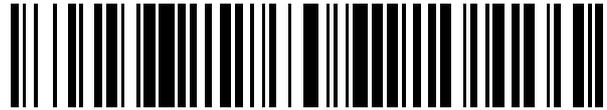


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 736 878**

51 Int. Cl.:

H04M 3/56

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.03.2016 PCT/US2016/025058**

87 Fecha y número de publicación internacional: **13.10.2016 WO16164233**

96 Fecha de presentación y número de la solicitud europea: **30.03.2016 E 16715972 (2)**

97 Fecha y número de publicación de la concesión europea: **17.04.2019 EP 3281396**

54 Título: **Gestión de audio de conferencia**

30 Prioridad:

**05.04.2015 US 201562143154 P
07.04.2015 US 201562144033 P
29.03.2016 US 201615083974**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
08.01.2020

73 Titular/es:

**QUALCOMM INCORPORATED, ATTN-
INTERNATIONAL IP ADMINISTRATION (100.0%)
5775 Morehouse Drive
San Diego, CA 92121-1714, US**

72 Inventor/es:

**ATTI, VENKATRAMAN S.;
SINDER, DANIEL J.;
LEUNG, NIKOLAI y
RAJENDRAN, VIVEK**

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 736 878 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

Gestión de audio de conferencia

5 **II. Campo**

[0001] La presente divulgación se refiere en general a gestión de audio de conferencia.

10 **III. Descripción de la técnica relacionada**

15 [0002] Los avances en la tecnología han dado como resultado dispositivos informáticos más pequeños y más potentes. Por ejemplo, existe actualmente una variedad de dispositivos informáticos personales portátiles, que incluyen dispositivos informáticos inalámbricos, tales como teléfonos inalámbricos portátiles, asistentes digitales personales (PDA) y dispositivos de radiolocalización que son pequeños, ligeros y fáciles de transportar por los usuarios. Más específicamente, los teléfonos inalámbricos portátiles, tales como los teléfonos móviles y los teléfonos de protocolo de Internet (IP), pueden transmitir paquetes de voz y datos a través de redes inalámbricas. Además, muchos de dichos teléfonos inalámbricos incluyen otros tipos de dispositivos que están incorporados en los mismos. Por ejemplo, un teléfono inalámbrico también puede incluir una cámara fotográfica digital, una cámara de video digital, un grabador digital y un reproductor de archivos de audio.

20 [0003] Tres o más participantes asociados con tres o más dispositivos inalámbricos correspondientes pueden usar una "llamada en conferencia" para comunicarse entre sí. Durante la llamada en conferencia, un primer dispositivo inalámbrico puede recibir una "transmisión de audio mixto" desde un mezclador de red. Por ejemplo, el mezclador de red puede recibir transmisiones de audio de cada participante de la llamada en conferencia. Después de recibir las transmisiones de audio, el mezclador de red puede mezclar una transmisión de audio de un segundo dispositivo inalámbrico con una transmisión de audio de un tercer dispositivo inalámbrico para generar una transmisión de audio mixto para proporcionar el primer dispositivo inalámbrico. Para generar la transmisión de audio mixto, el mezclador de red descodifica cada transmisión de audio entrante, mezcla las transmisiones de audio descodificado y recodifica las transmisiones de audio descodificado mixto. Sin embargo, el procesamiento de audio asociado con la generación de la transmisión de audio mixto puede introducir un retardo. Además, debido a que los participantes de la conferencia reciben una transmisión de audio mixto desde un mezclador de red central, puede que no sea posible ajustar el audio del participante individualmente. Por ejemplo, si las partes A, B y C están en una conferencia, la parte A puede ser incapaz de enfatizar la voz de la parte B sobre la voz de la parte C usando técnicas de ajuste de ganancia o procesamiento espacial, porque el mezclador de red proporciona a la parte A una única transmisión mezclada que incluye voz tanto de la parte B como de la parte C.

25 [0004] La Solicitud de Patente Europea con número de publicación EP 1954019 A1 describe sistemas y procedimiento para proporcionar sonido espacial simulado en sesiones de comunicación de voz de grupo en un dispositivo de comunicación inalámbrica que es uno de dos o más conectados a una red de comunicación inalámbrica. Un módulo de control configura el dispositivo de comunicación inalámbrica para recibir una entrada de audio de un usuario del dispositivo de comunicación inalámbrica, generar paquetes de datos de audio desde la entrada de audio y transmitir los paquetes de datos de audio de la red de comunicación inalámbrica. El módulo de control configura además el dispositivo de comunicación inalámbrica para recibir paquetes de datos de audio espacial y reproducir los datos de audio espacial contenidos en los paquetes de audio espacial utilizando un dispositivo de salida de audio estéreo.

30 **IV. Sumario**

35 [0005] Se divulgan los sistemas y procedimientos de gestión de audio durante una conferencia. De acuerdo con una implementación particular, para habilitar el ajuste de audio de los participantes individuales de la conferencia, un mezclador de red puede transmitir transmisiones de audio individuales a los participantes en lugar de generar una transmisión mixta. Por ejemplo, en una conferencia a tres partes, la parte A puede recibir transmisiones de audio separadas de la parte B y la parte C, respectivamente. Sin embargo, debido a la variación en las condiciones de la red, las transmisiones de la parte B y la parte C pueden no estar sincronizados entre sí, lo cual puede llevar a un dispositivo de la parte A de salida a ser temporalmente erróneo. Para ilustrar, si la transmisión de la parte B se retarda, la parte A puede escuchar una pregunta formulada por la parte B después de escuchar una respuesta a esa pregunta por la parte C.

40 [0006] De acuerdo con las técnicas descritas en el presente documento, un dispositivo móvil puede utilizar la gestión de memoria intermedia para sincronizar transmisiones recibidas desde diferentes dispositivos. Por ejemplo, se puede proporcionar una primera transmisión de audio de un primer participante de la conferencia a una primera memoria intermedia (por ejemplo, una memoria intermedia de eliminación de fluctuaciones) del dispositivo móvil, y se puede proporcionar una segunda transmisión de audio de un segundo participante de la conferencia a una segunda memoria intermedia del dispositivo móvil. Al recibir la primera y la segunda transmisiones de audio, un controlador de retardo puede comparar las marcas de tiempo de la primera transmisión de audio con las marcas de tiempo de la segunda transmisión de audio para sincronizar el audio del primer participante de la conferencia con el audio del segundo participante de la conferencia. La sincronización del audio puede satisfacer los requisitos de retardo de la Memoria

Descriptiva Técnica (TS) del Proyecto de Asociación de Tercera Generación (3GPP) 26.114 durante la conferencia. Por ejemplo, el controlador de retardo puede proporcionar una señal de control a las memorias intermedias primera y segunda basándose en las marcas de tiempo, de modo que la primera memoria intermedia emite paquetes del Protocolo de transferencia en tiempo real (RTP) de la primera transmisión de audio aproximadamente al mismo tiempo que la segunda memoria intermedia emite los paquetes RTP correspondientes de la segunda transmisión de audio.

[0007] Después de la sincronización, la primera transmisión de audio y la segunda transmisión de audio se pueden descodificar por separado para generar el primer audio descodificado y el segundo audio descodificado, respectivamente. El primer y el segundo audio descodificado se pueden mezclar para generar una transmisión de audio de salida en el dispositivo móvil. En algunas implementaciones, se puede aplicar una primera función de transferencia relacionada con la cabeza (HRTF, por sus siglas en inglés) al primer audio descodificado para ajustar una primera ganancia del primer audio descodificado y para dirigir espacialmente (p. ej., panoramizar) el primer audio descodificado basándose en los ajustes definidos por el usuario. Además, se puede aplicar una segunda HRTF al segundo audio descodificado para ajustar una segunda ganancia del segundo audio descodificado y para dirigir espacialmente el segundo audio descodificado basándose en los ajustes definidos por el usuario. Por lo tanto, proporcionar la primera y la segunda transmisión de audio al dispositivo móvil (a diferencia de mezclar las transmisiones de audio en un mezclador de red y proporcionar una transmisión de audio mixto resultante al dispositivo móvil) puede permitir que el dispositivo móvil pueda controlar las características (por ejemplo, características de ganancia y características de panoramización espacial) de transmisiones de audio individuales. En particular, la panoramización de audio puede permitir que el dispositivo de renderización elija variar o ajustar los niveles de audio de los participantes mediante el ajuste de las ganancias de mezcla. Además, la espacialización de audio puede permitir que el dispositivo de renderización elija diferentes funciones HRTF para distribuir espacialmente (por ejemplo, asientos virtuales) a los participantes, lo cual puede mejorar la discriminación entre el hablante y la palabra y permitir una mejor comprensión en el caso de personas que hablan simultáneamente. Además, se puede reducir (por ejemplo, eliminar) un retardo asociado con la mezcla de transmisiones de audio en el mezclador de red.

[0008] De acuerdo con un ejemplo de las técnicas descritas en el presente documento, un procedimiento para la gestión de audio durante una conferencia incluye la recepción, en una primera memoria intermedia de un dispositivo móvil, de una primera transmisión de audio desde un primer dispositivo asociado con un primer participante de la conferencia. El procedimiento también incluye recibir, en una segunda memoria intermedia del dispositivo móvil, una segunda transmisión de audio de un segundo dispositivo asociado con un segundo participante de la conferencia. El procedimiento incluye además generar una señal de control en un controlador de retardo del dispositivo móvil. La señal de control se proporciona a la primera memoria intermedia y a la segunda memoria intermedia para sincronizar el primer audio de memoria intermedia que se envía desde la primera memoria intermedia con el segundo audio de memoria intermedia que se envía desde la segunda memoria intermedia.

[0009] De acuerdo con otro ejemplo de las técnicas descritas en el presente documento, un dispositivo móvil incluye una primera memoria intermedia configurada para recibir una primera transmisión de audio desde un primer dispositivo asociado con un primer participante de una conferencia. El dispositivo móvil también incluye una segunda memoria intermedia configurada para recibir una segunda transmisión de audio de un segundo dispositivo asociado con un segundo participante de la conferencia. El dispositivo móvil incluye además un controlador de retardo configurado para generar una señal de control. La señal de control se proporciona a la primera memoria intermedia y a la segunda memoria intermedia para sincronizar el primer audio de memoria intermedia que se envía desde la primera memoria intermedia con el segundo audio de memoria intermedia que se envía desde la segunda memoria intermedia.

[0010] De acuerdo con otro ejemplo de las técnicas descritas en el presente documento, un medio legible por ordenador no transitorio incluye instrucciones para la gestión de audio durante una conferencia. Las instrucciones, cuando son ejecutadas por un procesador en un dispositivo móvil, hacen que el procesador realice operaciones. Las operaciones incluyen recibir, en una primera memoria intermedia, una primera transmisión de audio de un primer dispositivo asociado con un primer participante de la conferencia. Las operaciones incluyen, además, recibir, en una segunda memoria intermedia, una segunda transmisión de audio de un segundo dispositivo asociado con un segundo participante de la conferencia. Las operaciones también incluyen la generación de una señal de control en un controlador de retardo. La señal de control se proporciona a la primera memoria intermedia y a la segunda memoria intermedia para sincronizar el primer audio de memoria intermedia que se envía desde la primera memoria intermedia con el segundo audio de memoria intermedia que se envía desde la segunda memoria intermedia.

[0011] De acuerdo con otro ejemplo de las técnicas descritas en el presente documento, un aparato incluye medios para recibir una primera transmisión de audio de un primer dispositivo asociado con un primer participante de una conferencia. El aparato también incluye medios para recibir una segunda transmisión de audio desde un segundo dispositivo asociado con un segundo participante de la conferencia. El aparato incluye además medios para generar una señal de control. La señal de control se proporciona a los medios para recibir la primera transmisión de audio y a los medios para recibir la segunda transmisión de audio para sincronizar el primer audio de memoria intermedia que se envía desde los medios para recibir la primera transmisión de audio con el segundo audio de memoria intermedia que se envía desde los medios para recibir la segunda transmisión de audio.

[0012] Las ventajas particulares proporcionadas por al menos una de las implementaciones divulgadas incluyen reducir el retardo asociado con la mezcla de múltiples transmisiones de audio en una conferencia en un mezclador de red. Por ejemplo, omitir la mezcla de audio en el mezclador de red puede mejorar (por ejemplo, reducir) el retardo durante la conferencia. Otras implementaciones, ventajas y características de la presente divulgación resultarán evidentes después de revisar la aplicación completa, que incluye las siguientes secciones: Breve descripción de los dibujos, Descripción detallada y Reivindicaciones.

V. Breve descripción de los dibujos

[0013]

- La FIG. 1A es un diagrama que ilustra un sistema que es operable para gestionar audio durante una conferencia;
- la FIG. 1B es un diagrama que ilustra otro sistema que es operable para gestionar audio durante una conferencia;
- la FIG. 1C es un diagrama que ilustra otro sistema que es operable para gestionar audio durante una conferencia;
- la FIG. 2A es un diagrama de un dispositivo móvil que es operable para gestionar audio durante una conferencia;
- la FIG. 2B es otra implementación del dispositivo móvil en la FIG. 2A;
- la FIG. 3 es un diagrama que ilustra el audio descodificado procesado utilizando funciones de transferencia relacionadas con la cabeza (HRTF);
- la FIG. 4 es un diagrama que ilustra el audio descodificado procesado utilizando HRTF y datos de seguimiento de cabeza;
- la FIG. 5 es un diagrama que ilustra un sistema que es operable para gestionar audio durante una conferencia utilizando una sesión de multidifusión de protocolo de Internet (IP);
- la FIG. 6 es un diagrama de flujo que ilustra un procedimiento para gestionar audio durante una conferencia; y
- la FIG. 7 es un diagrama de bloques de un dispositivo móvil operable para realizar operaciones de procesamiento de señales de acuerdo con los sistemas, diagramas y procedimientos de las FIGs. 1A-6.

VI. Descripción detallada

[0014] Se describen técnicas particulares de la presente divulgación con referencia a los dibujos. En la descripción, las características comunes se designan mediante números de referencia comunes en todos los dibujos.

[0015] Con referencia a la FIG. 1A, se muestra una implementación particular de un sistema 100 que es operable para gestionar el audio durante una conferencia. El sistema 100 incluye un dispositivo móvil 102, un primer dispositivo 110, un segundo dispositivo 120, un tercer dispositivo 130 y un dispositivo de red 190. En un ejemplo, el dispositivo móvil 102, el primer dispositivo 110, el segundo dispositivo 120 y el tercer dispositivo 130 pueden incluir un equipo de usuario (UE) que sea compatible con un estándar de Proyecto de Asociación de Tercera Generación (3GPP). Aunque se ilustra una conferencia a cuatro partes en la FIG. 1A, debe entenderse que en implementaciones alternativas, más o menos de cuatro partes pueden participar en una conferencia.

[0016] El dispositivo móvil 102, el primer dispositivo 110, el segundo dispositivo 120, y el tercer dispositivo 130 pueden estar participando en una conferencia (por ejemplo, una llamada en conferencia de audio). El dispositivo de red 190 puede configurarse como audio de ruta (por ejemplo, transmisiones de audio) de un dispositivo a otro dispositivo conectado a la llamada en conferencia.

[0017] Para ilustrar, el dispositivo móvil 102 puede estar configurado para generar una transmisión de audio 104 que está codificada en el dispositivo móvil 102. La transmisión de audio 104 puede incluir voz (por ejemplo, voz del usuario) y/o ruido de fondo. La transmisión de audio 104 puede proporcionarse al dispositivo de red 190, y el dispositivo de red 190 puede dirigir la transmisión de audio 104 (o una versión transcodificada/procesada de la misma) a los otros dispositivos 110, 120, 130. El primer dispositivo 110 puede configurarse para generar una primera transmisión de audio 114 que está codificada en el primer dispositivo 110. La primera transmisión de audio 114 puede incluir voz de un primer participante de la conferencia (por ejemplo, un usuario del primer dispositivo 110) y/o ruido de fondo en el primer dispositivo 110. La primera transmisión de audio 114 puede proporcionarse al dispositivo de red 190, y el dispositivo de red 190 puede dirigir la primera transmisión de audio 114 (o una versión transcodificada/procesada de la misma) a los otros dispositivos 102, 120, 130. El segundo dispositivo 120 puede configurarse para generar una segunda transmisión de audio 124 que está codificada en el segundo dispositivo 120. La segunda transmisión de audio 124 puede incluir voz de un segundo participante de la conferencia (por ejemplo, un usuario del segundo dispositivo 120) y/o ruido de fondo en el segundo dispositivo 120. La segunda transmisión de audio 124 puede

proporcionarse al dispositivo de red 190, y el dispositivo de red 190 puede dirigir la segunda transmisión de audio 124 (o una versión transcodificada/procesada de la misma) a los otros dispositivos 102, 110, 130. El tercer dispositivo 130 puede configurarse para generar una tercera transmisión de audio 134 que está codificada en el tercer dispositivo 130. La tercera transmisión de audio 134 puede incluir voz de un tercer participante de la conferencia (por ejemplo, un usuario del tercer dispositivo 130) y/o ruido de fondo en el tercer dispositivo 130. La tercera transmisión de audio 134 se puede proporcionar al dispositivo de red 190, y el dispositivo de red 190 puede dirigir la tercera transmisión de audio 134 (o una versión transcodificada/procesada de la misma) a los otros dispositivos 102, 110, 120.

[0018] Aunque cada transmisión de audio 104, 114, 124, 134 se describe como que incluye voz de un usuario particular/participante de la conferencia, en otros ejemplos, una o más transmisiones de audio 104, 114, 124, 134 puede incluir voz a partir de múltiples participantes de la conferencia. Por ejemplo, dos o más de los dispositivos pueden estar relativamente cerca uno del otro y pueden "captar" voz de varios participantes de la conferencia. Además, o de forma alternativa, las transmisiones de audio pueden incluir voz de múltiples participantes de la conferencia asociados con un solo equipo de usuario (por ejemplo, un solo dispositivo móvil). Por ejemplo, varios participantes de la conferencia pueden hablar en un solo dispositivo móvil (o fijo), y el dispositivo móvil (o fijo) puede generar una transmisión de audio que incluye voz de los múltiples participantes de la conferencia. Para ilustrar, la primera transmisión de audio 114 puede ser una sola transmisión de audio que incluye voz desde múltiples altavoces. Por ejemplo, si cuatro personas están hablando en el primer dispositivo 110, la primera transmisión de audio 114 puede ser una transmisión de audio mono, una transmisión de audio estéreo, una transmisión de audio de 4 canales (por ejemplo, un canal por altavoz), etc.

[0019] Como se describió anteriormente, el dispositivo de red 190 puede dirigir las transmisiones de audio codificado 114, 124, 134 de los dispositivos 110, 120, 130, respectivamente, al dispositivo móvil 102. De acuerdo con las técnicas descritas en el presente documento, el dispositivo móvil 102 puede realizar una mezcla de audio de modo que la mezcla de audio se omita en el dispositivo de red 190. La mezcla de audio en el dispositivo móvil 102 se describe con mayor detalle con respecto a la FIG. 2A. Por lo tanto, en lugar de descodificar las transmisiones de audio 114, 124, 134, mezclando las transmisiones de audio descodificado para generar una transmisión de audio mixto, recodificando la transmisión de audio mixto para generar una transmisión de audio recodificado, y proporcionando la transmisión de audio recodificado al dispositivo móvil 102, el dispositivo de red 190 puede dirigir cada transmisión de audio codificado 114, 124, 134 al dispositivo móvil 102.

[0020] De acuerdo con un ejemplo, el dispositivo de red 190 puede operar como un "interruptor maestro" para realizar las actividades de señalización entre los otros dispositivos 102, 110, 120, 130. Por ejemplo, el dispositivo de red 190 puede recibir las transmisiones de audio 104, 114, 124, 134 de los dispositivos 102, 110, 120, 130, respectivamente, y puede negociar controles de velocidad de transmisión de bits asociados con la transmisión de las transmisiones de audio 104, 114, 124, 134. Al recibir las transmisiones de audio, el dispositivo de red 190 puede negociar velocidades de transmisión de bits a las cuales las transmisiones de audio particulares se dirigen a dispositivos particulares. Como ejemplo ilustrativo no limitativo, el dispositivo de red 190 puede negociar (con el dispositivo móvil 102) una primera velocidad de transmisión de bits a la que se proporciona la primera transmisión de audio 114 al dispositivo móvil 102, una segunda velocidad de transmisión de bits a la que se proporciona la segunda transmisión de audio 124 al dispositivo móvil 102, y una tercera velocidad de transmisión de bits a la que se proporciona la tercera transmisión de audio 134 al dispositivo móvil 102. El dispositivo de red 190 también puede ser capaz de negociar anchos de banda (por ejemplo, anchos de banda de enlace ascendente y anchos de banda de enlace descendente) a los que las señales (por ejemplo, transmisiones de audio) pueden comunicarse basándose en las capacidades del dispositivo individual. Por ejemplo, basándose en las capacidades del codificador/descodificador (CÓDEC) de cada dispositivo 102, 110, 120, 130, el dispositivo de red 190 puede ser capaz de negociar un ancho de banda en el que se proporcionan transmisiones de audio a un dispositivo en particular.

[0021] Omitir la mezcla de audio en el dispositivo de red 190 y realizar la mezcla de audio en los dispositivos individuales 102, 110, 120 y 130 puede reducir las degradaciones de la calidad que de otro modo podrían estar asociadas con una cadena de procesamiento de audio en el dispositivo de red 190. Por ejemplo, las degradaciones de la calidad del audio debidas a las operaciones de descodificación en el dispositivo de red 190, las operaciones de mezcla en el dispositivo de red 190 y las operaciones de recodificación en el dispositivo de red 190 pueden reducirse. Por lo tanto, al realizar la mezcla de audio en los dispositivos individuales 102, 110, 120 y 130 (a diferencia del dispositivo de red 190), se pueden omitir las operaciones de codificación en tándem y las operaciones de transcodificación. Además, la pérdida de la percepción binaural puede reducirse. Por ejemplo, si la primera transmisión de audio 114 es una transmisión de audio estéreo y las otras transmisiones de audio 124, 134 son transmisiones de audio mono, el dispositivo móvil 102 puede conservar las cualidades estéreo de la primera transmisión de audio 114 si las transmisiones de audio individuales 114, 124 134 se envían al dispositivo móvil 102. Sin embargo, realizar una mezcla de audio (por ejemplo, descodificar, mezclar y recodificar) en el dispositivo de red 190 y proporcionar una transmisión de audio mixto al dispositivo móvil 102 puede aumentar la probabilidad de que las cualidades estéreo de la primera transmisión de audio 114 se "pierdan". Por ejemplo, puede que no haya garantía de que el dispositivo de red 190 mantenga las cualidades estéreo de la primera transmisión de audio 114 durante la mezcla de audio.

[0022] Además, omitir la mezcla de audio en el dispositivo de red 190 puede reducir el retardo en una cadena de procesamiento de audio (por ejemplo, una cadena de transcodificación). Por ejemplo, realizar una mezcla de audio en

el dispositivo de red 190 puede requerir que el dispositivo de red 190 soporte memorias intermedias de eliminación de fluctuaciones que agreguen retardo a la cadena de procesamiento de audio. También se pueden evitar las recodificaciones múltiples omitiendo la mezcla de audio en el dispositivo de red 190. Por ejemplo, para generar una transmisión mixta para el dispositivo móvil 102, el dispositivo de red 190 puede recodificar una transmisión de audio mixto que incluye (o se genera a partir de) las transmisiones de audio 114, 124, 134. Como otro ejemplo, para generar una transmisión mixta para el primer dispositivo 110, el dispositivo de red 190 puede recodificar una transmisión de audio mixto que incluye (o se genera a partir de) las transmisiones de audio 104, 124, 134. Se pueden realizar operaciones de recodificación similares para proporcionar audio mixto a los otros dispositivos 120, 130. El sistema 100 puede evitar tales operaciones de recodificación omitiendo la mezcla de audio en el dispositivo de red 190.

[0023] En implementaciones particulares, como se describe con respecto a las FIGs. 2A-4, el dispositivo móvil 102 puede aplicar funciones personalizadas de transferencia relacionadas con la cabeza (HRTF, por sus siglas en inglés) a las transmisiones de audio 114, 124, 134. Por lo tanto, la dirección espacial se puede realizar utilizando las HRTF para habilitar el audio multicanal (por ejemplo, audio del canal izquierdo, audio del canal derecho, o una combinación de los mismos) que exceda las capacidades de audio estéreo "tradicionales". Además, como se describe con respecto a las FIGs. 2A y 4, el hecho de omitir la mezcla de audio en el dispositivo de red 190 puede permitir la dirección espacial basada en las características de seguimiento de cabeza en el dispositivo móvil 102. En un escenario en el que la conferencia incluye transmisiones de audio y vídeo, el hecho de omitir la mezcla de audio en el dispositivo de red 190 también puede permitir que el dispositivo móvil 102 sincronice audio y vídeo si el vídeo no se transcodifica en un dispositivo de red, como el dispositivo de red 190.

[0024] En una implementación particular, los sistemas y procedimientos descritos pueden soportar la sincronización de audio/vídeo para la comunicación, que puede realizarse usando una variedad de técnicas. De acuerdo con un ejemplo de sincronización de audio/vídeo, las transmisiones de audio y las transmisiones de vídeo correspondientes pueden transcodificarse en el dispositivo de red 190. De acuerdo con otro ejemplo de sincronización de audio/vídeo, las transmisiones de vídeo de cada dispositivo 102, 110, 120, 130 pueden ser transmitidas a otros dispositivos dentro del sistema 100 por el dispositivo de red 190, y las transmisiones de audio 104, 114, 124, 134 de cada dispositivo 102, 110, 120, 130 puede transcodificarse en el dispositivo de red 190 para generar transmisiones de audio mixto. Por ejemplo, el dispositivo de red 190 puede generar una transmisión de audio mixto que incluye (o que se genera a partir de) las transmisiones de audio 114, 124, 134 y puede enviar la transmisión de audio mixto al dispositivo móvil 102. Las transmisiones de vídeo individuales (asociadas con los dispositivos 110, 120, 130) retransmitidos al dispositivo móvil 102 pueden incluir paquetes RTP que tienen marcas de tiempo generadas desde los dispositivos 110, 120, 130 porque el dispositivo de red 190 omite las operaciones de transcodificación en las transmisiones de vídeo. Sin embargo, la transmisión de audio mixto puede tener una marca de tiempo diferente a las marcas de tiempo de las transmisiones de vídeo debido a las operaciones de transcodificación en el dispositivo de red 190. El dispositivo móvil 120 puede determinar (por ejemplo, rastrear) el retardo entre las marcas de tiempo de las transmisiones de vídeo y la marca de tiempo de la transmisión de audio mixto. Después de determinar el retardo, el dispositivo móvil 102 puede ajustar la transmisión de audio mixto y/o las transmisiones de vídeo para sincronizar el audio y el vídeo.

[0025] De acuerdo con otro ejemplo de sincronización de audio/vídeo, las transmisiones de vídeo de cada dispositivo 102, 110, 120, 130 pueden ser transmitidas a otros dispositivos en el sistema 100 por el dispositivo de red 190, y las transmisiones de audio 104, 114, 124, 134 de cada dispositivo 102, 110, 120, 130 también pueden ser transmitidas a otros dispositivos dentro del sistema 100 por el dispositivo de red 190. De acuerdo con este ejemplo, las operaciones de transcodificación para audio y vídeo se omiten en el dispositivo de red 190. Debido a que las transmisiones de vídeo y las transmisiones de audio 104, 114, 124, 134 se envían en paquetes RTP separados, puede haber una desviación (por ejemplo, un retardo) entre un paquete de vídeo RTP y un paquete de audio RTP correspondiente. Por ejemplo, un paquete de audio RTP de la primera transmisión de audio 114 y un paquete de vídeo RTP correspondiente del primer dispositivo 110 pueden tener diferentes marcas de tiempo. Al recibir el paquete de audio RTP y el paquete de vídeo RTP correspondiente, el dispositivo móvil 102 puede sincronizar el paquete de audio RTP y el paquete de vídeo RTP correspondiente.

[0026] De acuerdo con las técnicas descritas en el presente documento, el dispositivo móvil 102 puede ser capaz de "negociar" (por ejemplo, ajustar o realizar negociaciones de protocolo de descripción de sesión (SDP) una velocidad de transmisión de bits y/o un ancho de banda de cada transmisión de audio 114, 124, 134 proporcionada al dispositivo móvil 102 durante la conferencia basándose en la configuración definida por el usuario, las capacidades de hardware del dispositivo móvil 102, o una combinación de los mismos. Para ilustrar, el dispositivo móvil 102 puede proporcionar una primera señal (no mostrada) al dispositivo de red 190 para ajustar una primera velocidad de transmisión de bits a la cual la primera transmisión de audio 114 se proporciona al dispositivo móvil 102, proporcionar una segunda señal (no mostrada) al dispositivo de red 190 para ajustar una segunda velocidad de transmisión de bits a la que se proporciona la segunda transmisión de audio 124 al dispositivo móvil 102, y/o proporcionar una tercera señal (no mostrada) al dispositivo de red 190 para ajustar una tercera velocidad de transmisión de bits a la cual se proporciona la tercera transmisión de audio 134 al dispositivo móvil 102. Como ejemplo ilustrativo no limitativo, el dispositivo móvil 102 puede enviar al dispositivo de red 190 señales que indican que un usuario del dispositivo móvil 102 prefiere una mayor calidad (por ejemplo, mayor ancho de banda) para la primera transmisión de audio 114 que para las otras transmisiones de audio 124, 134 (por ejemplo, para enfatizar la voz del participante de la conferencia que utiliza el primer dispositivo 110).

[0027] De acuerdo con otra implementación, cada dispositivo en el sistema 100 puede negociar velocidades de transmisión de bits y/o anchos de banda "directamente" con otros dispositivos en el sistema 100 (por ejemplo, omitiendo las negociaciones en el dispositivo de red 190). Como ejemplo no limitativo, el dispositivo móvil 102 puede negociar directamente con el primer dispositivo 110 para ajustar la primera velocidad de transmisión de bits a la que se proporciona la primera transmisión de audio 114 al dispositivo móvil 102. De acuerdo con este ejemplo no limitativo, el dispositivo de red 190 puede funcionar como un "dispositivo de gestión" y puede supervisar las velocidades de transmisión de bits a las que se reciben las transmisiones de audio entrantes 104, 114, 124, 134. Sin embargo, las negociaciones se realizan en los dispositivos 102, 110, 120, 130 y no en el dispositivo de red 190.

[0028] En un escenario particular, el dispositivo de red 190 puede operar como un "paquete agrupador" y puede agrupar paquetes RTP para un dispositivo particular en el sistema 100. Para ilustrar, el dispositivo de red 190 puede agrupar (por ejemplo, paquetes de) la transmisión de audio 104, la segunda transmisión de audio 124 y la tercera transmisión de audio 134 en un "paquete agrupado" que se proporcionará al primer dispositivo 110. El dispositivo de red 190 puede insertar una cabecera RTP en el paquete agrupado y puede enviar el paquete agrupado al primer dispositivo 110. De acuerdo con este escenario, el primer dispositivo 110 puede utilizar una única memoria intermedia de eliminación de fluctuaciones para procesar la transmisión de paquetes agrupados (RTP) recibidos desde el dispositivo de red 190. El dispositivo de red 190 puede asignar identificadores (ID) de dispositivo en el paquete agrupado para dar instrucciones a los elementos de procesamiento dentro del primer dispositivo 110 en cuanto a qué transmisión de audio 104, 124, 134 está asociada con cada dispositivo 102, 120, 130. Por ejemplo, cada paquete dentro de una agrupación puede incluir un identificador de por qué dispositivo fue generado el paquete. En una implementación particular, los paquetes dentro de un agrupamiento pueden incluir sus propias marcas de tiempo. El agrupamiento en su conjunto puede o no incluir una marca de tiempo. Por lo tanto, en implementaciones particulares, la gestión de la memoria intermedia de eliminación de fluctuaciones puede incluir la utilización de marcas de tiempo dentro de las transmisiones así como marcas de tiempo entre transmisiones, como se describe adicionalmente con referencia a la FIG. 2A.

[0029] En un ejemplo, el ancho de banda en el dispositivo móvil 102 para recibir las transmisiones de audio 114, 124, 134 puede depender de una serie de altavoces activos en un momento dado. Por ejemplo, basándose en las capacidades de hardware del dispositivo móvil 102 (por ejemplo, limitaciones de ancho de banda), el dispositivo móvil 102 puede tener un límite de ancho de banda para recibir y procesar audio desde el dispositivo de red 190. A medida que aumenta el número de altavoces activos en un momento dado, el ancho de banda del recurso disponible (por ejemplo, hardware) en el dispositivo móvil 102 puede disminuir.

[0030] Para reducir las limitaciones de ancho de banda, la conferencia se puede "auto-regular" para reducir el número de altavoces activos en un momento dado. Típicamente, no hay muchos oradores activos al mismo tiempo; de lo contrario, la conversación de la conferencia puede ser cada vez más difícil de seguir. Si un participante en un dispositivo particular no está hablando en un momento dado, las tramas de datos en la transmisión de audio correspondiente generada por ese dispositivo pueden incluir tramas de datos de transmisión discontinua (DTX) y pueden tener una velocidad de transmisión de bits relativamente baja (por ejemplo, aproximadamente 0,3 kilobits por segundo (kbps)) indicativa de las características de ruido de fondo. Por ejemplo, si el primer participante en el primer dispositivo 110 es silencioso en el momento dado, la velocidad de datos media de la primera transmisión de audio 114 puede ser de aproximadamente 0,3 kbps, basándose en el envío de una trama de 2,4 kbps cada 8 tramas (por ejemplo, $2,4 \text{ kbps}/8 = 0,3 \text{ kbps}$).

[0031] Suponga que un altavoz activo tiene una velocidad de transmisión de bits de enlace ascendente de 13,2 kbps de acuerdo con un indicador de señal de protocolo (SID) con una frecuencia de cada 8 tramas. En un ejemplo no limitativo, el protocolo puede ser un protocolo de Servicios de voz mejorados (EVS). Cuando hay un altavoz activo (por ejemplo, el primer participante en el primer dispositivo 110), la velocidad de transmisión de bits de enlace ascendente media para el primer dispositivo 110 puede ser de 13,2 kbps, y las velocidades de transmisión de bits de enlace ascendente medias para cada uno de los otros dispositivos 102, 120, 130 puede ser de 0,3 kbps (por ejemplo, la velocidad de transmisión de bits DTX). Por lo tanto, el ancho de banda total medio de la red de enlace ascendente puede ser de aproximadamente $13,2 + 0,3 + 0,3 + 0,3 = 14,1 \text{ kbps}$. Cuando la mezcla de audio se realiza en los dispositivos participantes (a diferencia de en el dispositivo de red 190), la velocidad de transmisión de bits del enlace descendente al dispositivo móvil 102 es de 13,8 kbps, la velocidad de transmisión de bits media del enlace descendente al primer dispositivo 110 es de 0,9 kbps, la velocidad de transmisión de bits del enlace descendente media al segundo dispositivo 120 es de 13,8 kbps, y la velocidad de transmisión de bits media del enlace descendente al tercer dispositivo 130 es de 13,8 kbps. Por lo tanto, la velocidad de transmisión de bits de enlace descendente total media puede ser de aproximadamente 42,3 kbps. La velocidad de transmisión de bits de enlace descendente total media puede ser menor que la velocidad de transmisión de bits de enlace descendente total media cuando la mezcla se realiza en el dispositivo de red 190.

[0032] Cuando dos participantes en la conferencia están hablando en un momento dado, cada altavoz activo puede tener una velocidad de transmisión de bits de enlace ascendente media de 13,2 kbps. Si el primer participante en el primer dispositivo 110 y el segundo participante en el segundo dispositivo 120 están hablando en el momento dado, la velocidad de transmisión de bits de enlace ascendente media para el primer dispositivo 110 puede ser de 13,2 kbps

y la velocidad de transmisión de bits de enlace ascendente media para el segundo dispositivo 120 puede ser de 13,2 kbps. Las velocidades de transmisión de bits de enlace ascendente medias para cada uno de los otros dispositivos 102, 130 pueden ser de 0,3 kbps (por ejemplo, la velocidad de transmisión de bits de DTX). Por lo tanto, el ancho de banda total media de la red de enlace ascendente puede ser de aproximadamente 27 kbps. Cuando la mezcla de audio se realiza en los dispositivos participantes (a diferencia de en el dispositivo de red 190), la velocidad de transmisión de bits de enlace descendente media para el dispositivo móvil 102 es 26,7 kbps, la velocidad de transmisión de bits de enlace descendente media para el primer dispositivo 110 es de 13,8 kbps, la velocidad de transmisión de bits de enlace descendente media para el segundo dispositivo 120 es de 13,8 kbps, y la velocidad de transmisión de bits de enlace descendente media para el tercer dispositivo 130 es de 26,7 kbps. Por lo tanto, la velocidad de transmisión de bits de enlace descendente total media puede ser de aproximadamente 82 kbps. La velocidad de transmisión de bits de enlace descendente total media puede ser menor que la velocidad de transmisión de bits de enlace descendente total media cuando la mezcla se realiza en el dispositivo de red 190.

[0033] Como se describió anteriormente, además de la naturaleza de "auto-regulación" de la conferencia, el dispositivo móvil 102 puede enviar señales al dispositivo de red 190 para reducir el ancho de banda de enlace descendente basándose en los ajustes definidos por el usuario. Como ejemplo no limitativo, si el usuario del dispositivo móvil 102 no quiere escuchar lo que el primer participante del primer dispositivo 110 tiene que decir, el usuario puede enviar una señal al dispositivo de red 190 para que proporcione la primera transmisión de audio 114 a una velocidad de transmisión de bits más baja (por ejemplo, una velocidad de codificación de trama activa reducida "siguiente", como 9.6 kbps como ejemplo ilustrativo no limitativo) para reducir las restricciones en una limitación de ancho de banda en el dispositivo móvil 102. En particular, este control de suelo implícito para audio puede basarse en la configuración y la señalización desde el dispositivo de renderización al dispositivo de red. Por ejemplo, al cliente de múltiples transmisiones en el dispositivo móvil se le puede permitir recibir más transmisiones de audio de las que puede descodificar simultáneamente en un momento dado. En tal caso, el cliente de múltiples transmisiones en el terminal puede tener medios para elegir qué transmisiones priorizar y cuáles ignorar. Esta selección se puede hacer en qué transmisiones no están en modo DTX. Las transmisiones de medios también se pueden priorizar basándose en el nivel activo o el volumen de la transmisión de audio. Sin embargo, esto requiere la descodificación de los medios de cada transmisión para determinar la transmisión más alta. De forma alternativa, un subconjunto de transmisiones de medios se puede descodificar periódicamente y se pueden usar *a priori* las estadísticas a largo plazo del nivel activo y la energía de trama por transmisión de medios para decidir qué transmisiones de audio se descodifican. Las transmisiones priorizadas pueden además mezclarse espacialmente para la representación.

[0034] Debido a que la mezcla de audio se realiza en el dispositivo móvil 102 en lugar de en el dispositivo de red 190, el dispositivo móvil 102 puede estar configurado para sincronizar el audio de cada transmisión de audio recibida 114, 124, 134 para preservar la calidad de audio global de la conferencia. Dichas operaciones de sincronización pueden usarse para compensar cantidades variables de retardo en la recepción de las transmisiones de audio 114, 124, 134 (por ejemplo, debido a diferentes condiciones de red para las diferentes transmisiones de audio). Como se describe con mayor detalle con respecto a la FIG. 2A, el dispositivo móvil 102 puede incluir un "controlador de retardo" que está configurado para generar una señal de control para sincronizar el audio recibido de cada transmisión de audio recibida 114, 124, 134. Por ejemplo, el controlador de retardo puede operar de manera que un retardo medio para cada memoria intermedia de eliminación de fluctuaciones (en el dispositivo móvil 102) que recibe una transmisión de audio correspondiente 114, 124, 134 es sustancialmente el mismo para satisfacer una tasa de error de trama (FER). Como ejemplo no limitativo, el controlador de retardo puede garantizar que el retardo medio para cada memoria intermedia de eliminación de fluctuaciones sea de 3 tramas, 5 tramas, 7 tramas, etc.

[0035] El sistema 100 de la FIG. 1A puede reducir las degradaciones de la calidad que de otro modo podrían estar asociadas con una cadena de procesamiento de audio en el dispositivo de red 190 al realizar el procesamiento de audio (por ejemplo, mezclar) en el dispositivo móvil 102 (y de manera similar en los dispositivos 110, 120, 130). Por ejemplo, las degradaciones de la calidad del audio debidas a las operaciones de descodificación en el dispositivo de red 190, las operaciones de mezcla en el dispositivo de red 190 y las operaciones de recodificación en el dispositivo de red 190 pueden reducirse. Por lo tanto, al realizar la mezcla de audio en los dispositivos individuales 102, 110, 120, 130, las operaciones de transcodificación se pueden omitir.

[0036] Con referencia a la FIG. 1B, se muestra una implementación particular de otro sistema 153 que es operable para gestionar el audio durante una conferencia. El sistema 153 incluye el dispositivo móvil 102, el primer dispositivo 110, el segundo dispositivo 120 y el tercer dispositivo 130.

[0037] El sistema 150 puede operar de una manera sustancialmente similar a la del sistema 100 de la FIG. 1A; sin embargo, las transmisiones de audio 104, 114, 124, 134 en el sistema 153 pueden dirigirse de un dispositivo a otro sin un sistema de red central (por ejemplo, el dispositivo de red 190 de la FIG. 1A). Por lo tanto, un retardo asociado con el direccionamiento en el dispositivo de red 190 puede reducirse en el sistema 153 de la FIG. 1B.

[0038] Con referencia a la FIG. 1C, se muestra una implementación particular de otro sistema 170 que es operable para gestionar el audio durante una conferencia. El sistema 170 incluye el dispositivo móvil 102, el primer dispositivo 110, el segundo dispositivo 120, el tercer dispositivo 130, un cuarto dispositivo 140, un quinto dispositivo 150 y una

pasarela 180 (por ejemplo, un mezclador de red). De acuerdo con una implementación, la pasarela 180 puede ser un dispositivo móvil. De acuerdo con otra implementación, la pasarela 180 puede ser un dispositivo fijo.

[0039] En la ilustración de la FIG. 1C, el cuarto dispositivo 140 y el quinto dispositivo 150 son dispositivos heredados. Por ejemplo, los dispositivos heredados 140, 150 pueden no ser capaces de realizar una mezcla de audio de múltiples transmisiones de audio de la manera descrita en las FIGs. 1A-1B con respecto a los otros dispositivos 102, 110, 120, 130 (por ejemplo, debido a restricciones de recursos en los dispositivos heredados 140, 150). En lugar de eso, los dispositivos heredados 140, 150 pueden configurarse para recibir una transmisión de audio única (por ejemplo, una transmisión de audio mixto 184) que incluye (o se genera a partir de) las transmisiones de audio 104, 114, 124, 134 de los otros dispositivos. Para ilustrar, el dispositivo de red 190 puede retransmitir las transmisiones de audio 104, 114, 124, 134 a la pasarela 180. La pasarela 180 puede realizar una mezcla de audio en las transmisiones de audio 104, 114, 124, 134 para generar la transmisión de audio mixto 184. Después de generar la transmisión de audio mixto 184, la pasarela 180 puede retransmitir la transmisión de audio mixto 184 al cuarto dispositivo 140 y al quinto dispositivo 150.

[0040] El cuarto dispositivo 140 puede estar configurado para generar una cuarta transmisión de audio 144 que está codificado en el cuarto dispositivo 140. La cuarta transmisión de audio 144 puede incluir voz (por ejemplo, voz del usuario) y/o ruido de fondo. La cuarta transmisión de audio 144 puede proporcionarse a la pasarela 180, la pasarela 180 puede dirigir la cuarta transmisión de audio 144 (o una versión transcodificada/procesada de la misma) al dispositivo de red 190, y el dispositivo de red 190 puede dirigir la cuarta transmisión de audio 144 (o una versión transcodificada/procesada de la misma) a los otros dispositivos 102, 110, 120, 130. El quinto dispositivo 150 puede configurarse para generar una quinta transmisión de audio 154 que está codificada en el quinto dispositivo 150. La quinta transmisión de audio 154 puede incluir voz (por ejemplo, voz del usuario) y/o ruido de fondo. La quinta transmisión de audio 154 puede proporcionarse a la pasarela 180, la pasarela 180 puede dirigir la quinta transmisión de audio 154 (o una versión transcodificada/procesada de la misma) al dispositivo de red 190, y el dispositivo de red 190 puede dirigir la quinta transmisión de audio 154 (o una versión transcodificada/procesada de la misma) a los otros dispositivos 102, 110, 120, 130.

[0041] Aunque la FIG. 1C representa la pasarela 180 dirigiendo la cuarta transmisión de audio 144 y la quinta transmisión de audio 154 como transmisiones de audio separadas; en otras implementaciones, la pasarela 180 puede realizar una mezcla de audio en la cuarta transmisión de audio 144 y la quinta transmisión de audio 154 para generar una mezcla de transmisión de audio. La transmisión de audio mixto puede dirigirse al dispositivo de red 190, y el dispositivo de red 190 puede transmitir la transmisión de audio mixto a los otros dispositivos 102, 110, 120, 130.

[0042] Además, aunque la FIG. 1C representa que el cuarto dispositivo 140 y el quinto dispositivo 150 reciben la misma transmisión de audio mixto 184, en otras implementaciones, el cuarto dispositivo 140 y el quinto dispositivo 150 pueden recibir diferentes transmisiones de audio mixto. Por ejemplo, una transmisión de audio mixto recibida por el cuarto dispositivo 140 puede incluir (o puede generarse a partir de) las transmisiones de audio 104, 114, 124, 134, 154 de los dispositivos 102, 110, 120, 130, 150, respectivamente. Por lo tanto, la pasarela 180 puede configurarse para mezclar la quinta transmisión de audio 154 con las otras transmisiones de audio 104, 114, 124, 134 para proporcionar una transmisión de audio mixto al cuarto dispositivo 140. De manera similar, una transmisión de audio mixto recibida por el quinto dispositivo 150 puede incluir (o puede generarse a partir de) las transmisiones de audio 104, 114, 124, 134, 144 de los dispositivos 102, 110, 120, 130, 140, respectivamente. Por lo tanto, la pasarela 180 puede configurarse para mezclar la cuarta transmisión de audio 144 con las otras transmisiones de audio 104, 114, 124, 134 para proporcionar una transmisión de audio mixto al quinto dispositivo 150.

[0043] Con referencia a la FIG. 2A, se muestra una implementación particular del dispositivo móvil 102. El dispositivo móvil 102 incluye un controlador de retardo 202, una primera memoria intermedia 210, un primer descodificador 212, una segunda memoria intermedia 220, un segundo descodificador 222, una tercera memoria intermedia 230, un tercer descodificador 232, un procesador de dirección espacial 240 y un módulo de seguimiento de cabeza 250. Cabe señalar que aunque la FIG. 2A ilustra componentes del dispositivo móvil 102, componentes similares pueden incluirse en los otros dispositivos 110, 120, 130 asociados con la conferencia.

[0044] En un ejemplo, cada memoria intermedia 210, 220, 230 puede ser una memoria intermedia de eliminación de fluctuaciones que se configura para almacenar en memoria intermedia una transmisión de audio correspondiente. Por ejemplo, la primera memoria intermedia 210 puede recibir la primera transmisión de audio 114 (por ejemplo, los primeros paquetes del Protocolo de transferencia en tiempo real (RTP)) del primer dispositivo 110 asociado con el primer participante de la conferencia, la segunda memoria intermedia 220 puede recibir la segunda transmisión de audio 124 (por ejemplo, los segundos paquetes RTP) del segundo dispositivo 120 asociado con el segundo participante de la conferencia, y la tercera memoria intermedia 230 puede recibir la tercera transmisión de audio 134 (por ejemplo, los terceros paquetes RTP) del tercer dispositivo 130 asociado con el tercer participante de la conferencia. De acuerdo con otra implementación, la primera memoria intermedia 210 puede recibir la primera transmisión de audio 114 del dispositivo de red 190 de la FIG. 1A, la segunda memoria intermedia 220 puede recibir la segunda transmisión de audio 124 del dispositivo de red 190, y la tercera memoria intermedia 230 puede recibir la tercera transmisión de audio 134 del dispositivo de red 190. De acuerdo con esta implementación (por ejemplo, la implementación de "retransmisión de red"), la marca de tiempo RTP de la primera transmisión de audio 114 recibida

por la primera memoria intermedia 210 es la misma que la marca de tiempo RTP proporcionada por el primer dispositivo 110, la marca de tiempo RTP de la segunda transmisión de audio 124 recibida por la segunda memoria intermedia 220 es la misma que la marca de tiempo RTP proporcionada por el segundo dispositivo 120, y la marca de tiempo RTP de la tercera transmisión de audio 134 recibida por la tercera memoria intermedia 230 es la misma que la marca de tiempo RTP proporcionada por el tercer dispositivo 130.

[0045] Aunque, el dispositivo móvil 102 de la FIG. 2A se muestra que incluye tres memorias intermedias 210, 220, 230 para las tres transmisiones de audio correspondientes 114, 124, 134, como se describió anteriormente con respecto a la FIG. 1A, en una implementación alternativa, el dispositivo de red 190 de la FIG. 1A puede agrupar paquetes de las transmisiones de audio 114, 124, 134 para generar una transmisión de audio agrupada, donde cada paquete de la transmisión de audio agrupada incluye los paquetes de las transmisiones de audio individuales 114, 124, 134 además de una cabecera RTP para el paquete de agrupamiento. En este escenario, el dispositivo móvil 102 puede utilizar una única memoria intermedia (por ejemplo, un única memoria intermedia de eliminación de fluctuaciones) que está configurada para recibir los paquetes de la transmisión de audio agrupada. Por ejemplo, un paquete de la transmisión de audio agrupada puede incluir una cabecera RTP, un primer paquete RTP (correspondiente a la primera transmisión de audio 114) que se identifica mediante una primera ID asignada al primer dispositivo 110, un segundo paquete RTP (correspondiente a la segunda transmisión de audio 124) que se identifica mediante un segundo ID asignado al segundo dispositivo 120, y un tercer paquete RTP (correspondiente a la tercera transmisión de audio 134) que se identifica mediante un tercer ID asignado al tercer dispositivo 130. La cabecera RTP para el paquete agrupado puede incluir una marca de tiempo. De forma alternativa, o además, los paquetes RTP primero, segundo y tercero pueden incluir sus propias cabeceras RTP y marcas de tiempo.

[0046] La primera memoria intermedia 210 puede estar configurada para almacenar en memoria intermedia la primera transmisión de audio 114 y emitir el primer audio de memoria intermedia 214 al primer descodificador 212. La segunda memoria intermedia 220 puede configurarse para almacenar en memoria intermedia la segunda transmisión de audio 124 y emitir un segundo audio de memoria intermedia 224 al segundo descodificador 222. La tercera memoria intermedia 230 puede configurarse para almacenar en memoria intermedia la tercera transmisión de audio 134 y emitir el tercer audio de memoria intermedia 234 al tercer descodificador 232. Cada memoria intermedia 210, 220, 230 puede incluir un algoritmo de cálculo de retardo de referencia (RDCA). Por ejemplo, la primera memoria intermedia 210 puede incluir un primer RDCA 211, la segunda memoria intermedia 220 puede incluir un segundo RDCA 221, y la tercera memoria intermedia 230 puede incluir un tercer RDCA 231. Cada RDCA 211, 221, 231 puede ser implementado por un procesador (por ejemplo, dentro de las memorias intermedias correspondientes 210, 220, 230) ejecutando instrucciones. El retardo para cada memoria intermedia 210, 220, 230 puede basarse en el correspondiente RDCA 211, 221, 231. Como se describe a continuación, el controlador del controlador de retardo 202 puede operar de modo que un retardo medio para cada memoria intermedia 210, 220, 230 sea sustancialmente el mismo para satisfacer un FER. Por ejemplo, el controlador de retardo 202 puede configurarse para modificar cada RDCA 211, 221, 231 para garantizar que el primer audio de memoria intermedia 214, el segundo audio de memoria intermedia 224 y el tercer audio de memoria intermedia 234 estén sustancialmente sincronizados.

[0047] El controlador de retardo 202 (por ejemplo, un sincronizador de memoria intermedia de eliminación de fluctuaciones) puede estar configurado para generar una señal de control 204 que se proporciona a cada memoria intermedia 210, 220, 230. Basándose en la señal de control 204, las memorias intermedias 210, 220, 230 pueden sincronizar la salida del primer audio de memoria intermedia 214, el segundo audio de memoria intermedia 224 y el tercer audio de memoria intermedia 234. De forma alternativa, se pueden proporcionar diferentes señales de control a cada una de las memorias intermedias 210, 220, 230. De acuerdo con una implementación, el controlador de retardo 202 puede determinar el retardo dentro de cada memoria intermedia 210, 220, 230. Por ejemplo, basándose en el primer RDCA 211, la primera memoria intermedia 210 puede determinar un primer retardo dentro de la primera memoria intermedia 210 y puede proporcionar información asociada con el primer retardo al controlador de retardo 202 a través de una señal de realimentación 205. La segunda memoria intermedia 220 puede determinar un segundo retardo dentro de la segunda memoria intermedia 220 basado en el segundo RDCA 221 y puede proporcionar información asociada con el segundo retardo al controlador de retardo 202 a través de la señal de realimentación 205. Además, la tercera memoria intermedia 230 puede determinar un tercer retardo dentro de la tercera memoria intermedia 230 basándose en el tercer RDCA 231 y puede proporcionar información asociada con el tercer retardo al controlador de retardo 202 a través de la señal de realimentación 205. El primer, segundo y tercer retardo pueden determinarse basándose en las marcas de tiempo dentro de las transmisiones dentro de las transmisiones de audio individuales. Como ejemplo no limitativo, el primer retardo puede basarse en las marcas de tiempo de RTP en la primera transmisión de audio 114 (por ejemplo, el primer retardo puede ser independiente de las marcas de tiempo de RTP en las otras transmisiones de audio 124, 134).

[0048] El controlador de retardo 202 puede utilizar la información acerca del primer, segundo y tercer retardo, para generar la señal de control 204. Aunque en la FIG. 2A se representa una única señal de control, se puede proporcionar una primera señal de control a la primera memoria intermedia 210, se puede proporcionar una segunda señal de control a la segunda memoria intermedia 220, y se puede proporcionar una tercera señal de control a la tercera memoria intermedia 230. La señal de control 204 (o múltiples señales de control) puede indicar a cada memoria intermedia 210, 220, 230 que se "acelere", aumente el retardo o mantenga el retardo actual de manera que cada memoria intermedia 210, 220, 230 tenga un retardo medio sustancialmente similar. "Acelerar" una memoria intermedia

puede incluir indicar a la memoria intermedia que "suelte" una o más tramas. Para ilustrar, la primera memoria intermedia 210 puede tener un retardo medio de dos tramas, la segunda memoria intermedia 220 puede tener un retardo medio de cuatro tramas, y la tercera memoria intermedia 230 puede tener un retardo medio de seis tramas. Basándose en estas estadísticas, el controlador de retardo 202 puede ordenar a la primera memoria intermedia 210 aumentar su retardo medio en dos tramas, ordenar a la segunda memoria intermedia 220 que mantenga su retardo medio actual, y ordenar a la tercera memoria intermedia 230 que acelere su retardo medio en dos tramas para que cada memoria intermedia 210, 220, 230 tenga un retardo medio de aproximadamente cuatro tramas. Además, los recursos de la memoria intermedia se pueden reasignar basándose en los ajustes de retardo. Por ejemplo, debido a que la primera memoria intermedia 210 necesita aumentar su retardo medio en dos tramas y la tercera memoria intermedia 230 necesita reducir su retardo en dos tramas, los recursos de la memoria intermedia se pueden reasignar entre la tercera memoria intermedia 230 y la primera memoria intermedia 210.

[0049] Por lo tanto, el controlador de retardo 202 puede utilizar marcas de tiempo entre transmisiones, además de marcas de tiempo dentro de las transmisiones, para ajustar el retardo de cada memoria intermedia 210, 220, 230. Por ejemplo, los retardos para cada memoria intermedia 210, 220, 230 pueden ajustarse basándose en la comparación de los retardos medios de cada memoria intermedia 210, 220, 230, y los retardos medios de cada memoria intermedia 210, 220, 230 se basan en las marcas de tiempo RTP de las transmisiones de audio correspondientes 114, 124, 134. Como ejemplo ilustrativo no limitativo, las marcas de tiempo de los diversos paquetes RTP recibidos de tres dispositivos "A", "B" y "C" pueden ser $t(A, 1)$, $t(A, 2)$, $t(A, 3)$... $t(A, N)$; $t(B, 1)$, $t(B, 2)$, $t(B, 3)$... $t(B, N)$; $t(C, 1)$, $t(C, 2)$, $t(C, 3)$... $t(C, N)$. Las marcas de tiempo dentro de las transmisiones pueden corresponder a la diferencia de tiempo entre dos paquetes de la misma transmisión RTP (por ejemplo, $t(A, 2)$ y $t(A, 1)$). El controlador de retardo 202 puede determinar y/o usar el retardo medio dentro de una transmisión y un retardo entre transmisiones entre las transmisiones para generar la señal de control 204.

[0050] De acuerdo con otra implementación, el controlador de retardo 202 puede comparar una primera marca de tiempo de la primera transmisión de audio 114 (por ejemplo, un primer paquete de RTP), una segunda marca de tiempo de la segunda transmisión de audio 124 (por ejemplo, un segundo paquete de RTP), y una tercera marca de tiempo de la tercera transmisión de audio 134 (por ejemplo, un tercer paquete RTP). En un ejemplo, la primera, segunda y tercera marcas de tiempo pueden basarse en una fuente de reloj común (por ejemplo, una fuente de reloj global) que se proporciona (por ejemplo, mediante un servidor de temporización u otra fuente de reloj) a 102, 110, 120, y 130. Las marcas de tiempo de las transmisiones de audio 114, 124, 134 pueden ser recibidas por el dispositivo móvil 102 sin cambios por el dispositivo de red 190 de la FIG. 1A. El controlador de retardo 202 puede determinar una primera diferencia de tiempo entre la primera marca de tiempo y la segunda marca de tiempo, una segunda diferencia de tiempo entre la segunda marca de tiempo y la tercera marca de tiempo, y una tercera diferencia de tiempo entre la tercera marca de tiempo y la primera marca de tiempo.

[0051] En una implementación particular, el controlador de retardo 202 determina las diferencias de tiempo basándose en la comparación de los paquetes individuales a partir de diferentes memorias intermedias. Por ejemplo, la primera diferencia de tiempo puede corresponder a una diferencia entre una marca de tiempo del paquete "más antiguo" (por ejemplo, el siguiente paquete que se enviará) en la primera memoria intermedia 210 y una marca de tiempo del paquete más antiguo en la segunda memoria intermedia 220. De manera similar, la segunda diferencia de tiempo puede corresponder a una diferencia entre la marca de tiempo del paquete más antiguo en la segunda memoria intermedia 220 y una marca de tiempo del paquete más antiguo en la tercera memoria intermedia 230. La tercera diferencia puede corresponder a una diferencia entre la marca de tiempo del paquete más antiguo en la tercera memoria intermedia 230 y la marca de tiempo del paquete más antiguo en la primera memoria intermedia 210.

[0052] Basándose en las diferencias de tiempo, el controlador de retardo 202 puede generar la señal de control 204 para controlar cuándo el audio de memoria intermedia 214, 224, 234 se emite desde las memorias intermedias 210, 220, 230, respectivamente, de modo que el audio de memoria intermedia 214, 224, 234 está sincronizado. Por ejemplo, la señal de control 204 (o las múltiples señales de control descritas anteriormente) pueden hacer que los paquetes/tramas en las memorias intermedias 210, 220 y 230 que tienen la misma marca de tiempo se envíen de manera sustancialmente concurrente, independientemente de cuándo dichos paquetes/tramas realmente fueron recibidos por el dispositivo móvil 102. Después de que se envía un paquete desde una de las memorias intermedias 210, 220, 230, el controlador de retardo 202 puede recalcular una o más de las diferencias de tiempo y modificar la señal de control 204 de la forma correspondiente. Al generar la señal de control 204 para sincronizar el audio de memoria intermedia 214, 224, 234, el controlador de retardo 202 puede ayudar a cada una de las memorias intermedias 210, 220, 230 a cumplir con los requisitos de retardo medio de la Memoria Descriptiva Técnica (TS) de 3GPP 26.114 durante la conferencia, como haciendo que el retardo medio a través de las memorias intermedias 210, 220, 230 sea comparable.

[0053] Como ilustración no limitante de la sincronización de primer audio de memoria intermedia 214 con el segundo audio de memoria intermedia 224, el controlador de retardo 202 puede indicar (a través de la señal de control 204) a la primera memoria intermedia 210 retardar la salida (por ejemplo, el siguiente paquete de) el primer audio de memoria intermedia 214 la primera diferencia de tiempo si la primera marca de tiempo indica una hora anterior a la segunda marca de tiempo. Si la segunda marca de tiempo indica una hora anterior a la primera marca de tiempo, el controlador de retardo 202 puede indicar (a través de la señal de control 204) a la segunda memoria intermedia 220 retardar la

salida (por ejemplo, el siguiente paquete de) el segundo audio de memoria intermedia 224 primera diferencia de tiempo.

5 **[0054]** Por lo tanto, el controlador de retardo 202 puede estar configurado para supervisar el rendimiento de cada memoria intermedia individual 210, 220, 230 de tal manera que se cumplan los requisitos de retardo de gestión de memoria intermedia de fluctuación de rendimiento mínimo (JBM) TS 26.114 durante la multi-sesión (por ejemplo, la conferencia). Además, el controlador de retardo 202 puede proporcionar información adicional de ajuste de la marca de tiempo a los RDCA 211, 221, 231 en cada memoria intermedia 210, 220, 230, de manera que el "manejo de la memoria intermedia" se realice de manera eficiente. Como se describió anteriormente, el controlador de retardo 202
10 puede recibir características de marca de tiempo intra de los RDCA 211, 221, 231 que se ejecutan en cada memoria intermedia 210, 220, 230 a través de la señal de realimentación 205, y el controlador de retardo 202 puede usar dicha información para gestionar las memorias intermedias 210, 220, 230. Por lo tanto, se puede implementar una metodología de "bucle cerrado" para gestionar el retardo de la memoria intermedia.

15 **[0055]** En respuesta a la recepción del primer audio de memoria intermedia 214, el primer descodificador 212 puede descodificar el primer audio de memoria intermedia 214 para generar el primer audio descodificado 216. El primer audio descodificado 216 se puede proporcionar al procesador de dirección espacial 240. Sin embargo, si una trama en el primer audio de memoria intermedia 214 es una trama DTX (o NODATA) (por ejemplo, correspondiente al ruido de fondo y/o silencio), las operaciones de descodificación en el primer descodificador 212 se pueden omitir para reducir el consumo de energía y ahorrar recursos de procesamiento. Cuando las operaciones de descodificación se omiten para una trama, el estado de descodificación/memorias para la trama activa descodificada previamente se puede mantener para su uso en la descodificación subsiguiente de una siguiente trama activa. En respuesta a la recepción del segundo audio de memoria intermedia 224, el segundo descodificador 222 puede descodificar el segundo audio de memoria intermedia 224 para generar el segundo audio descodificado 226. El segundo audio descodificado 226
20 también se puede proporcionar al procesador de dirección espacial 240. Si una trama en el segundo audio de memoria intermedia 224 es una trama DTX (o NODATA), las operaciones de descodificación en el segundo descodificador 222 pueden omitirse para reducir el consumo de energía y ahorrar recursos de procesamiento. En respuesta a la recepción del tercer audio de memoria intermedia 234, el tercer descodificador 232 puede descodificar el tercer audio de memoria intermedia 234 para generar el tercer audio descodificado 236. El tercer audio descodificado 236 puede proporcionarse además al procesador de dirección espacial 240. Si una trama en el tercer audio de memoria intermedia 234 es una trama DTX, las operaciones de descodificación en el tercer descodificador 232 pueden omitirse para reducir el consumo de energía y ahorrar recursos de procesamiento. En una implementación particular, el dispositivo móvil 102 puede determinar si una trama (por ejemplo, un paquete) es una trama DTX (o NODATA) basada en la longitud del paquete RTP.
35

[0056] El procesador de dirección espacial 240 puede estar configurado para realizar una primera operación de dirección espacial en el primer audio descodificado 216 para proyectar el primer audio descodificado 216 desde un altavoz en un primer ángulo (α_1). Por ejemplo, el procesador de dirección espacial 240 puede aplicar una primera HRTF 242 al primer audio descodificado 216 para dirigir (por ejemplo, panoramizar) el primer audio descodificado 216
40 de acuerdo con el primer ángulo (α_1). Una ilustración de cómo dirigir el primer audio descodificado 216 de acuerdo con el primer ángulo (α_1) se representa en la FIG. 3. La primera HRTF 242 también puede ser utilizada por el procesador de dirección espacial 240 para ajustar una primera ganancia del primer audio descodificado 216. De acuerdo con una implementación, la primera ganancia y el primer ángulo (α_1) pueden ajustarse basándose en los ajustes definidos por el usuario en el dispositivo móvil 102. Por ejemplo, si el usuario determina que la voz del primer participante asociado con el primer dispositivo 110 es más importante que la voz asociada con los otros participantes de la conferencia, el usuario puede indicar al dispositivo móvil 102 que aumente la primera ganancia del primer audio descodificado 216. Basándose en la indicación del usuario, el procesador de dirección espacial 240 puede usar la primera HRTF 242 para aumentar la primera ganancia del primer audio descodificado 216.
45

50 **[0057]** El procesador de dirección espacial 240 puede también estar configurado para realizar una segunda operación de dirección espacial en el segundo audio descodificado 226 para proyectar el segundo audio descodificado 226 a través del altavoz en un segundo ángulo (α_2). Por ejemplo, el procesador de dirección espacial 240 puede aplicar una segunda HRTF 244 al segundo audio descodificado 226 para dirigir (por ejemplo, panoramizar) el segundo audio descodificado 226 de acuerdo con el segundo ángulo (α_2). Una ilustración de cómo dirigir el segundo audio descodificado 226 de acuerdo con el segundo ángulo (α_2) se representa en la FIG. 3. La segunda HRTF 244 también
55 puede ser utilizada por el procesador de dirección espacial 240 para ajustar una segunda ganancia del segundo audio descodificado 226. De acuerdo con una implementación, la segunda ganancia y el segundo ángulo (α_2) pueden ajustarse basándose en los ajustes definidos por el usuario en el dispositivo móvil 102. Por ejemplo, si el usuario determina que la voz del segundo participante asociado con el segundo dispositivo 120 es menos importante que la voz asociada con los otros participantes de la conferencia, el usuario puede indicar al dispositivo móvil 102 que disminuya la segunda ganancia del segundo audio descodificado 226 (o silenciar el segundo audio descodificado 226). Basándose en la indicación del usuario, el procesador de dirección espacial 240 puede usar la segunda HRTF 244 para disminuir la segunda ganancia del segundo audio descodificado 226.
60

65 **[0058]** El procesador de dirección espacial 240 puede también estar configurado para realizar una tercera operación de dirección espacial en el tercer audio descodificado 236 para proyectar el tercer audio descodificado 236 a través

del altavoz con un tercer ángulo (α_3). Por ejemplo, el procesador de dirección espacial 240 puede aplicar una tercera HRTF 246 al tercer audio descodificado 236 para dirigir (por ejemplo, panoramizar) el tercer audio descodificado 236 de acuerdo con el tercer ángulo (α_3). Una ilustración de la dirección del tercer audio descodificado 236 de acuerdo con el tercer ángulo (α_3) se representa en la FIG. 3. La tercera HRTF 246 también puede ser utilizada por el procesador de dirección espacial 240 para ajustar una tercera ganancia del tercer audio descodificado 236. De acuerdo con una implementación, la tercera ganancia y el tercer ángulo (α_3) pueden ajustarse basándose en los ajustes definidos por el usuario en el dispositivo móvil 102. Por ejemplo, si el usuario determina que la voz del tercer participante asociado con el tercer dispositivo 130 es menos importante que la voz asociada con los otros participantes de la conferencia, el usuario puede indicar al dispositivo móvil 102 que disminuya la tercera ganancia del tercer audio descodificado 236 (o silenciar el tercer audio descodificado 236). Basándose en la indicación del usuario, el procesador de dirección espacial 240 puede usar la tercera HRTF 246 para disminuir la tercera ganancia del tercer audio descodificado 236.

[0059] Cada HRTF 242, 244, 246 puede ser genéricos o puede ser "personalizado" por un usuario del dispositivo móvil 102. Por ejemplo, el usuario puede indicar una ubicación (por ejemplo, un ángulo) en la que una transmisión de audio en particular 114, 124, 134 tiene que proyectarse, basándose en el número de transmisiones de audio entrantes 114, 124, 134 y basándose en las capacidades de hardware del dispositivo móvil 102. Por lo tanto, la primera HRTF 242 puede ser diferente de la segunda HRTF 244, y la segunda HRTF 244 puede ser diferente de la tercera HRTF 246. Por ejemplo, el usuario puede indicar al dispositivo móvil 102 (a través de la primera HRTF 242) que proyecte el primer audio descodificado 216 en el lado izquierdo del "altavoz" (por ejemplo, un campo de escucha percibido del usuario) indicar al dispositivo móvil 102 (a través de la segunda HRTF 244) que proyecte el segundo audio descodificado 226 en el lado derecho del altavoz, e indicar al dispositivo móvil 102 (a través de la tercera HRTF 246) que proyecte el tercer audio descodificado 236 en el medio. En otra implementación, el usuario puede indicar a un mezclador de red (por ejemplo, el dispositivo de red 190 de la FIG. 1A) que use una HRTF específica si el mezclador de red está realizando una mezcla de audio. Por ejemplo, el dispositivo de red 190 u otro mezclador de red puede rellenarse con HRTF, que pueden negociarse durante la configuración de la conferencia.

[0060] Por lo tanto, el procesador de dirección espacial 240 puede aplicar las HRTF 242, 244, 246 al audio descodificado 216, 226, 236, respectivamente, para dirigir espacialmente el audio descodificado 216, 226, 236 de tal manera que parece que los participantes de la conferencia asociados con el audio descodificado 216, 226, 236 están distribuidos espacialmente en una configuración particular. Las HRTF 242, 244, 246 pueden usarse para realizar una panoramización de audio en el dispositivo móvil 102. La panoramización de audio puede reducir la "fatiga" para el usuario del dispositivo móvil 102 cuando hablan varios participantes de la conferencia. Por ejemplo, si el primer participante (por ejemplo, el participante asociado con el primer audio descodificado 216) y el tercer participante (por ejemplo, el participante asociado con el tercer audio descodificado 236) están hablando simultáneamente, la panoramización de audio puede hacer que parezca al usuario que el primer participante habla en una ubicación y el tercer participante habla en otra ubicación, lo cual puede reducir la fatiga experimentada por el usuario.

[0061] La panoramización de audio también puede permitir que el usuario del dispositivo móvil 102 varíe los niveles de audio (por ejemplo, las ganancias) de los participantes antes de mezcla de HRTF (por ejemplo, antes de generar una señal estéreo 270, como se describe a continuación). Por ejemplo, el usuario del dispositivo móvil 102 puede dar mayor importancia al audio del primer participante (por ejemplo, el primer audio descodificado 216) en relación con el audio de los otros participantes y puede ajustar selectivamente la ganancia de mezcla en cada HRTF 242, 244, 246. En un ejemplo, el usuario del dispositivo móvil 102 puede silenciar todo el audio descodificado 226, 236, excepto el primer audio descodificado 216 cuando varios participantes están hablando simultáneamente. Como se explica con respecto a la FIG. 1A, el usuario del dispositivo móvil 102 también puede enviar una señal al dispositivo de red 190 para gestionar la velocidad de transmisión de bits y el ancho de banda de audio para cada transmisión de audio 114, 124, 134 basándose en las preferencias de los usuarios y las capacidades de hardware del dispositivo móvil 102.

[0062] En una implementación particular, el procesamiento espacial en el dispositivo móvil 102 puede refinarse aún más basándose en los datos de seguimiento de cabeza (por ejemplo, datos que indican movimiento de cabeza asociado con el usuario del dispositivo móvil 102). Por ejemplo, el módulo de seguimiento de cabeza 250 puede rastrear un cambio (por ejemplo, una rotación) de la cabeza del usuario del dispositivo móvil 102. Por ejemplo, el módulo de seguimiento de cabeza 250 puede rastrear una orientación de la cabeza del usuario y proporcionar una señal (por ejemplo, que indica una cantidad de desplazamiento (β) en la orientación) al procesador de dirección espacial 240. El procesador de dirección espacial puede ajustar cada HRTF 242, 244, 246 según la cantidad de desplazamiento (β) para acomodar los ángulos respectivos en los que el audio descodificado 216, 226, 236 se proyecta en el altavoz. Por lo tanto, el procesador de dirección espacial 240 puede usar el primer ángulo (α_1) y la cantidad de desplazamiento (β) como entradas para la primera HRTF 242, el segundo ángulo (α_2) y la cantidad de cambio (β) como entradas para la segunda HRTF 244 y el tercer ángulo (α_3) y la cantidad de desplazamiento (β) como entradas para la tercera HRTF 246.

[0063] El procesador de dirección espacial 240 puede emitir (por ejemplo, proporcionar al altavoz) la señal estéreo 270 que incluye cada audio descodificado 216, 226, 236 que se ajusta por la respectiva HRTF 242, 244, 246. Aunque la FIG. 2A ilustra que una señal estéreo 270 es emitida por el procesador de dirección espacial 240, en otra implementación, el procesador de dirección espacial 240 puede emitir tres señales mono (no mostradas). Por ejemplo, una primera señal mono puede incluir el primer audio descodificado 216 que se ajusta con la primera HRTF 242, la

segunda señal mono puede incluir el segundo audio descodificado 226 que se ajusta con la segunda HRTF 244 y la tercera señal mono puede incluir el tercer audio descodificado 236 que se ajusta mediante la tercera HRTF 246.

5 **[0064]** El dispositivo móvil 102 de la FIG. 2A puede satisfacer los requisitos de retardo de 3GPP TS 26.114 durante la conferencia al sincronizar el audio de memoria intermedia 214, 224, 234. Por ejemplo, el controlador de retardo 202 puede generar la señal de control 204 para sincronizar el audio de memoria intermedia 214, 224, 234 basándose en las marcas de tiempo en los paquetes RTP de las transmisiones de audio correspondientes 114, 124, 134. Además, el dispositivo móvil 102 puede permitir que la panoramización de audio reduzca la "fatiga" cuando varios participantes de la conferencia hablan simultáneamente. Por ejemplo, el dispositivo móvil 102 puede dirigir espacialmente el audio descodificado 216, 226, 236 de manera que parece que los participantes de la conferencia asociados con el audio descodificado 216, 226, 236 están distribuidos espacialmente.

15 **[0065]** Con referencia a la FIG. 2B, se muestra otra implementación del dispositivo móvil 102. Además de los componentes representados en la FIG. 2A, el dispositivo móvil 102 representado en la FIG. 2B puede incluir un enésima memoria intermedia 280, un enésimo descodificador 282 y una enésima HRTF 284 implementada dentro del procesador de dirección espacial 240. Por lo tanto, de acuerdo con los sistemas y procedimientos descritos, un dispositivo puede utilizar una única memoria intermedia de eliminación de fluctuaciones (por ejemplo, en el caso de procesar una transmisión RTP agrupada), dos memorias intermedias de eliminación de fluctuaciones, tres memorias intermedias de eliminación de fluctuaciones o N memorias intermedias de eliminación de fluctuaciones (por ejemplo, donde N es un número entero mayor o igual a 4). Por ejemplo, si N es igual a siete, el dispositivo móvil 102 representado en la FIG. 2B puede incluir siete memorias intermedias (configuradas para recibir siete transmisiones de audio correspondientes), siete descodificadores y siete HRTF implementados dentro del procesador de dirección espacial 240.

25 **[0066]** La enésima memoria intermedia 280 puede operar en una manera sustancialmente similar a las memorias intermedias 210, 220, 230. Por ejemplo, la enésima memoria intermedia puede estar configurada para recibir una enésima transmisión de audio 292 desde un dispositivo en la conferencia (o desde el dispositivo de red 190 de la FIG. 1A) y para emitir el enésimo audio de memoria intermedia 294 basándose en la señal de control 204. El enésimo audio de memoria intermedia 294 puede sincronizarse con el audio de memoria intermedia 214, 224, 234 de las otras memorias intermedias 210, 220, 230, respectivamente. El enésimo codificador 282 puede funcionar de forma sustancialmente similar a los codificadores 212, 222 y 232. Por ejemplo, el enésimo descodificador 282 puede descodificar el enésimo audio de memoria intermedia 294 para generar el enésimo audio descodificado 296. El enésimo audio descodificado 296 se puede proporcionar al procesador de dirección espacial 240. El procesador de dirección espacial 240 también puede configurarse para realizar una enésima operación de dirección espacial en el enésimo audio descodificado 296 para proyectar el enésimo audio descodificado 296 desde el altavoz en un enésimo ángulo (α_N). Por ejemplo, el procesador de dirección espacial 240 puede aplicar la enésima HRTF 284 al enésimo audio descodificado 296 para dirigir (por ejemplo, panoramizar) el enésimo audio descodificado 296 de acuerdo con el enésimo ángulo (α_N).

40 **[0067]** Con referencia a la FIG. 3, se muestra un diagrama que ilustra un ejemplo de una disposición espacial del audio descodificado 216, 226, 236 después de aplicar las HRTF 242, 244, 246. En una implementación particular, las HRTF 242, 244, 246 pueden precomputarse basándose en una asignación de ubicación de altavoz virtual "preferida". Por ejemplo, la primera HRTF 242 puede dirigir espacialmente el primer audio descodificado 216 para que se perciba como si viniera del lado izquierdo de un altavoz (por ejemplo, proveniente del primer ángulo (α_1)). De manera similar, la segunda HRTF 244 puede dirigir espacialmente el segundo audio descodificado 226 para que se perciba como si procediera del lado izquierdo central del altavoz (por ejemplo, proveniente del segundo ángulo (α_2)), y la tercera HRTF 246 puede dirigir espacialmente el tercer audio descodificado 236 para percibirse como si viniera del lado derecho del altavoz (p. ej., proveniente del tercer ángulo (α_3)). Como se describe con respecto a la FIG. 2A, las HRTF 242, 244, 246 también pueden incluir control de ganancia para enfatizar un altavoz "preferido" en relación con otros altavoces.

50 **[0068]** Con referencia a la FIG. 4, se muestra un diagrama que ilustra un ejemplo de una disposición espacial del audio descodificado 216, 226, 236 después de aplicar las HRTF 242, 244, 246 y los datos de seguimiento de cabeza. En la FIG. 4, el módulo de seguimiento de cabeza 250 puede detectar que la cabeza del usuario se desplazó la cantidad de desplazamiento (β). Basándose en la información de seguimiento de cabeza, el procesador de dirección espacial 240 puede desplazar un campo de sonido que incluye el audio descodificado 216, 226, 236 la cantidad de desplazamiento (β) como se ilustra en la FIG. 4. Por lo tanto, la percepción de dónde se descodifica el audio descodificado 216, 226, 236 en la FIG. 3 puede permanecer sustancialmente sin cambios cuando el usuario del dispositivo móvil 102 desplaza la cabeza.

60 **[0069]** Con referencia a la FIG. 5, se muestra una implementación particular de un sistema 500 que es operable para gestionar audio durante una conferencia usando una sesión de multidifusión de protocolo de Internet (IP). El sistema 500 incluye el dispositivo móvil 102, el primer dispositivo 110 y el segundo dispositivo 120.

65 **[0070]** De acuerdo con una aplicación particular de operación, el dispositivo móvil 102 puede iniciar una llamada con el primer dispositivo 110. El dispositivo móvil 102 puede iniciar posteriormente una referencia de protocolo de descripción de sesión (SDP) para que el primer dispositivo 110 inicie una sesión de multidifusión con el segundo

dispositivo 120. El segundo dispositivo 120 también puede iniciar una sesión con el dispositivo móvil 102. En una implementación particular, si N nodos participan en la conferencia, pueden iniciarse $N*(N-1)$ sesiones por medio (por ejemplo, para 3 nodos, $3*2=6$ sesiones pueden iniciarse para audio, otras $3*2=6$ sesiones pueden iniciarse para vídeo, etc.). En la sesión de multidifusión de la FIG. 5, no hay ningún dispositivo de red implícito, tal como el dispositivo de red 190 de la FIG. 1A. Por lo tanto, la señalización del protocolo de inicio de sesión (SIP) y la negociación SDP pueden basarse en un CÓDEC soportado por cada dispositivo 102, 110, 120. La señalización SIP y las negociaciones SDP se pueden usar para seleccionar un CÓDEC de audio, una velocidad de transmisión de bits, un ancho de banda de audio, etc.

[0071] De acuerdo con una implementación, uno o más de los dispositivos 102, 110, 120 puede funcionar como una pasarela (por ejemplo, la pasarela de red) basándose en los recursos de hardware que están disponibles. Si un dispositivo funciona como una pasarela, un requisito de CÓDEC puede ser relajado. Por ejemplo, si un CÓDEC del primer dispositivo 110 no es compatible para descodificar la transmisión de audio 104 del dispositivo móvil 102, el segundo dispositivo 120 puede funcionar como una pasarela y descodificar la transmisión de audio 104, recodificar la transmisión de audio en un formato soportado por el primer dispositivo 110, y proporcionar la transmisión de audio recodificada al primer dispositivo 110.

[0072] Con referencia a la FIG. 6, se muestra un diagrama de flujo de un procedimiento 600 para gestionar audio durante una conferencia. El procedimiento 600 puede ser realizado por el dispositivo móvil 102 de las FIGs. 1A-2B, el primer dispositivo 110 de las FIGs. 1A-1C, el segundo dispositivo 120 de las FIGs. 1A-1C, el tercer dispositivo 130 de las FIGs. 1A-1C, o una combinación de los mismos.

[0073] El procedimiento 600 incluye la recepción, en una primera memoria intermedia de un dispositivo móvil, de una primera transmisión de audio desde un primer dispositivo asociado con un primer participante de una conferencia, a 602. Por ejemplo, refiriéndose a la FIG. 2A, la primera memoria intermedia 210 puede recibir la primera transmisión de audio 114 (por ejemplo, paquetes RTP) del primer dispositivo 110.

[0074] El procedimiento 600 puede incluir también recibir, en una segunda memoria intermedia del dispositivo móvil, una segunda transmisión de audio desde un segundo dispositivo asociado con un segundo participante de la conferencia, en 604. Por ejemplo, refiriéndose a la FIG. 2A, la segunda memoria intermedia 220 puede recibir la segunda transmisión de audio 124 (por ejemplo, paquetes RTP) del segundo dispositivo 120.

[0075] Una señal de control puede generarse en un controlador de retardo del dispositivo móvil, en 606. La señal de control se puede proporcionar a la primera memoria intermedia y a la segunda memoria intermedia para sincronizar el primer audio de memoria intermedia que se envía desde la primera memoria intermedia con el segundo audio de memoria intermedia que se envía desde la segunda memoria intermedia. Por ejemplo, refiriéndose a la FIG. 2A, el controlador de retardo 202 puede comparar la primera marca de tiempo de (por ejemplo, un primer paquete de) la primera transmisión de audio 114 con la segunda marca de tiempo de (por ejemplo, un paquete más antiguo de) la segunda transmisión de audio 124 y determinar una diferencia de tiempo entre la primera marca de tiempo y la segunda marca de tiempo. Si la primera marca de tiempo indica una hora anterior a la segunda marca de tiempo, la señal de control 204 puede indicar a la primera memoria intermedia 210 que retarde la salida del primer audio de memoria intermedia 214 la diferencia de tiempo para sincronizar el primer audio de memoria intermedia 214 con el segundo audio de memoria intermedia 224. Si la segunda marca de tiempo indica una hora anterior a la primera marca de tiempo, la señal de control 204 puede indicar a la segunda memoria intermedia 220 que retarde la salida del segundo audio de memoria intermedia 224 la diferencia de tiempo para sincronizar el primer audio de memoria intermedia 214 con el segundo audio de memoria intermedia 224. El controlador de retardo puede tener en cuenta los tiempos de llegada entre paquetes entre los paquetes [A1, A2,...] y los paquetes [B1, B2,...] de los participantes A y B y estimar la profundidad de memoria intermedia de eliminación de fluctuaciones variable necesaria para el almacenamiento en memoria intermedia de los medios del participante A y el participante B mientras se mantiene la pérdida de paquetes (o la ocultación provocada por la fluctuación) baja y se cumplen los requisitos mínimos de rendimiento de 3GPP TS 26.114. Por ejemplo, el participante A puede estar en malas condiciones de canal de radio y tiene una alta velocidad de pérdida de paquetes con baja fluctuación, mientras que el participante B está en buena condición de canal de radio con una velocidad de pérdida de paquetes muy baja pero con alta fluctuación. El controlador de retardo tiene en cuenta las características de fluctuación/pérdida de los paquetes del participante A y B, para asignar, por ejemplo, una mayor profundidad de la memoria intermedia para los paquetes del participante B, de modo que la pérdida total del paquete no exceda la de la pérdida experimentada por el participante A. El mecanismo del controlador de retardo anterior también se puede extender para casos que incluyen la recepción de más de dos transmisiones de medios.

[0076] De acuerdo con una implementación, el procedimiento 600 puede incluir además la descodificación, en un primer descodificador del dispositivo móvil, del primer audio de memoria intermedia para generar el primer audio descodificado. Por ejemplo, refiriéndose a la FIG. 2A, el primer descodificador 212 puede descodificar el primer audio almacenado 214 para generar el primer audio descodificado 216. El procedimiento 600 también puede incluir descodificar, en un segundo descodificador del dispositivo móvil, el segundo audio de memoria intermedia para generar un segundo audio descodificado. Por ejemplo, refiriéndose a la FIG. 2A, el segundo descodificador 222 puede descodificar el segundo audio almacenado 224 para generar el segundo audio descodificado 226.

[0077] De acuerdo con una implementación, el procedimiento 600 puede incluir la realización de una primera operación de dirección espacial en el primer audio descodificado para proyectar el primer audio descodificado de un altavoz en un primer ángulo. Por ejemplo, refiriéndose a la FIG. 2A, el procesador de dirección espacial 240 puede realizar la primera operación de dirección espacial en el primer audio descodificado 216 para proyectar el primer audio descodificado 216 desde el altavoz en el primer ángulo (α_1). Para ilustrar, el procesador de dirección espacial 240 puede aplicar la primera HRTF 242 al primer audio descodificado 216 para dirigir (por ejemplo, panoramizar) el primer audio descodificado 216 de acuerdo con el primer ángulo (α_1). La primera HRTF 242 también puede ser utilizada por el procesador de dirección espacial 240 para ajustar la primera ganancia del primer audio descodificado 216.

[0078] De acuerdo con una implementación del procedimiento 600, un primer descodificador puede recibir la señal de control desde el controlador de retardo y un segundo descodificador también puede recibir la señal de control desde el controlador de retardo. El primer descodificador puede descodificar un primer paquete de datos asociado con la primera transmisión de audio basándose en la señal de control u omitir las operaciones de descodificación en el primer paquete de datos basándose en la señal de control. De manera similar, el segundo descodificador puede descodificar un segundo paquete de datos asociado con la segunda transmisión de audio basándose en la señal de control u omitir las operaciones de descodificación en el segundo paquete de datos basándose en la señal de control. De acuerdo con una implementación, el primer descodificador y el segundo descodificador son descodificadores diferentes. El primer descodificador puede ser soportado por todos los participantes de la conferencia y el segundo descodificador puede ser soportado por un subconjunto de participantes de la conferencia. De acuerdo con otra implementación, el primer descodificador y el segundo descodificador incluyen un descodificador similar que funciona en diferentes modos.

[0079] El procedimiento 600 también puede incluir la realización de una segunda operación de dirección espacial en el segundo audio descodificado para proyectar el segundo audio descodificado del altavoz en un segundo ángulo. Por ejemplo, refiriéndose a la FIG. 2A, el procesador de dirección espacial 240 puede realizar la segunda operación de dirección espacial en el segundo audio descodificado 226 para proyectar el segundo audio descodificado 226 desde el altavoz en el segundo ángulo (α_2). Para ilustrar, el procesador de dirección espacial 240 puede aplicar la segunda HRTF 244 al segundo audio descodificado 226 para dirigir (por ejemplo, panoramizar) el segundo audio descodificado 226 de acuerdo con el segundo ángulo (α_2). La segunda HRTF 244 también puede ser utilizada por el procesador de dirección espacial 240 para ajustar la segunda ganancia del segundo audio descodificado 226.

[0080] De acuerdo con una implementación, el procedimiento 600 puede incluir proporcionar una primera señal a un dispositivo de red para ajustar una primera velocidad de transmisión de bits de la primera transmisión de audio. Por ejemplo, refiriéndose a las FIGs. 1A y 2A, el dispositivo de red 190 puede dirigir la primera transmisión de audio 114 desde el primer dispositivo 110 a la primera memoria intermedia 210 del dispositivo móvil 102, y el dispositivo de red 190 puede dirigir la segunda transmisión de audio 124 desde el segundo dispositivo 120 a la segunda memoria intermedia 220 del dispositivo móvil 102. El dispositivo móvil 102 puede proporcionar la primera señal al dispositivo de red 190 para ajustar la primera velocidad de transmisión de bits a la que se proporciona la primera transmisión de audio 114 al dispositivo móvil 102. El procedimiento 600 también puede incluir proporcionar una segunda señal al dispositivo de red para ajustar una segunda velocidad de transmisión de bits de la segunda transmisión de audio. Por ejemplo, refiriéndose a la FIG. 1A, el dispositivo móvil 102 puede proporcionar la segunda señal al dispositivo de red 190 para ajustar una segunda velocidad de transmisión de bits a la que se proporciona la segunda transmisión de audio 124 al dispositivo móvil 102.

[0081] El procedimiento 600 de la FIG. 6 puede satisfacer los requisitos de retardo de 3GPP TS 26.114 durante la conferencia al sincronizar el audio de memoria intermedia 214, 224, 234. Por ejemplo, el controlador de retardo 202 puede generar la señal de control 204 para sincronizar el audio de memoria intermedia 214, 224, 234 basándose en las marcas de tiempo en los paquetes RTP de las transmisiones de audio correspondientes 114, 124, 134. Además, el procedimiento 600 puede permitir que la panoramización de audio reduzca la "fatiga" cuando varios participantes de la conferencia hablan simultáneamente. Por ejemplo, el procedimiento 600 puede permitir el direccionamiento espacial del audio descodificado 216, 226, 236, de modo que parece que los participantes de la conferencia asociados con el audio descodificado 216, 226, 236 están distribuidos espacialmente.

[0082] En algunos ejemplos, el procedimiento 600 de la FIG. 6 puede implementarse a través de hardware (por ejemplo, un dispositivo FPGA, un ASIC, etc.) de una unidad de procesamiento, tal como una unidad central de procesamiento (CPU), un DSP o un controlador, a través de un dispositivo de firmware, o cualquier combinación de los mismos. En un ejemplo, un procesador que ejecuta instrucciones puede realizar el procedimiento 600 de la FIG. 6, como se describe con respecto a la FIG. 7.

[0083] Con referencia a la FIG. 7, se muestra un diagrama de bloques de una implementación ilustrativa particular del dispositivo móvil 102. En una implementación particular, el dispositivo 102 incluye un procesador 706 (por ejemplo, una CPU). El dispositivo móvil 102 puede incluir uno o más procesadores adicionales 710 (por ejemplo, uno o más DSP).

[0084] El dispositivo móvil 102 puede incluir una memoria 732 y un controlador inalámbrico 740 acoplado a una antena 742. El controlador inalámbrico 740 incluye el controlador de retardo 202 de las FIGs. 2A-2B, la primera memoria intermedia 210 de las FIGs. 2A-2B, el primer descodificador 212 de las FIGs. 2A-2B, la segunda memoria intermedia 220 de las FIGs. 2A-2B, el segundo descodificador 222 de las FIGs. 2A-2B, la tercera memoria intermedia 230 de las FIGs. 2A-2B, el tercer descodificador 232 de las FIGs. 2A-2B, y el procesador de dirección espacial 240 de las FIGs. 2A-2B. La lógica del controlador de retardo 202 de las FIGs. 2A-2B también se puede implementar dentro del procesador 706, o el uno o más procesadores adicionales 710. Por lo tanto, la primera memoria intermedia 210 en el controlador inalámbrico 740 puede configurarse para recibir la primera transmisión de audio 114 a través de la antena 742, la segunda memoria intermedia 220 en el controlador inalámbrico 740 puede configurarse para recibir la segunda transmisión de audio 124 a través de la antena 742, y la tercera memoria intermedia 230 en el controlador inalámbrico 740 puede configurarse para recibir la tercera transmisión de audio 134 a través de la antena 742. Para satisfacer los requisitos de retardo del 3GPP TS 26.114 durante una conferencia, se puede proporcionar una señal de control (no mostrada) a las memorias intermedias 210, 220, 230 para sincronizar el audio de memoria intermedia que se envía desde las memorias intermedias 210, 220, 230, como se describió anteriormente.

[0085] El dispositivo móvil 102 puede incluir una pantalla 728 acoplada a un controlador de pantalla 726. Se puede acoplar un altavoz 736, un micrófono 738 o ambos a un CÓDEC 734. El CÓDEC 734 puede incluir un convertidor de digital a analógico (DAC) 702 y un convertidor de analógico a digital (ADC) 704. En un ejemplo, la salida del procesador de dirección espacial (por ejemplo, una señal estéreo) puede proporcionarse a uno o más procesadores adicionales 710 para procesamiento adicional y al CÓDEC 734. La señal estéreo puede convertirse en una señal analógica a través del DAC 702 y emitirse por el altavoz 736.

[0086] La memoria 732 puede incluir instrucciones ejecutables 760 por el procesador 706, los procesadores 710, el CÓDEC 734, el controlador inalámbrico 740 y los componentes de los mismos, o una combinación de los mismos, para llevar a cabo el procedimiento 600 de la FIG. 6. La memoria 732 o el uno o más componentes del procesador 706, los procesadores 710, el controlador inalámbrico 740 y/o el CÓDEC 734 pueden ser un medio no transitorio legible por ordenador que incluya instrucciones (por ejemplo, las instrucciones 760) que, cuando son ejecutadas por un ordenador (por ejemplo, un procesador en el CÓDEC 734, el procesador 706 y/o los procesadores 710) hacen que el ordenador realice el procedimiento 600 de la FIG. 6.

[0087] Aunque el controlador de retardo 202, las memorias intermedias 210, 220, 230, los descodificadores 212, 222, 232, y el procesador de dirección espacial 240 se representan como estando dentro del controlador inalámbrico 740, en otras implementaciones, uno o más de las memorias intermedias 210, 220, 230, los descodificadores 212, 222, 232, o el procesador de dirección espacial 240 pueden estar dentro del CÓDEC 734, el procesador 710, el uno o más procesadores 706 adicionales, o en otro componente del dispositivo móvil 102.

[0088] En una implementación particular, el dispositivo móvil 102 puede incluirse en un dispositivo de sistema en paquete o sistema en chip 722, tal como un módem de estación móvil (MSM). En una implementación particular, el procesador 706, el procesador 710, el controlador de visualización 726, la memoria 732, el CÓDEC 734 y el controlador inalámbrico 740 pueden estar incluidos en un sistema en paquete o el dispositivo de sistema en chip 722. En una implementación particular, un dispositivo de entrada 730, tal como una pantalla táctil y/o un teclado, y una fuente de alimentación 744, están acoplados al dispositivo de sistema en chip 722. Además, en una implementación particular, como se ilustra en la FIG. 7, la pantalla 728, el dispositivo de entrada 730, el altavoz 736, el micrófono 738, la antena 742 y la fuente de alimentación 744 son externos con respecto al dispositivo de sistema en chip 722. Sin embargo, cada uno de la pantalla 728, el dispositivo de entrada 730, el altavoz 736, el micrófono 736, la antena 742 y la fuente de alimentación 744 se pueden acoplar a un componente del dispositivo de sistema en chip 722, tal como una interfaz o un controlador. En un ejemplo ilustrativo, el dispositivo móvil 102 corresponde a un dispositivo de comunicación móvil, un teléfono inteligente, un teléfono celular, un ordenador portátil, un ordenador, un ordenador tipo, un asistente digital personal, una pantalla, un televisor, una consola de juegos, un reproductor de música, una radio, un reproductor de vídeo digital, un reproductor de disco óptico, un sintonizador, una cámara, un dispositivo de navegación, un sistema descodificador, un sistema codificador o cualquier combinación de los mismos.

[0089] Aunque la FIG. 7 representa componentes del dispositivo móvil 102, otros dispositivos descritos en el presente documento (por ejemplo, el primer dispositivo 110 de las FIGs. 1A-1C, el segundo dispositivo 120 de las FIGs. 1A-1C, y/o el tercer dispositivo 130 de las FIGs. 1A-1C) puede incluir componentes similares a los ilustrados en la FIG. 7 para el dispositivo móvil 102.

[0090] Implementaciones adicionales de la divulgación se presentan en el presente documento en la forma de un apéndice. Debe entenderse que tales implementaciones pueden utilizarse en su lugar de, o además de, ciertas implementaciones ilustradas y descritas con referencia a las FIGs. 1A-7.

[0091] Conjuntamente con las implementaciones descritas, se divulga un aparato que incluye medios para recibir una primera transmisión de audio desde un primer dispositivo asociado con un primer participante de una conferencia. Por ejemplo, entre los medios para recibir la primera transmisión de audio pueden incluirse la primera memoria intermedia 210 de las FIGs. 2A-2B y 7, el controlador inalámbrico 740 de la FIG. 7, uno o más dispositivos configurados para

recibir la primera transmisión de audio (por ejemplo, un procesador que ejecuta instrucciones en un medio de almacenamiento legible por ordenador no transitorio), o cualquier combinación de los mismos.

5 **[0092]** El aparato puede incluir también medios para recibir una segunda transmisión de audio desde un segundo dispositivo asociado con un segundo participante de la conferencia. Por ejemplo, entre los medios para recibir la segunda transmisión de audio se puede incluir la segunda memoria intermedia 220 de las FIGs. 2A-2B y 7, el controlador inalámbrico 740 de la FIG. 7, uno o más dispositivos configurados para recibir la segunda transmisión de audio (por ejemplo, un procesador que ejecuta instrucciones en un medio de almacenamiento legible por ordenador no transitorio), o cualquier combinación de los mismos.

10 **[0093]** El aparato también puede incluir medios para generar una señal de control. La señal de control puede proporcionarse a los medios para recibir la primera transmisión de audio y a los medios para recibir la segunda transmisión de audio para sincronizar el primer audio de memoria intermedia que se envía desde los medios para recibir la primera transmisión de audio con el segundo audio de memoria intermedia que se envía desde los medios para recibir la segunda transmisión de audio. Por ejemplo, los medios para generar la señal de control pueden incluir el controlador de retardo 202 de las FIGs. 2A-2B y 7, el controlador inalámbrico 740 de la FIG. 7, el procesador 706 de la FIG. 7, el uno o más procesadores adicionales 710 de la FIG. 7, uno o más dispositivos configurados para generar la señal de control (por ejemplo, un procesador que ejecuta instrucciones en un medio de almacenamiento legible por ordenador no transitorio), o cualquier combinación de los mismos.

15 **[0094]** Los expertos en la técnica apreciarán además que los diversos bloques lógicos, configuraciones, módulos, circuitos y pasos de algoritmo ilustrativos descritos en relación con las implementaciones divulgadas en el presente documento pueden implementarse como hardware electrónico, software informático ejecutado por un dispositivo de procesamiento tal como un procesador de hardware, o combinaciones de ambos. Hasta aquí se han descrito diversos componentes, bloques, configuraciones, módulos, circuitos y pasos ilustrativos, en general, en lo que respecta a su funcionalidad. Que dicha funcionalidad se implemente como hardware o software ejecutable depende de la aplicación particular y de las restricciones de diseño impuestas al sistema global. Los expertos en la técnica pueden implementar la funcionalidad descrita de distintas maneras para cada aplicación particular, pero no se debería interpretar que dichas decisiones de implementación suponen apartarse del alcance de la presente divulgación.

20 **[0095]** Los pasos de un procedimiento o algoritmo descritos en relación con las implementaciones divulgadas en el presente documento pueden realizarse directamente en hardware, en un módulo de software ejecutado por un procesador o en una combinación de los dos. Un módulo de software puede residir en un dispositivo de memoria, tal como una memoria de acceso aleatorio (RAM), una memoria de acceso aleatorio magnetorresistiva (MRAM), una MRAM de transferencia de par de giro (STT-MRAM), una memoria flash, una memoria de solo lectura (ROM), una memoria de solo lectura programable (PROM), una memoria de solo lectura programable y borrrable (EPROM), una memoria de solo lectura programable y borrrable eléctricamente (EEPROM), unos registros, un disco duro, un disco extraíble o una memoria de solo lectura de disco compacto (CD-ROM). Un dispositivo de memoria a modo de ejemplo está acoplado al procesador de tal manera que el procesador puede leer información de, y escribir información en, el dispositivo de memoria. De forma alternativa, el dispositivo de memoria puede estar integrado en el procesador. El procesador y el medio de almacenamiento pueden residir en un ASIC. El ASIC puede residir en un dispositivo informático o en un terminal de usuario. De forma alternativa, el procesador y el medio de almacenamiento pueden residir como componentes discretos en un dispositivo informático o en un terminal de usuario.

35 **[0096]** La descripción anterior de las implementaciones divulgadas se proporciona para permitir que un experto en la técnica elabore o use las implementaciones divulgadas. Diversas modificaciones de estas implementaciones resultarán fácilmente evidentes para los expertos en la técnica, y los principios definidos en el presente documento se pueden aplicar a otros ejemplos sin apartarse del alcance de la divulgación. Por tanto, la presente divulgación no pretende limitarse a los aspectos mostrados en el presente documento, sino que se le ha de conceder el alcance más amplio posible compatible con los principios y características novedosas, según lo definido en las reivindicaciones siguientes.

REIVINDICACIONES

1. Un procedimiento (600) para gestionar el audio durante una conferencia, con el procedimiento (600) que comprende:
 - 5 recibir (602), en una primera memoria intermedia (210) de un dispositivo móvil (102), una primera transmisión de audio (114) de un primer dispositivo (110) asociado con un primer participante de la conferencia;
 - 10 recibir (604), en una segunda memoria intermedia (220) del dispositivo móvil (102), una segunda transmisión de audio (124) de un segundo dispositivo (120) asociado con un segundo participante de la conferencia; **caracterizado por**
 - 15 generar (606) una señal de control (204) en un controlador de retardo (202) del dispositivo móvil (102), con la señal de control (204) proporcionada a la primera memoria intermedia (210) y a la segunda memoria intermedia (220) para sincronizar el primer audio de memoria intermedia (214) que se envía desde la primera memoria intermedia (210) con el segundo audio de memoria intermedia (224) que se envía desde la segunda memoria intermedia (220).
- 20 2. El procedimiento (600) de acuerdo con la reivindicación 1, que comprende, además:
 - comparar, en el dispositivo móvil (102), una primera marca de tiempo de la primera transmisión de audio (114) con una segunda marca de tiempo de la segunda transmisión de audio (124), con la primera marca de tiempo y la segunda marca de tiempo basadas en una fuente de reloj común; y
 - 25 determinar, en el dispositivo móvil (102), una diferencia de tiempo entre la primera marca de tiempo y la segunda marca de tiempo.
- 30 3. El procedimiento (600) de la reivindicación 2, en el que la señal de control (204) indica a la primera memoria intermedia (210) que retarde la salida del primer audio de memoria intermedia (214) la diferencia de tiempo si la primera marca de tiempo indica un tiempo anterior a la segunda marca de tiempo; o en el que la señal de control (204) indica a la segunda memoria intermedia (220) que retarde la salida del segundo audio de memoria intermedia (224) la diferencia de tiempo si la segunda marca de tiempo indica una hora anterior a la primera marca de tiempo.
- 35 4. El procedimiento (600) de acuerdo con la reivindicación 1, que comprende, además:
 - descodificar, en un primer descodificador (212) del dispositivo móvil (102), el primer audio de memoria intermedia (214) para generar el primer audio descodificado (216); y
 - 40 descodificar, en un segundo descodificador (222) del dispositivo móvil (102), el segundo audio de memoria intermedia (224) para generar el segundo audio descodificado (226).
- 45 5. El procedimiento (600) de la reivindicación 4, que comprende además:
 - realizar, en el dispositivo móvil (102), una primera operación de dirección espacial en el primer audio descodificado (216) para proyectar el primer audio descodificado (216) desde un altavoz en un primer ángulo; y
 - 50 realizar, en el dispositivo móvil (102), una segunda operación de dirección espacial en el segundo audio descodificado (226) para proyectar el segundo audio descodificado (226) desde el altavoz en un segundo ángulo.
- 55 6. El procedimiento (600) de la reivindicación 5, en el que realizar la primera operación de dirección espacial comprende aplicar una primera función de transferencia relacionada con la cabeza (HRTF) al primer audio descodificado (216), y en el que realizar la segunda operación de dirección espacial comprende aplicar una segunda HRTF al segundo audio descodificado (226).
- 60 7. El procedimiento (600) de la reivindicación 5, en el que el primer ángulo y el segundo ángulo se basan en ajustes definidos por el usuario.
8. El procedimiento (600) de la reivindicación 7, que comprende además desplazar el primer ángulo y el segundo ángulo una cantidad de desplazamiento en respuesta a la detección del movimiento de cabeza asociado con un usuario del dispositivo móvil (102).
- 65 9. El procedimiento (600) de acuerdo con la reivindicación 4, que comprende, además:

ajuste de una primera ganancia del primer audio descodificado (216); y

ajuste de una segunda ganancia del segundo audio descodificado (226), en el que la primera ganancia y la segunda ganancia se ajustan basándose en los ajustes definidos por el usuario.

10. El procedimiento (600) de la reivindicación 1, en el que la primera transmisión de audio (114) se dirige desde el primer dispositivo (110) a la primera memoria intermedia (210) a través de un dispositivo de red (190), y en el que la segunda transmisión de audio (124) se dirige desde el segundo dispositivo (120) a la segunda memoria intermedia (220) a través del dispositivo de red (190).

11. El procedimiento (600) de acuerdo con la reivindicación 10, que comprende, además:

proporcionar una primera señal al dispositivo de red (190) para ajustar una primera velocidad de transmisión de bits de la primera transmisión de audio (114); y

proporcionar una segunda señal al dispositivo de red (190) para ajustar una segunda velocidad de transmisión de bits de la segunda transmisión de audio (124).

12. El procedimiento (600) de la reivindicación 11, en el que la primera velocidad de transmisión de bits y la segunda velocidad de transmisión de bits se ajustan basándose en los ajustes definidos por el usuario, las capacidades de hardware del dispositivo móvil (102), o una combinación de las mismas.

13. El procedimiento (600) de la reivindicación 1, en el que cada uno del dispositivo móvil (102), el primer dispositivo (110) y el segundo dispositivo (120) comprende un equipo de usuario, UE, que es compatible con un estándar de Proyecto de Asociación de Tercera Generación, 3GPP; o en el que la primera transmisión de audio (114) se recibe a través de una antena del dispositivo móvil (102); o en el que la primera memoria intermedia (210), la segunda memoria intermedia (220) y el controlador de retardo (202) están incluidos en un módem del dispositivo móvil (102); o en el que la primera memoria intermedia (210) comprende una primera memoria intermedia de eliminación de fluctuaciones, y en el que la segunda memoria intermedia (220) comprende una segunda memoria intermedia de eliminación de fluctuaciones.

14. Un dispositivo móvil (102) para gestionar audio durante una conferencia, con el dispositivo móvil (102) que comprende:

una primera memoria intermedia (210) configurada para recibir una primera transmisión de audio (114) desde un primer dispositivo (110) asociado con un primer participante de una conferencia;

una segunda memoria intermedia (220) configurada para recibir una segunda transmisión de audio (124) desde un segundo dispositivo (120) asociado con un segundo participante de la conferencia; **caracterizado por que** el dispositivo comprende además

un controlador de retardo (202) configurado para generar una señal de control (204), con la señal de control proporcionada (204) a la primera memoria intermedia (210) y a la segunda memoria intermedia (220) para sincronizar el primer audio de memoria intermedia (214) que se envía desde la primera memoria intermedia (210) con el segundo audio de memoria intermedia (224) que se envía desde la segunda memoria intermedia (220).

15. Un medio legible por ordenador no transitorio que comprende instrucciones para gestionar el audio durante una conferencia, con las instrucciones, cuando son ejecutadas por un procesador (706) en un dispositivo móvil (102), que hacen que el procesador (706) realice operaciones que comprenden:

recibir, en una primera memoria intermedia (210), una primera transmisión de audio (114) de un primer dispositivo (110) asociado con un primer participante de la conferencia;

recibir, en una segunda memoria intermedia (220), una segunda transmisión de audio (124) de un segundo dispositivo (120) asociado con un segundo participante de la conferencia; con las operaciones **caracterizadas** además **por**

generar una señal de control (204) en un controlador de retardo (202), con la señal de control (204) proporcionada a la primera memoria intermedia (210) y a la segunda memoria intermedia (220) para sincronizar el primer audio de memoria intermedia (214) que se envía desde la primera memoria intermedia (210) con segundo audio de memoria intermedia (224) que se envía desde la segunda memoria intermedia (220).

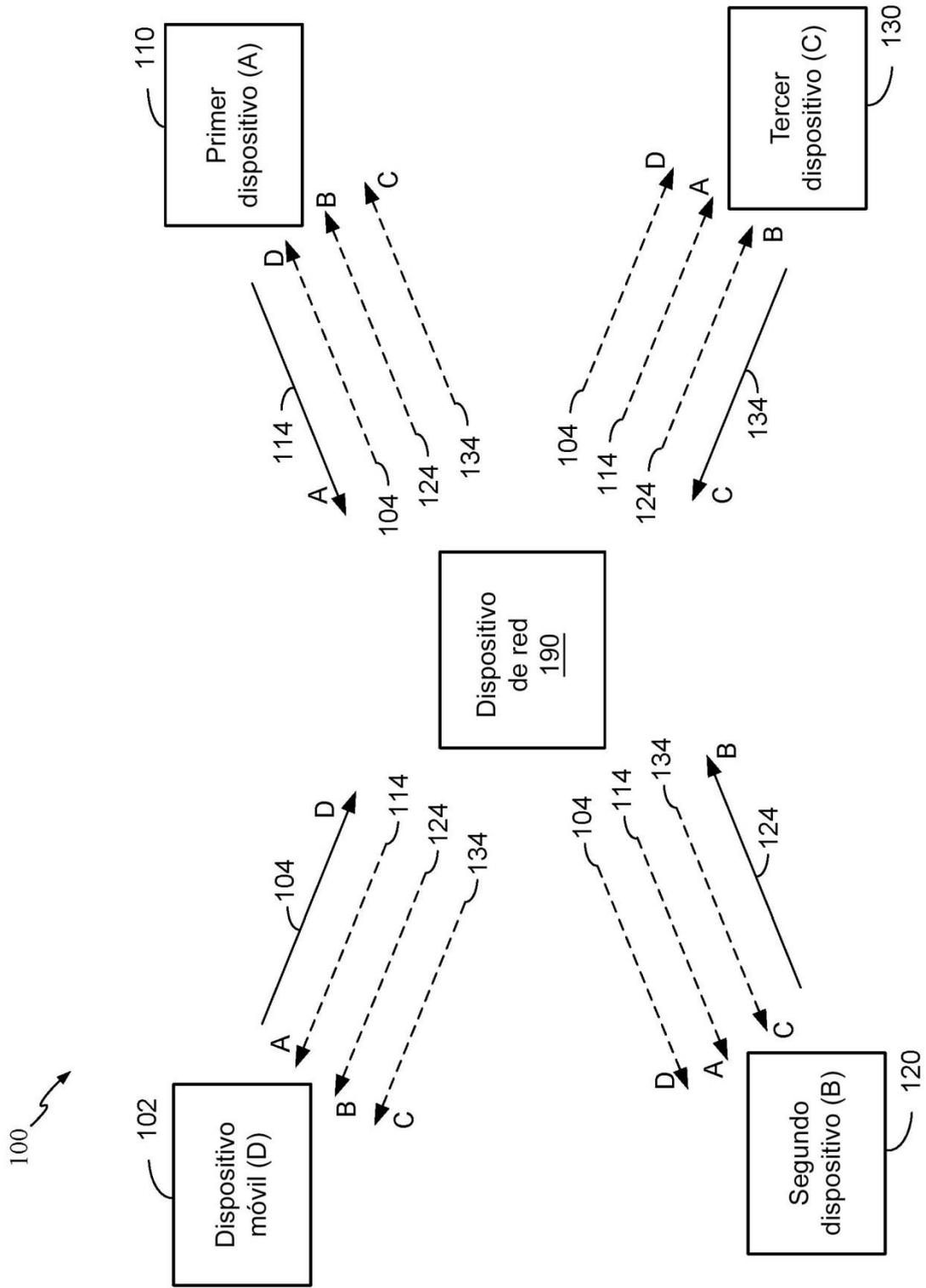


FIG. 1A

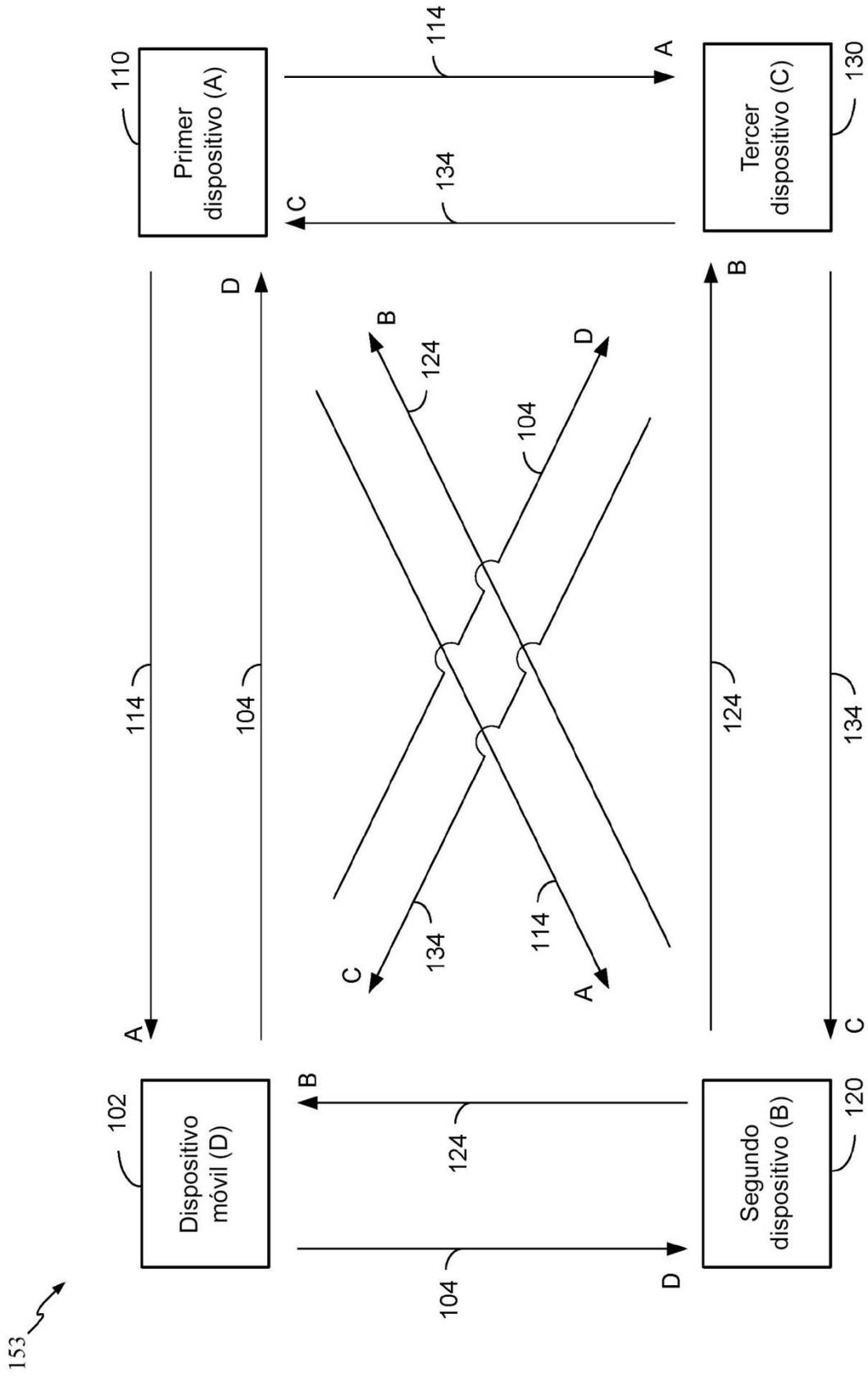


FIG. 1B

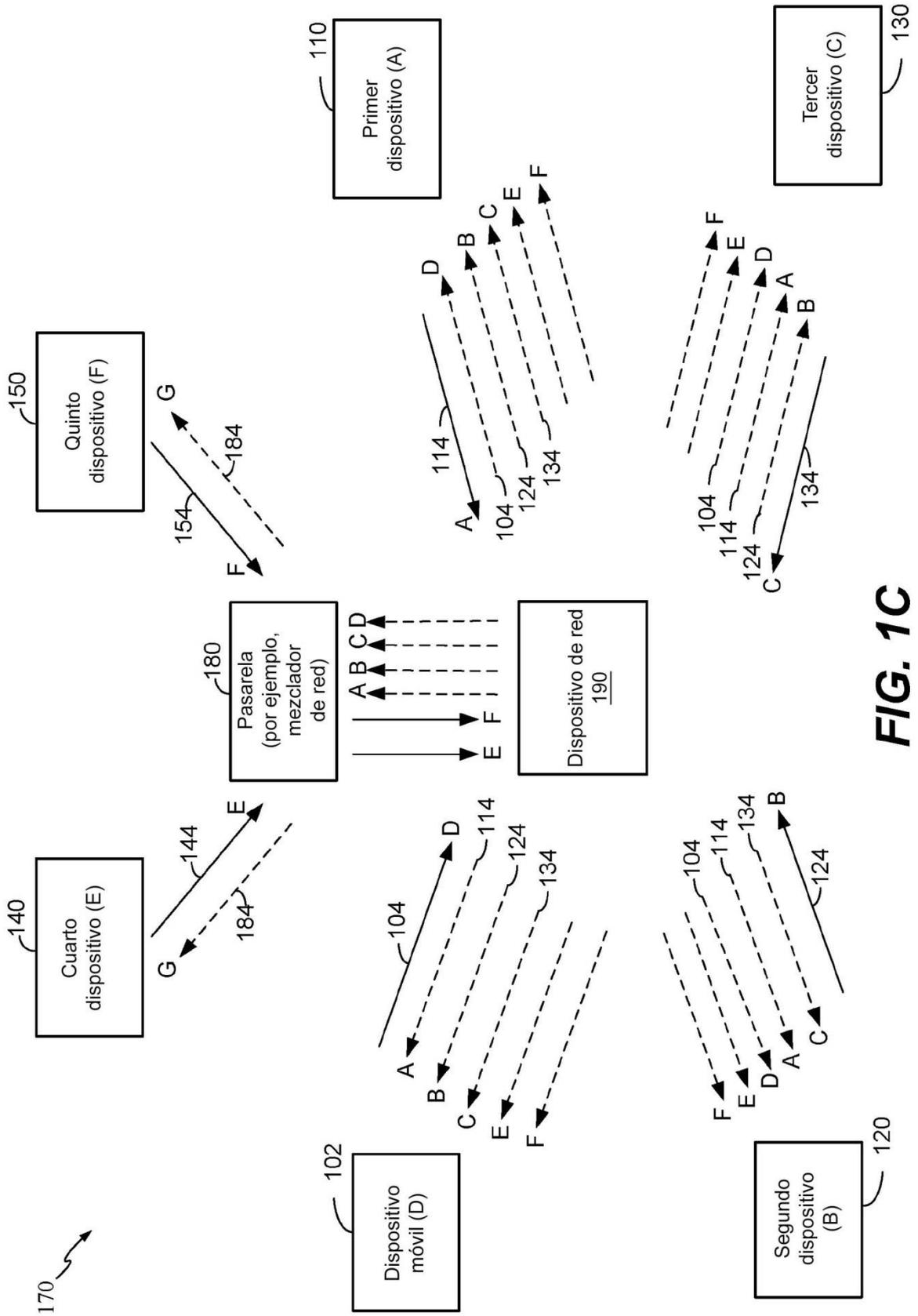


FIG. 1C

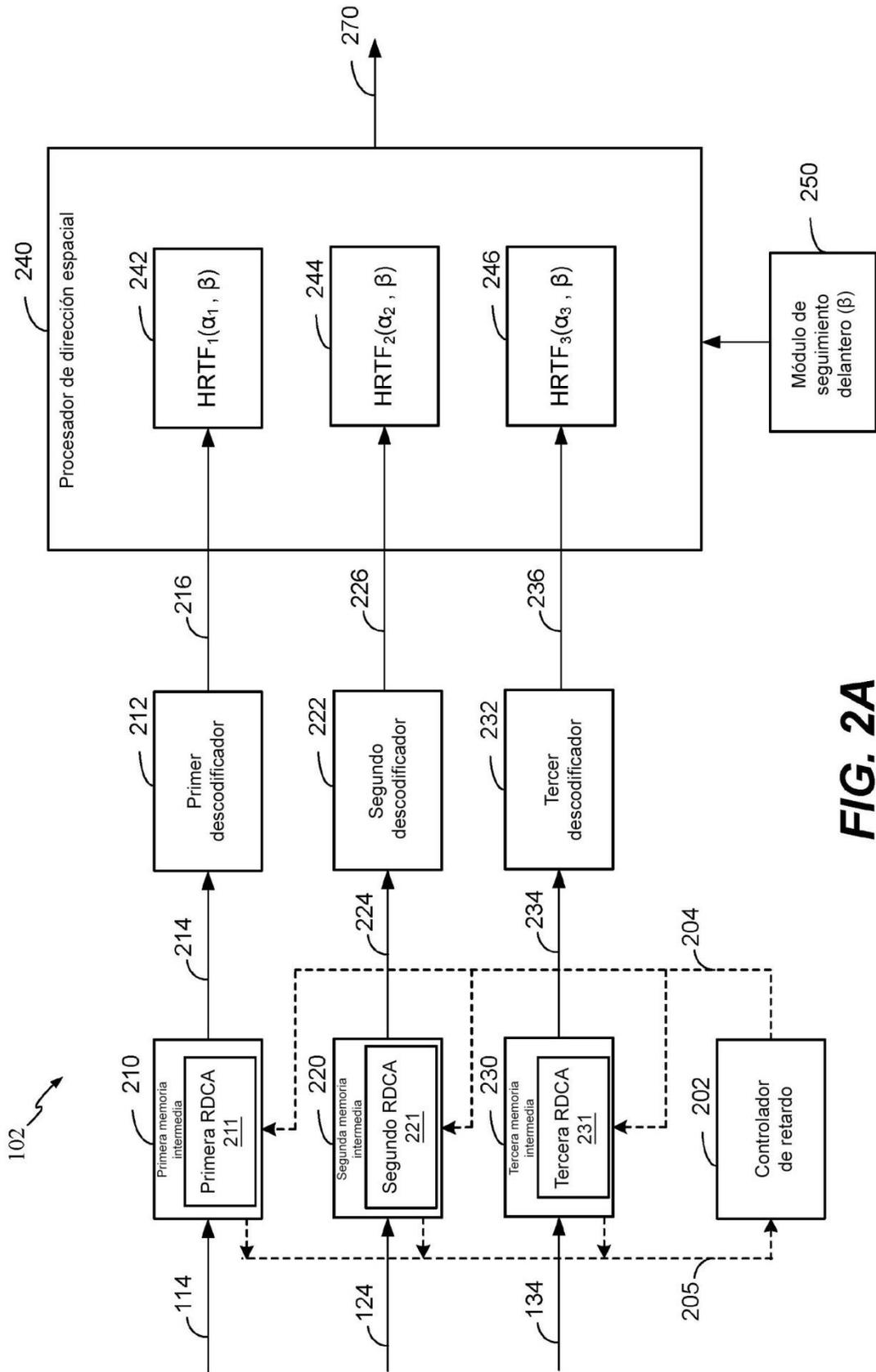


FIG. 2A

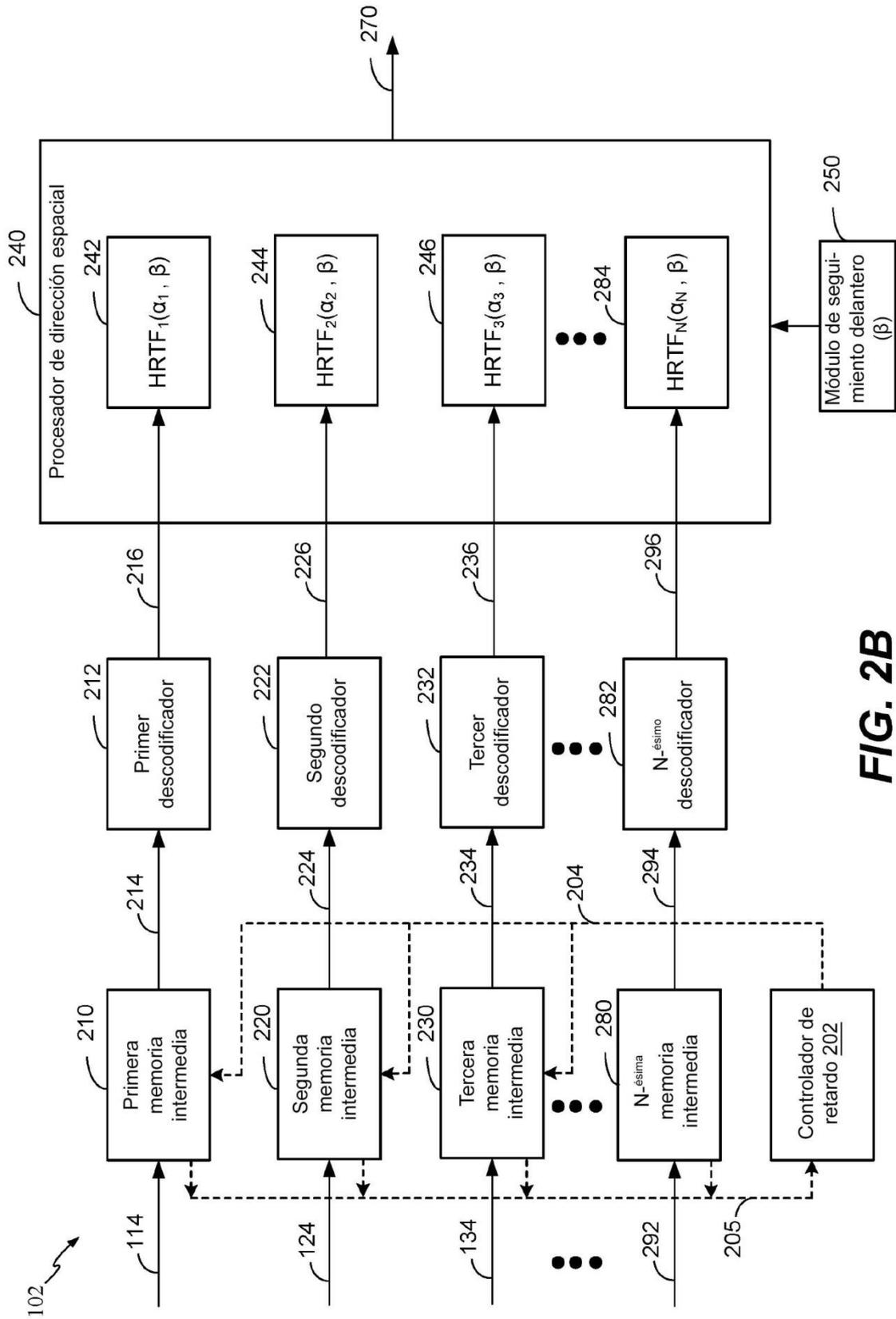


FIG. 2B

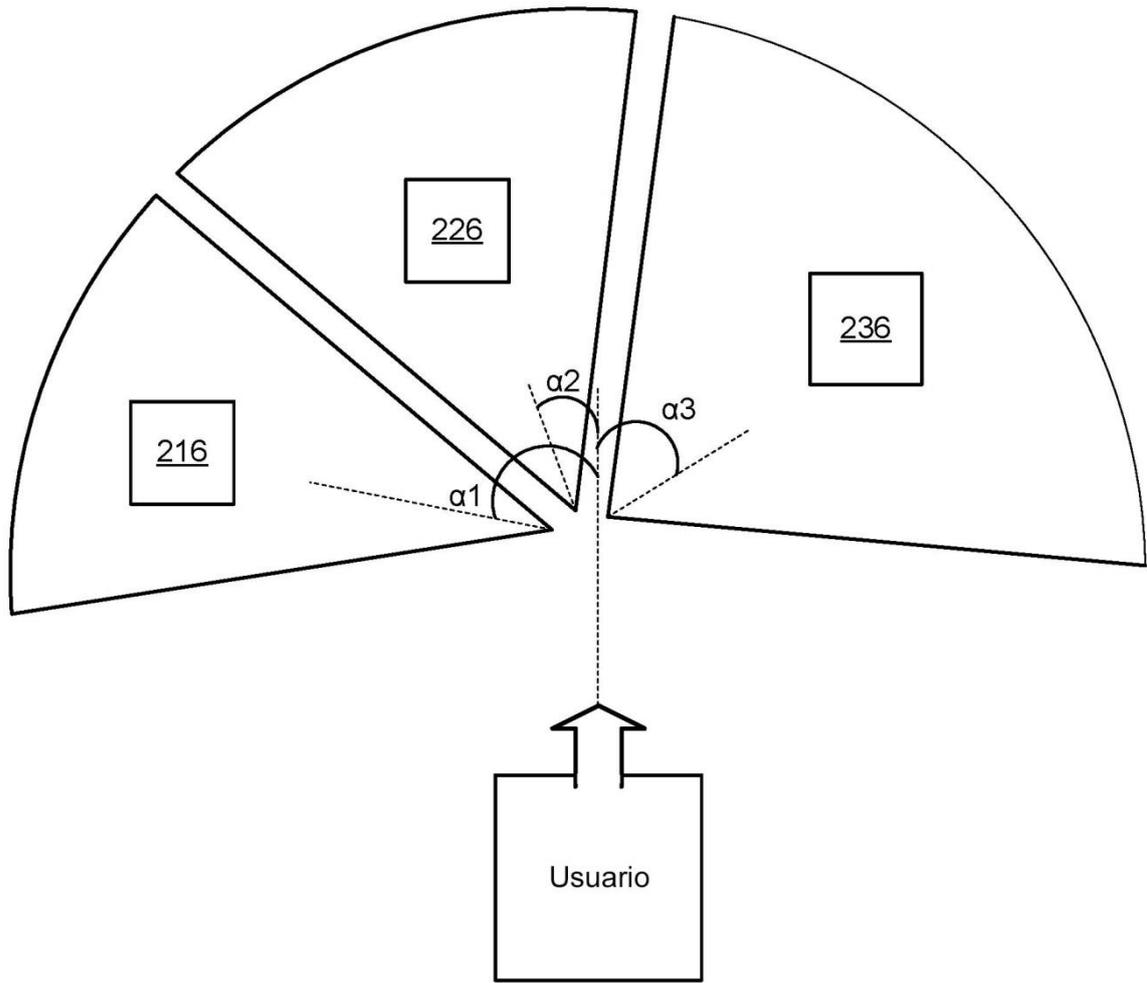


FIG. 3

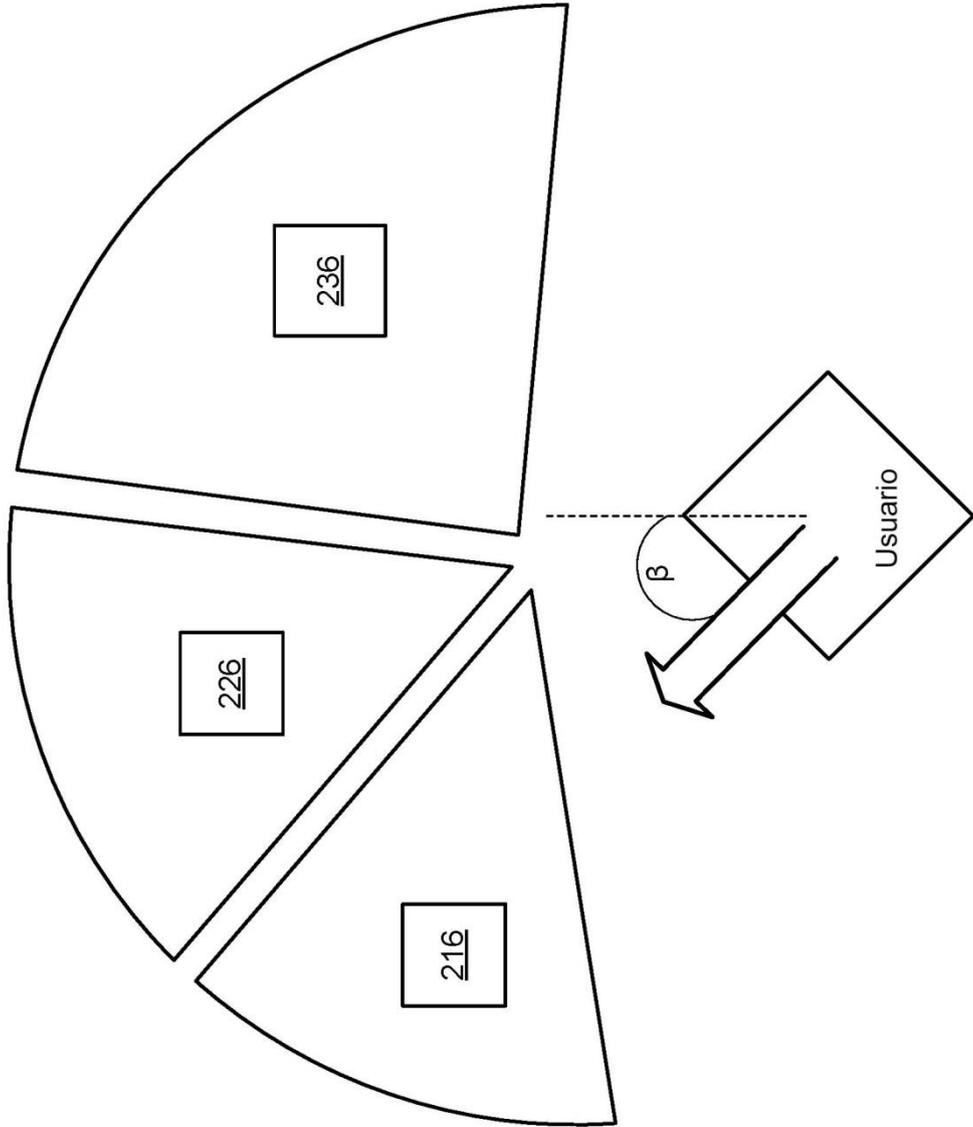


FIG. 4

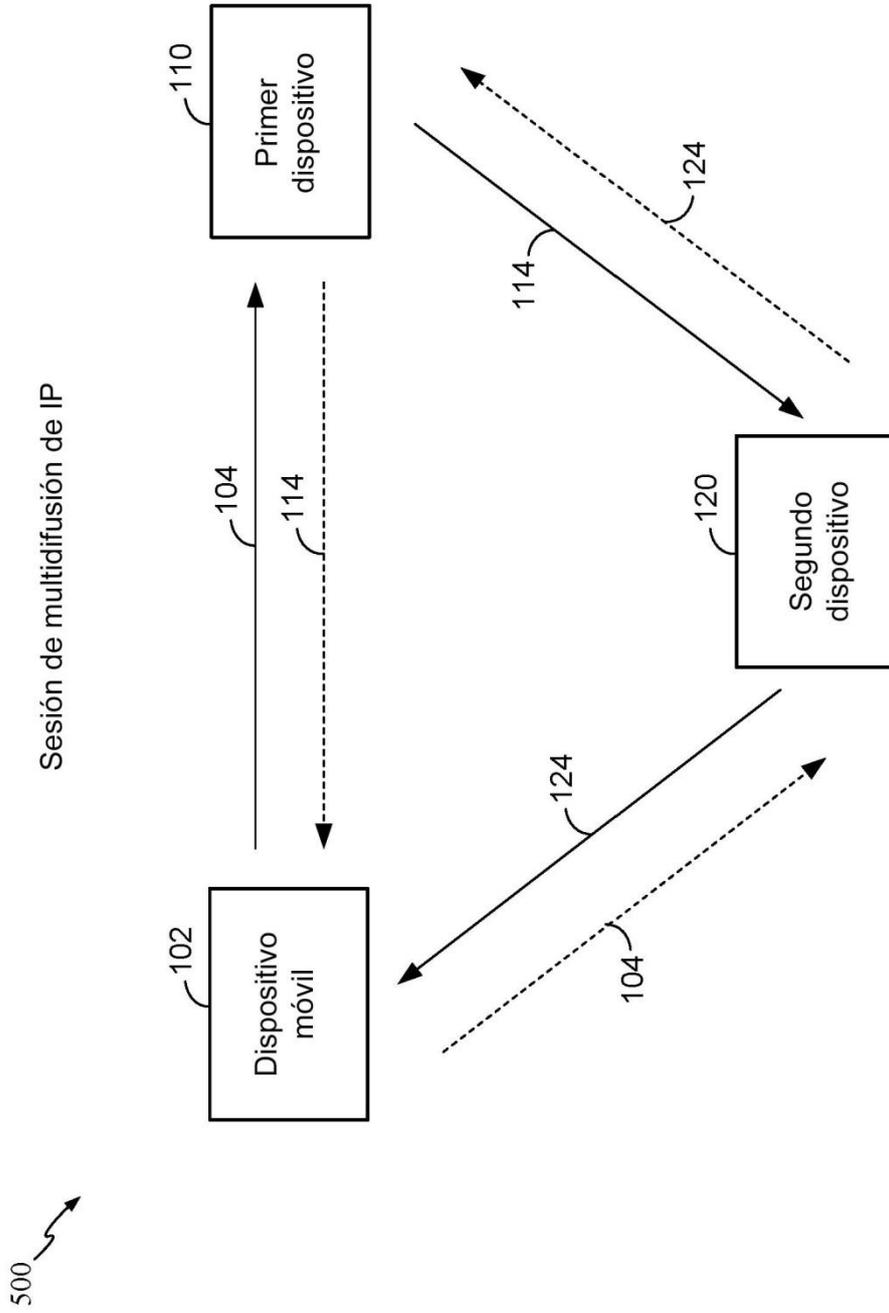


FIG. 5

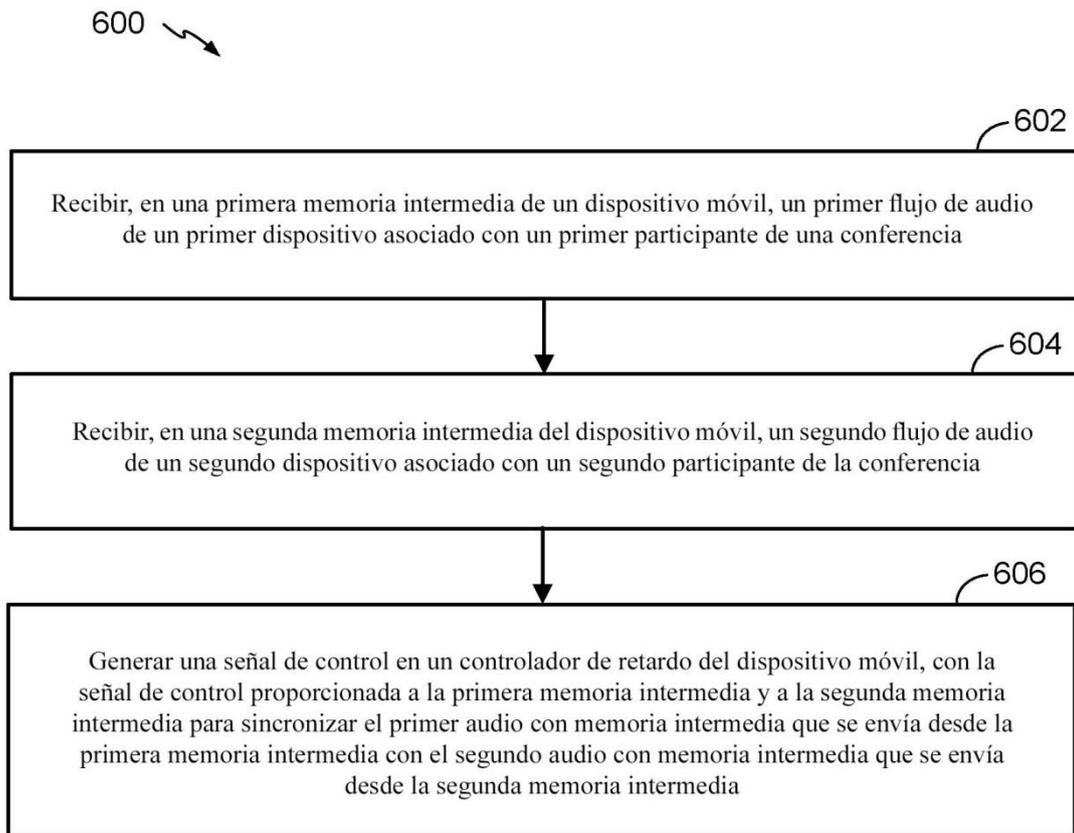


FIG. 6

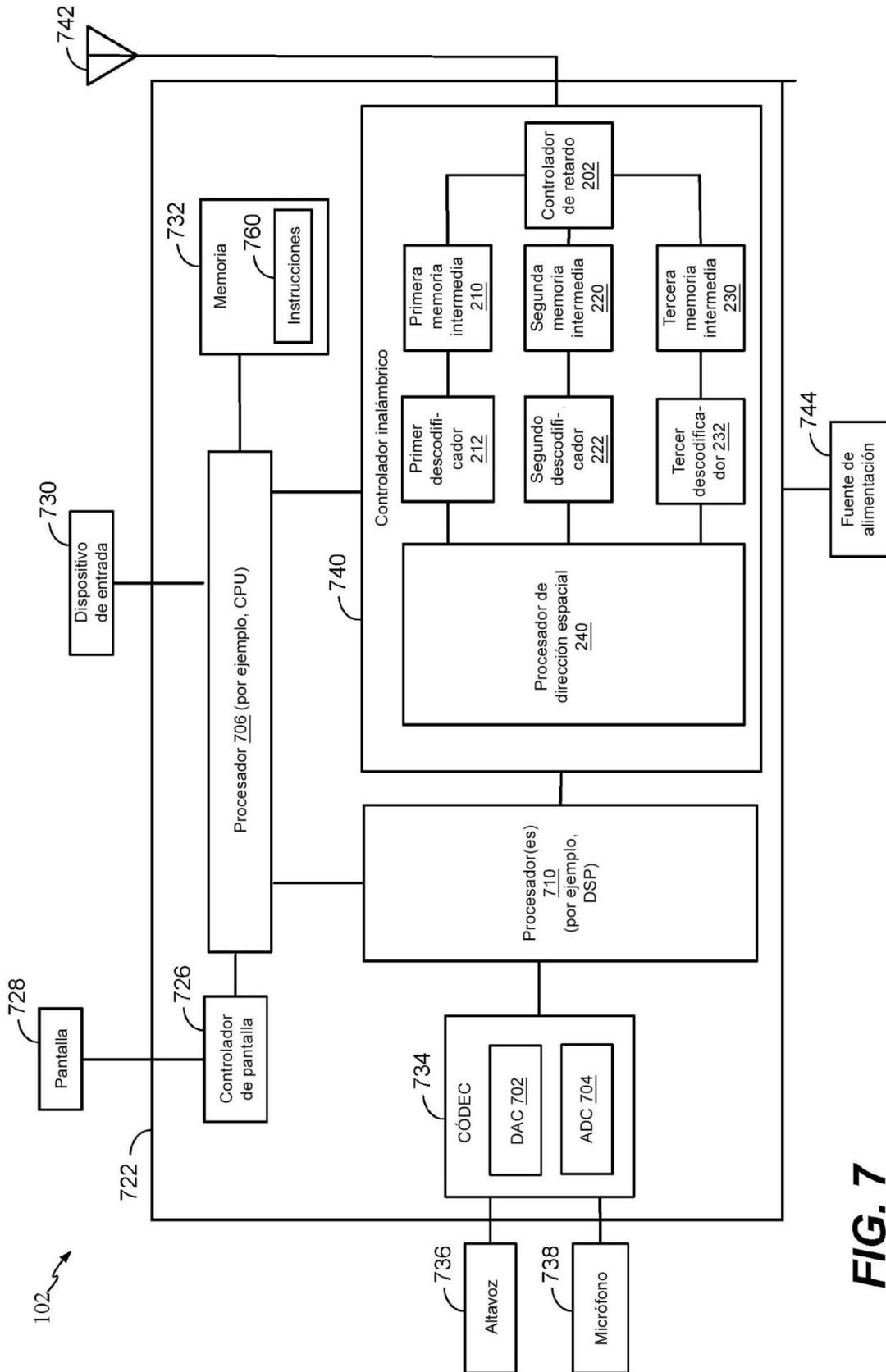


FIG. 7