

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 978 218**

51 Int. Cl.:

G10L 19/008 (2013.01)

H04S 3/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **07.10.2019 PCT/US2019/055009**
- 87 Fecha y número de publicación internacional: **16.04.2020 WO20076708**
- 96 Fecha de presentación y número de la solicitud europea: **07.10.2019 E 19794343 (4)**
- 97 Fecha y número de publicación de la concesión europea: **20.03.2024 EP 3864651**

54 Título: **Transformar señales de audio capturado en diferentes formatos a un número reducido de formatos para simplificar operaciones de codificación y decodificación**

30 Prioridad:

08.10.2018 US 201862742729 P

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

09.09.2024

73 Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (50.0%)
1275 Market Street
San Francisco, CA 94103, US y
DOLBY INTERNATIONAL AB (50.0%)

72 Inventor/es:

BRUHN, STEFAN;
ECKERT, MICHAEL;
TORRES, JUAN FELIX;
BROWN, STEFANIE y
MCGRATH, DAVID S.

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 978 218 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

Transformar señales de audio capturado en diferentes formatos a un número reducido de formatos para simplificar operaciones de codificación y decodificación

5

Campo técnico

Realizaciones de la presente divulgación se refieren generalmente a procesamiento de señal de audio y, más específicamente, a distribución de señales de audio capturado.

10

Antecedentes

El desarrollo del estándar de codificador/descodificador ("códec") de voz y vídeo se ha centrado recientemente en desarrollar un códec para servicios inmersivos de voz y audio (IVAS). Se espera que IVAS soporte un rango de capacidades de servicio, tal como funcionamiento desde mono a estéreo hasta codificación, decodificación y renderización de audio totalmente inmersivo. Un códec IVAS adecuado también proporciona una alta robustez ante errores en pérdida de paquetes y fluctuación de retardo bajo diferentes condiciones de transmisión. IVAS está destinado a ser soportado por una amplia gama de dispositivos, terminales y nodos de red, incluyendo pero no limitados a teléfonos móviles e inteligentes, tabletas electrónicas, ordenadores personales, teléfonos de conferencia, salas de conferencia, dispositivos de realidad virtual y realidad aumentada, dispositivos de cine en casa y otros dispositivos adecuados. Debido a que estos dispositivos, terminales y nodos de red pueden tener diversas interfaces acústicas para captura y renderización de sonido, puede no ser práctico para un códec IVAS abordar todas las diversas formas en que se captura y renderiza una señal de audio.

15

20

25

El documento EP2873254 se refiere a compensación de posición de altavoz con codificación jerárquica de audio 3D. Un primer conjunto de canales de audio se transforma en base a un modelo de onda esférica y luego se transforma en el dominio de la frecuencia para generar un segundo conjunto de canales de audio. El documento US2008/0319764 se refiere a un método para determinar un modo de codificación espacial de datos de audio. El método se puede implementar en una videoconferencia VoIP.

30

Sumario

Un objeto de la invención es superar las deficiencias de la técnica anterior. Este objeto de la invención se resuelve mediante las reivindicaciones independientes. Realizaciones específicas se definen en las reivindicaciones dependientes.

35

Las realizaciones divulgadas permiten convertir señales de audio capturado en diversos formatos mediante diversos dispositivos de captura en un número limitado de formatos que pueden ser procesados mediante un códec, por ejemplo, un códec IVAS.

40

En algunas realizaciones, una unidad de simplificación integrada en un dispositivo de audio recibe una señal de audio. Esa señal de audio puede ser una señal capturada por uno o más dispositivos de captura de audio acoplados con el dispositivo de audio. La señal de audio puede ser, por ejemplo, el audio de una videoconferencia entre personas en diferentes ubicaciones. La unidad de simplificación determina si la señal de audio está en un formato que no está soportado por una unidad de codificación del dispositivo de audio, comúnmente denominada como un "codificador". Por ejemplo, la unidad de simplificación puede determinar si la señal de audio está o no en un formato mono, estéreo o espacial estándar o propietario. En base a determinar que la señal de audio está en un formato que no está soportado por la unidad de codificación, la unidad de simplificación convierte la señal de audio en un formato que está soportado por la unidad de codificación. Por ejemplo, si la unidad de simplificación determina que la señal de audio está en un formato espacial propietario, la unidad de simplificación puede convertir la señal de audio en un formato "intermedio" espacial soportado por la unidad de codificación. La unidad de simplificación transfiere la señal de audio convertida a la unidad de codificación.

45

50

55

Una ventaja de las realizaciones divulgadas es que la complejidad de un códec, por ejemplo, un códec IVAS, se puede reducir reduciendo un número potencialmente grande de formatos de captura de audio a un número limitado de formatos, por ejemplo, mono, estéreo y espacial. Como resultado, el códec se puede desplegar en una variedad de dispositivos independientemente de las capacidades de captura de audio de los dispositivos.

60

Estos y otros aspectos, características, y realizaciones pueden expresarse como métodos, aparato, sistemas, componentes, productos de programa, medios o pasos para realizar una función, y de otras formas.

65

En algunas implementaciones, una unidad de simplificación de un dispositivo de audio recibe una señal de audio en un primer formato. El primer formato es uno de un conjunto de múltiples formatos de audio soportados por el dispositivo de audio. La unidad de simplificación determina si el primer formato está soportado por un codificador del dispositivo de audio. De acuerdo con el primer formato que no está soportado por el codificador, la unidad de

5 simplificación convierte la señal de audio en un segundo formato que está soportado por el codificador. El segundo formato es una representación alternativa del primer formato. La unidad de simplificación transfiere la señal de audio en el segundo formato al codificador. El codificador codifica la señal de audio. El dispositivo de audio almacena la señal de audio codificada o transmite la señal de audio codificada a uno o más dispositivos distintos.

10 Convertir la señal de audio en el segundo formato puede incluir generar metadatos para la señal de audio. Los metadatos pueden incluir una representación de una porción de la señal de audio. Codificar la señal de audio puede incluir codificar la señal de audio en el segundo formato en un formato de transporte soportado por un segundo dispositivo. El dispositivo de audio puede transmitir la señal de audio codificada al transmitir los metadatos que comprenden una representación de una porción de la señal de audio no soportada por el segundo formato.

15 En algunas implementaciones, determinar, por la unidad de simplificación, si la señal de audio está en el primer formato puede incluir determinar un número de dispositivos de captura de audio y una posición correspondiente de cada dispositivo de captura usado para capturar la señal de audio. Cada uno de los uno o más dispositivos puede ser configurado para reproducir la señal de audio del segundo formato. Al menos uno de los uno o más dispositivos distintos puede no ser capaz de reproducir la señal de audio del primer formato.

20 El segundo formato puede representar la señal de audio como un número de objetos de audio en una escena de audio, de los cuales ambos dependen de un número de canales de audio para transportar información espacial. El segundo formato puede incluir metadatos para transportar una porción adicional de información espacial. El primer formato y el segundo formato pueden ambos ser formatos de audio espacial. El segundo formato puede ser un formato de audio espacial y el primer formato puede ser un formato mono asociado con metadatos o un
25 formato estéreo asociado con metadatos. El conjunto de múltiples formatos de audio soportados por el dispositivo de audio puede incluir múltiples formatos de audio espacial. El segundo formato puede ser una representación alternativa del primer formato y se caracteriza además por permitir un grado comparable de calidad de experiencia.

30 En algunas implementaciones, una unidad de renderizado de un dispositivo de audio recibe una señal de audio en un primer formato. La unidad de renderizado determina si el dispositivo de audio es capaz de reproducir la señal de audio en el primer formato. En respuesta a determinar que el dispositivo de audio es incapaz de reproducir la señal de audio en el primer formato, la unidad de renderizado adapta la señal de audio que esté disponible en un segundo formato. La unidad de renderizado transfiere la señal de audio en el segundo formato
35 para renderización.

En algunas implementaciones, convertir, por la unidad de renderizado, la señal de audio en el segundo formato puede incluir usar metadatos que incluyen una representación de una porción de la señal de audio no soportada por un cuarto formato usado para codificar en combinación con la señal de audio en un tercer formato. Aquí, el
40 tercer formato corresponde al término "primer formato" en el contexto de la unidad de simplificación, que es uno de un conjunto de múltiples formatos de audio soportados en el lado del codificador. El cuarto formato corresponde al término "segundo formato" en el contexto de la unidad de simplificación, que es un formato soportado por el codificador, y que es una representación alternativa del tercer formato. Aquí y en cualquier parte más de esta memoria descriptiva, los términos primero, segundo, tercero y cuarto se usan para identificación y no son necesariamente indicativos de un orden particular.
45

En algunas implementaciones, una unidad de simplificación recibe, desde una unidad de preprocesamiento acústico, señales de audio en múltiples formatos. La unidad de simplificación recibe, desde un dispositivo, atributos del dispositivo, los atributos incluyen indicaciones de uno o más formatos de audio soportados por el
50 dispositivo. El uno o más formatos de audio incluyen al menos uno de un formato mono, un formato estéreo, o un formato espacial. La unidad de simplificación convierte las señales de audio en un formato de ingesta que es una representación alternativa de uno o más formatos de audio. La unidad de simplificación proporciona la señal de audio convertida a una unidad de codificación para su procesamiento aguas abajo. Cada una de la unidad de preprocesamiento acústico, la unidad de simplificación, y la unidad de codificación puede incluir uno o más
55 procesadores de ordenador.

En algunas implementaciones, un sistema de codificación incluye una unidad de captura configurada para capturar una señal de audio, una unidad de preprocesamiento acústico configurada para realizar operaciones que comprenden preprocesar la señal de audio, un codificador y una unidad de simplificación. La unidad de
60 simplificación está configurada para realizar las siguientes operaciones. La unidad de simplificación recibe, de la unidad de preprocesamiento acústico, una señal de audio en un primer formato. El primer formato es uno de un conjunto de múltiples formatos de audio soportados por el codificador. La unidad de simplificación determina si el primer formato es soportado por el codificador. En respuesta a determinar que el primer formato no es soportado por el codificador, la unidad de simplificación convierte la señal de audio en un segundo formato que es soportado por el codificador. La unidad de simplificación transfiere la señal de audio en el segundo formato al
65 codificador. El codificador está configurado para realizar operaciones que incluyen codificar la señal de audio y al

menos una de almacenar la señal de audio codificada o transmitir la señal de audio codificada a otro dispositivo.

5 En algunas implementaciones, convertir la señal de audio al segundo formato incluye generar metadatos para la señal de audio. Los metadatos pueden incluir una representación de una porción de la señal de audio no soportada por el segundo formato. Las operaciones del codificador pueden incluir además transmitir la señal de audio codificada al transmitir los metadatos que incluyen una representación de una porción de la señal de audio no soportada por el segundo formato.

10 En algunas implementaciones, el segundo formato representa la señal de audio como un número de objetos en una escena de audio y un número de canales para transportar información espacial. En algunas implementaciones, preprocesar la señal de audio puede incluir uno o más de realizar cancelación de ruido, realizar cancelación de eco, reducir un número de canales de la señal de audio, aumentar el número de canales de audio de la señal de audio, o generar metadatos acústicos.

15 En algunas implementaciones, convertir la señal de audio al segundo formato puede incluir usar metadatos que incluyen una representación de una porción de la señal de audio no soportada por un cuarto formato usado para codificar en combinación con la señal de audio un tercer formato. Aquí, el tercer formato corresponde al término "primer formato" en el contexto de la unidad de simplificación, que es uno de un conjunto de múltiples formatos de audio soportados en el lado del codificador. El cuarto formato corresponde al término "segundo formato" en el contexto de la unidad de simplificación, que es un formato soportado por el codificador, y que es una representación alternativa del tercer formato.

20 Estos y otros aspectos, características, y realizaciones resultarán evidentes a partir de las siguientes descripciones, incluidas las reivindicaciones.

25 **Breve descripción de los dibujos**

30 En los dibujos, se muestran para facilitar la descripción disposiciones u ordenamientos específicos de elementos esquemáticos, tales como aquellos que representan dispositivos, unidades, bloques de instrucciones y elementos de datos. Sin embargo, los expertos en la técnica deben entender que el orden o disposición específicos de los elementos esquemáticos en los dibujos no implica que se requiera un particular orden o secuencia de procesamiento, o separación de procesos. Además, la inclusión de un elemento esquemático en un dibujo no implica que se requiera tal elemento en todas las realizaciones o que las características representadas por tal elemento no puedan incluirse o combinarse con otros elementos en algunas realizaciones.

35 Además, en los dibujos, donde elementos de conexión, tales como líneas continuas o discontinuas o flechas, se usan para ilustrar una conexión, relación o asociación entre dos o más elementos esquemáticos distintos, la ausencia de cualquiera de tales elementos de conexión no implica que ninguna conexión, relación o asociación pueda existir. En otras palabras, algunas conexiones, relaciones, o asociaciones entre elementos no se muestran en los dibujos como para no oscurecer la divulgación. Además, para facilitar la ilustración, se usa un único elemento de conexión para representar múltiples conexiones, relaciones o asociaciones entre elementos. Por ejemplo, donde un elemento de conexión representa una comunicación de señales, datos o instrucciones, los expertos en la técnica deben entender que tal elemento representa una o múltiples rutas de señal, como pueda ser necesario, para afectar la comunicación.

40 La figura 1 ilustra diversos dispositivos que pueden estar soportados por el sistema IVAS, de acuerdo con algunas realizaciones de la presente divulgación.

50 La figura 2A es un diagrama de bloques de un sistema para transformar una señal de audio capturada a un formato listo para codificar, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 2B es un diagrama de bloques de un sistema para transformar de regreso el audio capturado a un formato de reproducción adecuado, de acuerdo con algunas realizaciones de la presente divulgación.

55 La figura 3 es un diagrama de flujo de acciones ejemplares para transformar una señal de audio a un formato soportado por una unidad de codificación, de acuerdo con algunas realizaciones de la presente divulgación.

60 La figura 4 es un diagrama de flujo de acciones ejemplares para determinar si una señal de audio está en un formato soportado por la unidad de codificación, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 5 es un diagrama de flujo de acciones ejemplares para transformar una señal de audio a un formato de reproducción disponible, de acuerdo con algunas realizaciones de la presente divulgación.

65 La figura 6 es otro diagrama de flujo de acciones ejemplares para transformar una señal de audio a un formato de reproducción disponible, de acuerdo con algunas realizaciones de la presente divulgación.

La figura 7 es un diagrama de bloques de una arquitectura de hardware para implementar las características descritas en referencia a las figuras 1-6, de acuerdo con algunas realizaciones de la presente divulgación.

5 **Descripción detallada**

En la siguiente descripción, a efectos de explicación, se establecen numerosos detalles específicos para proporcionar una comprensión profunda de la presente divulgación. Será evidente, sin embargo, que la presente divulgación se puede practicar sin estos detalles específicos.

10

Ahora se hará referencia en detalle a realizaciones, ejemplos de las cuales se ilustran en los dibujos que se acompañan. En la siguiente descripción detallada, se establecen numerosos detalles específicos para proporcionar una comprensión profunda de las diversas realizaciones descritas. Sin embargo, será evidente para un experto en la técnica que las diversas realizaciones descritas pueden practicarse sin estos detalles específicos. En otros casos, métodos, procedimientos, componentes y circuitos bien conocidos, no se han descrito en detalle como para no oscurecer innecesariamente aspectos de las realizaciones. Varias características se describen a continuación que pueden usarse independientemente una de otra o con cualquier combinación de otras características.

15

20

Como se usa en el presente documento, el término “incluye” y sus variantes deben leerse como términos abiertos que significan “incluye, pero no se limita a”. El término “o” debe leerse como “y/o” a menos que el contexto indique claramente lo contrario. El término “en base a” debe leerse como “en base al menos en parte a”.

25

La figura 1 ilustra diversos dispositivos que pueden estar soportados por el sistema IVAS. En algunas implementaciones, estos dispositivos se comunican a través del servidor de llamadas 102 que puede recibir señales de audio desde, por ejemplo, una red telefónica pública conmutada (PSTN) o un dispositivo de red móvil terrestre pública (PLMN) ilustrado por el dispositivo PSTN/OTRA PLMN 104. Este dispositivo puede usar el estándar G.711 y/o G.722 para la compresión y descompresión de audio (voz). Un dispositivo 104 generalmente es capaz de capturar y renderizar solo audio mono. El sistema IVAS está habilitado para soportar también equipos de usuario heredados 106. Esos dispositivos heredados pueden incluir dispositivos de servicios de voz mejorados (EVS), dispositivos que soportan estándar de codificación de voz a audio de banda ancha de múltiples tasas adaptativa (AMR-WB), dispositivos que soportan banda estrecha de múltiples tasas adaptativa (AMR-NB) y otros dispositivos adecuados. Estos dispositivos normalmente renderizan y capturan audio solo en mono.

30

35

El sistema IVAS también está habilitado para soportar equipos de usuario que capturan y renderizan señales de audio en diversos formatos, incluidos formatos de audio mejorados. Por ejemplo, el sistema IVAS está habilitado para soportar dispositivos de captura y renderizado estéreo (por ejemplo, equipo de usuario 108, ordenador portátil 114, y sistema de sala de conferencias 118), dispositivos de captura mono y renderizado binaural (por ejemplo, dispositivo de usuario 110 y dispositivo de ordenador 112), dispositivos de captura y renderizado inmersivos (por ejemplo, equipamiento de uso de sala de conferencias 116), dispositivos de captura estéreo y renderizado inmersivo (por ejemplo, cine en casa 120), captura mono y renderizado inmersivo (por ejemplo, equipo de realidad virtual (VR) 122), ingesta de contenido inmersivo 124, y otros dispositivos adecuados. Para soportar todos estos formatos directamente, el códec para el sistema IVAS necesitaría ser muy complejo y costoso de instalar. De este modo, sería deseable un sistema para simplificar el códec antes de la etapa de codificación.

40

45

Aunque, la descripción que sigue se centra en un sistema y códec IVAS, las realizaciones divulgadas son aplicables a cualquier códec para cualquier sistema de audio donde existe una ventaja en reducir un gran número de formatos de captura de audio a un número menor para reducir la complejidad del códec de audio o por cualquier otra razón deseada.

50

La figura 2A es un diagrama de bloques de un sistema 200 para transformar señales de audio capturadas a un formato listo para codificar, de acuerdo con algunas realizaciones de la presente divulgación. La unidad de captura 210 recibe una señal de audio de uno o más dispositivos de captura, por ejemplo, micrófonos. Por ejemplo, la unidad de captura 210 puede recibir una señal de audio de un micrófono (por ejemplo, señal mono), de dos micrófonos (por ejemplo, señal estéreo), de tres micrófonos, o de otro número y configuración de dispositivos de captura de audio. La unidad de captura 210 puede incluir personalizaciones por una o más terceras partes, donde las personalizaciones pueden ser particulares de los dispositivos de captura usados.

55

60

En algunas implementaciones, se captura una señal de audio mono con un micrófono. La señal mono se puede capturar, por ejemplo, con teléfono PSTN/PLMN 104, equipo de usuario heredado 106, dispositivo de usuario 110 con auriculares manos libres, dispositivo ordenador 112 con unos auriculares conectados, y el equipo de realidad virtual 122, como se ilustra en la figura 1.

65

En algunas implementaciones, la unidad de captura 210 recibe audio estéreo capturado usando diversas técnicas de grabación/micrófono. El audio estéreo puede ser capturado por, por ejemplo, equipo de usuario 108,

ordenador portátil 114, sistema de sala de conferencias 118, y cine en casa 120. En un ejemplo, el audio estéreo se captura con dos micrófonos direccionales en la misma ubicación colocados en un ángulo extendido de aproximadamente noventa grados o más. El efecto estéreo resulta de las diferencias de nivel entre canales. En otro ejemplo, el audio estéreo es capturado por dos micrófonos desplazados espacialmente. En algunas implementaciones, los micrófonos desplazados espacialmente son micrófonos omnidireccionales. El efecto estéreo en esta configuración resulta del nivel entre canales y de las diferencias de tiempo entre canales. La distancia entre los micrófonos tiene una influencia considerable en la amplitud estéreo percibida. En otro ejemplo más, el audio es capturado con dos micrófonos direccionales con un desplazamiento de diecisiete centímetros y un ángulo extendido de ciento diez grados. Este sistema a menudo se denomina un sistema de micrófono estéreo de Office de Radiodiffusion Télévision Française ("ORTF"). Otro sistema de captura estéreo más incluye dos micrófonos con características diferentes que están dispuestos de tal forma que la señal de un micrófono es la señal media y la otra la señal lateral. Esta disposición a menudo se denomina grabación del lado medio (M/S). El efecto estéreo de las señales de M/S se basa normalmente en diferencias de nivel entre canales.

En algunas implementaciones, la unidad de captura 210 recibe audio capturado usando técnicas de múltiples micrófonos. En estas implementaciones, la captura de audio implica una disposición de tres o más micrófonos. Esta disposición generalmente es requerida para capturar audio espacial y también puede ser eficaz para realizar supresión de ruido ambiental. A medida que aumenta el número de micrófonos, el número de detalles de una escena espacial que pueden ser capturados por los micrófonos también aumenta. En algunos casos, la precisión de la escena capturada también mejora cuando aumenta el número de micrófonos. Por ejemplo, diversos equipos de usuario (UE) de la figura 1 operados en modo de manos libres puede utilizar múltiples micrófonos para producir una señal de audio mono, estéreo o espacial. Además, un ordenador portátil abierto 114 con múltiples micrófonos puede usarse para producir una captura estéreo. Algunos fabricantes lanzan ordenadores portátiles con dos a cuatro micrófonos de sistemas microelectromecánicos ("MEMS") que permiten captura estéreo. La captura de audio inmersiva con múltiples micrófonos se puede implementar, por ejemplo, en el equipo de usuario de la sala de conferencias 216.

El audio capturado generalmente pasa por una etapa de preprocesamiento antes de ser ingerido en un códec de voz o audio. De este modo, la unidad de preprocesamiento acústico 220 recibe una señal de audio de la unidad de captura 210. En algunas implementaciones, la unidad de preprocesamiento acústico 220 realiza procesamiento de cancelación de ruido y eco, mezcla descendente y mezcla ascendente de canal (por ejemplo, reduciendo o aumentando un número de canales de audio) y/o cualquier tipo de procesamiento espacial. La salida de señal de audio de la unidad de preprocesamiento acústico 220 es generalmente adecuada para codificación y transmisión a otros dispositivos. En algunas implementaciones, el diseño específico de la unidad de preprocesamiento acústico 220 lo realiza un fabricante del dispositivo ya que depende de las especificaciones de la captura de audio con un dispositivo particular. Sin embargo, los requisitos establecidos por especificaciones de interfaz acústica pertinentes pueden establecer límites para estos diseños y asegurar que se cumplen ciertos requisitos de calidad. El preprocesamiento acústico se realiza con un propósito de producir uno o más diferentes tipos de señales de audio o formatos de entrada de audio que soporta un códec IVAS para permitir los diversos casos de uso objetivo de IVAS o niveles de servicio. Dependiendo de los requisitos específicos del servicio IVAS asociados con estos casos de uso, un códec IVAS puede ser requerido para soportar formatos mono, estéreo y espacial.

Generalmente, el formato mono se usa cuando es el único formato disponible, por ejemplo en base al tipo de dispositivo de captura, por ejemplo si las capacidades de captura del dispositivo de envío son limitadas. Para señales de audio estéreo, la unidad de preprocesamiento acústico 220 convierte las señales capturadas en una representación normalizada que cumple convenciones específicas (por ejemplo, convención de ordenación de canales izquierda-derecha). Para la captura estéreo M/S, este proceso puede implicar, por ejemplo, una operación de matriz como para que la señal se represente usando la convención izquierda-derecha. Después del preprocesamiento, la señal estéreo cumple ciertas convenciones (por ejemplo, convención izquierda-derecha). Sin embargo, se elimina información sobre dispositivos de captura estéreo específicos (por ejemplo, número de micrófono y configuración).

Para formatos espaciales, el tipo de señales de entrada espacial o formatos de audio espacial específicos obtenidos después de preprocesamiento acústico pueden depender del tipo de dispositivo de envío y sus capacidades para capturar audio. Al mismo tiempo, los formatos de audio espacial que pueden ser requeridos por los requisitos de servicio IVAS incluyen formato espacial de baja resolución, espacial de alta resolución, formato de audio espacial asistido por metadatos (MASA), y el formato de transporte (HTF) Ambisonics de orden superior ("HOA") o incluso formatos adicionales de audio espacial. La unidad de preprocesamiento acústico 220 de un dispositivo de envío con capacidades de audio espacial, de este modo, debe estar preparada para proporcionar una señal de audio espacial en formato adecuado que cumpla estos requisitos.

Los formatos espaciales de baja resolución incluyen WXY espacial, Ambisonics de primer orden ("FOA") y otros formatos. El formato WXY espacial se refiere a una representación de audio en formato B plano de primer orden de tres canales, con el componente de altura omitido (Z). Este formato es útil para escenarios de conferencias inmersivas y telefonía inmersiva eficiente con tasa de bits donde los requisitos de resolución espacial no son muy

altos y donde el componente de altura espacial puede considerarse irrelevante. El formato es especialmente útil para teléfonos de conferencia, ya que permite a clientes receptores realizar una renderización inmersiva de la escena de la conferencia capturada en una sala de conferencias con múltiples participantes. Asimismo, el formato es para uso de servidores de conferencias que disponen espacialmente a participantes de la conferencia en una sala de reuniones virtual. Por el contrario, FOA contiene el componente de altura (Z) como la señal del cuarto componente. Representaciones FOA son relevantes para aplicaciones de VR de baja tasa.

Formatos espaciales de alta resolución incluyen formatos espaciales basados en escena, objeto y canal. Dependiendo del número de señales de componentes de audio involucradas, cada uno de estos formatos permite representar audio espacial con una resolución virtualmente ilimitada. Por diversas razones (por ejemplo, limitaciones de tasa de bits y limitaciones de complejidad), sin embargo, existen limitaciones prácticas para relativamente pocas señales de componente (por ejemplo, doce). Adicionales formatos espaciales incluyen o pueden depender de formatos MASA o HTF.

Requerir que un dispositivo que soporta IVAS soporte el gran número y variedad de formatos de entrada de audio discutidos anteriormente puede resultar en costos sustanciales en términos de complejidad, huella de memoria, pruebas de implementación, y mantenimiento. Sin embargo, no todos los dispositivos tendrán la capacidad o se beneficiarían de soportar todos los formatos de audio. Por ejemplo, puede haber dispositivos habilitados para IVAS que solo soportan estéreo, pero no soportan captura espacial. Otros dispositivos pueden solo soportar entrada espacial de baja resolución, mientras que una clase adicional de dispositivos puede soportar solo captura HOA. De este modo, diferentes dispositivos sólo harían uso de ciertos subconjuntos de los formatos de audio. Por lo tanto, si el códec IVAS tuviera que soportar codificación directa de todos los formatos de audio, el códec IVAS se volvería innecesariamente complejo y costoso.

Para resolver este problema, el sistema 200 de la figura 2A incluye una unidad de simplificación 230. La unidad de preprocesamiento acústico 220 transfiere la señal de audio a la unidad de simplificación 130. En algunas implementaciones, la unidad de preprocesamiento acústico 220 genera metadatos acústicos que se transfieren a la unidad de simplificación 230 junto con la señal de audio. Los metadatos acústicos pueden incluir datos relacionados con la señal de audio (por ejemplo, metadatos de formato tales como mono, estéreo, espacial). Los metadatos acústicos también pueden incluir datos de cancelación de ruido y otros datos adecuados, por ejemplo, relacionados con las propiedades físicas o geométricas de la unidad de captura 210.

La unidad de simplificación 230 convierte diversos formatos de entrada soportados por un dispositivo a un conjunto común reducido de formatos de ingesta de códec. Por ejemplo, el códec IVAS puede soportar tres formatos de ingesta: mono, estéreo, y espacial. Mientras los formatos mono y estéreo son similares o idénticos a los formatos respectivos como producidos por la unidad de preprocesamiento acústico, el formato espacial puede ser un formato "intermedio". Un formato intermedio es un formato que puede representar con precisión cualquier señal de audio espacial obtenida de la unidad de preprocesamiento acústico 220 y discutida anteriormente. Esto incluye audio espacial representado en cualquier formato basado en escena, objeto y canal (o una combinación de los mismos). En algunas implementaciones, el formato intermedio puede representar la señal de audio como un número de objetos en una escena de audio y un número de canales para transportar información espacial para esa escena de audio. Además, el formato intermedio puede representar MASA, HTF u otros formatos de audio espacial. Un formato intermedio espacial adecuado puede representar audio espacial como m objetos y enésima orden HOA ("mObj+HOAn"), donde m y n son números enteros bajos, incluido cero.

El proceso 300 de la figura 3 ilustra acciones ejemplares para transformar datos de audio de un primer formato a un segundo formato. En 302, la unidad de simplificación 230 recibe una señal de audio, por ejemplo, de la unidad de preprocesamiento acústico 220. Como se discutió anteriormente, la señal de audio recibida de la unidad de preprocesamiento acústico 220 puede ser una señal que tuvo procesamiento de cancelación de ruido y eco realizado así como procesamiento de mezcla descendente y de mezcla ascendente de canal realizado, por ejemplo reduciendo o aumentando un número de canales de audio. En algunas implementaciones, la unidad de simplificación 230 recibe metadatos acústicos junto con la señal de audio. Los metadatos acústicos pueden incluir indicación de formato, y otra información como se discutió anteriormente.

En 304, la unidad de simplificación 230 determina si la señal de audio está en un primer formato que está soportado o no por una unidad de codificación 240 del dispositivo de audio. Por ejemplo, la unidad de detección de formato de audio 232, como se muestra en la figura 2A, puede analizar la señal de audio recibida de la unidad de preprocesamiento acústico 220 e identificar un formato de la señal de audio. Si la unidad de detección de formato de audio 232 determina que la señal de audio está en un formato mono o estéreo, la unidad de simplificación 230 pasa la señal a la unidad de codificación 240. Sin embargo, si la unidad de detección de formato de audio 232 determina que la señal está en un formato espacial, la unidad de detección de formato de audio 232 pasa la señal de audio a la unidad de transformación 234. En algunas implementaciones, la unidad de detección de formato de audio 232 puede usar los metadatos acústicos para determinar el formato de la señal de audio.

En algunas implementaciones, la unidad de simplificación 230 determina si la señal de audio está en el primer

formato determinando un número, configuración o posición de dispositivos de captura de audio (por ejemplo, micrófonos) usados para capturar la señal de audio. Por ejemplo, si la unidad de detección de formato de audio 232 determina que la señal de audio es capturada por un único dispositivo de captura (por ejemplo, un único micrófono), la unidad de detección de formato de audio 232 puede determinar que es una señal mono. Si la

5 unidad de detección de formato de audio 232 determina que la señal de audio es capturada por dos dispositivos de captura en un ángulo específico entre sí, la unidad de detección de formato de audio 232 puede determinar que la señal es una señal estéreo.

La figura 4 es un diagrama de flujo de acciones ejemplares para determinar si una señal de audio está en un formato soportado por la unidad de codificación, de acuerdo con algunas realizaciones de la presente divulgación. En 402, la unidad de simplificación 230 accede a la señal de audio. Por ejemplo, la unidad de detección de formato de audio 232 puede recibir la señal de audio como entrada. En 404, la unidad de simplificación 230 determina la configuración de captura acústica del dispositivo de audio, por ejemplo, un número de micrófonos y su configuración posicional usada para capturar la señal de audio. Por ejemplo,

10 15 la unidad de detección de formato de audio 232 puede analizar la señal de audio y determinar que se colocaron tres micrófonos en diferentes ubicaciones dentro de un espacio. En algunas implementaciones, la unidad de detección de formato de audio 232 puede usar metadatos acústicos para determinar la configuración de captura acústica. Es decir, la unidad de preprocesamiento acústico 220 puede crear metadatos acústicos que indiquen la posición de cada dispositivo de captura y el número de dispositivos de captura. Los metadatos también pueden contener descripciones de propiedades de audio detectadas, tales como dirección o directividad de una fuente de sonido. En 406, la unidad de simplificación 230 compara la configuración de captura acústica con una o más configuraciones de captura acústica almacenadas. Por ejemplo, las configuraciones de captura acústica almacenadas pueden incluir un número y posición de cada micrófono para identificar una configuración específica (por ejemplo, mono, estéreo o espacial). La unidad de simplificación 230 compara cada una de esas

20 25 configuraciones de captura acústica con la configuración de captura acústica de la señal de audio.

En 408, la unidad de simplificación 230 determina si la configuración de captura acústica coincide con una configuración de captura acústica almacenada asociada con un formato espacial. Por ejemplo, la unidad de simplificación 230 puede determinar un número de micrófonos usados para capturar la señal de audio y sus ubicaciones en un espacio. La unidad de simplificación 230 puede comparar esos datos con configuraciones conocidas almacenadas para formatos espaciales. Si la unidad de simplificación 230 determina que no hay coincidencia con un formato espacial, que puede ser una indicación de que el formato de audio es mono o estéreo, el proceso 400 pasa a 412, donde la unidad de simplificación 230 transfiere la señal de audio a una unidad de codificación 240. Sin embargo, si la unidad de simplificación 230 identifica el formato de audio como perteneciente al conjunto de formatos espaciales, el proceso 400 pasa a 410, donde la unidad de simplificación 230 convierte la señal de audio a un formato intermedio.

30 35

Volviendo a hacer referencia a la figura 3, en 306, la unidad de simplificación 230, de acuerdo con determinar que la señal de audio está en un formato que no está soportado por la unidad de codificación, convierte la señal de audio en un segundo formato que está soportado por la unidad de codificación. Por ejemplo, la unidad de transformación 234 puede transformar la señal de audio a un formato intermedio. El formato intermedio representa con precisión una señal de audio espacial representada originalmente en cualquier formato basado en escena, objeto y canal (o una combinación de los mismos). Además, el formato intermedio puede representar MASA, HTF u otro formato adecuado. Por ejemplo, un formato que puede servir como formato intermedio espacial puede representar audio como m objetos y enésima orden HOA ("mObj+HOAn"), donde m y n son números enteros bajos, incluido cero. El formato intermedio puede de este modo implicar la representación del audio con formas de onda (señales) y metadatos que pueden capturar propiedades explícitas de la señal de audio.

40 45

En algunas implementaciones, la unidad de transformación 234, cuando convierte la señal de audio en el segundo formato, genera metadatos para la señal de audio. Los metadatos pueden estar asociados con una porción de la señal de audio en el segundo formato, por ejemplo, metadatos de objetos que incluyen posiciones de uno o más objetos. Otro ejemplo es cuando el audio se capturó usando un conjunto propietario de dispositivos de captura y donde el número y configuración de los dispositivos no están soportados o no están representados de manera eficiente por la unidad de codificación y/o el formato intermedio. En tales casos, la unidad de transformación 234 puede generar metadatos. Los metadatos pueden incluir al menos uno de metadatos de transformación o metadatos acústicos. Los metadatos de transformación pueden incluir un subconjunto de metadatos asociado con una porción del formato que no está soportado por el proceso de codificación y/o el formato intermedio. Por ejemplo, los metadatos de transformación pueden incluir configuraciones del dispositivo para la configuración de captura (por ejemplo, micrófono) y/o configuraciones de dispositivo para la configuración del dispositivo de salida (por ejemplo, altavoz) cuando la señal de audio se reproduce en un sistema que está configurado para emitir específicamente el audio capturado por la configuración propietaria. Los metadatos, que se originan en la unidad de preprocesamiento acústico 220 y/o en la unidad de transformación 234, también pueden incluir metadatos acústicos, que describen ciertas propiedades de la señal de audio tales como una dirección espacial desde la que llega el sonido capturado, una directividad o una difusividad del sonido. En este ejemplo, se puede determinar que el audio es espacial, en formato espacial, aunque representado como una

50 55 60 65

señal mono o estéreo con metadatos adicionales. En este caso, las señales mono o estéreo y los metadatos se propagan al codificador 240.

5 En 308, la unidad de simplificación 230 transfiere la señal de audio en el segundo formato a la unidad de codificación. Como se ilustra en la figura 2A, si la unidad de detección de formato de audio 232 determina que el audio está en formato mono o estéreo, la unidad de detección de formato de audio 232 transfiere la señal de audio a la unidad de codificación. Sin embargo, si la unidad de detección de formato de audio 232 determina que la señal de audio está en un formato espacial, la unidad de detección de formato de audio 232 transfiere la señal de audio a la unidad de transformación 234. La unidad de transformación 234, después de transformar el audio espacial a, por ejemplo, el formato intermedio, transfiere la señal de audio a la unidad de codificación 240. En algunas implementaciones, la unidad de transformación 234 transfiere metadatos de transformación y metadatos acústicos, además de la señal de audio, a la unidad de codificación 240.

15 La unidad de codificación 240 recibe la señal de audio en el segundo formato (por ejemplo, el formato intermedio) y codifica, la señal de audio en el segundo formato, en un formato de transporte. La unidad de codificación 240 propaga la señal de audio codificada a alguna entidad de envío que la transmite a un segundo dispositivo. En algunas implementaciones, la unidad de codificación 240 o entidad posterior almacena la señal de audio codificada para su posterior transmisión. La unidad de codificación 240 puede recibir la señal de audio en formato mono, estéreo o intermedio y codificar esas señales para el transporte de audio. Si la señal de audio está en el formato intermedio y la unidad de codificación recibe metadatos de transformación y/o metadatos acústicos desde la unidad de simplificación 230, la unidad de codificación transfiere los metadatos de transformación y/o metadatos acústicos al segundo dispositivo. En algunas implementaciones, la unidad de codificación 240, codifica los metadatos de transformación y/o metadatos acústicos en una señal específica que el segundo dispositivo puede recibir y descodificar. Entonces, la unidad de codificación envía la señal de audio codificada al transporte de audio para transportarla a uno o más otros dispositivos. De este modo, cada dispositivo (por ejemplo, de los dispositivos en la figura 1) es capaz de codificar la señal de audio en el segundo formato (por ejemplo, el formato intermedio), pero los dispositivos generalmente no son capaces de codificar la señal de audio en el primer formato.

30 En una realización, la unidad de codificación 240 (por ejemplo, el códec IVAS descrito anteriormente) opera con señales de audio mono, estéreo o espacial proporcionadas por la etapa de simplificación. La codificación se realiza dependiendo de una selección de modo de códec que puede ser en base a uno o más de los niveles de servicio IVAS negociados, las capacidades del dispositivo del lado de envío y recepción, y la tasa de bits disponible.

35 El nivel de servicio puede, por ejemplo, incluir telefonía estéreo IVAS, conferencias inmersivas IVAS, transmisión de VR generada por el usuario IVAS, u otro nivel de servicio adecuado. Un determinado formato de audio (mono, estéreo, espacial) puede ser asignado a un nivel de servicio IVAS específico para el cual se elige un modo adecuado de funcionamiento de códec IVAS.

40 Además, el modo de funcionamiento del códec IVAS se puede seleccionar en respuesta a las capacidades del lado del dispositivo de envío y recepción. Por ejemplo, dependiendo de las capacidades del dispositivo de envío, la unidad de codificación 240 puede ser incapaz de acceder a una señal de ingesta espacial, por ejemplo, porque la unidad de codificación 240 solo está provista de una señal mono o estéreo. Además, un intercambio de capacidad de extremo a extremo o una solicitud de modo de códec correspondiente puede indicar que el extremo de recepción tiene ciertas limitaciones de renderizado que hacen innecesario codificar y transmitir una señal de audio espacial o viceversa. En otro ejemplo, otro dispositivo puede solicitar audio espacial.

50 En algunas implementaciones, un intercambio de capacidades de extremo a extremo no puede resolver completamente las capacidades del dispositivo remoto. Por ejemplo, el punto de codificación puede no tener información sobre si la unidad de descodificación, a veces denominada como descodificador, será para un único altavoz mono, altavoces estéreo o si se renderizará de forma binaural. El escenario de renderizado real puede variar durante una sesión de servicio. Por ejemplo, el escenario de renderizado puede cambiar si cambia el equipo de reproducción conectado. En un ejemplo, puede no haber intercambio de capacidades de extremo a extremo porque el dispositivo receptor no está conectado durante la sesión de codificación IVAS. Esto puede producirse en el servicio de correo de voz o en los servicios de transmisión de contenido de realidad virtual (generados por el usuario). Otro ejemplo en el que las capacidades del dispositivo de recepción se desconocen o no se pueden resolver debido a ambigüedades, es un codificador único que necesita soportar múltiples puntos finales. Por ejemplo, en una conferencia IVAS o distribución de contenido de realidad virtual, un punto final puede usar unos auriculares y otro final puede renderizar a altavoces estéreo.

65 Una forma de abordar este problema es asumir la menor capacidad posible del dispositivo de recepción y seleccionar un modo de funcionamiento de códec IVAS correspondiente, que, en ciertos casos, puede ser mono. Otra forma de abordar este problema es requerir que el descodificador IVAS, incluso si el codificador es operado en un modo que soporta audio espacial o estéreo, deduzca una señal de audio descodificada que se puede renderizar en dispositivos con capacidad de audio inferior respectivamente. Es decir, una señal codificada como

señal de audio espacial también debería ser descodificable tanto para renderizado estéreo como mono. Del mismo modo, una señal codificada como estéreo también debería ser descodificable para renderización mono.

5 Por ejemplo, en conferencias IVAS, un servidor de llamadas solo debería necesitar realizar una única codificación y enviar la misma codificación a múltiples puntos finales, algunos de los cuales pueden ser binaurales y otros pueden ser estéreo. De este modo, una única codificación de dos canales puede soportar tanto la renderización en, por ejemplo, el ordenador portátil 114 y el sistema de sala de conferencias 118 con altavoces estéreo como la renderización inmersiva con presentación binaural en el dispositivo de usuario 110 y el equipo de realidad virtual 122. De este modo, una única codificación puede soportar ambas consecuencias simultáneamente. Como resultado, una implicación es que la codificación de dos canales soporta tanto la retransmisión de altavoces estéreo como la retransmisión renderizada binaural con una única codificación.

15 Otro ejemplo implica extracción mono de alta calidad. El sistema puede soportar extracción de una señal mono de alta calidad de una señal de audio estéreo o espacial codificada. En algunas implementaciones, es posible extraer un flujo de bits de códec de servicios de voz mejorada ("EVS") para descodificación mono, por ejemplo, usando el descodificador EVS estándar.

20 Alternativa o adicionalmente al nivel de servicio y las capacidades del dispositivo, la tasa de bits disponible es otro parámetro que puede controlar la selección del modo de códec. En algunas implementaciones, la tasa de bits necesita aumentar con la calidad de la experiencia que se puede ofrecer en el extremo de recepción y con el número asociado de componentes de la señal de audio. Con las tasas de bits más bajas, sólo es posible la renderización de audio mono. El códec EVS ofrece funcionamiento mono hasta 5,9 kilobits por segundo. A medida que aumenta la tasa de bits, se puede lograr servicio de mayor calidad. Sin embargo, la calidad de codificación ("QoE") sigue siendo limitada debido a funcionamiento y renderización sólo mono. El siguiente nivel superior de QoE es posible con estéreo de dos canales (convencional). Sin embargo, el sistema requiere una tasa de bits más alta que la tasa de bits mono más baja para ofrecer calidad útil, porque ahora hay dos componentes de señal de audio para transmitir. La experiencia de sonido espacial requiere una QoE más alta que la estéreo. En el extremo inferior del rango de tasa de bits, esta experiencia se puede habilitar con una representación binaural de la señal espacial que puede denominarse "estéreo espacial". Estéreo espacial se basa en pre-renderización binaural del lado del codificador (con apropiadas funciones de transferencia relacionadas con la cabeza ("HRTF")) de la ingesta de señal de audio espacial en el codificador (por ejemplo, la unidad de codificación 240) y es probablemente la representación espacial más compacta porque se compone de sólo dos señales de componentes de audio. Debido a que el estéreo espacial transporta más información perceptual, la tasa de bits requerida para lograr una calidad suficiente probablemente es mayor que la tasa de bits necesaria para una señal estéreo convencional. Sin embargo, la representación de estéreo espacial puede tener limitaciones en relación con la personalización de la renderización en el extremo de recepción. Estas limitaciones pueden incluir una restricción al renderizado con auriculares, al uso de un conjunto preseleccionado de HRTF, o al renderizado sin seguimiento de la cabeza. Incluso una QoE mayor a tasas de bits más altas está habilitada por un modo de códec para codificar la señal de audio en un formato espacial que no depende de pre-renderización binaural en el codificador y más bien representa el formato intermedio espacial ingerido. Dependiendo de la tasa de bits, el número de señales de componentes de audio representadas de ese formato puede ser ajustado. Por ejemplo, esto puede dar como resultado una representación espacial más o menos potente que va desde el WXY espacial hasta formatos de audio espacial de alta resolución, como se discutió anteriormente. Esto permite una resolución espacial de baja a alta dependiendo de la tasa de bits disponible y ofrece la flexibilidad para abordar una amplia gama de escenarios de renderizado, incluido el binaural con seguimiento de cabeza. Este modo se denomina modo "espacial versátil".

50 En algunas implementaciones, el códec IVAS opera a las tasas de bits del códec EVS, es decir, en un rango de 5,9 a 128 kilobits por segundo. Para funcionamiento estéreo de baja tasa con transmisión en entornos con ancho de banda limitado, se pueden requerir tasas de bits de hasta 13,2 kbps. Este requisito podría estar sujeto a viabilidad técnica usando un códec IVAS particular y posiblemente aún permitir una operación atractiva de servicio IVAS. Para funcionamiento estéreo espacial de baja tasa con transmisión en entornos con ancho de banda limitado, las tasas de bits más bajas que permiten renderización espacial y renderización estéreo simultánea pueden ser posibles hasta 24,4 kilobits por segundo. Para funcionamiento en modo espacial versátil, es probable que sea posible una resolución espacial baja (espacial WXY, FOA) de hasta 24,4 kilobits por segundo, a la que, sin embargo, se podría lograr la calidad de audio como en el modo de funcionamiento estéreo espacial.

60 Haciendo referencia ahora a la figura 2B, un dispositivo de recepción recibe un flujo de transporte de audio que incluye la señal de audio codificada. La unidad de descodificación 250 del dispositivo de recepción recibe la señal de audio codificada (por ejemplo, en un formato de transporte codificado por un codificador) y la descodifica. En algunas implementaciones, la unidad de descodificación 250 recibe la señal de audio codificada en uno de cuatro modos: mono, estéreo (convencional), estéreo espacial o espacial versátil. La unidad de descodificación 250 transfiere la señal de audio a la unidad de renderizado 260. La unidad de renderizado 260 recibe la señal de audio de la unidad de descodificación 250 para renderizar la señal de audio. Es notable que generalmente no hay necesidad de recuperar el primer formato de audio espacial original ingerido en la unidad

de simplificación 230. Esto permite ahorros significativos en la complejidad del descodificador y/o en la huella de memoria de una implementación de descodificador IVAS.

5 La figura 5 es un diagrama de flujo de acciones ejemplares para transformar una señal de audio a un formato de reproducción disponible, de acuerdo con algunas realizaciones de la presente divulgación. En 502, la unidad de renderizado 260 recibe una señal de audio en un primer formato. Por ejemplo, la unidad de renderizado 260 puede recibir la señal de audio en los siguientes formatos: mono, estéreo convencional, estéreo espacial, espacial versátil. En algunas implementaciones, la unidad de selección de modo 262 recibe la señal de audio. La unidad de selección de modo 262 identifica el formato de la señal de audio. Si la unidad de selección de modo 10 262 determina que el formato de la señal de audio está soportado por la configuración de reproducción, la unidad de selección de modo 262 transfiere la señal de audio al renderizador 264. Sin embargo, si la unidad de selección de modo determina que la señal de audio no está soportada, la unidad de selección de modo realiza un procesamiento adicional. En algunas implementaciones, la unidad de selección de modo 262 selecciona una unidad de descodificación diferente.

15 En 504, la unidad de renderizado 260 determina si el dispositivo de audio es capaz de reproducir la señal de audio en un segundo formato que está soportado por la configuración de reproducción. Por ejemplo, la unidad de renderizado 260 puede determinar (por ejemplo, en base al número de altavoces y/u otros dispositivos de salida y su configuración y/o metadatos asociados con el audio descodificado) que la señal de audio está en formato 20 estéreo espacial, pero el dispositivo de audio es capaz de reproducir el audio recibido solo en mono. En algunas implementaciones, no todos los dispositivos en el sistema (por ejemplo, como se ilustra en la figura 1) son capaces de reproducir la señal de audio en el primer formato, pero todos los dispositivos son capaces de reproducir la señal de audio en un segundo formato.

25 En 506, la unidad de renderizado 260, en base a determinar que el dispositivo de salida es capaz de reproducir la señal de audio en el segundo formato, adapta la descodificación de audio para producir una señal en el segundo formato. Como alternativa, la unidad de renderizado 260 (por ejemplo, la unidad de selección de modo 262 o el renderizador 264) puede usar metadatos, por ejemplo, metadatos acústicos, metadatos de transformación o una combinación de metadatos acústicos y metadatos de transformación, para adaptar la señal 30 de audio al segundo formato. En 508, la unidad de renderizado 260 transfiere la señal de audio en el primer formato soportado o en el segundo formato soportado para salida de audio (por ejemplo, a un controlador que interactúa con un sistema de altavoces).

35 En algunas implementaciones, la unidad de renderizado 260 convierte la señal de audio al segundo formato usando metadatos que incluyen una representación de una porción de la señal de audio no soportada por el segundo formato en combinación con la señal de audio en el primer formato. Por ejemplo, si la señal de audio se recibe en formato mono y los metadatos incluyen información de formato espacial, la unidad de renderizado puede convertir la señal de audio en el formato mono a un formato espacial usando los metadatos.

40 La figura 6 es otro diagrama de bloques de acciones ejemplares para transformar una señal de audio a un formato de reproducción disponible, de acuerdo con algunas realizaciones de la presente divulgación. En 602, la unidad de renderizado 260 recibe una señal de audio en un primer formato. Por ejemplo, la unidad de renderizado 260 puede recibir la señal de audio en un formato mono, estéreo convencional, estéreo espacial o 45 espacial versátil. En algunas implementaciones, la unidad de selección de modo 262 recibe la señal de audio. En 604, la unidad de renderizado 260 recupera las capacidades de salida de audio (por ejemplo, capacidades de reproducción de audio) del dispositivo de audio. Por ejemplo, la unidad de reproducción 260 puede recuperar un número de altavoces, su configuración de posición, y/o la configuración de otros dispositivos de reproducción disponibles para reproducción. En algunas implementaciones, la unidad de selección de modo 262 realiza la operación de recuperación.

50 En 606, la unidad de renderizado 260 compara las propiedades de audio del primer formato con las capacidades de salida del dispositivo de audio. Por ejemplo, la unidad de selección de modo 262 puede determinar que la señal de audio está en un formato estéreo espacial (por ejemplo, en base a metadatos acústicos, metadatos de transformación, o una combinación de metadatos acústicos y los metadatos de transformación) y el dispositivo 55 de audio es capaz de reproducir la señal de audio solo en formato estéreo convencional a través de un sistema de altavoces estéreo (por ejemplo, en base a la configuración de los altavoces y otros dispositivos de salida). La unidad de renderizado 260 puede comparar las propiedades de audio del primer formato con las capacidades de salida del dispositivo de audio. En 608, la unidad de renderizado 260 determina si las capacidades de salida del dispositivo de audio coinciden con las propiedades de salida de audio del primer formato. Si las capacidades de salida del dispositivo de audio no coinciden con las propiedades de audio del primer formato, el proceso 600 pasa a 610 donde la unidad de renderizado 260 (por ejemplo, la unidad de selección de modo 262) realiza acciones para obtener la señal de audio en un segundo formato. Por ejemplo, la unidad de renderizado 260 puede adaptar la unidad de descodificación 250 para descodificar el audio recibido en el segundo formato o la 60 unidad de renderizado puede usar metadatos acústicos, metadatos de transformación o una combinación de metadatos acústicos y metadatos de transformación para transformar el audio de formato estéreo espacial al segundo formato soportado, que es estéreo convencional en el ejemplo dado. Si las capacidades de salida del 65

dispositivo de audio coinciden con las propiedades de salida de audio del primer formato, o después de la operación de transformación 610, el proceso 600 pasa a 612, donde la unidad de renderizado 260 (por ejemplo, usando el renderizador 264) transfiere la señal de audio, que ahora se garantiza que está soportada, al dispositivo de salida.

5

La figura 7 muestra un diagrama de bloques de un sistema de ejemplo 700 adecuado para implementar realizaciones de ejemplo de la presente divulgación. Como se muestra, el sistema 700 incluye una unidad central de procesamiento (CPU) 701 que es capaz de realizar diversos procesos de acuerdo con un programa almacenado en, por ejemplo, una memoria de sólo lectura (ROM) 702 o un programa cargado desde, por ejemplo, una unidad de almacenamiento 708 a una memoria de acceso aleatorio (RAM) 703. En la RAM 703, los datos requeridos cuando la CPU 701 realiza los diversos procesos también se almacenan, como se requiere. La CPU 701, la ROM 702 y la RAM 703 están conectadas entre sí a través de un bus 704. Una interfaz de entrada/salida (E/S) 705 también está conectada al bus 704.

10

Los siguientes componentes están conectados a la interfaz de E/S 705: una unidad de entrada 706, que puede incluir un teclado, un ratón o similar; una unidad de salida 707 que puede incluir un visualizador tal como un visualizador de cristal líquido (LCD) y uno o más altavoces; incluyendo la unidad de almacenamiento 708 un disco duro, u otro dispositivo de almacenamiento adecuado; y una unidad de comunicación 709 que incluye una tarjeta de interfaz de red tal como una tarjeta de red (por ejemplo, cableada o inalámbrica).

20

En algunas implementaciones, la unidad de entrada 706 incluye uno o más micrófonos en diferentes posiciones (dependiendo del dispositivo anfitrión) que permiten captura de señales de audio en diversos formatos (por ejemplo, mono, estéreo, espacial, inmersivo, y otros formatos adecuados).

25

En algunas implementaciones, la unidad de salida 707 incluye sistemas con diversos números de altavoces. Como se ilustra en la figura 1, la unidad de salida 707 (dependiendo de las capacidades del dispositivo anfitrión) puede renderizar señales de audio en diversos formatos (por ejemplo, mono, estéreo, inmersivo, binaural, y otros formatos adecuados).

30

unidad de comunicación 709 está configurada para comunicarse con otros dispositivos (por ejemplo, a través de una red). Una unidad 710 también se conecta a la interfaz de E/S 705, como se requiere. Un medio extraíble 711, tal como un disco magnético, un disco óptico, un disco magnetoóptico, una unidad flash u otro medio extraíble adecuado se monta en la unidad 710, de modo que un programa de ordenador leído desde el mismo se instala en la unidad de almacenamiento 708, como se requiere. Una persona experta en la técnica entendería que, aunque se describe que el sistema 700 incluye los componentes descritos anteriormente, en aplicaciones reales es posible agregar, eliminar y/o reemplazar algunos de estos componentes y todas estas modificaciones o alteraciones caen dentro del alcance de la presente divulgación.

35

De acuerdo con realizaciones de ejemplo de la presente divulgación, los procesos descritos anteriormente pueden implementarse como programas de software de ordenador o en un medio de almacenamiento legible por ordenador. Por ejemplo, realizaciones de la presente divulgación incluyen un producto de programa de ordenador que incluye un programa de ordenador incorporado de forma tangible en un medio legible por máquina, incluyendo el programa de ordenador código de programa para realizar métodos. En tales realizaciones, el programa de ordenador puede descargarse y montarse desde la red a través de la unidad de comunicación 709, y/o instalarse desde el medio extraíble 711.

40

45

Generalmente, diversas realizaciones de ejemplo de la presente divulgación pueden ser implementadas en hardware o circuitos de propósito especial (por ejemplo, circuitos de control), software, lógica o cualquier combinación de los mismos. Por ejemplo, la unidad de simplificación 230 y otras unidades discutidas anteriormente pueden ejecutarse mediante los circuitos de control (por ejemplo, una CPU en combinación con otros componentes de la figura 7), de este modo, la circuitería de control puede estar realizando las acciones descritas en esta divulgación. Algunos aspectos pueden implementarse en hardware, mientras que otros aspectos pueden implementarse en firmware o software que puede ejecutarse mediante un controlador, microprocesador u otro dispositivo informático (por ejemplo, circuitería de control). Mientras diversos aspectos de las realizaciones de ejemplo de la presente divulgación se ilustran y describen como diagramas de bloques, diagramas de flujo, o usando alguna otra representación pictórica, se apreciará que los bloques, aparato, sistemas, técnicas o métodos descritos en el presente documento pueden implementarse en, como ejemplos no limitativos, hardware, software, firmware, circuitos o lógica de propósito especial, hardware o controlador u otros dispositivos informáticos de propósito general, o alguna combinación de los mismos.

50

55

60

Además, diversos bloques mostrados en los diagramas de flujo pueden verse como pasos de método, y/o como operaciones que resultan del funcionamiento de un código de programa de ordenador, y/o como una pluralidad de elementos de circuito lógico acoplados construidos para llevar a cabo la o las funciones asociadas. Por ejemplo, realizaciones de la presente divulgación incluyen un producto de programa de ordenador que incluye un programa de ordenador incorporado de forma tangible en un medio legible por máquina, conteniendo el programa de ordenador códigos de programa configurados para llevar a cabo los métodos descritos

65

anteriormente.

- 5 En el contexto de la divulgación, un medio legible por máquina puede ser cualquier medio tangible que pueda contener, o almacenar un programa para su uso por o en conexión con un sistema, aparato o dispositivo de ejecución de instrucciones. El medio legible por máquina puede ser un medio de señal legible por máquina o un medio de almacenamiento legible por máquina. Un medio legible por máquina puede ser no transitorio y puede incluir, pero no limitarse a, un sistema, aparato o dispositivo electrónico, magnético, óptico, electromagnético, de infrarrojo o semiconductor, o cualquier combinación adecuada de los anteriores. Ejemplos más específicos del medio de almacenamiento legible por máquina incluirían una conexión eléctrica que tiene uno o más cables, un 10 disquete de ordenador portátil, un disco duro, una memoria de acceso aleatorio (RAM), una memoria de sólo lectura (ROM), una memoria de sólo lectura programable y borrable (EPROM o memoria flash), una fibra óptica, una memoria de sólo lectura de disco compacto portátil (CD-ROM), un dispositivo de almacenamiento óptico, un dispositivo de almacenamiento magnético, o cualquier combinación adecuada de los anteriores.
- 15 El código de programa de ordenador para llevar a cabo métodos de la presente divulgación puede escribirse en cualquier combinación de uno o más lenguajes de programación. Estos códigos de programas de ordenador pueden proporcionarse a un procesador de un ordenador de propósito general, ordenador de propósito especial, u otro aparato de procesamiento de datos programable que tenga circuitería de control, de tal forma que los 20 códigos de programa, cuando los ejecuta el procesador del ordenador u otro aparato de procesamiento de datos programable, hacen que se implementen las funciones/operaciones especificadas en los diagramas de flujo y/o diagramas de bloques. El código de programa puede ejecutarse completamente en un ordenador, parcialmente en el ordenador, como un paquete de software independiente, parcialmente en el ordenador y parcialmente en un ordenador remoto o completamente en el ordenador o servidor remoto o distribuido en uno o más ordenadores remotos y/o servidores.
- 25

REIVINDICACIONES

1. Un método que comprende:

5 recibir (paso 602), por una unidad de simplificación (230) en un dispositivo de envío, desde una unidad de preprocesamiento acústico (220), señales de audio en una pluralidad de formatos de renderización de audio;

10 recibir (paso 604), por la unidad de simplificación (230), desde un dispositivo de recepción, atributos del dispositivo de recepción, incluyendo los atributos indicaciones de uno o más formatos de audio soportados por el dispositivo de recepción, incluyendo el uno o más formatos de audio al menos uno de un formato mono, un formato estéreo, o un formato espacial;

15 convertir (paso 610), por la unidad de simplificación, las señales de audio en un formato de ingesta que es una representación alternativa de los uno o más formatos de audio; y

proporcionar, por la unidad de simplificación, la señal de audio convertida, a una unidad de codificación (240) para procesamiento aguas abajo,

20 codificar, por la unidad de codificación, la señal de audio de formato de ingesta en una señal de audio codificada en un formato de transporte descodificable por el dispositivo de recepción,

en el que cada una de la unidad de preprocesamiento acústico, la unidad de simplificación y la unidad de codificación comprende uno o más procesadores de ordenador.

25 2. El método de acuerdo con la reivindicación 1, que comprende además:

transmitir la señal de audio codificada al dispositivo de recepción.

30 3. El método de la reivindicación 1 o 2, en el que convertir la señal de audio en el formato de ingesta comprende generar metadatos para la señal de audio, en el que los metadatos comprenden una representación de una porción de la señal de audio.

35 4. El método de la reivindicación 2, que comprende además transmitir la señal de audio codificada al transmitir los metadatos que comprenden una representación de una porción de la señal de audio.

5. El método de la reivindicación 1, en el que el formato de ingesta representa la señal de audio como un número de objetos de audio en una escena de audio de los cuales ambos dependen de un número de canales de audio para transportar información espacial.

40 6. El método de la reivindicación 5, en el que el formato de ingesta comprende además metadatos para transportar una porción adicional de información espacial.

45 7. El método de cualquier reivindicación anterior, en el que el formato de ingesta está caracterizado además por permitir un grado comparable de calidad de experiencia.

8. Un aparato que comprende:

uno o más procesadores de ordenador; y

50 uno o más medios de almacenamiento no transitorios que almacenan instrucciones que, cuando son ejecutadas por el uno o más procesadores de ordenador, hacen que el uno o más procesadores de ordenador realicen operaciones de cualquiera de las reivindicaciones 1 a 7.

9. Un sistema de codificación que comprende:

55 equipo de usuario configurado para capturar señales de audio en una pluralidad de formatos de renderización;

una unidad de preprocesamiento acústico configurada para realizar operaciones que comprenden preprocesar las señales de audio;

60 un codificador; y

una unidad de simplificación configurada para realizar operaciones que comprenden:

65 recibir, desde la unidad de preprocesamiento acústico, las señales de audio;

recibir desde un dispositivo de recepción, atributos del dispositivo de recepción, incluyendo los atributos indicaciones de uno o más formatos de audio soportados por el dispositivo de recepción, incluyendo el uno o más formatos de audio al menos uno de un formato mono, un formato estéreo, o un formato espacial;

- 5 convertir la señal de audio en un formato de ingesta que es una representación alternativa del uno o más formatos de audio; y

transferir la señal de audio en el formato de ingesta al codificador,

- 10 en el que el codificador está configurado para realizar operaciones que comprenden:

codificar la señal de audio de formato de ingesta en una señal de audio codificada en un formato de transporte decodificable por el dispositivo de recepción; y

- 15 almacenar la señal de audio codificada o transmitir la señal de audio codificada a otro dispositivo.

10. El sistema de codificación de la reivindicación 9, en el que convertir la señal de audio al formato de ingesta comprende generar metadatos para la señal de audio, en el que los metadatos comprenden una representación de una porción de la señal de audio.

20

11. El sistema de codificación de la reivindicación 10, comprendiendo además las operaciones del codificador transmitir la señal de audio codificada al transmitir los metadatos que comprenden una representación de una porción de la señal de audio.

25

12. El sistema de codificación de la reivindicación 10, en el que el formato de ingesta representa el audio de la señal de audio como un número de objetos en una escena de audio y un número de canales para transportar información espacial.

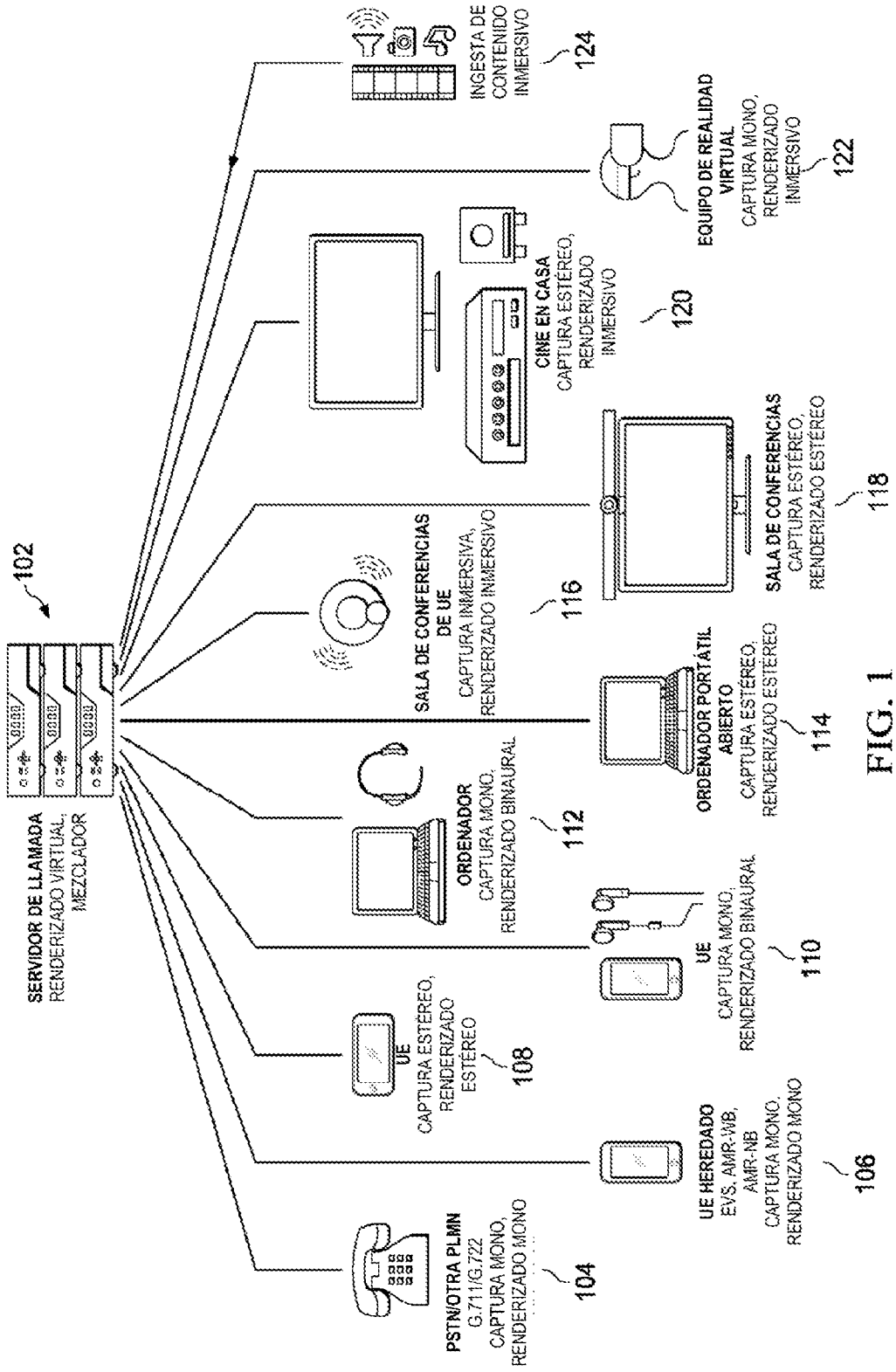


FIG. 1

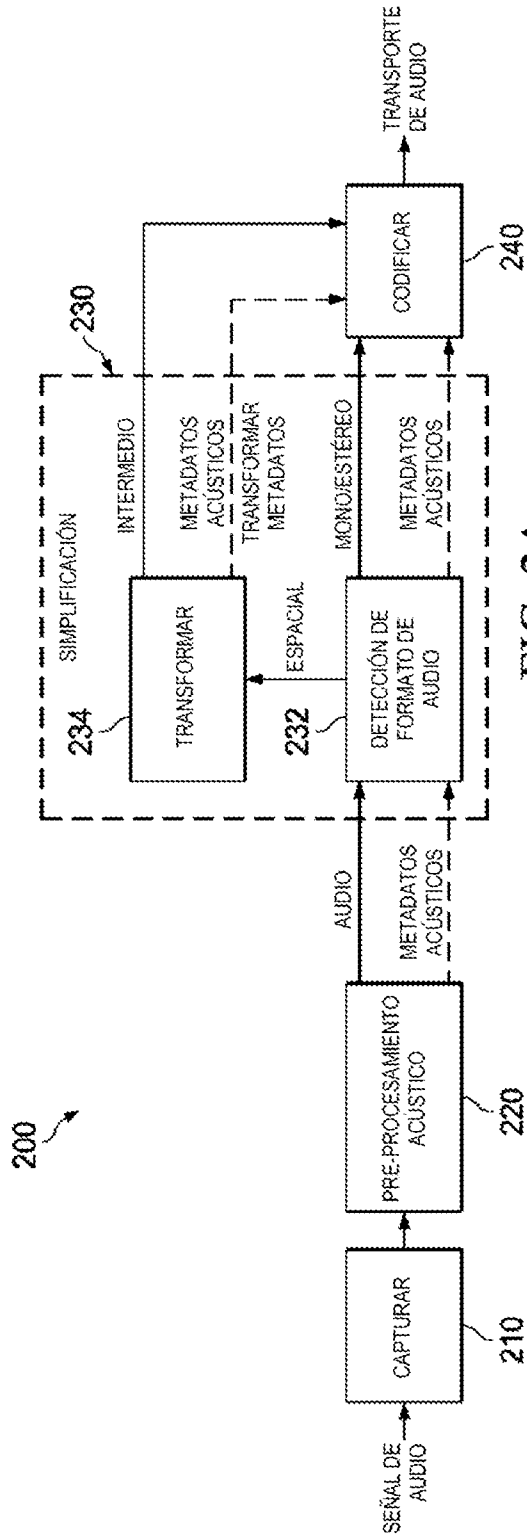


FIG. 2A

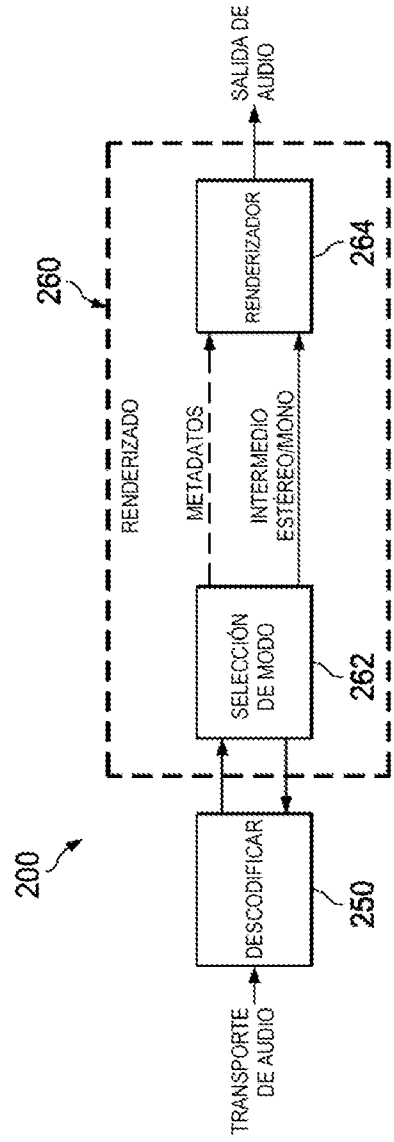


FIG. 2B

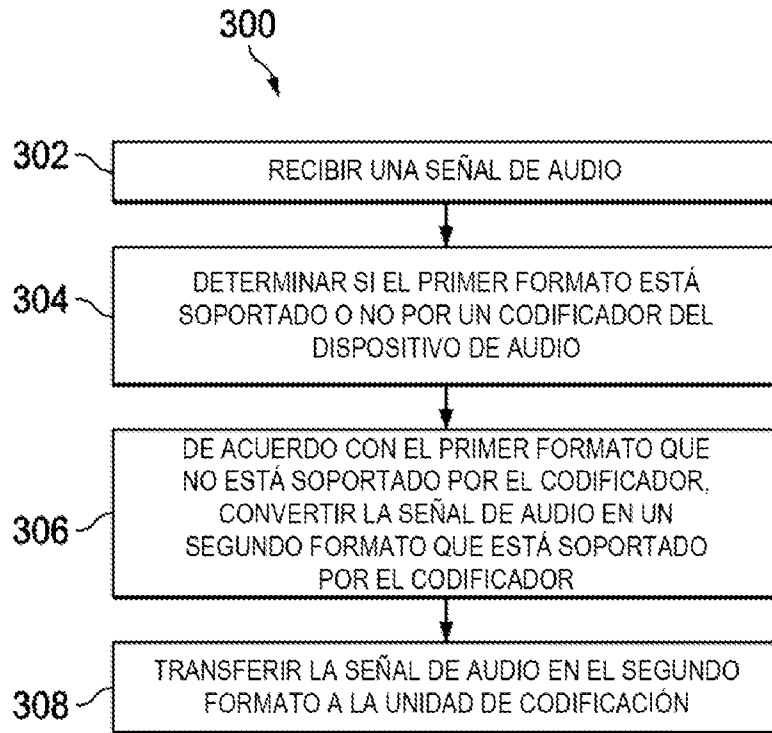


FIG. 3

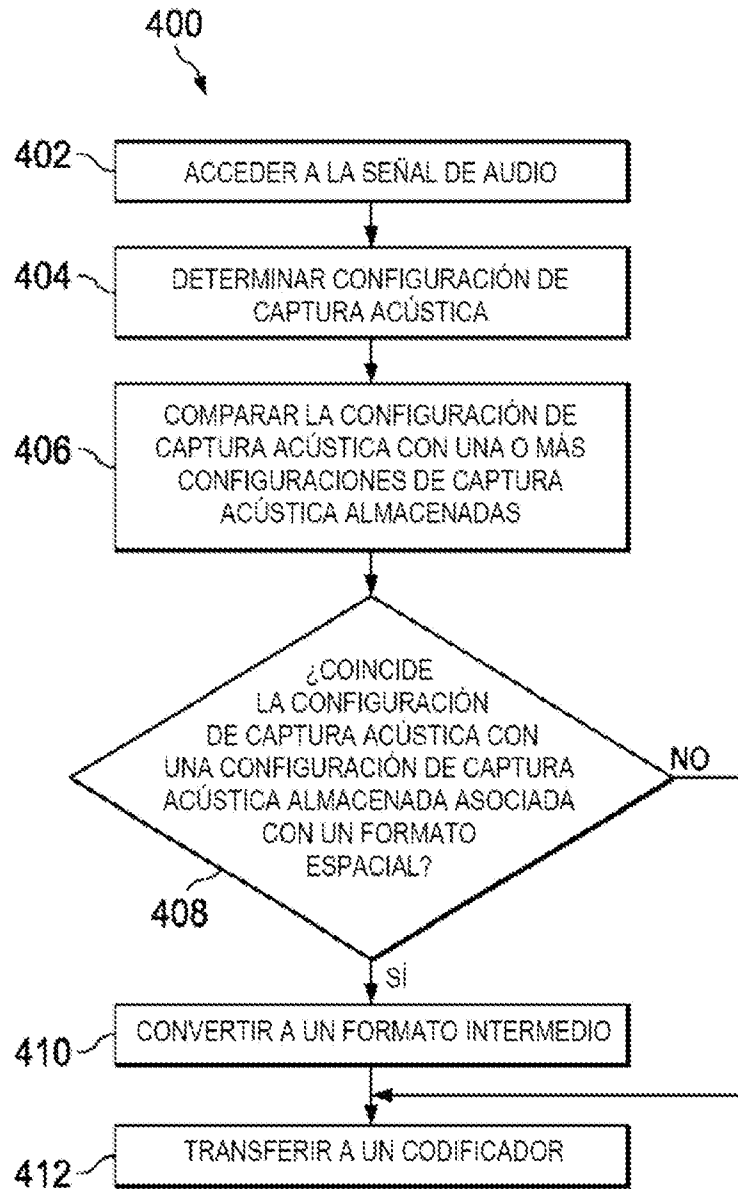


FIG. 4

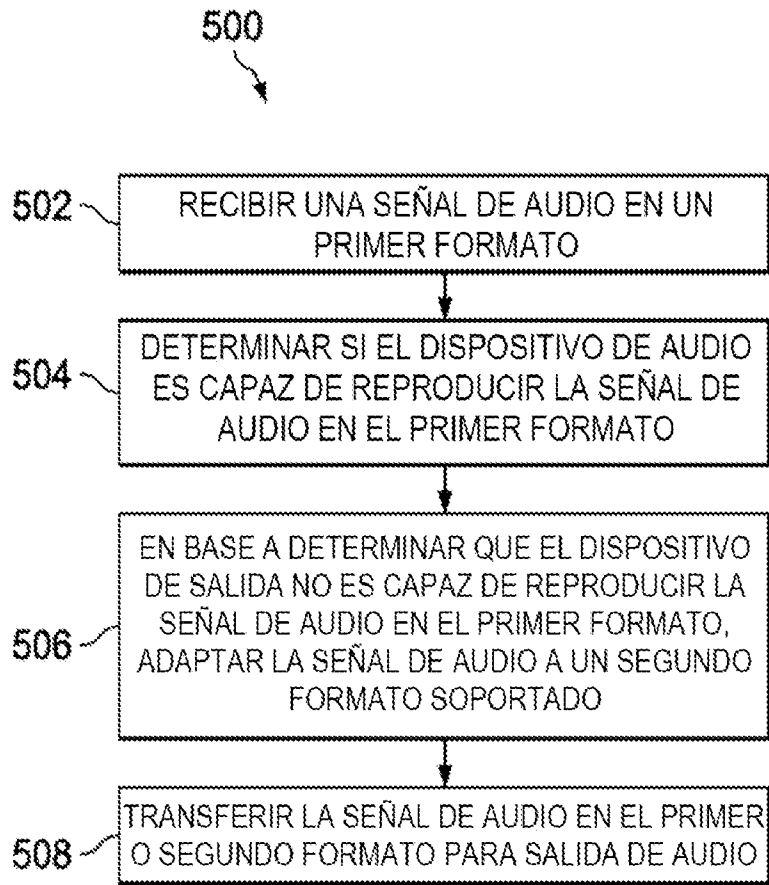


FIG. 5

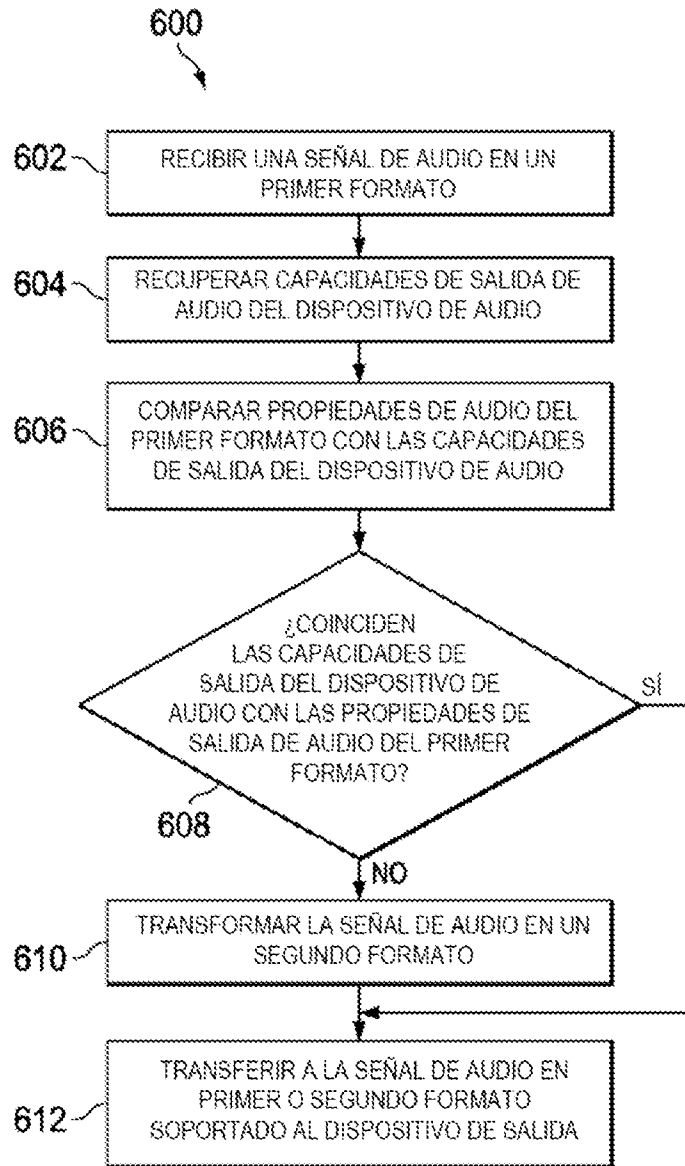


FIG. 6

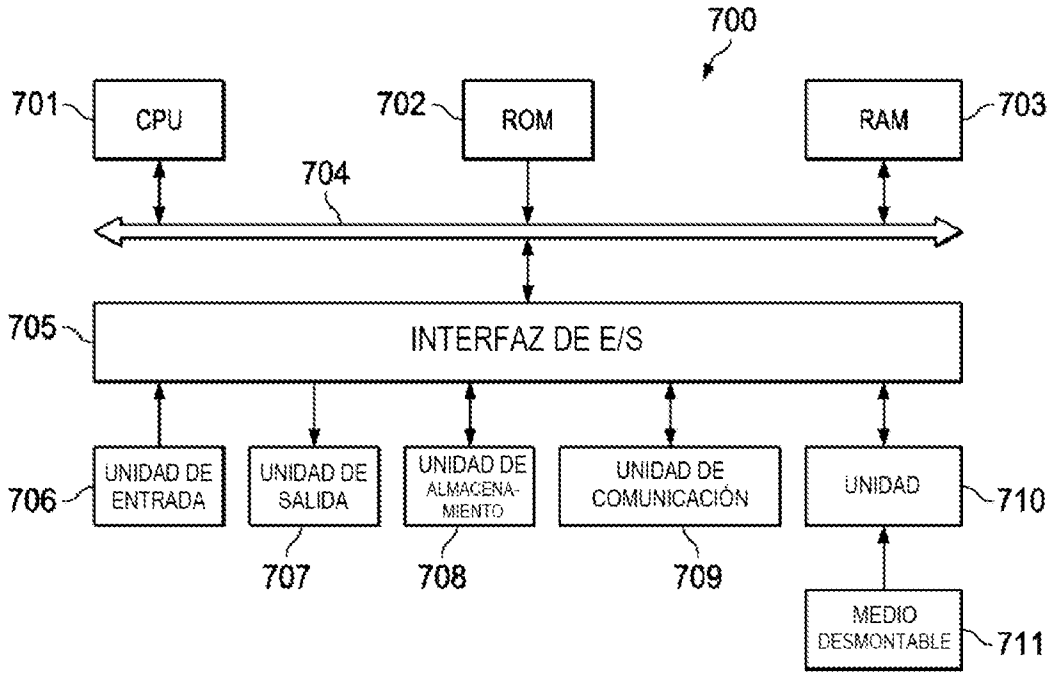


FIG. 7