



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 298 841**

51 Int. Cl.:

H04Q 7/38 (2006.01)

H04M 3/523 (2006.01)

H04M 3/533 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **04798672 .4**

86 Fecha de presentación : **25.11.2004**

87 Número de publicación de la solicitud: **1695586**

87 Fecha de publicación de la solicitud: **30.08.2006**

54

Título: **Procedimiento y sistema para transmitir mensajes de voz.**

30

Prioridad: **03.12.2003 GB 0328035**

45

Fecha de publicación de la mención BOPI:
16.05.2008

45

Fecha de la publicación del folleto de la patente:
16.05.2008

73

Titular/es: **BRITISH TELECOMMUNICATIONS
public limited company
81 Newgate Street
London EC1A 7AJ, GB**

72

Inventor/es: **Scahill, Francis James y
Ringland, Simon Patrick Alexander**

74

Agente: **Curell Suñol, Marcelino**

ES 2 298 841 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

Procedimiento y sistema para transmitir mensajes de voz.

5 **Campo técnico**

La presente invención se refiere a un procedimiento y a un sistema de comunicaciones que utiliza tecnología de reconocimiento de voz para analizar un mensaje de voz y determinar de este modo su destino previsto.

10 **Antecedentes de la invención y técnica anterior**

Los sistemas de mensajería de voz semidúplex basados en paquetes móviles son conocidos en la técnica. Denominados de forma coloquial sistemas “pulsar para hablar” (PTT), en Estados Unidos se encuentran disponibles en el mercado desde hace algunos años, gracias a Nextel Communications, con la marca de servicio “Direct Connect”.

15 Dichos sistemas PTT también han sido desarrollados para funcionar dentro de un entorno de protocolo de internet (IP), con sistemas de voz sobre IP (VoIP). Específicamente, son conocidos en la técnica tanto los servicios generales de radio por paquetes (GPRS) como el acceso múltiple por división de código (CDMA) basados en sistemas VoIP PTT, así como los producidos por Motorola (ver http://www.motorola.com/mediacenter/news/detail/0,1958_3069_2512_23,00.html) y Qualcomm (ver http://www.qualcomm.com/press/releases/2002/020111_qchat_voip.html).

20 Cuando se utiliza un sistema PTT, normalmente el usuario selecciona el receptor previsto en una lista de una libreta de direcciones guardada en su propio microteléfono utilizando una interfaz gráfica y los controles de usuario del propio dispositivo, como es bien conocido en la técnica. También se conoce la marcación por voz de los servicios PTT, no obstante, y un ejemplo de dispositivo de la técnica anterior que proporciona dicha funcionalidad es el adaptador de bolsillo producido por Cellport Systems Inc. de Boulder, CO, para los teléfonos móviles Motorola iDEN i1000 e i1000 plus. La guía de usuario para el adaptador de bolsillo Cellport puede encontrarse en http://www.cellport.com/adapterquides/nextel_i1000_PAG.pdf. Como se expone en la misma, este marcador de voz comprende las palabras código predeterminadas que debe pronunciar el usuario, seguidas de la identificación (por ejemplo el número, pero alternativamente un código de marcado rápido) del receptor con el cual el usuario desea conectar, antes de decir el mensaje de voz que el usuario desea enviar. Por ejemplo, en el sistema Cellport, al utilizar la marcación por voz el usuario deberá pronunciar las palabras “Cellport, dial, pound, pound, 6284”. A continuación el adaptador repite las palabras reconocidas “pound, pound, 6284” y acto seguido se efectúa el proceso de conexión. El usuario puede entonces decir su mensaje pulsando el botón PTT en la forma usual.

25 No obstante, incluso con esta funcionalidad de marcación por voz, sigue existiendo una “fase de marcación” separada, en la cual el usuario debe seleccionar el receptor previsto, ya sea utilizando una interfaz gráfica normal, ya sea mediante la interfaz de marcación por voz, y hasta que esta fase de marcación se ha completado y se ha establecido la conexión el usuario no puede empezar a decir su primer mensaje. Por lo tanto, esta fase separada de marcación introduce una demora hasta que el usuario puede empezar a hablar y además requiere una interacción adicional del usuario con el dispositivo consistente en navegar por la pantalla gráfica o hablar según los protocolos de marcación por voz.

30 La patente US nº 6.075.844 da a conocer un sistema de mensajería en el cual se graba un mensaje en un dispositivo de grabación y posteriormente se envía al receptor. La persona que graba el mensaje dice el nombre del receptor al principio del mismo y, basándose en este enunciado el sistema decide donde debe enviarlo después de cargarlo en una red de comunicaciones.

35 El documento EP 1 041 779 A2 da a conocer una red de telefonía basada en el Protocolo de Internet (Voz sobre IP). Se conecta un primer teléfono a un servidor de directorios a través del canal de voz. El teléfono presenta un único botón multifunción. El llamante puede pulsar este botón para proporcionar el nombre de la persona a la cual desea llamar. A continuación, el servidor de directorios, basándose en la entrada de voz, determina el nombre del receptor a partir de una entrada de voz enviada por el canal de voz y devuelve al primer teléfono la dirección IP correspondiente al nombre reconocido. Seguidamente, el primer teléfono, basándose en dicha dirección IP, intenta establecer una llamada con un segundo teléfono correspondiente a la parte llamada. Si dicha parte responde a la llamada, puede efectuarse una comunicación entre los dos teléfonos. Alternativamente, el servidor de encaminamiento puede conectar el primer teléfono con el segundo teléfono. En este documento, la etapa de marcación por voz tiene lugar antes que la etapa de llamada real.

40 Otro documento WO01/78245 da a conocer un procedimiento de mensajería móvil caracterizado porque un usuario pulsa una tecla del teléfono móvil y pronuncia un enunciado que comprende una instrucción, por ejemplo “hablar”, un nombre de un receptor y una palabra clave de mensaje. Basándose en la palabra clave, se selecciona un mensaje de entre una lista de mensajes predeterminados y se envía al receptor identificado.

65

Sumario de la invención

La invención se propone mejorar el funcionamiento descrito anteriormente eliminando la fase de marcación separada de la interfaz de usuario. Más particularmente, la invención utiliza el reconocimiento de voz y la tecnología asociada para analizar un mensaje de voz para identificar el receptor previsto del mensaje y para transmitir el mensaje, o por lo menos una variante del mismo (por ejemplo un texto de un mensaje tal como lo obtiene el reconocedor de voz), al receptor previsto a través de la red. Esto permite al usuario simplemente pulsar el botón PTT de su microteléfono y decir inmediatamente su mensaje (preferentemente incluyendo dentro del mismo alguna indicación del receptor previsto, por ejemplo el nombre o similar), sin tener que esperar una fase previa de marcación separada.

Considerando lo que se ha dicho anteriormente, en un primer aspecto se dispone un procedimiento de comunicaciones que comprende las etapas siguientes:

recibir en un servidor encaminador de una red de comunicaciones un flujo de audio que contiene un enunciado, cuando un usuario pulsa un botón de un microteléfono y empieza a hablar;

almacenar en una memoria intermedia el flujo de audio recibido;

realizar un proceso de reconocimiento de voz en el flujo de audio recibido para reconocer el enunciado contenido en el mismo;

determinar, si es posible, un receptor previsto del flujo de audio en función del enunciado reconocido; y

si se ha determinado un receptor previsto, transmitir el flujo de audio al receptor previsto determinado utilizando un servicio de comunicaciones en semidúplex provisto por una red de conmutación e paquetes.

Otros aspectos también disponen un programa informático según la reivindicación 9, un medio de grabación legible por ordenador según la reivindicación 10 y un sistema de comunicaciones según la reivindicación 11.

Como se ha establecido anteriormente, la invención ofrece la ventaja de no ser necesario esperar a la realización de una fase de marcación separada antes de que el usuario pueda decir el mensaje. Esto mejora la experiencia de comunicación del usuario y hace que el servicio de comunicaciones en semidúplex le resulte más agradable y fácil de utilizar.

En una forma de realización de la invención, cuando en la etapa de determinación, a partir del enunciado reconocido, se determinan uno o más receptores posibles, el procedimiento comprende además las etapas siguientes:

indicar uno o más receptores previstos posibles a un usuario; y

recibir una señal de selección procedente del usuario indicando el, o los, receptor(es) previsto(s) posible(s) determinado(s) a los cuáles debe transmitirse el mensaje.

De este modo, cuando la etapa de determinación ha determinado uno o más receptores potenciales previstos para un mensaje, puede solicitarse al usuario la aclaración de cuáles de los receptores previstos posible identificados deben recibir el mensaje. Preferentemente, para mayor facilidad de interconexión, la etapa de indicación comprende además la generación de un aviso audio vocal correspondiente a la pluralidad de receptores previstos posibles; y una transmisión del aviso audio vocal generado al usuario. Esta característica adicional tiene en cuenta un aviso de salida de audio del dispositivo del usuario.

Además, en formas de realización de la invención, el proceso de reconocimiento de voz se realiza preferentemente solamente en una parte del mensaje de voz recibido. Esta característica reconoce que es probable que el receptor previsto de un mensaje se indicará probablemente al inicio de un mensaje (por ejemplo, si consideramos el mensaje "Hola Roger, ¿vendrás al bar esta noche?", el receptor previsto ("Roger") se identifica en la primera frase), y por lo tanto pueden conservarse los recursos de reconocimiento de voz efectuando sólo la cantidad de reconocimiento necesaria.

Por otra parte, en formas de realización de la invención, se incluyen las etapas adicionales de recepción e indicación de la identidad del usuario que ha generado el mensaje y selección de una gramática vocal dependiente del usuario que debe utilizarse con el proceso de reconocimiento de voz, pudiendo dicha gramática comprender de forma codificada los datos de la libreta de direcciones y similares.

Adicionalmente, las formas de realización de la invención pueden comprender, además, las etapas de recepción de una señal del usuario para la activación del reconocimiento de voz, dependiendo de la recepción de dicha señal la realización de las etapas de reconocimiento de voz y de determinación. Esta funcionalidad permite al usuario indicar de forma explícita cuando un mensaje es un mensaje para un nuevo receptor, y por lo tanto que deben realizarse las etapas de reconocimiento de voz y de determinación del receptor. De este modo, también se mejora la eficacia de la utilización de los recursos de reconocimiento de voz y además mejora el funcionamiento global de la invención, ya que las etapas de reconocimiento de voz y determinación del receptor sólo se realizarán en mensajes (normalmente

primeros mensajes de una serie) para los cuales existe una elevada probabilidad de que se mencione en ellos alguna clase de identificación del receptor, por ejemplo el nombre o similar, y por lo tanto una elevada probabilidad de que pueda determinarse el receptor previsto.

5

Breve descripción de los dibujos

Otras características y ventajas de la presente invención se pondrán de manifiesto a partir de la siguiente descripción de una forma de realización de la misma, proporcionada únicamente a título de ejemplo, y considerada conjuntamente con los dibujos adjuntos, en los cuales:

10

la figura 1 es un diagrama de bloques de un sistema arquitectónico de una forma de realización de la presente invención;

15

la figura 2(a) es un diagrama de flujo que ilustra las etapas del procedimiento implicadas en la forma de realización de la invención; y

la figura 2(b) es un diagrama de flujo continuación del diagrama de flujo de la figura 2(a).

20

Descripción de la forma de realización

A continuación, se describe una forma de realización de la presente invención haciendo referencia a las figuras 1, 2(a) y 2(b).

25

La forma de realización de la invención dispone un servicio pulsar para hablar (PTT) controlado por la voz. Más particularmente, la forma de realización está prevista para disponer un servicio de comunicaciones de pulsar para hablar que puede utilizar cualquiera de las tecnologías de comunicaciones PTT ya conocidas en la técnica y citadas en la parte introductoria de esta memoria, y añadir al mismo la funcionalidad que permite que las comunicaciones pulsar para hablar sean dirigidas a un receptor o destino previsto sin esperar a que se realice una fase de marcación explícita. En lugar de ello, en las formas de realización según la invención, se realiza el reconocimiento de voz en los mensajes pronunciados y se aplica una gramática vocal para determinar los receptores o el destino previstos para el mensaje, enviándose a continuación dicho mensaje al receptor o destino previsto determinado de este modo.

35

Considerando lo que se ha mencionado anteriormente, la figura 1 representa un diagrama de bloques arquitectónico global del sistema que ilustra los elementos principales del sistema de una forma de realización según la presente invención. Por lo tanto, con referencia a la figura 1, dentro de la forma de realización según la invención se dispone un servidor de encaminamiento de audio 14 para recibir flujos de señales de audio digitales transportadas por un servicio de comunicaciones PTT de una red (no mostrada) desde microteléfonos con función PTT, así como información del identificador de línea llamante (CLI). El servidor encaminador de audio 14 se dispone con una memoria intermedia de audio 142, que consiste en un medio de almacenamiento de datos como por ejemplo una RAM, un disco duro, un medio de almacenamiento óptico o similar, la cual está prevista para guardar temporalmente en ella cualquier mensaje de audio recibido. Adicionalmente, la forma de realización presenta un servidor de reconocimiento de voz 18 dispuesto para recibir audio digital desde la memoria intermedia de audio 142 del servidor encaminador de audio junto con la información del CLI y también para recibir gramática vocal y datos de léxico para su utilización en un proceso de reconocimiento de voz, desde una libreta de direcciones y una base de datos gramaticales 20. Durante la utilización, el servidor de reconocimiento de voz 18 ejecuta una aplicación de reconocimiento de voz para aplicar una gramática específica del usuario al audio digital recibido de la memoria intermedia de audio digital 142, para reconocer cualquier enunciado pronunciado y determinar el receptor previsto. Debe tenerse en cuenta que la aplicación de reconocimiento de voz ejecutada por el servidor de reconocimiento de voz puede ser cualquier aplicación de reconocimiento de voz actualmente conocida en la técnica, pero preferentemente una aplicación de reconocimiento de voz independiente del hablante. Un software adecuado de reconocimiento de voz que se encontraba disponible antes de la fecha de prioridad y que puede ser utilizado por el servidor de reconocimiento de voz 18 en la presente forma de realización es el Nuance 7, de Nuance Communications Inc., de 1005 Hamilton Court, Menlo Park, CA 94025.

55

El servidor de reconocimiento de voz 18 está dispuesto, además, para pasar el resultado del reconocimiento de voz, que preferentemente consiste en un conjunto de pares de valores clave que representan los valores de segmentos gramaticales en la voz reconocida junto con sus valores de seguridad de reconocimiento asociados, a un servidor de determinación del receptor 16. El servidor de determinación del receptor está dispuesto para recibir los pares de valores clave y efectuar la acción adecuada según los pares de valores clave devueltos por el reconocedor, como se describirá posteriormente. Una de las posibles acciones que el servidor de determinación del receptor puede llevar a cabo consiste en pasar una dirección en forma de un identificador del Número Marcado (DNI) al servidor encaminador de audio 14. Adicionalmente, el servidor de determinación del receptor 16 está dispuesto, además, para recibir datos del identificador de línea llamante (CLI) de los microteléfonos de usuario (descritos posteriormente) y también para enviar información de precandidatos a los microteléfonos de usuario, como se describirá más adelante.

65

Considerando de nuevo la libreta de direcciones y la base de datos gramatical 20, dicha base de datos guarda, para cada usuario del sistema registrado, una gramática de reconocimiento de voz que codifica datos de la libreta

ES 2 298 841 T3

de direcciones correspondientes a los nombres de posibles receptores y a sus DNI respectivos. Para cada usuario registrado se almacena una gramática específica de usuario registrada. A continuación, se muestra un ejemplo de formato de una gramática:

```
5      Names
      (
      [
10         ( bob ) { return("Bob Smith +447711123456")}
          ( peter jones ) { return("Peter Jones +447722123456") } w
          ( pete ) { return("Pete Brown +447733123456") }
      ]
15     )

      Phonemes:filler
      [
20     ph1
      ph2
      ...
      ph41
25     ]

      Fillers:filler
      [
30         Phonemes
      @-@
      ]

35     EndCal1
      [
      end call
      over and out
40     ]

      Overall
      [
45         (?hi +Names:n ?(it's Bob) *Fillers) {<action "placecall">
          <recipient $n>}
          *Fillers
          EndCal1 {<action "endcall">}
50     ]
```

La creación de la gramática de marcación puede efectuarse mediante cualquiera de las técnicas bien conocidas que utilizan texto (por ejemplo obteniendo una copia de la libreta de direcciones del teléfono del usuario) o una combinación de la libreta de direcciones para la parte del número y una entrada de voz para definir la pronunciación del nombre.

La anterior descripción se refiere a los diversos servidores que preferentemente se encuentran comprendidos en una red que suministra el servicio de comunicaciones PTT semidúplex o forman parte de la misma. No obstante, para describir con mayor detalle el funcionamiento de la forma de realización, también es necesario describir las características especiales de los microteléfonos móviles de usuario adaptados para su utilización con la presente invención. La figura 1 también ilustra un microteléfono de esta clase en forma de estación móvil A (10). También muestra la estación móvil B (12), pero ésta puede consistir en un microteléfono adaptado para PTT como los que ya son conocidos en la técnica.

Por lo tanto, haciendo referencia a la estación móvil A (10), la estación móvil A (10) está provista de un cliente de audio PTT A (102), que es el software adecuado para controlar la estación móvil 10 para proporcionar una funcionalidad PTT convencional. Es decir, el cliente de audio PTT A (102) permite que la estación móvil 10 utilice un servicio PTT de forma convencional. No obstante, la diferencia principal con la técnica anterior es que el cliente de audio

ES 2 298 841 T3

PTT A (102) está dispuesto para enviar datos de audio digitalizados producidos por el microteléfono A al servidor encaminador de audio 14, junto con el identificador de línea llamante de la estación móvil 10.

Adicionalmente, con la estación móvil 10 se suministra un cliente selector visual A (104). Se trata de un programa de software adicional que está dispuesto para interactuar con el servidor de determinación del receptor 16 dentro de la red, con el fin de enviar al mismo el identificador de línea llamante (CLI) de la estación móvil 10 cuando se inicia primero una llamada PTT, y también para recibir una lista de identidades de receptor candidatas del servidor de determinación del receptor 16, en el caso de que se determine más de un receptor previsto. Además, el cliente selector visual A (104) está dispuesto para visualizar una lista de precandidatos para el usuario en una pantalla de la estación móvil A y permitir al usuario seleccionar el receptor previsto. A continuación la información de la selección se retransmite de vuelta al servidor de determinación del receptor 16.

Dentro de la estación móvil B (12) convencional se dispone un cliente de audio B (122), que es esencialmente el mismo que el cliente de audio PTT A (102) dispuesto en la estación móvil 10. El cliente de audio PTT B (122) se encuentra dispuesto para dotar a la estación móvil B (12) de la funcionalidad PTT convencional y la única diferencia entre el cliente de audio PTT B (122) y la técnica anterior es que el cliente de audio PTT B (122) está dispuesto para recibir un mensaje PTT de la memoria intermedia de audio 142 que forma parte del servidor encaminador de audio 14. En todos los demás aspectos, la estación móvil B 12 puede ser convencional.

A continuación, se describe el funcionamiento de la forma de realización de la figura 1 respecto a las figuras 2(a) y (b).

En esta descripción de ejemplo del funcionamiento de la forma de realización suponemos que la estación móvil A no presenta ninguna llamada PTT en curso y desea enviar un mensaje PTT a la estación móvil B. Es decir, el estado actual de la estación A es que no ha enviado ni recibido ninguna llamada PTT a ninguna otra estación o de ninguna otra estación durante por lo menos un período de espera del PTT (normalmente 20 segundos). De este modo, dentro de la forma de realización de la invención, para iniciar una llamada desde la estación móvil A a la estación móvil B en la etapa 2.2 el usuario A pulsa el botón PTT. La pulsación del botón PTT de la estación móvil A hace que el cliente de audio PTT 102 empiece a ejecutarse, así como el cliente selector visual 104. En la etapa 2.4, el cliente selector visual 104 se conecta con el servidor de determinación del receptor 16 y envía el identificador de línea llamante (CLI) de la estación móvil A al servidor de determinación del receptor 16. A continuación (o por lo menos simultáneamente o antes, el orden de las etapas 2.4 y 2.6 no es importante) el cliente de audio PTT 102 se conecta con el servidor encaminador de audio 14 y empieza a fluir audio digitalizado al servidor encaminador de audio 14. En este punto, en la etapa 2.6, el usuario A de la estación móvil A pronuncia el mensaje de audio que desea transmitir mediante el servicio PTT y la estación móvil A digitaliza y empaqueta el mensaje para transferirlo por el flujo de audio al servidor encaminador de audio 14. Esta digitalización y empaquetado es bien conocida en la técnica. Además de transferir el audio al servidor de audio 14, el cliente de audio PTT 102 también envía el identificador de línea llamante (CLI) de la estación móvil A al servidor encaminador de audio 14.

En la etapa 2.8, el servidor encaminador de audio almacena el flujo de audio recibido en la memoria intermedia de audio 142 y también envía una copia del flujo de audio al servidor de reconocimiento de voz 18, en la etapa 2.10. Simultáneamente, el servidor encaminador de audio 14 también envía el identificador de línea llamante de la estación móvil A al servidor de reconocimiento de voz 18.

A continuación, en la etapa 2.12, el servidor de reconocimiento de voz 18 utiliza el identificador de línea llamante recibido de la estación móvil A para acceder a la libreta de direcciones y la base de datos gramaticales 20, para recuperar de la misma la gramática específica de usuario allí almacenada para la estación móvil. Se apreciará que la gramática y el léxico para el reconocimiento de voz se guardan en la libreta de direcciones y en la base de datos de gramática 20 indexada por CLI, para tener en cuenta la gramática y el léxico específicos para la estación móvil A que deben recuperarse.

A continuación, en la etapa 2.16, el servidor de reconocimiento de voz 18 efectúa el proceso de reconocimiento de voz en el flujo de audio recibido del servidor encaminador de audio 14. Debe tenerse en cuenta que el servidor de reconocimiento de voz 18 puede efectuar el reconocimiento de voz en el flujo de audio recibido cuando se recibe el flujo o, alternativamente, puede esperar hasta detectar silencio en el flujo antes de empezar el reconocimiento [o el final del flujo cuando se ha liberado el botón PTT]. Esta elección dependerá del software exacto de reconocimiento de voz seleccionado para utilizarlo en el servidor de reconocimiento de voz 18. El proceso de reconocimiento de voz realizado por el servidor de reconocimiento de voz 18 actúa para reconocer el enunciado del usuario contenido en el flujo de audio recibido desde el servidor encaminador de audio 14, utilizando la gramática de reconocimiento para el usuario para guiar el proceso de reconocimiento. Dentro de la forma de realización, a continuación, el servidor de reconocimiento de voz devuelve información clave al servidor de determinación del receptor a través del los segmentos Nuance NL (cuando se utiliza el reconocedor Nuance 7 mencionado anteriormente). Así, por ejemplo, para el enunciado "corto y fuera" el reconocedor devolvería un par de valores clave de *acción=final de llamada*, mientras que para "Hola Bob, soy Bob" el reconocedor devolvería dos pares de valores clave *acción=establecimiento de llamada* y *receptor="Bob Smith +447711 i23456"*, como se ha determinado mediante la gramática de usuario. Con los pares de valores clave, el reconocedor también dispone un valor de seguridad de reconocimiento indicativo de hasta qué punto es aceptado el reconocimiento de cualquier individuo que conduce a un par de valores clave particular que son correctos. Estos valores de seguridad son bien conocidos en la técnica.

ES 2 298 841 T3

Una vez efectuado el reconocimiento, y enviados los pares de valores clave y los valores de seguridad al servidor de determinación del receptor, en la etapa 2.20 el servidor de determinación del receptor efectúa una evaluación de los valores de seguridad de los pares de valores clave devueltos. En este caso los valores de seguridad pueden compararse con un valor umbral o similar y si la evaluación indica que el reconocedor está seguro de los resultados el proceso puede pasar a efectuar la etapa 2.22. En la etapa 2.22, el servidor de determinación del receptor 16 envía los DNI del (de los) receptor(es) previsto(s) determinado(s) y obtenido(s) de los pares de valores clave recibidos al servidor encaminador de audio 14 y, además, como confirmación, al cliente selector visual 104 en la estación móvil A. A continuación, en la etapa 2.24, el servidor encaminador de audio transmite el mensaje de audio almacenado en la memoria intermedia desde la memoria intermedia de audio 142 al (a los) receptor(es) identificado(s) por el (los) DNI recibido(s) del servidor de determinación del receptor, utilizando el servicio de comunicaciones PTT. En este caso, suponemos que los DNI recibidos del servidor de determinación del receptor identifican la estación móvil B, en cuyo caso el servidor encaminador de audio transmite el mensaje de audio desde la memoria intermedia de audio 142 al cliente de audio PTT 122 en la estación móvil B, a través de la red normal adaptada para PTT. Simultáneamente, en la etapa 2.26, el cliente selector visual 104 en la estación móvil A visualiza el (los) DNI determinado(s) para el usuario A en la pantalla de la estación móvil A, como confirmación de que el mensaje ha sido enviado adecuadamente. En este punto, por lo tanto, el servidor encaminador de audio ha establecido una llamada PTT entre la estación móvil A y la estación móvil B y, a continuación, las comunicaciones PTT pueden continuar de la forma convencional.

Es importante tener en cuenta que es posible para el servidor de reconocimiento de voz reconocer de forma segura dos o más receptores previstos, cuando se habla con dos o más identificadores de receptor (por ejemplo el mensaje "Pete, Bob, soy David), en cuyo caso tanto Pete como Bob son receptores previstos). Gracias a la gramática de reconocimiento, puede enviarse de vuelta el DNI de ambos o de todos los receptores previstos y, gracias al reconocimiento seguro de ambos o todos los receptores queda claro que el mensaje estaba previsto para ambos o todos los receptores. En tal caso el servidor de determinación del receptor controla el servidor encaminador de audio para establecer una llamada PTT de grupo con cada uno de los receptores previstos determinados (Pete y Bob en el ejemplo). Por lo tanto, esta característica permite grupos de llamada para llamadas de grupo que deben definirse de forma dinámica, sencillamente indicando en el mensaje los nombres de cada uno de los receptores previstos que deben ser partes de la llamada de grupo.

Volviendo a la etapa 2.20, si la evaluación realizada en la misma no indica la existencia de un resultado seguro, en la etapa 2.28 el servidor de determinación del receptor realiza una evaluación adicional para determinar si se han obtenido o no uno o más resultados no seguros devueltos por el reconocedor de voz. Si lo que ocurre es que no se han devuelto pares de valores clave, el proceso de reconocimiento de voz ha fallado. En este caso, en la etapa 2.36 el servidor de determinación del receptor envía un mensaje al cliente selector visual 104 en la estación móvil A indicando que el reconocimiento ha fallado y a continuación se visualiza un mensaje de fallo de reconocimiento para el usuario en la estación móvil A. En tal caso, a continuación el usuario A debe seleccionar el receptor previsto para este mensaje utilizando la interfaz de usuario gráfica convencional.

No obstante, en caso contrario, si la evaluación de la etapa 2.28 indica que existen uno o más resultados no seguros se invita al usuario a confirmar el o los resultado(s) no seguros. Por lo tanto, en la etapa 2.30 el servidor de determinación del receptor 16 envía una lista de uno o más resultados no seguros al cliente selector visual 104 en la estación móvil A. A continuación, el cliente selector visual 104 visualiza la lista recibida para el usuario en la pantalla de usuario de la estación móvil A y, en la etapa 2.32, el usuario selecciona el receptor previsto de la lista. Seguidamente, el cliente selector visual 104 envía de nuevo la información de la selección al servidor de determinación del receptor 16 en la etapa 2.34. El servidor de determinación del receptor recibe la información de la selección y a continuación vuelve a la etapa 2.22 en la cual se envía el DNI seleccionado (o los DNI cuando se seleccionan múltiples receptores previstos) al servidor encaminador de audio. A continuación el proceso pasa a la etapa 2.24 y avanza tal como se ha descrito anteriormente.

Por lo tanto, a la vista de la descripción anterior, se apreciará que la forma de realización de la invención ofrece comunicaciones PTT adecuadas sin que el usuario deba someterse a la fase de marcación y esperar la conexión posterior. A este respecto, la invención utiliza el tiempo de espera inherente en las comunicaciones PTT y en particular en las implantaciones VoIP de las mismas, y explota este tiempo de espera para realizar el reconocimiento de voz y la posterior determinación del receptor previsto en el mensaje, para permitir la selección automática del receptor o del destino. Por lo tanto, la invención proporciona una experiencia de usuario mejorada superior y más extensa a la proporcionada por los servicios de comunicaciones PTT convencionales conocidos en la técnica.

Se apreciará que en la forma de realización descrita pueden efectuarse diversas modificaciones para generar otras formas de realización. Por ejemplo, en otra forma de realización, para iniciar el proceso de reconocimiento y determinación del receptor, el usuario A de la estación móvil A debe enviar una señal de activación desde la estación móvil A, por ejemplo mediante una doble pulsación del botón PTT. Esta "doble pulsación" haría que el cliente selector visual A (104) enviará una señal de activación al servidor de determinación del receptor 16, que a su vez enviaría una señal de activación al servidor de reconocimiento de voz 18. Una operación de activación explícita de esta clase puede resultar conveniente para evitar que la invención funcione en circunstancias inesperadas.

Como variante adicional, en otras formas de realización, el usuario A puede efectuar una llamada de grupo utilizando la invención. Las llamadas de grupo utilizando PTT son conocidas por si mismas en la técnica, y se incluyen para su utilización en formas de realización de la invención incluyendo dentro de cada dirección de usuario un nombre

ES 2 298 841 T3

de grupo, junto con los números de teléfono asociados, que forman parte del grupo. En funcionamiento, el usuario A pronuncia el nombre del grupo, que es reconocido por el servidor de reconocimiento de voz, y el nombre de grupo almacenado aplicado a la gramática de usuario para determinar los DNI del grupo. Si se reconoce el nombre de grupo, el servidor de determinación del receptor envía cada uno de los DNI pertenecientes al grupo al servidor encaminador de audio 14, que a continuación conecta la llamada PTT de grupo de forma convencional.

En otra forma de realización, el servidor de reconocimiento de voz está dispuesto para reconocer solamente los primeros pocos segundos de un mensaje, para conservar los recursos de reconocimiento de voz. Esta característica se basa en la premisa de que en la mayoría de saludos se pronuncia el nombre del receptor dentro de un límite de tiempo de este tipo (por ejemplo los saludos: “Hola Pete”, “Qué tal, Bob”, “Buenos días, Pete”, etc.). Los límites de tiempo de reconocimiento de entre 3 a 5 segundos resultan suficientes para este fin.

Con respecto a la selección de receptores previstos en el caso de resultados no seguros, en la forma de realización anterior describimos el cliente selector visual que visualiza la lista de precandidatos para el usuario para selección visual. No obstante, en otras formas de realización, el servidor de determinación del receptor puede incluir un programa sintetizador de voz que se utiliza para generar avisos de audio referentes a las selecciones disponibles, que a continuación son encaminadas al cliente de audio PTT 102 en el microteléfono para ser tratadas por el usuario. Debe tenerse en cuenta que esto puede realizarse simultáneamente con la visualización de la lista de precandidatos por el cliente selector visual para presentar las selecciones tanto en interfaz de audio como visual, o alternativamente puede sustituir a la selección visual. Para obtener la selección hablada por el usuario de un receptor previsto (por ejemplo, el usuario dice “Sí” cuando se lee el nombre del receptor previsto y (opcionalmente) “No” cuando suena el nombre de un receptor distinto del previsto, o alternativamente el usuario dice “Bob Smith” para distinguir entre Bob Smith y Bob Jones, previamente designados sencillamente como “Bob”), el cliente de audio puede transmitir cualquier respuesta de usuario al servidor de reconocimiento de voz a través del servidor encaminador de audio para reconocimiento de las respuestas.

Finalmente, como característica opcional, el servidor encaminador de audio puede disponerse en otras formas de realización de la invención para cortar el audio que ha sido reconocido y utilizarlo para seleccionar el receptor previsto del mensaje y transmitir solamente la parte del mensaje no utilizada para la determinación del receptor previsto.

Aunque la descripción anterior se centra en el funcionamiento de la invención antes de establecer una llamada PTT, en otro modo, las formas de realización de la invención pueden funcionar durante la llamada PTT (es decir, dentro del período de tiempo de espera de PTT, cuando los mensajes están esperando para viajar en doble sentido entre dos o más partes en una llamada) para detectar una frase “fin de llamada” predeterminada, como por ejemplo “corto y fuera” o “llamada final”, y para operar terminando la llamada. En este modo de operación, el flujo de audio encaminado a través del servidor encaminador de audio se copia en el servidor de reconocimiento de voz, el cual realiza el reconocimiento de voz en cada mensaje enviado para detectar la frase de final de llamada predeterminada. Debe tenerse en cuenta que puede predefinirse más de una frase de final de llamada y el servidor de reconocimiento de voz puede detectar cualquiera de las frases predeterminadas. Si se detecta una de estas frases, el servidor de reconocimiento de voz lo indica al servidor encaminador de audio, el cual termina la llamada. Simultáneamente, el servidor de reconocimiento de voz puede indicar al servidor de determinación del receptor 16 que puede enviar la señal (“llamada finalizada” al cliente selector visual 104 en el terminal de usuario. En tal caso, el cliente selector visual 104 a continuación visualiza un mensaje de “llamada finalizada” para el usuario en la pantalla de la estación móvil 10. De forma similar, puede enviarse al terminal móvil una salida “llamada finalizada” utilizando un mensaje de audio sintetizado o almacenado.

En la forma de realización descrita anteriormente, los móviles A y B se comunican con los servidores utilizando una red celular inalámbrica. En formas de realización alternativas podría utilizarse, en su lugar, una red de acceso inalámbrico no celular, por ejemplo LAN, Wi-Fi y Bluetooth inalámbricas. En otras formas de realización alternativas, uno o ambos terminales pueden ser fijos (por ejemplo un ordenador personal).

Si el contexto no exige claramente lo contrario, en las reivindicaciones y la descripción los términos “comprender”, “comprendiendo” y similares deben considerarse en sentido inclusivo y no exclusivo o exhaustivo, es decir, en el sentido de “incluido pero no limitado a”.

ES 2 298 841 T3

REIVINDICACIONES

1. Procedimiento de comunicaciones que comprende las etapas siguientes:

5 recibir (etapa 2.6) en un servidor encaminador de una red de comunicaciones un flujo de audio que contiene un enunciado, cuando un usuario pulsa un botón de un microteléfono y empieza a hablar;

almacenar en una memoria intermedia (etapa 2.8) el flujo de audio recibido;

10 realizar un proceso de reconocimiento de voz (etapa 2.16) en el flujo de audio recibido para reconocer el enunciado contenido en el mismo;

determinar (etapa 2.16), si es posible, un receptor previsto del flujo de audio en función del enunciado reconocido;

15 y

si se ha determinado un receptor previsto, transmitir (etapa 2.24) el flujo de audio al receptor previsto determinado, utilizando un servicio de comunicaciones en semidúplex proporcionado por una red de conmutación de paquetes.

20 2. Procedimiento según la reivindicación 1, en el que cuando en la etapa de determinación se determinan uno o más receptores previstos posibles a partir de un enunciado reconocido, el procedimiento comprende asimismo las etapas siguientes:

25 indicar (etapa 2.30) uno o más receptores previstos posibles a un usuario; y

recibir (etapa 2.34) una señal de selección procedente del usuario indicando uno, o más, receptores previstos posibles determinados a los cuáles debe transmitirse el mensaje.

30 3. Procedimiento según la reivindicación 2, en el que la etapa de indicación comprende asimismo la generación de un aviso audio vocal correspondiente a uno o más receptores previstos posibles; y la transmisión al usuario del aviso audio vocal generado.

35 4. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que cuando la etapa de determinación determina una pluralidad de receptores previstos, el flujo de audio se transmite a cada uno de los receptores determinados utilizando una función de llamada de grupo del servicio de comunicaciones en semidúplex.

5. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que el proceso de reconocimiento de voz se realiza únicamente en una parte del flujo de audio recibido.

40 6. Procedimiento según cualquiera de las reivindicaciones anteriores, que comprende asimismo las etapas siguientes: recibir (etapa 2.10) una indicación de la identidad de un usuario que ha generado el flujo de audio; y seleccionar (etapa 2.12) una gramática vocal dependiente del usuario para su utilización en el proceso de reconocimiento de voz en función de la identidad del usuario.

45 7. Procedimiento según cualquiera de las reivindicaciones anteriores, que comprende asimismo la etapa de recepción de una señal del usuario para la activación del reconocimiento de voz, en el que la etapas de reconocimiento de voz y determinación se realizan en función de la recepción de dicha señal.

50 8. Procedimiento según cualquiera de las reivindicaciones anteriores, que comprende además las etapas siguientes:

controlar los flujos de audio transportados por el servicio de comunicaciones en semidúplex;

55 realizar un proceso de reconocimiento de voz (etapa 2.16) en los flujos de audio controlados para determinar los enunciados respectivos contenidos en los mismos; y,

si se determina que cualquiera de los flujos de audio contiene un enunciado predeterminado, indicar que el servicio de comunicaciones en semidúplex debería dejar de transportar el flujo de audio.

60 9. Programa informático o sucesión de programas informáticos dispuestos para que al ser ejecutados por un sistema informático induzcan al programa informático a realizar todas las etapas del procedimiento según cualquiera de las reivindicaciones anteriores.

65 10. Medio de almacenamiento legible por ordenador que almacena un programa informático o un programa cualquiera o diversos programas cualesquiera de una sucesión de programas informáticos según la reivindicación 9.

ES 2 298 841 T3

11. Sistema de comunicaciones que comprende:

5 un microteléfono dispuesto en funcionamiento para recibir un flujo de audio que contiene un enunciado y enviar dicho flujo de audio a un servidor encaminador de una red de comunicaciones cuando un usuario pulsa un botón de dicho microteléfono y empieza a hablar;

unos medios de almacenamiento (142) para el almacenamiento intermedio del flujo de audio recibido;

10 un reconocedor de voz (18) dispuesto en funcionamiento para reconocer el enunciado contenido dentro del flujo de audio recibido;

unos medios de determinación del receptor (16) dispuestos para determinar, si es posible, un receptor previsto del flujo de audio en función del enunciado reconocido; y

15 unos medios (14) para transmitir el flujo de audio a un receptor previsto determinado utilizando un servicio de comunicaciones en semidúplex proporcionado por una red de conmutación de paquetes, si se ha determinado el receptor previsto.

12. Sistema según la reivindicación 11, que comprende asimismo:

20 unos medios indicadores (16, 104) para indicar a un usuario uno o más receptores previstos posibles; y

unos medios (16) para recibir del usuario una señal de selección que indique uno o más de los posibles receptores previstos determinados a los cuales debería transmitirse el flujo de audio.

25 13. Sistema según la reivindicación 12, en el que los medios indicadores (16) comprenden asimismo:

unos medios de generación de avisos de audio para generar un aviso audio vocal correspondiente a uno o más receptores previstos posibles; y una salida (102) para transmitir al usuario el aviso audio vocal generado.

30 14. Sistema según cualquiera de las reivindicaciones 11 a 13, en el que cuando los medios de determinación del receptor (16) determinan una pluralidad de receptores previstos, los medios de transmisión (14) está dispuestos asimismo para transmitir el flujo de audio a cada uno de los receptores determinados utilizando una función de llamada de grupo del servicio de comunicaciones en semidúplex.

35 15. Sistema según cualquiera de las reivindicaciones 11 a 14, en el que el reconocedor de voz (18) funciona solamente en una parte del flujo de audio recibido.

16. Sistema según cualquiera de las reivindicaciones 11 a 15 que comprende asimismo:

40 unos medios (16) para recibir una indicación de la identidad de un usuario que ha generado el flujo de audio; y

unos medios de selección de gramática (18, 20) para seleccionar una gramática vocal que depende del usuario para utilizarla en el proceso de reconocimiento de voz en función de la identidad del usuario.

45 17. Sistema según cualquiera de las reivindicaciones 11 a 16, que comprende asimismo unos medios destinados a recibir del usuario una señal de activación del reconocimiento de voz, en el que los medios de reconocimiento de voz y de determinación del receptor pueden funcionar dependiendo de la recepción de dicha señal.

50 18. Sistema según cualquiera de las reivindicaciones 11 a 17, que comprende asimismo unos medios (14, 18) para controlar flujos de audio transportados por el servicio de comunicaciones semidúplex, estando dispuesto asimismo el reconocedor de voz (18) para efectuar un proceso de reconocimiento de voz en los flujos de audio controlados para determinar los respectivos enunciados contenidos en los mismos; comprendiendo asimismo el sistema unos medios de señalización (18) para indicar que el servicio de comunicaciones semidúplex debe dejar de transportar flujos de audio, si se determina que alguno de los flujos de audio contiene un enunciado predeterminado.

60

65

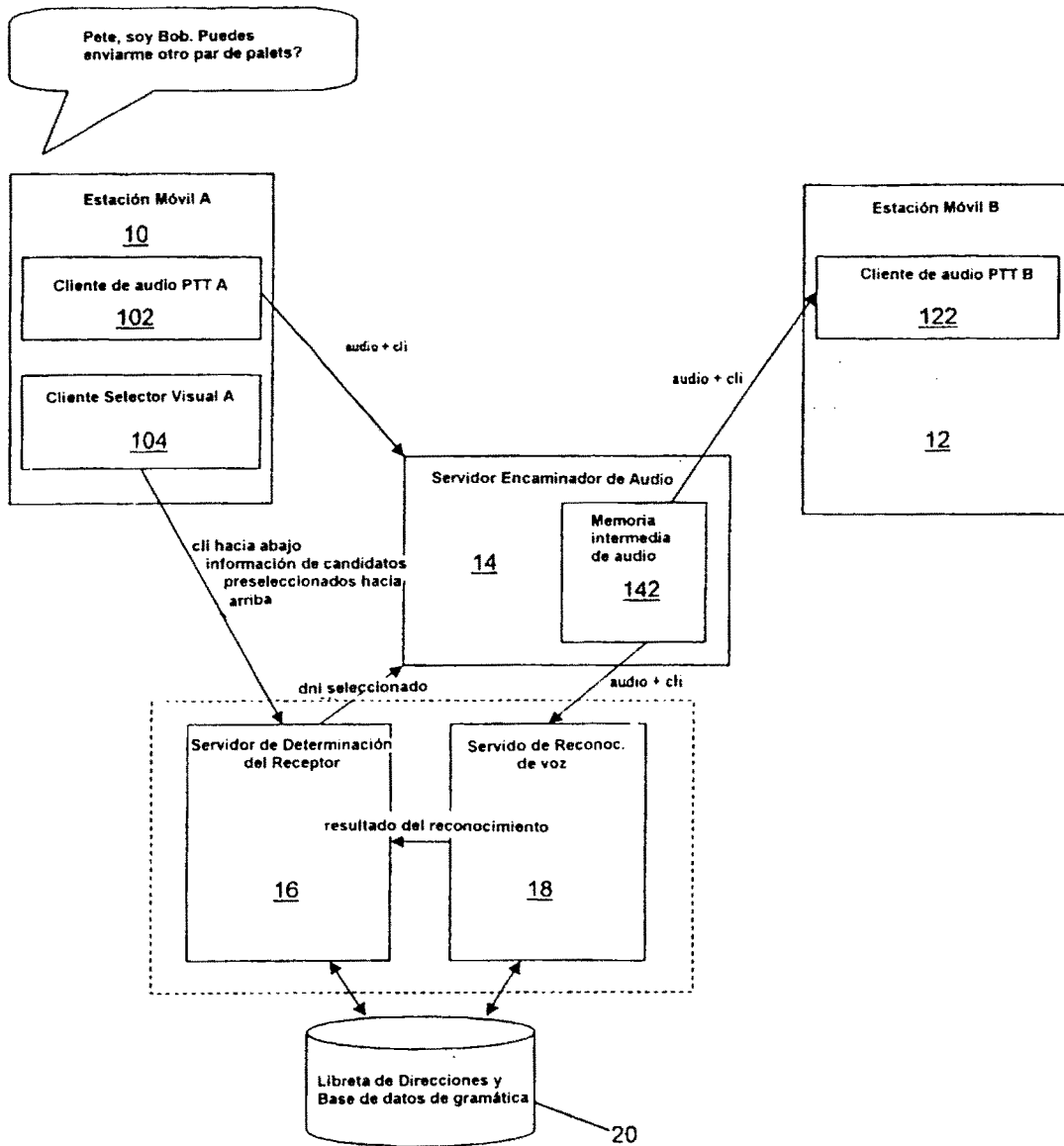
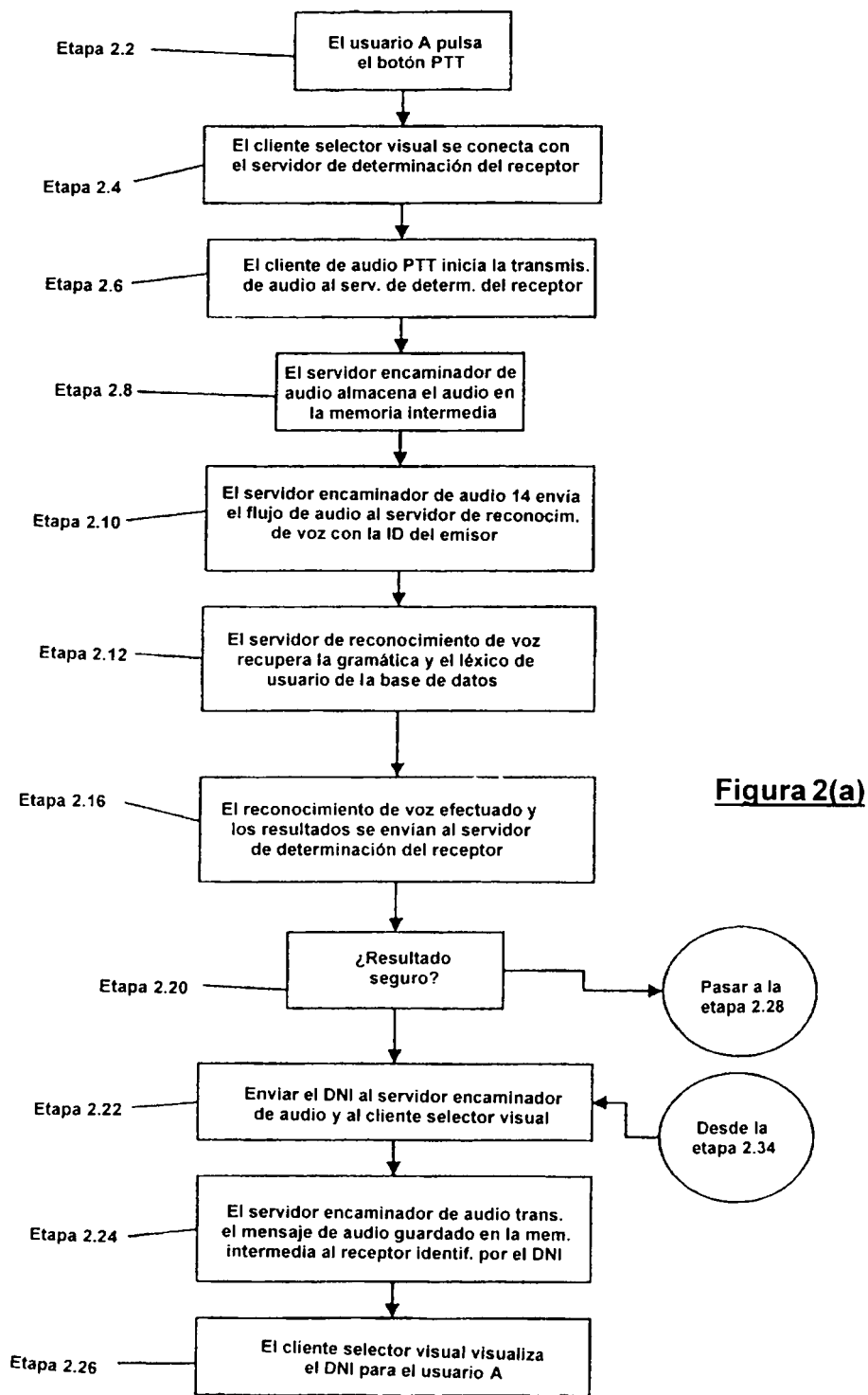


Figura 1



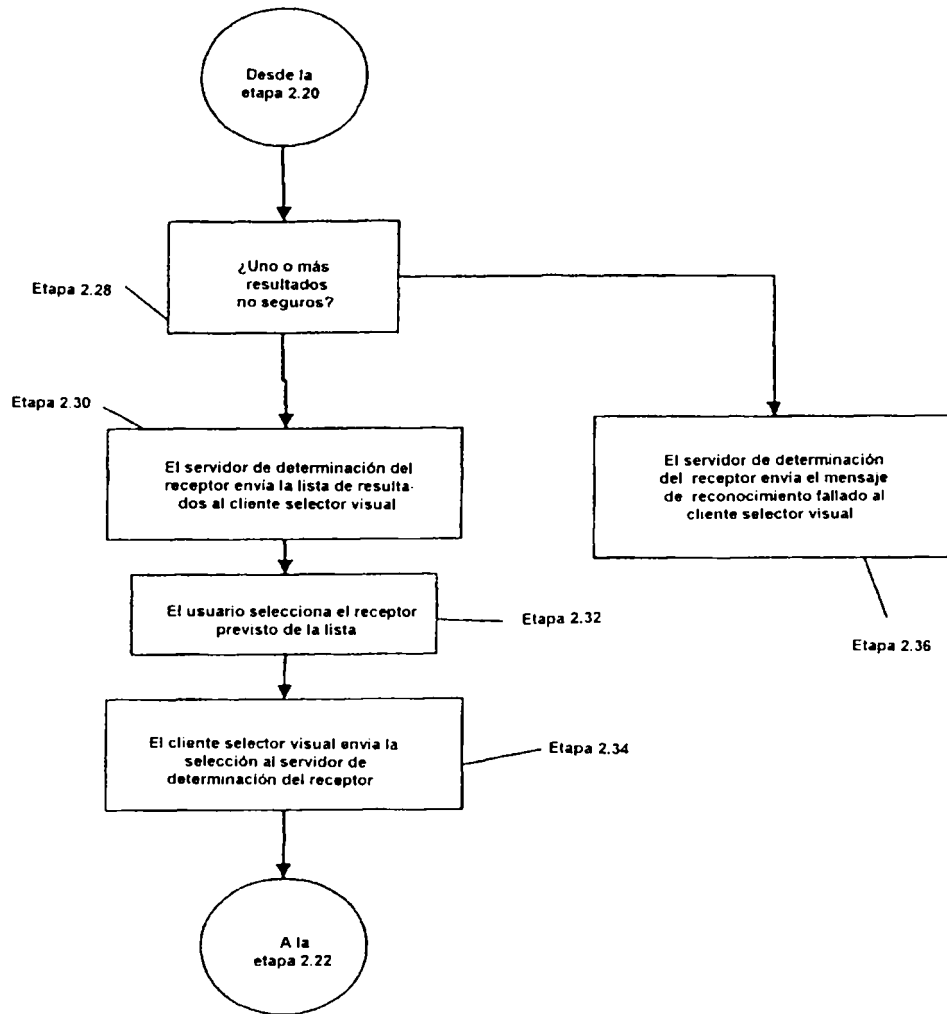


Figura 2(B)