichos sistemas sensores pueden ampliarse todavía más permitiendo la posición de los controladores y/o los modificadores acústicos de los recintos para ser automáticamente ajustables mediante servomecanismos electromecánicos. Lo que antecede permitiría el cambio de la direccionalidad de los controladores en el tiempo de ejecución para adaptar su posicionamiento en el entorno de escucha en relación con las paredes y otros controladores ("dirección activa"). De modo similar, cualesquiera modificadores acústicos (tales como deflectores, bocinas o guías de ondas) podrían sintonizarse para proporcionar las respuesta de frecuencia y de fase correctas para una reproducción óptima en cualquier configuración del entorno de escucha ("sintonización activa"). Ambas funciones de dirección activa y sintonía activa podrían realizarse durante la configuración del entorno de escucha inicial (p.ej., en conjunción con el sistema de configuración de sala automática/EQ automática) o durante la reproducción en respuesta al contenido que se presenta.

#### Interconexión bidireccional

Una vez configurados, los altavoces deben conectarse al sistema de presentación. Las interconexiones tradicionales suelen ser de dos tipos: entrada al nivel de altavoz para altavoces pasivos y entrada a nivel de línea para altavoces activos. Según se ilustra en la Figura 4C, el sistema de audio adaptativo 450 incluye una función de interconexión bidireccional. Esta interconexión se materializa dentro de un conjunto de conexiones físicas y lógicas entre la etapa de presentación 454 y el amplificador/altavoz 458 y las etapas microfónicas 460. La capacidad para dirigir múltiples controladores en cada armario de altavoces se soporta por estas interconexiones inteligentes entre la fuente sonora y el altavoz. La interconexión bidireccional permite la transmisión de señales desde la fuente sonora (presentador) al altavoz con la inclusión de las señales de control y de las señales de audio. La señal desde el altavoz a la fuente del sonido consiste en señales de control y señales de audio, en donde las señales de audio, en este caso, tienen un origen de audio desde los micrófonos incorporados opcionales. La potencia puede proporcionarse también como para de la interconexión bidireccional, al menos para el caso en donde no se alimentan por separado los altavoces/controladores.

La Figura 10 es un diagrama 1000 que ilustra la composición de una interconexión bidireccional en conformidad con una forma de realización. La fuente sonora 1002, que puede representar un dispositivo presentador más una cadena de procesadores de sonido/amplificadores, está lógicamente y físicamente acoplada al armario de altavoces 1004 por intermedio de un par de enlaces de interconexión 1006 y 1008. La interconexión 1006 desde la fuente sonora 1002 a los controladores 1005 dentro del armario de altavoces 1004 comprende una señal electroacústica para cada controlador, una o más señales de control y una potencia opcional. La interconexión 1008 desde el armario de altavoces 1004 de nuevo hacia la fuente del sonido 1002 comprende señales sonoras procedentes del micrófono 1007 u otros sensores para calibración del presentador u otra funcionalidad de procesamiento de sonido similar. La interconexión de realimentación 1008 contiene también algunas definiciones y parámetros de controladores que se

utilizan por el presentador para modificar o procesar las señales sonoras establecidas para los controladores sobre la interconexión 1006.

5 En una forma de realización, a cada controlador en cada uno de los armarios del sistema se le asigna un  
 10 identificador (p.ej., una asignación numérica) durante la configuración del sistema. Cada armario de altavoces  
 (recinto) puede ser también identificado de forma única. Esta asignación numérica se utiliza por el armario de  
 altavoces para determinar qué señal de audio se envía a qué controlador dentro del armario. La asignación se  
 memoriza en el armario de altavoces en un dispositivo de memoria adecuado. Como alternativa, cada controlador  
 puede configurarse para memorizar su propio identificador en la memoria local. En otra alternativa, tal como una en  
 15 la que los controladores/altavoces no tengan ninguna capacidad de memorización local, los identificadores pueden  
 memorizarse en la etapa de presentación u otro componente dentro de la fuente del sonido 1002. Durante un  
 proceso de descubrimiento de altavoces, cada altavoz (o una base de datos central) se consulta por la fuente de  
 sonido para conocer su perfil. El perfil define algunas definiciones de controladores incluyendo el número de  
 controladores en un armario de altavoces u otra red definida, las características acústicas de cada controlador (p.ej.,  
 20 tipo de controlador, respuesta de frecuencia, etc.), la posición x,y,z del centro de cada controlador en relación con el  
 centro de la cara frontal del armario de altavoces, el ángulo de cada controlador con respecto a un plano definido  
 (p.ej., techo, suelo, ejes verticales del armario, etc.) y el número de micrófonos y características microfónicas. Otro  
 controlador pertinente y parámetros de micrófono/sensor pueden también definirse a este respecto. En una forma de  
 realización, las definiciones de controladores y el perfil del armario de altavoces puede expresarse como uno o más  
 documentos XML utilizados por el presentador.

En una posible puesta en práctica, se crea una red de control de Protocolo Internet (IP) entre la fuente de sonido  
 1002 y el armario de altavoces 1004. Cada armario de altavoces y fuente del sonido actúa como un punto final de  
 red único y se les proporciona una dirección local-enlace en la inicialización o activación inicial. Un mecanismo de  
 autodescubrimiento tal como una red de configuración cero (zeroconf) puede utilizarse para permitir a la fuente de  
 25 sonido localizar cada altavoz en la red. Las redes de configuración de ceros son un ejemplo de un proceso que crea  
 automáticamente una red IP utilizable sin intervención manual del operador o servidores de configuración  
 especiales, y se pueden utilizar otras técnicas similares. Dado un sistema de red inteligente, múltiples fuentes  
 pueden residir en la red IP como los altavoces. Esto permite que múltiples fuentes exciten directamente los  
 30 altavoces sin encaminar el sonido a través de una fuente de audio “maestra” (p.ej., receptor A/V tradicional). Si otra  
 fuente intenta direccionar los altavoces, se realizan comunicaciones entre todas las fuentes para determinar qué  
 fuente está actualmente “activa”, si el hecho de estar activa es necesario y si el control puede ser objeto de  
 transición a una nueva fuente de sonido. A las fuentes se les puede asignar una prioridad durante la fabricación  
 sobre la base de su clasificación, a modo de ejemplo, una fuente de telecomunicaciones puede tener una más alta  
 35 prioridad que una fuente de actividades recreativas. En un entorno de múltiples espacios, tal como un entorno  
 residencial típico, todos los altavoces dentro del entorno global pueden residir en una red única, pero pueden no  
 necesitar direccionarse de forma simultánea. Durante el establecimiento y configuración automática, el nivel acústico  
 proporcionado sobre la interconexión 1008 puede utilizarse para determinar qué altavoces están situados en el  
 mismo espacio físico. Una vez que se determine esta información, los altavoces pueden agruparse en  
 40 agrupamientos denominados clúster. En este caso, los identificadores IDs de clúster pueden asignarse y formar  
 parte de las definiciones del controlador. El ID de clúster se envía a cada altavoz, y cada clúster puede direccionarse  
 simultáneamente por la fuente de sonido 1002.

Según se ilustra en la Figura 10, una señal de potencia opcional puede transmitirse a través de la interconexión  
 45 bidireccional. Los altavoces pueden ser pasivos (que necesitan alimentación externa desde la fuente del sonido) o  
 activos (que requieren alimentación desde una toma de energía eléctrica). Si el sistema de altavoces consiste en  
 altavoces activos sin soporte inalámbrico, la entrada al altavoz consiste en una entrada Ethernet cableada en  
 cumplimiento con la norma IEEE 802.3. Si el sistema de altavoces consiste en altavoces activos con soporte  
 50 inalámbrico, la entrada al altavoz consiste en una entrada Ethernet inalámbrica en cumplimiento con la norma IEEE  
 802.11 o como alternativa, un estándar inalámbrico especificado por la organización de WISA. Altavoces pasivos  
 pueden proporcionarse por señales de alimentación adecuadas proporcionadas directamente por la fuente del  
 sonido.

#### 55 Configuración y calibración del sistema

Según se ilustra en la Figura 4C, la funcionalidad del sistema de audio adaptativo incluye una función de calibración  
 462. Esta función está habilitada por el micrófono 1007 y los enlaces de interconexión 1008 ilustrados en la Figura  
 10. La función del componente microfónico en el sistema 1000 es medir la respuesta de los controladores  
 individuales en el entorno de escucha con el fin de derivar una respuesta del sistema global. Múltiples topologías  
 60 microfónicas pueden utilizarse para esta finalidad, incluyendo un micrófono único o una red de micrófonos. El caso  
 más simple es donde un micrófono de medición omnidireccional único, situado en el centro del entorno de escucha,  
 se utiliza para medir la respuesta de cada controlador. Si el entorno de escucha y las condiciones de reproducción  
 garantizan un análisis más refinado, pueden utilizarse, en su lugar, múltiples micrófonos. La localización más  
 adecuada para múltiples micrófonos es dentro de los armarios de altavoces físicos de la configuración de altavoces  
 65 particulares que se utiliza en el entorno de escucha. Los micrófonos instalados en cada recinto permiten al sistema  
 medir la respuesta de cada controlador, en múltiples posiciones en un entorno de escucha. Una alternativa a esta



topología es utilizar múltiples micrófonos de medición omnidireccional situados en localizaciones más probables dentro del entorno de escucha.

Los micrófonos se utilizan para permitir la configuración automática y la calibración de los algoritmos de post-procesamiento y del presentador. En el sistema de audio adaptativo, el presentador es responsable de convertir un objeto híbrido y un flujo de audio basado en canal en señales de audio individuales designadas para controladores direccionables específicos, dentro de uno o más altavoces individuales. El componente de post-procesamiento puede incluir: funcionalidades de retardo, ecualización, ganancia, virtualización de altavoces y mezcla ascendente. La configuración de los altavoces representa una información frecuentemente crítica que el componente del presentador puede utilizar para convertir un objeto híbrido y flujo de audio basado en canal en señales de audio individuales por controlador para proporcionar una reproducción óptima del contenido de audio. La información de configuración del sistema incluye: (1) el número de altavoces físicos en el sistema, (2) el número de controladores individualmente direccionables en cada altavoz, y (3) la posición y la dirección de cada controlador individualmente direccionable, en relación con la geometría del entorno de escucha. Otras características son también posibles. La Figura 11 ilustra la función de una configuración automática y de un componente de calibración del sistema, en conformidad con una forma de realización. Según se ilustra en el diagrama 1100, una red matricial 1102 de uno o más micrófonos proporciona información acústica al componente de configuración y de calibración 1104. Esta información acústica captura algunas características pertinentes del entorno de escucha. El componente de configuración y de calibración 1104 proporciona, entonces, esta información al dispositivo presentador 1106 y cualesquiera componentes de post-procesamiento pertinentes 1108 de modo que las señales de audio que se envían, en última instancia, a los altavoces sean ajustadas y optimizadas para el entorno de escucha.

El número de altavoces físicos en el sistema y el número de controladores individualmente direccionables en cada altavoz son las propiedades de los altavoces físicos. Estas propiedades se transmiten directamente desde los altavoces a través de la interconexión bidireccional 456 al presentador 454. El presentador y los altavoces utilizan un protocolo de descubrimiento común, de modo que cuando altavoces están conectados o desconectados del sistema, la presentación es notificada del cambio y puede reconfigurar el sistema en consecuencia.

La geometría (tamaño y forma) del entorno de escucha es un elemento necesario de información en el proceso de configuración y de calibración. La geometría puede determinarse en varias maneras diferentes. En un módulo de configuración manual, la anchura, la longitud y la altura del cubo de aglutinación mínima para el entorno de escucha se introducen en el sistema por el oyente o un técnico a través de una interfaz de usuario que proporciona entrada al presentador u otra unidad de procesamiento dentro del sistema de audio adaptativo. Varias diferentes técnicas y herramientas de interfaz de usuario pueden utilizarse para esta finalidad. A modo de ejemplo, la geometría del entorno de escucha puede enviarse al presentador mediante que un programa que efectúa un mapeado o trazos automáticos de la geometría del entorno de escucha. Dicho sistema puede utilizar una combinación de visión por ordenador, sonar y mapeado físico basado en láser 3D.

El presentador utiliza la posición de los altavoces dentro de la geometría del entorno de escucha para derivar las señales de audio para cada controlador individualmente direccionable, incluyendo los controladores directos y reflejados (proyección hacia arriba). Los controladores directos son los que tienen como objetivo que la mayoría de su modelo de dispersión intersecte la posición de escucha antes de difundirse por una o más superficies reflectantes (tal como suelo, pared o techo). Los controladores reflejados son los que tienen como objetivo que la mayoría de sus modelos de dispersión sean reflejados antes de intersectar la posición de escucha según se ilustra en la Figura 6. Si un sistema está en un modo de configuración manual, las coordenadas 3D para cada controlador director puede introducirse en el sistema por intermedio de una interfaz UI. Para los controladores reflejados, las coordenadas 3D de la reflexión primaria se introducen en la UI. Láseres o técnicas similares pueden utilizarse para visualizar el modelo de dispersión de los controladores difundidos sobre las superficies del entorno de escucha, de modo que las coordenadas 3D puedan medirse e introducirse manualmente en el sistema.

La posición del controlador y su apuntamiento orientativo se suele realizar utilizando técnicas manuales o automáticas. En algunos casos, sensores inerciales pueden incorporarse en cada altavoz. En este modo, el altavoz central es designado como el altavoz "maestro" y su medición de brújula se considera como la referencia. Los otros altavoces pueden transmitir los modelos de dispersión y las posiciones de brújulas para cada uno de sus controladores individualmente direccionables. Acoplada con la geometría del entorno de escucha, entre el ángulo de referencia del altavoz central y cada controlador adicional proporciona información suficiente para que el sistema pueda determinar automáticamente si un controlador es directo o reflejado.

La configuración de posiciones de altavoces puede ser completamente automatizada si se utiliza un micrófono posicional 3D (esto es, la técnica Ambisonic). En este modo, el sistema envía una señal de prueba a cada controlador y registra la respuesta. Dependiendo del tipo de micrófono, las señales pueden necesitar transformarse en una representación de x, y, z. Estas señales se analizan para encontrar las componentes de x, y, y z de la primera llegada dominante. Acoplada con la geometría del entorno de escucha, suele proporcionar información suficiente para el sistema para establecer automáticamente las coordenadas 3D para todas las posiciones de altavoces, directas o reflejadas. Dependiendo de la geometría del entorno de escucha, una combinación híbrida de los tres modos descritos para configurar las coordenadas de los altavoces puede ser más efectiva que utilizar

solamente una técnica.

La información de configuración de altavoces es una componente requerida para configurar el presentador. La información de calibración de altavoces es también necesaria para configurar la cadena de post-procesamiento: retardo, ecualización y ganancia. La Figura 12 es un diagrama de flujo que ilustra las etapas del proceso de realización de la calibración automática de altavoces utilizando un micrófono único, en conformidad con una forma de realización. En este modo, el retardo, la ecualización y la ganancia se calculan automáticamente por el sistema utilizando un micrófono de multidifusión omnidireccional único situado en la parte media de la posición de escucha. Según se ilustra en el diagrama 1200, el proceso se inicia midiendo la respuesta de impulsos del espacio para cada controlador único solo, en el bloque 1202. El retardo para cada controlador se calcula luego con la averiguación de la compensación del valor máximo de la correlación cruzada de la respuesta de impulsos acústica (capturada con el micrófono) con respuesta de impulsos eléctricos directamente capturados, bloque 1204. En el bloque 1206, el retardo calculado se aplica a la respuesta de impulsos de captura directa (referencia). El proceso determina, entonces, los valores de banda ancha y de ganancia por banda que, cuando se aplican a una respuesta de impulsos medida, dan lugar a una diferencia mínima entre ella y la respuesta de impulsos de captura directa (referencia), bloque 1208. Lo que antecede puede realizarse tomando la FFT con función de ventana de la respuesta de impulsos de referencia y medidos, calculando las relaciones de magnitud por contenedor entre las dos señales, aplicando un filtro de la mediana a las relaciones de magnitud por contenedor, calculando los valores de ganancia por banda promediando la ganancia para la totalidad de los contenedores que caen completamente dentro de una banda, calcular una ganancia de banda base tomando la media de todas las ganancias por banda, restando la ganancia de banda ancha desde las ganancias por banda y aplicando la curva X de pequeño espacio (-2 dB/octava por encima de 2 kHz). Una vez que se determinen los valores de la ganancia en el bloque 1208, el proceso determina los valores del retardo finales restando el retardo mínimo de los demás, de modo que al menos un controlador en el sistema tendrá siempre un retardo adicional nulo, bloque 1210.

En el caso de calibración automática utilizando múltiples micrófonos, el retardo, la ecualización y la ganancia se calculan automáticamente por el sistema utilizando múltiples micrófonos de medición omnidireccionales. El proceso es prácticamente idéntico a la técnica de micrófono único, con la excepción de que se repite para cada uno de los micrófonos, y los resultados son promediados.

#### Aplicaciones alternativas

En lugar de poner en práctica un sistema de audio adaptativo en un entorno de escucha completo o sala, es posible poner en práctica aspectos del sistema de audio adaptativo en aplicaciones más localizadas, tales como televisión, ordenadores, consolas de juego o dispositivos similares. Este caso se basa efectivamente en altavoces que están dispuestos de forma matricial en un plano sin relieve que corresponde a la pantalla de visión o superficie del monitor. La Figura 13 ilustra el uso de un sistema de audio adaptativo en un caso de uso de televisión y barra de sonido, a modo de ejemplo. En general, el caso de uso de televisión proporciona retos operativos para crear una experiencia de audio inmersiva sobre la base de la calidad frecuentemente reducida de los equipos (altavoces de TV, barra de sonido, altavoces, etc.) y las localizaciones/configuraciones de altavoces, que pueden limitarse en términos de resolución espacial (es decir, sin altavoces posteriores o circundantes). El sistema 1300 de la Figura 13 incluye altavoces en las localizaciones izquierda y derecha de la televisión estándar (TV-L y TV-R) así como controladores de proyección hacia la izquierda, hacia la derecha o hacia arriba (TV-LH y TV-RH). La televisión 1302 puede incluir también una barra de sonido 1304 o altavoces en alguna clase de red de altura. En general, la magnitud y la calidad de los altavoces de televisión se reducen debido a limitaciones de coste y opciones de diseño en comparación con los altavoces autónomos o de entorno residencial. El uso de la denominada virtualización dinámica, sin embargo, puede ayudar a superar estas deficiencias. En la Figura 13, el efecto de virtualización dinámica se ilustra para los altavoces TV-L y TV-R, de modo que las personas en una posición de escucha específica 1308 oirían elementos horizontales asociados con objetos de audio adecuados individualmente presentados en el plano horizontal. Además, los elementos de altura asociados con objetos de audio adecuados serán presentados correctamente por intermedio del audio reflejado transmitido por los controladores LH y RH. El uso de la virtualización estéreo en los altavoces izquierdo y derecho de la televisión es similar a los altavoces de entorno residencial L y R en donde una experiencia de usuario de virtualización de altavoces dinámicos potencialmente inmersivos puede ser posible mediante el control dinámico de los parámetros de los algoritmos de virtualización de altavoces basados en la información espacial del objeto proporcionada por el contenido de audio adaptativo. Esta virtualización dinámica puede utilizarse para crear la percepción de objetos que se desplazan a lo largo de las partes laterales en el entorno de escucha.

El entorno de televisión puede incluir también un altavoz HRC según se ilustra dentro de la barra de sonido 1304. Dicho altavoz HRC puede ser una unidad direccionable que permite obtener panorámicas por intermedio de la red HRC. Lo que antecede puede ser beneficioso (en particular para grandes pantallas) al tener una red de canales centrales de proyección frontal con altavoces individualmente direccionables que permiten panorámicas discretas de objetos de audio mediante una red que establece la coincidencia del movimiento de los objetos de vídeo en la pantalla. Este altavoz se ilustra también como teniendo altavoces de proyección lateral. Estos podrían activarse y utilizarse si el altavoz se utiliza como una barra de sonido de modo que los controladores de proyección lateral proporcionen más inmersión debido a la falta de altavoces envolventes o posteriores. El concepto de virtualización

dinámica es también conocido para el altavoz de barra de sonido /HRC. La virtualización dinámica se ilustra para los altavoces L y R en los lados más alejados de la red de altavoces de proyección frontal. De nuevo, lo que antecede podría utilizarse para crear la percepción de objetos que se desplazan a lo largo de las partes laterales del entorno de escucha. Este altavoz central modificado podría incluir también más altavoces y poner en práctica un haz sonoro direccionable con zonas acústicas controladas por separado. También conocido en la puesta en práctica a modo de ejemplo de la Figura 13 es un altavoz NFE 1306 situado en frente de la posición de escucha principal 1308. La inclusión del altavoz NFE puede proporcionar una mayor envolvente por el sistema de audio adaptativo desplazando el sonido alejándose de la parte frontal del entorno de escucha y más próxima al oyente.

Con respecto a la presentación de auriculares, el sistema de audio adaptativo mantiene la intención original del creador haciendo coincidir los HRTFs con la posición espacial. Cuando se reproduce audio a través de los auriculares, una virtualización espacial binaural puede conseguirse por la aplicación de una denominada Función de Transferencia Relacionada con la Cabeza (HRTF), que procesa la señal de audio y añade pistas perceptuales que crean la percepción del audio que se reproduce en el espacio tridimensional y no sobre auriculares estéreo estándar. La exactitud de la reproducción espacial depende de la selección de la función HRTF adecuada que puede variar sobre la base de varios factores, incluyendo la posición espacial de los canales de audio u objetos que se presentan. La utilización de la información espacial proporcionada por el sistema de audio adaptativo puede dar lugar a la selección de uno o un número variable continuo de HRTFs que representan un espacio 3D para mejorar todavía más la experiencia de la reproducción.

El sistema facilita también la adición guiada, la presentación binaural tridimensional y la virtualización. De forma similar al caso de la presentación espacial, utilizando nuevos tipos y localizaciones de altavoces que sean nuevos y modificados, es posible mediante el uso de HRTFs tridimensionales crear pistas para simular el sonido de audio procedente del plano horizontal y del eje vertical. Formatos de audio anteriores que proporcionan solamente presentación de información de localización de altavoces fijas han sido más limitados. Con la información del formato de audio adaptativo, un sistema de auriculares de presentación binaural, tridimensional, tiene información detallada y de utilidad que puede utilizarse para dirigir qué elementos del audio son adecuados para su presentación en los planos horizontal y vertical. Algún contenido puede basarse en el uso de altavoces aéreos para proporcionar un mayor sentido de envolvente. Esos objetos de audio y la información podrían utilizarse para la presentación binaural que se percibe por encima de la cabeza del oyente cuando se utilizan auriculares. La Figura 14 ilustra una representación simplificada de una experiencia de virtualización de auriculares binaurales tridimensional para uso en un sistema de audio adaptativo, en conformidad con una forma de realización. Según se ilustra en la Figura 14, un conjunto de auriculares 1402 utilizado para reproducir audio desde un sistema de audio adaptativo incluye señales de audio 1404 en el plano x, y, estándar así como en el plano z de modo que la altura asociada con algunos objetos de audio o sonidos sea reproducida de modo que sonaría de forma similar a su origen por encima o por debajo de los sonidos originados en el sistema x, y.

#### Definiciones de metadatos

En una forma de realización, el sistema de audio adaptativo incluye componentes que generan metadatos a partir del formato de audio espacial original. Los métodos y componentes del sistema 300 comprenden un sistema de presentación de audio configurado para procesar uno o más flujos de bits que contienen elementos de audio basados en canales convencionales y elementos de codificación de objetos de audio. Una nueva capa de extensión que contiene los elementos de codificación de objetos de audio se define y añade a uno de entre el flujo de bits de códec de audio basados en canal o el flujo de bits de objetos de audio. Este método habilita a los flujos de bits, que incluyen la capa de extensión a procesarse por los presentadores para uso con diseños de controladores y altavoces existentes o con altavoces de la siguiente generación que utilizan controladores individualmente direccionables y definiciones de controladores. El contenido de audio espacial procedente del procesador de audio espacial comprende objetos de audio, canales y metadatos de posiciones. Cuando se presenta un objeto, se le asigna uno o más altavoces en función de los metadatos de posiciones y la localización de los altavoces de reproducción. Metadatos adicionales pueden asociarse con el objeto para modificar la localización de reproducción o limitar, de cualquier otro modo, los altavoces que han de utilizarse para la reproducción. Los metadatos se generan en la estación de trabajo de audio en respuesta a las entradas de mezcla del ingeniero de sonido para proporcionar colas de espera de la presentación que controlan los parámetros espaciales (p.ej., posición, velocidad, intensidad, timbre, etc.) y especifican qué controladores o altavoces, en el entorno de escucha, reproducen su sonido respectivo durante la exhibición. Los metadatos se asocian con los respectivos datos de audio en la estación de trabajo para el empaquetado y transporte por el procesador de audio espacial.

La Figura 15 es una tabla que ilustra algunas definiciones de metadatos para uso en un sistema de audio adaptativo para entornos de escucha en conformidad con una forma de realización. Según se ilustra en la tabla 1500, las definiciones de metadatos incluyen: tipo de contenido de audio, definiciones de controladores (número, características, posición, ángulo de proyección), señales de control para sintonización/dirección activa e información de calibración que incluye información del espacio de escucha y de los altavoces.

Características y capacidades

Según se indicó con anterioridad, el ecosistema de audio adaptativo permite al creador de contenidos incorporar la intención espacial de la mezcla (posición, magnitud, velocidad, etc.) dentro de los flujos de bits mediante metadatos. Esto permite una magnitud increíble de flexibilidad en la reproducción espacial de audio. Desde un punto de vista de la presentación espacial, el formato de audio adaptativo permite al creador de contenidos adaptar la mezcla a la posición exacta de los altavoces en el entorno de escucha para evitar una distorsión espacial causada por la geometría del sistema de reproducción que no es idéntico al sistema de autoría de edición. En sistemas de reproducción de audio actuales, en donde solamente se envía audio para un canal de altavoces, la intención del creador de contenidos es desconocida para las localizaciones en el entorno de escucha que no sean localizaciones de altavoces fijas. Bajo el paradigma de canales/altavoces actuales, la única información que se conoce es que un canal de audio específico debe enviarse a un altavoz específico que tiene una localización predefinida en un entorno de escucha. En el sistema de audio adaptativo, que utiliza metadatos transmitidos a través de los conductos de distribución y creación, el sistema de reproducción puede utilizar esta información para reproducir el contenido en una manera que coincida con la intención original del creador de contenidos. A modo de ejemplo, la relación entre altavoces es conocida para diferentes objetos de audio. Proporcionado la localización espacial para un objeto de audio, la intención del creador de contenidos es conocida y esta intención puede ser objeto de "mapeado" de correspondencia en la configuración de altavoces, incluyendo su localización. Con el sistema de presentación de audio de tipo dinámico, esta presentación puede actualizarse y mejorarse añadiendo altavoces adicionales.

El sistema permite también añadir una presentación espacial tridimensional guiada. Se han realizado numerosos intentos para crear una experiencia de presentación de audio más inmersiva mediante el uso de nuevos diseños y configuraciones de altavoces. Incluyen el uso de altavoces bipolares y dipolares, controladores de proyección lateral, proyección posterior y proyección hacia arriba. Con los sistemas de localización de altavoces fijos y de canal anteriores, la determinación de qué elementos de audio deben enviarse a estos altavoces modificados es relativamente difícil. Con el uso de un formato de audio adaptativo, un sistema de presentación tiene información útil y detallada de qué elementos del audio (objetos o cualquier otro) son adecuados para enviarse a nuevas configuraciones de altavoces. Es decir, el sistema permite el control sobre qué señales de audio se envían a los controladores de proyección frontal y qué señales se envían a los controladores de proyección hacia arriba. A modo de ejemplo, el contenido cinematográfico de audio adaptativo se basa principalmente en el uso de altavoces aéreos para proporcionar un mayor sentido de envolvente. Estos objetos de audio e información pueden enviarse a los controladores de proyección hacia arriba para proporcionar un audio reflejado en el entorno de escucha para crear un efecto similar.

El sistema permite también la adaptación de la mezcla a la configuración de hardware exacta del sistema de reproducción. Existen numerosos diferentes tipos de altavoces y configuraciones posibles en el equipo de presentación tales como televisiones, entornos residenciales, barras acústicas, sistemas de reproductores de música portátiles, etc. Cuando se envían estos sistemas, con la información de localización de audio específica del canal (esto es, canal izquierdo y derecho o audio multicanal estándar), el sistema debe procesar el audio para su adaptación adecuada a las capacidades del equipo de presentación. Un ejemplo típico es cuando un audio estéreo estándar (izquierdo/derecho) se envía a una barra de sonido, que tiene más de dos altavoces. En los sistemas de audio actuales en donde solamente se envía audio para un canal de altavoces, la intención del creador de contenidos es desconocida y una experiencia de audio más inmersiva hecha posible por el equipo mejorado debe crearse por algoritmos que hagan supuestos de cómo modificar el audio para su reproducción en el hardware. Un ejemplo de lo que antecede es el uso de PLII, PLII-z, o Envolvente de la Siguiete Generación para "mezclar" audio basado en canal para más altavoces que el número original de alimentaciones de canal. Con el sistema de audio adaptativo, utilizando metadatos transmitidos a través de los conductos de creación y distribución, un sistema de reproducción puede utilizar esta información para reproducir el contenido en una manera que coincida más estrechamente con la intención original del creador de contenidos. A modo de ejemplo, algunas barras acústicas tienen altavoces de proyección lateral para crear un sentido de envolvente. Con el audio adaptativo, la información espacial y la información del tipo de contenido (esto es, diálogo, música, efectos ambientales, etc.) pueden utilizarse por la barra de sonido cuando se controla por un sistema de presentación, tal como un receptor de TV o A/V para enviar solamente el audio adecuado a estos altavoces de proyección lateral.

La información espacial transmitida por el audio adaptativo permite la presentación dinámica del contenido con un conocimiento de la localización y tipo de los altavoces presentes. Además, la información sobre la relación de los oyentes o del oyente para los equipos de reproducción de audio está ahora potencialmente disponible y puede utilizarse en la presentación. La mayor parte de las consolas de juego incluyen un accesorio de cámara y procesamiento de imagen inteligente que puede determinar la posición e identidad de una persona en el entorno de escucha. Esta información puede utilizarse por un sistema de audio adaptativo para modificar la presentación para transmitir, con mayor exactitud, la intención creativa del creador de contenidos sobre la base de la posición del oyente. A modo de ejemplo, en casi todos los casos, el audio presentado para reproducción supone que el oyente está situado en un "punto ideal", que suele ser equidistante de cada altavoz y la misma posición en la que el mezclador de sonidos estaba situado durante la creación de contenidos. Sin embargo, en numerosas ocasiones, la persona no está en esta posición ideal y su experiencia no coincide con la intención creativa del mezclador. Un ejemplo típico es cuando un oyente está sentado en el lado izquierdo del entorno de escucha en una silla o sofá. Para este caso, el sonido que se reproduce desde los altavoces más cercanos a la izquierda serán percibidos como siendo más fuertes y sesgando la percepción espacial de la mezcla de audio a la izquierda. Mediante el

entendimiento de la posición del oyente el sistema podría ajustar la presentación del audio para hacer más bajo el nivel de sonido en los altavoces izquierdos y elevar el nivel de los altavoces derechos para reequilibrar la mezcla de audio y hacerla más perceptualmente correcta. Retardando el audio para compensar la distancia del oyente desde el llamado punto ideal, es también posible. La posición del oyente podría detectarse mediante el uso de una cámara o un control a distancia modificado con alguna señalización incorporada que señalaría la posición del oyente para el sistema de presentación.

Además de utilizar altavoces estándar y localizaciones de altavoces para direccionar la posición de escucha, es también posible utilizar tecnologías de direccionamiento de haces para crear campos sonoros “zonas” que varían dependiendo de la posición del oyente y de su contenido. La formación de haces de audio utiliza una red de altavoces (normalmente 8 a 16 altavoces horizontalmente espaciados) y utiliza una manipulación de fase y procesamiento para crear un haz acústico direccionable. La red de altavoces de formación de haces permite la creación de zonas de audio en donde el audio es principalmente audible que puede utilizarse para dirigir sonidos u objetos específicos con procesamiento selectivo hacia una localización espacial específica. Un caso de uso obvio es procesar el diálogo en una pista sonora utilizando un algoritmo de post-procesamiento de mejora del diálogo y un dirigir ese objeto de audio directamente a un usuario que está en una escucha impedida.

#### Codificación matricial y mezcla espacial

En algunos casos, los objetos de audio pueden ser un componente deseado de contenido de audio adaptativo; sin embargo, sobre la base de las limitaciones del ancho de banda, puede no ser posible enviar objetos de audio y audio de altavoces/canal. Anteriormente, se ha utilizado la codificación matricial para transmitir más información de audio que es posible para un sistema de distribución dado. A modo de ejemplo, éste era el caso en los días iniciales de la industria cinematográfica en donde fue creado audio de multicanales por los mezcladores de sonido pero los formatos de películas solamente proporcionaban audio estéreo. La codificación matricial fue utilizada para la mezcla inteligente del audio de multicanal para dos canales estéreo, que fueron luego procesados con algunos algoritmos para recrear una aproximación estrecha de la mezcla multicanales desde el audio estéreo. De modo similar, es posible una mezcla inteligente de objetos de audio en los canales de altavoces base y mediante el uso de metadatos de audio adaptativos y algoritmos envolventes de la siguiente generación sensibles a la frecuencia y al tiempo sofisticados para extraer los objetos y presentarlos de forma espacialmente correcta con un sistema de presentación de audio adaptativo.

Además, cuando existen limitaciones de ancho de banda del sistema de transmisión para el audio (aplicaciones inalámbricas 3G y 4G, a modo de ejemplo), existen también ventajas al transmitir lechos auditivos multicanales espacialmente diversos que se codifican matricialmente junto con los objetos de audio individuales. Un caso de uso de dicha metodología de transmisión sería para la transmisión de una difusión de eventos deportivos con dos lechos auditivos distintos y múltiples objetos de audio. Los lechos auditivos podrían representar el audio multicanal capturado en dos secciones de gradas de equipos diferentes y los objetos de audio podrían representar diferentes anunciadores que pueden tener simpatías por un equipo o el otro. La utilización de una codificación estándar de una representación 5.1 de cada lecho auditivo junto con dos o más objetos podría superar las limitaciones de ancho de banda del sistema de transmisión. En este caso, si cada uno de los lechos auditivos 5.1 fueran codificados matricialmente para una señal estéreo, en tal caso, los dos lechos que fueron originalmente capturados como canales 5.1 podrían transmitirse como lecho auditivo de dos canales 1, lecho auditivo de dos canales 2, objeto 1 y objeto 2 como solamente cuatro canales en lugar de los canales 5.1 + 5.1 + 2 o 12.1.

#### Procesamiento dependiente de la posición y del contenido

El ecosistema de audio adaptativo permite al creador de contenidos crear objetos de audio individuales y añadir información sobre el contenido que puede transmitirse al sistema de reproducción. Esto permite una gran cantidad de flexibilidad en el procesamiento de audio antes de la reproducción. El procesamiento puede adaptarse a la posición y tipo de objeto mediante un control dinámico de la virtualización de altavoces sobre la base de una posición y tamaño del objeto. La virtualización de altavoces se refiere a un método de procesamiento de audio de modo que un altavoz virtual sea percibido por un oyente. Este método se suele utilizar para la reproducción de altavoces estéreo cuando el audio origen es un audio multicanal que incluye alimentaciones de canales de altavoces envolventes. El procesamiento de altavoces virtuales modifica el audio de canal de altavoz envolvente de tal manera que cuando se reproduce en altavoces estéreo, los elementos de audio envolventes se virtualizan en la parte lateral y posterior del oyente como si estuviera un altavoz allí situado. Actualmente, los atributos de localización de la posición del altavoz virtual son estáticos porque la localización prevista de los altavoces envolventes era fija. Sin embargo, con el contenido de audio adaptativo, las localizaciones espaciales de objetos de audio diferentes son dinámicas y distintas (esto es, únicas para cada objeto). Es posible que el post-procesamiento tal como la virtualización de altavoces virtuales pueda controlarse ahora en una manera más informada controlando dinámicamente los parámetros tales como el ángulo posicional del altavoz para cada objeto y combinando luego las salidas presentadas de varios objetos virtualizados para crear una experiencia de audio más inmersiva que represente más estrecha la intención del mezclador de sonidos.

Además de la virtualización horizontal estándar de objetos de audio, es posible utilizar pistas de altura perceptuales

que procesan audio de objetos dinámicos y de canal fijo y obtienen la percepción de la reproducción de altura de audio a partir de un par estándar de altavoces estéreo en la localización normal del plano horizontal.

5 Algunos efectos o procesos de mejora pueden ser conscientemente aplicados a tipos adecuados de contenidos de audio. A modo de ejemplo, una mejora del diálogo puede aplicarse a objetos de diálogos solamente. La mejora del diálogo se refiere a un método de procesamiento de audio que contiene diálogo de modo que se aumente y/o mejore la audibilidad y/o inteligibilidad del diálogo. En numerosos casos, el procesamiento de audio que se aplica al diálogo es inadecuado para el contenido de audio no de diálogo (p.ej., música, efectos ambientales, etc.) y pueden dar lugar a un artefacto audible no deseable. Con el audio adaptativo, un objeto de audio podría contener solamente el diálogo en un elemento de contenido y puede etiquetarse en consecuencia de modo que una solución de presentación aplicará, de forma selectiva, la mejora del diálogo a solamente el contenido de diálogo. Además, si el objeto de audio es solamente diálogo (y no una mezcla de diálogo y otro contenido, lo que suele ser el caso), entonces, el procesamiento de la mejora del diálogo puede procesar el diálogo exclusivamente (con lo que se limita que se realice cualquier procesamiento sobre cualquier otro contenido).

15 De modo similar, la gestión de respuesta de audio o de ecualización puede personalizarse también para características de audio específicas. A modo de ejemplo, la gestión de bajos (filtrado, atenuación, ganancia) orientada a un objeto específico basado en su tipo. La gestión de bajos se refiere al aislamiento y procesamiento selectivo de solamente la frecuencia de bajos (o inferiores) en un elemento de contenido particular. Con los sistemas de audio actuales y los mecanismos de entrega, éste es un proceso “a ciegas” que se aplica a la totalidad del audio. Con un audio adaptativo, los objetos de audio específicos en los que la gestión de bajos es adecuada pueden identificarse por metadatos y aplicar secuencialmente el procesamiento de presentación.

25 El sistema de audio adaptativo facilita también la compresión del margen dinámico basado en el objeto. Las pistas de audio tradicionales tienen la misma duración que su propio contenido, mientras que un objeto de audio pudiera producirse durante una cantidad de tiempo limitada en el contenido. Los metadatos asociados con un objeto pueden contener información relacionada con el nivel sobre su amplitud de señal media y máxima, así como su tiempo de inicio o ataque (en particular, para material transitorio). Esta información permitirá a un compresor adaptar mejor su compresión y constante de tiempo (ataque, liberación, etc.) para adaptarse mejor al contenido.

30 El sistema facilita también la ecualización automática de la sala de altavoces. La acústica del entorno de escucha y altavoces desempeña un importante papel en la introducción de coloración audible al sonido con el consiguiente impacto sobre el timbre del sonido reproducido. Además, las acústicas son dependientes de la posición debido a las reflexiones del entorno de escucha y a las variaciones de la directividad de los altavoces y debido a esta variación, el timbre percibido variará notablemente para diferentes posiciones de escucha. Una función de AutoEQ (ecualización automática de salas), proporcionada en el sistema, ayuda a mitigar algunas de estas anomalías mediante una medición automática espectral de la sala de altavoces y la ecualización, compensación del retardo automatizada (que proporciona una creación de imágenes adecuada y posiblemente una detección de localización de altavoces relativa basada en mínimos cuadráticos) y ajustes del nivel, con la redirección de bajos basada en la capacidad del espacio de los altavoces, así como la segmentación óptima de los altavoces principales con los altavoces de graves o subwoofers. En una sala de entorno residencial u otro entorno de escucha, el sistema de audio adaptativo incluye algunas funciones adicionales, tales como: (1) cálculo de curva objetivo automatizado basado en la acústica de la sala de reproducción (que se considera un problema abierto en la investigación para la ecualización en los entornos de escucha residenciales), (2) la influencia del control del decaimiento modal utilizando un análisis de tiempo-frecuencia, (3) el entendimiento de los parámetros derivados de las mediciones que rigen las envolventes/espacios/anchura de fuentes/inteligibilidad y control de estos parámetros para proporcionar la mejor experiencia de escucha posible, (4) filtrado direccional que incorpora modelos de auriculares para la adaptación del timbre entre los altavoces frontales y “otros” y (5) la detección de posiciones espaciales de los altavoces en una configuración discreta en relación con el oyente y el re-mapeado espacial (p.ej., un denominado Summit inalámbrico sería un ejemplo). La desadaptación en el timbre entre altavoces es especialmente revelada en algún contenido de panorámicas entre un altavoz de anclaje frontal (p.ej., central) y los altavoces de sonido envolvente/posteriores/de anchura/de altura.

55 En un sentido global, el sistema de audio adaptativo permite también una experiencia de reproducción de audio/vídeo muy atractiva, en particular, con grandes tamaños de pantallas en un entorno residencial, si la localización espacial reproducida de algunos elementos de audio coincide con los elementos de imagen en la pantalla. Un ejemplo es tener el diálogo en un programa cinematográfico o de televisión que coincida espacialmente con una persona o personaje que esté hablando en la pantalla. Con el audio basado en canal de altavoces normal, no existe un método fácil para determinar en dónde el diálogo debe situarse espacialmente para coincidir con la localización de la persona o personaje en la pantalla. Con la información de audio disponible en un sistema de audio adaptativo, este tipo de alineación de audio/visual podría lograrse fácilmente, incluso en sistemas de entorno residencial que se caracterizan por pantallas de tamaño mayor. La alineación posicional visual y espacial de audio podría utilizarse también para objetos de diálogo/no personaje tales como vehículos, camiones, animación, etc.

65 El ecosistema de audio adaptativo permite también la mejora de la gestión de los contenidos, permitiendo a un creador de contenidos crear objetos de audio individuales y añadir información sobre el contenido que puede

transmitirse al sistema de reproducción. Lo que antecede permite una mayor flexibilidad en la gestión de contenidos de audio. Desde un punto de vista de gestión de contenidos, el audio adaptativo permite varias acciones tales como cambiar el idioma del contenido de audio solamente sustituyendo un objeto de diálogo para reducir el tamaño del fichero de contenidos y/o reducir el tiempo de descarga. Los programas cinematográficos, televisivos y otros programas de actividades recreativas suelen distribuirse a un nivel internacional. Esta circunstancia suele requerir que el idioma en el elemento de contenido cambie dependiendo de dónde será reproducido (francés para películas que se muestren Francia, alemán para programas de TV que se muestren en Alemania, etc.). Actualmente, esta circunstancia suele requerir la creación de una pista sonora de audio completamente independiente, empaquetada y distribuida para cada idioma. Con el sistema de audio adaptativo, y el concepto inherente de objeto de audio, el diálogo para un elemento de contenido podría ser un objeto de audio independiente. Esto permite que el idioma del contenido sea fácilmente modificado sin necesidad de actualizar o modificar otros elementos de la pista sonora de audio tal como música, efectos, etc. Esto no se aplicaría solamente a idiomas extranjeros sino también a un lenguaje inadecuado para determinada audiencia, publicidad dirigida, etc.

Aspectos del entorno de audio aquí descritos representan la reproducción del contenido de audio o audiovisual a través de altavoces adecuados y dispositivos de reproducción pertinentes y puede representar cualquier entorno en el que un oyente esté experimentando la reproducción del contenido capturado, tal como sala cinematográfica, sala de conciertos, teatro de exteriores, un entorno residencial, cabinas de escucha, vehículos, consola de juegos, sistema de auriculares o telefónica, sistema de megafonía (PA) o cualquier otro entorno de reproducción. Aunque las formas de realización han sido descritas principalmente con respecto a ejemplos y puestas en práctica en un entorno residencial, en el que el contenido de audio espacial está asociado con el contenido de televisión, conviene señalar que las formas de realización podrían ponerse en práctica también en otros sistemas. El contenido de audio espacial, que comprende audio basado en el objeto y audio basado en el canal, puede utilizarse en conjunción con cualquier contenido relacionado (audio, vídeo, gráficos, etc. asociados) o puede constituir un contenido de audio autónomo. El entorno de reproducción puede ser cualquier entorno de escucha adecuado desde auriculares o monitores de campos cercanos para salas pequeñas o grandes, vehículos instalaciones al aire libre, salas de concierto, etc.

Aspectos de los sistemas aquí descritos pueden ponerse en práctica en un entorno de red de procesamiento de sonido basado en ordenador adecuado para procesar ficheros de audio digitales o digitalizados. Partes del sistema de audio adaptativo pueden incluir una o más redes que comprenden cualquier número deseado de máquinas individuales, incluyendo uno o más controladores (no ilustrados) que sirven para memorizar y enrutar los datos transmitidos entre los ordenadores. Dicha red puede construirse sobre varios protocolos de red diferentes y puede ser Internet, una Red de Área Amplia (WAN), una Red de Área Local (LAN), o cualquiera de sus combinaciones. En una forma de realización en la que la red comprenda Internet, una o más máquinas pueden configurarse para acceder a Internet a través de los programas del explorador de la red.

Uno o más de los componentes, bloques, procesos u otros componentes funcionales pueden ponerse en práctica mediante un programa informático que controle la ejecución de un dispositivo informático basado en el procesador del sistema. Conviene señalar que las diversas funciones aquí dadas a conocer pueden describirse utilizando cualquier número de combinaciones de hardware, firmware y/o datos y/o instrucciones que se materializan en varios soportes legibles por ordenador o legibles por máquina, en términos de su comportamiento operativo, transferencia de registro, componente lógico y/o otras características. Los soportes legibles por ordenador en los que puede materializarse dichos datos y/o instrucciones formateadas incluyen, sin limitación, soportes físicos (no transitorios), soportes de almacenamiento no volátil en varias formas tales como soportes de memorización óptico, magnético o de semiconductores.

A no ser que el contexto lo requiera claramente de otro modo, mediante la descripción y las reivindicaciones, los términos “comprende”, “comprendiendo” y similares han de interpretarse en un sentido inclusivo a diferencia de un sentido exclusivo o exhaustivo; es decir, en un sentido de “incluir, sin limitación”. Los términos que utilizan singular o plural también incluyen el plural o el singular respectivamente. Además, los términos “aquí”, “en el presente”, “anterior”, “siguiente” y términos de significado similar se refieren a esta solicitud como un conjunto y no a cualesquiera partes particulares de la misma. Cuando la palabra “o” se utiliza en referencia a una lista de dos o más elementos, dicha palabra cubre la totalidad de las interpretaciones siguientes de la palabra: cualquiera de los elementos en la lista, todos los elementos en la lista y cualquier combinación de los elementos en la lista.

Aunque una o más puestas en práctica han sido descritas a modo de ejemplo y en términos de las formas de realización específicas, ha de entenderse que una o más puestas en práctica no están limitadas a las formas de realización dadas a conocer. Por el contrario, están previstas para cubrir varias modificaciones y disposiciones similares que podrían ser evidentes para los expertos en esta técnica. Por lo tanto, el alcance de las reivindicaciones adjuntas debe estar conforme con la más amplia interpretación con el fin de abarcar todas dichas modificaciones y disposiciones similares.

**REIVINDICACIONES**

- 5           1. Un sistema (500, 1300) que sirve para realizar una presentación del sonido utilizando elementos sonoros reflejados, que comprende:
- 10           una red de controladores de audio para distribución alrededor de un entorno de escucha (800), en donde al menos un controlador de la red de controladores de audio es un controlador de proyección hacia arriba (706), que está configurado para proyectar ondas sonoras hacia una o más superficies del entorno de escucha para reflexión hacia una zona de escucha dentro del entorno de escucha;
- 15           un dispositivo de presentación (310) configurado para recibir y procesar un flujo de bits que incluye flujos de audio y uno o más conjuntos de metadatos que están asociados con cada uno de los flujos de audio y que especifican un emplazamiento de reproducción en el entorno de escucha de un flujo de audio respectivo, en donde los flujos de audio comprenden uno o más flujos de audio reflejados y uno o más flujos de audio directos; y
- 20           un componente de reproducción acoplado al dispositivo de presentación y configurado para presentar los flujos de audio a una pluralidad de señales de alimentación de audio que corresponde a la red de controladores de audio en conformidad con los uno o más conjuntos de metadatos y en donde los uno o más flujos de audio reflejados se transmiten a al menos un controlador de proyección hacia arriba,
- 25           caracterizado por cuanto que el sistema realiza un procesamiento de señal para introducir pistas de altura de percepción en los uno o más flujos de audio reflejados transmitidos a al menos un controlador de proyección hacia arriba.
- 30           2. El sistema según la reivindicación 1, en donde cada controlador de audio de la red de controladores de audio es direccionable, de manera única, en conformidad con un protocolo de comunicación utilizado por el dispositivo de presentación y el componente de reproducción.
- 35           3. El sistema según la reivindicación 2, en donde el al menos un controlador de audio comprende uno de entre: un controlador de proyección lateral y un controlador de proyección hacia arriba y en donde el al menos un controlador de audio está materializado, además, en uno de entre: un controlador autónomo dentro de un recinto acústico de altavoces y un controlador colocado próximo a uno o más controladores de proyección frontal en un recinto acústico unitario.
- 40           4. El sistema según la reivindicación 3, en donde la red de controladores de audio comprende controladores que están distribuidos alrededor del entorno de escucha en conformidad con una configuración de sonido de ambiente envolvente definido.
- 45           5. El sistema según la reivindicación 4, en donde el entorno de escucha comprende un entorno residencial y en donde el dispositivo de presentación y el componente de reproducción comprende parte de un sistema de audio residencial y en donde, además, los flujos de audio comprenden un contenido de audio seleccionado de entre el grupo constituido por: un contenido cinematográfico transformado para reproducción en un entorno residencial, contenido de televisión, contenido generado por el usuario, contenido de juegos informáticos y música.
- 50           6. El sistema según la reivindicación 4, en donde un conjunto de metadatos asociado con el flujo de audio transmitido al por lo menos un controlador define una o más características que pertenecen a la reflexión.
- 55           7. El sistema según la reivindicación 6, en donde el conjunto de metadatos complementa un conjunto de metadatos base que incluye elementos de metadatos asociados con un flujo basado en el objeto de información de audio espacial, y en donde los elementos de metadatos para el flujo basado en el objeto especifican parámetros espaciales que controlan la reproducción de un sonido basado en objeto correspondiente y comprende uno o más de entre: posición de sonido, anchura del sonido y velocidad del sonido.
- 60           8. El sistema según la reivindicación 7, en donde el conjunto de metadatos incluye, además, elementos de metadatos asociados con un flujo basado en los canales de la información de audio espacial y en donde los elementos de metadatos asociados con cada flujo basado en los canales comprende designaciones de canales de sonido de ambiente envolvente de los controladores de audio en la configuración de sonido de ambiente envolvente definida.
- 65           9. El sistema según la reivindicación 6, en donde el al menos un controlador está asociado con un micrófono situado en el entorno de escucha, estando el micrófono configurado para transmitir información de audio de configuración que encapsula características del entorno de escucha hacia un componente de calibración acoplado al dispositivo de presentación y en donde la información de audio de configuración se utiliza por el dispositivo de presentación para definir o modificar el conjunto de metadatos asociado con el flujo de audio transmitido al por lo menos un controlador de audio.



5 **10.** El sistema según la reivindicación 1, en donde el al menos un controlador comprende uno de entre: un transductor de audio manualmente ajustable dentro de un recinto que es ajustable con respecto al ángulo de proyección de sonido relativo a un plano de implantación del entorno de escucha y un transductor de audio controlable eléctricamente en el interior de un recinto que es automáticamente ajustable con respecto al ángulo de proyección del sonido.

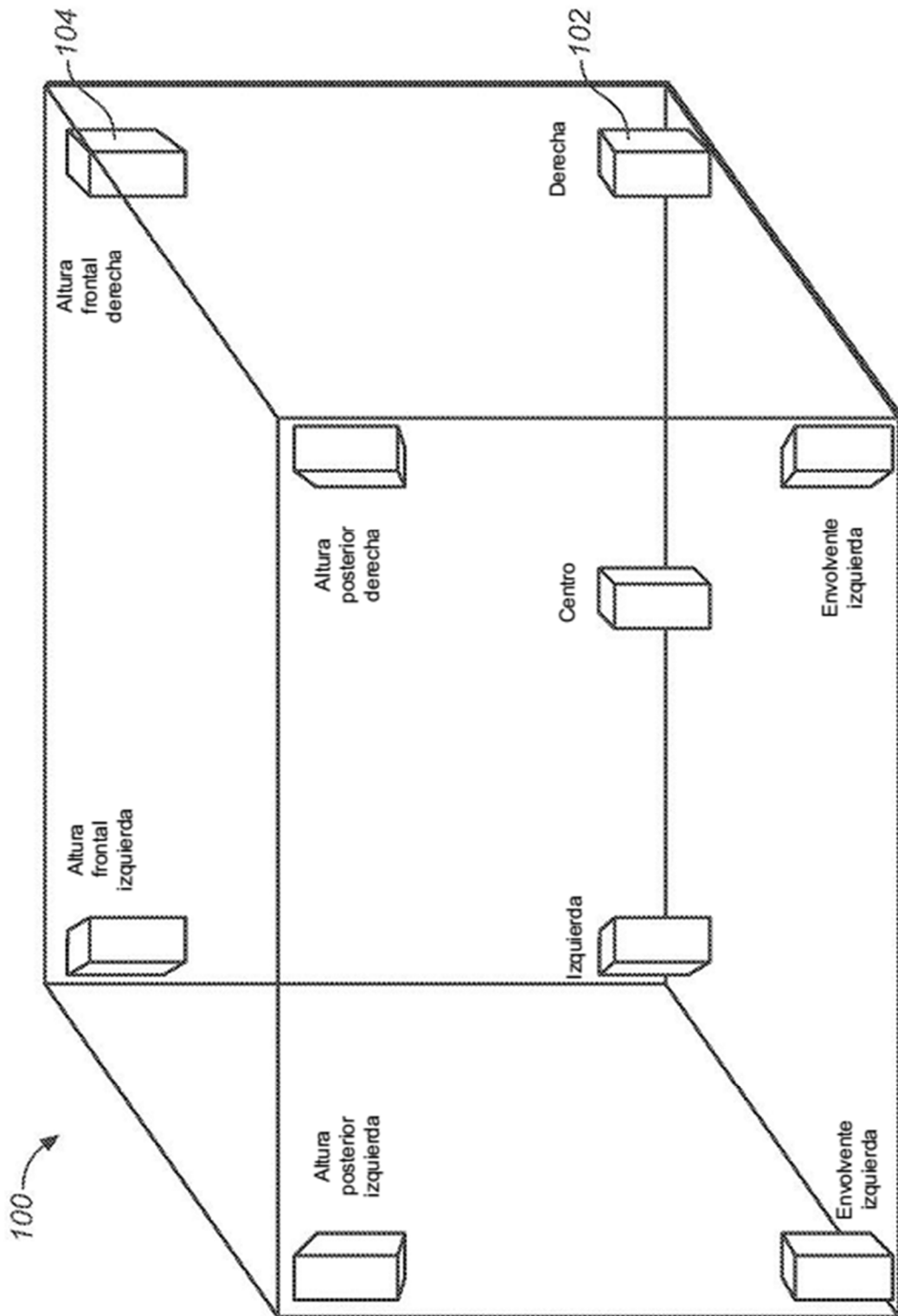
10 **11.** Un método en un sistema (500, 1300) que sirve para presentar el sonido utilizando elementos sonoros reflejados, comprendiendo el sistema una red de controladores de audio para distribución alrededor de un entorno de escucha (800), en donde al menos un controlador de la red de controladores de audio es un controlador de proyección hacia arriba (706), que está configurado para proyectar ondas sonoras hacia una o más superficies del entorno de escucha para reflexión a una zona de escucha dentro del entorno de escucha, comprendiendo dicho método:

15 recibir y procesar un flujo de bits que incluye flujos de audio y uno o más conjuntos de metadatos que están asociados con cada uno de los flujos de audio y que especifican un emplazamiento de reproducción en el entorno de escucha de un flujo de audio respectivo, en donde los flujos de audio comprenden uno o más flujos de audio reflejados y uno o más flujos de audio directos; y

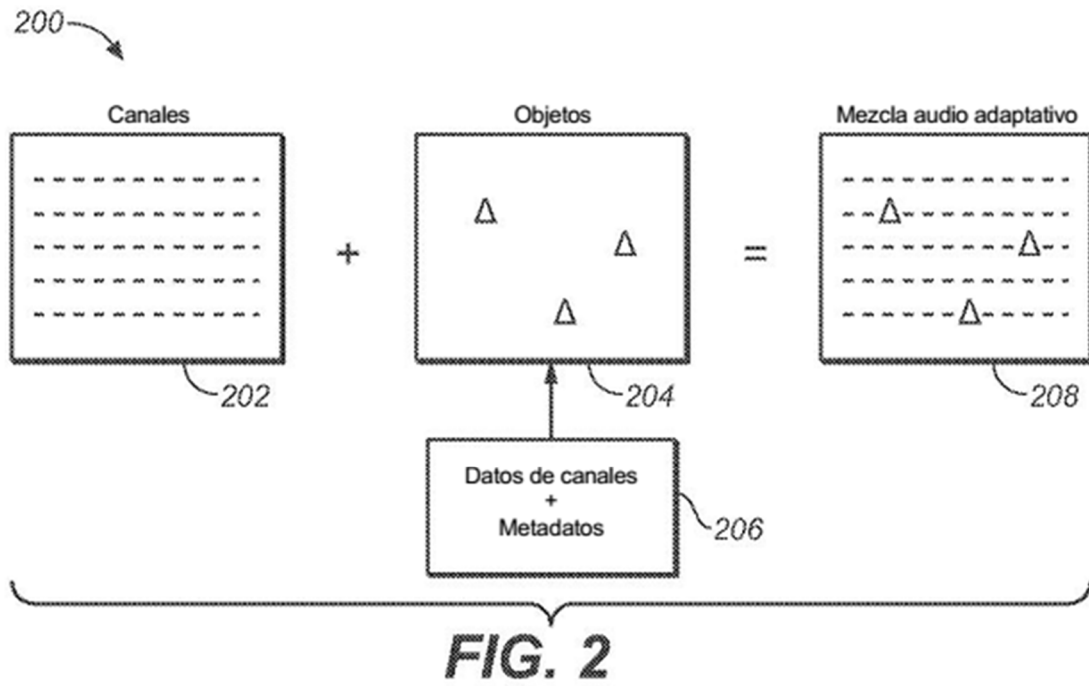
20 presentar los flujos de audio a una pluralidad de señales de alimentación de audio que corresponden a la red de controladores de audio en función con los uno o más conjuntos de metadatos y en donde los uno o más flujos de audio reflejados se transmiten al por lo menos un controlador de proyección hacia arriba.

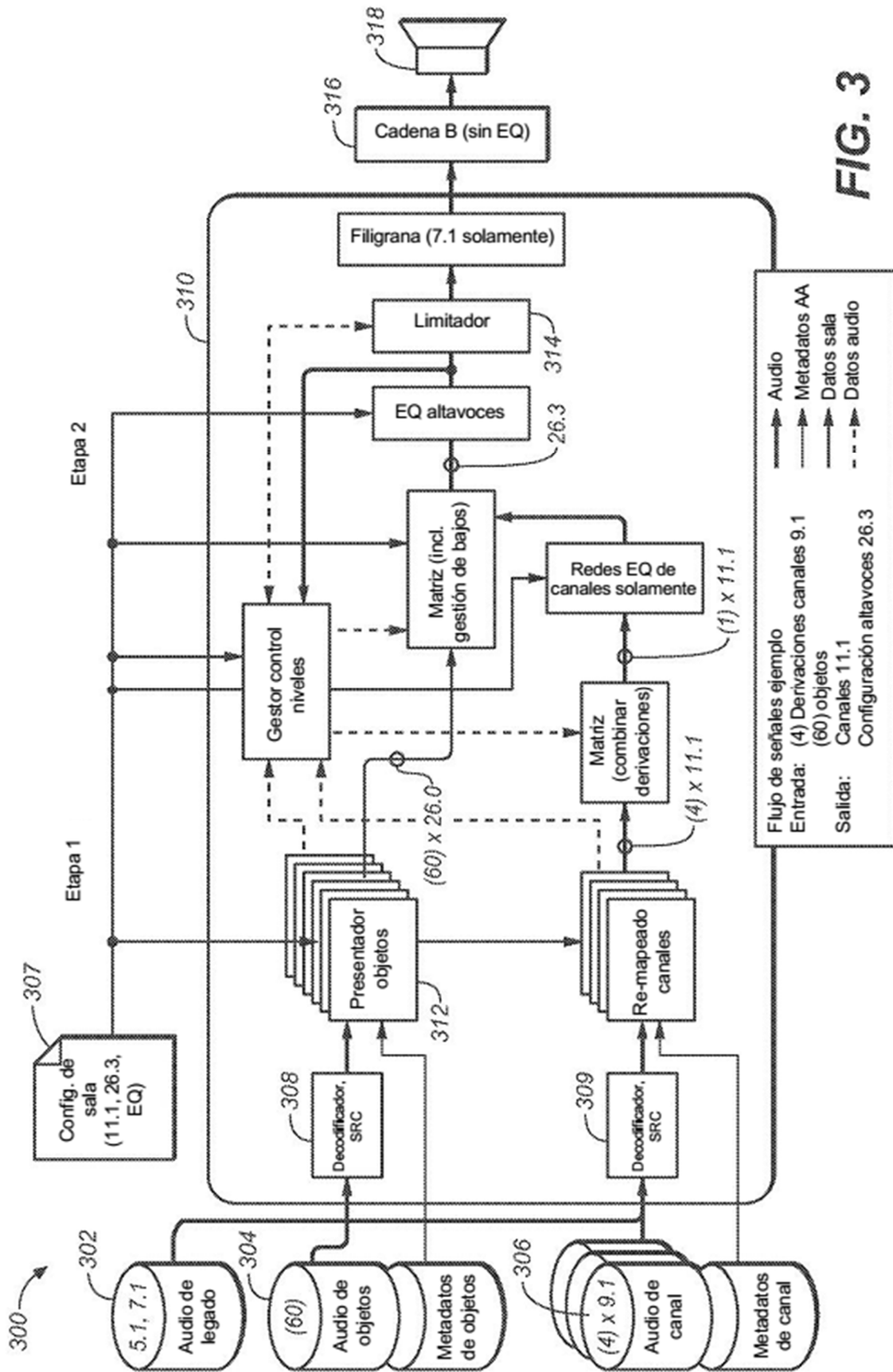
25 caracterizado por cuanto que realiza un procesamiento de señal para introducir pistas de altura de percepción en el uno o más flujos de audio reflejados transmitidos a al menos un controlador de proyección hacia arriba.

**12.** Un soporte legible por ordenador que comprende un programa informático, estando el programa informático configurado para controlar la ejecución de un dispositivo informático basado en procesador de un sistema para presentar sonido de manera que se realice el método descrito en la reivindicación 11.



**FIG. 1**





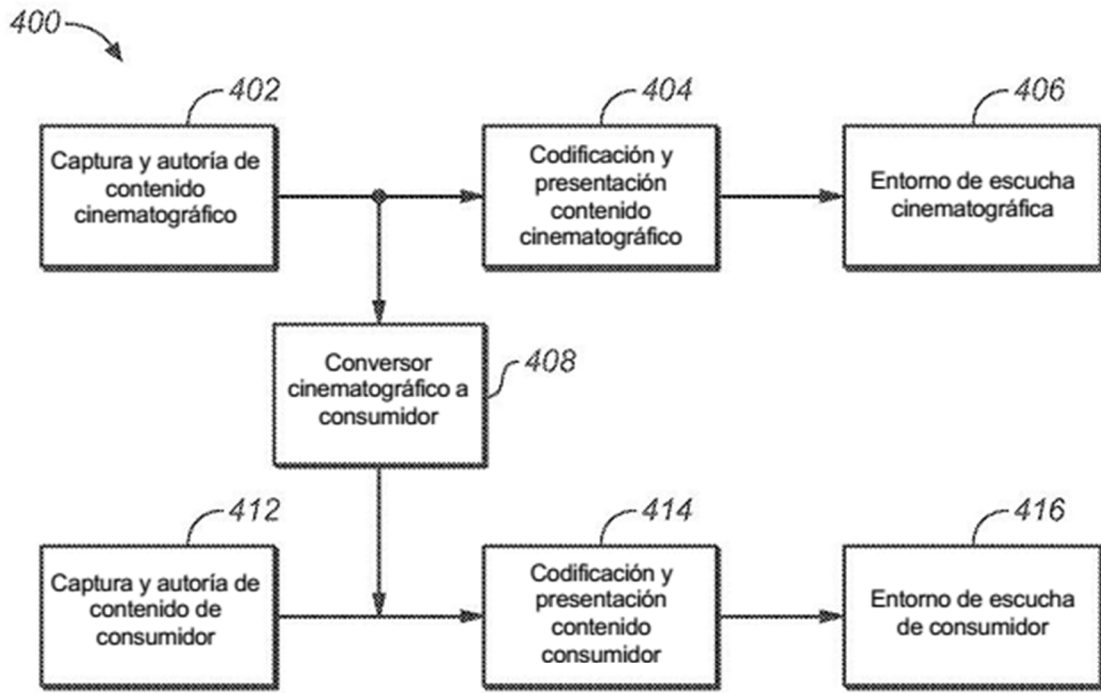


FIG. 4A

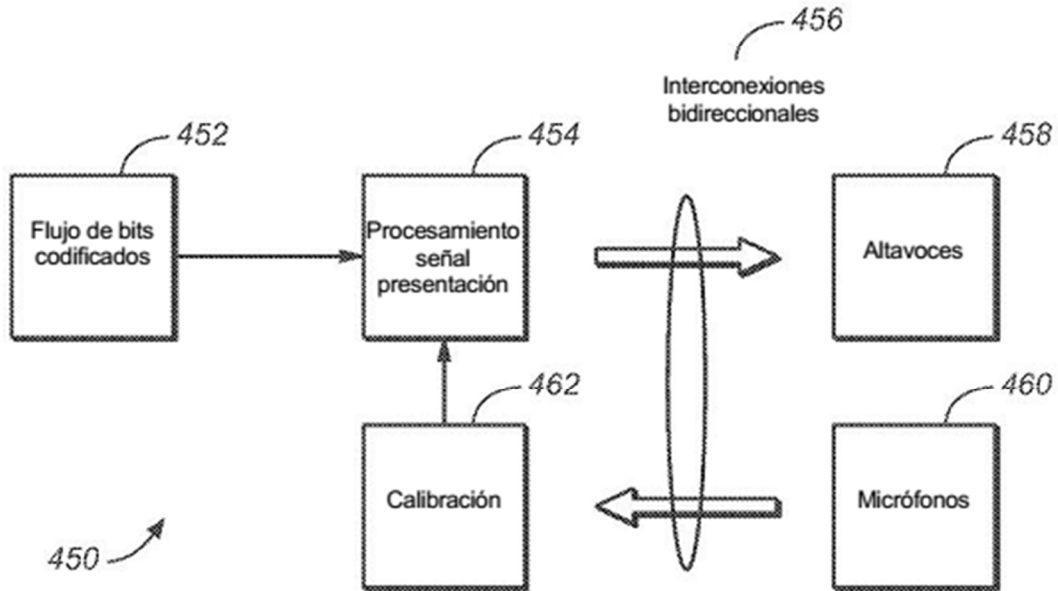


FIG. 4C

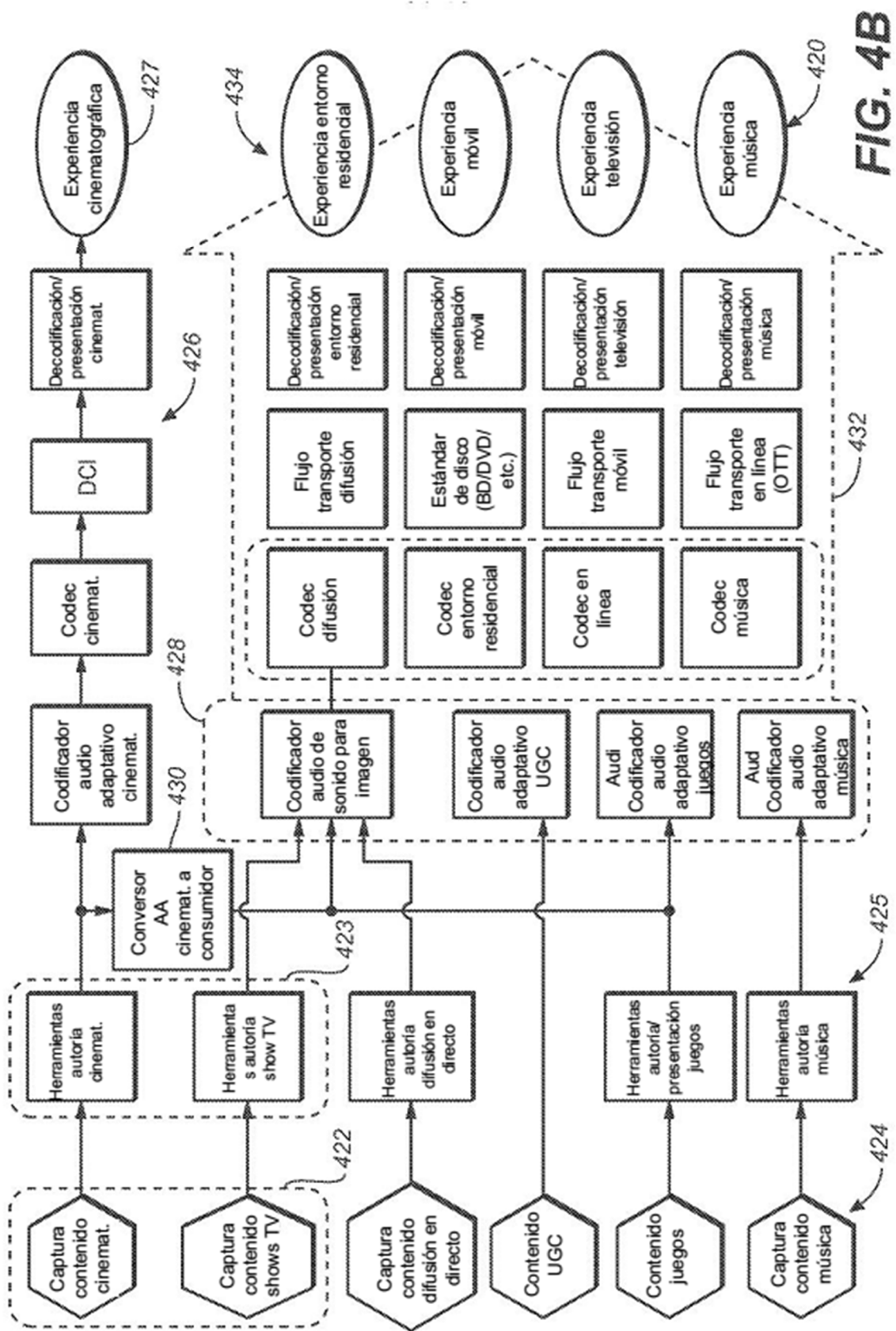
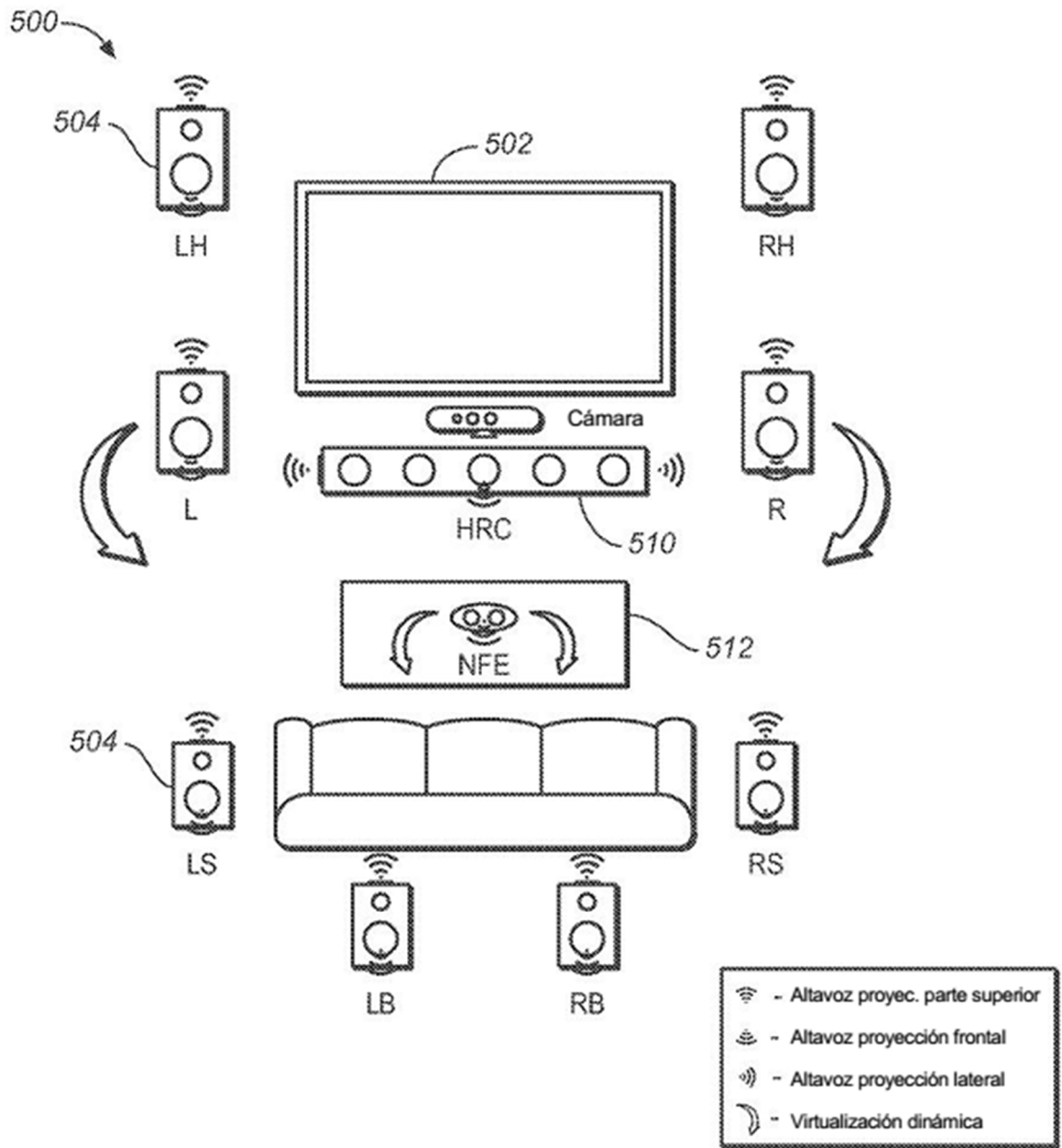
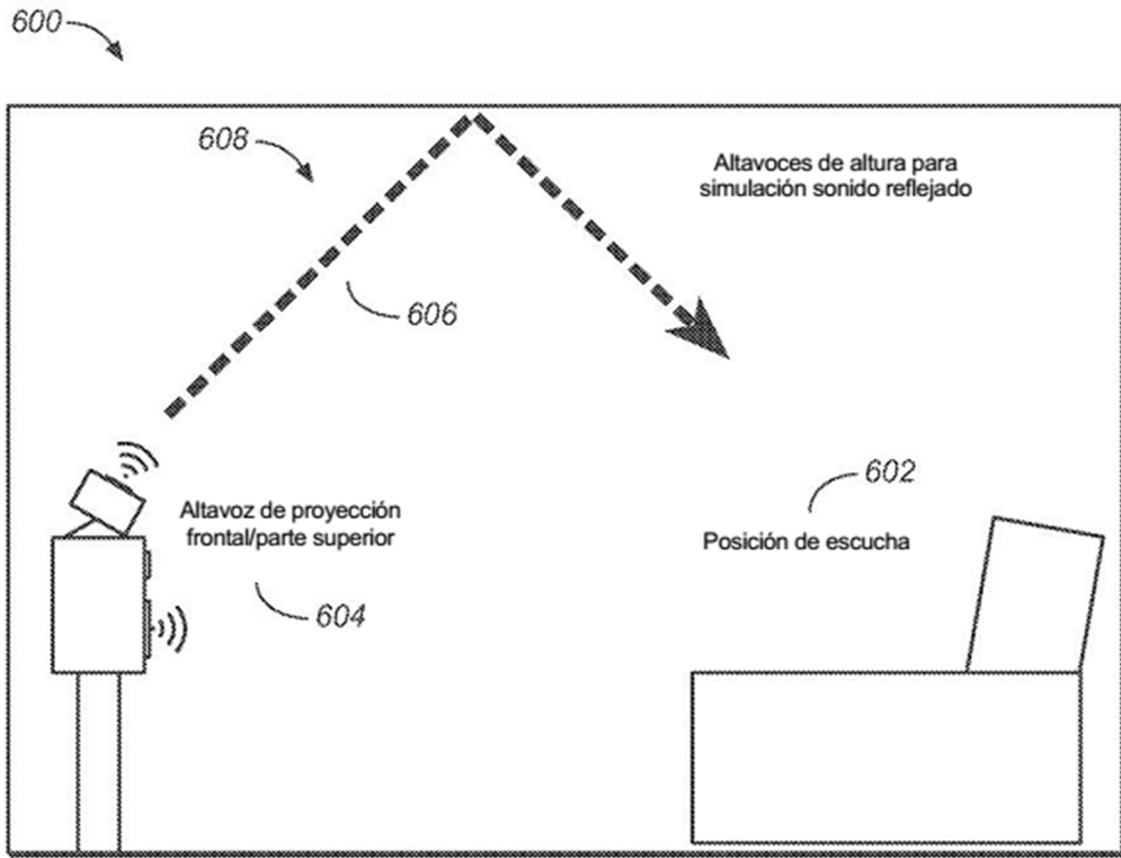


FIG. 4B

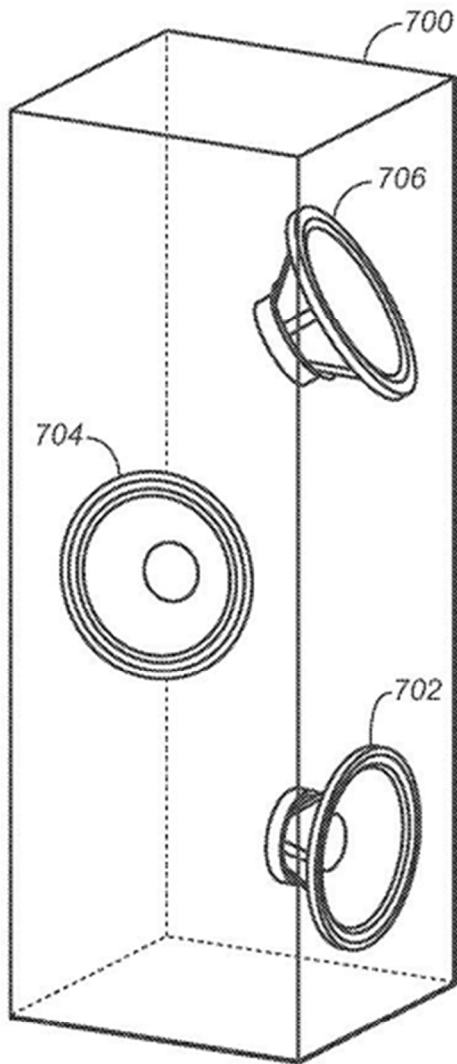


**FIG. 5**

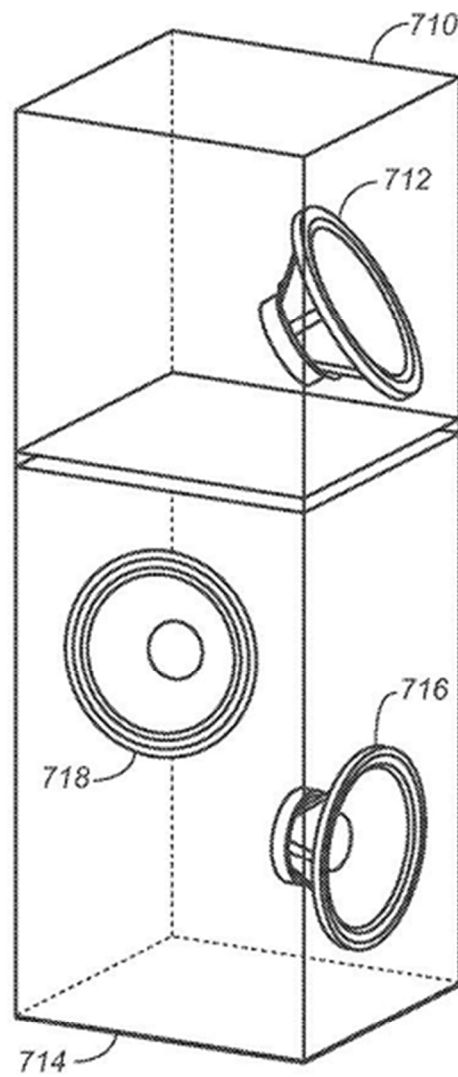


**FIG. 6**

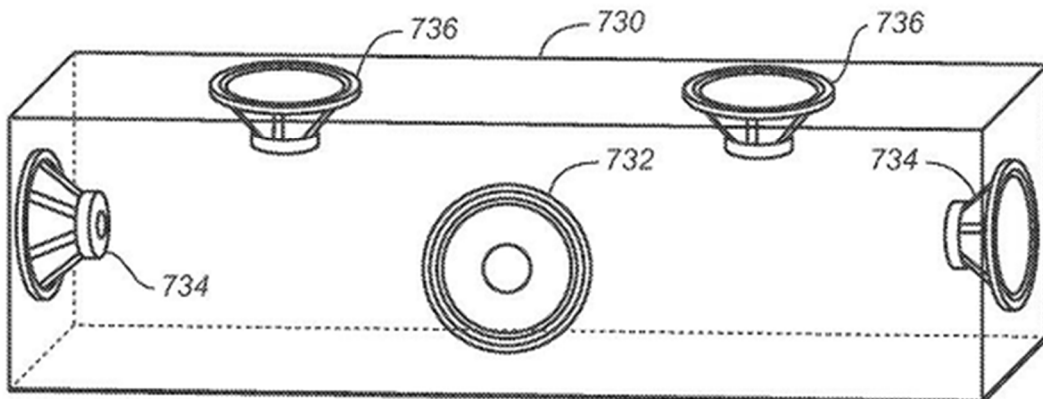




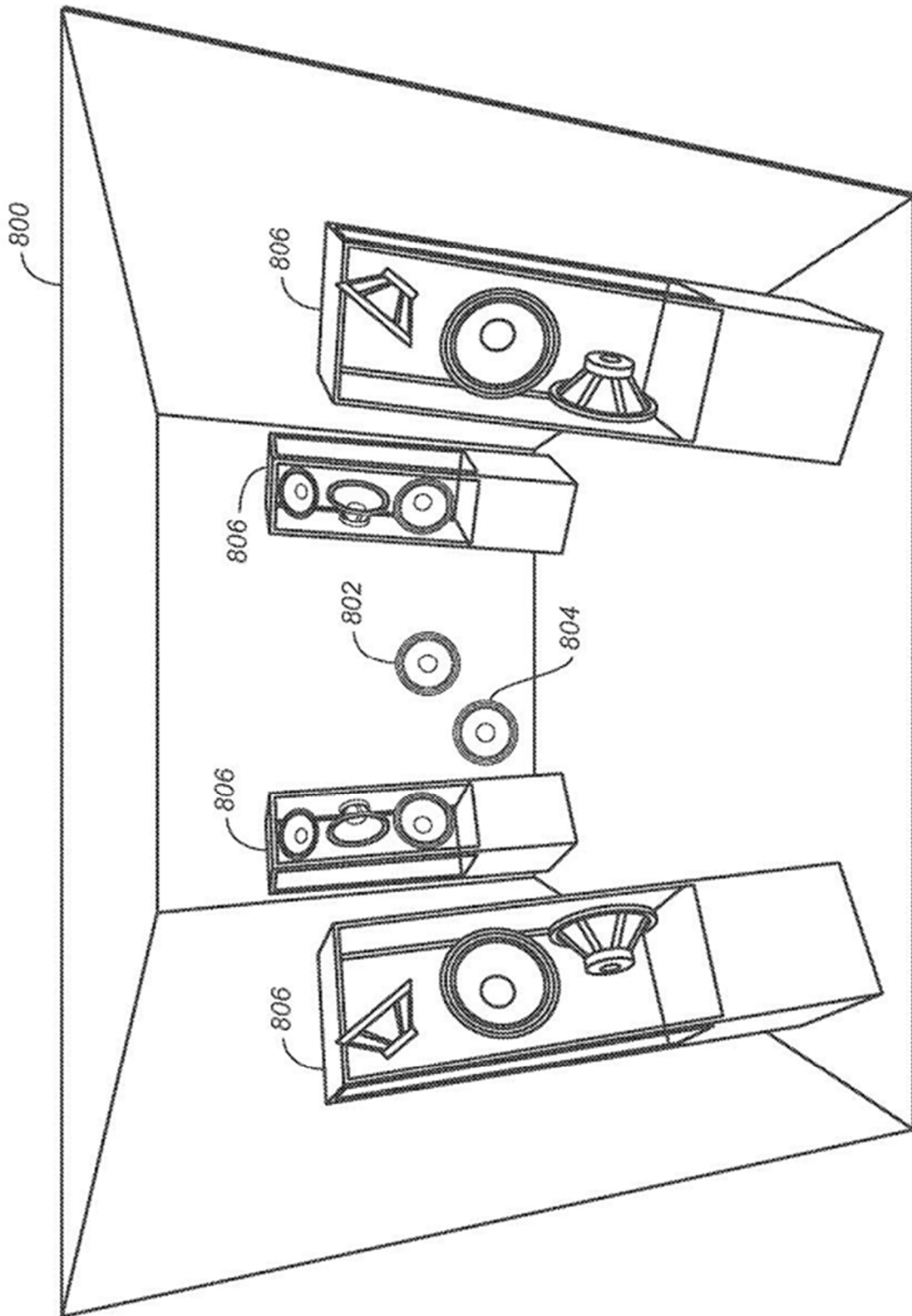
**FIG. 7A**



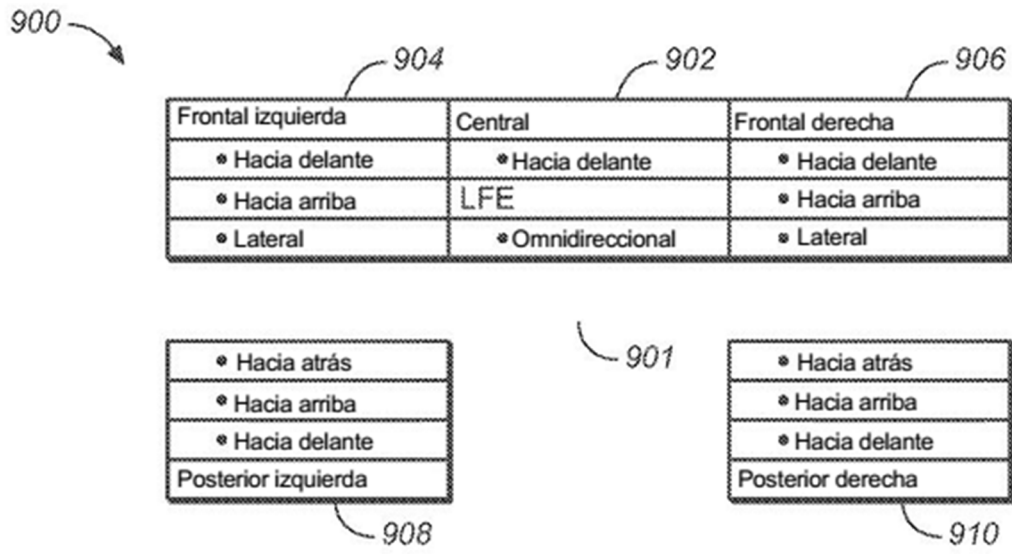
**FIG. 7B**



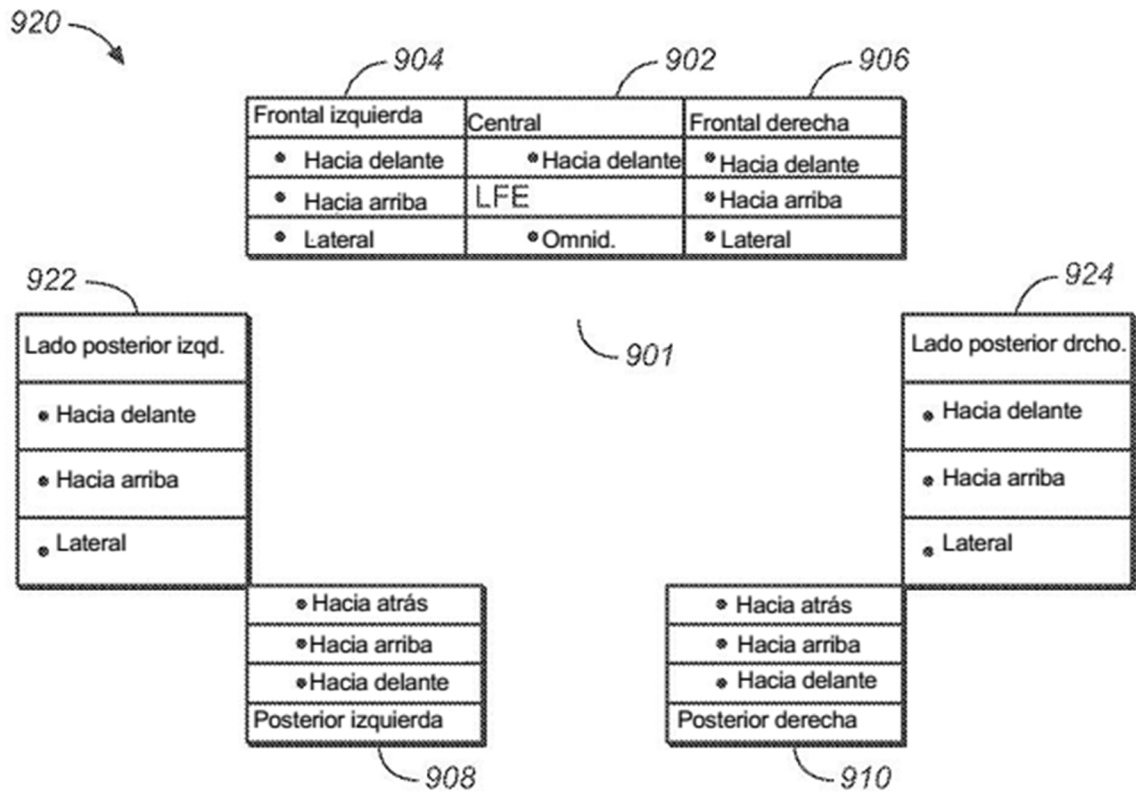
**FIG. 7C**



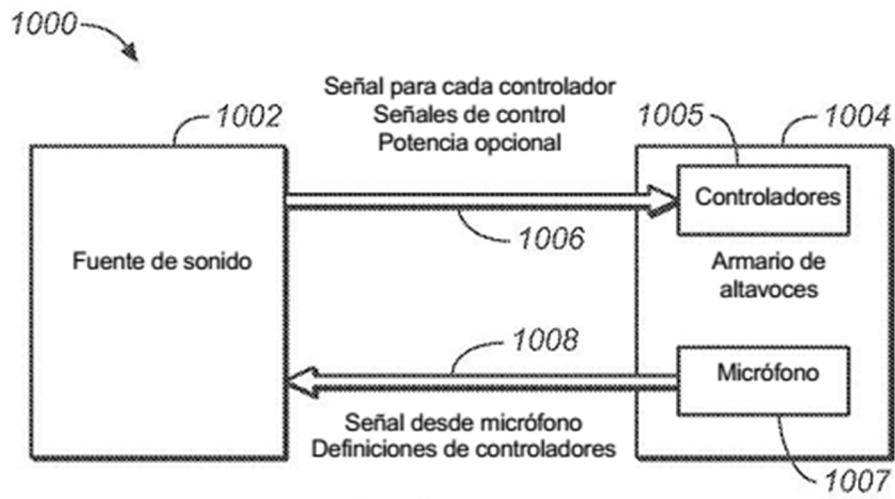
**FIG. 8**



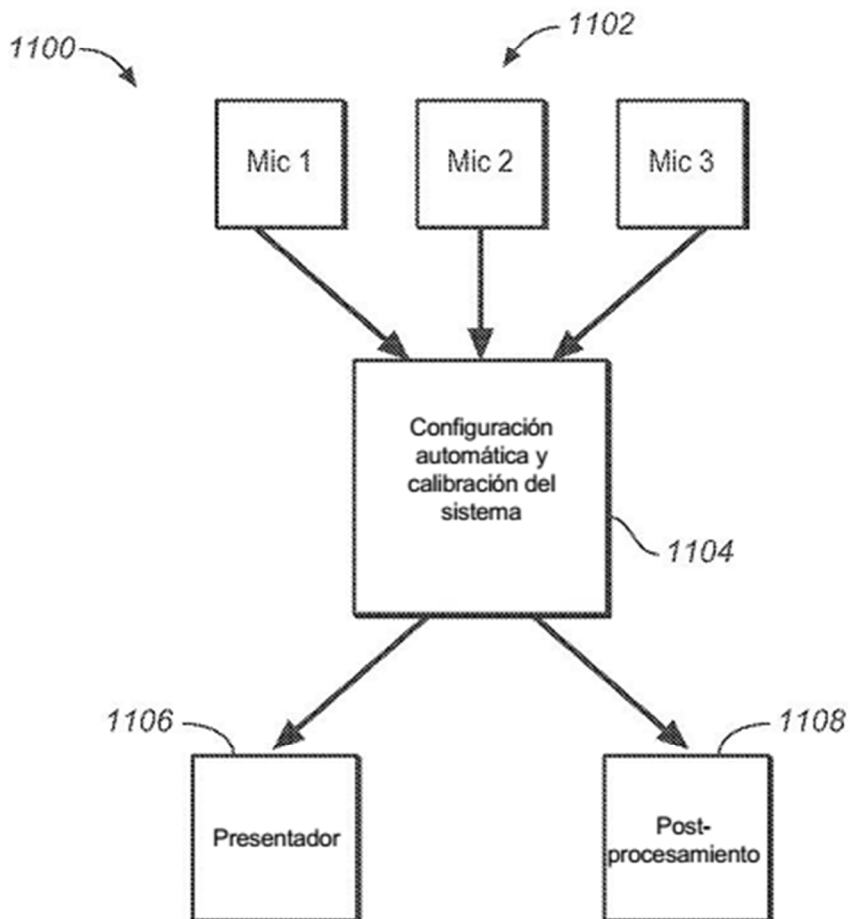
**FIG. 9A**



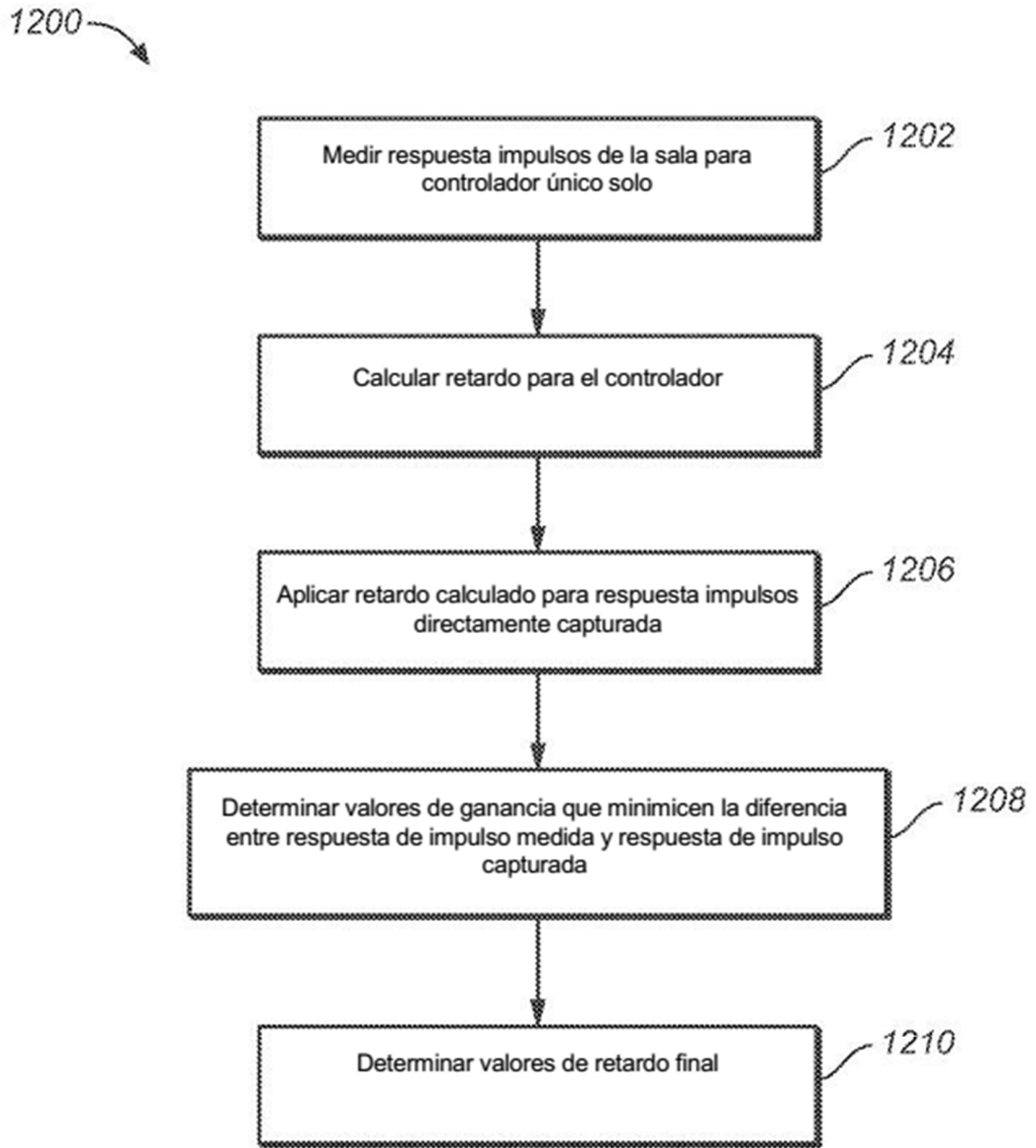
**FIG. 9B**



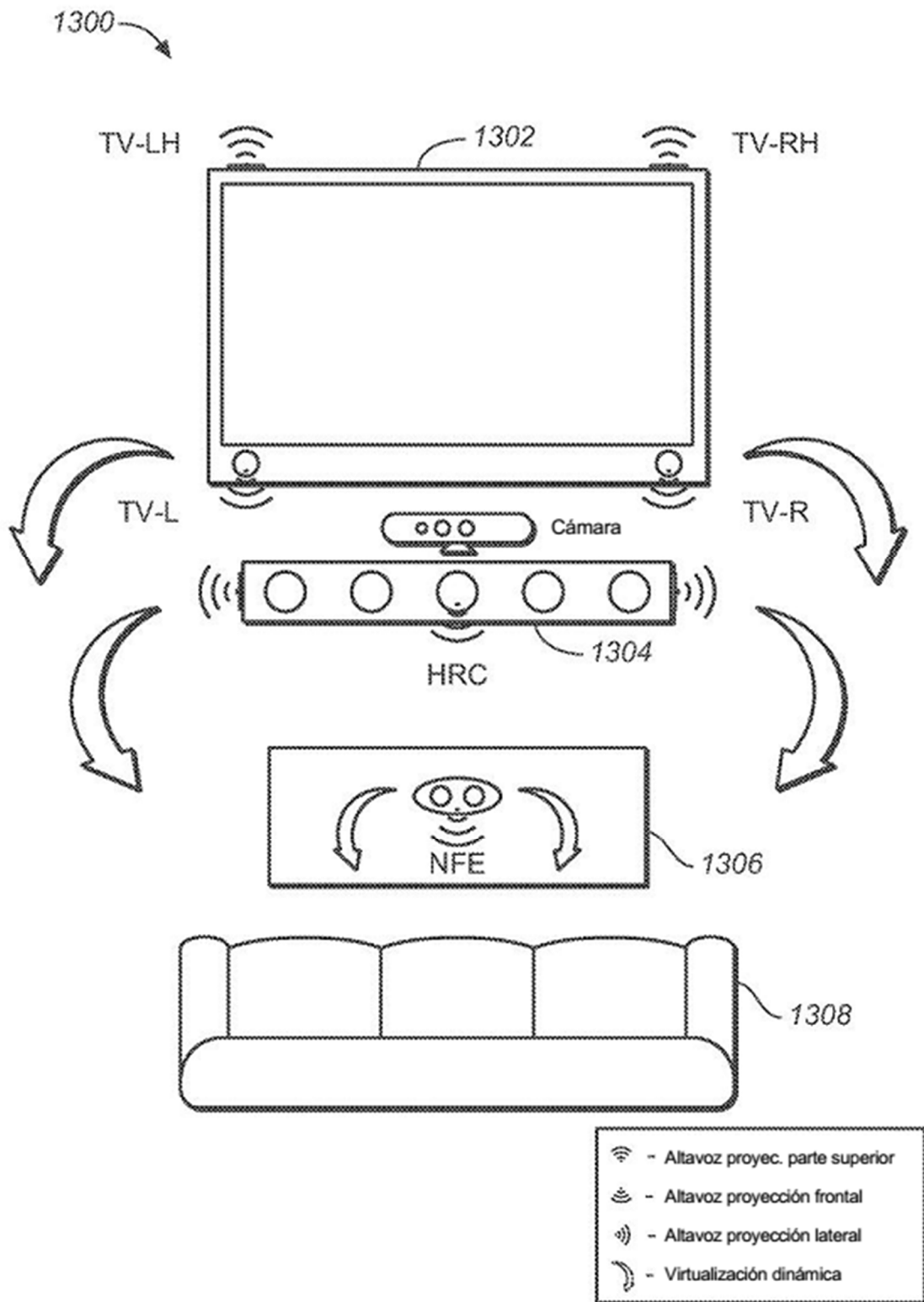
**FIG. 10**



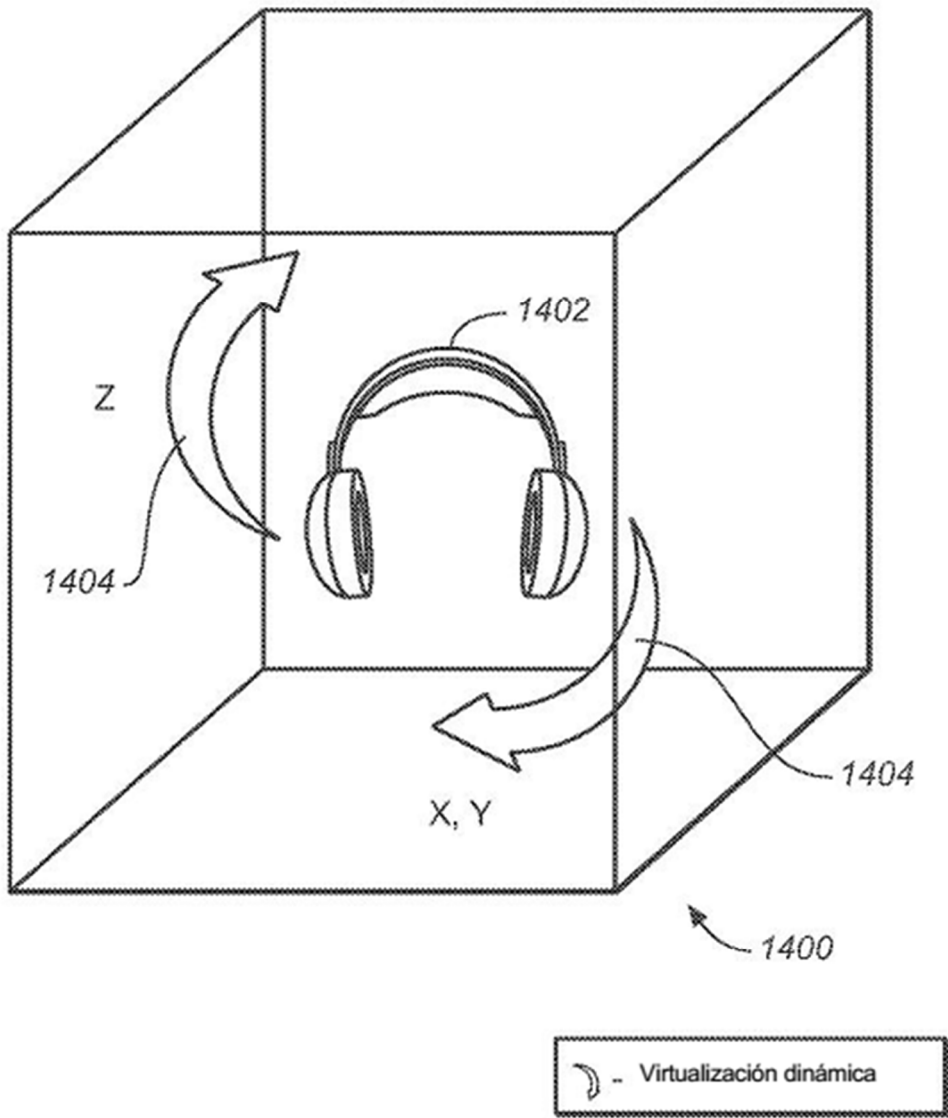
**FIG. 11**



**FIG. 12**



**FIG. 13**



**FIG. 14**

1500

Tipo de metadatos	Elementos de metadatos
Tipo de contenido de audio	Diálogo/música/ambiente/efectos Directo/difundido/reflejado
Definiciones de controladores	Número de controladores Características acústicas Posición de los controladores Ángulo de los controladores
Señales de control	Dirección activa Sintonía activa
Información de calibración	Tipo y localización de sensores Dimensiones de la sala Características ambientales Información de creación de imágenes Localizaciones de altavoces Localizaciones de altavoces

FIG. 15

1600

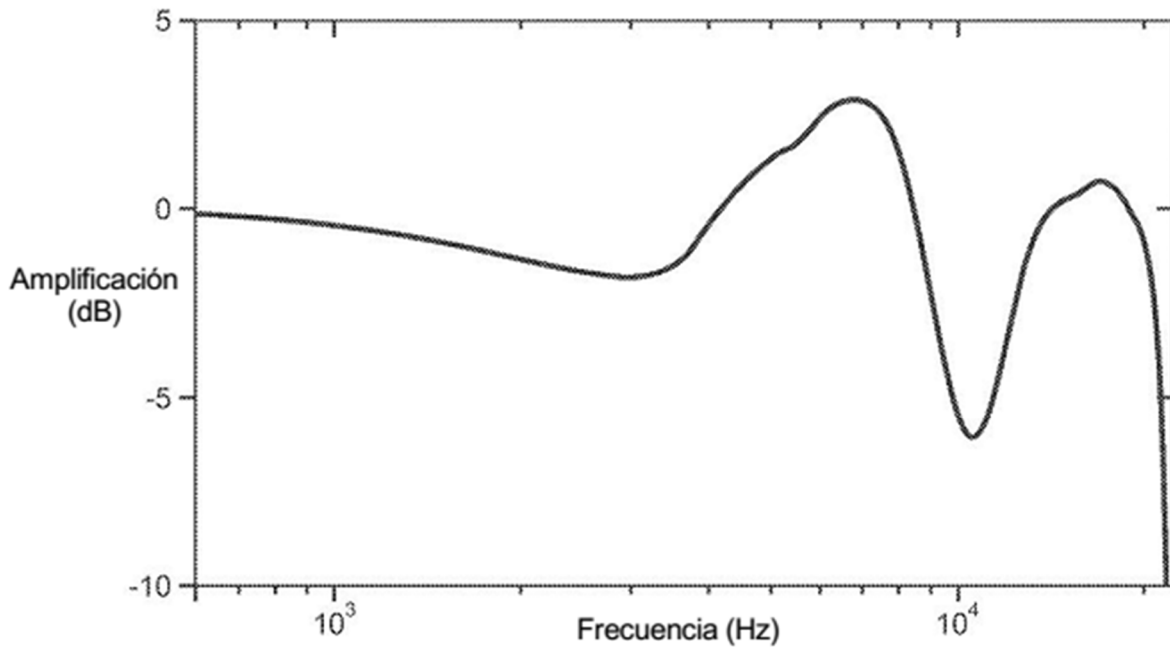


FIG. 16