

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 989 982**

51 Int. Cl.:

H04L 65/1069 (2012.01)

H04L 65/403 (2012.01)

H04L 65/80 (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.06.2022** **PCT/US2022/073156**

87 Fecha y número de publicación internacional: **29.12.2022** **WO22272304**

96 Fecha de presentación y número de la solicitud europea: **24.06.2022** **E 22747551 (4)**

97 Fecha y número de publicación de la concesión europea: **03.07.2024** **EP 4360290**

54 Título: **Método y sistema para teleconferencia usando dispositivos móviles coordinados**

30 Prioridad:

24.06.2021 US 202163202796 P

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

28.11.2024

73 Titular/es:

AFINITI, LTD. (100.0%)
Crawford House, 50 Cedar Avenue
Hamilton HM11, BM

72 Inventor/es:

SHAHMEER, MUHAMMAD;
AHMAD, KHUBAIB;
BASHIR, MUHAMMAD USMAN y
SAJJAD, MUHAMMAD ABDULLAH

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 989 982 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

Método y sistema para teleconferencia usando dispositivos móviles coordinados

Campo técnico

5 La presente divulgación se refiere, en general, a sistemas de comunicación. Más particularmente, los aspectos de la presente divulgación se refieren a una aplicación que permite la configuración de un sistema de comunicación de sala-a-sala usando dispositivos móviles.

Antecedentes

10 Las teleconferencias han adquirido recientemente una mayor importancia, especialmente con el advenimiento de la pandemia COVID-19. Las teleconferencias tradicionales han implicado el uso de un sistema telefónico con hardware especializado para los participantes en diferentes salas. Sin embargo, los cambios en la accesibilidad de la transmisión de vídeo e Internet han cambiado la infraestructura fundamental para las teleconferencias. Por ejemplo, el mundo está evolucionando desde GSM y PSTN a sistemas basados en voz sobre protocolo de Internet (VoIP). VoIP ha reducido drásticamente el coste de la comunicación al compartir la infraestructura de red entre datos y voz. Algunos servicios VoIP son gratuitos. Este modelo permite a los usuarios descargar aplicaciones de software a un dispositivo informático. Dicho software permite realizar llamadas de bajo coste a otros usuarios del servicio, e incluso, en algunos casos, permite realizar llamadas domésticas a teléfonos fijos y a teléfonos móviles sin cargo.

20 Los recientes avances no se han integrado en las teleconferencias, tal como las situaciones en las que hay varios participantes en una sala que se comunican con otro grupo remoto. En dichos casos, frecuentemente deben usarse los sistemas telefónicos tradicionales, ya que los diferentes sistemas basados en VoIP son incompatibles entre sí y, de esta manera, no pueden implementarse de manera fiable entre diferentes grupos de participantes. Por ejemplo, varias personas pueden estar amontonadas alrededor de un único teléfono móvil o teléfono con altavoz en una sala para facilitar una teleconferencia con participantes remotos.

25 De esta manera, existe la necesidad de una aplicación de dispositivo de usuario que permita el establecimiento de un sistema de teleconferencia ad hoc entre participantes en diferentes ubicaciones. Existe una necesidad adicional de establecer una comunicación entre dispositivos en una sala y una comunicación con dispositivos externos mediante el establecimiento de un dispositivo líder. Existe una necesidad adicional de una rutina para seleccionar una única secuencia de entre múltiples secuencias de audio en una sala para la comunicación con un dispositivo externo.

30 El documento US 2016/0014373 describe un sistema y un método de videoconferencia/audiokonferencia basado en la nube que realiza una agregación basada en la ubicación de una transmisión multimedia full-dúplex para organizar múltiples conexiones a una conferencia telefónica originadas desde la misma ubicación física o una ubicación compartida. El sistema de videoconferencia/audiokonferencia basado en la nube realiza una sincronización de las señales de audio del micrófono y del altavoz de las conexiones en la misma ubicación. De esta manera, puede realizarse una conferencia telefónica con múltiples dispositivos de usuario que realizan conexiones desde la misma ubicación física.

35 El documento WO 2021/119090 describe un puente de conferencia que organiza a los clientes colocados en grupos, elige un orador elegido por cada grupo y envía señales solo a los oradores elegidos. Los oradores elegidos mezclan el audio procedente de otros grupos, lo comparten en el interior de sus grupos usando conexiones locales de baja latencia y reproducen el audio después de un retraso.

Sumario

40 La invención está definida por un método para establecer un sistema de teleconferencia entre dos ubicaciones según la reivindicación 1, un dispositivo electrónico configurado para realizar las operaciones del método según la reivindicación 14 y un medio de almacenamiento legible por ordenador no transitorio según la reivindicación 15.

Las posibles características del método se indican en las reivindicaciones dependientes. Las realizaciones en la descripción que no están comprendidas dentro del alcance de las reivindicaciones deben considerarse como ejemplos ilustrativos para una mejor comprensión de la invención.

45 Otro ejemplo ilustrativo y relacionado divulgado es un método para establecer un sistema de teleconferencia entre dos ubicaciones. Se transmite una primera señal entre un dispositivo de comunicación líder y un dispositivo de comunicación cliente. Cada dispositivo de comunicación incluye un transceptor de comunicación de red, un altavoz y un micrófono. Se recibe una primera entrada de audio en el dispositivo líder generada desde una fuente de audio. La primera entrada de audio se analiza y se convierte en una primera pluralidad de paquetes de audio. Se recibe una segunda entrada de audio en el dispositivo cliente generada desde la fuente de audio. La segunda entrada de audio se analiza y se convierte en una segunda pluralidad de paquetes de audio. El dispositivo líder determina una primera métrica de al menos un primer paquete de audio en la primera pluralidad de paquetes de audio. El dispositivo líder determina una segunda métrica de al menos un segundo paquete de audio en la segunda pluralidad de paquetes de audio. Se determina si la primera métrica es o no mayor que la segunda métrica. Al menos un primer paquete de audio en la primera pluralidad de paquetes de audio y un número predeterminado de paquetes de audio restantes en la primera pluralidad de paquetes de audio se transmiten a un

dispositivo remoto en base a la determinación de que la primera métrica es mayor que la segunda métrica.

Una implementación adicional del otro método ilustrativo y ejemplar relacionado es una realización en la que el método incluye además transmitir el al menos un segundo paquete de audio en la segunda pluralidad de paquetes de audio y un número predeterminado de paquetes de audio restantes en la segunda pluralidad de paquetes de audio al dispositivo remoto en base a la determinación de que la primera métrica no es mayor que la segunda métrica. Otra implementación es una en la que el método ejemplar incluye recibir una señal visual desde el dispositivo remoto; y mostrar la señal visual en las pantallas del primer dispositivo y del segundo dispositivo. Otra implementación es una en la que la transmisión al dispositivo remoto se realiza usando un protocolo de voz sobre Internet a través de Internet. Otra implementación es una en la que el método ejemplar incluye además recibir una segunda señal desde el dispositivo remoto y emitir una tercera señal de audio en el dispositivo de comunicación líder en base a la segunda señal. El método incluye también emitir una cuarta señal de audio en el dispositivo cliente en base a la segunda señal, en el que la cuarta señal de audio está desfasada con relación a la tercera señal de audio. Otra implementación es una en la que la segunda señal comprende una señal de audio. Otra implementación es una en la que cada uno de entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente son teléfonos inteligentes. Otra implementación es una en la que la transmisión de una primera señal entre un dispositivo de comunicación líder y un dispositivo de comunicación cliente comprende además que el dispositivo de comunicación cliente establezca una comunicación de red con el dispositivo de comunicación líder mediante la recepción de información de red a través de un indicador visual mostrado por el dispositivo de comunicación líder. Otra implementación es una en la que la transmisión de una primera señal entre un dispositivo de comunicación líder y un dispositivo de comunicación cliente comprende además que el dispositivo de comunicación cliente establezca una comunicación de red con el dispositivo de comunicación líder mediante la recepción de información de red a través de una comunicación inalámbrica. Otra implementación es una en la que la primera métrica se basa en una energía media del al menos un paquete de audio. Otra implementación es una en la que la primera señal comprende una transmisión que usa un canal de datos Web RTC. Otra implementación es una en la que la transmisión del al menos un primer paquete de audio en la primera pluralidad de paquetes de audio y un número predeterminado de paquetes de audio restantes en la primera pluralidad de paquetes de audio a un dispositivo remoto comprende una transmisión que usa un canal de audio WebRTC. Otra implementación es una en la que la primera señal comprende una primera frecuencia. Otra implementación es una en la que el método ejemplar incluye además transmitir una segunda señal que comprende una segunda frecuencia entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente; y sincronizar el dispositivo de comunicación líder y el dispositivo de comunicación cliente en base a al menos una de entre la primera señal y la segunda señal.

Otro ejemplo ilustrativo y relacionado divulgado es un método para configurar un sistema de teleconferencia entre múltiples dispositivos móviles en una sala. Se establece un canal de comunicación de red entre un primer dispositivo móvil de entre los múltiples dispositivos móviles y un segundo dispositivo móvil de entre los múltiples dispositivos móviles. El primer dispositivo móvil se designa como un dispositivo líder. El segundo dispositivo móvil se designa como un dispositivo cliente. Se recibe una comunicación con protocolo de voz sobre Internet desde un dispositivo remoto al primer dispositivo móvil a través de Internet. La comunicación con protocolo de voz sobre Internet desde el primer dispositivo móvil se transmite al segundo dispositivo móvil. La comunicación con protocolo de voz sobre Internet se reproduce en el primer dispositivo móvil y en el segundo dispositivo móvil.

Una implementación adicional del otro método ilustrativo y ejemplar relacionado es una realización en la que la comunicación de red entre los dispositivos móviles primero y segundo se establece mediante la recepción de información de red a través de un indicador visual mostrado por el primer dispositivo móvil. Otra implementación es una en la que la comunicación de red entre los dispositivos móviles primero y segundo se establece mediante la recepción de información de red a través de una comunicación inalámbrica desde el primer dispositivo móvil. Otra implementación es una en la que el método ejemplar incluye además sincronizar el primer dispositivo móvil y el segundo dispositivo móvil para sincronizar la reproducción de la comunicación recibida desde el dispositivo remoto.

Otro ejemplo ilustrativo y relacionado divulgado es un medio de almacenamiento legible por ordenador no transitorio que incluye instrucciones que, cuando son ejecutadas por al menos un procesador de un primer dispositivo móvil, causan que el primer dispositivo móvil realice operaciones que incluyen establecer un canal de comunicación de red entre un segundo dispositivo móvil y el primer dispositivo móvil. Las operaciones incluyen designar el primer dispositivo móvil como un dispositivo líder; y designar un segundo dispositivo móvil como un dispositivo cliente. Las operaciones incluyen recibir una comunicación con protocolo de voz sobre Internet desde un dispositivo remoto a través de Internet. Las operaciones incluyen el envío de la comunicación desde el dispositivo remoto al segundo dispositivo móvil con instrucciones para reproducir la comunicación en un tiempo seleccionado. Las operaciones incluyen reproducir la comunicación desde el dispositivo remoto en el primer dispositivo móvil en sincronización con el tiempo seleccionado.

Otro ejemplo ilustrativo y relacionado divulgado es un sistema de teleconferencia para participantes en una sala. El sistema incluye la recepción por parte de un primer dispositivo móvil de una comunicación con protocolo de voz sobre Internet desde un dispositivo remoto a través de Internet. El primer dispositivo móvil se designa como un dispositivo líder. Un segundo dispositivo móvil se designa como un dispositivo cliente mediante el establecimiento de un canal de comunicación de red con el primer dispositivo móvil. El primer dispositivo móvil está configurado para enviar la comunicación desde el dispositivo remoto al segundo dispositivo móvil. Los dispositivos móviles primero y segundo están configurados para reproducir la comunicación recibida desde el dispositivo remoto.

- Otro ejemplo ilustrativo y relacionado divulgado es un dispositivo móvil para configurar un sistema de teleconferencia. El dispositivo móvil incluye un micrófono; un altavoz; una interfaz de red; y un procesador acoplado al micrófono, al altavoz y a la interfaz de red. El procesador está configurado para recibir una comunicación con protocolo de voz sobre Internet desde un dispositivo remoto a través de la interfaz de red. El procesador está configurado para establecer una comunicación con un segundo dispositivo móvil en la sala. El procesador está configurado para enviar la comunicación desde el dispositivo remoto al segundo dispositivo móvil. El procesador está configurado para reproducir la comunicación recibida desde el dispositivo remoto en el altavoz en sincronización con la reproducción por parte del segundo dispositivo móvil de la comunicación recibida.
- Tal como se ha descrito anteriormente, un método para establecer un sistema de teleconferencia entre dos ubicaciones se indica en la reivindicación 1.
- Las implementaciones pueden incluir una o más de las siguientes características. Por ejemplo, en algunas implementaciones, el método comprende determinar que la primera métrica es mayor que la segunda métrica, y la transmisión del conjunto de paquetes de audio comprende transmitir al menos un paquete de audio en la primera pluralidad de paquetes de audio y el número predeterminado de paquetes de audio adicionales en la primera pluralidad de paquetes de audio al dispositivo remoto en base a la determinación de que la primera métrica es mayor que la segunda métrica.
- En algunas implementaciones, el método comprende determinar que la primera métrica no es mayor que la segunda métrica, y la transmisión del conjunto de paquetes de audio comprende transmitir al menos un paquete de audio en la segunda pluralidad de paquetes de audio y el número predeterminado de paquetes de audio adicionales en la segunda pluralidad de paquetes de audio al dispositivo remoto en base a la determinación de que la primera métrica no es mayor que la segunda métrica.
- En algunas implementaciones, el método comprende: recibir una señal visual desde el dispositivo remoto; y mostrar la señal visual en las pantallas del dispositivo líder y del dispositivo cliente.
- En algunas implementaciones, la transmisión al dispositivo remoto se realiza usando un protocolo de voz sobre Internet a través de Internet.
- En algunas implementaciones, el método comprende: recibir una segunda señal desde el dispositivo remoto; emitir una tercera señal de audio en el dispositivo de comunicación líder en base a la segunda señal; y emitir una cuarta señal de audio en el dispositivo cliente en base a la segunda señal, en el que la cuarta señal de audio está desfasada con relación a la tercera señal de audio.
- En algunas implementaciones, la segunda señal comprende una señal de audio.
- En algunas implementaciones, cada uno de entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente son teléfonos inteligentes.
- En algunas implementaciones, la transmisión de la primera señal entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente comprende además el establecimiento, por parte del dispositivo de comunicación líder, de una comunicación de red con el dispositivo de comunicación cliente en base a la recepción por parte del dispositivo de comunicación cliente de información de red a través de un indicador visual mostrado por el dispositivo de comunicación líder.
- En algunas implementaciones, la transmisión de una primera señal entre un dispositivo de comunicación líder y un dispositivo de comunicación cliente comprende además el establecimiento de una comunicación de red, por parte del dispositivo de comunicación líder, con el dispositivo de comunicación cliente en base a la recepción por parte del dispositivo de comunicación cliente de información de red a través de una comunicación inalámbrica.
- En algunas implementaciones, la primera métrica se basa en una energía media del al menos un paquete de audio.
- En algunas implementaciones, la primera señal comprende una transmisión que usa un canal de datos Web RTC.
- En algunas implementaciones, la transmisión del al menos un primer paquete de audio en la primera pluralidad de paquetes de audio y un número predeterminado de paquetes de audio restantes en la primera pluralidad de paquetes de audio a un dispositivo remoto comprende una transmisión que usa un canal de audio WebRTC.
- En algunas implementaciones, la primera señal comprende una primera frecuencia.
- En algunas implementaciones, el método comprende: transmitir una segunda señal que comprende una segunda frecuencia entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente; y sincronizar el dispositivo de comunicación líder y el dispositivo de comunicación cliente en base a al menos una de entre la primera señal y la segunda señal.
- En algunas implementaciones, el método comprende la obtención, por parte del dispositivo de comunicación líder, de un código para una llamada o conferencia que incluye una comunicación a través de una red entre el dispositivo de comunicación líder y al menos otro dispositivo de comunicación líder. La transmisión de la primera señal comprende la

emisión, por parte del dispositivo de comunicación líder, del código. El método comprende además, después de emitir el código, permitir la participación del dispositivo de comunicación cliente en la llamada o conferencia en respuesta a una comunicación que el dispositivo de comunicación cliente proporciona en base al código, de manera que el dispositivo de comunicación líder reciba datos de audio grabados por el dispositivo de comunicación cliente y el dispositivo de comunicación líder proporcione datos de audio desde el al menos otro dispositivo de comunicación líder al dispositivo de comunicación cliente.

En algunas implementaciones, el código identifica (i) la llamada o conferencia y (ii) una sala, de entre múltiples salas implicadas en la llamada o conferencia, que corresponde al dispositivo de comunicación líder.

En algunas implementaciones, el código es un código óptico legible por máquina.

En algunas implementaciones, el código es un código de respuesta rápida (QR).

En algunas implementaciones, el método comprende: el mantenimiento, por parte del dispositivo de comunicación líder, de la comunicación con cada uno de los múltiples dispositivos remotos que están implicados simultáneamente en una llamada o conferencia; la recepción, por parte del dispositivo de comunicación líder, de datos de audio desde cada uno de los dispositivos remotos; el mezclado, por parte del dispositivo de comunicación líder, de los datos de audio recibidos desde los múltiples dispositivos líderes; y la provisión, por parte del dispositivo de comunicación líder, de los datos de audio mezclados al dispositivo de comunicación cliente.

En algunas implementaciones, a los paquetes de audio en la primera pluralidad de paquetes de audio se les asignan números de secuencia y a los paquetes de audio en la segunda pluralidad de paquetes de audio se les asignan números de secuencia; en el que el método comprende además correlacionar los números de secuencia de la primera pluralidad de paquetes de audio y de la segunda pluralidad de paquetes de audio; y en el que el al menos un paquete de audio de la primera pluralidad de paquetes de audio y el al menos un paquete de audio de la segunda pluralidad de paquetes de audio se seleccionan en base a la correlación de los números de secuencia.

En algunas implementaciones, la primera métrica es una medida de claridad del al menos un paquete de audio de la primera pluralidad de paquetes de audio; en el que la segunda métrica es una medida de claridad del al menos un paquete de audio de la segunda pluralidad de paquetes de audio; y en el que el método comprende la determinación, por parte del dispositivo de comunicación líder, de si seleccionar el conjunto de paquetes de audio a transmitir al dispositivo remoto desde la primera pluralidad de paquetes de audio o la segunda pluralidad de paquetes de audio en base a cuál de entre la primera métrica y la segunda métrica indica una mayor claridad.

En algunas implementaciones, la determinación de la primera métrica comprende determinar la primera métrica en base a una serie de múltiples paquetes de audio desde la primera pluralidad de paquetes de audio que codifican un audio que ocurre durante un período de tiempo; la determinación de la segunda métrica comprende determinar la segunda métrica en base a una serie de múltiples paquetes de audio desde la segunda pluralidad de paquetes de audio que codifican un audio que ocurre durante el período de tiempo y el número predeterminado de paquetes de audio es mayor que un número de paquetes de audio evaluados para determinar la primera métrica o la segunda métrica.

En algunas implementaciones, el dispositivo de comunicación líder está situado en una sala, y el dispositivo de comunicación líder está configurado para realizar una serie de evaluaciones de transmisiones de paquetes de audio desde múltiples dispositivos en la sala y está configurado para variar cuál de las secuencias se transmite al dispositivo remoto en base a los resultados de las evaluaciones. Cada una de las evaluaciones comprende (i) generar métricas para paquetes de audio de las secuencias respectivas que corresponden a un período de tiempo de evaluación que tiene una duración predeterminada y (ii) seleccionar una de las secuencias en base a las métricas. Para cada una de las evaluaciones, el dispositivo de comunicación líder está configurado para transmitir, al dispositivo remoto, una secuencia contigua de paquetes desde la secuencia seleccionada que incluye (i) los paquetes de audio de la secuencia seleccionada que corresponden al período de tiempo de evaluación y (ii) paquetes de audio adicionales de la secuencia seleccionada que representan una duración de tiempo predeterminada después del final del período de tiempo de evaluación.

En algunas implementaciones, la duración de tiempo predeterminada es mayor que el periodo de tiempo de evaluación.

En algunas implementaciones, el método comprende además sincronizar los datos de audio de reproducción recibidos por el dispositivo de comunicación líder desde el dispositivo remoto a través de la red con la reproducción de los datos de audio por uno o más dispositivos remotos.

Otros ejemplos ilustrativos de estos y otros aspectos analizados en el presente documento incluyen sistemas, aparatos y programas informáticos correspondientes codificados en dispositivos de almacenamiento informático, configurados para realizar las acciones de los métodos. Un sistema de uno o más ordenadores puede configurarse de tal manera en virtud de software, firmware, hardware o una combinación de los mismos instalados en el sistema que, durante el funcionamiento, causan que el sistema realice las acciones. Uno o más programas informáticos pueden configurarse de tal manera en virtud de tener instrucciones que, cuando son ejecutadas por un aparato de procesamiento de datos, causan que el aparato realice las acciones.

El sumario anterior no pretende representar cada realización o cada aspecto de la presente divulgación. Más bien, el

sumario anterior simplemente proporciona un ejemplo de algunos de los aspectos y características novedosos establecidos en el presente documento. Las características y ventajas anteriores, y otras características y ventajas de la presente divulgación, serán fácilmente evidentes a partir de la siguiente descripción detallada de las realizaciones y los modos representativos para llevar a cabo la presente invención, cuando se consideran con relación a los dibujos adjuntos y las reivindicaciones adjuntas.

Breve descripción de los dibujos

La divulgación se comprenderá mejor a partir de la siguiente descripción de realizaciones ejemplares junto con una referencia a los dibujos adjuntos, en los que

La Fig. 1A es un diagrama de un sistema de teleconferencia ad hoc configurado entre dispositivos de usuario móviles de participantes en múltiples salas;

La Fig. 1B es un diagrama de la comunicación de una señal de audio en el sistema de la Fig. 1A;

La Fig. 2 es un diagrama de bloques de un dispositivo de usuario móvil ejemplar;

La Fig. 3A es un diagrama de flujo de la rutina para la comunicación entre un dispositivo líder y otro dispositivo de usuario móvil designado como un dispositivo líder cuando se recibe una señal desde un dispositivo de comunicación externo;

La Fig. 3B es un diagrama de flujo del módulo de preprocesamiento de audio usado en la Fig. 3A;

La Fig. 4A es un diagrama de flujo de la grabación y transmisión de señales de audio en un dispositivo móvil líder;

La Fig. 4B es un diagrama de los módulos de un módulo de procesamiento de audio (APM);

La Fig. 4C es un diagrama de flujo de recepción y reproducción de señales de audio recibidas desde dispositivos líderes remotos;

La Fig. 5 es un diagrama de flujo del proceso de un dispositivo cliente que se conecta con un dispositivo líder y la transmisión de señales de audio desde el dispositivo cliente al dispositivo líder;

La Fig. 6 es un diagrama de flujo del proceso ejemplar de un dispositivo cliente que recibe señales de audio desde un dispositivo líder para reproducir datos de audio;

La Fig. 7 es un diagrama de flujo del proceso combinado del procesamiento de los datos de audio recibidos por un dispositivo líder desde los dispositivos cliente y la transmisión de señales de audio desde el dispositivo líder a otro dispositivo líder;

La Fig. 8 es un diagrama de flujo del proceso de establecimiento de comunicación entre el dispositivo líder y los dispositivos cliente para grabar una señal de audio sincronizados entre sí;

La Fig. 9 es un diagrama del proceso de selección de paquetes de las mismas señales de audio desde diferentes dispositivos cliente para la selección del micrófono;

La Fig. 10 es un diagrama de flujo del proceso de selección de una de las secuencias de la misma señal de audio realizado por el dispositivo líder;

La Fig. 11 es un diagrama de flujo del algoritmo de sincronización de reproducción de un dispositivo líder; y

La Fig. 12 es un diagrama de flujo del algoritmo de sincronización de reproducción de un dispositivo cliente.

La presente divulgación es susceptible de diversas modificaciones y formas alternativas. Algunas realizaciones representativas se han mostrado a modo de ejemplo en los dibujos y se describirán en detalle en el presente documento. La invención está definida por las reivindicaciones adjuntas.

Descripción detallada de las realizaciones ilustradas

Las presentes invenciones pueden llevarse a la práctica de muchas formas diferentes. Las realizaciones representativas se muestran en los dibujos y se describirán en detalle en el presente documento. La presente divulgación es un ejemplo o ilustración de los principios de la presente divulgación, y no pretende limitar los amplios aspectos de la divulgación a las realizaciones ilustradas. En este sentido, los elementos y limitaciones que se divulgan, por ejemplo, en las secciones Resumen, Sumario y Descripción Detallada, pero que no se establecen explícitamente en las reivindicaciones, no deberían incorporarse a las reivindicaciones, individual o colectivamente, por implicación, inferencia o de otra manera. Para los propósitos de la presente descripción detallada, a menos que se rechace específicamente, el singular incluye el plural y viceversa; y la expresión "que incluye" significa "que incluye, sin limitación". Además, las palabras de aproximación, tales como "aproximadamente", "casi", "sustancialmente", "aproximadamente" y similares, pueden usarse en el presente documento de manera que signifiquen "en", "cerca de" o "casi en", o "dentro del 3-5% de", o "dentro de tolerancias de fabricación aceptables", o cualquier combinación lógica de los mismos, por ejemplo.

El objetivo de la presente divulgación es proporcionar una solución de teleconferencia en línea, mediante la cual todos los dispositivos de usuario de una sala de reuniones puedan unirse a todos los dispositivos de usuario de otra sala de reuniones remota. El sistema ejemplar se basa en una aplicación móvil VoIP disponible para diferentes sistemas operativos, tales como Apple iOS o Google Android. El sistema permite la integración de sistemas telefónicos más tradicionales, tales como GSM o PSTN, y el sistema divulgado actualmente permite la integración de otras aplicaciones VoIP.

El sistema ejemplar es una aplicación móvil que proporciona una solución de conferencia para múltiples participantes en diferentes salas mediante dispositivos de usuario móviles individuales. La aplicación móvil ejemplar y la infraestructura asociada proporcionan comunicación sala-a-sala, que es una solución de conferencia sin interrupciones sin hardware especializado, que solo requiere dispositivos de usuario móviles. El sistema ejemplar puede reemplazar los teléfonos de escritorio convencionales y los "teléfonos estrella" (por ejemplo, teléfonos con altavoz Polycom) en las salas de conferencias. El dispositivo de usuario móvil de cada participante proporciona un micrófono y un altavoz, que reemplazan el micrófono y el altavoz de un teléfono estrella conocido. La aplicación móvil ejemplar incluye la selección del micrófono y el altavoz, la sincronización en la reproducción, la cancelación de ruido, el control automático de ganancia en base a un factor calculado, la cancelación del eco acústico y la configuración de QoS (calidad de servicio) para gestionar la prioridad de los paquetes VoIP, por ejemplo, DSCP (punto de código de servicio diferenciado) para señales de audio. La aplicación móvil permite la configuración de un único dispositivo líder desde uno de los dispositivos de usuario móviles en una sala. Los otros dispositivos móviles en la sala están asociados como dispositivos cliente al dispositivo líder, que se comunica a través de una red de comunicación (por ejemplo, Internet) con uno o más dispositivos líderes situados en otras salas. Cada dispositivo líder recibe y procesa los datos de audio locales (por ejemplo, audio grabado en su propia sala) y distribuye los datos de audio procesados a través de la red a los otros dispositivos líderes en la conferencia. Cada dispositivo líder recibe y procesa también los datos de audio grabados de manera remota proporcionados por los otros dispositivos líderes, y distribuye los datos de audio entre los dispositivos cliente en su sala para una reproducción sincronizada. La comunicación entre los dispositivos líderes proporciona el intercambio de datos necesarios para establecer y mantener una conferencia entre múltiples salas, cada una de las cuales tiene múltiples dispositivos participantes.

Un ejemplo de los principios descritos es el producto CONFU. El sistema de teleconferencia ad hoc ejemplar proporciona una solución sin interrupciones que reemplaza los teléfonos de escritorio convencionales, los teléfonos estrella u otro hardware especializado en una sala de conferencias. El sistema proporciona un escenario de localización de sonido integrado y proporciona a los participantes una experiencia de audioconferencia fluida en un escenario de varias salas.

La Fig. 1A es un diagrama de bloques de un sistema 100 de comunicación ejemplar que puede establecerse entre participantes en salas 110, 112 y 114 diferentes. El sistema 100 permite una comunicación de conferencia entre los participantes en dos o más salas físicamente separadas, y en el ejemplo ilustrado se muestran tres de dichas salas 110, 112 y 114. Cada una de las salas 110, 112 y 114 incluye múltiples participantes, tales como los participantes 120a, 120b, 120c y 120d en la sala 110; los participantes 122a, 122b, 122c en la sala 112; y los participantes 124a y 124b en la sala 114. Cada uno de los participantes tiene acceso a un dispositivo de usuario móvil que es operable para la comunicación que les permite participar en una teleconferencia con los otros participantes en el sistema 100. En este ejemplo, los dispositivos de usuario incluyen dispositivos 130a-130c móviles, tales como teléfonos inteligentes correspondientes a los participantes 120a-120c; dispositivos 132a-132c móviles, tales como teléfonos inteligentes correspondientes a los participantes 122a-122c; y dispositivos 134a-134b móviles, tales como teléfonos inteligentes correspondientes a los participantes 124a-124b. Un participante, tal como el participante 120d, puede usar otros dispositivos 136 informáticos equipados con micrófonos o altavoces (por ejemplo, tabletas u ordenadores portátiles).

La aplicación ejemplar está instalada en cada uno de los dispositivos 130a-130c, 132a-132c y 134a-134b de usuario móviles mediante medios convencionales, tales como mediante la descarga de la aplicación desde la nube (por ejemplo, desde la App Store de Apple o la tienda Google Play). Cuando cada dispositivo 130a-130c, 132a-132c y 134a-134b móvil accede a la aplicación ejemplar, la aplicación ejemplar puede ofrecer enlaces correspondientes a las salas de conferencias disponibles al dispositivo móvil. El dispositivo móvil puede solicitar unirse a una de las salas de conferencias, por ejemplo, haciendo clic el usuario en un enlace asociado con esa sala de conferencias. En algunos ejemplos, la aplicación ejemplar identifica las salas de conferencias como disponibles en base a la proximidad geográfica de la sala de conferencias al dispositivo móvil o en base a una lista de invitados de la sala de conferencias que incluye o bien el dispositivo móvil o bien la información asociada con un usuario del dispositivo móvil. En algunos ejemplos, un dispositivo móvil, que ya está asociado con una sala de conferencias, tiene un código QR y una capacidad de unión mediante NFC en la interfaz de usuario generada en la pantalla del dispositivo móvil mediante la aplicación ejemplar. Tal como se describe más adelante con relación a la Fig. 5, otros posibles participantes pueden escanear el código QR mediante sus dispositivos móviles respectivos o simplemente pueden pulsar en el dispositivo móvil, que ya está asociado con una sala de conferencias, para activar la capacidad de unión mediante NFC para incorporarse a la reunión. En algunos ejemplos, la aplicación ejemplar puede crear una nueva sala de conferencias asociada con el dispositivo móvil. Se entiende también que no es necesario que la "sala de conferencias" sea una sala física; puede ser cualquier espacio en el que uno o más dispositivos móviles que usan la aplicación ejemplar en proximidad entre sí puedan coordinarse entre sí para una teleconferencia. Por consiguiente, la aplicación ejemplar proporciona una solución de conferencia centralizada sin interrupciones para satisfacer las necesidades de conferencia convencionales de los participantes.

En este ejemplo, múltiples participantes 120a-120d están presentes en la primera sala 110, múltiples participantes 122a-122c están presentes en la segunda sala 112 y múltiples participantes 124a-124b están presentes en la tercera sala 114. La sala 112 está físicamente separada de las salas 110, 114 y, de esta manera, puede estar en el mismo edificio o en una

ubicación geográficamente remota que tenga comunicación de red (por ejemplo, conexiones celulares o Wi-Fi a Internet) con los dispositivos móviles en las salas 110, 114. Aunque en el sistema 100 se representan en total tres salas 110, 112 y 114, la presente divulgación contempla que un sistema según la presente divulgación pueda tener cualquier número de salas. Aunque las salas 110, 112 y 114 se representan con dos a cuatro participantes, la presente divulgación contempla que cada sala 110, 112 y 114 pueda tener cualquier cantidad de participantes. Además, para los propósitos de la presente divulgación, los dispositivos móviles que se encuentran en el interior de la misma sala se denominan "locales entre sí", mientras que los dispositivos móviles que se encuentran en salas diferentes entre sí se denominan "remotos entre sí".

Cada sala 110, 112 y 114 tiene un dispositivo líder. El dispositivo líder en cada sala, una vez seleccionado, transmite toda la comunicación desde esa sala a las otras salas en el sistema 100 y recibe toda la comunicación desde las otras salas en el sistema 100.

En un ejemplo, un dispositivo móvil realiza una llamada desde una primera sala (por ejemplo, la sala 110) a al menos una segunda sala (por ejemplo, las salas 112 y/o la sala 114). En este ejemplo, el dispositivo móvil que ha realizado la llamada se convierte en el dispositivo líder para esa sala.

En otro ejemplo, los dispositivos móviles en cada sala intercambian información entre sí para determinar qué dispositivo móvil será un dispositivo líder para dicha sala. Por ejemplo, los dispositivos 130a-130c móviles intercambian información con el resto de dispositivos móviles en la sala 110 para determinar un líder preferido. En algunos ejemplos, el dispositivo móvil de entre los dispositivos 130a-130c móviles que tiene la mayor potencia de procesamiento se selecciona como el dispositivo líder. En algunos ejemplos, un dispositivo móvil se selecciona como el dispositivo líder en base a cualquiera de entre: un nivel de carga de batería, una intensidad de conectividad inalámbrica, una potencia de procesamiento, RAM móvil u otras características del dispositivo móvil.

En algunos ejemplos, el estado "líder" de un dispositivo móvil puede transferirse a un dispositivo móvil diferente durante una teleconferencia, el cual se convierte en el nuevo dispositivo líder. La transferencia puede ser en respuesta a una de las características o métricas indicadas anteriormente, tal como un cambio en el nivel de carga de batería relativo entre los dos dispositivos móviles.

Por ejemplo, en la sala 110, el dispositivo 130a móvil se selecciona como el dispositivo líder. Por lo tanto, el dispositivo 130a móvil transmite la comunicación desde cualquiera de los dispositivos 130a-130c móviles y el dispositivo 136 informático a las salas 112 y 114. Los otros dispositivos, tales como los dispositivos 130b, 130c móviles y el dispositivo 136 informático, se convierten en dispositivos cliente asociados con el dispositivo 130a líder. Por ejemplo, en la sala 112, el dispositivo 132a móvil se selecciona como el dispositivo líder. Por lo tanto, el dispositivo 132a móvil transmite la comunicación desde cualquiera de los dispositivos 132a-132c móviles a las salas 110 y 114. Los otros dispositivos 132b, 132c móviles se convierten en dispositivos cliente asociados con el dispositivo 132a líder. Por ejemplo, en la sala 114, el dispositivo 134a móvil se selecciona como el dispositivo líder. Por lo tanto, el dispositivo 134a móvil transmite las comunicaciones desde el dispositivo 134b móvil a las salas 110 y 112. El otro dispositivo 134b móvil se convierte en un dispositivo cliente asociado con el dispositivo 134a líder.

En algunos ejemplos, un participante usa su dispositivo móvil (por ejemplo, el participante asociado con el dispositivo líder) para llamar desde una de las salas 110, 112, 114 a (1) otro participante en una sala diferente, o (2) una sala diferente en el sistema. En algunos ejemplos, la aplicación ejemplar está configurada para conectar de manera comunicativa salas preseleccionadas en base a la información desde un participante antes de que comience la llamada (por ejemplo, todas las salas 110, 112, 114 en el sistema 100 están configuradas para conectarse). En todos los ejemplos, se hace referencia a la comunicación entre salas en el sistema como "llamada".

La llamada puede iniciarse usando un protocolo VoIP, SIP o un protocolo estándar o propietario similar para la comunicación a través de una red de datos. La comunicación se establece entre la sala 110 y la sala 112 a través de los dispositivos 130a y 132a líderes en este ejemplo mediante una comunicación de protocolo de red externa a través de una red 140, tal como Internet. Tal como se explicará, todos los participantes en la sala 110 pueden comunicarse con todos los participantes en la sala 112 y viceversa. Los dispositivos 130a y 132a líderes comparten información de tokens NTS, información de identificación de sala correspondiente para cada dispositivo líder e información de identificación de reunión usando un servicio de "backend", tal como una base 142 de datos en la nube de Google Firebase a través de la red 140. De manera similar a la comunicación establecida entre las salas 110 y 112, la comunicación se establece entre todas las salas restantes en el sistema 100 (por ejemplo, la comunicación se establece entre las salas 110 y 114, y entre las salas 112 y 114).

La Fig. 1B muestra una ruta ejemplar de una comunicación de audio en la que una señal de audio desde la sala 110 se transmite a las salas 112 y 114. La Fig. 1B demuestra que, según las realizaciones de la presente divulgación, un dispositivo líder en una sala recibe señales de audio desde los otros dispositivos líderes en las otras salas. Por ejemplo, cada señal de audio se transmite en una serie de paquetes, tal como se describe más adelante. Cada dispositivo líder transmite la señal de audio recibida a los dispositivos cliente asociados con dicho dispositivo líder. El dispositivo líder dirige todos los dispositivos móviles en una única sala para reproducir la señal de audio recibida en sincronización en todos los dispositivos móviles de la sala.

La comunicación entre los dispositivos líderes, tales como los dispositivos 130a y 132b móviles en la Fig. 1B, se produce

a través de Internet 140. En algunos ejemplos, la comunicación entre el dispositivo líder, tal como el dispositivo 130a móvil, y los dispositivos cliente, los otros dispositivos 130b-130d móviles en una sala individual, tal como la sala 110, es a través de una red ad hoc establecida en una cuadrícula estructurada. Por ejemplo, el dispositivo líder puede comunicarse con los dispositivos cliente en la misma sala usando un canal de comunicación inalámbrico de corto alcance, tal como Bluetooth u otro enlace de comunicación de radiofrecuencia directo de dispositivo a dispositivo. Como otro ejemplo, el dispositivo líder puede comunicarse con los dispositivos cliente usando Wi-Fi. Los dispositivos móviles de una sala pueden conectarse en la misma conexión de internet, es decir, un router Wi-Fi o un punto de acceso móvil.

Cuando uno de los participantes en la sala 110 habla (por ejemplo, el participante 120b, tal como se muestra en la Fig. 1B), los micrófonos en cada uno de los dispositivos reciben y graban una señal de audio. Por ejemplo, el dispositivo 130b móvil graba la señal 1 de audio, el dispositivo 130c móvil graba la señal 2 de audio, el dispositivo 130a móvil graba la señal 3 de audio y el teléfono tradicional graba la señal 4 de audio. Por lo tanto, todas las señales 1, 2, 3 y 4 de audio capturan el audio generado por el participante 120b. Sin embargo, las señales 1, 2, 3 y 4 de audio grabadas pueden diferir entre sí, por ejemplo, en base a (1) una sensibilidad del micrófono en cada dispositivo móvil, (2) un retardo de tiempo de un dispositivo móvil que recibe la señal de audio según una posición del dispositivo móvil en la sala, o (3) si la señal de audio encontró una interferencia antes de ser grabada por un dispositivo móvil. Por ejemplo, el dispositivo 130a móvil está más lejos del participante 120b que el dispositivo 130b móvil y, por consiguiente, la señal 3 de audio grabada tendrá un retardo de tiempo con respecto a la señal 1 de audio grabada. Por ejemplo, el dispositivo 130b móvil puede tener un micrófono mejor que el dispositivo 136 informático y, por consiguiente, la señal 1 de audio tiene mejor claridad que la señal 4 de audio. Por ejemplo, puede haber muebles u otras barreras físicas entre el participante 120b y el dispositivo 130a móvil y, por consiguiente, la señal 3 de audio grabada puede tener una claridad baja. Por ejemplo, la señal de audio desde el participante 120b puede recibir eco debido a los bordes de la sala 110, y el dispositivo 130a móvil puede grabar tanto la señal de audio directa desde el participante 120b como la señal de audio de eco, lo que resulta en que la señal 3 de audio tenga una claridad baja.

Cada dispositivo 130b, 130c y 136 transmite su audio grabado al dispositivo 130a líder. De esta manera, la aplicación ejemplar ejecutada en el dispositivo 130a líder selecciona una de las señales 1, 2, 3, 4 de audio grabadas para transmitir la misma a otras salas en el sistema 100. La aplicación ejemplar selecciona la señal de audio desde cada uno de los dispositivos 130a-130c y 136 en base a qué señal de audio tiene las mejores métricas. Por ejemplo, la aplicación ejemplar evalúa qué señal 1, 2, 3, 4 de audio es la más clara y/o la más fuerte. En algunos ejemplos, la señal de audio grabada seleccionada por la aplicación ejemplar se recibe desde un dispositivo que está físicamente más cerca de la persona que habla, por ejemplo, el participante 120b. Por ejemplo, el dispositivo 130a líder limpia la señal de audio y envía una señal 150 de audio limpiada a los teléfonos 132a y 134a líderes a través de la red 140 donde la señal 150 de audio limpiada se transmite a su vez a todos los dispositivos cliente en las salas 112 y 114 respectivas.

El sistema ejemplar incluye dos características básicas de (1) un modo de conferencia, tal como se ha descrito anteriormente con relación a la Fig. 1B, y (2) un modo de reunión. En el modo reunión, el dispositivo líder de cada sala recibe una señal y coordina la reproducción en cada uno de los otros dispositivos en la sala. Por ejemplo, en el modo reunión, un vídeo con datos tanto de audio como visuales se reproduce de manera sincrónica en cada dispositivo en cada sala.

El modo de conferencia tiene dos tipos de comunicación. Uno de los tipos de comunicación es desde un dispositivo móvil designado como un dispositivo líder en una primera sala a otro dispositivo móvil designado como un dispositivo líder en una segunda sala. El otro tipo de comunicación es una comunicación ad hoc entre el dispositivo móvil designado como el dispositivo líder y los otros dispositivos móviles en la sala que están designados como teléfonos cliente. En este ejemplo, un sistema de comunicación web en tiempo real (WebRTC) proporciona la plataforma básica para la comunicación. El sistema ejemplar añade ciertos módulos adicionales al WebRTC para cumplir los requisitos para las características operativas de teleconferencia descritas en el presente documento.

En este ejemplo, la red 140 permite la comunicación entre cualquiera de los dispositivos de usuario en la Fig. 1A. Los participantes pueden acceder a los dispositivos de usuario móviles y pueden ejecutar la aplicación de teleconferencia ejemplar para permitir la configuración de la teleconferencia sin hardware de teleconferencia previo. Los dispositivos de usuario móviles pueden incluir prácticamente cualquier dispositivo informático preferiblemente móvil equipado con un micrófono y un altavoz que esté configurado para enviar y recibir información a través de una red con capacidad inalámbrica, tal como la red 140.

La red 140 está configurada para permitir comunicaciones entre un dispositivo informático con otro dispositivo informático. La red 140 puede estar habilitada para emplear cualquier forma de medio legible por ordenador para comunicar información desde un dispositivo electrónico a otro. En un conjunto de LANs interconectadas, incluidas las basadas en diferentes arquitecturas y protocolos, un enrutador y/o dispositivo de puerta de enlace actúa como un enlace entre las LANs, que permite el envío de mensajes entre dispositivos informáticos. Además, los enlaces de comunicación en el interior de las LANs incluyen típicamente un par de cables trenzados o un cable coaxial, mientras que los enlaces de comunicación entre las redes pueden utilizar líneas telefónicas analógicas; líneas digitales dedicadas completas o fraccionarias incluyendo T1, T2, T3 y T4; redes digitales de servicios integrados (ISDNs); líneas de abonado digitales (DSL); enlaces inalámbricos, incluyendo enlaces satelitales; u otros enlaces de comunicación conocidos por las personas expertas en la técnica. Además, los ordenadores remotos y otros dispositivos electrónicos relacionados pueden conectarse de manera remota a LANs o WANs a través de un módem y un enlace telefónico temporal.

La red 140 puede incluir además cualquiera de una diversidad de subredes inalámbricas que pueden superponerse además a redes ad-hoc independientes y similares, para proporcionar una conexión orientada a la infraestructura. Dichas subredes pueden incluir redes de malla, redes LAN inalámbricas (WLAN), redes celulares y similares. La red 140 puede incluir también un sistema autónomo de terminales, pasarelas, enrutadores y similares conectados por enlaces de radio inalámbricos o transceptores inalámbricos. Estos conectores pueden configurarse para moverse libremente y de forma aleatoria y organizarse de forma arbitraria, de manera que la topología de la red 108 pueda cambiar de forma rápida y arbitraria.

La red 140 puede emplear además múltiples tecnologías de acceso, incluyendo acceso de radio de cuarta (4G) o quinta (5G) generación para sistemas celulares; WLAN; enrutadores inalámbricos (WR) en malla; y similares. Las tecnologías de acceso, tales como 4G, 5G, y las redes de acceso futuras pueden permitir una cobertura de área amplia para dispositivos móviles, tales como los dispositivos de usuario con varios grados de movilidad. Por ejemplo, la red 108 puede permitir una conexión de radio a través de un acceso a red de radio, tal como el sistema global para comunicación móvil (GSM), servicios generales de radio por paquetes (GPRS), Entorno GSM de datos mejorados (EDGE), acceso múltiple por división de código de banda ancha (WCDMA), CDMA2000 y similares. La red 108 puede construirse también para su uso con diversos otros protocolos de comunicación por cable e inalámbricos, incluyendo TCP/IP, UDP, SIP, SMS, WAP, CDMA, TDMA, EDGE, UMTS, GPRS, GSM, UWB, WiMax, IEEE 802.11x y similares. En esencia, la red 108 puede incluir prácticamente cualquier mecanismo de comunicación por cable y/o inalámbrico mediante el cual la información puede viajar entre un dispositivo informático y otro dispositivo informático, red y similares.

La Fig. 2 es un diagrama de bloques de los componentes de un dispositivo de usuario móvil, tal como el dispositivo 130a de usuario móvil en las Figs. 1A y 1B. El dispositivo 130a de usuario móvil incluye un procesador 210 de aplicaciones, una fuente 212 de alimentación, una pantalla 214, un procesador 216 de banda base (o procesadores de comunicación inalámbrica similares, tales como un procesador Wi-Fi o Bluetooth) y un CÓDEC 218. En este ejemplo, la pantalla 214 es una pantalla táctil LCD que puede permitir al usuario controlar las aplicaciones ejecutadas por el procesador 210 de aplicaciones mediante entradas táctiles, así como ver los gráficos generados por el procesador 210 de aplicaciones. La pantalla 214 puede controlarse mediante un controlador 220 de pantalla táctil. El procesador 210 de aplicaciones puede estar acoplado a varios dispositivos, tales como una cámara 222, y otras interfaces, tales como un puerto de comunicación, etc.

El procesador 216 de banda base recibe señales desde un transmisor receptor 230 de red que permite las comunicaciones con la red 140 en la Fig. 1A y un receptor 232 de georreferenciación que permite la recepción de datos de posicionamiento para determinar la ubicación del dispositivo 130a móvil. El procesador 216 de banda base procesa las señales y está acoplado al CÓDEC 218, que convierte las señales para su uso por el procesador 210 de aplicaciones. El CÓDEC 218 decodifica también las señales de audio recibidas desde un micrófono 240 y codifica las señales de datos para su emisión por un altavoz 242 para funciones tales como una aplicación telefónica ejecutada por el procesador 210 de aplicaciones.

Los procesadores 210 y 216 pueden implementarse, de manera conveniente, usando uno o más sistemas informáticos de propósito general, microprocesadores, procesadores de señales digitales, microcontroladores, circuitos integrados específicos de la aplicación (ASIC), dispositivos lógicos programables (PLD), dispositivos lógicos programables por campo (FPLD), matrices de puertas programables por campo (FPGA) y similares, programados según las enseñanzas descritas e ilustradas en el presente documento, tal como apreciarán las personas expertas en las técnicas informáticas, de software y de redes.

El software del sistema operativo y otras aplicaciones se almacenan en la memoria 250 de solo lectura (ROM), memoria 252 de acceso aleatorio (RAM) y un dispositivo 254 de almacenamiento de memoria para su acceso por parte del procesador 210 de aplicaciones. En este ejemplo, el dispositivo 254 de almacenamiento de memoria es una memoria flash, pero pueden usarse otros dispositivos de memoria. Las aplicaciones almacenadas en el dispositivo 254 de almacenamiento de memoria incluyen la aplicación de teleconferencia, que crea gráficos de interfaz en la pantalla e interactúa con una aplicación de navegación. Otras formas de aplicación pueden incorporar los principios que se explican a continuación. En este ejemplo, la aplicación de teleconferencia puede precargarse en el dispositivo 130a de usuario móvil o puede ofrecerse como una aplicación que puede descargarse al dispositivo 130a de usuario móvil desde un dispositivo de red, tal como un servidor, a través de la red 140.

El dispositivo 254 de almacenamiento de memoria incluye un medio legible por máquina en el que se almacenan uno o más conjuntos de instrucciones (por ejemplo, software) que materializan una cualquiera o más de las metodologías o funciones descritas en el presente documento. Las instrucciones pueden residir también, completa o al menos parcialmente, en el interior del dispositivo 254 de almacenamiento de memoria, la ROM 250, la RAM 252 y/o en el interior de los procesadores 210 o 216 durante su ejecución por el dispositivo 130a de usuario móvil. Las instrucciones pueden transmitirse o recibirse además a través de una red, tal como la red 140 en la Fig. 1A, a través del transmisor receptor 230 de red. Aunque el medio legible por máquina se muestra en un ejemplo como un único medio, debe considerarse que la expresión "medio legible por máquina" incluye un único medio o múltiples medios (por ejemplo, una base de datos centralizada o distribuida, y/o cachés y servidores asociados) que almacenan los uno o más conjuntos de instrucciones. Puede considerarse también que la expresión "medio legible por máquina" incluye cualquier medio que sea capaz de almacenar, codificar o contener un conjunto de instrucciones para su ejecución por la máquina y que causen que la máquina realice una cualquiera o más de las metodologías de las diversas realizaciones, o que sea capaz de almacenar, codificar o contener estructuras de datos utilizadas por o asociadas con dicho conjunto de instrucciones. Por consiguiente,

puede considerarse que la expresión "medio legible por máquina" incluye, pero no está limitada a, memorias de estado sólido, medios ópticos y medios magnéticos.

Una diversidad de diferentes tipos de dispositivos de almacenamiento de memoria, tales como una memoria de acceso aleatorio (RAM) o una memoria de solo lectura (ROM) en el sistema o un disquete, disco duro, CD ROM, DVD ROM, flash u otro medio legible por ordenador que es leído y/o escrito por un sistema magnético, óptico u otro sistema de lectura y/o escritura que está acoplado al procesador, pueden usarse para la memoria o las memorias en el dispositivo 130a de usuario.

La Fig. 3A es un diagrama de flujo de la rutina para la comunicación entre un dispositivo cliente y el dispositivo de usuario móvil designado como un dispositivo líder cuando se recibe una señal desde un dispositivo de comunicación externo (por ejemplo, un dispositivo líder en otra sala). Tal como se muestra en la Fig. 3A, puede producirse una comunicación entre dispositivos líderes, tales como los dispositivos 130a y 132a de usuario móviles ejemplares que están designados como dispositivos líderes en la Fig. 1A (310). Cuando un dispositivo líder intenta iniciar una llamada con el otro dispositivo líder, el dispositivo líder iniciador puede obtener un token de servicio transversal de red (NTS) desde la base 142 de datos u otro servicio "backend". Los ejemplos del servicio "backend" pueden incluir la función AWS Lambda. Los tokens pueden ser credenciales efímeras que permiten el acceso a los recursos NAT ("traducción de direcciones de red") transversal usando reenvío (TURN) desde los clientes WebRTC y VoIP. Los tokens proporcionan una manera segura de crear sesiones de comunicación con navegadores web o aplicaciones de dispositivos móviles sin exponer las credenciales de autenticación en un entorno del lado del cliente. Todos los tokens pueden tener una vida útil limitada para proporcionar protección contra abusos. En este ejemplo, la vida útil del token puede configurarse para cualquier tiempo hasta 24 horas. Dicho token solo se requiere para conferencias multiparte entre dispositivos líderes, que son remotos entre sí, para gestionar temas relacionados con la traducción de direcciones de red (NAT).

Para la comunicación ad hoc, no se requiere el token NTS. Por ejemplo, la presente divulgación contempla dispositivos cliente que usan utilidades transversales de sesión para NAT (STUN), ya que no hay ningún requisito de NAT transversal en la comunicación ad hoc.

La rutina determina si un token se obtiene o no con éxito, (314). Si un token NTS no se obtiene con éxito, la rutina vuelve a ejecutar el bucle e intenta obtener un token NTS (312). Si un token NTS se obtiene con éxito (314), se establece una conexión entre dos dispositivos líderes y el dispositivo líder con el que se contacta empieza a sonar (316). La configuración de un punto final en una conexión WebRTC se denomina una descripción de sesión. Cuando un usuario inicia una llamada WebRTC a otro usuario, se crea una descripción especial denominada oferta. Esta descripción especial incluye toda la información acerca de la configuración propuesta por la parte llamante para la llamada. A continuación, el destinatario responde con una respuesta, que es una descripción de su final de la llamada (318). En el mecanismo respuesta de oferta, la descripción de la sesión incluye información acerca del tipo de medios enviados, su formato, el protocolo de transferencia usado, la dirección IP y el puerto del punto final, y otra información necesaria para describir un punto final de transferencia de medios. Esta información se intercambia y se almacena usando el protocolo de descripción de sesión (SDP) en la base de datos en tiempo real en la nube (320). ICE es un protocolo que permite que dos dispositivos establezcan una conexión interactiva usando ofertas y respuestas, incluso si los dos dispositivos están separados por un dispositivo de traducción de direcciones de red (NAT). La parte llamante introduce sus candidatos ICE en la base 142 de datos en tiempo real firebase mientras que la parte llamada extrae los candidatos ICE desde la base 142 de datos en tiempo real firebase (322). De esta manera, ambos dispositivos comparten la información necesaria para intercambiar datos multimedia entre sí. Este intercambio se gestiona usando un mecanismo de señalización que se realiza usando firebase en tiempo real (324). De esta manera, se establece un canal de comunicación entre la parte llamante y la parte llamada.

A continuación, el módulo de dispositivo de audio del dispositivo líder se inicializa (326). El módulo del dispositivo de audio incluye tanto una grabadora de audio como un reproductor de audio. Por consiguiente, el dispositivo líder está configurado para grabar y reproducir de manera continua paquetes de audio. Después de la inicialización del módulo del dispositivo de audio, el dispositivo líder graba señales de audio desde su micrófono 240 en la Fig. 2 y envía las señales de audio grabadas para un preprocesamiento de audio adicional (328) (descrito más detalladamente a continuación con referencia a la Fig. 3B). Estos paquetes se transmiten a otros dispositivos móviles remotos designados como dispositivos líderes (330).

Por consiguiente, la Fig. 3A demuestra una comunicación mejorada entre dispositivos líderes basada en la canalización de canales de audio WebRTC divulgada en la Fig. 3A y los procesos divulgados para el canal de audio de WebRTC. Concretamente, la presente divulgación proporciona ventajas sobre los procesos convencionales ya que los dispositivos ad hoc (por ejemplo, dispositivos cliente) envían y reciben datos de audio a través del canal de datos WebRTC después del procesamiento requerido.

Con referencia ahora a la Fig. 3B, se muestra un sistema 300B ejemplar que realiza el preprocesamiento de la etapa 328. Por ejemplo, el sistema 300B es un módulo WebRTC. El sistema 300B incluye un módulo 352 de remezclado y remuestreo, un módulo 354 de procesamiento de audio (APM), un módulo 356 de codificación y un módulo 358 de encriptación. El APM 354 incluye un módulo 362 de cancelación de eco acústico (AEC), un módulo 364 de control automático de ganancia (AGC) y un módulo 366 de reducción de ruido adaptativo (ANR).

Durante el preprocesamiento, los fragmentos grabados de la señal de audio recibida se codifican usando el módulo 356 de codificación, que en este ejemplo es un codificador Opus. Después de codificar los datos, se generan paquetes

transmisibles a partir de los fragmentos grabados de la señal de audio recibida; a continuación, los paquetes transmisibles se encriptan en el módulo 358 de encriptación.

Además, la presente divulgación contempla que el sistema 300B esté configurado para realizar todo el preprocesamiento de las señales de audio en el dispositivo líder local, antes de transmitir estos paquetes de audio a los dispositivos líderes remotos. El módulo 352 de remezclado y remuestreo procesa las señales de audio grabadas del dispositivo líder local. Por ejemplo, la señal de audio grabada es multicanal, y el módulo 352 de remezclado y remuestreo diezma y mezcla la señal de audio en un único canal para el procesamiento interior por el APM 354.

Con referencia de nuevo a la Fig. 3A, la etapa 330 de transmisión de audio incluye además una etapa de formación de paquetes para el audio grabado. Una capa de transporte y una dirección IP se adjuntan en la etapa de generación de paquetes. La presente divulgación contempla que los paquetes puedan contener cualquier cantidad de datos de audio, preferiblemente que contengan entre 5 ms y 50 ms de datos de audio, y más preferiblemente que contengan aproximadamente 10 ms de datos de audio para dispositivos Android y 40 ms de datos de audio para dispositivos iOS. Estos paquetes contienen metadatos de audio (por ejemplo, una cabecera), así como audio codificado. Los metadatos de audio pueden incluir información acerca del paquete, tal como un número de secuencia del paquete y una marca de tiempo.

El proceso global de grabación y transmisión de señales de audio y de formación de paquetes de audio en un dispositivo líder se muestra en la Fig. 4A como el proceso 400A. El proceso 400A incluye primero grabar una señal de audio en un módulo grabador Oboe del dispositivo líder (410). A continuación, el dispositivo líder calcula una energía cuadrática media (412) y una media móvil exponencial (EMA) (414) de la señal grabada, y pasa la señal grabada a través de un filtro de media móvil (416). El paquete grabado se transmite a una memoria intermedia del dispositivo de audio del dispositivo líder (418). Los paquetes grabados se remezclan y remuestrean (420). A continuación, los paquetes grabados son transmitidos por la memoria intermedia del dispositivo de audio al APM 354 en el orden predeterminado para el preprocesamiento (422).

Con referencia ahora a la Fig. 48, el APM 354 incluye el módulo 364 de control automático de ganancia (AGC), el módulo 362 de cancelación de eco acústico (AEC), un módulo 480 de filtro pasa alta, un módulo 482 de cancelación de ruido adaptativo (ANC), un módulo 484 de detector de actividad de voz (VAD) y un circuito 486 de inteligibilidad del habla. Por ejemplo, el APM 354, tal como se ha descrito en la Fig. 4B, proporciona detalles adicionales de los módulos incluidos en un APM, según las diversas realizaciones de la presente divulgación. El módulo 364 AGC amplifica la voz y el sonido. El módulo 362 AEC elimina el eco, la reverberación y los sonidos añadidos no deseados de la señal de audio. El módulo 480 de filtro pasa alta deja pasar las señales con una frecuencia superior a una frecuencia umbral y atenúa las señales con frecuencias inferiores a la frecuencia umbral. El módulo 482 ANC elimina el ruido no deseado de la señal. El módulo 484 VAD y el circuito 486 de inteligibilidad del habla usan algoritmos de procesamiento del habla y/o algoritmos de cálculo de energía para determinar si hay o no voces humanas en el audio grabado. Por ejemplo, el módulo 484 VAD es parte del módulo 482 ANC o está acoplado de manera comunicativa al mismo. Por ejemplo, el circuito 486 de inteligibilidad del habla es parte del módulo 362 AEC o está acoplado de manera comunicativa al mismo. En algunos ejemplos, no mostrados, el APM 354 incluye además un módulo preamplificador para amplificar señales de bajo nivel a un nivel operativo estándar.

Con referencia a la Fig. 4A, a continuación, se codifica (424) la señal de audio procesada. Los paquetes son formados por el módulo de formación de paquetes (426). Por ejemplo, la señal de audio grabada se divide en una serie de paquetes de datos de audio. En algunos ejemplos, cada uno de los paquetes es similar en longitud a los demás (por ejemplo, preferiblemente tienen una diferencia de 5 ms de datos de audio entre sí y, más preferiblemente, tienen una diferencia de 1 ms de datos de audio entre sí). Cada paquete se forma con una cabecera RTP y una carga útil, tal como se describe más adelante en el presente documento. El módulo 358 de encriptación encripta los paquetes (428). Un tranceptor envía los paquetes a la red 140 (por ejemplo, Internet); a continuación, el mensaje se transmite a través de la red 140 y se enruta al otro o a los otros dispositivos líderes (430).

Con referencia ahora a la Fig. 4C, el otro o los otros dispositivos líderes reciben los paquetes desde una interfaz de red (432). Los paquetes se desencriptan (434). Los datos se extraen de los paquetes desencriptados (436) a una cabecera RTP y una carga útil. Los datos de cabecera de protocolo en tiempo real y los datos de carga útil extraídos se envían a un módulo NetEq de un canal de audio del WebRTC para gestionar la fluctuación de la red (438), si corresponde, de manera que la reproducción se sincronice entre las salas remotas. A continuación, los paquetes se almacenan en un almacenamiento intermedio antifluencia (440) y, a continuación, se decodifican (442). A continuación, estos paquetes se pasan a un módulo de posprocesamiento para mantener la señal de referencia para AEC (444). A continuación, se mezclan los paquetes desde otros dispositivos móviles remotos designados como dispositivos líderes (446). Por ejemplo, si un dispositivo líder recibe paquetes desde múltiples dispositivos líderes remotos, los paquetes se mezclan para formar un único paquete para su reproducción en la sala local del dispositivo líder. El único paquete desde un dispositivo líder, o el único paquete mezclado desde los N dispositivos líderes (en el caso de una comunicación N - N) se envía a continuación a toda la memoria intermedia del dispositivo de audio del dispositivo líder (448). A continuación, el dispositivo líder determina si los dispositivos cliente locales están conectados (450). Si los dispositivos cliente locales están conectados, entonces el dispositivo líder transmite el único paquete a los dispositivos cliente locales (452). Si los dispositivos locales no están conectados, el dispositivo líder pasa a la etapa (454).

Cada dispositivo cliente y el dispositivo líder usan un reproductor, tal como un reproductor Oboe, para reproducir los datos de audio (454). Por ejemplo, la reproducción de la etapa 454 incluye decodificar cada paquete recibido y extraer los valores de protocolo de serialización de datos y de carga útil. En algunos ejemplos de la presente divulgación, se usan

decodificadores WebRTC para realizar la etapa 454.

La Fig. 5 muestra el proceso de grabación de señales de audio y de formación de paquetes de audio en un dispositivo cliente como el proceso 500 de grabación. El dispositivo móvil del usuario escanea un código QR mostrado en el dispositivo móvil designado como el dispositivo líder o recibe la información desde el QR mediante un mecanismo NFC desde el dispositivo líder (510). Por ejemplo, el código QR incluye información de identificación de la sala e información de identificación de la reunión.

Una vez recibida la información del código QR o la comunicación NFC, ese dispositivo cliente calcula un valor de protocolo de tiempo de red (NTP), tal como se describe más adelante con relación a la Fig. 8. A continuación, el micrófono y el altavoz del dispositivo cliente se activan (512) y el dispositivo cliente graba una señal de audio (514).

El dispositivo cliente calcula la energía cuadrática media de la señal de audio (516). Se calcula una media móvil exponencial (EMA) sobre la señal de audio para suavizar las variaciones en la energía de la señal de audio (518). A continuación, la señal de audio pasa a un filtro de media móvil (520). La señal de audio se codifica usando el codificador opus (522). El codificador opus decide si la señal es una voz silenciosa o activa en base a la detección de actividad de voz y la probabilidad de presencia de voz (524). Si nadie está hablando, lo que indica una voz silenciosa en el micrófono, entonces se añade ruido rosa o no se añade ningún ruido (por ejemplo, una secuencia de bits vacíos o '0') a la señal de audio; la secuencia de bits vacíos o '0' reemplaza los datos grabados originales (526). Por consiguiente, la secuencia de bits vacíos o '0' ahorra ancho de banda de la red y ayuda a ahorrar potencia de procesamiento, lo cual es más ventajoso que enviar los datos de audio originales. La señal de audio (bien con el ruido rosa añadido o bien con los datos de audio originales) se usa a continuación para formar un paquete (528). Si la señal de audio es una voz activa, los paquetes se forman a partir de la señal de voz activa y los datos de protocolo de serialización de datos contienen cualquier información necesaria para sincronizar el dispositivo cliente o líder con los otros dispositivos locales. A continuación, los paquetes creados a partir de la señal de audio se envían al dispositivo líder (530).

Además, los dispositivos cliente reciben datos de señal de audio desde el dispositivo líder local. La Fig. 6 muestra el proceso 600 de reproducción de cómo los dispositivos de usuario cliente procesan y emiten la señal de audio recibida desde el dispositivo líder local. Los paquetes de voz se reciben en el dispositivo líder y se envían al dispositivo cliente (610). Estos paquetes provienen del canal de datos WebRTC y, a continuación, se desenscriptan (612). El dispositivo cliente realiza extracciones de datos de cabecera y de carga útil desde los paquetes recibidos según una extracción de modelo de protocolo de serialización de datos (614). A continuación, los paquetes se envían al almacenamiento intermedio antifluctuación Speex para gestionar la fluctuación de la red (616). A continuación, la señal se decodifica con un decodificador Opus en este ejemplo (618). Se ejecuta el algoritmo 368 de sincronización para garantizar la sincronización con los otros dispositivos de usuario móviles en la sala (620). A continuación, el dispositivo cliente reproduce la señal de voz (622).

La Fig. 7 muestra el proceso 700 de recepción y transmisión para que un dispositivo líder procese los datos de audio y los envíe a un dispositivo líder remoto. Un primer dispositivo líder llama a un segundo (y remoto) dispositivo líder (710), este intento inicia la llamada en el segundo dispositivo líder. Después de que el usuario del segundo dispositivo líder acepte la llamada del primer dispositivo líder, el micrófono y los altavoces del primer dispositivo líder se activan para recibir y reproducir señales de audio (712). La señal de audio recibida desde el micrófono del primer dispositivo líder se graba y se envía al almacenamiento intermedio del primer dispositivo líder (714). Una vez enviada una señal de audio a la memoria intermedia, se calcula la energía de la señal de audio (716). Se calcula una media móvil exponencial (EMA) (718) sobre la señal de audio para suavizar la variación de energía de la señal de audio. La energía calculada se envía al filtro de media móvil (720). El filtro de media móvil calcula una media del paquete actual con respecto a los datos recopilados antes del paquete actual, durante un período de tiempo umbral (por ejemplo, 1 segundo de datos). A continuación, el primer dispositivo líder determina si los dispositivos cliente están conectados o no al dispositivo líder (722).

Si los dispositivos cliente están conectados con el primer dispositivo líder, entonces un decodificador opus del dispositivo líder decodifica las señales de audio desde los dispositivos cliente conectados (724). A continuación, se ejecuta un algoritmo de selección sobre las señales de audio decodificadas (726). El algoritmo de selección selecciona un paquete para la transmisión de entre los paquetes correspondientes desde todos los dispositivos locales en base a NTP, SSN, los valores de tamaño del almacenamiento intermedio, el algoritmo Bargeln y otras consideraciones, tal como se describe en el presente documento. En algunos ejemplos, la etapa 726 selecciona un paquete de entre los paquetes correspondientes que tiene el valor de energía más alto o el volumen más alto. Todos los paquetes seleccionados se pasan a través de un algoritmo de Bargeln, que se usa para: (1) minimizar los falsos intercambios entre el cliente y el dispositivo líder; y (2) minimizar los cortes de las señales de voz. La señal de audio de altavoz decodificada seleccionada se pasa al canal de audio del primer dispositivo líder para un preprocesamiento adicional y una formación de paquetes (728) antes de transmitir la misma al segundo (y remoto) dispositivo líder a través de Internet (730).

Si los dispositivos cliente no están conectados, entonces no se realiza un algoritmo de selección y la señal de audio del primer dispositivo líder se pasa al canal de audio para su preprocesamiento adicional y formación de paquetes (728) antes de transmitir la misma al dispositivo líder remoto a través de Internet (730).

Cada dispositivo móvil debe sincronizarse con cada dispositivo móvil local para que los dispositivos locales graben audio al mismo tiempo. Típicamente, se realiza un proceso de sincronización en el dispositivo líder para todos los dispositivos

locales. La Fig. 8 muestra un proceso 800 de sincronización para establecer una comunicación entre el dispositivo líder y los dispositivos cliente para grabar la misma señal de audio al mismo tiempo (por ejemplo, los SSNs de cada dispositivo local se correlacionarán entre sí). El NTP es un protocolo de red para la sincronización de relojes entre sistemas informáticos a través de redes de datos de latencia variable con conmutación de paquetes. El NTP está diseñado para sincronizar todos los ordenadores participantes con una precisión de unos pocos milisegundos con respecto al tiempo del sistema del dispositivo líder. El NTP puede conseguir una precisión superior a un milisegundo en redes de área local en condiciones ideales. Las rutas asimétricas y la congestión de la red pueden causar errores de 100 ms o más. Después de calcular un valor NTP, cada uno de los dispositivos cliente sincroniza su tiempo según el dispositivo líder, de manera que los dispositivos cliente graban y reproducen una señal de audio de manera sincronizada entre sí y con el dispositivo líder. Por consiguiente, el cálculo del valor NTP permite la sincronización de los dispositivos una vez que se contabiliza la latencia del hardware.

Un dispositivo cliente escanea el código QR mostrado por el dispositivo líder o se une a la comunicación con el dispositivo líder a través de NFC (810). El dispositivo cliente calcula un valor de protocolo de tiempo de red (NTP) (812). El algoritmo de sincronización determina si se ha calculado o no un valor NTP con un valor umbral (814). Por ejemplo, el valor umbral es 5 ms. Si se calcula un valor NTP que es mayor que el valor umbral, entonces el dispositivo cliente vuelve a la etapa 812. Si se calcula un valor NTP que es menor o igual que el valor umbral, entonces el dispositivo cliente pasa a la etapa 816.

A continuación, el dispositivo cliente realiza una sincronización grabada de manera que el micrófono del dispositivo cliente grabará al mismo tiempo que todos los demás dispositivos locales, en base al NTP (816). Por ejemplo, el NTP permite la sincronización de los relojes en tiempo real de todos los dispositivos locales en una red ad hoc. El dispositivo cliente determina los números (SSNs) de secuencia sincronizados en base al NTP (818). Los SSNs se asocian con cada paquete de audio para que el orden de los paquetes se mantenga y sea identificable; los SSNs son usados también por el algoritmo de selección tal como se ha descrito anteriormente para garantizar que el audio se seleccione en base a los paquetes que se graban de manera coherente para evitar duplicar la voz cuando los dispositivos líderes intercambian la selección de la señal de audio entre diferentes dispositivos locales. A continuación, la relación del SSN se formula con la cantidad de datos almacenados en el almacenamiento intermedio grabado (820). En base al SSN, se predice el tiempo NTP en el que se inicia la grabación en todos los dispositivos cliente (822). Por consiguiente, entonces todos los dispositivos locales están sincronizados en la grabación con respecto al dispositivo líder (824).

El NTP utiliza un sistema de semicapas, jerárquico, de fuentes de tiempo. Cada nivel de esta jerarquía se denomina estrato y se le asigna un número que comienza con cero para el reloj de referencia en la parte superior. Un servidor sincronizado con un servidor de estrato n se ejecuta en el estrato n + 1. El número representa la distancia desde el reloj de referencia y se usa para prevenir dependencias cíclicas en la jerarquía. El estrato no siempre es una indicación de calidad o fiabilidad; es común encontrar fuentes de tiempo del estrato 3 que son de mayor calidad que otras fuentes de tiempo del estrato 2.

Por ejemplo, un dispositivo cliente sondea de manera regular su dispositivo líder asociado para determinar un cálculo de NTP. El cliente debe calcular su desfase temporal y el retraso de ida y vuelta. El desfase θ temporal, la diferencia en tiempo absoluto entre los dos relojes, está definido por:

$$\theta = \left| \frac{(t_1 - t_0) + (t_2 - t_3)}{2} \right|$$

y el retraso δ de ida y vuelta está determinado por:

$$\delta = (t_3 - t_0) + (t_2 - t_1)$$

En estas ecuaciones, t_0 es la marca de tiempo del cliente para transmitir un paquete de solicitud; t_1 es la marca de tiempo del servidor para recibir el paquete de solicitud transmitido al servidor por el cliente; t_2 es la marca de tiempo del servidor para transmitir el paquete de respuesta; y t_3 es la marca de tiempo del cliente para recibir el paquete de respuesta transmitido al cliente por el servidor.

Para derivar la expresión para el desfase, tenga en cuenta que, para el paquete de solicitud,

$$t_0 + \theta + \delta/2 = t_1$$

y para el paquete de respuesta,

$$t_0 + \theta - \delta/2 = t_2$$

Resolviendo para θ se obtiene la definición del desfase temporal. Los valores de θ y δ se pasan a través de los filtros y se someten a un análisis estadístico. Los valores atípicos se descartan hasta que solo quede un número umbral de candidatos (por ejemplo, tres candidatos restantes). Se deriva una estimación del desfase temporal a partir de los candidatos restantes. A continuación, se ajusta el reloj para reducir el desfase gradualmente, creando un bucle de retroalimentación. La sincronización precisa se consigue cuando las rutas tanto entrantes como salientes entre el cliente y el servidor tienen un retardo nominal simétrico. Si las rutas no tienen un retraso nominal común, existe un sesgo sistemático de la mitad de la

diferencia entre los tiempos de viaje hacia adelante y hacia atrás.

Las Figs. 9 y 10 se describirán en conjunto. La Fig. 9 demuestra la formación y transmisión de paquetes a partir de señales de audio grabadas para los dispositivos 130a-130c móviles. La Fig. 10 demuestra un proceso 1000 de selección, en el que uno de los audios grabados desde cada uno de los dispositivos 130a-130c móviles se selecciona para su transmisión a los dispositivos remotos.

Con referencia primero a la Fig. 9, los dispositivos 130a, 130b y 130c pueden ser los dispositivos 130a, 130b y 130c de la sala 110 del sistema 100 de las Figs. 1A-1B. Uno de los participantes en la sala 110 puede empezar a hablar. La señal de audio resultante es detectada por los micrófonos de todos los dispositivos 130a-130c de usuario móviles. Cada uno de los dispositivos 130a-130c móviles graba la señal de audio y divide la señal de audio grabada en paquetes mediante la rutina en las Figs. 5 o 7. De esta manera, el dispositivo 130a líder produce un primer conjunto 910 de paquetes a partir de la señal de audio grabada. El dispositivo 130b cliente produce otro conjunto 920 de paquetes a partir de la señal de audio grabada. El dispositivo 130c cliente produce otro conjunto 930 de paquetes a partir de la señal de audio grabada.

Con referencia ahora a la Fig. 10, la presente divulgación contempla un proceso 1000 ejemplar de selección de secuencia. Todos los dispositivos de usuario móviles reciben una señal de audio desde la misma fuente (1010). Debido a un retraso de tiempo u otros factores relacionados con la sala, tal como se ha descrito anteriormente, la señal de audio recibida en cada dispositivo puede variar ligeramente y, por consiguiente, D1 recibe y graba la señal 1 de audio, D2 recibe y graba la señal 2 de audio y D3 recibe y graba la señal 3 de audio.

Cada uno de los dispositivos D1, D2 de usuario móviles convierte la señal en paquetes para su transmisión al teléfono líder (1012); por ejemplo, un proceso ejemplar para la formación de paquetes a partir de audio grabado se ha descrito anteriormente con relación a las Figs. 5-7. Después de la formación, cada paquete se transmite desde los dispositivos cliente al dispositivo L1 líder. Por ejemplo, D1 transmite la señal 1 de audio a L1 en paquetes A1-An durante un período de tiempo. Por ejemplo, D2 transmite la señal 2 de audio a L1 en paquetes B1-Bn durante un período de tiempo. Por ejemplo, L1 recibe los paquetes A1 y B1 al mismo tiempo que se forma el paquete C1, L1 recibe los paquetes A2 y B2 al mismo tiempo que se forma el paquete C2, y L1 recibe los paquetes An y Bn al mismo tiempo que se forma el paquete Cn. Los tiempos para la recepción de paquetes desde D1 y D2 varían según la fluctuación de la red y el proceso 800 de sincronización descrito anteriormente con relación a la Fig. 8.

El dispositivo líder realiza el procesamiento sobre cada paquete mediante el módulo 354 de preprocesamiento descrito en la Fig. 3 (1014). Una vez recibidos un número umbral de paquetes, el dispositivo líder selecciona el dispositivo móvil correspondiente a al menos un paquete con las mejores métricas de procesamiento durante un período de tiempo de evaluación umbral (1016). Por ejemplo, el dispositivo líder evalúa el dispositivo móvil correspondiente a al menos un paquete con las mejores métricas durante un período de tiempo umbral (por ejemplo, el período de tiempo puede corresponder a 200-250 ms de datos de audio o 7-11 paquetes). La métrica de procesamiento puede ser la señal más alta en términos de decibelios o la señal con la mayor claridad, o de otro modo como se describe el presente documento. Por ejemplo, si los paquetes A1-A11 son superiores a los paquetes B1-B11 y superiores a los paquetes C1-C11, entonces se selecciona la secuencia 930 de audio desde el dispositivo 130c en la etapa 1016.

A continuación, el dispositivo líder transmite la secuencia de paquetes de audio seleccionada a la red 140 para su transmisión a los dispositivos líderes remotos; los dispositivos líderes remotos transmiten a continuación la secuencia de audio recibida a sus dispositivos clientes locales correspondientes (1018). La secuencia de audio seleccionada se transmite durante un periodo de tiempo de transmisión umbral. Por ejemplo, si se selecciona la secuencia 930 de audio según la etapa 1016, entonces la etapa 1018 permite la transmisión de la secuencia 930 de audio a la red 140 durante dicho periodo de tiempo umbral. En algunos ejemplos, el periodo de tiempo de transmisión umbral en la etapa 1018 es más largo que el periodo de tiempo de evaluación umbral en la etapa 1016. Por ejemplo, el periodo de tiempo de transmisión umbral está comprendido entre 300-700 ms, y es preferiblemente de aproximadamente 500 ms.

El dispositivo líder puede realizar repetidamente las etapas 1014 a 1018 para reevaluar periódicamente qué secuencia de audio transmitir a los otros dispositivos líderes. En este proceso, el dispositivo líder no necesita evaluar todos los datos de audio desde cada dispositivo en la sala. Por el contrario, el dispositivo líder puede evaluar solo subconjuntos de los datos de audio recibidos, tales como las secuencias de paquetes que representan ventanas de tiempo particulares usadas como periodos de evaluación. Estas ventanas de evaluación están separadas de manera que las secuencias evaluadas de paquetes de audio tengan secuencias de paquetes de audio no evaluados entre las mismas. En base a cada evaluación, el dispositivo líder selecciona una secuencia de audio a transmitir durante una parte limitada de la llamada, y el dispositivo L1 líder puede transmitir una cantidad predeterminada de paquetes de audio (por ejemplo, paquetes que representan una duración predeterminada del audio grabado) de la secuencia de audio seleccionada después de cada evaluación. El dispositivo L1 líder transmite una cantidad de paquetes de audio que representan una cantidad de tiempo predeterminada que excede la duración del periodo de evaluación. Por ejemplo, después de cada evaluación, el dispositivo L1 líder puede transmitir una parte de la secuencia de audio seleccionada que incluye (i) la secuencia de los paquetes de audio de la secuencia seleccionada que corresponden al periodo de evaluación, y (ii) una cantidad predeterminada de paquetes de audio adicionales de la secuencia seleccionada que siguen a los paquetes de audio correspondientes al periodo de evaluación.

En muchos casos, la transmisión de más datos de audio de los que se incluyen en un periodo de evaluación puede

proporcionar diversas ventajas. Por ejemplo, usando una duración predeterminada para cada período de transmisión, el dispositivo líder puede proporcionar audio en segmentos que proporcionan cada uno al menos una duración mínima para la reproducción. Esto ayuda a mantener la secuencia de audio de salida global sin interrupciones y mejora la calidad de audio al evitar una conmutación demasiado frecuente entre las secuencias de audio. La transmisión de al menos una duración mínima de datos de audio (por ejemplo, para una longitud fija de 500 ms) para cada evaluación de selección de secuencia, proporciona una consistencia que aumenta la inteligibilidad de las secuencias de audio transmitidas a otras salas.

Además, la limitación de la evaluación de las secuencias de audio a determinados períodos de evaluación (por ejemplo, períodos de evaluación más cortos que los períodos de transmisión) reduce la carga de procesamiento en el dispositivo líder. Esto contribuye a una mayor eficiencia y a un menor consumo de energía del dispositivo líder, que frecuentemente es un dispositivo móvil alimentado por batería. En lugar de evaluar todos los paquetes de todos los dispositivos, el dispositivo líder puede evaluar paquetes de audio que representan una fracción de la duración total de una llamada. Por ejemplo, si el dispositivo líder usa períodos de evaluación, cada uno con una duración de 250 ms, y si el dispositivo líder transmite 500 ms de datos de audio para cada evaluación (por ejemplo, los 250 ms de datos de audio evaluados, así como los 250 ms de datos de audio posteriores), el procesamiento para evaluar los datos de audio grabados en la sala es la mitad que si se evaluaran todos los datos de audio. La eficiencia puede aumentarse con diferentes relaciones de períodos de evaluación a períodos de transmisión. Por ejemplo, el uso de un período de evaluación de 200 ms y un período de transmisión de 800 ms usa solo una cuarta parte de la cantidad de procesamiento que la evaluación de todos los paquetes de audio. Este beneficio de eficiencia puede ser especialmente útil a medida que aumenta el número de dispositivos en una sala y, de esta manera, el número de secuencias de audio a evaluar.

Usando estas técnicas, pueden establecerse la duración de la transmisión o la frecuencia entre evaluaciones para proporcionar una alta capacidad de respuesta durante una llamada. Por ejemplo, la duración de la transmisión puede ser de un segundo o menos, lo que permite reevaluar qué secuencia de audio se emitirá al menos una vez por segundo o más frecuentemente. Esto puede permitir que el dispositivo líder varíe qué secuencia de audio se transmite, tal como conmutar entre secuencias de audio a medida que los usuarios de los diferentes dispositivos cliente alternan el habla o a medida que el movimiento de los dispositivos en la sala cambia qué secuencia de audio proporciona la calidad más alta.

Como un ejemplo más detallado, el dispositivo L1 líder recibe secuencias de audio desde uno o más dispositivos D1, D2 en la sala. A medida que se reciben las secuencias de audio, el dispositivo L1 líder usa las marcas de tiempo y los números de secuencia transmitidos con las secuencias de audio para correlacionar los datos de audio desde los múltiples dispositivos. Por ejemplo, tal como se muestra en la Fig. 9, el dispositivo L1 líder puede determinar una alineación de paquetes de audio desde cada uno de los dispositivos L1, D1 y D2 con respecto al tiempo, de manera que se conozcan los paquetes de audio que representan los mismos tiempos. El dispositivo L1 líder puede almacenar y recuperar también valores de parámetros que establecen las características de la evaluación. Estos parámetros pueden establecer la duración del período de evaluación, para especificar el período de tiempo o el número de paquetes que se evaluarán para cada decisión de selección de secuencia. Los parámetros pueden establecer también la duración del período de transmisión, para especificar una cantidad predeterminada de datos de audio a transmitir por cada evaluación, independientemente de si se especifica como una cantidad total de datos de audio por cada evaluación o como una cantidad de paquetes de audio adicionales a proporcionar además de los paquetes de audio evaluados en el período de evaluación.

Como un ejemplo, pueden generarse paquetes de audio para cada codificación de 25 ms de datos de audio. El dispositivo L1 líder puede recuperar los valores de parámetros almacenados que establecen el período de evaluación en 8 paquetes o 200 ms de datos de audio, y puede establecer el período de transmisión en 20 paquetes o 500 ms de datos de audio. En muchos casos, el período de transmisión es el doble de largo que el período de evaluación o más largo (por ejemplo, tres veces, cuatro veces, cinco veces, etc.). Como resultado, la cantidad predeterminada de paquetes de audio adicionales usados después de cada selección de secuencia de audio puede ser una duración de datos de audio que es tan grande como o mayor que la duración de los datos de audio evaluados en el período de evaluación correspondiente. Por ejemplo, para cada evaluación de una ventana de datos de audio de 200 ms, el dispositivo líder puede transmitir 500 ms de datos de audio desde la secuencia seleccionada (por ejemplo, los 200 ms de paquetes de datos de audio evaluados en la secuencia seleccionada, más los 300 ms de paquetes de datos de audio posteriores que siguen inmediatamente a los paquetes de datos de audio en el período de evaluación).

Con referencia a la Fig. 9, el dispositivo líder puede establecer un período de evaluación de t_0 a t_8 . El dispositivo L1 líder puede evaluar cada una de las secuencias de audio respectivas en la sala, por ejemplo, el conjunto 910 de paquetes desde el dispositivo L1 líder, el conjunto 920 de paquetes desde el dispositivo D1 cliente y el conjunto 930 de paquetes desde el dispositivo D2 cliente. La evaluación puede incluir la generación de una o más métricas para cada una de las secuencias de audio usando los paquetes en el período de evaluación. De esta manera, el dispositivo L1 líder evalúa una o más métricas para el grupo de paquetes A1-A8, una o más métricas para el grupo de paquetes B1-B8 y una o más métricas para el grupo de paquetes C1-C8. Las métricas para cada grupo de paquetes pueden incluir puntuaciones para el volumen, la claridad, el nivel de actividad de voz, etc. El dispositivo L1 líder evalúa las secuencias de audio comparando las puntuaciones para las diferentes secuencias de audio para determinar cuál de ellas tiene voz u otro audio con la calidad más alta.

Después de seleccionar la secuencia de audio en base a la evaluación, el dispositivo L1 líder transmite una cantidad predeterminada de paquetes según lo especificado por el valor o los valores de los parámetros almacenados que

establecen la duración del período de transmisión. Por ejemplo, en base a la evaluación de los paquetes A1-A8, B1-B8 y C1-C8, el dispositivo L1 líder puede seleccionar la secuencia de audio del dispositivo D1 cliente como la que tiene la actividad de voz con la calidad más alta. Usando un período de transmisión predeterminado de 500 ms, el dispositivo líder envía 500 ms de datos de audio desde el dispositivo D1 cliente (por ejemplo, los paquetes B1-B20). Esto incluye los paquetes de audio B1-B8 en el período de evaluación, así como los paquetes adicionales B9-B20 que no se evaluaron en la evaluación. Después del final del período de transmisión, el dispositivo L1 líder empieza entonces la siguiente evaluación. Por ejemplo, el dispositivo L1 líder evaluaría grupos A21-A28, B21-B28 y C21-C28 de paquetes en un segundo período de evaluación de la misma duración predeterminada usada anteriormente, por ejemplo, 200 ms. El dispositivo L1 líder seleccionaría cuál de las secuencias de audio proporciona los mejores datos de audio en el segundo período de evaluación, antes de transmitir paquetes en un segundo período de transmisión que tiene la misma duración predeterminada usada en el primer período de transmisión. Si la segunda evaluación muestra que el dispositivo L1 líder tiene la mejor grabación de audio, entonces el dispositivo L1 líder transmitirá 500 pulgadas de datos de audio desde su propia secuencia, por ejemplo, los paquetes A21-A40. De esta manera, el dispositivo L1 líder puede reevaluar periódicamente qué secuencia de audio tiene la mayor actividad y calidad de voz, y puede cambiar la secuencia de audio fuente que se transmite en consecuencia.

La aplicación ejemplar realiza también la cancelación de reverberación para los dispositivos de usuario en una sala. La reverberación se produce cuando las señales de audio reverberan desde las paredes, los muebles u otros obstáculos físicos de las salas. El eco se produce cuando la salida de audio de un dispositivo se recibe en un micrófono o grabadora del mismo dispositivo; la presente divulgación contempla también que el eco incluya la salida de audio sincronizada desde otros dispositivos locales que se recibe en el micrófono/grabadora del dispositivo. Cada dispositivo móvil tiene capacidades de cancelación de eco inherentes al dispositivo que reducen la retroalimentación durante las llamadas/conferencias de audio y vídeo. Por ejemplo, cada dispositivo está configurado para determinar su impulso, que es la diferencia entre la señal de audio emitida por el dispositivo y la entrada de señal de audio recibida del dispositivo. Debido a que las señales de audio se sincronizan mediante los métodos divulgados para su reproducción en todos los dispositivos locales, la diferencia en la entrada y salida de audio en cada dispositivo es insignificante, por lo que la cancelación de eco de cada dispositivo puede cancelar inherentemente cualquier eco o reverberación recibidos en el dispositivo que provengan de la reproducción sincrónica desde otros dispositivos locales. Esta ventajosa cancelación de eco solo es posible debido a que los métodos divulgados proporcionan una sincronización superior a la de los métodos convencionales.

La Fig. 11 muestra un proceso 1100 para el algoritmo de sincronización de reproducción para un dispositivo líder. Tal como se muestra en la Fig. 11, los paquetes de audio son recibidos por un dispositivo líder local desde un dispositivo líder remoto a través de Internet (1110). El tiempo de reproducción para este paquete de audio se calcula en base a un almacenamiento intermedio de tamaño fijo añadido para compensar el retraso de red entre el dispositivo líder local y los dispositivos cliente locales. A continuación, el tiempo de reproducción calculado se envía a los dispositivos cliente desde el dispositivo líder local junto con una versión codificada de los paquetes de audio (1112). A continuación, la carga útil extraída se inserta en un almacenamiento intermedio de tamaño fijo en el dispositivo líder (1114). La longitud del almacenamiento intermedio puede ser fija, pero, para satisfacer la latencia de hardware del dispositivo líder, entonces el tamaño del almacenamiento intermedio se ajusta una vez. Esto no tiene efecto sobre el tiempo de reproducción proyectado que se envía a los dispositivos cliente. A continuación, se extraen los datos del almacenamiento intermedio y los datos se extraen desde los paquetes de audio para su reproducción (1116). A continuación, los datos de carga útil se proporcionan al dispositivo líder para su reproducción usando un reproductor oboe (1118).

En la Fig. 12 se muestra un proceso 1200 del algoritmo de sincronización de reproducción para los dispositivos cliente. Los paquetes de audio se reciben en un dispositivo cliente local desde un dispositivo líder local (1210) y se colocan en un almacenamiento intermedio antifluencia Speex de un dispositivo cliente (1212). Se extrae un paquete desde este almacenamiento intermedio tras las solicitudes de reproducción (1214). Los datos de carga útil se decodifican y el tiempo de reproducción adjunto proporcionado por el dispositivo líder se compara con el tiempo actual (1216). Si la diferencia es tolerable, el paquete se envía para su reproducción (1220). Si la diferencia no es tolerable, el algoritmo de sincronización en el dispositivo cliente ajusta los paquetes de audio para alinear el tiempo (1218). El tiempo actual en los dispositivos cliente se define como:

Tiempo actual = Tiempo del sistema + desfase NTP + latencia del hardware.

A continuación, los datos se envían de vuelta al reproductor oboe después de que el algoritmo de sincronización ajuste los mismos para su reproducción (1220).

Aunque los ejemplos anteriores se refieren a señales de audio, debe entenderse que los mismos principios se pueden aplicar a las secuencias de imágenes usadas en videoconferencias. Tal como se ha explicado anteriormente, cada uno de los dispositivos móviles puede incluir una cámara que permite la captura de una secuencia de vídeo del participante. El teléfono líder puede combinar las diferentes secuencias de vídeo y puede enviarlas a otros participantes a través de los teléfonos líderes respectivos en otras salas. El teléfono líder puede recibir y enrutar también secuencias de vídeo externas recibidas a los dispositivos cliente en la sala. Las secuencias de vídeo respectivas pueden mostrarse en las pantallas de los dispositivos móviles en la sala.

Los diagramas de flujo en las Figs. 3-8 y 10 son representativos de instrucciones legibles por máquina ejemplares para procesar las señales de audio desde dentro de la red ad hoc y desde dispositivos externos. En este ejemplo, las

instrucciones legibles por máquina comprenden un algoritmo para su ejecución por: (a) un procesador; (b) un controlador; y/o (c) uno o más dispositivos de procesamiento adecuados. El algoritmo puede materializarse en un software almacenado en medios tangibles, tales como memoria flash, CD-ROM, disquete, disco duro, disco de vídeo digital (versátil) (DVD) u otros dispositivos de memoria. Sin embargo, las personas con conocimientos ordinarios en la materia apreciarán fácilmente que el algoritmo completo y/o partes del mismo pueden ser ejecutados de manera alternativa por un dispositivo distinto de un procesador y/o pueden materializarse en firmware o hardware dedicado de una manera bien conocida (por ejemplo, puede materializarse en un circuito integrado específico de la aplicación [ASIC], un dispositivo lógico programable [PLD], un dispositivo lógico programable por campo [FPLD], una matriz de puertas programables por campo [FPGA], lógica discreta, etc.). Por ejemplo, cualquiera o la totalidad de los componentes de las interfaces pueden implementarse mediante software, hardware y/o firmware. Además, algunas o todas las instrucciones legibles por máquina representadas por los diagramas de flujo pueden implementarse manualmente. Además, aunque los algoritmos ejemplares se describen con referencia a los diagramas de flujo ilustrados en las Figs. 3-8 y 10, las personas con conocimientos ordinarios en la materia apreciarán fácilmente que pueden usarse de manera alternativa muchos otros métodos de implementación de las instrucciones legibles por máquina ejemplares. Por ejemplo, puede cambiarse el orden de ejecución de los bloques y/o pueden cambiarse, eliminarse o combinarse algunos de los bloques descritos.

Tal como se usan en la presente solicitud, los términos "componente", "módulo", "sistema" o similares, se refieren generalmente a una entidad relacionada con un ordenador, ya sea hardware (por ejemplo, un circuito), una combinación de hardware y software, software o una entidad relacionada con una máquina operativa con una o más funcionalidades específicas. Por ejemplo, un componente puede ser, pero no está limitada a ser, un proceso que se ejecuta en un procesador (por ejemplo, un procesador de señales digitales), un procesador, un objeto, un ejecutable, un hilo de ejecución, un programa y/o un ordenador. A modo de ilustración, tanto una aplicación que se ejecuta en un controlador, como el controlador, pueden ser un componente. Uno o más componentes pueden residir en el interior de un proceso y/o un hilo de ejecución, y un componente puede estar localizado en un ordenador y/o puede estar distribuido entre dos o más ordenadores. Además, un "dispositivo" puede tener forma de hardware diseñado especialmente, hardware generalizado especializado mediante la ejecución de software en el mismo que permite que el hardware realice una función específica, software almacenado en un medio legible por ordenador o una combinación de los mismos.

La terminología usada en el presente documento tiene el propósito de describir realizaciones particulares únicamente y no pretende limitar la invención. Tal como se usan en el presente documento, las formas singulares "un", "una" y "el/la" pretenden incluir también las formas plurales, a menos que el contexto indique claramente lo contrario. Además, en la medida en que los términos "incluyendo", "incluye", "que tiene", "tiene", "con" o variantes de los mismos, se usan en la descripción detallada y/o en las reivindicaciones, dichos términos pretenden ser inclusivos de una manera similar al término "que comprende".

A menos que se defina de otra manera, todos los términos (incluyendo los términos técnicos y científicos) usados en el presente documento tienen el mismo significado que entiende comúnmente una persona con conocimientos ordinarios en la materia. Además, los términos, tales como los definidos en los diccionarios de uso común, deben interpretarse como si tuvieran un significado que sea coherente con su significado en el contexto de la técnica pertinente, y no se interpretarán en un sentido idealizado o excesivamente formal a menos que se defina expresamente de esa manera en el presente documento.

Aunque anteriormente se han descrito diversas realizaciones de la presente invención, debe entenderse que se han presentado solo a modo de ejemplo y no a modo de limitación. La invención está definida por las reivindicaciones adjuntas.

REIVINDICACIONES

1. Procedimiento para establecer un sistema de teleconferencia entre dos ubicaciones, que comprende:

transmitir, por parte de un dispositivo de comunicación líder, una primera señal entre el dispositivo (130, 132, 134) de comunicación líder y un dispositivo (130, 132, 134) de comunicación cliente, en el que cada uno de entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente incluye un transceptor de comunicación de red, un altavoz (242) y un micrófono (240), en el que el dispositivo de comunicación líder y el dispositivo de comunicación cliente están situados en una primera sala;

generar (410, 412, 414, 416, 418, 420, 422, 424), por parte del dispositivo de comunicación líder, primeros datos de audio en base a una entrada de audio que proviene de una fuente de audio y es recibida por el micrófono (240) del dispositivo de comunicación líder;

dividir (426), por parte del dispositivo de comunicación líder, los primeros datos de audio en una primera pluralidad de paquetes de audio;

recibir, por parte del dispositivo de comunicación líder, una segunda pluralidad de paquetes de audio desde el dispositivo de comunicación cliente, en el que la segunda pluralidad de paquetes de audio incluye segundos datos de audio que el dispositivo de comunicación cliente ha generado en base a la entrada de audio procedente de la fuente de audio y es recibida por el micrófono del dispositivo de comunicación cliente;

determinar, por parte del dispositivo de comunicación líder, una primera métrica para al menos un paquete de audio en la primera pluralidad de paquetes de audio;

determinar, por parte del dispositivo de comunicación líder, una segunda métrica para al menos un paquete de audio en la segunda pluralidad de paquetes de audio;

comparar, por parte del dispositivo de comunicación líder, la primera métrica con la segunda métrica;

y

transmitir, por parte del dispositivo de comunicación líder, un conjunto de paquetes de audio a un dispositivo remoto situado en una segunda sala,

en el que el conjunto de paquetes de audio se selecciona en base a la comparación de la primera métrica con la segunda métrica para incluir i) el al menos un paquete de audio en la primera pluralidad de paquetes de audio y un número predeterminado de paquetes de audio adicionales en la primera pluralidad de paquetes de audio o ii) el al menos un paquete de audio en la segunda pluralidad de paquetes de audio y un número predeterminado de paquetes de audio adicionales en la segunda pluralidad de paquetes de audio.

2. Método según cualquiera de las reivindicaciones anteriores, en el que: el conjunto transmitido de paquetes de audio incluye i) el al menos un paquete de audio en la primera pluralidad de paquetes de audio y el número predeterminado de paquetes de audio adicionales en la primera pluralidad de paquetes de audio; y el método comprende además determinar que la primera métrica es mayor que la segunda métrica, en el que la transmisión del conjunto de paquetes de audio al dispositivo remoto comprende transmitir el al menos un paquete de audio en la primera pluralidad de paquetes de audio y el número predeterminado de paquetes de audio adicionales en la primera pluralidad de paquetes de audio al dispositivo remoto en base a la determinación de que la primera métrica es mayor que la segunda métrica.

3. Método según la reivindicación 1, en el que: el conjunto transmitido de paquetes de audio incluye ii) el al menos un paquete de audio en la segunda pluralidad de paquetes de audio y el número predeterminado de paquetes de audio adicionales en la segunda pluralidad de paquetes de audio; y el método comprende además determinar que la primera métrica no es mayor que la segunda métrica, en el que la transmisión del conjunto de paquetes de audio comprende transmitir el al menos un paquete de audio en la segunda pluralidad de paquetes de audio y el número predeterminado de paquetes de audio adicionales en la segunda pluralidad de paquetes de audio al dispositivo remoto en base a la determinación de que la primera métrica no es mayor que la segunda métrica.

4. Método según cualquiera de las reivindicaciones anteriores, que comprende, además:

recibir una señal visual desde el dispositivo remoto; y

mostrar la señal visual en las pantallas del dispositivo líder y del dispositivo cliente.

5. Método según cualquiera de las reivindicaciones anteriores, en el que la transmisión al dispositivo remoto se realiza usando un protocolo de voz sobre Internet a través de Internet.

6. Método según cualquiera de las reivindicaciones anteriores, que comprende, además:

recibir una segunda señal desde el dispositivo remoto;

emitir una tercera señal de audio en el dispositivo de comunicación líder en base a la segunda señal; y

emitir una cuarta señal de audio en el dispositivo cliente en base a la segunda señal, en el que la cuarta señal de audio está desfasada con relación a la tercera señal de audio, y

opcionalmente en el que la segunda señal comprende una señal de audio.

5 7. Método según cualquiera de las reivindicaciones anteriores, en el que cada uno de entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente son teléfonos inteligentes.

8. Método según cualquiera de las reivindicaciones anteriores, en el que la transmisión de la primera señal entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente comprende además establecer, por parte del dispositivo de comunicación líder, una comunicación de red con el dispositivo de comunicación cliente en base a
10 que el dispositivo de comunicación cliente recibe información de red a través de:

un indicador visual mostrado por el dispositivo de comunicación líder; o

una comunicación inalámbrica.

9. Método según cualquiera de las reivindicaciones anteriores, en el que:

la primera métrica se basa en una energía media del al menos un paquete de audio; o

15 la primera señal comprende una transmisión que usa un canal de datos Web RTC; o

la primera métrica se basa en una energía media del al menos un paquete de audio y la primera señal comprende una transmisión que usa un canal de datos Web RTC.

10. Método según cualquiera de las reivindicaciones anteriores, en el que: el conjunto transmitido de paquetes de audio incluye i) el al menos un paquete de audio en la primera pluralidad de paquetes de audio y el número
20 predeterminado de paquetes de audio adicionales en la primera pluralidad de paquetes de audio; y transmitir el al menos un primer paquete de audio en la primera pluralidad de paquetes de audio y el número predeterminado de paquetes de audio restantes en la primera pluralidad de paquetes de audio al dispositivo remoto comprende una transmisión que usa un canal de audio Web RTC.

11. Método según cualquiera de las reivindicaciones anteriores, en el que la primera señal comprende una primera
25 frecuencia, y

opcionalmente en el que el método comprende además transmitir una segunda señal que comprende una segunda frecuencia entre el dispositivo de comunicación líder y el dispositivo de comunicación cliente y sincronizar el dispositivo de comunicación líder y el dispositivo de comunicación cliente en base a al menos una de entre la primera señal y la segunda señal.

12. Método según cualquiera de las reivindicaciones anteriores, que comprende además obtener, por parte del dispositivo de comunicación líder, un código para una llamada o conferencia que incluye una comunicación a través
30 de una red entre el dispositivo de comunicación líder y al menos otro dispositivo de comunicación líder;

en el que la transmisión de la primera señal comprende proporcionar, por parte del dispositivo de comunicación líder, la emisión del código; y

35 en el que el método comprende además, después de proporcionar la emisión del código, habilitar la participación del dispositivo de comunicación cliente en la llamada o conferencia en respuesta a una comunicación que el dispositivo de comunicación cliente proporciona en base al código, de manera que el dispositivo de comunicación líder reciba datos de audio grabados por el dispositivo de comunicación cliente y el dispositivo de comunicación líder proporcione datos de audio desde el al menos otro dispositivo de comunicación líder al dispositivo de
40 comunicación cliente, y

opcionalmente, en el que el código identifica i) la llamada o conferencia y ii) una sala, de entre múltiples salas (110, 112, 114) implicadas en la llamada o conferencia, que corresponde al dispositivo de comunicación líder, por ejemplo, en el que el código es un código óptico legible por máquina o un código "QR" de respuesta rápida.

13. Método según cualquiera de las reivindicaciones anteriores, en el que la primera métrica es una medida de la
45 claridad del al menos un paquete de audio de la primera pluralidad de paquetes de audio;

en el que la segunda métrica es una medida de la claridad del al menos un paquete de audio de la segunda pluralidad de paquetes de audio; y

50 en el que el método comprende determinar, por parte del dispositivo de comunicación líder, si seleccionar o no el conjunto de paquetes de audio a transmitir al dispositivo remoto de entre la primera pluralidad de paquetes de audio o la segunda pluralidad de paquetes de audio en base a cuál de entre la primera métrica y la segunda métrica

indica una mayor claridad.

14. Dispositivo electrónico configurado para realizar las operaciones del método según cualquiera de las reivindicaciones 1-13.

- 5 15. Medio de almacenamiento legible por ordenador no transitorio que almacena instrucciones que, cuando son ejecutadas por al menos un procesador de un dispositivo, causan que el dispositivo realice las operaciones del método según cualquiera de las reivindicaciones 1-13.

FIG. 1A

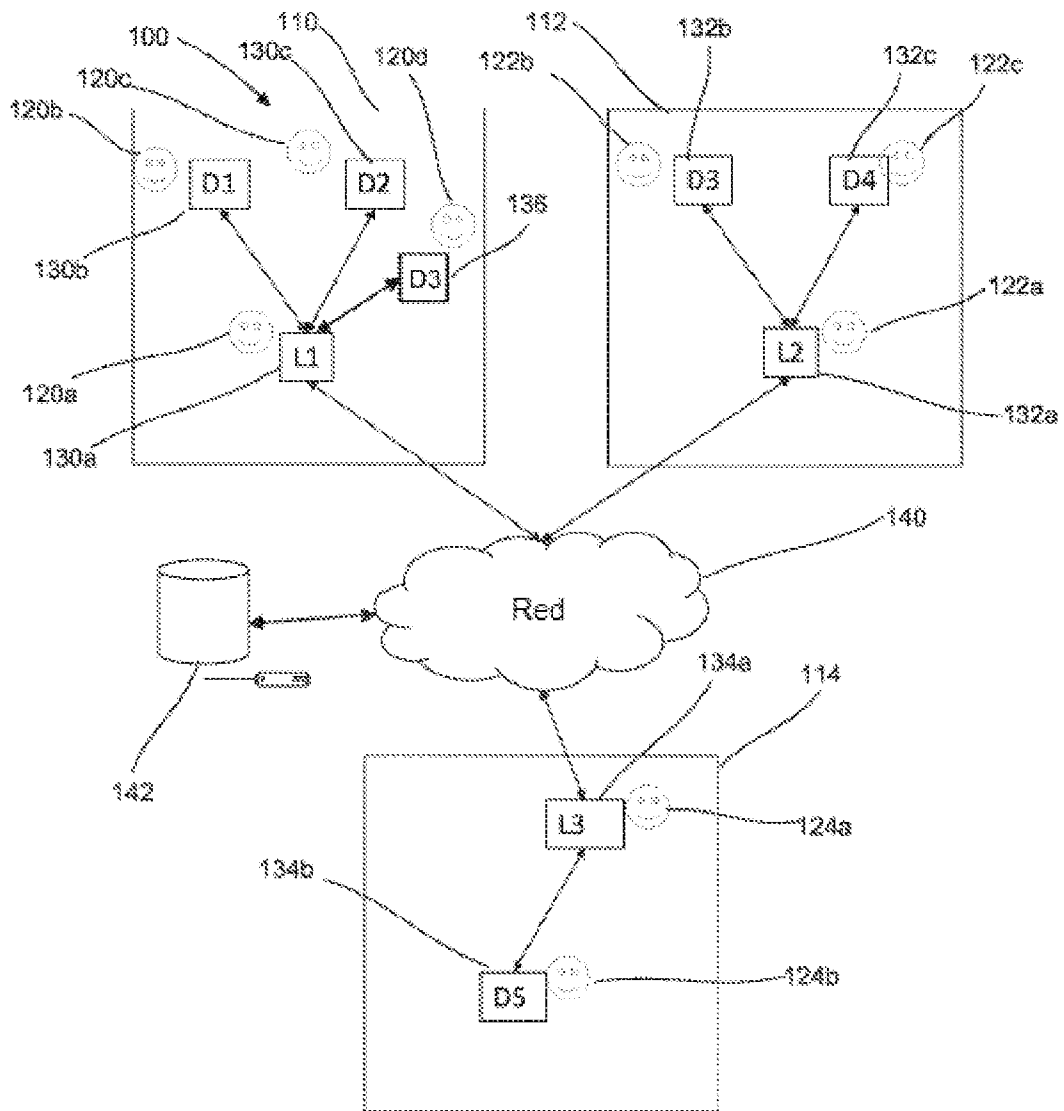


FIG. 1B

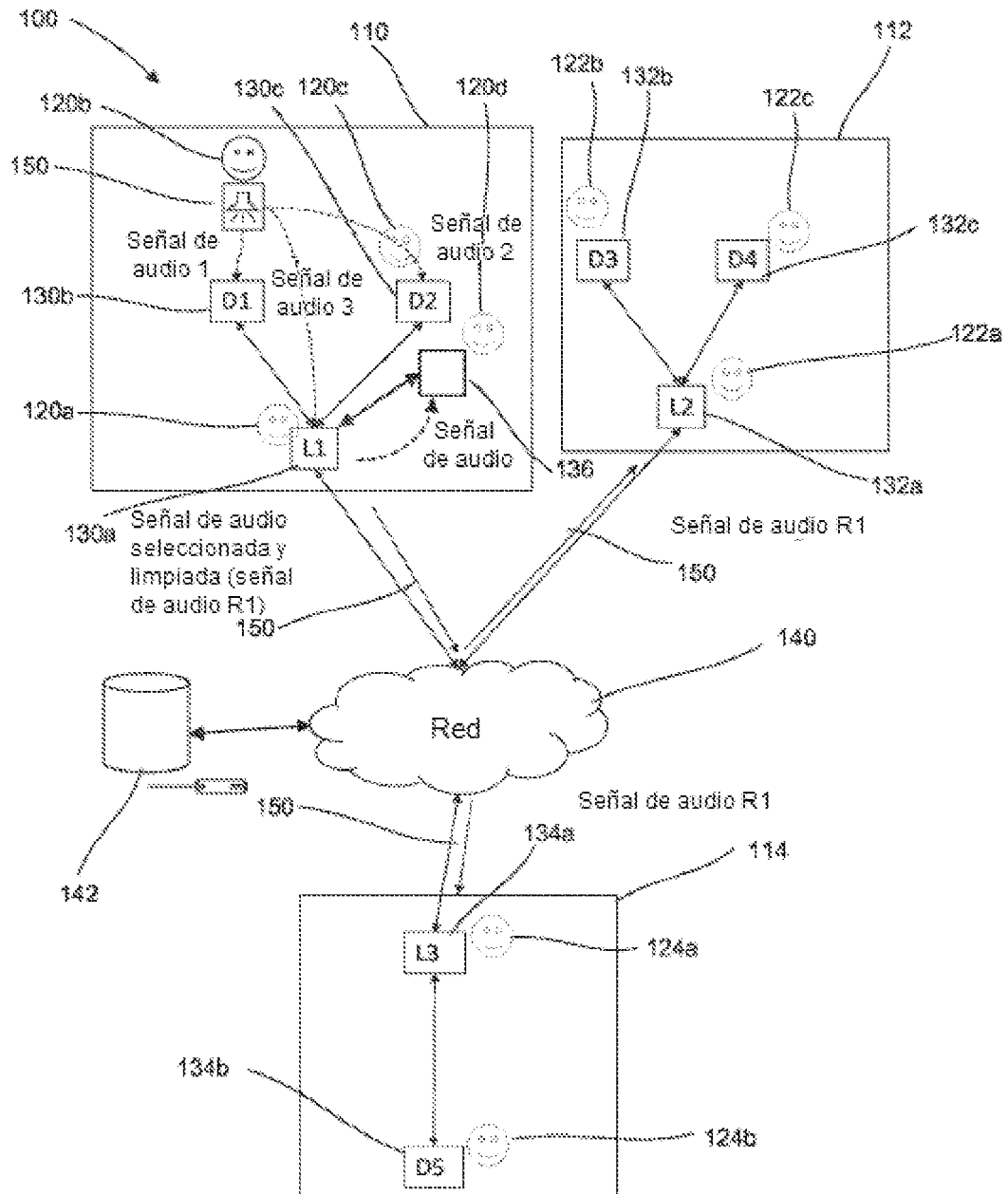
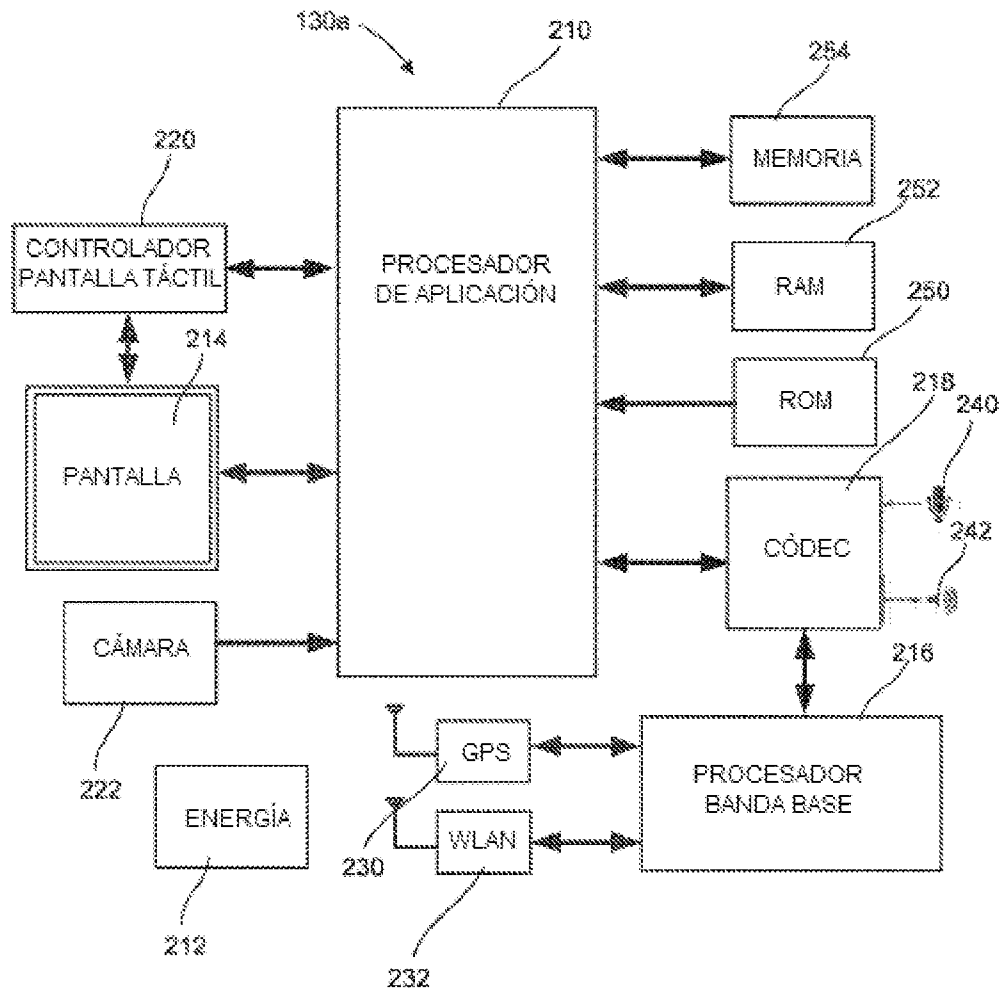


FIG. 2



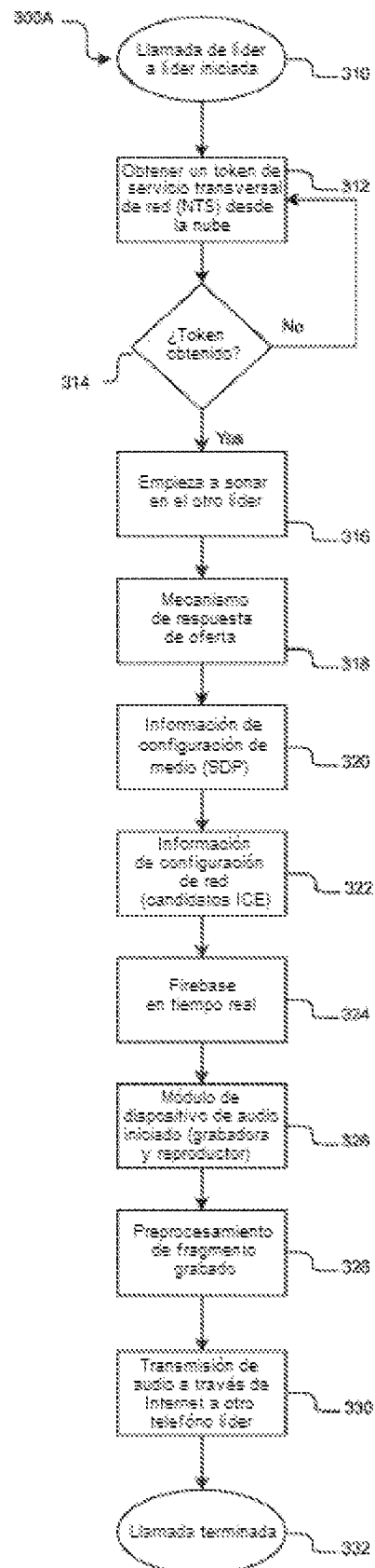


FIG. 3A

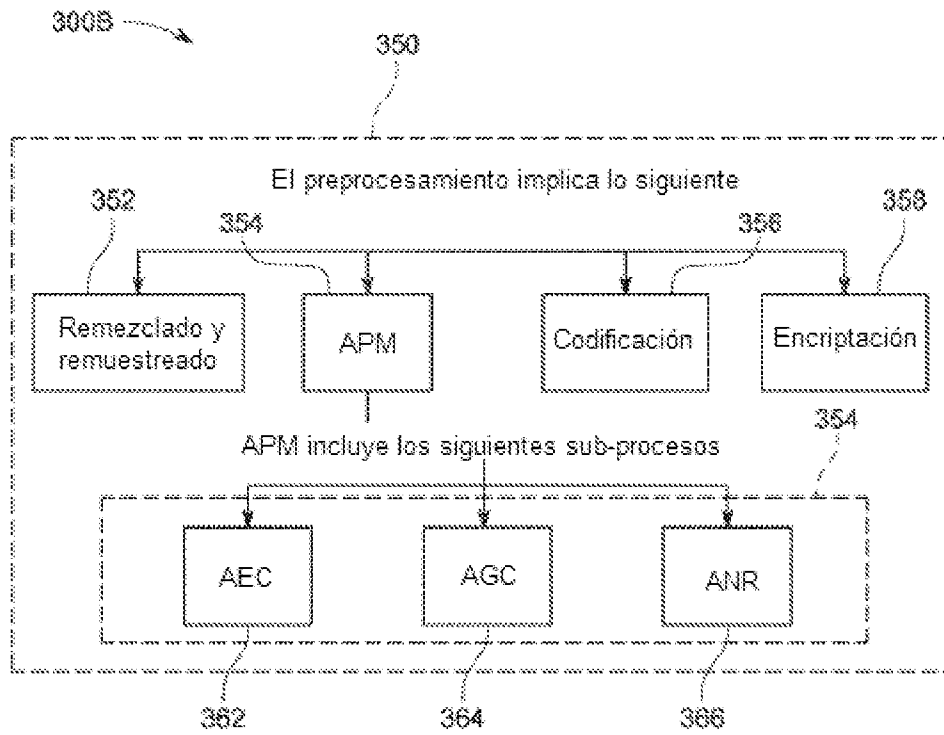


FIG. 3B

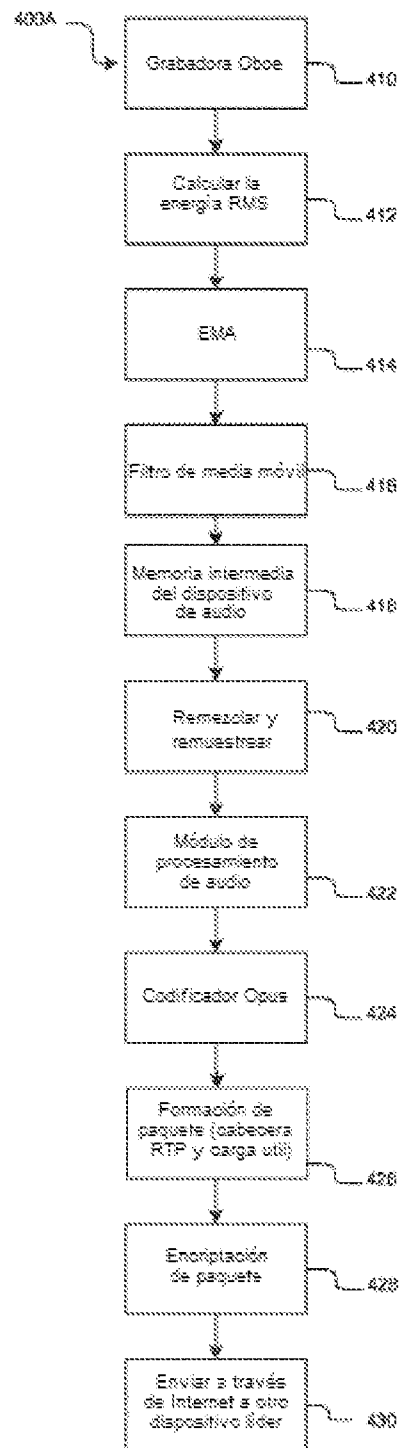


FIG. 4A

400B

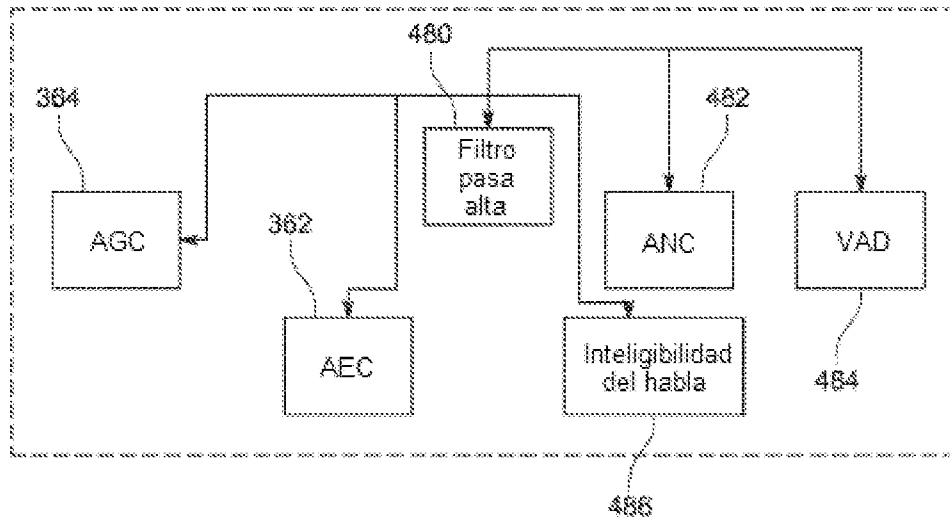


FIG. 4B

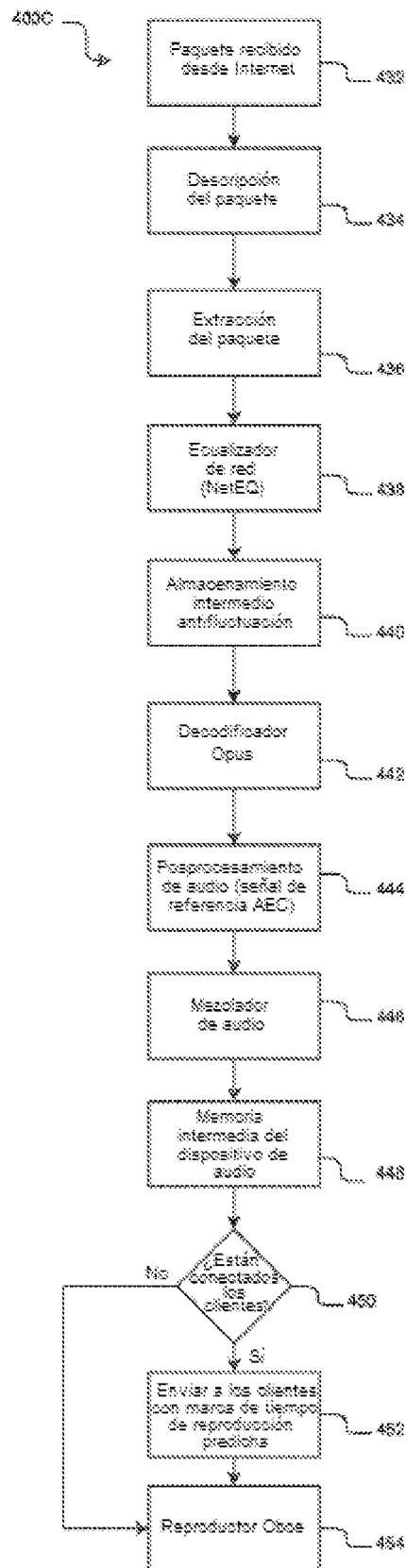


FIG. 4C

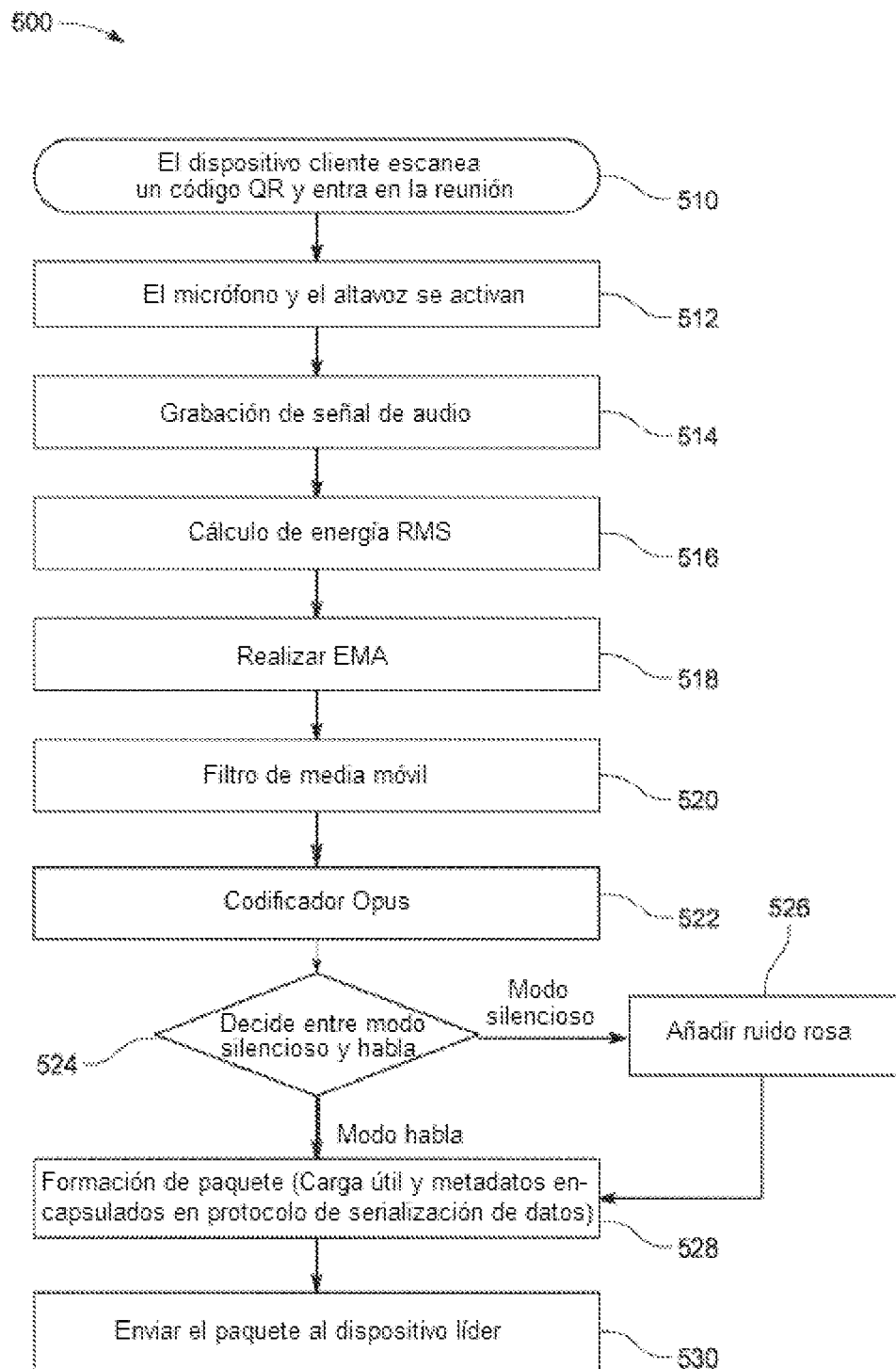


FIG. 5

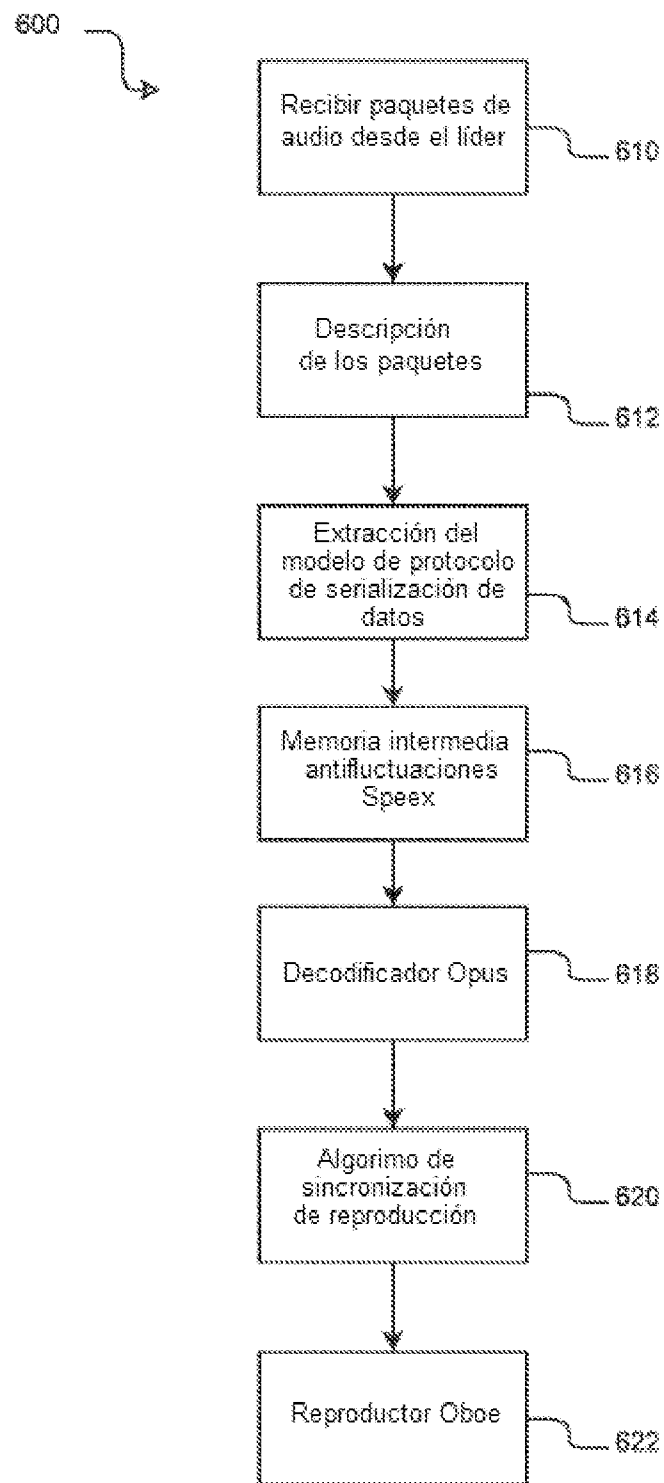


FIG. 6

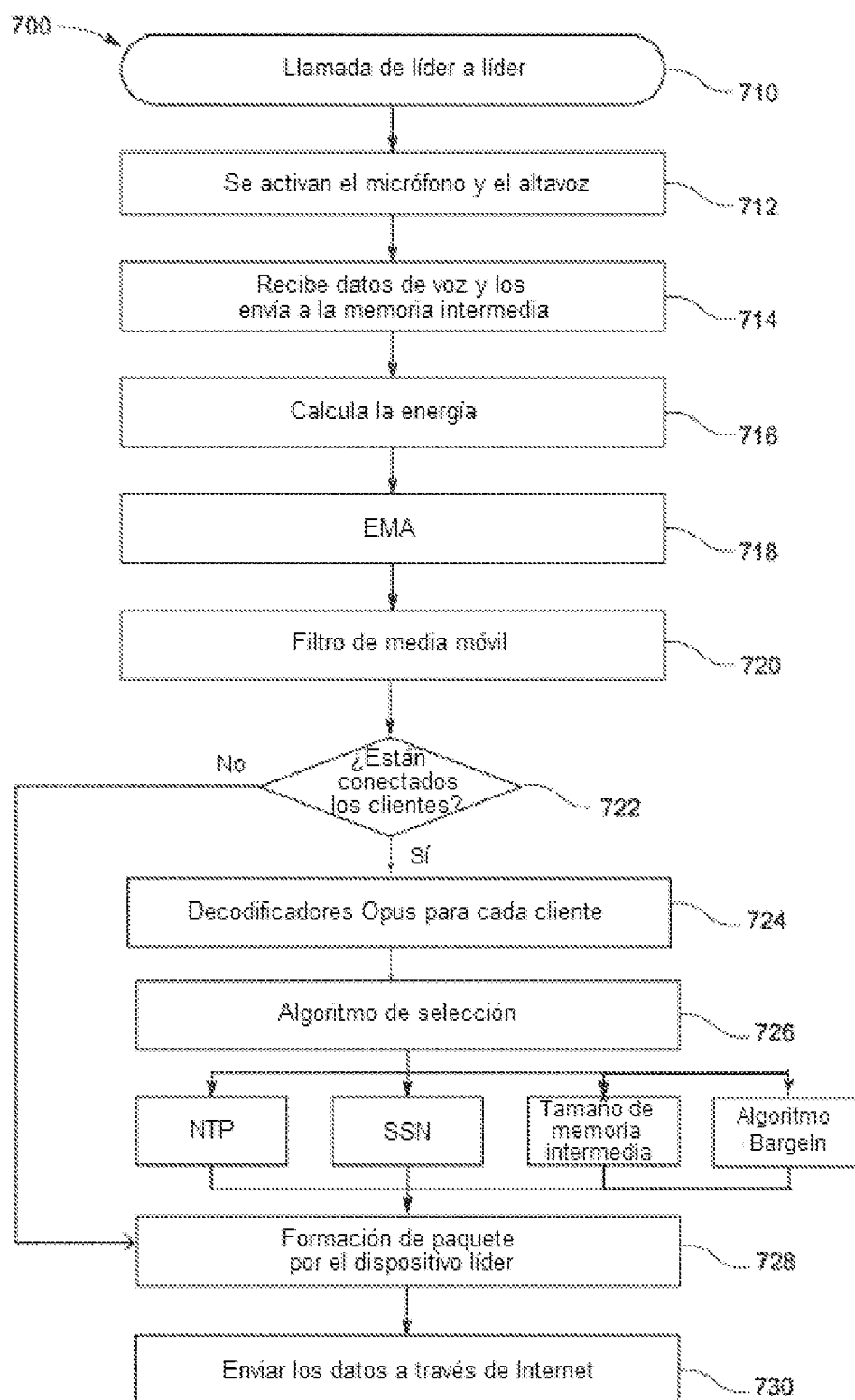


FIG. 7

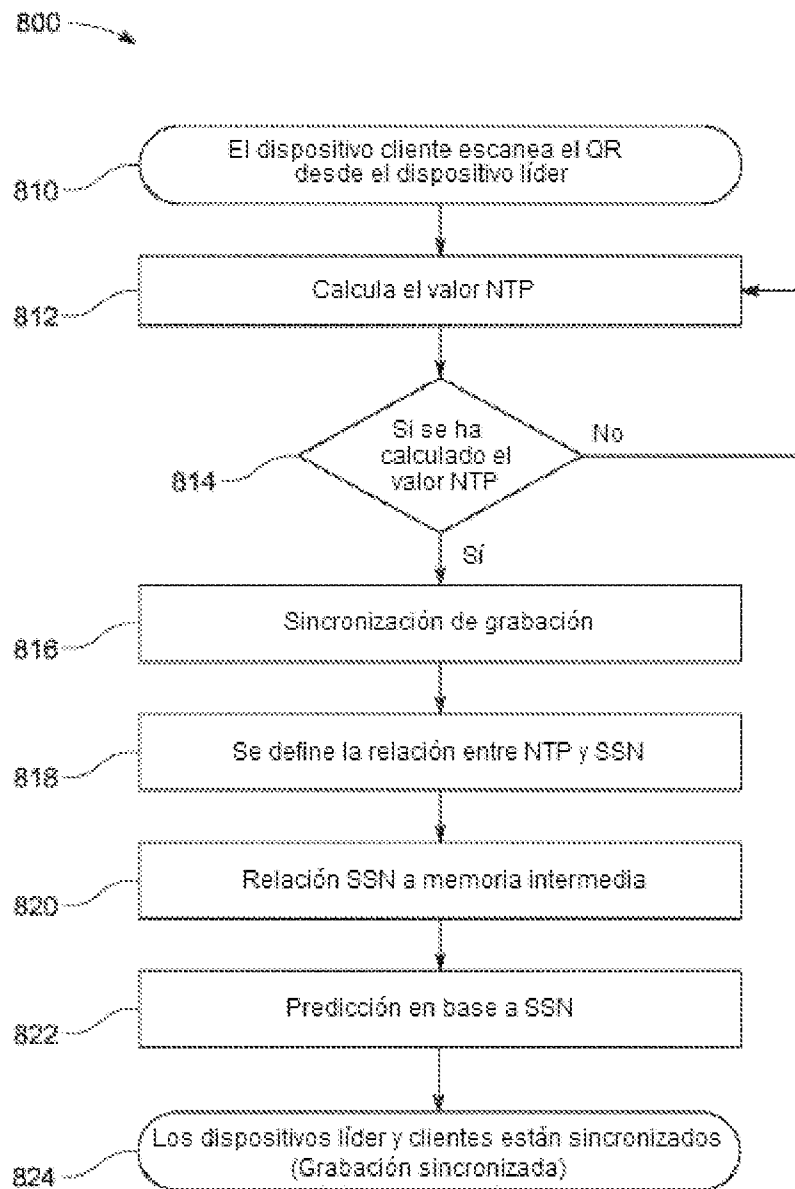


FIG. 8

FIG. 9

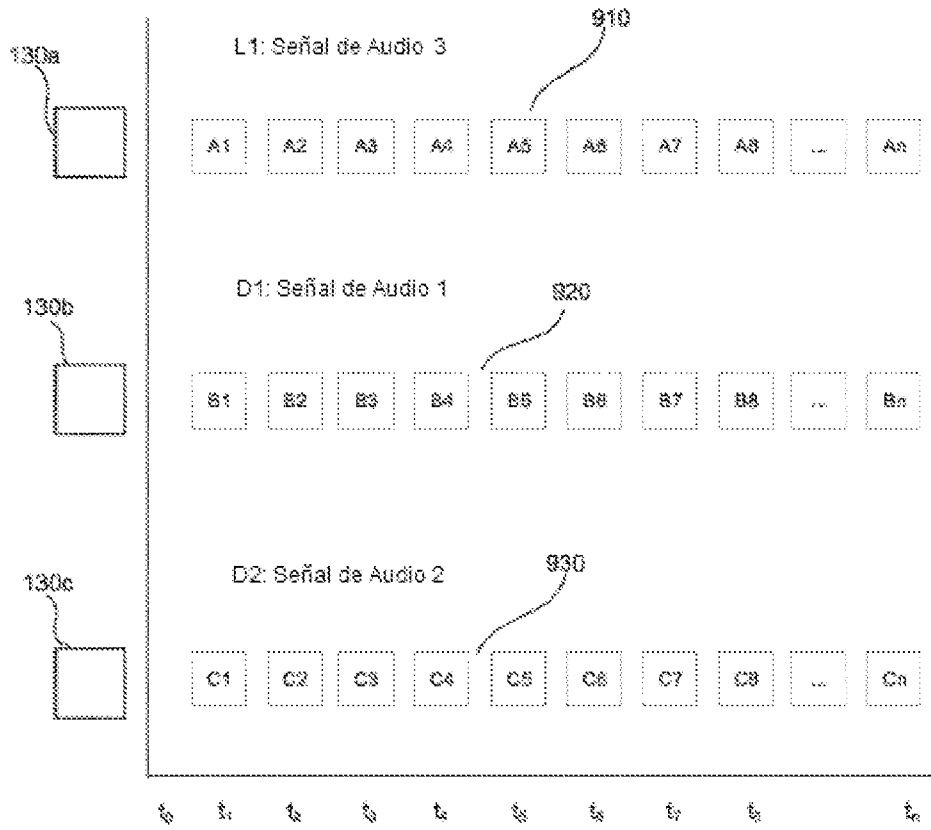
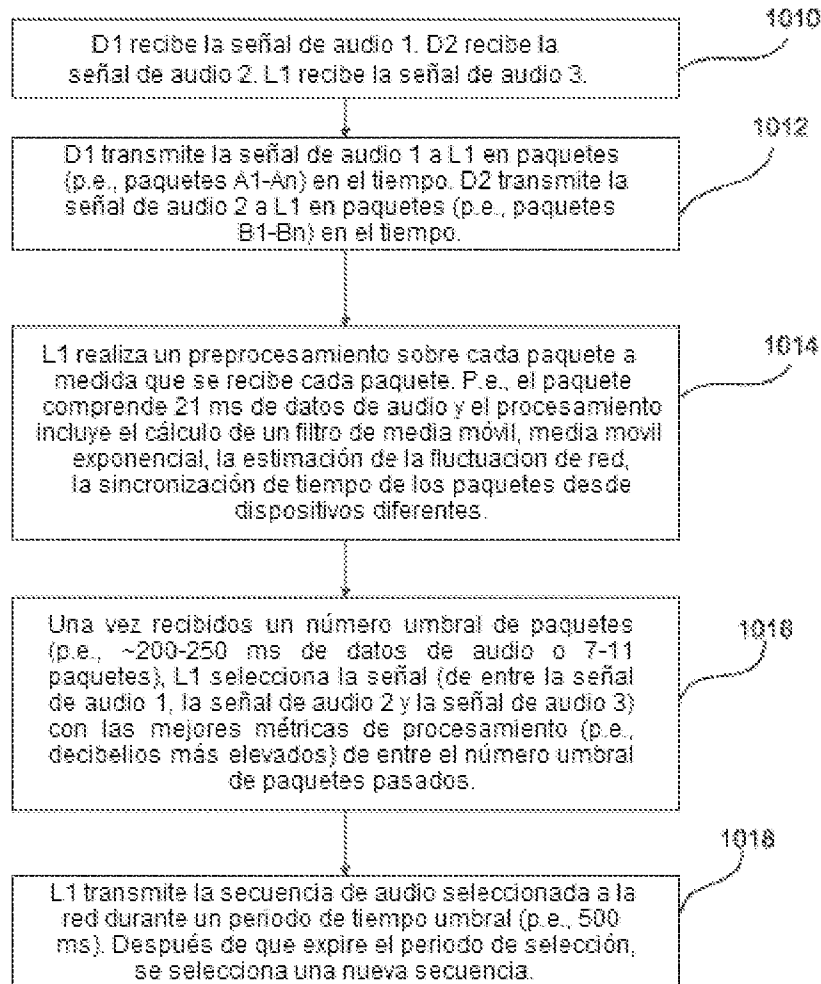


FIG. 10



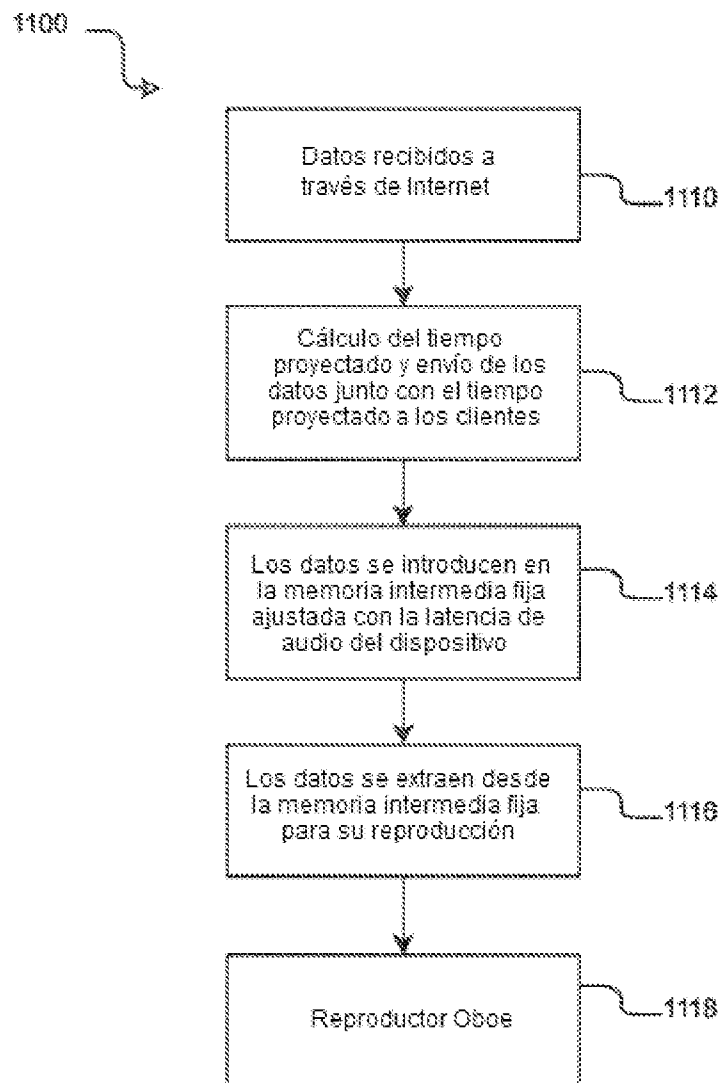


FIG. 11

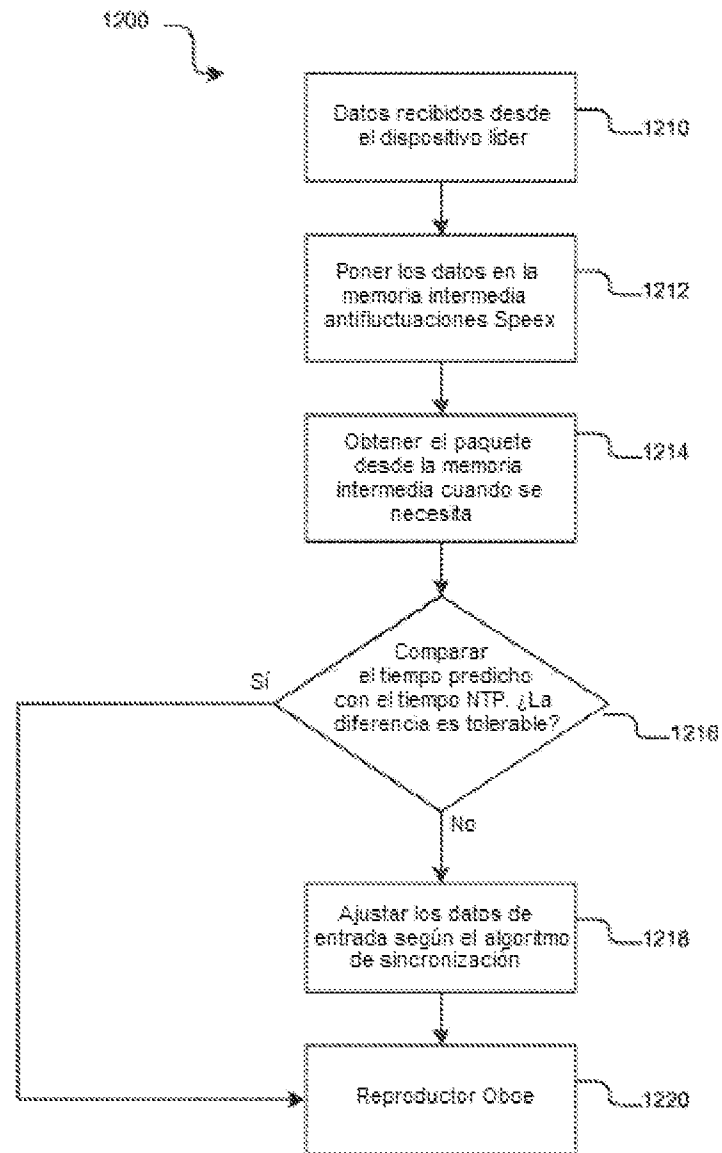


FIG. 12