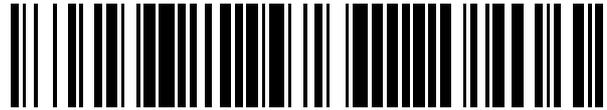


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 395 717**

51 Int. Cl.:

G10L 15/28 (2006.01)

G10L 19/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.02.2004 E 04708630 (1)**

97 Fecha y número de publicación de la concesión europea: **14.11.2012 EP 1595244**

54 Título: **Cuantificación de la frecuencia fundamental para el reconocimiento de voz distribuido**

30 Prioridad:

07.02.2003 US 360581

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

14.02.2013

73 Titular/es:

**MOTOROLA MOBILITY, LLC (50.0%)
600 North US Highway 45
Libertyville, IL 60048 , US y
INTERNATIONAL BUSINESS MACHINES
CORPORATION (50.0%)**

72 Inventor/es:

**RAMABADRAN, TENKASI, V. y
SORIN, ALEXANDER**

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 395 717 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

Cuantificación de la frecuencia fundamental para el reconocimiento de voz distribuido

5 Referencia cruzada a una solicitud relacionada

La presente solicitud de patente está relacionada con la solicitud de patente de Estados Unidos co-pendiente y de propiedad común número 10/360.582, número de expediente del mandatario CML00872M, titulada "Class Quantization For Distributed Speech Recognition", presentada en la misma fecha con la presente solicitud de patente.

Campo de la invención

La presente invención se refiere, en general, al campo de los sistemas de reconocimiento de voz distribuidos, y más específicamente se refiere al reconocimiento de voz distribuido para las comunicaciones de ancho de banda estrecho y comunicaciones inalámbricas.

Antecedentes de la invención

Con la llegada de los buscapersonas y los teléfonos móviles la industria de los servicios inalámbricos ha crecido hasta convertirse en una industria de billones de dólares. La mayor parte de los ingresos de los proveedores de servicios inalámbricos (WSP) proceden de las suscripciones. Como tal, la capacidad de un WSP para ejecutar una red con éxito depende de la calidad del servicio proporcionado a los abonados a través de una red que tiene un ancho de banda limitado. Para este fin, los WSP están constantemente buscando maneras de mitigar la cantidad de información que se transmite a través de la red mientras se mantiene una alta calidad del servicio a los abonados.

Recientemente, el reconocimiento de voz ha disfrutado de éxito en la industria del servicio inalámbrico. El reconocimiento de voz se usa para una diversidad de aplicaciones y servicios. Por ejemplo, un abonado del servicio inalámbrico puede estar provisto de una característica de marcación rápida mediante la que el abonado dice el nombre de un destinatario de una llamada en el dispositivo inalámbrico. El nombre del destinatario se reconoce usando el reconocimiento de voz y se inicia una llamada entre el abonado y el destinatario. En otro ejemplo, la información (411) del llamante puede utilizar el reconocimiento de voz para reconocer el nombre de un destinatario a quien un abonado está intentando hacer una llamada.

Como el reconocimiento de voz gana aceptación en la comunidad inalámbrica, el reconocimiento de voz distribuido (DSR) ha surgido como una tecnología emergente. El DSR se refiere a una trama en la que se distribuye la extracción de características y las partes de reconocimiento de patrones de un sistema de reconocimiento de voz. Esto es, el sistema de extracción de características y las partes de reconocimiento de patrones del reconocimiento de voz se realiza mediante dos unidades de procesamiento diferentes en dos localizaciones diferentes. Específicamente, el proceso de extracción de características se realiza en el extremo frontal, es decir, el dispositivo inalámbrico, y el proceso de reconocimiento de patrones se realiza en el extremo trasero, es decir, por el proveedor de servicios inalámbricos. El DSR mejora el reconocimiento de voz para las tareas más complicadas, tales como la reserva de avión automatizada con la información de vuelo hablada o las transacciones de corretaje con características similares.

El Instituto Europeo de Normas de Telecomunicaciones (ETSI) promulga un conjunto de normas para el DSR. Las normas DRS del ETSI ES 201 108 (de abril de 2000) y ES 202 050 (de julio de 2002) definen la extracción de características y algoritmos de compresión en el extremo frontal. Estas normas, sin embargo, no incorporan reconstrucción de voz en el extremo trasero, lo que puede ser importante en algunas aplicaciones. Como resultado, se han liberado los nuevos elementos WI-030 y 034-WI de trabajo por el ETSI para extender las normas anteriores (ES 201 108 y ES 202 050, respectivamente), para incluir la reconstrucción de voz en el extremo trasero, así como el reconocimiento del lenguaje tonal.

En las normas de DSR actuales, las características que se extraen, se comprimen, y se transmiten al extremo trasero son 13 coeficientes cepstrales en las frecuencias de Mel (MFCC), C0 - C12, y el logaritmo de la trama de energía, log-E. Estas características se actualizan cada 10 ms o 100 veces por segundo. En las propuestas de las normas extendidas (es decir, los elementos de trabajo descritos anteriormente), la información de frecuencia fundamental y de clase (o sonoridad) se obtienen también para cada trama y se transmiten, además, a los MFCC y al log-E. Esto aumenta la cantidad de información que se transmite por el dispositivo inalámbrico en la red y consume ancho de banda adicional. De esta manera, es deseable que la representación de la información de clase y de la frecuencia fundamental sea tan compacta como sea posible con el fin de mantener baja la velocidad de bits.

Ha sido un problema continuo representar la información de frecuencia fundamental de forma compacta sin sacrificar exactitud y robustez frente a los errores del canal de comunicación. Uno de los enfoques para cuantificar la frecuencia fundamental se presenta en el documento de Estados Unidos 6199037 B1. En general, los codificadores de voz (por ejemplo, el codificador predictivo lineal de excitación mixta, que es la norma federal de Estados Unidos a

2400 bps) cuantifican, de manera absoluta la información de la frecuencia fundamental usando 7 o más bits por trama. En las normas DSR extendidas, es importante mantener la velocidad de bits adicional, debido a la información de frecuencia fundamental y de clase, tan baja como sea posible. Se han adoptado una combinación de técnicas absolutas y diferenciales para cuantificar la información de período de la frecuencia fundamental usando solamente 6 bits por trama, de esta manera se ahorra al menos 1 bit por trama. Sin embargo, esto puede generar, potencialmente, problemas en términos de exactitud y robustez para los errores del canal.

Por lo tanto, existe una necesidad de superar los problemas con la técnica anterior como se ha tratado anteriormente.

Sumario de la invención

Brevemente, de acuerdo con la presente invención, se desvela un sistema, método, programa informático y medio legible por ordenador para cuantificar la información de la frecuencia fundamental de audio de acuerdo con las reivindicaciones adjuntas 1, 6, 11 y 21. En una realización de la presente invención, el método para un sistema de procesamiento de información incluye capturar audio que represente una trama numerada de una pluralidad de tramas numeradas. El método incluye además calcular una clase de la trama, en la que una clase es una cualquiera de entre una clase sonora y una clase sorda. Si la trama es una clase sonora, se calcula una frecuencia fundamental para la trama. Si la trama es una trama numerada par y una clase sonora, se calcula una palabra clave de una primera longitud cuantificando la frecuencia fundamental de la trama de manera absoluta. Si la trama es una trama numerada par y una clase sorda, se calcula una palabra clave de la primera longitud que indique una trama de clase sorda. Si la trama es una trama numerada impar y una clase sonora, y al menos una de las tres tramas inmediatamente anteriores a la trama es fiable, se calcula una palabra clave de una segunda longitud cuantificando la frecuencia fundamental de la trama de manera diferencial que hace referencia a una frecuencia fundamental cuantificada de la trama fiable anterior más cercana, en la que la primera longitud es mayor que la segunda longitud. Si la trama es una trama numerada impar y una clase sonora, y cada una de las tres tramas inmediatamente anteriores a la trama no es fiable, se calcula una palabra clave de la segunda longitud cuantificando la frecuencia fundamental de la trama de manera absoluta. Si la trama es una trama numerada impar y una clase sorda, se calcula una palabra clave de la segunda longitud que indique una trama de clase sorda.

En esta realización una trama numerada par es fiable si es una clase sonora y una trama numerada impar es fiable si es una clase sonora y la frecuencia fundamental de la trama se cuantifica de manera absoluta o se cuantifica de manera diferencial en referencia a una frecuencia fundamental de la trama inmediatamente anterior.

En otra realización de la presente invención, un sistema de procesamiento de información para cuantificar la información de la frecuencia fundamental de audio, incluye un micrófono para capturar audio que representa una trama numerada de una pluralidad de tramas numeradas. El sistema de procesamiento de información incluye además un procesador de señal digital para calcular una clase de la trama, en la que una clase es una cualquiera de entre una clase sonora y una clase sorda. Si la trama es una clase sonora, el procesador calcula una frecuencia fundamental para la trama. Si la trama es una trama numerada par y una clase sonora, el procesador calcula una palabra clave de una primera longitud cuantificando la frecuencia fundamental de la trama de manera absoluta. Si la trama es una trama numerada par y una clase sorda, el procesador calcula una palabra clave de la primera longitud que indique una trama de clase sorda. Si la trama es una trama numerada impar y una clase sonora, y al menos una de las tres tramas inmediatamente anteriores a la trama es fiable, el procesador calcula una palabra clave de una segunda longitud cuantificando la frecuencia fundamental de la trama de manera diferencial que hace referencia a una frecuencia fundamental cuantificada de la trama fiable anterior más cercana, en la que la primera longitud es mayor que la segunda longitud. Si la trama es una trama numerada impar y una clase sonora, y cada una de las tres tramas inmediatamente anteriores a la trama no es fiable, el procesador calcula una palabra clave de la segunda longitud cuantificando la frecuencia fundamental de la trama de manera absoluta. Si la trama es una trama numerada impar y una clase sorda, el procesador calcula una palabra clave de la segunda longitud que indique una trama de clase sorda.

En esta realización una trama numerada par es fiable si es una clase sonora y una trama numerada impar es fiable si es una clase sonora y la frecuencia fundamental de la trama se cuantifica de manera absoluta o se cuantifica de manera diferencial en referencia a una frecuencia fundamental de la trama inmediatamente anterior.

Las realizaciones preferidas de la presente invención son ventajosas porque sirven para disminuir la cantidad de bits usados para transmitir información de audio a través de una red de comunicaciones. Esto es beneficioso porque las redes de comunicaciones poseen un ancho de banda limitado. El ahorro de bits se convierte en hacer más disponible el ancho de banda para los abonados actuales o adicionales. De esta manera, la presente invención proporciona tanto una mejora en el rendimiento de red como un aumento de la calidad de las comunicaciones.

Breve descripción de los dibujos

La figura 1 es un diagrama de bloques que ilustra una red para el reconocimiento de voz distribuido de acuerdo con una realización preferida de la presente invención

La figura 2 es un diagrama de bloques detallado de un sistema de comunicación inalámbrico para el reconocimiento de voz distribuido de acuerdo con una realización preferida de la presente invención.

La figura 3 es un diagrama de bloques que ilustra un dispositivo móvil para un sistema de comunicación inalámbrico de acuerdo con una realización preferida de la presente invención.

5 La figura 4 es un diagrama de bloques que ilustra el intercambio de datos de un dispositivo inalámbrico con el proveedor del servicio inalámbrico.

La figura 5 es un diagrama de flujo funcional que muestra el proceso de reconocimiento de voz distribuido global de acuerdo con una realización preferida de la presente invención.

10 La figura 6 es un diagrama de flujo funcional que muestra un proceso de cuantificación de clase y de frecuencia fundamental de acuerdo con una realización preferida de la presente invención.

La figura 7 es un diagrama de bloques que ilustra las asignaciones de bits convencionales para los procesos de cuantificación de clase y de frecuencia fundamental.

La figura 8 es un diagrama de bloques que ilustra las asignaciones de bits para los procesos de cuantificación de clase y de frecuencia fundamental de acuerdo con una realización preferida de la presente invención.

15 Las figuras 9A, 9B y 9C son un diagrama de flujo funcional que muestra otro proceso de cuantificación de la frecuencia fundamental de acuerdo con una realización preferida de la presente invención.

La figura 10 es un diagrama de bloques de un sistema de procesamiento de información útil para implementar una realización preferida de la presente invención.

20 Descripción detallada

La presente invención, de acuerdo con una realización preferida, supera ventajosamente los problemas de la técnica anterior mediante la reducción efectiva del número de bits usados en la cuantificación de la frecuencia fundamental, como se discutirá en detalle a continuación.

25 I. Información general

La figura 1 es un diagrama de bloques que ilustra una red para el reconocimiento de voz distribuido (DSR) de acuerdo con una realización preferida de la presente invención. La figura 1 muestra un servidor de red o proveedor 30 102 de servicios inalámbricos que funciona en una red 104, que conecta el servidor/proveedor 102 de servicios inalámbricos con los clientes 106 y 108. En una realización de la presente invención, la figura 1 representa un sistema informático de red, que incluye un servidor 102, una red 104 y los equipos cliente 106 a 108. En una primera realización, la red 104 es una red de circuitos conmutados, tal como la red telefónica de servicio público (PSTN). Como alternativa, la red 104 es una red de conmutación de paquetes. La red de conmutación de paquetes es una 35 red de área amplia (WAN), tal como la Internet global, una WAN privada, una red de área local (LAN), una red de telecomunicaciones o cualquier combinación de las redes mencionadas. En otra alternativa, la red 104 es una red cableada, una red inalámbrica, una red de difusión o una red punto a punto.

40 En la primera realización, el servidor 102 y los equipos clientes 106 y 108 comprenden uno o más ordenadores personales (PC) (por ejemplo, las estaciones de trabajo IBM o los PC compatibles que ejecutan el sistema operativo de Microsoft Windows 95/98/2000/ME/CE/NT/XP, los equipos Macintosh que ejecutan el sistema operativo Mac OS, los PC que ejecutan el sistema operativo LINUX o equivalente), o cualquier otro dispositivo de procesamiento informático. Como alternativa, el servidor 102 y los equipos clientes 106 y 108 incluyen uno o más sistemas de 45 servidor (por ejemplo, las estaciones de trabajo Sun Ultra que ejecutan el sistema operativo SunOS o el sistema operativo AIX, las estaciones de trabajo IBM RS/6000 y los servidores que ejecutan el sistema operativo AIX o los servidores que ejecutan el sistema operativo LINUX).

50 En otra realización de la presente invención, la figura 1 representa un sistema de comunicación inalámbrico, que incluye un proveedor 102 de servicios inalámbricos, una red 104 inalámbrica y los dispositivos 106 a 108 inalámbricos. El proveedor 102 de servicios inalámbricos es un servicio de telefonía móvil analógica de primera generación, un servicio de telefonía móvil digital de segunda generación o un servicio de telefonía móvil con capacidad de internet de tercera generación.

55 En esta realización, la red 104 inalámbrica es una red de telefonía móvil, una red de dispositivos móviles de mensajería de texto, una red de buscapersonas, o similar. Además, la norma de comunicaciones de la red 104 inalámbrica de la figura 1 es el acceso múltiple por división de código (CDMA), el acceso múltiple por división de tiempo (TDMA), el sistema global para comunicaciones móviles (GSM), el servicio general de radiocomunicaciones por paquetes (GPRS), el acceso múltiple por división de frecuencia (FDMA) o similares. La red 104 inalámbrica 60 soporta cualquier número de dispositivos 106 a 108 inalámbricos, que son teléfonos móviles, dispositivos de mensajería de texto, equipos portátiles, buscapersonas, localizadores, o similares.

65 En esta realización, el proveedor 102 de servicios inalámbricos incluye un servidor, que comprende uno o más ordenadores personales (PC) (por ejemplo, las estaciones de trabajo IBM o los PC compatibles que ejecutan el sistema operativo de Microsoft Windows 95/98/2000/ME/CE/NT/XP, los equipos Macintosh que ejecutan el sistema operativo Mac OS, los PC que ejecutan el sistema operativo LINUX o equivalente), o cualquier otro dispositivo de procesamiento informático. En otra realización de la presente invención, el servidor 102 del proveedor de servicios

inalámbricos es uno o más sistemas de servidor (por ejemplo, las estaciones de trabajo Sun Ultra que ejecutan el sistema operativo SunOS o el sistema operativo AIX, las estaciones de trabajo IBM RS/6000 y los servidores que ejecutan el sistema operativo AIX o los servidores que ejecutan el sistema operativo LINUX).

5 Como se explicó anteriormente, el DSR se refiere a una trama en la que se distribuye la extracción de características y las partes de reconocimiento de patrones de un sistema de reconocimiento de voz. Esto es, la extracción de características y las partes de reconocimiento de patrones del sistema de reconocimiento de voz se realiza mediante dos unidades de procesamiento diferentes en dos localizaciones diferentes. Específicamente, el proceso de extracción de características se realiza por el extremo frontal, por ejemplo, los dispositivos 106 a 108 inalámbricos, y
10 el proceso de reconocimiento de patrones se realiza por el extremo trasero, por ejemplo, mediante un servidor 102 del proveedor de servicios inalámbricos. El proceso de extracción de características, tal como se realiza en el extremo frontal mediante los dispositivos 106 a 108 inalámbricos, se describe con mayor detalle a continuación.

15 La figura 2 es un diagrama de bloques detallado de un sistema de comunicación inalámbrico para el DSR de acuerdo con una realización preferida de la presente invención. La figura 2 es un diagrama de bloques más detallado del sistema de comunicación inalámbrico descrito con referencia a la figura 1 anterior. El sistema de comunicación inalámbrico de la figura 2 incluye un controlador 201 acoplado a las estaciones base 202, 203, y 204. Además, el sistema de comunicación inalámbrico de la figura 2 se interconecta a una red externa a través de una interfaz 206 telefónica. Las estaciones base 202, 203, y 204 soportan, individualmente, partes de un área de
20 cobertura geográfica que contiene las unidades de abonado o los transceptores (es decir, los dispositivos inalámbricos) 106 y 108 (véase la figura 1). Los dispositivos 106 y 108 inalámbricos interconectan con las estaciones base 202, 203 y 204 usando un protocolo de comunicación, tal como CDMA, FDMA, CDMA, GPRS y GSM.

25 El área de cobertura geográfica del sistema de comunicación inalámbrico de la figura 2 está dividida en regiones o celdas, que se suministran, individualmente, mediante las estaciones base 202, 203, y 204 (también denominadas en el presente documento como servidores de celda). Un dispositivo inalámbrico que funciona dentro del sistema de comunicación inalámbrico selecciona un servidor de celda específico como su interfaz principal para recibir y transmitir las funciones dentro del sistema. Por ejemplo, el dispositivo 106 inalámbrico tiene al servidor 202 de celda como su servidor de celda principal, y el dispositivo 108 inalámbrico tiene al servidor 204 de celda como su servidor
30 de celda principal. Preferentemente, un dispositivo inalámbrico selecciona un servidor de celda que proporciona la mejor interfaz de comunicación en el sistema de comunicación inalámbrico. Por lo común, esto dependerá de la calidad de la señal de las señales de comunicación entre un dispositivo inalámbrico y un servidor de celda específico.

35 Como un dispositivo inalámbrico se mueve entre diversas localizaciones geográficas en el área de cobertura, puede necesitarse una transferencia o traspaso a otro servidor de celda, que funcionará entonces como servidor de celda principal. Un dispositivo inalámbrico monitoriza las señales de comunicación de las estaciones base que suministran las celdas vecinas para determinar el nuevo servidor más adecuado para los propósitos de transferencia. Además de monitorizar la calidad de una señal transmitida desde un servidor de celda vecino, el dispositivo inalámbrico
40 también monitoriza la información del código de color transmitido asociado con la señal transmitida para identificar rápidamente qué servidor de celda vecino es la fuente de la señal transmitida.

45 La figura 3 es un diagrama de bloques que ilustra un dispositivo inalámbrico para un sistema de comunicación inalámbrico de acuerdo con una realización preferida de la presente invención. La figura 3 es un diagrama de bloques más detallado de un dispositivo inalámbrico descrito con referencia a las figuras 1 y 2 anteriores. La figura 3 muestra un dispositivo 106 inalámbrico, tal como se muestra en la figura 1. En una realización de la presente invención, el dispositivo 106 inalámbrico es una radio de dos vías capaz de recibir y transmitir señales de radiofrecuencia a través de un canal de comunicación conforme a un protocolo de comunicaciones, tal como el CDMA, FDMA, CDMA, GPRS o GSM. El dispositivo 106 inalámbrico funciona conforme al control de un controlador
50 302 que conmuta el dispositivo 106 inalámbrico entre los modos de recepción y transmisión. En el modo de recepción, el controlador 302 acopla una antena 316, a través de un interruptor 314 de transmisión/recepción, a un receptor 304. El receptor 304 decodifica las señales recibidas y proporciona esas señales decodificadas al controlador 302. En el modo de transmisión, el controlador 302 acopla la antena 316, a través del interruptor 314, a un transmisor 312.

55 El controlador 302 hace funcionar el transmisor y el receptor de acuerdo con las instrucciones almacenadas en la memoria 310. Las instrucciones almacenadas incluyen un algoritmo de programación de medición de la celda vecina. La memoria 310 es una memoria flash, otra memoria no volátil, una memoria de acceso aleatorio (RAM), una memoria de acceso aleatorio dinámico (DRAM) o similares. Un módulo 311 temporizador proporciona información de sincronización al controlador 302 para seguir los eventos programados. Además, el controlador 302 puede utilizar la información de tiempo del módulo 311 temporizador para seguir la programación de las transmisiones del servidor de celda vecina y la información del código de color transmitido.
60

65 Cuando se programa una medición de una celda vecina, el receptor 304, conforme al control del controlador 302, monitoriza los servidores de celda vecina y recibe un "indicador de calidad de la señal recibida" (RSQI). El circuito 308 RSQI genera señales de RSQI que representan la calidad de la señal de las señales transmitidas por cada

servidor de celda monitorizado. Cada señal RSQI se convierte en información digital mediante un conversor 306 analógico a digital y se proporciona como entrada al controlador 302. Usando la información de código de color y el indicador de calidad de la señal recibida asociado, el dispositivo 106 inalámbrico determina el servidor de celda vecina más apropiado para su uso como un servidor de celda principal cuando la transferencia es necesaria.

5 El procesador 320 en la figura 3 realiza diversas funciones tales como las funciones atribuidas al reconocimiento de voz distribuido, descrito en mayor detalle a continuación. En diversas realizaciones de la presente invención, el procesador 320 en la figura 3 es un procesador único o más de un procesador para realizar las tareas descritas anteriormente.

10 II. Reconocimiento de voz distribuido

La figura 4 es un diagrama de bloques que ilustra un intercambio de datos de un dispositivo 106 inalámbrico con el proveedor 102 de servicios inalámbricos. La figura 4 es un diagrama de bloques más detallado de un dispositivo 106 inalámbrico descrito con referencia a las figuras 1 y 2 anteriores. Se muestran también algunas de las funciones que se realizan en el dispositivo 106 inalámbrico. La figura 4 muestra un dispositivo 106 inalámbrico y la red 104 inalámbrica, tal como se muestra en la figura 1. También se muestran el procesador 320 y el transmisor 312 del dispositivo 106 inalámbrico, como se muestra en la figura 3.

20 En una realización de la presente invención, el dispositivo 106 inalámbrico incluye también un micrófono 404 para recibir el audio 402. A continuación, el audio 402 recibido se procesa por el procesador 320. Entre los procesos realizados por el procesador 320, se cuantifican la clase y la frecuencia fundamental de una trama de audio 402 por el procesador 320. La clase y la frecuencia fundamental de una trama de audio 402 se incorpora en al menos una palabra clave que se incluye en un paquete 406. A continuación, el paquete 406 se transmite mediante el transmisor 312 a través de la red 104 a un servidor o al proveedor 102 de servicios inalámbricos. Los procesos de cuantificación de la clase y de la frecuencia fundamental de una trama de audio 402 y la generación de, al menos, una palabra clave se describen con mayor detalle a continuación.

30 La figura 5 es un diagrama de flujo funcional que muestra el proceso de reconocimiento de voz distribuido global de acuerdo con una realización preferida de la presente invención. El diagrama de flujo funcional de la figura 5 representa el proceso, en un cliente 106, de realización de la extracción de características del audio de entrada y el proceso, en el servidor o en el proveedor 102 de servicios inalámbricos, de realización del reconocimiento de patrones. El diagrama de flujo funcional de la figura 5 comienza con la etapa 502 y fluye directamente a la etapa 504.

35 En la etapa 504, el cliente 106 recibe audio para su transmisión al servidor 102. En una realización en la que el sistema de la figura 1 representa una red inalámbrica, tal como se describe en la figura 2, en la etapa 504 el dispositivo 106 inalámbrico recibe el audio 402 (véase la figura 4) a través de un micrófono 404. A continuación, en la etapa 506, el cliente 106, procede a realizar la extracción de las características del audio que se recibió. La extracción de características incluye la cuantificación de los valores de la frecuencia fundamental y de la clase para una trama. La extracción de características se describe en mayor detalle a continuación. En la realización de la red inalámbrica, en la etapa 506, el procesador 320 (véase la figura 3) del dispositivo 106 inalámbrico procede a realizar la extracción de las características en el audio 402.

45 En la etapa 508, el cliente 106 genera el paquete de datos que incluye las características extraídas para su transmisión al servidor 102. La generación del paquete de datos se describe con mayor detalle a continuación. En la realización de la red inalámbrica, en la etapa 508, el dispositivo 106 inalámbrico genera el paquete 406 de datos que incluye las características extraídas. A continuación, en la etapa 510, el cliente 106, procede a transmitir el paquete de datos al servidor 102. En la realización de la red inalámbrica, en la etapa 510, el transmisor 312 del dispositivo 106 inalámbrico procede a transmitir el paquete 406 de datos al proveedor de la red inalámbrica o al servidor 102 a través de la red 104 inalámbrica.

50 En la etapa 512, el servidor 102 recibe el paquete de datos enviado por el cliente 106 y, en la etapa 514, el servidor 102 procede a reconstruir el audio en base al paquete de datos. En la etapa 516, el servidor 102 realiza el reconocimiento de voz en base al paquete de datos recibido desde el cliente 106. En la realización de la red inalámbrica, en la etapa 512, el proveedor de servicios inalámbricos o el servidor 102 recibe el paquete 406 de datos enviado por el dispositivo 106 inalámbrico y, en la etapa 514, el proveedor de servicios inalámbricos o el servidor 102 procede a reconstruir el audio en base al paquete 406 de datos. En la etapa 516, el proveedor de servicios inalámbricos o el servidor 102 realiza el reconocimiento de voz en base al paquete 406 de datos recibido desde el dispositivo 106 inalámbrico. En la etapa 518, el flujo funcional de la figura 5 se para.

III. Creación del paquete de datos

A. Primer esquema de cuantificación

65

En las propuestas de las normas DSR extendidas (en respuesta a los elementos de trabajo WI-030 y WI-034), la información (o sonoridad) de clase de una trama puede tomar cuatro valores posibles: 1) sin voz, es decir, silencio o ruido de fondo, 2) sorda, 3) sonora mixta y 4) sonora total. Estos cuatro valores se pueden dividir, generalmente, en dos categorías: no sonora (que incluye las clases sin voz y sorda) y sonora (que incluye las clases sonora mixta y sonora total). Para cuantificar la información de clase para una trama, se necesitan, normalmente, 2 bits ya que hay cuatro valores de clase que representar y 2 bits pueden representar cuatro valores. La información de la frecuencia fundamental de una trama puede tomar cualquier valor en el intervalo desde aproximadamente 19 a aproximadamente 140 muestras. Para cuantificar la información de la frecuencia fundamental de una trama, por ejemplo, como valores enteros, se necesitan, normalmente, 7 bits ya que hay 122 valores de la frecuencia fundamental que representar y 7 bits pueden representar 128 valores.

En una realización preferida de la presente invención, mediante la combinación de la información de clase y la información de la frecuencia fundamental, se puede ahorrar un bit por trama como sigue. Una de las palabras clave de la frecuencia fundamental de 7-bits (la palabra clave con todo ceros, por ejemplo) se usa para indicar que la clase es no sonora (es decir, bien sin voz o sorda). Las restantes 127 palabras clave de las frecuencias fundamentales no nulas de 7-bits se usan para representar los diferentes valores de la frecuencia fundamental, así como, la información de que la clase es sonora (es decir, bien sonora mixta o sonora total). Esto es, una de las palabras clave de la frecuencia fundamental de 7-bit indica que la clase es no sonora mientras que las palabras clave restantes indican que la clase es sonora. Por consiguiente, un bit de clase es ahora suficiente para eliminar la ambigüedad entre las dos clases no sonoras (es decir, entre las clases sin voz y sorda) y entre las dos clases sonoras (es decir, entre las clases sonora mixta o sonora total).

La siguiente tabla muestra un ejemplo de los valores de la palabra clave de la frecuencia fundamental de 7-bits y los valores de la palabra clave de 1-bit usados para indicar la información de la frecuencia fundamental y de clase, como se ha descrito anteriormente.

Tipo de Clase	Palabra clave de la frecuencia fundamental de 7-bits	Palabra clave de la clase de 1-bit
Sin voz	todas a cero	0
Sorda	todas a cero	1
Sonora mixta	ninguna a cero	0
Sonora total	ninguna a cero	1

De esta manera, en resumen, de acuerdo con el presente ejemplo, el número total de bits usados para representar la información de clase y de la frecuencia fundamental es de 8 bits. Esto está en contra de los 9 bits que habrían sido necesarios para representar la información de clase y de la frecuencia fundamental, si la información de clase y de la frecuencia fundamental se hubiese cuantificado por separado (es decir, 7-bits para la frecuencia fundamental y 2 bits para la clase; véase más arriba). La penalización provocada por este esquema es que una de las 128 palabras clave de la frecuencia fundamental se ha usado para indicar la información de clase. Debido a la relativa poca importancia de una palabra clave, esto tiene muy poco (e insignificante) impacto en la exactitud de la cuantificación de la frecuencia fundamental.

La figura 6 es un diagrama de flujo funcional que muestra un proceso de cuantificación de clase y de la frecuencia fundamental de acuerdo con una realización preferida de la presente invención. El diagrama de flujo funcional de la figura 6 indica el proceso, en un cliente 106, de cálculo de la información de clase y de la frecuencia fundamental y de generación de un paquete de datos que representa a los mismos. La figura 6 describe con más detalle las etapas 506 y 508 de la figura 5. El diagrama de flujo funcional de la figura 6 comienza con la etapa 602 y fluye, directamente, a la etapa 604.

En la etapa 604, el cliente 106 calcula el valor de la frecuencia fundamental del audio que se recibió. En la realización ejemplar de la red inalámbrica, en la etapa 604, el dispositivo 106 inalámbrico (más específicamente, el procesador 320 del dispositivo 106 inalámbrico) calcula el valor de la frecuencia fundamental del audio 402 que se recibió a través del micrófono 404. En la etapa 606, el cliente 106 realiza la cuantificación de la frecuencia fundamental en base al valor de la frecuencia fundamental del audio. En un ejemplo, se realiza la cuantificación de la frecuencia fundamental de 7-bits, como se ha descrito anteriormente. En la realización de la red inalámbrica, en la etapa 606, el dispositivo 106 inalámbrico realiza la cuantificación de la frecuencia fundamental en base al valor de la frecuencia fundamental del audio 402.

En la etapa 608, el cliente 106 realiza el cálculo de la clase del audio. En la etapa 610, el cliente 106 determina si la clase que se calculó en la etapa 608 es de la categoría no sonora, es decir, bien una clase sin voz o una clase sorda. Si el resultado de la determinación de la etapa 610 es positivo, entonces el control de flujo pasa a la etapa 612. Si el resultado de la determinación de la etapa 610 es negativo, entonces el control de flujo pasa a la etapa 616.

En la etapa 612, el cliente 106 pone una palabra clave de la frecuencia fundamental a una palabra clave de la frecuencia fundamental predefinida que indica una trama de categoría de clase no sonora (es decir, una clase sin voz o una clase sorda). En un ejemplo, el cliente 106 pone una palabra clave de la frecuencia fundamental de 7-bit a

todo ceros – la palabra clave de la frecuencia fundamental predefinida que indica una trama de clase sin voz o de clase sorda. En la etapa 614, el cliente 106 pone una palabra clave de clase para indicar la clase de una trama de categoría de clase no sonora (es decir, bien una clase sin voz o una clase sorda). En un ejemplo, el cliente 106 pone una palabra clave de clase de 1-bit que indica, bien una clase sin voz o una clase sorda.

5 En la etapa 616, el cliente 106 pone una palabra clave de la frecuencia fundamental al valor de la frecuencia fundamental generado cuantificando la frecuencia fundamental realizada en la etapa 604. En un ejemplo, el cliente 106 pone una palabra clave de la frecuencia fundamental de 7-bits al valor de la frecuencia fundamental generado por la cuantificación de la frecuencia fundamental realizada en la etapa 604. En la etapa 618, el cliente 106 pone una palabra clave de clase para indicar la clase de una trama de categoría de clase sonora (es decir, bien sonora mixta o sonora total). En un ejemplo, el cliente 106 pone una palabra clave de clase de 1-bit que indica bien una clase sonora mixta o una clase sonora total. En la etapa 620, el flujo funcional de la figura 6 se para.

15 La figura 7 es un diagrama de bloques que ilustra las asignaciones de bits convencionales para un proceso de cuantificación de la clase y de la frecuencia fundamental. La figura 7 muestra que siete bits representan la cuantificación 702 de la frecuencia fundamental. De esta manera, se usan 128 valores posibles para representar los datos de periodo de la frecuencia fundamental de una trama de audio. La figura 7 también muestra que dos bits representan la cuantificación 704 de clase. De esta manera, se usan los cuatro valores posibles para representar la clase de una trama de audio. Las cuatro clases se definen: sin voz, sorda, sonora mixta y sonora total. Por lo tanto, de acuerdo con la figura 7, se usan un total de nueve bits para representar los valores de cuantificación de la frecuencia fundamental y de clase para una trama de audio.

25 La figura 8 es un diagrama de bloques que indica las asignaciones de bits para el proceso de cuantificación de clase y de la frecuencia fundamental de acuerdo con una realización preferida de la presente invención. La figura 8 muestra las asignaciones de bits para la cuantificación de clase y de la frecuencia fundamental de acuerdo con la presente invención. La figura 8 muestra que siete bits representan la cuantificación 802 de la frecuencia fundamental. Sin embargo, de los 128 valores posibles disponibles con los siete bits, sólo se usan 127 valores para representar los datos de periodo de la frecuencia fundamental de una trama de audio. El valor restante se usa para indicar la información de clase, específicamente, la categoría de clase no sonora (es decir, la clase sin voz o la clase sorda). Preferentemente, el único valor 802 usado para indicar la categoría de clase es todo ceros, y este valor indica una trama de la clase de categoría no sonora (es decir, la clase sin voz o la clase sorda). Los otros 127 valores de 802 se usan para indicar el valor de la frecuencia fundamental de una trama de categoría sonora (es decir, la clase sonora mixta o la clase sonora total).

35 La figura 8 también muestra que un bit representa la cuantificación 804 de clase. Esto es una ventaja significativa de la presente invención. De esta manera, los dos valores posibles, o palabras clave, se usan adicionalmente para representar la clase de una trama de audio. Los dos valores posibles se usan para diferenciar entre la categoría de clase no sonora (es decir, la clase sin voz o la clase sorda) y entre la categoría de clase sonora (es decir, la clase sonora mixta o la clase sonora total). Preferentemente, un valor cero de 804 se usa para indicar una clase sin voz si 802 es todo ceros, un valor cero de 804 se usa para indicar una clase sonora mixta si 802 no es todo ceros, un valor de uno de 804 se usa para indicar una clase sorda si 802 es todo ceros, y un valor de uno de 804 se usa para indicar una clase sonora total si 802 no es todo ceros (véase la tabla anterior). Por lo tanto, de acuerdo con la figura 8, se usan un total de ocho bits para representar los valores de cuantificación de la frecuencia fundamental y de clase para una trama de audio.

45 B. Segundo esquema de cuantificación

50 En las propuestas de las normas DSR extendidas, se estima el periodo de la frecuencia fundamental para cada trama y se actualiza cada 10 ms (o 100 veces por segundo). El periodo de la frecuencia fundamental estimada puede tomar valores fraccionados e intervalos desde aproximadamente 19 a aproximadamente 140 muestras a una velocidad de muestreo de 8 kHz. Por lo tanto, la frecuencia fundamental estimada oscila desde aproximadamente 57 Hz a aproximadamente 420 Hz. Cuando se realiza el proceso de cuantificación de la frecuencia fundamental, es deseable para lograr la exactitud, es decir, un error de cuantificación mínimo, una velocidad de bits baja, y robustez frente a errores de canal.

55 En una realización preferida de la presente invención, todas las tramas numeradas como pares (empezando con la primera trama numerada a cero) se cuantifican usando 7 bits y todas las tramas numeradas como impares se cuantifican usando 5 bits. Consideradas como pares de trama, la primera trama en un par de trama que se cuantifica usando 7 bits y la segunda trama en un par de trama que se cuantifica usando 5 bits. De esta manera, el número medio de bits por trama es de seis. Esto corresponde a una velocidad de bits de 600 bps debido, únicamente, a la cuantificación de la frecuencia fundamental.

65 Se usa un esquema de cuantificación absoluta para las tramas numeradas como pares. Fuera de los 128 palabras clave disponibles, una palabra clave (es decir, la palabra clave todo ceros) se usa para la transmisión de información de la clase de categoría no sonora, es decir, para indicar que la trama es sin voz o sorda. Las restantes 127 palabras clave se usan para la cuantificación del periodo de la frecuencia fundamental. Este esquema se ha descrito

con mayor detalle anteriormente. El intervalo de la frecuencia fundamental desde aproximadamente 19 a aproximadamente 140 muestras se divide por igual (en el dominio logarítmico) en 127 regiones y los puntos medios de estas regiones se eligen como los niveles de reconstrucción. Para cualquier valor dado de la frecuencia fundamental, se elige el valor de la frecuencia fundamental cuantificada correspondiente como el nivel de reconstrucción más cercano en el dominio lineal. Las 127 palabras clave se asignan una a una a los 127 niveles de reconstrucción. El error de cuantificación máximo con este diseño cuantificador es de aproximadamente el 0,8%.

Para las tramas numeradas como impares, se usa, la mayoría de las veces, un esquema de cuantificación diferencial. Sin embargo, en determinadas situaciones (como se muestra, a continuación, en la tabla), también se usa un esquema de cuantificación absoluta. Para el esquema de cuantificación diferencial, tiene que seleccionarse una referencia de modo que la diferencia entre el valor del periodo de la frecuencia fundamental de la trama actual y el valor de referencia (o más apropiadamente, la relación de los dos valores) pueda cuantificarse. Aunque el periodo de la frecuencia fundamental cuantificada de la trama anterior proporciona la mejor referencia posible, esta trama puede que no siempre sea una categoría de clase sonora (es decir, bien sonora mixta o sonora total). Por lo tanto, el valor del periodo de la frecuencia fundamental cuantificada de una de las tres tramas anteriores, se selecciona como referencia. Esto es, la cuantificación diferencial de la segunda trama del par de trama se realiza usando el valor del periodo de la frecuencia fundamental cuantificada de la primera trama del par de trama o cualquiera de los dos valores del periodo de la frecuencia fundamental cuantificados del par de trama anterior como referencia.

En el lado del servidor, es importante limitar la propagación de errores debidos a un error en uno de los valores del periodo de la frecuencia fundamental descodificada. Para este propósito, se identifica cada valor de la frecuencia fundamental cuantificada en el lado del cliente como que puede ser fiable (R) o no fiable (U) para servir como una referencia. Cada valor de la frecuencia fundamental cuantificada de manera absoluta se considera como fiable. Cada valor de la frecuencia fundamental cuantificada de manera diferencial se considera fiable si la referencia usada para su cuantificación es el valor del periodo de la frecuencia fundamental cuantificada de la primera trama del mismo par de trama. Ya que los valores del periodo de la frecuencia fundamental de las tramas vecinas están, generalmente, cercanos entre sí, los valores del periodo de la frecuencia fundamental cercanos al valor de referencia se cuantifican de una manera fina y los valores del periodo de la frecuencia fundamental más lejanos al valor de referencia se cuantifican de una manera tosca. Los niveles de cuantificación elegidos para la cuantificación diferencial dependerán de cuál de los tres valores de la frecuencia fundamental cuantificados anteriores se ha elegido como referencia, así como el valor de referencia. La siguiente tabla indica cómo se cuantifican los valores del periodo de la frecuencia fundamental de las tramas numeradas como impares.

P(-2)	P(-1)	P(0)	Acción Tomada
0	0	0	Cuantificación absoluta
0	1U	0	Cuantificación absoluta
*	*	1	Cuantificación diferencial: Referencia P(0)
*	1R	0	Cuantificación diferencial: Referencia P(-1)
1	0	0	Cuantificación diferencial: Referencia P(-2)
1	1U	0	Cuantificación diferencial: Referencia P(-2)

En la tabla anterior, el valor que va a cuantificarse es el P(1), el valor del periodo de la frecuencia fundamental de la segunda trama de un par de trama. El valor de referencia es el valor del periodo de la frecuencia fundamental cuantificada de una de las tres tramas anteriores, es decir, P(0), el valor del periodo de la frecuencia fundamental cuantificada de la primera trama del mismo par de trama, P(-1), el valor del periodo de la frecuencia fundamental cuantificada de la segunda trama del par de trama anterior, y P(-2), el valor del periodo de la frecuencia fundamental cuantificada de la primera trama del par de trama anterior.

En la tabla, un valor de "0" indica que la trama correspondiente es una clase de categoría no sonora (es decir, la clase sin voz o la clase sorda). Un valor de "1" indica que la trama correspondiente es una categoría de clase sonora (es decir, la clase sonora mixta o la clase sonora total) y su valor del periodo de la frecuencia fundamental cuantificada puede usarse como una referencia. Para el P(-1) cuantificado diferencial, también se tiene "1R" y "1U" para indicar si el valor del periodo de la frecuencia fundamental cuantificada es fiable o no fiable, respectivamente. Un "*" indica que el periodo de la frecuencia fundamental cuantificada es intrascendente, es decir, el valor puede ser un "0" o un "1", y no hace una diferencia. Como condiciones iniciales, se supone que P(-1) = 0 y P(-2) = 0, tanto en el codificador (es decir, el cliente 106) como en el descodificador (es decir, el servidor 102). La última columna indica si la frecuencia fundamental se cuantificó de manera absoluta o de manera diferencial, y si fue de manera diferencial, la trama de referencia que se usó.

Cuando las tres tramas anteriores son una clase de categoría no sonora o cuando el único valor de referencia disponible es el P(-1) no fiable, P(1) se cuantifica de manera absoluta usando 5 bits. Una palabra clave, tal como la palabra clave todo ceros, se usa para indicar que la trama es de una clase de categoría no sonora. Las restantes 31 palabras clave se usan para cuantificar el periodo P(1) de la frecuencia fundamental de una manera similar a la

usada para cuantificar los valores del periodo de la frecuencia fundamental de las tramas numeradas como pares. La opción de cuantificación absoluta que usa 5 bits se elige, típicamente, para la primera trama de un segmento sonoro o para algunas tramas mal clasificadas que pertenecen a una condición de fondo ruidoso. En ambos casos, el error de cuantificación ligeramente más grande que resulta del uso de sólo 5 bits no provoca ninguna pérdida significativa de la calidad de voz o la inteligibilidad. El uso de sólo 5 bits ayuda a limitar la propagación de los errores de descodificación como se explica más adelante.

Cuando la primera trama de un par de trama es de una clase de categoría sonora, entonces el correspondiente valor del periodo de la frecuencia fundamental cuantificada se elige siempre como referencia, sin tener en cuenta los valores de $F(-1)$ y $P(-2)$. De acuerdo con una realización ejemplar, de las 32 palabras clave posibles (usando 5 bits de cuantificación del valor del periodo de la frecuencia fundamental), una palabra clave, tal como la palabra clave todo ceros, se usa para indicar que la trama actual es sin voz / sorda. Se usan veintisiete palabras clave para cubrir un intervalo pequeño de la frecuencia fundamental alrededor del valor de referencia de una manera logarítmica (similar a la cuantificación absoluta de 7-bits discutida anteriormente). Los dos extremos del intervalo de la frecuencia fundamental representan los niveles de reconstrucción. Los restantes cuatro niveles se usan para cuantificar de manera tosca el resto del intervalo de la frecuencia fundamental como se indica en la tabla. Nótese que los cuatro niveles elegidos dependen del valor de $P(0)$. Por ejemplo, si $P(0)$ es pequeño, entonces los cuatro niveles son mayores que $P(0)$. Por otra parte, si $P(0)$ es grande, entonces los cuatro niveles son más pequeños que $P(0)$.

Cuando la primera trama de un par de trama es de la clase de categoría no sonora, entonces se elige o bien $P(-1)$ o $P(-2)$ como referencia. Si $P(-1)$ corresponde a una trama de la clase de categoría sonora y es fiable, entonces se elige como referencia, independientemente del valor de $P(-2)$. Si $P(-1)$ corresponde a una trama de la clase de categoría no sonora o corresponde a una trama de la clase de categoría sonora pero es no fiable, y $P(-2)$ corresponde a una trama de la clase de categoría sonora, entonces $P(-2)$ se elige como referencia. Si $P(-1)$ o $P(-2)$ se eligen como referencia, el método de cuantificación es similar. Una de las palabras clave, tal como la palabra clave todo ceros, se usa para indicar que la trama actual es de la clase de categoría no sonora. Se usan veinticinco palabras clave para cubrir un intervalo pequeño de la frecuencia fundamental alrededor del valor de referencia de una manera logarítmica (similar a la cuantificación absoluta de 7-bits discutida anteriormente). Los dos valores finales del intervalo de la frecuencia fundamental representan los niveles de reconstrucción. Los restantes seis niveles se usan para cuantificar de manera tosca el resto del intervalo de la frecuencia fundamental.

El esquema de cuantificación anterior satisface los requisitos de exactitud, velocidad de bits baja y de solidez de la siguiente manera. Cuantificando los valores del periodo de la frecuencia fundamental de las tramas numeradas como pares con 7 bits y las de las tramas numeradas como impares con 5 bits, se ahorra un promedio de 1 bit por trama, es decir, 100 bits por segundo. Al mismo tiempo, la exactitud no se ve comprometida. La cuantificación absoluta de siete bits es suficientemente exacta. La cuantificación absoluta de cinco bits se usa, típicamente, para la primera trama de un segmento sonoro y para algunas tramas de fondo ruidoso. En ambos casos, la falta de exactitud no es crítica y no afecta a la calidad o la inteligibilidad de la voz reconstruida de ninguna manera significativa. Con la cuantificación diferencial de 5-bits, los valores del periodo de la frecuencia fundamental, que se acercan al valor de referencia, se cuantifican con bastante exactitud. Estos son los valores del periodo de la frecuencia fundamental de alta probabilidad. Los valores del periodo de la frecuencia fundamental, que están más lejos del valor de referencia son de baja probabilidad y se cuantifican de manera tosca. Una vez más, el error más grande en la cuantificación de estos valores no es crítico y no afecta significativamente a la calidad o inteligibilidad de la voz reconstruida.

La propagación del error en la presente invención se limita identificando los valores del periodo de la frecuencia fundamental cuantificada de manera diferencial como fiables y no fiables y mediante el uso de la cuantificación absoluta de 5-bit de las tramas numeradas como impares cuando no hay disponible un valor de referencia, o el único valor de referencia disponible es no fiable. Por ejemplo, considérese la situación en la que se han borrado un número pares de tramas. Este es el tipo más común de situación de error de canal para un canal DSR. Supóngase que los bits correspondientes a las tramas que siguen a las borradas se han recibido correctamente. Si la primera trama del par de trama que sigue a las borradas es una trama sonora, entonces no hay, en absoluto, propagación del error. Esto es debido a que la primera trama siempre se cuantifica de manera absoluta (usando 7 bits) y la segunda trama se cuantifica de manera diferencial usando el valor del periodo de la frecuencia fundamental cuantificada de la primera trama como referencia. También, las tramas siguientes no dependen de ninguna de las tramas borradas.

Si la primera trama es de una clase de categoría no sonora, entonces la segunda trama no puede decodificarse correctamente a menos que sea también una clase de categoría no sonora. Esto es porque el segundo valor de la frecuencia fundamental de la trama se podría haber cuantificado de manera diferencial usando el valor de la frecuencia fundamental cuantificada de una de las últimas tramas borradas como referencia. En este caso, el error se propaga a la segunda trama que sigue a las borradas.

Si la tercera trama es de una clase de categoría sonora, entonces la propagación del error cesa debido a los valores del periodo de la frecuencia fundamental cuantificada de todas las tramas que siguen a la tercera trama que no

dependen de las tramas borradas o del par de trama recibido correctamente que sigue a las borradas. Si la tercera trama es de una clase de categoría no sonora, entonces el valor del periodo de la frecuencia fundamental cuantificada de la cuarta trama puede descodificarse con éxito, porque debe haberse cuantificado de manera absoluta dado que las tramas primera y tercera son de una clase de categoría no sonora y la segunda trama es no fiable. Por lo tanto, la propagación del error que sigue al borrado de uno o más pares de trama cesa después de dos tramas a lo sumo.

De manera similar, se puede demostrar que cualquier error en el valor del periodo de la frecuencia fundamental decodificada de una trama numerada par (debido a los errores de bit aleatorios) puede propagarse hasta tres tramas a lo sumo. Además, cualquier error en el valor del periodo de la frecuencia fundamental decodificada de una trama numerada impar (debido a los errores de bit aleatorios) puede propagarse hasta dos tramas a lo sumo.

Las figuras 9A, 9B y 9C son un diagrama de flujo funcional que muestra otro proceso de cuantificación de la frecuencia fundamental de acuerdo con una realización preferida de la presente invención. El diagrama de flujo funcional de las figuras 9A, 9B y 9C indica el proceso, en un cliente 106, de cálculo de la información de la frecuencia fundamental de una trama, de generación de un paquete de datos que representa a la misma y de continuar con la siguiente trama. Las figuras 9A, 9B y 9C describen con más detalle las etapas 506 y 508 de la figura 5. El diagrama de flujo funcional de las figuras 9A, 9B y 9C comienza con la etapa 902 (en la figura 9A) y fluye, directamente, a la etapa 904.

En la etapa 903, el cliente 106 calcula el valor de la frecuencia fundamental del audio para la trama actual. En la realización de la red inalámbrica, en la etapa 903, el dispositivo 106 inalámbrico (más específicamente, el procesador 320 del dispositivo 106 inalámbrico) calcula el valor de la frecuencia fundamental del audio 402 que se recibió a través del micrófono 404. En la etapa 904, el cliente 106 determina si la trama actual es una trama par o impar. Si el resultado de la determinación de la etapa 904 es par, entonces el control fluye a la etapa 910. Si el resultado de la determinación de la etapa 904 es impar, entonces el control fluye a la etapa 905. En la etapa 905, la trama actual es una trama impar y de esta manera, el cliente 106 procede a buscar una trama de referencia adecuada para utilizarse para la cuantificación de la frecuencia fundamental diferencial. En la etapa 906, el control fluye, directamente, a la etapa 916 (B) de la figura 9 B.

En la etapa 910, el cliente 106 realiza la cuantificación de la frecuencia fundamental de manera absoluta en base al valor de la frecuencia fundamental del audio. En un ejemplo, la cuantificación de la frecuencia fundamental de manera absoluta de 7-bits se realiza, como se ha descrito anteriormente. En la realización de la red inalámbrica, en la etapa 910, el dispositivo 106 inalámbrico realiza la cuantificación de la frecuencia fundamental de manera absoluta en base al valor de la frecuencia fundamental del audio 402.

En la etapa 912, el cliente 106 pone una palabra clave de la frecuencia fundamental al valor de la frecuencia fundamental generado por la cuantificación de la frecuencia fundamental de manera absoluta realizada en la etapa 910. En un ejemplo, el cliente 106 pone una palabra clave de la frecuencia fundamental de 7-bits al valor de la frecuencia fundamental generado por la cuantificación de la frecuencia fundamental de manera absoluta realizada en la etapa 910. En la etapa 915 (E), el control fluye directamente a la etapa 914. En la etapa 914, el proceso de cuantificación de la frecuencia fundamental avanza hacia la trama siguiente y el control fluye, directamente, a la etapa 903.

En la etapa 916 (B) de la figura 9B, el control fluye, directamente, a la etapa 917. En la etapa 917, el cliente 106 determina si la clase de la trama inmediatamente anterior a la trama "0" actual es de la clase de categoría sonora (es decir, la clase sonora mixta o la clase sonora total). Téngase en cuenta que en las figuras 9B y 9C, la trama actual se designa trama "0", la trama inmediatamente anterior a la trama "0" es la trama "-1", la trama inmediatamente anterior a la trama "-1" es la trama "-2" y la trama inmediatamente anterior a la trama "-2" es la trama "-3". Si el resultado de la determinación de la etapa 917 es positivo, entonces el control fluye a la etapa 940. Si el resultado de la determinación de la etapa 917 es negativo, entonces el control fluye a la etapa 920. En la etapa 920, el cliente 106 avanza hacia la trama anterior para continuar buscando una trama de referencia adecuada para utilizarse para la cuantificación de la frecuencia fundamental de manera diferencial.

En la etapa 927, el cliente 106 determina si la clase de trama "-2" es de la clase de categoría sonora (es decir, la clase sonora mixta o la clase sonora total). Si el resultado de la determinación de la etapa 927 es positivo, entonces el control fluye a la etapa 928. Si el resultado de la determinación de la etapa 927 es negativo, entonces el control fluye a la etapa 930. En la etapa 928, el cliente 106 determina si el valor de la frecuencia fundamental de la trama "-2" se cuantificó de manera absoluta. Si el resultado de la determinación de la etapa 928 es positivo, entonces el control fluye a la etapa 940. Si el resultado de la determinación de la etapa 928 es negativo, entonces el control fluye a la etapa 929. En la etapa 929, el cliente 106 determina si el valor de la frecuencia fundamental de la trama "-2" se cuantificó de manera diferencial y es fiable (esto es, se cuantificó de manera diferencial y se referenció la trama inmediatamente anterior). Si el resultado de la determinación de la etapa 929 es positivo, entonces el control fluye a la etapa 940. Si el resultado de la determinación de la etapa 929 es negativo, entonces el control fluye a la etapa 930.

En la etapa 930, el cliente 106 avanza hacia la trama anterior para continuar buscando una trama de referencia adecuada para utilizarse para la cuantificación diferencial de la frecuencia fundamental. En la etapa 937, el cliente 106 determina si la clase de trama "-3" es de la clase de categoría sonora (es decir, la clase sonora mixta o la clase sonora total). Si el resultado de la determinación de la etapa 937 es positivo, entonces el control fluye a la etapa 940. Si el resultado de la determinación de la etapa 937 es negativo, entonces el control fluye a la etapa 942. La etapa 940 fluye, directamente, a la etapa 960 (C) de la figura 9C y la etapa 942 fluye, directamente, a la etapa 950 (D) de la figura 9C.

En la etapa 950 (D) de la figura 9C, el control fluye, directamente, a la etapa 952. En la etapa 952, se determina que no se ha encontrado una trama de referencia adecuada para la cuantificación diferencial de la trama "0" actual. En la etapa 956, el cliente 106 realiza la cuantificación de la frecuencia fundamental de manera absoluta en base al valor de la frecuencia fundamental del audio. En un ejemplo, la cuantificación de la frecuencia fundamental de manera absoluta de 5-bits se realiza, como se ha descrito anteriormente. En la realización de la red inalámbrica, en la etapa 956, el dispositivo 106 inalámbrico realiza la cuantificación de la frecuencia fundamental de manera absoluta en base al valor de la frecuencia fundamental del audio 402. En la etapa 958, el cliente 106 pone una palabra clave al valor de la frecuencia fundamental generada por la cuantificación de la frecuencia fundamental de manera absoluta realizada en la etapa 956. En un ejemplo, el cliente 106 pone una palabra clave de la frecuencia fundamental de 5-bits al valor de la frecuencia fundamental generado por la cuantificación de la frecuencia fundamental de manera absoluta realizada en la etapa 956.

En la etapa 960 (C) de la figura 9C, el control fluye directamente a la etapa 962. En la etapa 962, se determina que se ha encontrado una trama de referencia adecuada para la cuantificación diferencial de la trama "0" actual. En la etapa 966, el cliente 106 realiza la cuantificación de la frecuencia fundamental de manera diferencial que hace referencia a la trama de referencia identificada. En un ejemplo, se realiza la cuantificación de la frecuencia fundamental de manera diferencial de 5-bits, como se ha descrito anteriormente. En la etapa 968, el cliente 106 pone una palabra clave de la frecuencia fundamental al valor de la frecuencia fundamental generado por la cuantificación diferencial realizada en la etapa 966. En un ejemplo, el cliente 106 pone una palabra clave de la frecuencia fundamental de 5-bits al valor de la frecuencia fundamental generado por la cuantificación de la frecuencia fundamental de manera diferencial realizada en la etapa 966.

En la etapa 970, el control fluye, directamente, a la etapa 915 (E) de la figura 9A. En la etapa 915 (E), el control fluye, directamente, a la etapa 914. En la etapa 914, el proceso de cuantificación de la frecuencia fundamental avanza hacia la trama siguiente y el control fluye, directamente, de vuelta a la etapa 903.

C. Revisión de la técnica anterior

En la norma de predicción lineal de excitación mixta (MELP) (un norma de telecomunicaciones), no hay distinción entre las tramas sin voz y sordas. Ambas clases se combinan entre sí y se indican mediante un valor del periodo de la frecuencia fundamental de cero. Se usan 4 bits adicionales para cuantificar la información de clase cuando el periodo de la frecuencia fundamental es mayor que cero, es decir, cuando una trama es de la clase de categoría sonora (por ejemplo, sonora mixta o sonora total). Estos 4 bits identifican la sonoridad en diferentes bandas del espectro del habla. El valor de la frecuencia fundamental se cuantifica de manera absoluta usando 7 bits. Por lo tanto, no hay un ahorro de bits en MELP, tal como se describe en la presente invención.

En LPC-10 (otra norma de telecomunicaciones), se usan 7 bits para indicar una trama de la clase de categoría sonora y la frecuencia fundamental. Hay 60 niveles de periodo de la frecuencia fundamental y 3 niveles usados para indicar que: 1) las dos medias tramas son de la clase de categoría no sonora (es decir, una clase sin voz y una clase sorda), 2) sólo la segunda mitad de la trama es de la clase de categoría sonora (es decir, clase sonora mixta y la clase sonora total) o 3) sólo la primera mitad de la trama es de la clase de categoría sonora. Por lo tanto, LPC-10 sólo distingue entre la clase de categoría sonora y la clase de categoría sorda. LPC-10 no distingue entre la clase de categoría sonora (es decir, entre la clases sin voz y sorda) o entre la clase de categoría sorda (es decir, entre las clases sonora mixta y sonora total). La presente invención extiende LPC-10 con la introducción de las clases sin voz y sorda conforme a la clase de categoría no sonora y las clases sonora mixta y sonora total conforme a las clases de categoría sonora.

IV. Implementaciones ejemplares

La presente invención se puede realizar en un soporte físico, en un soporte lógico, o en una combinación de soporte físico y soporte lógico en los clientes 106, 108, o el servidor 102 de la figura 1. Puede realizarse un sistema de acuerdo con una realización preferida de la presente invención, tal como se describe en las figuras 5, 6, 9A, 9B y 9C, de una manera centralizada en un sistema informático, o de una manera distribuida en la que los distintos elementos se dispersan en diversos sistemas informáticos interconectados. Cualquier tipo de sistema informático - u otro aparato adaptado para llevar a cabo los métodos descritos en el presente documento - es adecuado. Una combinación típica de soporte físico y soporte lógico podría ser un sistema informático de propósito general con un programa informático que, cuando está cargado y ejecutado, controla el sistema informático de tal manera que realiza los métodos descritos en el presente documento.

- Una realización de la presente invención puede incorporarse también en un producto de programa informático (en los clientes 106 y 108 y el servidor 102), que comprende todas las características que permiten la implementación de los métodos descritos en el presente documento, y que, cuando se cargan en un sistema informático, es capaz de llevar a cabo estos métodos. Los medios de programa informático o el programa informático tal como se usa en la presente invención indica cualquier expresión, en cualquier lenguaje, código o notación, de un conjunto de instrucciones destinadas a hacer que un sistema que tiene una capacidad de procesamiento de información realice una función específica, bien directamente o después de una o ambas de las siguientes a) la conversión a otro lenguaje, código o notación; y b) la reproducción en una forma material diferente.
- 10 Un sistema informático puede incluir, entre otras cosas, uno o más equipos y al menos un medio legible por ordenador, lo que permite a un sistema informático, leer datos, instrucciones, mensajes o paquetes de mensajes y otra información legible por ordenador desde el medio legible por ordenador. El medio legible por ordenador puede incluir una memoria no volátil, tal como la ROM, memoria Flash, memoria de la unidad de disco, unidad de CD-ROM y otro almacenamiento permanente. Además, un medio legible por ordenador puede incluir, por ejemplo, el
- 15 almacenamiento volátil, tal como la RAM, memoria intermedia, memoria caché y circuitos de red. Adicionalmente, el medio legible por ordenador puede comprender información legible por ordenador en un medio de estado transitorio tal como un enlace de red y/o una interfaz de red, incluyendo una red cableada o una red inalámbrica, que permiten que un sistema informático lea tal información legible por ordenador.
- 20 La figura 10 es un diagrama de bloques de un sistema informático útil para implementar una realización de la presente invención. El sistema informático de la figura 10 es una representación más detallada de los clientes 106 y 108 y del servidor 102. El sistema informático de la figura 10 incluye uno o más procesadores, tal como el procesador 1004. El procesador 1004 está conectado a una infraestructura 1002 de comunicación (por ejemplo, un bus de comunicaciones, una barra sobrecruzada, o una red). Se describen diversas realizaciones de soporte lógico en términos de este sistema informático ejemplar. Después de leer esta descripción, se hará evidente para un experto en la materia cómo implementar la invención usando otros sistemas informáticos y/o arquitecturas informáticas.
- 25 El sistema informático puede incluir una interfaz 1008 de visualización que envía gráficos, texto y otros datos desde la infraestructura 1002 de comunicación (o desde una memoria intermedia de trama no mostrada) para visualizar en la unidad 1010 de visualización. El sistema informático también incluye una memoria 1006 principal, preferentemente memoria de acceso aleatorio (RAM), y también puede incluir una memoria 1012 secundaria. La memoria 1012 secundaria puede incluir, por ejemplo, una unidad de disco 1014 duro y/o una unidad 1016 de almacenamiento removible, que representa una unidad de disquete, una unidad de cinta magnética, una unidad de disco óptico, etc. La unidad 1016 de almacenamiento removible lee y/o escribe en una unidad de almacenamiento 1018 removible de una manera bien conocida por los expertos en la materia. La unidad de almacenamiento 1018 removible, representa un disquete, una cinta magnética, un disco óptico, etc., que se lee y se escribe mediante la unidad 1016 de almacenamiento removible. Como se apreciará, la unidad de almacenamiento 1018 removible incluye un medio de almacenamiento usable informático que tiene almacenado en él mismo el soporte lógico informático y/o los datos.
- 30 El sistema informático puede incluir una interfaz 1008 de visualización que envía gráficos, texto y otros datos desde la infraestructura 1002 de comunicación (o desde una memoria intermedia de trama no mostrada) para visualizar en la unidad 1010 de visualización. El sistema informático también incluye una memoria 1006 principal, preferentemente memoria de acceso aleatorio (RAM), y también puede incluir una memoria 1012 secundaria. La memoria 1012 secundaria puede incluir, por ejemplo, una unidad de disco 1014 duro y/o una unidad 1016 de almacenamiento removible, que representa una unidad de disquete, una unidad de cinta magnética, una unidad de disco óptico, etc. La unidad 1016 de almacenamiento removible lee y/o escribe en una unidad de almacenamiento 1018 removible de una manera bien conocida por los expertos en la materia. La unidad de almacenamiento 1018 removible, representa un disquete, una cinta magnética, un disco óptico, etc., que se lee y se escribe mediante la unidad 1016 de almacenamiento removible. Como se apreciará, la unidad de almacenamiento 1018 removible incluye un medio de almacenamiento usable informático que tiene almacenado en él mismo el soporte lógico informático y/o los datos.
- 35 En realizaciones alternativas, la memoria 1012 secundaria puede incluir otros medios similares para permitir que los programas informáticos u otras instrucciones se carguen en el sistema informático. Tales medios pueden incluir, por ejemplo, una unidad 1022 de almacenamiento removible y una interfaz 1020. Entre los ejemplos de tales medios se pueden incluir un programa de cartucho y una interfaz de cartucho (tal como la que se encuentra en los dispositivos de los videojuegos), un chip de memoria removible (tal como una EPROM o PROM) y el zócalo asociado, y otras unidades 1022 de almacenamiento removible e interfaces 1020 que permiten que el soporte lógico y los datos se transfieran desde la unidad 1022 de almacenamiento removible al sistema informático.
- 40 El sistema informático también puede incluir una interfaz 1024 de comunicaciones. La interfaz 1024 de comunicaciones permite que el soporte lógico y los datos se transfieran entre el sistema informático y los dispositivos externos. Entre los ejemplos de la interfaz 1024 de comunicaciones se pueden incluir un módem, una interfaz de red (como una tarjeta Ethernet), un puerto de comunicaciones, una ranura y la tarjeta PCMCIA, etc. El soporte lógico y los datos transferidos a través de la interfaz 1024 de comunicaciones están en forma de señales que pueden ser, por ejemplo, electrónicas, electromagnéticas, ópticas u otras señales capaces de recibirse mediante la interfaz 1024 de comunicaciones. Estas señales se proporcionan a la interfaz 1024 de comunicaciones a través de una ruta 1026 de comunicaciones (es decir, un canal). Este canal 1026 lleva las señales y puede implementarse usando alambre o cable, fibra óptica, una línea telefónica, un enlace de teléfono móvil, un enlace de RF y/u otros canales de comunicaciones.
- 45 En este documento, los términos "medio de programa informático", "medio usable por ordenador", "medio legible por máquina" y "medio legible por ordenador" se usan para referirse, generalmente, a medios tales como la memoria 1006 principal y la memoria 1012 secundaria, la unidad 1016 de almacenamiento removible, un disco duro instalado en la unidad 1014 de comunicaciones (es decir, un canal). Este canal 1026 lleva las señales y puede implementarse usando alambre o cable, fibra óptica, una línea telefónica, un enlace de teléfono móvil, un enlace de RF y/u otros canales de comunicaciones.
- 50 En este documento, los términos "medio de programa informático", "medio usable por ordenador", "medio legible por máquina" y "medio legible por ordenador" se usan para referirse, generalmente, a medios tales como la memoria 1006 principal y la memoria 1012 secundaria, la unidad 1016 de almacenamiento removible, un disco duro instalado en la unidad 1014 de comunicaciones (es decir, un canal). Este canal 1026 lleva las señales y puede implementarse usando alambre o cable, fibra óptica, una línea telefónica, un enlace de teléfono móvil, un enlace de RF y/u otros canales de comunicaciones.
- 55 En este documento, los términos "medio de programa informático", "medio usable por ordenador", "medio legible por máquina" y "medio legible por ordenador" se usan para referirse, generalmente, a medios tales como la memoria 1006 principal y la memoria 1012 secundaria, la unidad 1016 de almacenamiento removible, un disco duro instalado en la unidad 1014 de comunicaciones (es decir, un canal). Este canal 1026 lleva las señales y puede implementarse usando alambre o cable, fibra óptica, una línea telefónica, un enlace de teléfono móvil, un enlace de RF y/u otros canales de comunicaciones.
- 60 En este documento, los términos "medio de programa informático", "medio usable por ordenador", "medio legible por máquina" y "medio legible por ordenador" se usan para referirse, generalmente, a medios tales como la memoria 1006 principal y la memoria 1012 secundaria, la unidad 1016 de almacenamiento removible, un disco duro instalado en la unidad 1014 de comunicaciones (es decir, un canal). Este canal 1026 lleva las señales y puede implementarse usando alambre o cable, fibra óptica, una línea telefónica, un enlace de teléfono móvil, un enlace de RF y/u otros canales de comunicaciones.
- 65 En este documento, los términos "medio de programa informático", "medio usable por ordenador", "medio legible por máquina" y "medio legible por ordenador" se usan para referirse, generalmente, a medios tales como la memoria 1006 principal y la memoria 1012 secundaria, la unidad 1016 de almacenamiento removible, un disco duro instalado en la unidad 1014 de comunicaciones (es decir, un canal). Este canal 1026 lleva las señales y puede implementarse usando alambre o cable, fibra óptica, una línea telefónica, un enlace de teléfono móvil, un enlace de RF y/u otros canales de comunicaciones.

5 desde el medio legible por ordenador. El medio legible por ordenador, por ejemplo, puede incluir memoria no volátil, tal como un disquete, ROM, memoria Flash, memoria de la unidad de disco, unidad de CD-ROM, y otro almacenamiento permanente. Es útil, por ejemplo, para transportar información, tal como datos e instrucciones informáticas, entre los sistemas informáticos. Además, el medio legible por ordenador puede comprender información legible por ordenador en un medio de estado transitorio tal como un enlace de red y/o una interfaz de red, incluyendo una red cableada o una red inalámbrica, que permiten a un equipo leer dicha información legible por ordenador.

10 Los programas informáticos (también llamados lógica de control informática) se almacenan en la memoria 1006 principal y/o en la memoria 1012 secundaria. Los programas informáticos también pueden recibirse a través de la interfaz 1024 de comunicaciones. Tales programas informáticos, cuando se ejecutan, permiten al sistema informático llevar a cabo las características de la presente invención como se discute en el presente documento. En particular, los programas informáticos, cuando se ejecutan, permiten al procesador 1004 realizar las características del sistema informático. En consecuencia, tales programas informáticos representan controladores del sistema informático.

V. Conclusión

20 Aunque se han descrito las realizaciones específicas de la invención, los expertos en la materia entenderán que pueden hacerse cambios a las realizaciones específicas sin apartarse del alcance de la invención de acuerdo con las reivindicaciones adjuntas. El alcance de la invención no se limita, por lo tanto, a las realizaciones específicas. Además, se pretende que las reivindicaciones adjuntas cubran cualquiera y todas las aplicaciones, modificaciones y realizaciones dentro del alcance de la presente invención.

REIVINDICACIONES

1. Un método para un sistema de procesamiento de información para cuantificar la información de la frecuencia fundamental de audio, que comprende:

5 capturar audio que representa una trama numerada de una pluralidad de tramas numeradas;
 calcular una clase de trama, en la que una clase es una cualquiera de entre una clase sonora y una clase sorda;
 si la trama es una clase sonora, calcular una frecuencia fundamental para la trama;
 10 si la trama es una trama numerada par y una clase sonora, calcular una palabra clave de una primera longitud cuantificando la frecuencia fundamental de la trama de manera absoluta;
 si la trama es una trama numerada par y una clase sorda, calcular una palabra clave de la primera longitud que indique una trama de clase sorda;
 15 si la trama es una trama numerada impar y una clase sonora, y al menos una de las tres tramas inmediatamente anteriores a la trama es fiable, calcular una palabra clave de una segunda longitud cuantificando la frecuencia fundamental de la trama diferencial que hace referencia a una frecuencia fundamental cuantificada de la trama fiable anterior más cercana, en la que la primera longitud es mayor que la segunda longitud;
 si la trama es una trama numerada impar y una clase sonora, y cada una de las tres tramas inmediatamente anteriores a la trama no es fiable, calcular una palabra clave de la segunda longitud cuantificando la frecuencia fundamental de la trama de manera absoluta; y
 20 si la trama es una trama numerada impar y una clase sorda, calcular una palabra clave de la segunda longitud que indique una trama de clase sorda;
 en el que una trama numerada par es fiable si es una clase sonora, y en el que una trama numerada impar es fiable si es una clase sonora y la frecuencia fundamental de la trama se cuantifica de manera absoluta o se cuantifica de manera diferencial en referencia a una frecuencia fundamental de la trama inmediatamente anterior.

2. El método de la reivindicación 1, en el que el sistema de procesamiento de información es uno cualquiera de entre un equipo cliente y un dispositivo inalámbrico, tal como un teléfono móvil.

3. El método de la reivindicación 2, en el que el audio se refiere a la voz.

4. El método de la reivindicación 3, en el que una trama incluye desde aproximadamente 20 a aproximadamente 30 milisegundos de audio, y las tramas sucesivas pueden solaparse entre si desde aproximadamente 10 a aproximadamente 15 milisegundos.

5. El método de la reivindicación 1, que comprende además:

40 transmitir la palabra clave que se calculó a un servidor.

6. Un sistema de procesamiento de información para cuantificar la información de la frecuencia fundamental de audio, que comprende:

45 un medio para capturar audio que representa una trama numerada de una pluralidad de tramas numeradas;
 un medio para calcular una clase de la trama, en la que una clase es una cualquiera de entre una clase sonora y una clase sorda;
 si la trama es una clase sonora, un medio para calcular una frecuencia fundamental para la trama;
 si la trama es una trama numerada par y una clase sonora, un medio para calcular una palabra clave de una primera longitud cuantificando la frecuencia fundamental de la trama de manera absoluta;
 50 si la trama es una trama numerada par y una clase sorda, un medio para calcular una palabra clave de la primera longitud que indique una trama de clase sorda;
 si la trama es una trama numerada impar y una clase sonora, y al menos una de las tres tramas inmediatamente anteriores a la trama es fiable, un medio para calcular una palabra clave de una segunda longitud cuantificando la frecuencia fundamental de la trama de manera diferencial que hace referencia a una frecuencia fundamental cuantificada de la trama fiable anterior más cercana, en la que la primera longitud es mayor que la segunda longitud;
 si la trama es una trama numerada impar y una clase sonora, y cada una de las tres tramas inmediatamente anteriores a la trama no es fiable, un medio para calcular una palabra clave de la segunda longitud cuantificando la frecuencia fundamental de la trama de manera absoluta; y
 60 si la trama es una trama numerada impar y una clase sorda, un medio para calcular una palabra clave de la segunda longitud que indique una trama de clase sorda;
 en el que una trama numerada par es fiable si es una clase sonora, y en el que una trama numerada impar es fiable si es una clase sonora y la frecuencia fundamental de la trama se cuantifica de manera absoluta o se cuantifica de manera diferencial en referencia a una frecuencia fundamental de la trama inmediatamente anterior.

7. El sistema de procesamiento de información de la reivindicación 6, en el que el sistema de procesamiento de información es uno cualquiera de entre un equipo cliente y un dispositivo inalámbrico, tal como un teléfono móvil.
8. El sistema de procesamiento de información de la reivindicación 7, en el que el audio se refiere a la voz.
- 5 9. El sistema de procesamiento de información de la reivindicación 6, que comprende además:
un transmisor para transmitir la palabra clave que se calculó a un servidor.
- 10 10. El sistema de procesamiento de información de la reivindicación 8, en el que una trama incluye desde aproximadamente 20 a aproximadamente 30 milisegundos de audio, y las tramas sucesivas pueden solaparse entre sí desde aproximadamente 10 a aproximadamente 15 milisegundos.
- 15 11. Un programa informático que comprende instrucciones de programa, adaptado para realizar las etapas del método de una cualquiera de las reivindicaciones 1-5, cuando dicho programa se ejecuta en un sistema de procesamiento de información.
12. Un medio legible por ordenador que lleva un programa informático de acuerdo con la reivindicación 11.

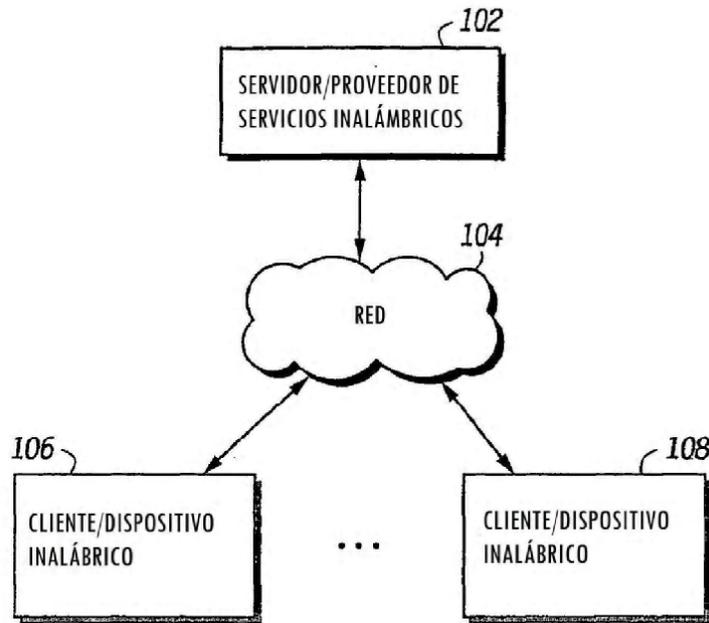


FIG. 1

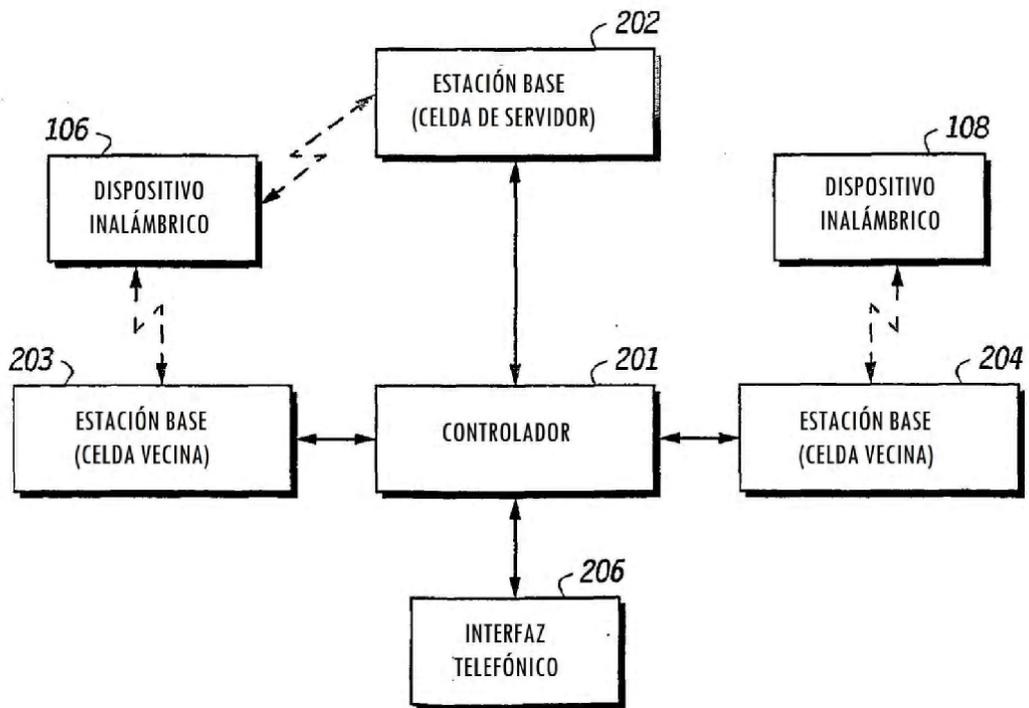


FIG. 2

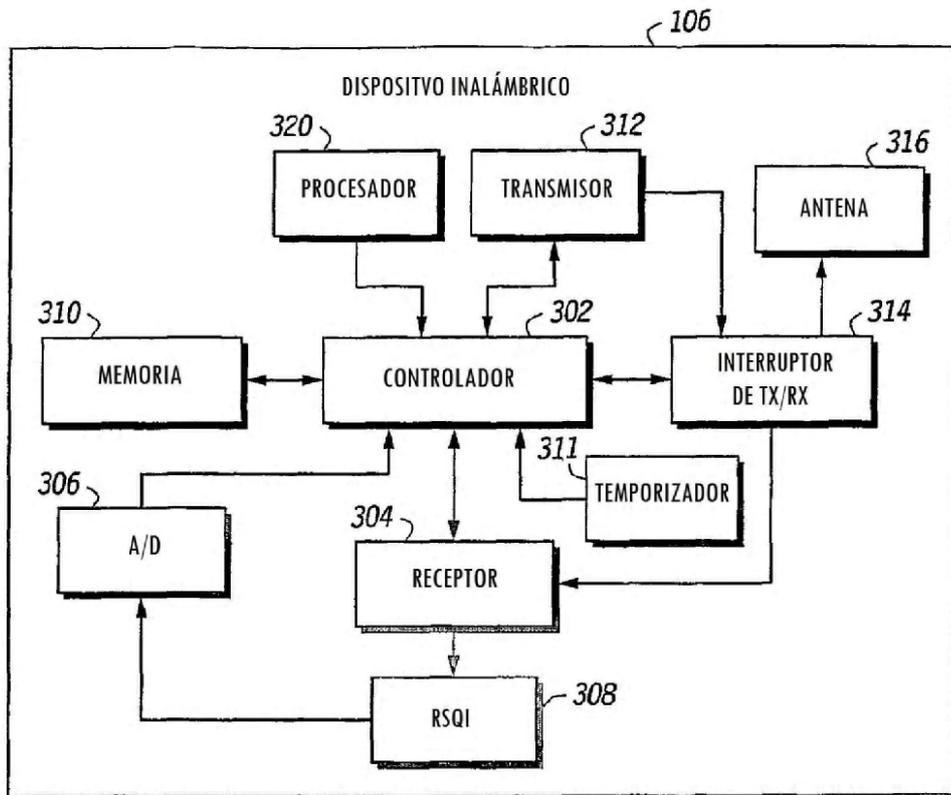


FIG. 3

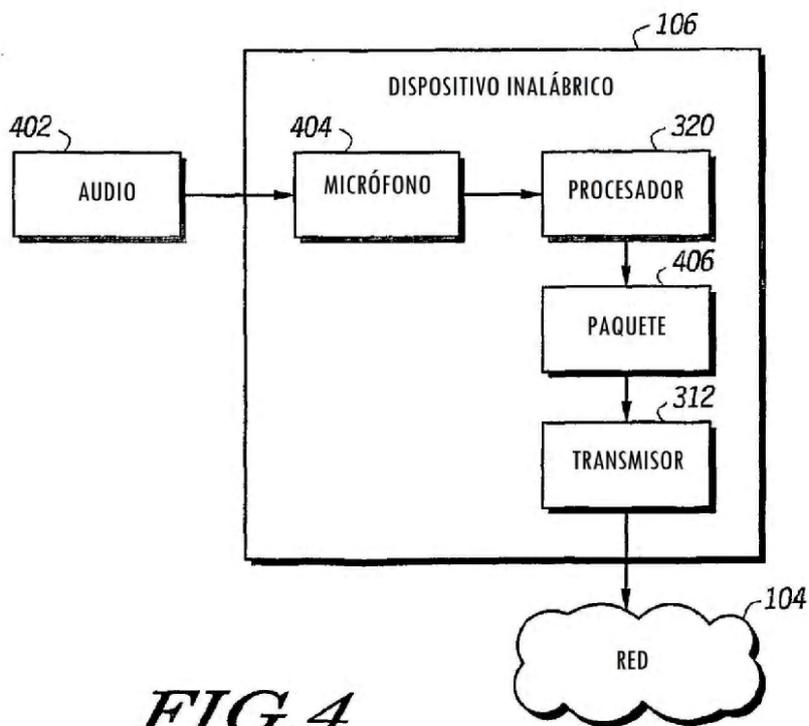


FIG. 4

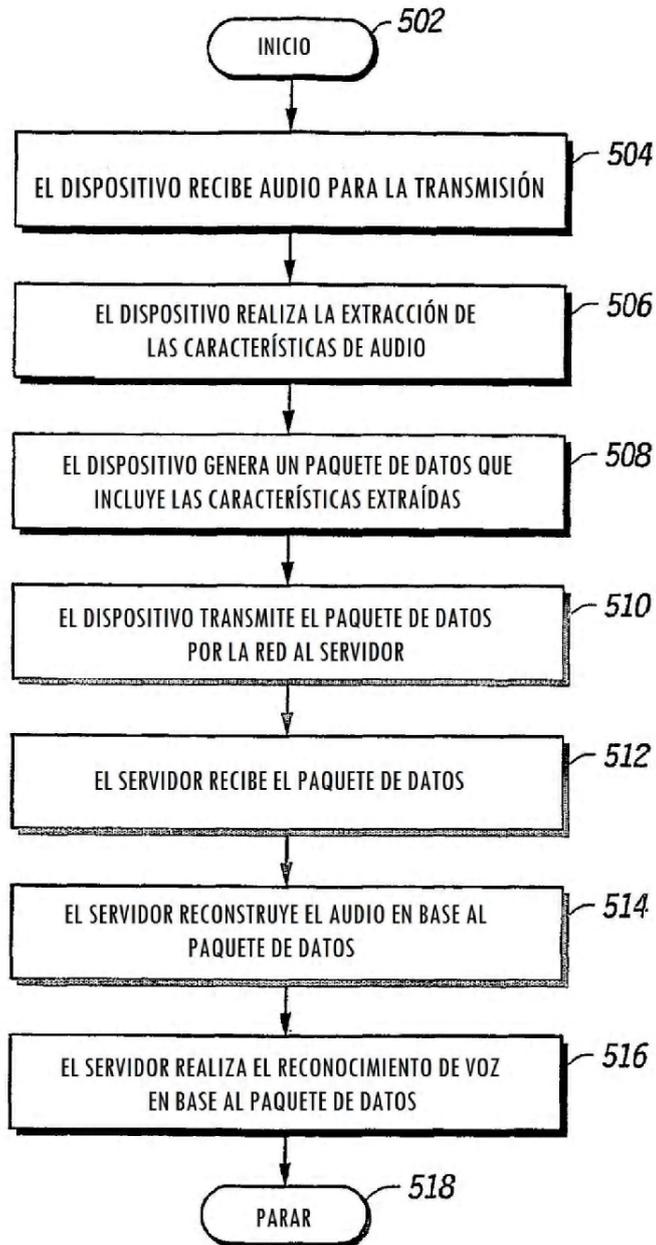


FIG.5

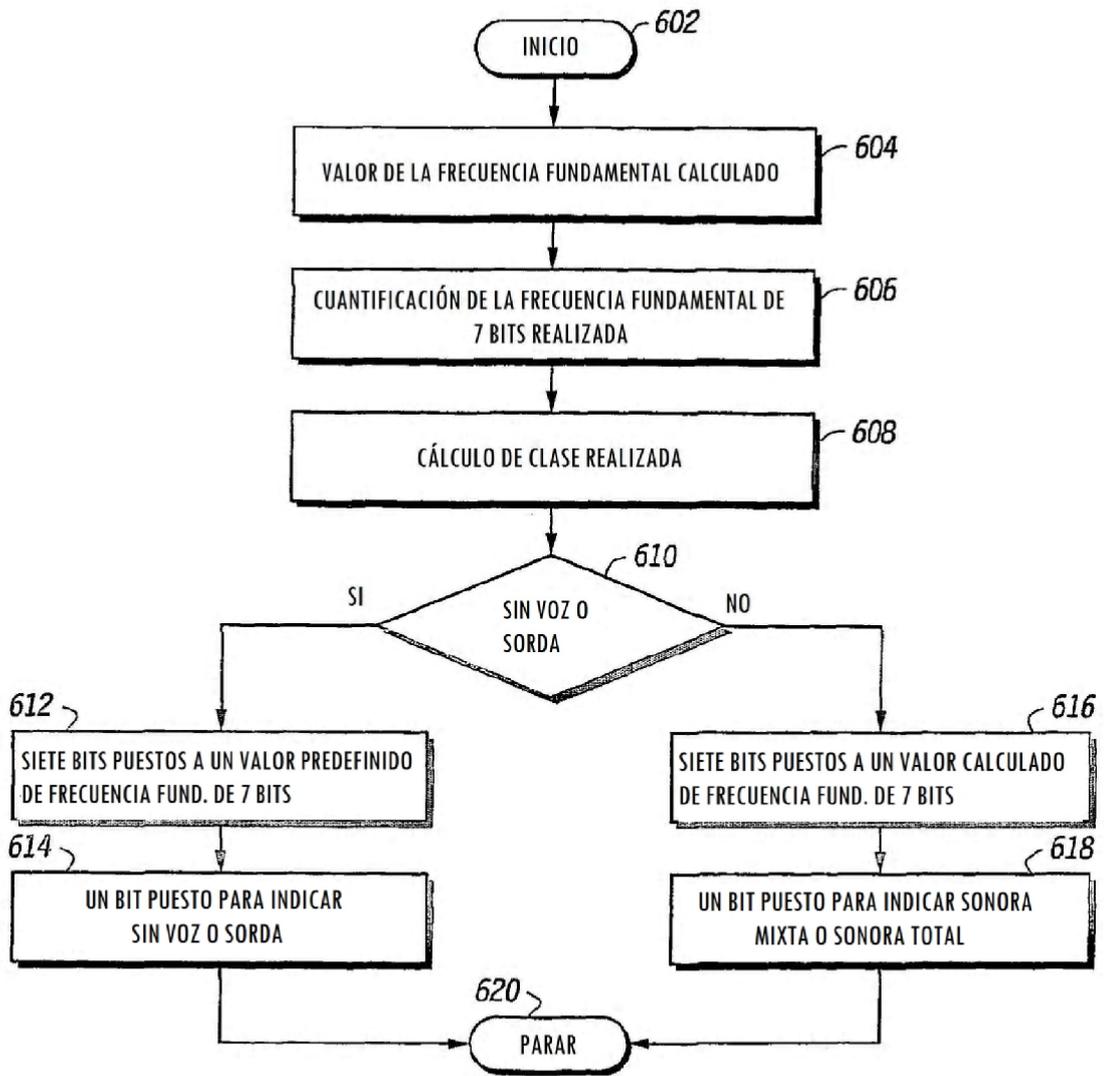
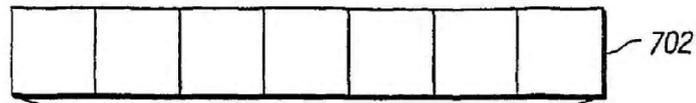
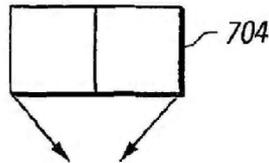


FIG.6



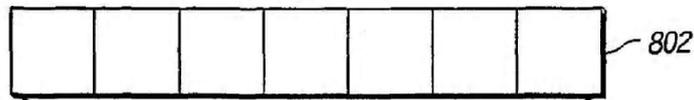
EN TOTAL 128 VALORES
LOS 128 VALORES: DATOS DE LA FRECUENCIA FUNDAMENTAL



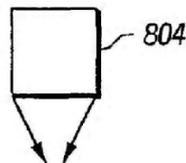
EN TOTAL 4 VALORES
LOS 4 VALORES: DATOS DE CLASE
- SIN VOZ
- SORDA
- SONORA MIXTA
- SONORA TOTAL

FIG. 7

- TÉCNICA ANTERIOR -



EN TOTAL 128 VALORES
127 VALORES: DATOS DE LA FRECUENCIA FUNDAMENTAL
1 VALOR: DATOS DE CLASE



EN TOTAL 2 VALORES
1 VALOR: SIN VOZ O SONORA MIXTA
1 VALOR: SORDA O SONORA TOTAL

FIG. 8

FIG. 9A

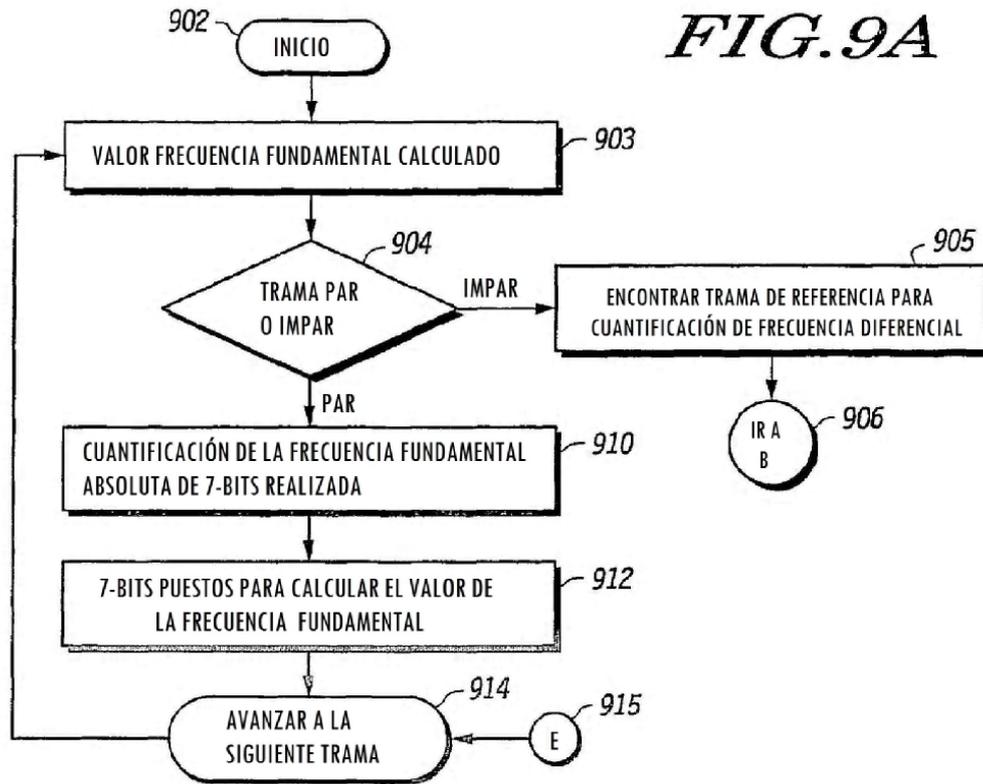
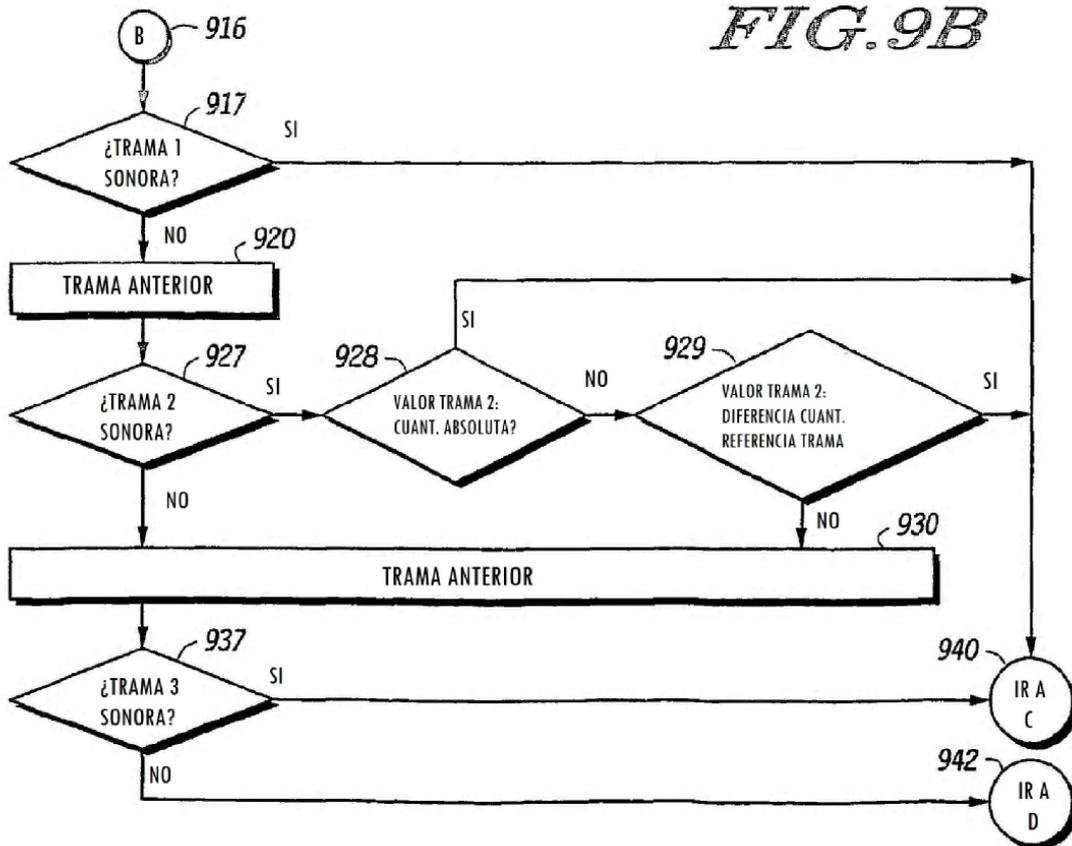


FIG. 9B



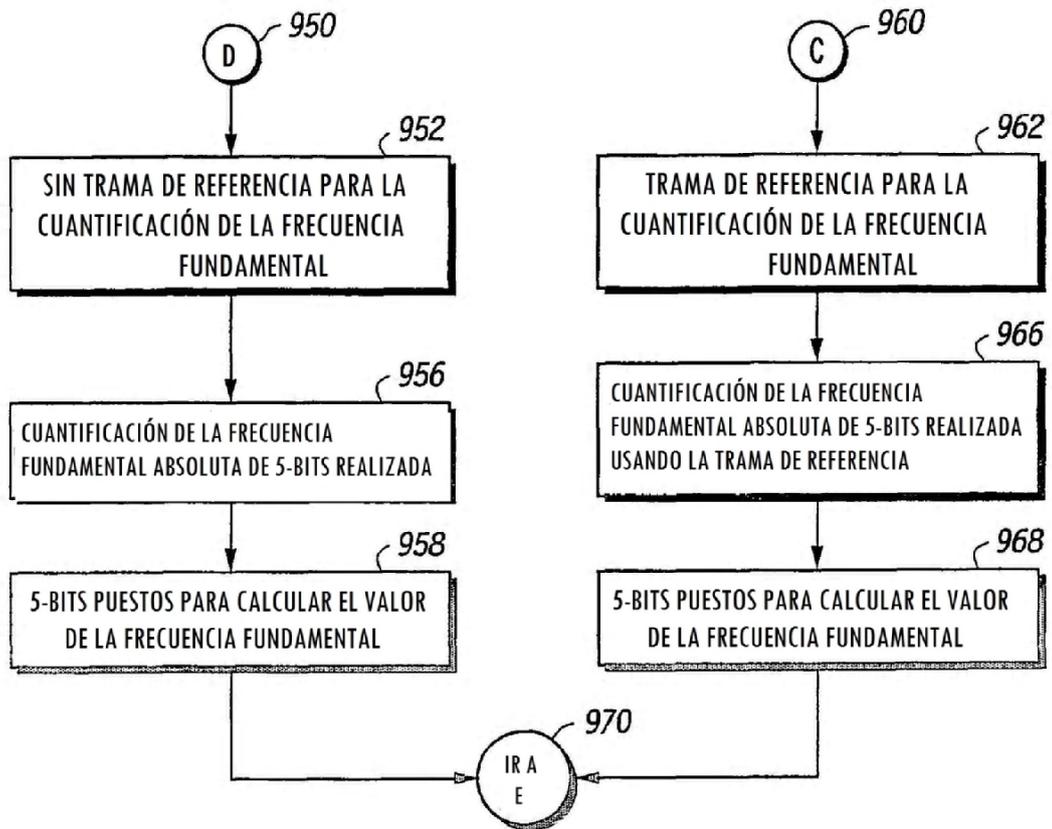


FIG. 9C

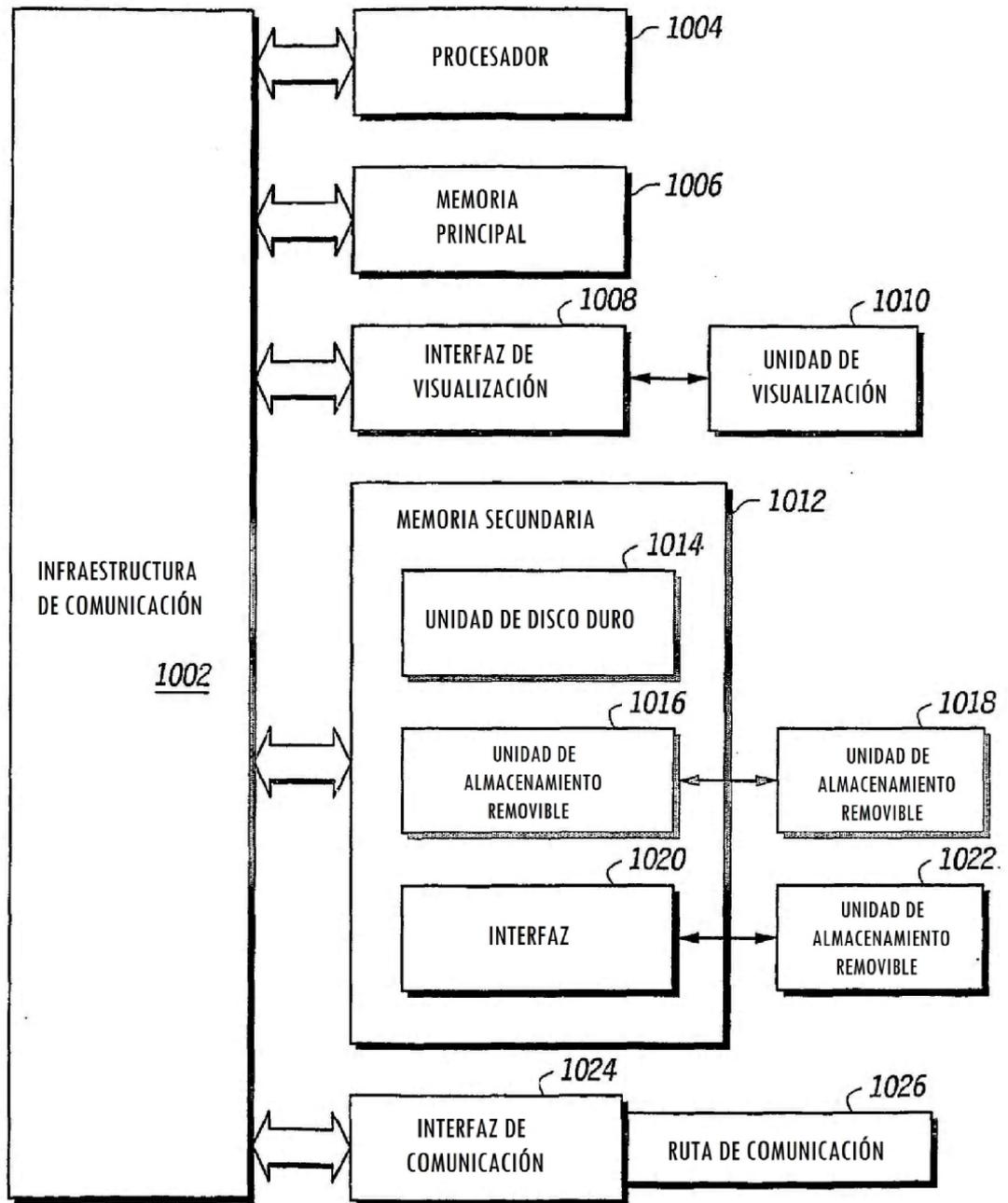


FIG.10