

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 967 189**

51 Int. Cl.:

**H04J 3/06** (2006.01)

**H04L 65/80** (2012.01)

**H04L 65/00** (2012.01)

**G10L 25/78** (2013.01)

**G10L 25/48** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.04.2015 E 19171996 (2)**

97 Fecha y número de publicación de la concesión europea: **01.11.2023 EP 3550743**

54 Título: **Control de memoria intermedia de fluctuación basado en monitorización de la fluctuación de retardo y la dinámica conversacional**

30 Prioridad:

**16.04.2014 CN 201410152754**  
**06.05.2014 US 201461989340 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**29.04.2024**

73 Titular/es:

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

72 Inventor/es:

**LI, KAI;**  
**SUN, XUEJING y**  
**SPITTLE, GARY**

74 Agente/Representante:

**LINAGE GONZÁLEZ, Rafael**

ES 2 967 189 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**

Control de memoria intermedia de fluctuación basado en monitorización de la fluctuación de retardo y la dinámica conversacional

5

**Referencia cruzada a solicitudes relacionadas**

Esta solicitud reivindica prioridad de la solicitud de patente china núm. 201410152754.9 (referencia: D13174CN01), presentada el 16 de abril de 2014, y de la solicitud de patente provisional de Estados Unidos núm. 61/989340 (referencia: D13174USP1), presentada el 6 de mayo de 2014.

10

Esta solicitud es una solicitud divisional europea de la solicitud de patente Euro-PCT EP15718706.3 (referencia: D13174EP01), presentada el 9 de abril de 2015.

**15 Campo técnico**

Esta divulgación se refiere al procesamiento de señales de audio. En particular, esta divulgación se relaciona con el procesamiento de señales de audio para telecomunicaciones, incluyendo, pero sin limitarse a, el procesamiento de señales de audio para videoconferencias o videoconferencias.

20

**Antecedentes**

La transmisión de voz a través de redes de paquetes está sujeta a variación de retardo, comúnmente conocida como fluctuación. La fluctuación puede medirse, por ejemplo, en términos de variación de tiempo entre llegadas (IAT) o variación de retardo de paquetes (PDV). La variación de IAT se puede medir de acuerdo con la diferencia de tiempo de recepción de paquetes adyacentes. La PDV puede, por ejemplo, medirse con referencia a intervalos de tiempo a partir de un dato o a un tiempo de recepción de paquetes "ancla". En las redes basadas en protocolo de Internet (IP), un retardo fijo puede atribuirse a retardos algorítmicos, de procesamiento y de propagación debidos al material y a la distancia, mientras que un retardo variable puede ser causado por la fluctuación del tráfico de red de IP, por diferentes rutas de transmisión a través de Internet, etc.

25

30

Los dispositivos VoIP (voz sobre protocolo de Internet) generalmente dependen de una "memoria intermedia de fluctuación" para contrarrestar el impacto negativo de la fluctuación. Al introducir un retardo adicional entre el tiempo en que se recibe un paquete de datos de audio y el tiempo en que se reproduce el paquete, una memoria intermedia de fluctuación tiene como objetivo transformar el flujo desigual de paquetes que llegan en un flujo regular de paquetes, de tal modo que las variaciones de retardo no causen degradación de la calidad del sonido perceptual para los usuarios finales. La comunicación por voz es altamente sensible a los retardos. De acuerdo con la recomendación G.114 de UIT, por ejemplo, el retardo unidireccional debe mantenerse por debajo de 150 ms para mantener una conversación normal, considerándose inaceptable por encima de 400 ms. Por lo tanto, el retardo adicional añadido por una memoria intermedia de fluctuación debe ser lo suficientemente pequeño como para evitar causar una degradación de la calidad del sonido perceptual. Desafortunadamente, una memoria intermedia de fluctuación pequeña conducirá a una pérdida de paquetes más frecuente cuando los paquetes lleguen más tarde de lo esperado debido a retardos en la red.

35

40

45

Ejemplos de memorias intermedias de fluctuación de la técnica anterior están en los documentos US 2008/0240004 A1 y US 2010/0265834 A1.

**Sumario**

50

La presente invención se establece en las reivindicaciones independientes adjuntas. Las realizaciones se establecen en las reivindicaciones dependientes.

De acuerdo con algunas implantaciones descritas en el presente documento, el método puede implicar la recepción de datos de audio. Los datos de audio pueden incluir paquetes de audio recibidos en tiempos reales de llegada de paquetes durante un intervalo de tiempo, que puede corresponder a un segmento de análisis de conversación. El segmento de análisis de conversación puede incluir una pluralidad de charlas. El método puede implicar analizar los datos de audio del segmento de análisis de conversación para determinar datos de dinámica de fluctuación de red y datos de interactividad conversacional. Los datos de dinámica de fluctuación de red pueden proporcionar una indicación de fluctuación en una red que retransmite los paquetes de datos de audio. Los datos de interactividad conversacional pueden proporcionar una indicación de interactividad entre participantes de una conversación representada por los datos de audio. El método puede implicar controlar el tamaño de memoria intermedia de fluctuación de acuerdo tanto con los datos de dinámica de fluctuación de red como con los datos de interactividad conversacional.

55

60

65

El análisis de los datos de audio para determinar los datos de dinámica de fluctuación de red puede implicar determinar al menos una variación de entre la variación de retardo de paquetes (PDV) y la variación de tiempo entre

llegadas (IAT) en base, al menos en parte, a los tiempos reales de llegada de paquetes. La determinación de la PDV puede implicar comparar los tiempos esperados de llegada de paquetes con los tiempos reales de llegada de paquetes.

5 De acuerdo con algunas implantaciones, el análisis de los datos de audio puede implicar determinar espectros de percentiles de tiempos de retardo de paquetes. La determinación de los datos de dinámica de fluctuación de red puede implicar la determinación de un espectro de entre percentiles de retardo de paquetes correspondiente a una diferencia entre un primer tiempo de retardo de paquetes de un primer espectro de percentiles y un segundo tiempo de retardo de paquetes de un segundo espectro de percentiles. En algunos ejemplos, analizar los datos de audio  
10 puede implicar determinar un espectro de tiempos de retardo de paquetes de acuerdo con las estadísticas de orden de la variación de retardo de paquetes. El espectro de tiempos de retardo de paquetes puede incluir tiempos de retardo de paquetes más cortos, tiempos de retardo de paquetes medios y tiempos de retardo de paquetes más largos. Determinar los datos de dinámica de fluctuación de red puede implicar determinar una diferencia entre uno de los tiempos de retardo de paquetes más grandes y uno de los tiempos de retardo de paquetes medianos. En algunas implantaciones, analizar los datos de audio para determinar los datos de dinámica de fluctuación de red  
15 puede implicar determinar una probabilidad de presencia de pico de retardo y/o una intensidad de pico de retardo.

En algunos ejemplos, analizar los datos de audio para determinar los datos de interactividad conversacional puede implicar determinar tiempos de intervención sencilla, durante los cuales puede estar hablando sólo un participante conversacional, tiempos de intervención doble, durante los cuales pueden estar hablando dos o más participantes conversacionales, y tiempos silenciosos compartidos, durante los cuales puede no estar hablando ninguno de los participantes conversacionales. Analizar los datos de audio para determinar los datos de interactividad conversacional puede implicar determinar al menos un ritmo de entre un ritmo de alternancia de hablantes y un ritmo de interrupción de hablantes.  
20

Algunos métodos pueden implicar recibir una indicación de silencio de hablante y/o una indicación de presentación. La determinación de los datos de interactividad conversacional puede implicar determinar la interactividad conversacional de acuerdo con al menos una indicación de entre la indicación de silencio de hablante o la indicación de presentación.  
25

En algunas implantaciones, analizar los datos de audio para determinar los datos de interactividad conversacional puede implicar determinar una medida de interactividad conversacional (CIM). La CIM puede, por ejemplo, basarse en reglas heurísticas y/o en entropía relativa conversacional.  
30

Por ejemplo, la CIM puede basarse, al menos en parte, en reglas heurísticas que implican la aplicación de un umbral para un ritmo de alternancia de hablantes, un umbral para tiempos de intervención sencilla, durante los cuales puede estar hablando un único participante conversacional solamente, un umbral para tiempos de intervención doble, durante los cuales pueden estar hablando dos o más participantes conversacionales, y/o un umbral para tiempos silenciosos compartido, durante los cuales puede no estar hablando ninguno de los participantes conversacionales.  
35

En algunas implantaciones, la CIM puede basarse al menos en parte en entropía relativa conversacional. La entropía relativa conversacional puede determinarse, al menos en parte, de acuerdo con las probabilidades de los estados conversacionales. Los estados conversacionales pueden incluir las probabilidades de tiempos de intervención sencilla, durante los cuales puede estar hablando sólo un participante conversacional, de tiempos de intervención doble, durante los cuales pueden estar hablando dos o más participantes conversacionales, y tiempos silenciosos compartido, durante los cuales puede no estar hablando ninguno de los participantes conversacionales.  
40

De acuerdo con algunas implantaciones, determinar los datos de interactividad conversacional puede implicar analizar la actividad conversacional de un solo participante conversacional. Por ejemplo, analizar la actividad conversacional del único participante conversacional puede implicar determinar si el único participante conversacional está interviniendo o no. Controlar el tamaño de memoria intermedia de fluctuación puede implicar establecer la memoria intermedia de fluctuación en un tamaño relativamente más pequeño cuando el único participante conversacional está interviniendo, y configurar la memoria intermedia de fluctuación en un tamaño relativamente mayor cuando el único participante conversacional no está interviniendo.  
45

En algunas implantaciones, controlar el tamaño de memoria intermedia de fluctuación puede implicar establecer una memoria intermedia de fluctuación en un tamaño relativamente mayor cuando los datos de dinámica de fluctuación de red indican más cantidad que la cantidad umbral de fluctuación de red. Por ejemplo, controlar el tamaño de memoria intermedia de fluctuación puede implicar configurar una memoria intermedia de fluctuación para un primer participante conversacional en un tamaño relativamente mayor cuando los datos de dinámica de fluctuación de red indican más cantidad que la cantidad umbral de fluctuación de red o cuando los datos de interactividad conversacional indican menos cantidad que la cantidad umbral de participación conversacional del primer participante conversacional.  
50

De acuerdo con algunas implantaciones, controlar el tamaño de memoria intermedia de fluctuación puede implicar configurar una memoria intermedia de fluctuación a un tamaño relativamente más pequeño cuando los datos de  
55

5 dinámica de fluctuación de red indican menos cantidad que la cantidad umbral de fluctuación de red o cuando los datos de interactividad conversacional indican al menos una cantidad umbral de interactividad conversacional. En algunos ejemplos, controlar el tamaño de memoria intermedia de fluctuación puede implicar configurar una memoria intermedia de fluctuación para un primer participante conversacional en un tamaño relativamente más pequeño cuando los datos de dinámica de fluctuación de red indican menos cantidad que la cantidad umbral de fluctuación de red o cuando los datos de interactividad conversacional indican al menos una cantidad umbral de participación conversacional por parte del primer participante conversacional. En algunos ejemplos, controlar el tamaño de memoria intermedia de fluctuación puede implicar asignar una ponderación relativamente menor a los datos de dinámica de fluctuación de red y asignar una ponderación relativamente mayor a los datos de interactividad conversacional.

15 De acuerdo con algunas implantaciones, controlar el tamaño de memoria intermedia de fluctuación puede implicar establecer un tamaño de memoria intermedia de fluctuación de acuerdo con uno de al menos tres modos de control de memoria intermedia de fluctuación. Por ejemplo, los modos de control de memoria intermedia de fluctuación pueden incluir un modo de pico, un modo de pérdida baja y un modo normal. Cada uno de los modos de control de memoria intermedia de fluctuación corresponde a un espectro de tamaños de memoria intermedia de fluctuación.

20 Al menos uno de los modos de control de memoria intermedia de fluctuación puede corresponder a datos de dinámica de fluctuación de red que indican al menos una cantidad umbral de fluctuación de red y datos de interactividad conversacional que indican al menos una cantidad umbral de interactividad conversacional. Al menos uno de los modos de control de memoria intermedia de fluctuación puede corresponder a datos de dinámica de fluctuación de red, que indican al menos una cantidad umbral de fluctuación de red, y datos de interactividad conversacional, que indican menos cantidad que la cantidad umbral de interactividad conversacional. Al menos uno de los modos de control de memoria intermedia de fluctuación puede corresponder a datos de dinámica de fluctuación de red, que indican menos cantidad que la cantidad umbral de fluctuación de red, y datos de interactividad conversacional, que indican al menos una cantidad umbral de interactividad conversacional. Al menos uno de los modos de control de memoria intermedia de fluctuación puede corresponder a datos de dinámica de fluctuación de red, que indican menos cantidad que la cantidad umbral de fluctuación de red, y datos de interactividad conversacional, que indican menos cantidad que la cantidad umbral de interactividad conversacional.

30 De acuerdo con algunas implantaciones, estos métodos y/u otros métodos divulgados en el presente documento pueden implantarse mediante uno o más medios no transitorios que tengan equipo lógico informático (software) almacenado en ellos. El software puede incluir instrucciones para controlar uno o más dispositivos para realizar tales métodos, al menos en parte.

35 Al menos algunos aspectos de la presente divulgación se pueden implantar mediante aparato. Por ejemplo, uno o más dispositivos pueden ser capaces de realizar, al menos en parte, los métodos divulgados en el presente documento. En algunas implantaciones, un aparato puede incluir un sistema de interfaz, un sistema de memoria que puede incluir una memoria intermedia de fluctuación y un sistema lógico. El sistema lógico puede ser capaz de recibir datos de audio mediante el sistema de interfaz. Los datos de audio pueden incluir paquetes de audio recibidos en tiempos reales de llegada de paquetes durante un intervalo de tiempo que puede corresponder a un segmento de análisis de conversación.

45 El sistema de interfaz puede incluir una interfaz de red, una interfaz entre el sistema lógico y el sistema de memoria y/o una interfaz de dispositivo externo. El sistema lógico puede incluir al menos un elemento de entre un procesador de un único o de múltiples microprocesadores de fines generales, un procesador de señales digitales (DSP), un circuito integrado de aplicación específica (ASIC), una matriz de puertas programables en campo (FPGA) u otro dispositivo lógico programable, lógica de puerta discreta o de transistor, o componentes discretos de equipo físico informático (hardware).

50 El sistema lógico puede ser capaz de analizar los datos de audio del segmento de análisis de conversación para determinar datos de dinámica de fluctuación de red y datos de interactividad conversacional. Los datos de dinámica de fluctuación de red pueden proporcionar una indicación de fluctuación en una red que retransmite los paquetes de datos de audio. Los datos de interactividad conversacional pueden proporcionar una indicación de interactividad entre participantes de una conversación representada por los datos de audio. El sistema lógico puede ser capaz de controlar un tamaño de memoria intermedia de fluctuación de acuerdo con los datos de dinámica de fluctuación de red y con los datos de interactividad conversacional. El intervalo de tiempo puede corresponder a un segmento de análisis de conversación que incluya una pluralidad de charlas.

60 En algunas implantaciones, el análisis de los datos de audio para determinar los datos de dinámica de fluctuación de red puede implicar determinar al menos una de entre la variación de retardo de paquetes (PDV) o la variación de tiempo entre llegadas (IAT) comparando los tiempos esperados de llegada de paquetes con los tiempos reales de llegada de paquetes. En algunos ejemplos, analizar los datos de audio para determinar los datos de dinámica de fluctuación de red puede implicar determinar al menos una probabilidad de presencia de pico de retardo o una intensidad de pico de retardo.

De acuerdo con algunas implantaciones, analizar los datos de audio para determinar los datos de interactividad conversacional puede implicar determinar tiempos de intervención sencilla, durante los cuales sólo puede estar hablando un único participante conversacional, tiempos de intervención doble, durante los cuales pueden estar hablando dos o más participantes conversacionales, y tiempos silenciosos compartido, durante los cuales puede no estar hablando ninguno de los participantes conversacionales. Analizar los datos de audio para determinar los datos de interactividad conversacional puede implicar determinar una medida de interactividad conversacional (CIM) en base a al menos una de las reglas heurísticas o a la entropía relativa conversacional.

Los detalles de una o más implantaciones del objeto descrito en esta memoria descriptiva se exponen en los dibujos que se acompañan y en la descripción siguiente. Otras características, aspectos y ventajas resultarán evidentes a partir de la descripción, los dibujos y las reivindicaciones. Obsérvese que las dimensiones relativas de las siguientes figuras pueden no estar dibujadas a escala.

**Breve descripción de los dibujos**

La figura 1A es un diagrama que ilustra esquemáticamente un ejemplo de un sistema de comunicación por voz en el que se pueden aplicar realizaciones de la solicitud.

La figura 1B es un diagrama que ilustra esquemáticamente otro ejemplo de un sistema de comunicación por voz en el que se pueden implantar aspectos de la solicitud.

La figura 2 es un diagrama de flujo que ilustra bloques de algunos métodos de control de memoria intermedia de fluctuación proporcionados en el presente documento.

La figura 3 proporciona un ejemplo de un modelo conversacional bipartito que proporciona algunos ejemplos de estados conversacionales.

La figura 4 es un diagrama de flujo que ilustra bloques de algunos métodos de control de memoria intermedia de fluctuación proporcionados en el presente documento.

La figura 5 es un diagrama de bloques que proporciona ejemplos de componentes de un aparato capaz de implantar diversos aspectos de esta divulgación.

La figura 6 es un diagrama de bloques que proporciona ejemplos de componentes de un aparato de procesamiento de audio.

Los números de referencia y designaciones similares en los diversos dibujos indican elementos similares.

**Descripción de realizaciones de ejemplo**

La siguiente descripción está dirigida a ciertas implantaciones con el fin de describir algunos aspectos innovadores de esta divulgación, así como ejemplos de contextos en los que se pueden implantar estos innovadores. Sin embargo, las enseñanzas contenidas en el presente documento se pueden aplicar de diversas maneras diferentes. Por ejemplo, aunque se describen diversas implantaciones en términos de ejemplos particulares de procesamiento de datos de audio, las enseñanzas del presente documento son ampliamente aplicables a otras implantaciones de procesamiento de datos de audio conocidas, así como a implantaciones de procesamiento de datos de audio que pueden introducirse en el futuro.

Lo que es más, las realizaciones descritas se pueden implantar en una variedad de hardware, software, soporte lógico inalterable (firmware), etc. Por ejemplo, los aspectos de la presente solicitud se pueden implantar en un sistema, en un dispositivo (por ejemplo, un teléfono celular, un teléfono portátil, reproductor multimedia portátil, un ordenador personal, un servidor, un descodificador de televisión, una grabadora de vídeo digital u otro reproductor multimedia), en un método o en un producto de programa de ordenador. En consecuencia, los aspectos de la presente solicitud pueden tomar la forma de una realización de hardware, de una realización de software (que incluye firmware, software residente, microcódigos, etc.) o de una realización que combine ambos aspectos de software y hardware. Tales realizaciones pueden denominarse "circuito", "módulo" o "sistema" en el presente documento. Además, aspectos de la presente solicitud pueden tomar la forma de un producto de programa de ordenador materializado en uno o más medios no transitorios que tengan un código de programa legible por ordenador materializado en el mismo. Tales medios no transitorios pueden incluir, por ejemplo, un disco duro, una memoria de acceso aleatorio (RAM), una memoria de sólo lectura (ROM), una memoria de sólo lectura programable y borrable (EPROM o memoria Flash), una memoria de sólo lectura de disco compacto portátil (CD-ROM), un dispositivo de almacenamiento óptico, un dispositivo de almacenamiento magnético o cualquier combinación adecuada de los anteriores. En consecuencia, las enseñanzas de esta divulgación no pretenden limitarse a las implantaciones mostradas en las figuras y/o descritas en el presente documento, sino que tienen, en cambio, una amplia aplicabilidad.

Los términos "bloque" y "paquete" se utilizan en el presente documento como sinónimos. En consecuencia, un "bloque de audio" o un "bloque de datos de audio" tendrá el mismo significado que un "paquete de audio" o un "paquete de datos de audio".

5 Como se usa en el presente documento, el término "memoria intermedia" puede referirse a una región de un dispositivo de memoria física utilizada para almacenar datos temporalmente, o a una memoria intermedia de datos lógica o virtual que "apunte" a una ubicación en una memoria física. Un "memoria intermedia de fluctuación" se referirá, en general, a una memoria intermedia lógica o física para almacenar tramas de audio recibidas. Aunque se utilizará generalmente una memoria intermedia de fluctuación para almacenar temporalmente datos de audio  
 10 codificados antes de un proceso de decodificación, la memoria intermedia de fluctuación puede almacenar diversas formas de paquetes de audio o tramas de audio, dependiendo de la implantación específica. Por lo tanto, en toda la memoria descriptiva, se interpretará que el término "memoria intermedia de fluctuación" incluye tanto una memoria intermedia de fluctuación que realmente almacene (o apunte a) tramas de audio, como una memoria intermedia de fluctuación que realmente almacene (o apunte a) diversas formas de paquetes (bloques) que posteriormente se  
 15 decodificará en tramas de audio antes de reproducirse o de alimentarse en componentes para su posterior procesamiento. Es posible que el proceso de decodificación no siempre se analice explícitamente en conexión con los procesos de almacenamiento en memoria intermedia, aunque la decodificación se realizará generalmente antes de la reproducción o "playback" de los datos de audio. En consecuencia, el término "trama" según se utiliza en el presente documento debe interpretarse extensamente como que incluye una trama ya decodificada de un paquete,  
 20 una trama todavía codificada en un paquete, un paquete en sí que incluye una o más tramas, o más de una trama codificada en un paquete o ya decodificadas del paquete. En otras palabras, en el contexto de la presente solicitud, el procesamiento que implica a una trama también puede interpretarse como un procesamiento que implica a un paquete, o como un procesamiento que implica simultáneamente a más de una trama contenida en un paquete.

25 En el contexto de la presente solicitud, el significado de la expresión "al mismo tiempo" (o similar) incluye, pero sin limitarse a, el significado literal exacto, y se interpretará como "dentro del mismo lapso/intervalo de tiempo de una granularidad predefinida". En la presente solicitud, por ejemplo, la granularidad predefinida puede ser el lapso de tiempo entre dos tramas/paquetes enviadas/os consecutivamente (tal lapso de tiempo puede denominarse "lapso de trama"), o el ritmo de sondeo de la red para comprobar la llegada de paquetes, o la granularidad de tiempo de  
 30 procesamiento, pero sin limitarse a éstos. Por ejemplo, se puede cuantificar el tiempo de llegada por duración/lapso de trama, por ejemplo, 20 ms. Alternativa, o adicionalmente, se puede representar el tiempo como correspondiente a un número de paquete entero. De manera similar, en el contexto de la presente solicitud, cuando implica a un punto de tiempo específico, dependiendo del contexto, puede también significar un lapso de tiempo de la granularidad predefinida. Además, cuando implica un tiempo específico  $t_i$  ( $i$  es un número entero) para una determinada trama, que se entenderá como un punto de tiempo, asumiendo para mayor claridad que indica el punto de tiempo en el que  
 35 comienza la trama en el lado del emisor, o que indica el punto de tiempo en el que comienza la recepción de la trama en el lado del receptor.

40 Como se mencionó anteriormente, una memoria intermedia de fluctuación puede contrarrestar el impacto negativo de la inestabilidad de red almacenando temporalmente paquetes recibidos de datos de audio (denominados también en el presente documento "paquetes de audio"), que pueden corresponder a datos de voz. En algunas implantaciones, la memoria intermedia de fluctuación puede almacenar paquetes de audio antes de que los paquetes de audio se proporcionen a un decodificador y, posteriormente, se reproduzcan, por ejemplo, mediante altavoces de un terminal de comunicación.

45 Determinar un tamaño de memoria intermedia de fluctuación adecuado puede suponer un desafío. Una memoria intermedia de fluctuación demasiado pequeña puede hacer que se pierda una cantidad inaceptable de paquetes de audio, particularmente durante tiempos de picos de retardo que pueden haberse originado, por ejemplo, por un aumento en la actividad de red. Sin embargo, una memoria intermedia de fluctuación demasiado larga puede  
 50 provocar retardos en la percepción. Ciertas conversaciones, tal como las conversaciones interactivas entre dos o más personas, pueden requerir un retardo relativamente menor con el fin de evitar molestar a los participantes de una conversación. Sin embargo, los participantes de otras conversaciones, tal como de una presentación, que es predominantemente una conversación unidireccional, son generalmente más tolerantes a un retardo inicial.

55 En consecuencia, en algunas implantaciones descritas en el presente documento, se puede controlar el tamaño de memoria intermedia de fluctuación con el fin de tratar estos dos tipos de conversaciones de manera diferente. Algunas de tales implantaciones proporcionan una gestión de memoria intermedia de fluctuación advertida del contexto capaz de controlar dinámicamente el tamaño de memoria intermedia de fluctuación en función de los datos de dinámica de fluctuación de red correspondientes al contexto de fluctuación de red a largo plazo.

60 La figura 1A es un diagrama que ilustra esquemáticamente un ejemplo de un sistema de comunicación por voz en el que se pueden aplicar realizaciones de la solicitud. Como se ilustra en la figura 1A, el participante conversacional A hace funcionar un terminal A de comunicación, y el participante conversacional B hace funcionar un terminal B de comunicación. Al igual que otros terminales de comunicación mostrados y descritos en el presente documento, los  
 65 terminales A y B de comunicación pueden, por ejemplo, incluir teléfonos, ordenadores personales, dispositivos electrónicos móviles (por ejemplo, teléfonos celulares, teléfonos inteligentes, tabletas, etc.) o similares. Los

terminales A y B de comunicación pueden incluir componentes tales como los que se describen más adelante (por ejemplo, con referencia a las figuras 5 y 6).

5 Durante una sesión de comunicación por voz, tal como una teleconferencia, el participante conversacional A y el participante conversacional B pueden participar entre sí mediante sus terminales A y B de comunicación. En este ejemplo, los terminales A y B de comunicación son capaces de comunicarse mediante un enlace 103 de datos. El enlace 103 de datos puede implantarse como una conexión punto a punto, como una red de comunicación, etc.

10 En este ejemplo, los terminales A y B de comunicación son capaces de realizar VAD (detección de actividad de voz) en bloques de audio de señales de audio capturadas por uno o más micrófonos. Si se detecta una presencia de voz en un bloque de audio, se puede realizar el procesamiento correspondiente (por ejemplo, aplicar una ganancia adecuada para datos de voz) en el bloque de audio mediante un sistema lógico del terminal de comunicación, y el bloque de audio se puede transmitir al terminal de comunicación de otro participante conversacional mediante el enlace 103 de datos. Si no se decide la presencia de voz en un bloque de audio, puede realizarse el procesamiento correspondiente (por ejemplo, aplicando una ganancia adecuada para datos que no son de voz) en el bloque de audio mediante un sistema lógico del terminal de comunicación, y el bloque de audio puede transmitirse al terminal de comunicación de otro participante conversacional mediante el enlace 103 de datos.

20 En este ejemplo, los terminales A y B de comunicación son capaces de indicar un tiempo silencioso al terminal de comunicación del otro participante conversacional. Como se utiliza en el presente documento, un "tiempo silencioso" es un tiempo durante el cual un participante de una conversación no habla. Durante un "tiempo silencioso", el terminal de comunicación de un participante conversacional puede detectar sonidos ajenos al habla, tal como el ruido ambiental. Los datos de audio correspondientes a tales sonidos pueden procesarse y transmitirse a uno o más terminales de comunicación. En algunas implantaciones, un tiempo silencioso puede indicarse transmitiendo metadatos de tiempo silencioso (por ejemplo, estableciendo un indicador o bit correspondiente), mientras que, en otras implantaciones, un tiempo silencioso puede indicarse no transmitiendo nada durante el período de tiempo correspondiente a un bloque de audio. En algunas implantaciones, los metadatos de tiempo silencioso pueden corresponder a la activación, por parte de un participante conversacional, de un control de "silencio" de un terminal de comunicación.

30 En esta implantación, los terminales A y B de comunicación son capaces de establecer y controlar memorias intermedias de fluctuación, que se representan como "JB" en las figuras 1A y 1B. Aquí, los terminales A y B de comunicación son capaces de recibir datos de audio codificados, por ejemplo, como bloques de audio, y almacenarlos como entradas en una memoria intermedia de fluctuación. Las entradas pueden corresponder a un tiempo, por ejemplo, a un tiempo en el que se reciben los bloques de audio. Los bloques de audio en la memoria intermedia de fluctuación pueden decodificarse y/o, de otro modo, procesarse para su reproducción por uno o más altavoces del terminal de comunicación. La recepción de metadatos de tiempo silencioso o de nada puede originar entradas vacías correspondientes en la memoria intermedia de fluctuación. Los terminales A y B de comunicación pueden ser capaces de controlar un tamaño de memoria intermedia de fluctuación como se muestra y describe en el presente documento.

45 La figura 1B es un diagrama que ilustra esquemáticamente otro ejemplo de un sistema de comunicación por voz en el que se pueden implantar aspectos de la solicitud. En este ejemplo, se puede realizar una conferencia por voz entre los participantes conversacionales A, B y C.

50 Como se ilustra en la figura 1B, el participante conversacional A hace funcionar un terminal A de comunicación, el participante conversacional B hace funcionar un terminal B de comunicación, y el participante conversacional C hace funcionar un terminal C de comunicación. Durante una sesión de conferencia por voz, el participante conversacional A, el participante conversacional B, y el participante conversacional C pueden hablar entre sí a través de sus terminales A, B y C de comunicación, respectivamente. Los terminales de comunicación ilustrados en la figura 1B pueden ser capaces de proporcionar esencialmente la misma funcionalidad que los ilustrados en la figura 1A, desde la perspectiva de los participantes conversacionales A, B y C. Aunque en la figura 1B se ilustran tres terminales de comunicación, otras implantaciones pueden implicar más o menos terminales de comunicación.

55 Sin embargo, en el ejemplo mostrado en la figura 1B, los terminales A, B y C de comunicación están configurados para comunicarse con otro dispositivo, que es un servidor en este ejemplo, a través de un enlace 113 de datos común o de enlaces 113 de datos separados. El enlace 113 de datos puede implantarse como una conexión punto a punto o como una red de comunicación. Los terminales A, B y C de comunicación pueden ser capaces de realizar VAD y procesamiento apropiado en bloques de audio de la señal de audio capturada por el terminal de comunicación, por ejemplo, como se describió anteriormente.

60 En esta implantación, los terminales A, B y C de comunicación son capaces de indicar un tiempo silencioso al servidor. En algunas implantaciones, un tiempo silencioso puede indicarse transmitiendo metadatos de tiempo silencioso (por ejemplo, estableciendo un indicador o bit correspondiente), mientras que en otras implantaciones un tiempo silencioso puede indicarse no transmitiendo nada durante el período de tiempo correspondiente a un bloque de audio. Los terminales A, B y C de comunicación pueden ser capaces de incluir una "marca de tiempo" o

metadatos de tiempo similares con un paquete de audio transmitido, que indique el tiempo de transmisión del paquete de audio.

5 En esta implantación, el servidor también es capaz de establecer y controlar memorias intermedias de fluctuación. En el ejemplo mostrado en la figura 1B, el servidor ha establecido memorias intermedias  $J_{B_A}$ ,  $J_{B_B}$  y  $J_{B_C}$  de fluctuación, correspondientes a cada uno de los terminales A, B y C de comunicación. Por ejemplo, el servidor puede ser capaz de controlar un tamaño de memoria intermedia de fluctuación como se describe en el presente documento. En esta implantación, el servidor es capaz de recibir los bloques de audio transmitidos por los terminales A, B y C de comunicación y de almacenarlos en entradas en las memorias intermedias de fluctuación  $J_{B_A}$ ,  $J_{B_B}$  y  $J_{B_C}$  correspondientes a los tiempos de los bloques de audio. Por ejemplo, el servidor puede ser capaz de almacenar los bloques de audio en entradas en las memorias intermedias de fluctuación  $J_{B_A}$ ,  $J_{B_B}$  y  $J_{B_C}$  correspondientes a las marcas de tiempo de los bloques de audio. La recepción de los metadatos del tiempo silencioso (o de nada) puede ocasionar las correspondientes entradas vacías en las memorias intermedias de fluctuación.

15 En este ejemplo, el servidor es capaz también de mezclar bloques de audio, correspondientes al mismo tiempo, de cada una de las memorias intermedias  $J_{B_A}$ ,  $J_{B_B}$  y  $J_{B_C}$  de fluctuación, en un bloque de audio mixto. Se pueden transmitir copias de los bloques de audio mixto a cada uno de los terminales A, B y C de comunicación. El servidor puede incluir uno o más tipos de marcas de tiempo con cada uno de los bloques de audio mixto. En algunos casos, por ejemplo, los terminales A y B de comunicación pueden estar enviando paquetes de audio al servidor. El servidor puede mezclar los paquetes de audio (ya sea en el dominio tiempo o en el dominio frecuencia) antes de enviar los paquetes mixtos al terminal C de comunicación. Que el servidor realice tal mezcla puede depender de diversos factores, tal como el ancho de banda, si el servidor está configurado para mezcla, si se desea mono o multicanal para el terminal C de comunicación, etc.

25 Los terminales A, B y C de comunicación son capaces de establecer y controlar memorias intermedias  $J_B$  de fluctuación en este ejemplo. Los terminales A, B y C de comunicación pueden recibir los bloques de audio mixtos desde el servidor y pueden almacenarlos en entradas de memoria intermedia de fluctuación correspondientes a los tiempos de los bloques de audio mixtos. Por ejemplo, las entradas de la memoria intermedia de fluctuación pueden corresponder a un tiempo en el que se reciben los bloques de audio. En cada terminal de comunicación, los bloques de audio en la memoria intermedia de fluctuación pueden ser decodificados y reproducidos por un sistema de altavoces del terminal de comunicación. Los terminales A, B y C de comunicación pueden ser capaces de controlar un tamaño de memoria intermedia de fluctuación como se divulga en el presente documento.

35 Se puede utilizar una memoria intermedia de fluctuación para transformar un flujo desigual de paquetes de audio de llegada en un flujo regular de paquetes de audio, de tal modo que las variaciones de retardo no originen una degradación de calidad perceptual a los participantes conversacionales. La determinación de un nivel adecuado de memoria intermedia de fluctuación implica generalmente un equilibrio entre el retardo promedio de memoria intermedia y el ritmo de pérdida de paquetes. Se han propuesto muchos algoritmos de gestión de memoria intermedia de fluctuación (JBM) basados en estadísticas, tales como los algoritmos JBM basados en la variación de retardo entre paquetes (IPDV) divulgados en la publicación de patente de Estados Unidos núm. 2009/0003369 y el algoritmo JBM basado en histograma divulgado en S. B. Moon, J. Kurose y D. Towsley, Packet Audio Playout Delay Adjustment: Performance Bounds and Algorithms, en *Multimedia Systems* (1998) 6:17-28.

45 Sin embargo, las implantaciones aquí descritas proporcionan métodos alternativos de control de memoria intermedia de fluctuación. Algunas implantaciones descritas en el presente documento implican el análisis de datos de audio para determinar datos de dinámica de fluctuación de red y datos de interactividad conversacional para el control de memoria intermedia de fluctuación advertido del contexto.

50 La figura 2 es un diagrama de flujo que ilustra bloques de algunos métodos de control de memoria intermedia de fluctuación proporcionados en el presente documento. El método 200 puede, por ejemplo, ser realizado (al menos en parte) por un servidor o por otro dispositivo tal que esté configurado para comunicarse con terminales de comunicación, tal como se describe anteriormente con referencia a la figura 1B. Sin embargo, algunos métodos proporcionados en el presente documento pueden realizarse (al menos en parte) por un terminal de comunicación. Al igual que con otros métodos descritos en el presente documento, los bloques del método 200 no se realizan necesariamente en el orden indicado. Lo que es más, algunas implantaciones del método 200 (y otros métodos divulgados en el presente documento) pueden incluir más o menos bloques de los indicados o descritos.

60 En este ejemplo, el método 200 comienza con el bloque 205, que implica recibir datos de audio durante un intervalo de tiempo que corresponde a un segmento de análisis de conversación. En este ejemplo, los datos de audio incluyen paquetes de audio recibidos en los tiempos reales de llegada de paquetes. El intervalo de tiempo puede ser, por ejemplo, un intervalo de tiempo de largo plazo o de corto plazo que incluye una pluralidad de charlas. Según se utiliza en el presente documento, el término "charla" corresponde a un segmento continuo (o substancialmente continuo) de habla entre "tiempos silenciosos compartidos" de una conversación. Aunque los paquetes de audio correspondientes a los tiempos silenciosos compartidos pueden incluir ruido de fondo, etc., el término "tiempo silencioso compartido" se utiliza en el presente documento para referirse a un tiempo durante el cual ningún participante conversacional está hablando. En algunas implantaciones, la longitud de un paquete o trama puede ser

del orden de decenas de milisegundos (por ejemplo, de 20 ms) y un segmento de análisis de conversación puede ser del orden de decenas de segundos, por ejemplo, de 20 segundos.

5 Aquí, el bloque 210 implica analizar los datos de audio del segmento de análisis de conversación para determinar datos de dinámica de fluctuación de red y datos de interactividad conversacional. Los datos de dinámica de fluctuación de red pueden, por ejemplo, proporcionar una indicación de fluctuación en una red que retransmite los paquetes de datos de audio. Los datos de interactividad conversacional pueden proporcionar una indicación de interactividad entre participantes de una conversación representada por los datos de audio. En esta implantación, el bloque 215 implica controlar un tamaño de memoria intermedia de fluctuación de acuerdo tanto con los datos de 10 dinámica de fluctuación de red como con los datos de interactividad conversacional.

El análisis de los datos de audio para determinar los datos de dinámica de fluctuación de red puede implicar determinar al menos una variación de entre la variación de retardo de paquetes (PDV) y la variación de tiempo entre llegadas (IAT) en base, al menos en parte, a los tiempos reales de llegada de paquetes. Los datos de dinámica de fluctuación de red pueden, por ejemplo, incluir un espectro de entre percentiles de retardo de paquetes, una 15 probabilidad de pico de retardo y/o una densidad de pico de retardo. Por ejemplo, en algunas implantaciones, la PDV para un número determinado de paquetes puede determinarse como la diferencia entre el tiempo de llegada real del paquete y el tiempo esperado de llegada del paquete.

20 Considérese, por ejemplo, un segmento de análisis de conversación que consta de M charlas y N paquetes de audio. En esta discusión, las siguientes variables representan las cantidades correspondientes que se establecen a continuación:

25  $t_k^i$  representa la marca de tiempo de emisor del  $i$ -ésimo paquete en la  $k$ -ésima charla;

$r_k^i$  representa la marca de tiempo de receptor del  $i$ -ésimo paquete en la  $k$ -ésima charla;

$n_k$ : número de paquetes recibidos en la  $k$ -ésima charla;

30  $p_k^i$  representa la marca de tiempo de playback del  $i$ -ésimo paquete en la  $k$ -ésima charla; y

$p_k^i - t_k^i$  representa el retardo de playback.

35 En algunas implantaciones, se puede realizar una indicación de llegada tardía de acuerdo con si la marca de tiempo de playback indica que el tiempo de reproducción/ playback del paquete de audio fue más temprano al tiempo de llegada del paquete de audio. Por ejemplo, una indicación de llegada tardía de un paquete de audio se indicará en forma binaria, de tal modo que una indicación de llegada tardía sea uno o cero, por ejemplo, como sigue:

$$I_k^i = \begin{cases} 1, & \text{si } p_k^i < r_k^i \\ 0, & \text{de otro modo} \end{cases} \quad (\text{Ecuación 1})$$

40 En la ecuación 1,  $I_k^i$  representa una indicación de llegada tardía para el  $i$ -ésimo paquete en la  $k$ -ésima charla. Sin embargo, en implantaciones alternativas, se puede utilizar una indicación de llegada tardía más granular. Por ejemplo, una indicación de llegada tardía puede indicar un grado de retraso de acuerdo con dos o más umbrales de tiempo.

45 En algunas de tales implantaciones, se puede determinar una variación de retardo de paquete de acuerdo con una ventana de tiempo que incluye los  $w$  paquetes de audio pasados que se han recibido, por ejemplo, de la siguiente manera:

$$50 \Delta d_n^{i-w+1}, \Delta d_n^{i-w+2}, \dots, \Delta d_n^i. \quad (\text{Ecuación 2})$$

En la ecuación 2,  $\Delta d_n^i$  representa el retardo de paquete del  $i$ -ésimo paquete de audio,  $\Delta d_n^{i-w+1}$  representa el retardo de paquete de un paquete de audio recibido  $(w+1)$  paquetes antes del  $i$ -ésimo paquete de audio,  $\Delta d_n^{i-w+2}$  representa el retardo de paquete del siguiente paquete de audio recibido, etc. El retardo del paquete puede determinarse, por ejemplo, de acuerdo con la PDV o con el IAT. El valor de  $w$  también puede denominarse en el presente documento "longitud de la ventana de análisis". La elección de  $w$  determina cómo de rápido puede adaptarse un algoritmo correspondiente a la variación de retardo. En consecuencia, el valor de  $w$  está sujeto a un equilibrio entre precisión y capacidad de respuesta. En algunas implantaciones,  $w$  puede estar en el espectro de 200

a 1000.

De acuerdo con algunas implantaciones, los paquetes de audio recibidos se pueden clasificar en un espectro de tiempos de retardo de paquetes, por ejemplo, de acuerdo con las estadísticas de orden de la variación de retardo de paquetes. Por ejemplo, las estadísticas de orden para las variaciones de retardo de paquetes mencionadas anteriormente se pueden representar como  $\Delta D^1, \Delta D^2, \dots, \Delta D^w$ , en donde  $\Delta D^1 \leq \Delta D^2 \leq \dots \leq \Delta D^w$ . En este ejemplo,  $\Delta D^1$  representa el retardo más pequeño y  $\Delta D^w$  representa el retardo más grande.

En consecuencia, tal espectro puede incluir tiempos de retardo de paquetes más cortos, tiempos de retardo de paquetes medios y tiempos de retardo de paquetes más largos. Determinar los datos de dinámica de fluctuación de red puede implicar determinar una diferencia entre los tiempos de retardo, por ejemplo, una diferencia entre uno de los tiempos de retardo de paquetes más grandes y uno de los tiempos de retardo de paquetes medios.

Algunas de tales implantaciones implican determinar espectros percentiles de tiempos de retardo de paquetes. En algunas de tales implantaciones, determinar los datos de dinámica de fluctuación de red puede implicar determinar un espectro de entre percentiles de retardo de paquetes correspondiente a una diferencia entre un tiempo de retardo de primer paquete de un primer espectro de percentiles (por ejemplo, para uno de los tiempos de retardo de paquetes más grandes) y un segundo tiempo de retardo de paquete de un segundo espectro de percentiles (por ejemplo, para uno de los tiempos de retardo de paquete medianos).

En algunas implantaciones, un espectro de entre percentiles se puede determinar de la siguiente manera:

$$\Delta D^r - \Delta D^k \quad (\text{Ecuación 3})$$

En algunas implantaciones,  $r$  puede seleccionarse de tal modo que  $\Delta D^r$  represente uno de los tiempos de retardo de paquetes más grandes, mientras que  $k$  puede seleccionarse de tal modo que  $\Delta D^k$  represente un tiempo de retardo en o cerca de la media de los tiempos de retardo de paquetes. En un ejemplo, el redondeo  $k$  se puede determinar de la siguiente manera:

$$r = \text{redondear}(0,995 \times w) \text{ y } k = \text{redondear}(0,5 \times w) \quad (\text{Ecuaciones 4 y 5})$$

En las Ecuaciones 4 y 5, "redondear" representa un proceso de redondeo a un número entero. En algunas implantaciones, los valores de redondeo  $k$  pueden determinarse de acuerdo con datos empíricos basados en experimentos que implican paquetes de audio transmitidos en diferentes tipos de redes reales.

La determinación de los datos de dinámica de fluctuación de red puede implicar determinar una intensidad de pico de retardo, por ejemplo, en base al número de paquetes que tuvieron un retardo más allá de un umbral, y la cantidad de retardo para esos paquetes, y/o determinar una probabilidad de presencia de pico de retardo. Por ejemplo, en algunas implantaciones, la probabilidad de presencia de un pico de retardo se puede determinar de la siguiente manera:

$$p = \frac{\sum_i I_n^i}{w} \quad (\text{Ecuación 6})$$

En la ecuación 6,  $p$  representa una probabilidad de presencia de pico de retardo y  $I_n^i$  representa un indicador de pico de retardo. En algunos ejemplos, el indicador de pico de retardo de la ecuación 6 se puede determinar de la siguiente manera:

$$I_n^i = \begin{cases} 0, & \text{si } \Delta d_n^i < \xi_m \\ 1, & \text{de otro modo} \end{cases} \quad (\text{Ecuación 7})$$

En la ecuación 7,  $\xi_m$  representa un umbral de pico de retardo. En algunas implantaciones, el umbral de pico de retardo puede estar en el espectro de 5 a 20 intervalos de paquetes, por ejemplo, 10 intervalos de paquetes. Por ejemplo, si un intervalo de tiempo esperado entre paquetes es de 20 ms, un umbral de pico de retardo de 10 correspondería a 200 ms.

Algunas implantaciones pueden implicar determinar la intensidad promedio del pico de retardo durante un intervalo de tiempo. En algunos ejemplos, la intensidad promedio del pico de retardo se puede determinar de la siguiente manera:

$$\lambda = \frac{\sum_{i=1}^w I_n^i \Delta d_i^i}{\sum_{i=1}^w I_n^i} \quad (\text{Ecuación 8})$$

En la ecuación 8,  $\lambda$  representa la intensidad promedio de pico de retardo para una longitud de ventana de análisis de  $w$  paquetes de audio.

5 Algunas implantaciones pueden implicar la combinación de más de un tipo de datos de dinámica de fluctuación de red. Por ejemplo, algunas implantaciones de detección de modo de pico (PMD) pueden implicar combinar más de un tipo de datos de dinámica de fluctuación de red para detectar un modo de pico de fluctuación de red. En algunos ejemplos, la PMD implicará una detección de modo de pico "a largo plazo" durante un segmento de análisis de conversación que incluye una pluralidad de charlas.

En algunos ejemplos, la PMD puede basarse, al menos en parte, en cálculos de espectro de entre percentiles (IPR) (por ejemplo, como se describió anteriormente), en intensidad de pico de retardo y/o en probabilidad de presencia de pico de retardo. En un ejemplo, la PMD se determina de la siguiente manera:

$$PMD = f(IPR, p, \lambda) = \begin{cases} 1, & \text{si } (IPR > IPR\_th \ \&\& \ p > p\_th) \ || \ (IPR > IPR\_th \ \&\& \ \lambda > \lambda\_th) \\ 0, & \text{de otro modo} \end{cases}$$

(Ecuación 9)

En la ecuación 9,  $IPR\_th$  representa un umbral del espectro de entre percentiles de estadísticas de orden,  $p\_th$  representa un umbral de probabilidad de presencia de picos de retardo y  $\lambda\_th$  representa un umbral de intensidad promedio de picos de retardo. En algunas implantaciones,  $IPR\_th$  puede estar en el espectro de 5-9, por ejemplo, 7. En algunos ejemplos,  $p\_th$  puede estar en el espectro de 0,03 a 0,07, por ejemplo, 0,05, y  $\lambda\_th$  puede estar en el espectro de 10-20, por ejemplo, 15. En consecuencia, en el ejemplo mostrado en la ecuación 9, un proceso de detección de modo de pico producirá un valor binario que corresponde a una determinación de "sí o no". Sin embargo, en implantaciones alternativas, se pueden determinar valores de detección de modo de pico relativamente más granulares. Por ejemplo, en algunas implantaciones, un espectro de dos o más valores de PMD puede corresponder a distintos grados de fluctuación de red.

Como se señaló anteriormente, diversas implantaciones descritas en el presente documento implican controlar un tamaño de memoria intermedia de fluctuación de acuerdo tanto con los datos de dinámica de fluctuación de red como con los datos de interactividad conversacional. El análisis anterior se ha centrado principalmente en la determinación de diversos tipos de datos de dinámica de fluctuación de red. A continuación se muestran diversos ejemplos de determinar datos de interactividad conversacional.

En algunas implantaciones, determinar datos de interactividad conversacional puede implicar determinar estados conversacionales. La figura 3 proporciona un ejemplo de un modelo conversacional bipartito que proporciona algunos ejemplos de estados conversacionales.

En la figura 3, las áreas sombreadas sobre el eje t de tiempo horizontal indican tiempos durante los cuales el participante conversacional "A" está interviniendo, mientras que las áreas sombreadas por debajo del eje de tiempo indican tiempos durante los cuales el participante conversacional "B" está interviniendo. Las áreas etiquetadas como estado A y estado B indican tiempos de "intervención sencilla" durante los cuales habla un único participante conversacional.

El área etiquetada como estado D indica un tiempo de "intervención doble" durante el cual el participante conversacional A y el participante conversacional B están hablando. En algunos casos, una conversación puede incluir tres o más participantes. En consecuencia, según se utiliza en el presente documento, un tiempo de "intervención doble" significa un tiempo durante el cual al menos dos participantes conversacionales están hablando.

El estado M de la figura 3 corresponde a un tiempo silencioso compartido durante el cual ni el participante A ni el participante B están hablando. En vista del hecho de que una conversación puede incluir a veces tres o más participantes, según se utiliza en el presente documento, un "tiempo silencioso compartido" significa un tiempo durante el cual ningún participante conversacional habla.

Sin embargo, determinar datos de interactividad conversacional puede implicar determinar otros tipos de estados conversacionales. En algunas implantaciones, determinar los datos de interactividad conversacional puede implicar determinar al menos un ritmo de entre el ritmo de alternancia de hablantes y el ritmo de interrupción de hablantes. Una interrupción exitosa del participante conversacional "A" por parte del participante conversacional "B" puede, por

ejemplo, inferirse mediante la secuencia estado A/estado D/estado B. Una interrupción fallida del participante conversacional "A" puede, por ejemplo, determinarse por una secuencia estado A/estado D/estado A. Algunas implantaciones pueden implicar enviar o recibir una indicación de silencio de hablante o una indicación de presentación, por ejemplo, como metadatos asociados con los paquetes de audio. Por ejemplo, una indicación de silencio de hablante o una indicación de presentación pueden corresponder a la recepción de la entrada de un participante conversacional, tal como presionar un botón de "silencio", introducir un código correspondiente a una indicación de silencio de hablante o una indicación de presentación (por ejemplo, mediante un teclado o un panel de teclas), etc. La determinación de los datos de interactividad conversacional puede basarse, al menos en parte, en la indicación de silencio de hablante y/o en la indicación de presentación.

En algunas implantaciones, analizar los datos de audio para determinar los datos de interactividad conversacional puede implicar determinar una medida de interactividad conversacional (CIM) que se corresponde con un grado de interactividad conversacional. En algunas implantaciones, la CIM puede basarse en reglas heurísticas y/o en entropía relativa conversacional.

En algunos ejemplos, la CIM puede basarse, al menos en parte, en reglas heurísticas que implican la aplicación de al menos un umbral de entre un umbral para un ritmo de alternancia de hablantes, un umbral para tiempos de intervención sencilla, durante los cuales sólo está hablando un único participante conversacional, un umbral para tiempos de intervención doble, durante los cuales están hablando dos o más participantes conversacionales, o un umbral para tiempos silenciosos compartidos, durante los cuales ningún participante conversacional está hablando. Para reglas heurísticas, el grado de interactividad conversacional puede basarse en un número de paquetes correspondientes a tiempos de intervención sencilla, un número de paquetes correspondientes a tiempos silenciosos compartidos, un número de paquetes correspondientes a tiempos de intervención doble y/o un número de veces en que los hablantes cambian (lo que puede denominarse en el presente documento "alternancia de hablantes") dentro de un segmento de análisis de conversación que incluye una pluralidad de charlas.

En uno de tales ejemplos, una CIM puede basarse en reglas heurísticas de la siguiente manera:

$$CIM(k) = \begin{cases} 0, & \left( \frac{\lambda_1(k)}{\lambda_2(k)} \in [0, 1/umbral\_ST] \cup [umbral\_ST, +\infty) \right) \parallel (\lambda_{MS}(k) \gg umbral\_MS) \parallel (SAR(k) \ll umbral\_SAR) \\ 1, & \text{de otro modo} \end{cases}$$

(Ecuación 10)

En el ejemplo de la ecuación 10,  $CIM(k)$  representa una medida de interactividad conversacional en un  $k$ -ésimo segmento de análisis de conversación,  $umbral\_ST$  representa un umbral para tiempos de intervención sencilla,  $umbral\_MS$  representa un umbral para tiempos silenciosos compartidos, y  $umbral\_SAR$  representa un umbral para alternancia de hablantes. En algunas implantaciones,  $umbral\_ST$  puede estar en el espectro de 8 a 12, por ejemplo, 10. En algunos ejemplos,  $umbral\_MS$  puede estar en el espectro de 0,7 a 0,9 (por ejemplo, 0,8) y  $umbral\_SAR$  puede estar en el espectro de 0,003 a 0,007 (por ejemplo, 0,005).

En el ejemplo mostrado en la ecuación 10, una medida de interactividad conversacional será un valor binario que puede imaginarse correspondiente a una determinación "interactiva" o "no interactiva". Sin embargo, en implantaciones alternativas, se pueden determinar valores relativamente más granulares de interactividad conversacional. Por ejemplo, en algunas implantaciones, un espectro de dos o más valores de CIM puede corresponder a grados que varían de interactividad conversacional.

En algunas implantaciones en las que la CIM se basa al menos en parte en entropía relativa conversacional, la entropía relativa conversacional puede basarse, al menos en parte, en probabilidades de estados conversacionales. En algunos ejemplos, los estados conversacionales pueden corresponderse con las probabilidades de tiempos de intervención sencilla, tiempos de intervención doble y tiempos silenciosos compartido. En uno de tales ejemplos, la entropía relativa de una conversación que implica al participante conversacional A y al participante conversacional B se puede determinar de la siguiente manera:

$$E(k) = \sum_{i \in \{A, B, M, D\}} -P_i(k) \log_2 \left( \frac{P_i(k)}{Q_i(k)} \right) \quad \text{(Ecuación 11)}$$

En la ecuación 11,  $E(k)$  representa una CIM conversacional basada en entropía relativa en un  $k$ -ésimo segmento de análisis de conversación, y  $P_i(k)$  representa una probabilidad de un estado conversacional en el  $k$ -ésimo segmento de análisis de conversación. En este ejemplo,  $P_A(k)$  representa la probabilidad de un tiempo de intervención sencilla para el participante conversacional A,  $P_B(k)$  representa la probabilidad de un tiempo de intervención sencilla para el participante conversacional B,  $P_M(k)$  representa la probabilidad de un tiempo silencioso compartido, durante el cual ninguno de los participantes conversacionales habla, y  $P_D(k)$  representa la probabilidad

de un tiempo de intervención doble, durante el cual ambos participantes conversacionales están hablando.

En la ecuación 11,  $Q_i(k)$  representa la probabilidad de que cada estado se corresponda con un modo de presentación. La determinación de la probabilidad de un modo de presentación puede basarse, al menos en parte, en la cantidad relativa de tiempo que un participante conversacional en particular está hablando, en la duración relativa de las charlas de un participante conversacional en relación con la duración de las charlas de otros participantes conversacionales, etc. Como se señaló anteriormente, algunas implantaciones pueden implicar enviar o recibir una indicación de silencio de hablante o una indicación de presentación, por ejemplo, como metadatos asociados con los paquetes de audio. Por lo tanto, la determinación de la probabilidad de un modo de presentación puede basarse, al menos en parte, en una indicación de silencio de hablante y/o en una indicación de presentación.

En algunos ejemplos, controlar un tamaño de memoria intermedia de fluctuación de acuerdo con los datos de dinámica de fluctuación de red y con los datos de interactividad conversacional (véase el bloque 215 de la figura 2) puede implicar configurar una memoria intermedia de fluctuación a un tamaño relativamente mayor cuando los datos de dinámica de fluctuación de red indican más de una cantidad umbral de fluctuación de la red y/o cuando los datos de interactividad conversacional indican menos cantidad que la cantidad umbral de interactividad conversacional. El bloque 215 puede implicar configurar una memoria intermedia de fluctuación a un tamaño relativamente más pequeño cuando los datos de dinámica de fluctuación de red indican menos cantidad que la cantidad umbral de fluctuación de red y/o cuando los datos de interactividad conversacional indican al menos una cantidad umbral de interactividad conversacional. En algunos ejemplos, el bloque 215 puede implicar configurar una memoria intermedia de fluctuación a un tamaño relativamente mayor cuando los datos de dinámica de fluctuación de red indican más cantidad que la cantidad umbral de fluctuación de red y/o cuando los datos de interactividad conversacional indican menos cantidad que la cantidad umbral de participación conversacional por un primer participante conversacional. En algunos ejemplos, el bloque 215 puede implicar configurar una memoria intermedia de fluctuación para el primer participante conversacional a un tamaño relativamente menor cuando los datos de dinámica de fluctuación de red indican menos cantidad que la cantidad umbral de fluctuación de red y/o cuando los datos de interactividad conversacional indican al menos una cantidad umbral de participación conversacional del primer participante conversacional.

Sin embargo, a los datos de dinámica de fluctuación de red y a los datos de interactividad conversacional se les puede dar o no el mismo peso en la determinación del tamaño de memoria intermedia de fluctuación, dependiendo de la implantación. En algunas implantaciones, por ejemplo, controlar el tamaño de memoria intermedia de fluctuación de acuerdo con los datos de dinámica de fluctuación de red y con los datos de interactividad conversacional puede implicar asignar una ponderación relativamente menor a los datos de dinámica de fluctuación de red y asignar una ponderación relativamente mayor a los datos de interactividad conversacional. Por ejemplo, si los datos de interactividad conversacional indican que existe una alta probabilidad de que una conversación esté en modo de presentación, se puede determinar un tamaño de memoria intermedia de fluctuación relativamente grande. Algunas implantaciones implican controlar el tamaño de memoria intermedia de fluctuación de acuerdo con uno de al menos tres modos de control de memoria intermedia de fluctuación, por ejemplo, como se describe a continuación. Cada uno de los modos de control de memoria intermedia de fluctuación corresponde a un espectro de tamaño de memoria intermedia de fluctuación. En algunas implantaciones, un tamaño de memoria intermedia de fluctuación "pequeño" puede estar en el espectro de decenas de milisegundos, por ejemplo, 5 ms, 10 ms, 25 ms, 30 ms, 35 ms, 40 ms, 45 ms, 50 ms, 55 ms, etc. En algunos ejemplos, un tamaño de memoria intermedia de fluctuación "grande" puede estar en el espectro de cientos de milisegundos, por ejemplo, 500 ms, 600 ms, 700 ms, 800 ms, 900 ms, 1000 ms, 1100 ms, 1200 ms, 1300 ms, 1400 ms, 1500 ms, etc. En algunas implantaciones, el tamaño de memoria intermedia de fluctuación puede fijarse hasta que los datos de dinámica de fluctuación de red y/o los datos de interactividad conversacional indiquen un cambio en el modo de control de memoria intermedia de fluctuación. Sin embargo, en otras implantaciones, el tamaño de memoria intermedia de fluctuación puede variar de manera adaptativa dentro de un espectro de tamaño durante el funcionamiento de un único modo de control de memoria intermedia de fluctuación, por ejemplo, como se analiza a continuación.

La figura 4 es un diagrama de flujo que ilustra bloques de algunos métodos de control de memoria intermedia de fluctuación proporcionados en el presente documento. El método 400 puede, por ejemplo, ser realizado (al menos en parte) por un servidor o por otro dispositivo tal que esté configurado para comunicarse con terminales de comunicación, tal como se describió anteriormente con referencia a la figura 1B. Sin embargo, algunos métodos proporcionados en el presente documento pueden ser realizados (al menos en parte) por un terminal de comunicación. Al igual que con otros métodos descritos en el presente documento, los bloques del método 400 no se realizan necesariamente en el orden indicado. Lo que es más, algunas implantaciones del método 400 pueden incluir más o menos bloques de los indicados o descritos.

Aquí, el método 400 comienza con el bloque 405, que implica recibir paquetes de audio de un segmento de análisis de conversación. En el bloque 410, se realiza un análisis de interactividad conversacional, que puede ser un análisis de interactividad conversacional a largo plazo. Como se utiliza en el presente documento, un intervalo de tiempo "largo plazo" incluye al menos una pluralidad de charlas. Si se determina en el bloque 415 que el análisis de interactividad conversacional indica un modo de presentación, el proceso continúa en el bloque 440, en el que se determina un modo de memoria intermedia de fluctuación de retardo largo/pérdida baja (relativamente grande). En

este ejemplo, el modo de retardo largo/pérdida baja se corresponde con un tamaño de memoria intermedia de fluctuación, o al menos con un espectro de tamaño de memoria intermedia de fluctuación. Por lo tanto, en este ejemplo, no es necesaria la detección de modo de pico para determinar el tamaño de memoria intermedia de fluctuación (o al menos un espectro de tamaño de memoria intermedia de fluctuación). Sin embargo, el tamaño de memoria intermedia de fluctuación puede variar de manera adaptativa durante un modo de retardo largo/pérdida baja. En algunas implantaciones, durante un modo de retardo largo/pérdida baja, los parámetros que pueden controlar un equilibrio entre el ritmo de pérdida de paquetes y el retardo de memoria intermedia de fluctuación se pueden ajustar para reducir la pérdida de paquetes a costa de aumentar el retarde de memoria intermedia.

5 Sin embargo, si se determina en el bloque 415 que el análisis de interactividad conversacional no indica un modo de presentación, el proceso continúa hasta el bloque 440 de detección del modo de pico, que implica analizar los paquetes de audio del segmento de análisis de conversación para determinar datos de dinámica de fluctuación de red.. En este ejemplo, el bloque 425 implica determinar si está presente un modo de pico (correspondiente a una alta fluctuación de red), en base a los datos de dinámica de fluctuación de red determinados en el bloque 420.

15 Si se determina en el bloque 425 que los datos de dinámica de fluctuación de red indican un modo de pico, el proceso continúa en el bloque 435, en el que se determina un espectro de tamaño de memoria intermedia de fluctuación en modo de pico. El tamaño de memoria intermedia de fluctuación en modo de pico puede ser un tamaño de memoria intermedia de fluctuación relativamente grande. En algunas implantaciones, el tamaño de memoria intermedia de fluctuación en modo de pico puede ser el mismo tamaño que el tamaño de memoria intermedia de fluctuación de retardo largo/pérdida baja correspondiente al bloque 440, mientras que, en otras implantaciones, el tamaño de memoria intermedia de fluctuación en modo de pico puede ser menor que el tamaño de memoria intermedia de fluctuación correspondiente a bloque 440. Sin embargo, el tamaño de memoria intermedia de fluctuación puede variar de manera adaptativa durante un modo de pico. En algunas implantaciones, durante un modo de pico, aumentará el retardo máximo de inicio de charla. En algunos ejemplos, pueden implantarse procesos asimétricos de "ataque" y "decaimiento" del control de memoria intermedia de fluctuación durante un modo de pico. Por ejemplo, en algunas implantaciones de modo de pico, un proceso de "ataque" de control de memoria intermedia de fluctuación puede permitir que el tamaño de memoria intermedia de fluctuación aumente rápidamente en respuesta a una fluctuación pico instantánea. En algunos ejemplos de modo de pico, el tamaño de memoria intermedia de fluctuación se puede controlar para reducir el ritmo de pérdida de paquetes cuando hay una llegada en ráfagas de paquetes con retardo después de un pico de retardo.

35 Si se determina en el bloque 425 que los datos de dinámica de fluctuación de red no indican un modo de pico, el proceso continúa en el bloque 430, en el que se determina un espectro de tamaño de memoria intermedia de fluctuación de "modo normal". En este ejemplo, el espectro de tamaño de memoria intermedia de fluctuación en "modo normal" es menor que el espectro de tamaño de memoria intermedia de fluctuación en modo de pico, y es más pequeño que el espectro de tamaño de memoria intermedia de fluctuación de retardo largo/pérdida baja. Sin embargo, el tamaño de memoria intermedia de fluctuación puede variar de manera adaptativa durante un modo normal. Por ejemplo, el tamaño de memoria intermedia de fluctuación puede variar de manera adaptativa de acuerdo con métodos conocidos por el experto en la técnica. El método 400 se puede realizar de manera continua, de tal modo que se puedan recibir y analizar paquetes de audio adicionales después de que se haya determinado un tamaño de memoria intermedia de fluctuación en el bloque 430, 435 o 440.

45 En algunas implantaciones, controlar el tamaño de memoria intermedia de fluctuación puede implicar establecer un tamaño de memoria intermedia de fluctuación de acuerdo con uno de al menos tres modos de control de memoria intermedia de fluctuación. Cada uno de los modos de control de memoria intermedia de fluctuación puede corresponder a un espectro de tamaño de memoria intermedia de fluctuación. Un ejemplo se muestra en la tabla 1, a continuación:

50 Tabla 1

Configuraciones de algoritmos	Modo de presentación	Modo de pico
Perfil de memoria intermedia de fluctuación en modo de retardo largo (pérdida baja)	1	1
Perfil de memoria intermedia de fluctuación en modo de retardo largo (pérdida baja)	1	0
Perfil de memoria intermedia de fluctuación sintonizado en modo de pico	0	1
Perfil de memoria intermedia de fluctuación en modo normal	0	0

55 En este ejemplo, el mayor de los espectros de tamaño de memoria intermedia de fluctuación corresponde a datos de dinámica de fluctuación de red que indican más de la cantidad umbral de fluctuación de red ("modo de pico") y datos de interactividad conversacional que indican menos cantidad que la cantidad umbral de interactividad conversacional ("modo de presentación"). Aquí, se establece un bit correspondiente a "modo de presentación" y un bit correspondiente a "modo de pico". En esta implantación, el tamaño más grande de la memoria intermedia de

fluctuación corresponde a un modo de retardo largo/pérdida baja. En algunas implantaciones, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen más de una cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen menos cantidad que la cantidad umbral de participación conversacional por parte de un único participante conversacional. En algunos ejemplos, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen más de una cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen menos cantidad que la cantidad umbral de participación conversacional de todos menos de uno de los participantes conversacionales, o de todos menos de un número umbral de participantes conversacionales. En algunos ejemplos, el tamaño más grande de memoria intermedia de fluctuación puede estar en el espectro de cientos de milisegundos, por ejemplo, 500 ms, 600 ms, 700 ms, 800 ms, 900 ms, 1000 ms, 1100 ms, 1200 ms, 1300 ms, 1400 ms, 1500 ms, etc. En algunas implantaciones, el tamaño de memoria intermedia de fluctuación puede fijarse en el tamaño más grande de memoria intermedia de fluctuación hasta que los datos de dinámica de fluctuación de red y/o los datos de interactividad conversacional indiquen un cambio. Sin embargo, en otras implantaciones, el tamaño de memoria intermedia de fluctuación puede variar de manera adaptativa dentro de un espectro de tamaño, por ejemplo, de acuerdo con los parámetros para el modo de retardo largo/pérdida baja descrito anteriormente.

En esta implantación, el mayor o el segundo mayor de los espectros de tamaño de memoria intermedia de fluctuación corresponde a datos de dinámica de fluctuación de red que indican menos cantidad que la cantidad umbral de fluctuación de red y a datos de interactividad conversacional que indican al menos una cantidad umbral de interactividad conversacional. Aquí, "modo de presentación" se establece en uno, y "modo de pico" se establece en cero. En algunas implantaciones, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen menos cantidad que la cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen menos cantidad que la cantidad umbral de participación conversacional por parte de un único participante conversacional. En algunos ejemplos, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen menos cantidad que la cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen menos cantidad que la cantidad umbral de participación conversacional de todos menos de uno de los participantes conversacionales, o de todos menos de un número umbral de participantes conversacionales. En esta implantación, el tamaño más grande de la memoria intermedia de fluctuación se corresponde con un modo de retardo largo/pérdida baja, que implica el mayor de al menos tres espectros de tamaño de memoria intermedia de fluctuación. Sin embargo, en implantaciones alternativas, esta condición puede corresponder al segundo mayor de al menos cuatro espectros de tamaño de memoria intermedia de fluctuación, o al mayor de dos espectros de tamaño de memoria intermedia de fluctuación.

En este ejemplo, el segundo tamaño más pequeño de la memoria intermedia de fluctuación corresponde a datos de dinámica de fluctuación de red que indican una cantidad superior a un umbral de fluctuación de red y datos de interactividad conversacional que indican una cantidad inferior a un umbral de interactividad conversacional. Aquí, "modo de presentación" se establece en cero y "modo de pico" se establece en uno. En algunas implantaciones, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen más de una cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen más de una cantidad umbral de participación conversacional por parte de un único participante conversacional. En algunos ejemplos, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen más de una cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen más de una cantidad umbral de participación conversacional por parte de un número umbral de participantes conversacionales. En este ejemplo, el segundo tamaño más pequeño de memoria intermedia de fluctuación corresponde a un tamaño de memoria intermedia de fluctuación en modo de pico. Sin embargo, en algunas implantaciones alternativas, el tamaño de memoria intermedia de fluctuación en modo de pico puede ser el mismo tamaño que la memoria intermedia de fluctuación que corresponde a un modo de retardo largo/pérdida baja. Como se señaló anteriormente, en algunas implantaciones, el tamaño de memoria intermedia de fluctuación puede ser variable durante un modo de pico. En algunas implantaciones, durante un modo de pico, aumentará el retardo máximo de inicio de charla.

En algunos ejemplos, se pueden implantar procesos asimétricos de "ataque" y "decaimiento" del control de memoria intermedia de fluctuación durante un modo de pico. Como se utiliza en el presente documento, el término "ataque" puede corresponder a una respuesta a un pico de retardo (por ejemplo, aumentando el tamaño de memoria intermedia de fluctuación) y "decaimiento" puede corresponder a un proceso que no es de ataque, tal como a un retorno a un tamaño de memoria intermedia de fluctuación más bajo después de un proceso de ataque. Por ejemplo, la aplicación de parámetros de suavizado asimétricos para procesos de ataque y decaimiento puede implicar la aplicación de un parámetro de suavizado de ataque si la PDV o el IAT es mayor que el tamaño de memoria intermedia de fluctuación actual. En algunas implantaciones, la aplicación de parámetros asimétricos de suavizado para los procesos de ataque y decaimiento puede implicar la aplicación de un parámetro de suavizado de decaimiento si la PDV o el IAT no es mayor que el tamaño de memoria intermedia de fluctuación actual. En algunas implantaciones de modo de pico, un proceso de "ataque" de control de memoria intermedia de fluctuación puede permitir que el tamaño de memoria intermedia de fluctuación aumente rápidamente en respuesta a una fluctuación pico instantánea. El tamaño de memoria intermedia de fluctuación se puede controlar para reducir el ritmo de pérdida de paquetes cuando se produce una llegada en ráfagas de paquetes retardados después de un pico de retardo.

- En esta implantación, el más pequeño de los tamaños de memoria intermedia de fluctuación corresponde a datos de dinámica de fluctuación de red que indican menos cantidad que la cantidad umbral de fluctuación de red y datos de interactividad conversacional que indican menos cantidad que la cantidad umbral de interactividad conversacional.
- 5 En algunas implantaciones, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen menos cantidad que la cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen más de una cantidad umbral de participación conversacional por parte de un único participante conversacional. En algunos ejemplos, este modo se indicará cuando los datos de dinámica de fluctuación de red indiquen menos cantidad que la cantidad umbral de fluctuación de red y cuando los datos de interactividad conversacional indiquen más de una cantidad umbral de participación conversacional por parte de un número umbral
- 10 de participantes conversacionales. Aquí, "modo de presentación" se establece en cero y "modo de pico" se establece en cero, lo que indica un "modo normal" de control de memoria intermedia de fluctuación y un espectro de tamaño de memoria intermedia de fluctuación de un modo normal correspondiente. En este ejemplo, el espectro de tamaño de memoria intermedia de fluctuación en modo normal es menor que el espectro de tamaño de memoria intermedia de fluctuación de retardo largo/pérdida baja. En algunas implantaciones, el espectro de tamaño de memoria intermedia de fluctuación en modo normal puede estar en el espectro de decenas de milisegundos, por ejemplo, 25 ms, 30 ms, 35 ms, 40 ms, 45 ms, 50 ms, 55 ms, etc. Sin embargo, el tamaño de memoria intermedia puede variar de manera adaptativa durante un modo normal. Por ejemplo, el tamaño de memoria intermedia de fluctuación puede variar de manera adaptativa de acuerdo con métodos conocidos por el experto en la técnica.
- 15
- 20 Aunque pueden implantarse diversos aspectos de la presente divulgación en un dispositivo, tal como un servidor, que pueda facilitar las comunicaciones entre múltiples terminales de comunicación, algunos aspectos pueden implantarse (al menos en parte) en un único terminal de comunicación. En algunos de tales ejemplos, determinar los datos de interactividad conversacional puede implicar analizar la actividad conversacional de un solo participante conversacional. Debido a que tales implantaciones pueden implicar analizar la actividad conversacional de un participante conversacional que está utilizando un terminal de comunicación particular, se puede hacer referencia al participante conversacional único como participante conversacional "local" o "de extremo cercano".
- 25
- 30 En algunas de tales implantaciones, analizar la actividad conversacional del único participante conversacional puede implicar determinar si el único participante conversacional está o no interviniendo. Controlar el tamaño de memoria intermedia de fluctuación puede implicar configurar la memoria intermedia de fluctuación en un espectro de tamaño relativamente menor cuando el único participante conversacional está interviniendo, y configurar la memoria intermedia de fluctuación en un espectro de tamaño relativamente mayor cuando el único participante conversacional no está interviniendo.
- 35
- 40 Configurar la memoria intermedia de fluctuación en un espectro de tamaño relativamente menor cuando el participante conversacional de extremo cercano está interviniendo puede permitir que otros participantes conversacionales de "extremo lejano" interrumpan y/o respondan rápidamente al participante conversacional de extremo cercano. Cuando el participante conversacional más cercano no está interviniendo, esta condición indica que es probable que otro participante conversacional esté interviniendo o que hay un tiempo silencioso compartido. En tales condiciones, la longitud de memoria intermedia de fluctuación del participante conversacional de extremo cercano se puede establecer en un espectro de tamaño relativamente mayor, porque puede ser aceptable un retardo mayor.
- 45
- 50 De acuerdo con algunas de tales realizaciones, se pueden usar estadísticas de la distribución de longitud de charla del participante conversacional de extremo cercano para controlar el algoritmo de la memoria intermedia de fluctuación. Por ejemplo, si las charlas del participante conversacional de extremo cercano tienden a ser relativamente largas, esto puede indicar que el participante conversacional de extremo cercano está haciendo una presentación.
- 55
- 60 La figura 5 es un diagrama de bloques que proporciona ejemplos de componentes de un aparato capaz de implantar diversos aspectos de esta divulgación. El aparato 500 puede ser (o puede ser una parte de), por ejemplo, un terminal de comunicación, un servidor, etc. En algunos ejemplos, el aparato puede implantarse en un componente de otro dispositivo. Por ejemplo, en algunas implantaciones el aparato 500 puede ser una tarjeta de línea.
- 65
- En este ejemplo, el aparato 500 incluye un sistema de interfaz 505, un sistema 510 de memoria y un sistema lógico 515. El sistema lógico 515 y/o el sistema 510 de memoria pueden ser capaces de establecer una o más memorias intermedias de fluctuación en el sistema 510 de memoria. El sistema 505 de interfaz puede incluir una interfaz de red, una interfaz entre el sistema lógico y el sistema de memoria y/o una interfaz de dispositivo externo (tal como una interfaz de bus serie universal (USB)). El sistema lógico 515 puede incluir, por ejemplo, un procesador de uno o múltiples microprocesadores de fines generales, un procesador de señales digitales (DSP), un circuito integrado de aplicación específica (ASIC), una matriz de puertas programables en campo (FPGA) u otro dispositivo lógico programable, lógica de transistor o puerta discreta, y/o componentes de hardware discretos.
- En este ejemplo, el sistema lógico 515 es capaz de recibir datos de audio mediante el sistema de interfaz. Los datos de audio pueden incluir paquetes de audio recibidos en tiempos reales de llegada de paquetes durante un intervalo

- de tiempo que corresponde a un segmento de análisis de conversación. El intervalo de tiempo puede ser un intervalo de tiempo a largo plazo que incluya una pluralidad de charlas. El sistema lógico 515 puede ser capaz de analizar los datos de audio para determinar los datos de dinámica de fluctuación de red y los datos de interactividad conversacional y de controlar un tamaño de memoria intermedia de fluctuación de acuerdo con los datos de
- 5 dinámica de fluctuación de red y con los datos de interactividad conversacional.
- El análisis de los datos de audio para determinar los datos de dinámica de fluctuación de red puede implicar determinar al menos una variación de entre la variación de retardo de paquetes (PDV) y la variación de tiempo entre llegadas (IAT) comparando los tiempos esperados de llegada de paquetes con los tiempos reales de llegada de
- 10 paquetes. Analizar los datos de audio para determinar los datos de dinámica de fluctuación de red puede implicar determinar al menos un elemento de entre una probabilidad de presencia de pico de retardo y una intensidad de pico de retardo.
- Analizar los datos de audio para determinar los datos de interactividad conversacional implica determinar tiempos de intervención sencilla, tiempos de intervención doble y tiempos silenciosos compartidos. En algunas implantaciones, analizar los datos de audio para determinar los datos de interactividad conversacional puede implicar determinar una medida de interactividad conversacional (CIM) en base a reglas heurísticas y/o entropía relativa conversacional.
- 15
- La figura 6 es un diagrama de bloques que proporciona ejemplos de componentes de un aparato de procesamiento de audio. En este ejemplo, el dispositivo 600 incluye un sistema 605 de interfaz. El sistema 605 de interfaz puede incluir una interfaz de red, tal como una interfaz de red inalámbrica. Alternativamente, o adicionalmente, el sistema 605 de interfaz puede incluir una interfaz de bus serie universal (USB) u otra tal interfaz.
- 20
- El dispositivo 600 incluye un sistema lógico 610. El sistema lógico 610 puede incluir un procesador, tal como un procesador de un solo microprocesador o de múltiples microprocesadores de fines generales. El sistema lógico 610 puede incluir un procesador de señales digitales (DSP), un circuito integrado de aplicación específica (ASIC), una matriz de puertas programables en campo (FPGA) u otro dispositivo lógico programable, lógica de transistor o puerta discreta, o componentes de hardware discretos, o combinaciones de ellos. El sistema lógico 610 puede configurarse para controlar los otros componentes del dispositivo 600. Por ejemplo, el sistema lógico 610 puede ser capaz de controlar el tamaño de una o de más memorias intermedias de fluctuación en el sistema 615 de memoria. Aunque no se muestra ninguna interfaz entre los componentes del dispositivo 600 de la figura 6, el sistema lógico 610 puede configurarse con interfaces para establecer comunicación con los otros componentes. Los otros componentes pueden estar configurados o no para comunicarse entre sí, según corresponda.
- 25
- El sistema lógico 610 puede configurarse para realizar análisis de datos de audio y funcionalidad de control de memoria intermedia de fluctuación, incluyendo, entre otras, la funcionalidad descrita en el presente documento. En algunas de tales implantaciones, el sistema lógico 610 puede configurarse para funcionar (al menos en parte) de acuerdo con el software almacenado en uno o más medios no transitorios. Los medios no transitorios pueden incluir memoria asociada con el sistema lógico 610, tal como memoria de acceso aleatorio (RAM) y/o memoria de sólo lectura (ROM). Los medios no transitorios pueden incluir memoria del sistema 615 de memoria. El sistema 615 de memoria puede incluir uno o más tipos adecuados de medios de almacenamiento no transitorios, tales como memoria flash, un disco duro, etc.
- 30
- El sistema lógico 610 puede configurarse para recibir tramas de datos de audio codificados mediante el sistema 605 de interfaz y para decodificar los datos de audio codificados. Alternativamente, o adicionalmente, el sistema lógico 610 puede configurarse para recibir tramas de datos de audio codificados mediante una interfaz entre el sistema 615 de memoria y el sistema lógico 610. El sistema lógico 610 puede configurarse para controlar el/los altavoz/altavoces 620 de acuerdo con datos de audio decodificados.
- 35
- El sistema 630 de exhibición visual puede incluir uno o más tipos adecuados de dispositivos de exhibición visual, dependiendo de la manifestación del dispositivo 600. Por ejemplo, el sistema 630 de exhibición visual puede incluir un dispositivo de exhibición visual de cristal líquido, un dispositivo de exhibición visual de plasma, un dispositivo de exhibición visual biestable, etc.
- 40
- El sistema 635 de entrada de usuario puede incluir uno o más dispositivos configurados para aceptar entradas de un usuario. En algunas implantaciones, el sistema 635 de entrada de usuario puede incluir una pantalla táctil que se superpone a un dispositivo de exhibición visual del sistema 630 de exhibición visual. El sistema 635 de entrada de usuario puede incluir un ratón, una bola de seguimiento, un sistema de detección de gestos, una palanca de mano, una o más interfaces gráficas de usuario y/o uno o más menús presentados en el sistema 630 de exhibición visual, botones, un teclado, interruptores, etc. En algunas implantaciones, el sistema 635 de entrada de usuario puede incluir el micrófono 625: un usuario puede dar órdenes de voz al dispositivo 600 mediante el micrófono 625. El sistema lógico puede configurarse para reconocimiento de voz y para controlar al menos algunos funcionamientos del dispositivo 600 de acuerdo con tales órdenes de voz.
- 45
- El sistema 640 de energía puede incluir uno o más dispositivos de almacenamiento de energía adecuados, tales como una batería de níquel-cadmio o una batería de iones de litio. El sistema 640 de energía puede configurarse
- 50
- 55
- 60
- 65

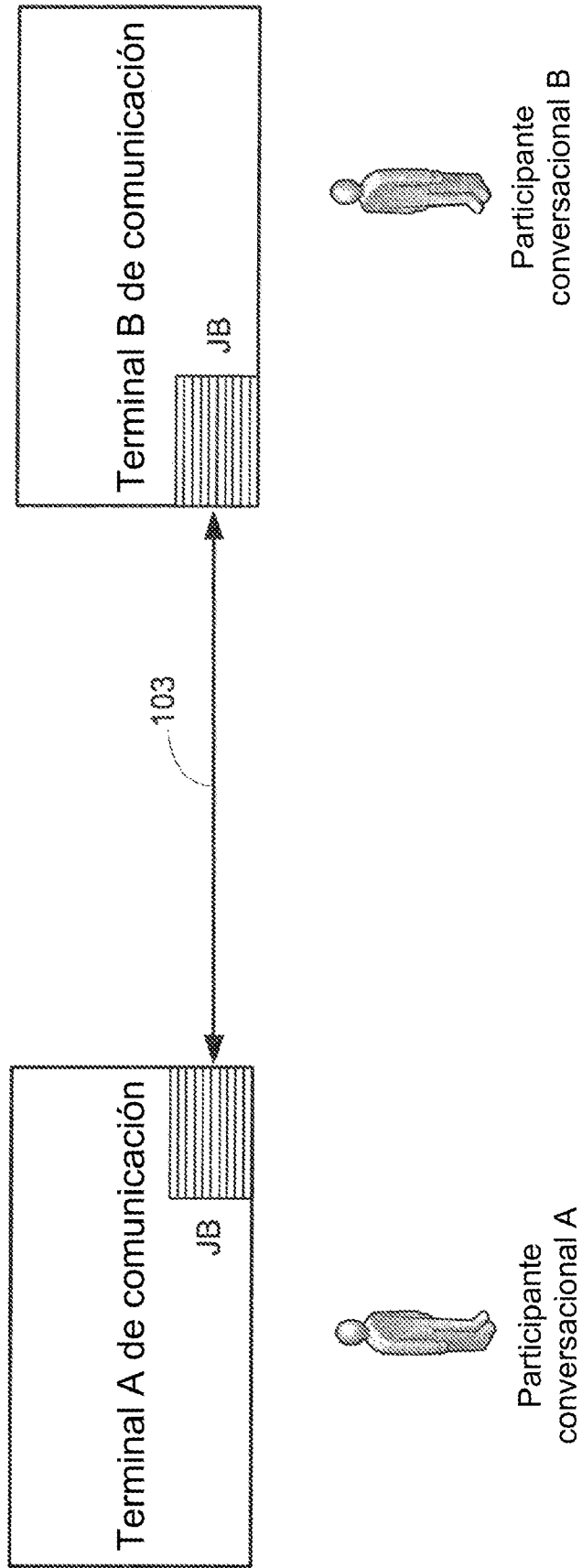
para recibir energía desde una toma de corriente.

5 Las diversas modificaciones a las implantaciones descritas en esta divulgación le resultarán fácilmente evidentes al experto en la técnica. Los principios generales definidos en el presente documento se pueden aplicar a otras implantaciones sin apartarse del alcance de esta divulgación. De este modo, no se pretende que las reivindicaciones se limiten a las implantaciones que se muestran en el presente documento, sino que van acordes con un más amplio alcance consistente con esta divulgación, con los principios y con las nuevas características descritas en el presente documento.

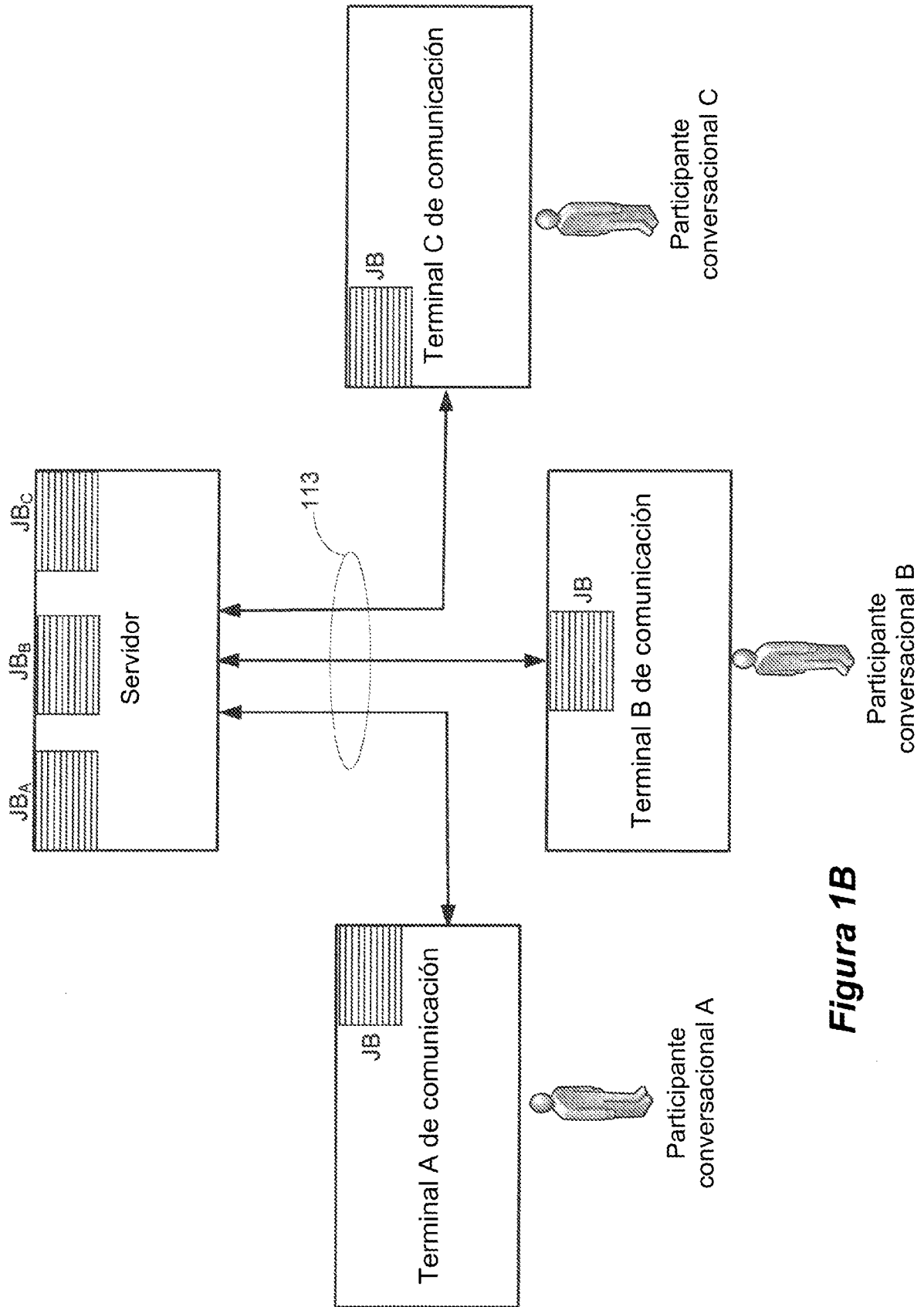
**REIVINDICACIONES**

1. Un método, que comprende:

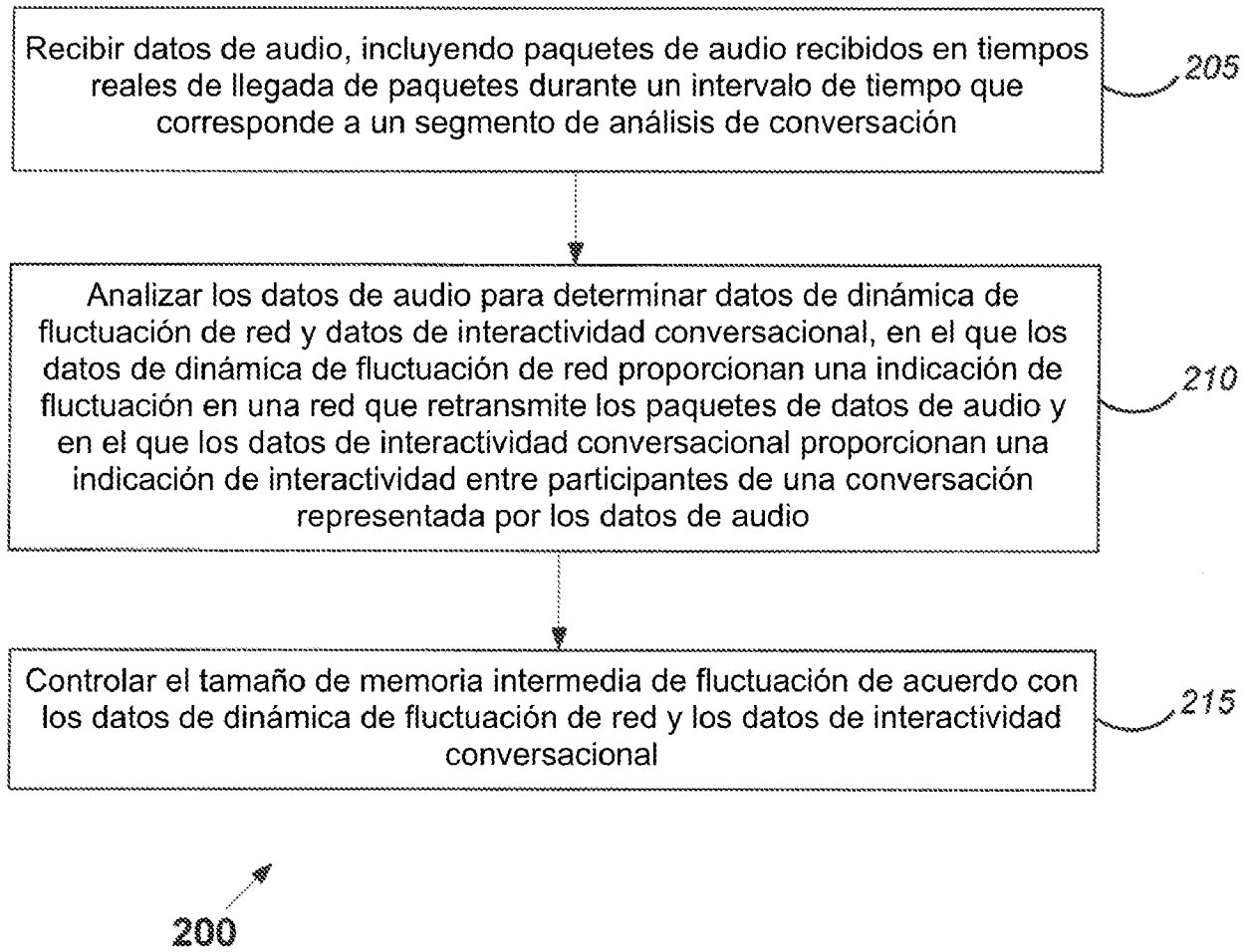
- 5 recibir tramas de datos de audio, estando las tramas de datos de audio incluidas en paquetes de datos de audio recibidos en tiempos reales de llegada de paquetes durante un intervalo de tiempo que corresponde a un segmento (205) de análisis de conversación;
- 10 analizar los datos de audio del segmento de análisis de conversación para determinar datos de dinámica de fluctuación de red y datos de interactividad conversacional, en el que los datos de dinámica de fluctuación de red proporcionan una indicación de fluctuación en una red que retransmite los paquetes de datos de audio y en el que los datos de interactividad conversacional proporcionan una indicación de interactividad entre participantes de una conversación representada por la trama (210) de datos de audio;
- 15 seleccionar, en base a la dinámica de fluctuación de red y a los datos (215) de interactividad conversacional de los datos de audio, un modo de control de memoria intermedia de fluctuación entre al menos tres modos de control de memoria intermedia de fluctuación, en el que cada modo de control de memoria intermedia de fluctuación corresponde a un espectro de tamaños de memoria intermedia de fluctuación; y
- 20 determinar un tamaño de memoria intermedia de fluctuación en base al modo de control de memoria intermedia de fluctuación seleccionado.
- 25 2. El método de la reivindicación 1, en el que determinar el tamaño de memoria intermedia de fluctuación comprende variar de manera adaptativa el tamaño de memoria intermedia de fluctuación dentro del espectro de tamaños para el modo de control de memoria intermedia de fluctuación seleccionado.
- 30 3. El método de una cualquiera de las reivindicaciones 1 a 2, en el que los al menos tres modos de control de memoria intermedia de fluctuación incluyen un modo de pico, un modo de pérdida baja y un modo normal,
- en el que el modo de pico se determina mediante el uso de cálculos entre percentiles, IPR, intensidad de pico de retardo y/o probabilidad de presencia de pico,
- en el que el modo de pérdida baja está relacionado con un modo de presentación,
- 35 en el que el modo normal es un modo sin pico y un modo sin presentación.
4. Un aparato configurado para realizar el método de cualquiera de las reivindicaciones 1-3.
- 40 5. Un producto de programa de ordenador que tiene instrucciones que, cuando se ejecutan por un dispositivo o sistema informático, hacen que dicho dispositivo o sistema informático realice el método de cualquiera de las reivindicaciones 1 a 3.



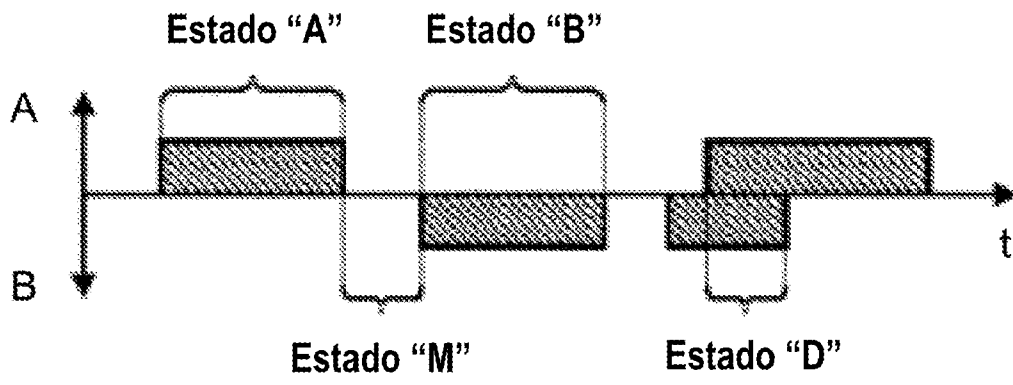
**Figura 1A**



**Figura 1B**



**Figura 2**



**Figura 3**

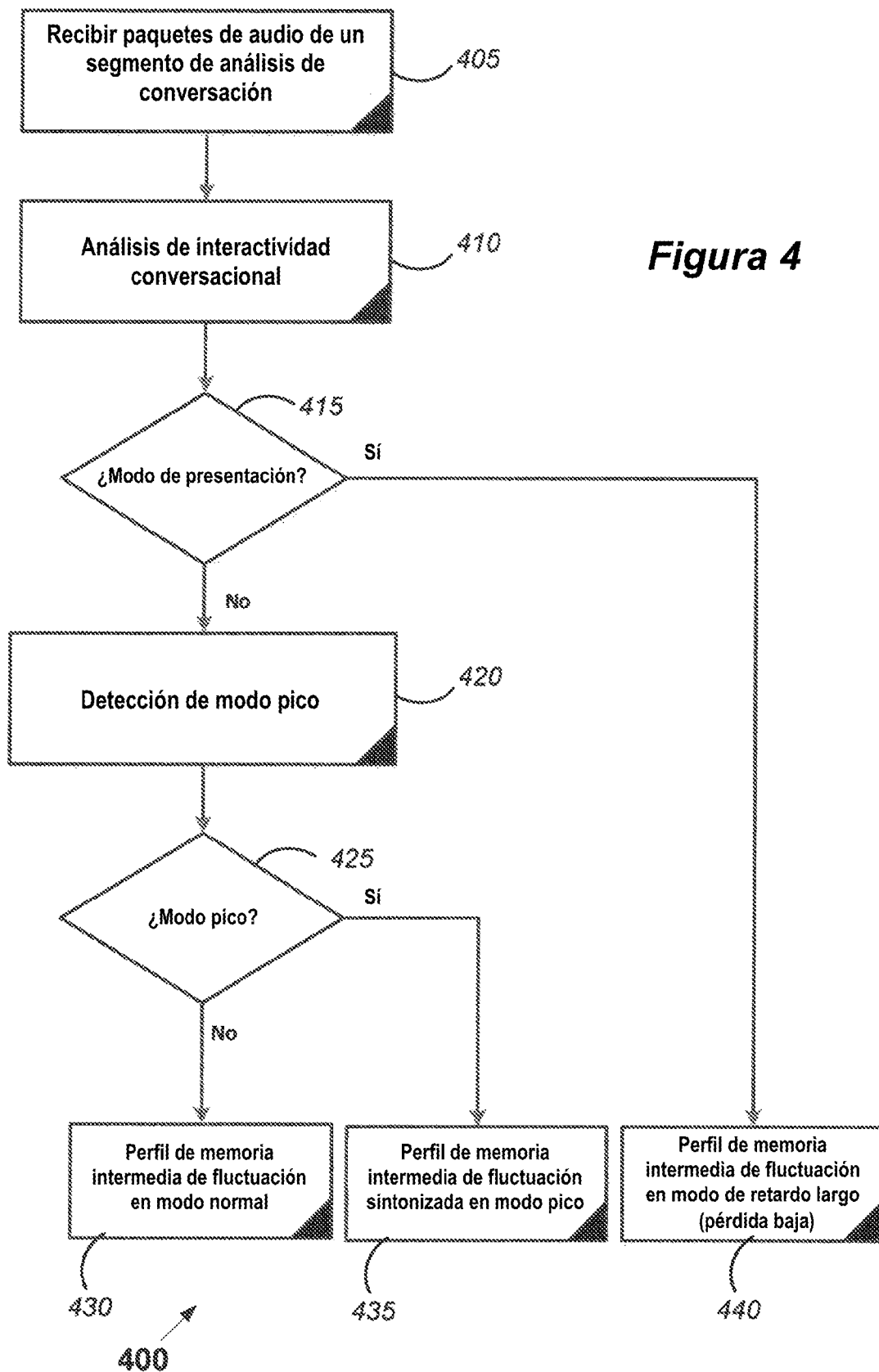
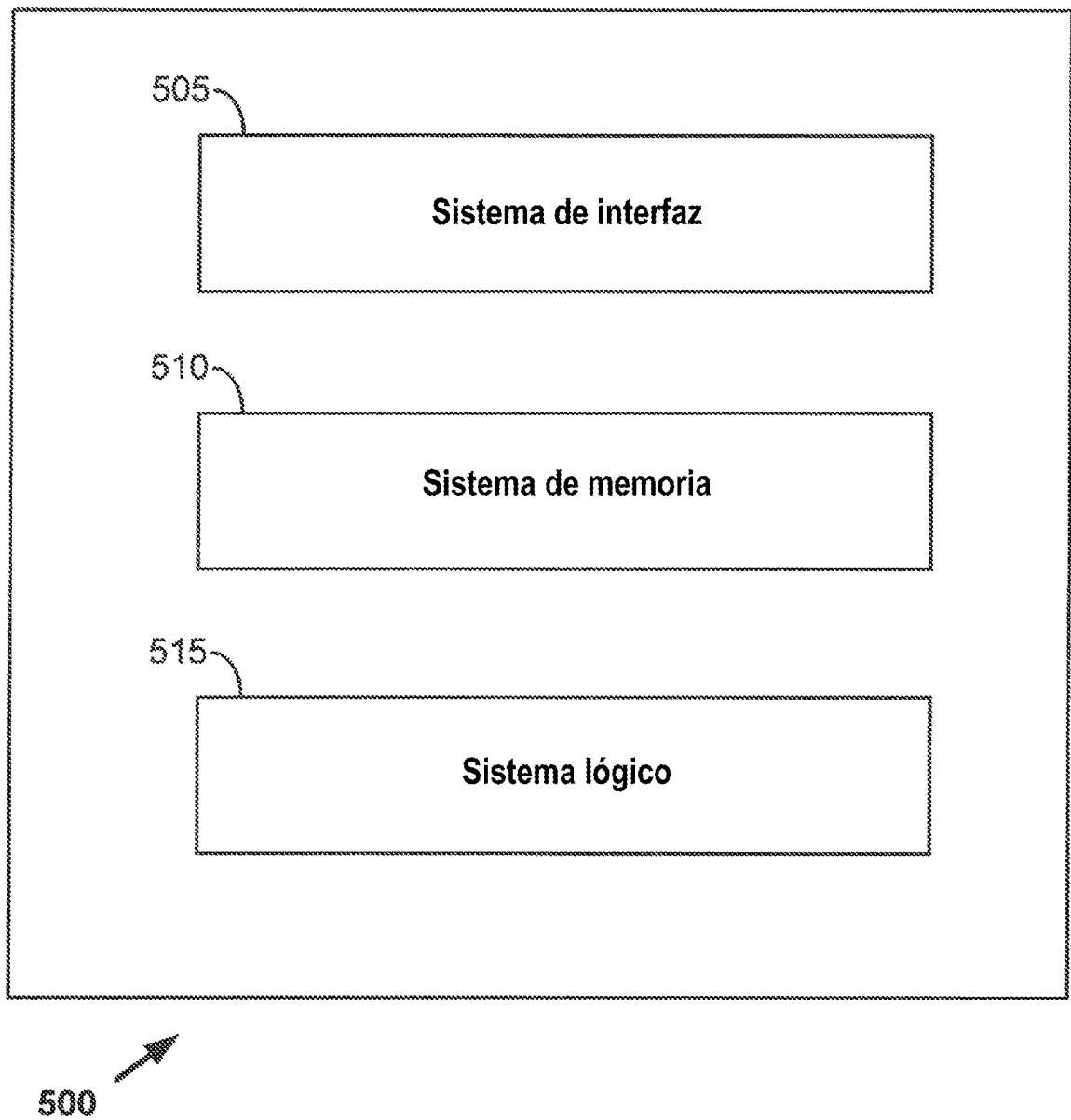
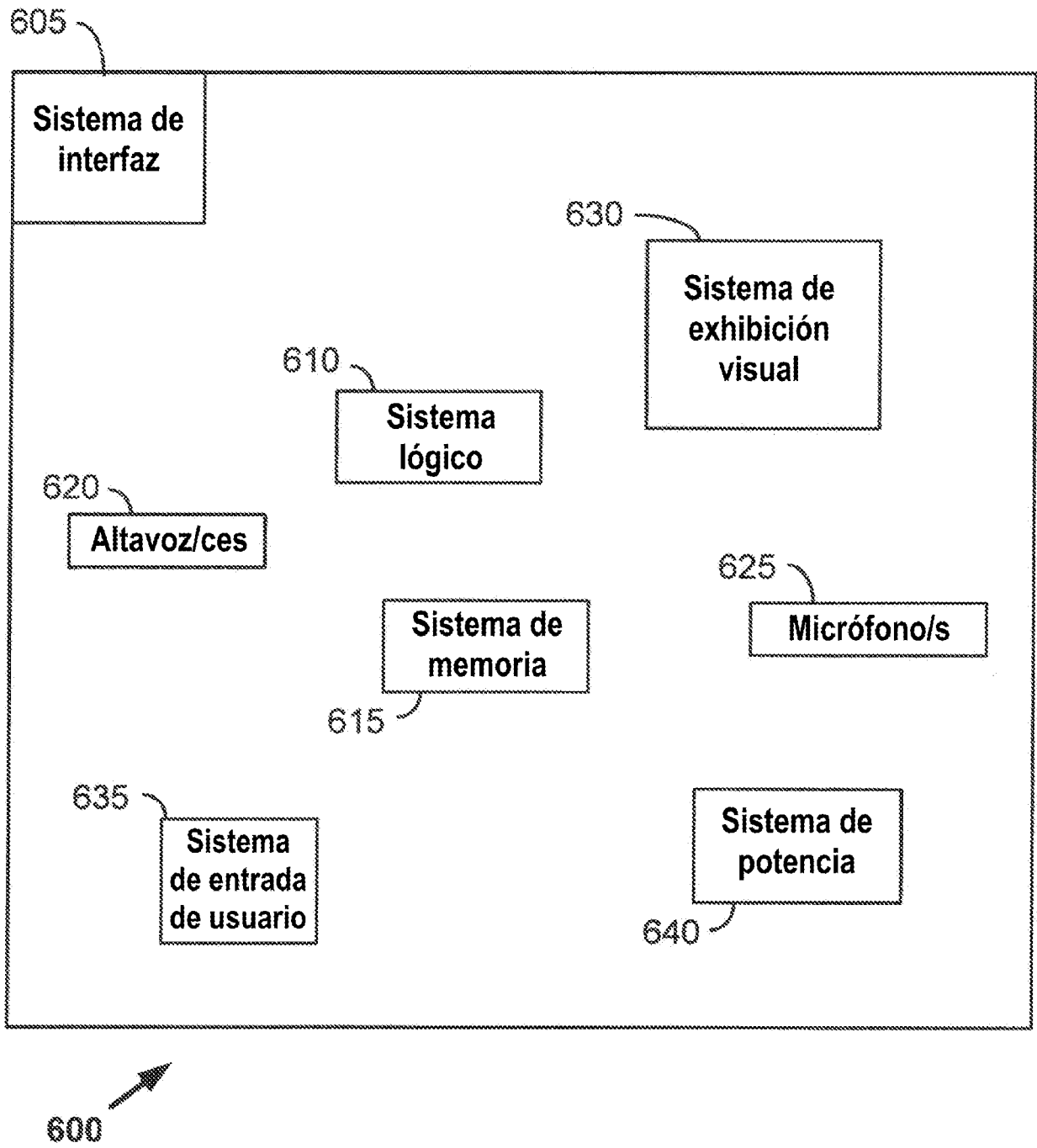


Figura 4



**Figura 5**



**Figura 6**