(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 291 440**

(51) Int. Cl.:
G10L 15/26 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **02703691 .2**
(86) Fecha de presentación : **12.02.2002**
(87) Número de publicación de la solicitud: **1362343**
(87) Fecha de publicación de la solicitud: **19.11.2003**

(54) Título: **Procedimiento, módulo, dispositivo y servidor para reconocimiento de voz.**

(30) Prioridad: **13.02.2001 FR 01 01910**

(45) Fecha de publicación de la mención BOPI:
01.03.2008

(45) Fecha de la publicación del folleto de la patente:
01.03.2008

(73) Titular/es: **Thomson Licensing**
46, quai Alphonse Le Gallo
92100 Boulogne-Billancourt, FR

(72) Inventor/es: **Soufflet, Frédéric y**
Tazine, Nour-Eddine

(74) Agente: **Arpe Fernández, Manuel**

DESCRIPCIÓN

Procedimiento, módulo, dispositivo y servidor para reconocimiento de voz.

La presente invención se refiere al ámbito de los interfaces de voz.

Más específicamente, la invención se refiere a la utilización de modelos de lenguaje y/o de unidades fonéticas en terminales que utilizan el reconocimiento de voz.

Los sistemas de información o de control utilizan con cada vez mayor frecuencia un interfaz de voz para agilizar y/o hacer más intuitiva la interacción con el usuario. Debido a la creciente complejidad de estos sistemas, las exigencias en lo tocante al reconocimiento de voz son cada vez mayores, tanto en lo que se refiere a la amplitud del reconocimiento (un vocabulario muy grande) como a la rapidez del reconocimiento (en tiempo real).

En el estado actual de la técnica se conocen procedimientos de reconocimiento de voz basados en la utilización de modelos de lenguaje (probabilidad de que una palabra dada del vocabulario de la aplicación siga a otra palabra o grupo de palabras en el orden cronológico de la frase) y de unidades fonéticas. Estas técnicas se describen especialmente en la obra de Frederik Jelinek "Statistical Methods for Speech Recognition" (o "métodos estadísticos de reconocimiento de voz") publicado en la editorial MIT Press en 1997.

Estas técnicas se basan en unos modelos de lenguaje y en unas unidades fonéticas que se generan a partir de muestras de voz representativas (resultantes, por ejemplo, de una población de usuarios de un terminal a los que se hacen pronunciar los comandos).

En la práctica, los modelos de lenguaje deben tener en cuenta el estilo de locución normalmente utilizado por un usuario del sistema, y concretamente, de sus "defectos": titubeos, falsos inicios, cambios de parecer,...

La calidad de un modelo de lenguaje utilizado influye enormemente en la fiabilidad del reconocimiento de voz. Esta calidad se mide normalmente mediante un índice denominado perplejidad del modelo de lenguaje, y que representa esquemáticamente el número de elecciones que debe efectuar el sistema para cada palabra decodificada. Cuanto más baja sea dicha perplejidad, mayor será la calidad.

El modelo de lenguaje es necesario para traducir la señal de voz en una sucesión textual de palabras, una etapa que a menudo es utilizada por los sistemas de diálogo. Por tanto, es necesario construir una lógica de comprensión que permita comprender la solicitud para responder a la misma.

- el denominado método estadístico N-grama que normalmente utiliza un bi-grama o tri-grama, y que consiste en suponer que la probabilidad de que se de una palabra en la frase depende exclusivamente de las N palabras que la preceden, independientemente de su contexto dentro de la frase.

Si tomamos el ejemplo del tri-grama para un vocabulario de 1000 palabras, sería necesario definir 1000^3 probabilidades para definir el modelo de lenguaje, lo cual es imposible. Las palabras se agrupan entonces en conjuntos que bien son definidos explícitamente por el diseñador del modelo o bien son deducidos mediante métodos auto-organizativos.

Este modelo de lenguaje se construye automáticamente a partir de un corpus de texto.

Este tipo de modelo de lenguaje se utiliza principalmente para los sistemas de dictado de voz, cuya funcionalidad última es traducir la señal de voz en un texto, sin que sea necesaria una fase de comprensión.

- El segundo método consiste en describir la sintaxis por medio de una gramática probabilística, normalmente una gramática no contextual definida en virtud de una serie de reglas descritas en la llamada Notación de Backus Naur o notación BNF, o una extensión de esta forma a las gramáticas contextuales. Las reglas que describen gramáticas suelen estar manuscritas. Este tipo de modelo de lenguaje es apropiado para las aplicaciones de comando y control, en las que la etapa de reconocimiento está seguida por una etapa para control de un aparato o de búsqueda de información en una base de datos.

El modelo de lenguaje de una aplicación describe el conjunto de las expresiones (por ejemplo, frases) que se va a hacer que reconozca la aplicación. Un inconveniente de la técnica anterior es que si el modelo de lenguaje es de mala calidad, el sistema de reconocimiento, aunque tenga muy buenas prestaciones a nivel de descodificación acústico-fonética, tendrá unas mediocres prestaciones con determinadas expresiones.

Los modelos de lenguaje de tipo estocástico no tienen, por decirlo adecuadamente, una definición clara de las expresiones que se encuentran en el modelo de lenguaje, ni de aquellas que se encuentran fuera de él. Simplemente, algunas expresiones tienen una probabilidad de darse que *a priori* es mayor que la de otras.

Los modelos de lenguaje basados en una gramática probabilística proponen una clara diferencia entre expresiones pertenecientes a los modelos de lenguaje y expresiones externas al modelo de lenguaje. En estos modelos existen expresiones que no podrán ser reconocidas jamás, independientemente de la calidad de los modelos fonéticos utilizados.

Normalmente, son expresiones que no tienen ningún sentido o cuyo sentido escapa al ámbito de aplicación del sistema desarrollado.

El documento EP-A-945851 utiliza un sistema cliente-servidor para adaptar el vocabulario disponible y distribuirlo a todos los clientes cuando uno de los clientes desea añadir una palabra al vocabulario.

Puede comprobarse que los modelos de lenguaje de tipo probabilista y sus derivados son los más eficaces para las aplicaciones de comando y control. Estas gramáticas suelen ser manuscritas, y una de las principales dificultades del desarrollo de los sistemas de diálogos es proponer un modelo de lenguaje de buena calidad.

En particular, en lo que respecta a los modelos basados en gramática, no es posible definir de forma exhaustiva un lenguaje, sobre todo si este último es susceptible de ser utilizado por una gran población (por ejemplo, el caso de un mando a distancia para electrodomésticos). No es posible tener en cuenta todas las expresiones y giros posibles (desde el estilo elevado al argot) y/o errores gramaticales...

La invención se refiere a un procedimiento y un sistema de reconocimiento de voz que permite modificar y mejorar a distancia un modelo de lenguaje, a partir de la grabación de expresiones no reconocidas por el sistema.

Más concretamente, la invención tiene por objeto un procedimiento para reconocimiento de voz ejecutado al menos en un terminal, utilizando dicho procedimiento para reconocimiento de voz un modelo de lenguaje caracterizado porque comprende las siguientes etapas:

- detección de al menos una expresión no reconocida en uno de los terminales;

- grabación en el terminal de datos representativos de la expresión no reconocida;

- transmisión de los datos grabados desde el terminal a un servidor remoto, a través de un primer canal de transmisión;

- análisis, desde el servidor remoto, de los datos y la información generada para la corrección del modelo de lenguaje, teniendo en cuenta al menos una parte de las expresiones no reconocidas; y

- transmisión a través de un segundo canal de transmisión del servidor, al menos a un terminal, de la información de corrección, a fin de permitir reconocer en el futuro al menos determinadas expresiones no reconocidas.

De este modo, la invención se basa en un enfoque nuevo e inventivo del reconocimiento de voz, que permite actualizar los diferentes elementos que permiten el reconocimiento de voz en función de expresiones no reconocidas a nivel local, ya que el servidor remoto dispone de importantes recursos (por ejemplo, humanos y/o capacidades de cálculo) que generan las informaciones de corrección.

Se observa que en este caso los modelos de lenguaje comprenden:

- los modelos de lenguaje en sentido estricto (este es el caso, por ejemplo, cuando los datos que constituyen el objeto del reconocimiento son de tipo puramente textual);

- los modelos formados por uno o varios modelos de lenguaje en el sentido estricto y por uno o varios conjuntos de unidades fonéticas (lo que se corresponde, concretamente, con el caso general de un reconocimiento de voz aplicado a muestras de voz).

La invención va más allá de la mera actualización de un vocabulario. Efectivamente, es posible que dicha expresión no sea reconocida aunque todas las palabras de una expresión figuren en el vocabulario utilizado por el modelo de lenguaje del terminal. Únicamente la actualización del propio modelo de lenguaje permite entonces reconocer esta expresión posteriormente. La actualización del vocabulario, que es una de las informaciones de las que se deriva el modelo de lenguaje, no es suficiente.

En este caso, las expresiones deben tomarse en un sentido amplio, y se refieren a cualquier expresión de voz que permita la interacción entre un terminal y su usuario. Las expresiones (en inglés, “utterance”) incluyen, concretamente, las frases, las locuciones, las palabras aisladas o no, palabras de códigos específicos del terminal, instrucciones, comandos...

Las informaciones de corrección pueden incluir concretamente informaciones que permitan modificar parcial o totalmente el modelo de lenguaje y/o unidades fonéticas presentes en cada terminal suprimiendo, remplazando o añadiendo elementos.

El servidor puede recibir de cada terminal unos datos que le permitan mejorar el modelo de lenguaje y/o las unidades fonéticas presentes en el terminal transmisor de datos, pero también en el resto de los terminales, ya que cada uno de los terminales se beneficia de la experiencia común adquirida por el servidor a partir de todos los terminales.

De este modo, la invención permite tener en cuenta estilos de lenguaje o giros propios de determinados usuarios (por ejemplo, la expresión "20 horas de la tarde" (un pleonasmo difícilmente previsible *a priori*) en lugar de "las 20 horas" o "las 8 de la tarde") y que no habían sido previstos durante la construcción del modelo de lenguaje ejecutado.

Por otra parte, la invención tiene en cuenta la evolución de las lenguas vivas (nuevos giros o expresiones...).

Debe observarse que la invención se aplica igualmente a los modelos de lenguaje de tipo estocástico y a los modelos de lenguaje de tipo gramático probabilista. Cuando la invención se aplica a los modelos de lenguaje de tipo estocástico, los datos de corrección suelen ser muy numerosos para influir sobre el reconocimiento, mientras que los datos de corrección para un modelo basado en una gramática probabilista pueden ser poco numerosos y tener una importante influencia sobre la eficacia y la fiabilidad del reconocimiento.

De acuerdo con una característica específica, el procedimiento se caracteriza porque los datos representativos de las expresiones no reconocidas incluyen una grabación vocal comprimida representativa de parámetros descriptivos de la señal acústica.

De este modo, la invención permite tener en cuenta ventajosamente los datos vocales emitidos en origen mediante un análisis detallado a nivel del servidor, limitando el volumen de los datos transmitidos al servidor remoto.

De acuerdo con una característica específica, el procedimiento se caracteriza porque durante la etapa de transmisión mediante el terminal, este transmite adicionalmente al servidor, al menos, una de las informaciones que forman parte del grupo, y que incluyen:

- informaciones de contexto de utilización del procedimiento para reconocimiento de voz cuando no se reconoce una expresión; e

- informaciones relativas al interlocutor que ha pronunciado una expresión no reconocida.

De este modo, se facilita el reconocimiento de voz de las expresiones no reconocidas por el terminal, que puede efectuarse a distancia.

Asimismo, puede efectuarse en función del contexto una verificación de la validez del contenido de las expresiones no reconocidas. (por ejemplo, un comando "grabación de la emisión" tiene sentido y por tanto es válido cuando el terminal al que se dirige es un magnetoscopio, y no tiene sentido en caso de un teléfono móvil).

De acuerdo con una característica específica, el procedimiento se caracteriza porque ejecuta una encriptación y/o codificación de los datos grabados y/o de las informaciones de corrección.

De este modo, los datos se aseguran eficazmente y permanecen confidenciales.

La invención se refiere también a un módulo para reconocimiento de voz que utiliza un modelo de lenguaje, y que incluye:

- un analizador adaptado para detectar expresiones no reconocidas;

- un grabador de los datos representativos de al menos una expresión no reconocida;

- un transmisor adaptado para transmitir los datos grabados a un servidor remoto; y

- un receptor de informaciones de corrección que permite corregir el modelo de lenguaje transmitido al módulo, y que permite reconocer en el futuro, al menos, ciertas expresiones no reconocidas por el módulo, habiendo sido transmitidas las informaciones de corrección por el servidor remoto, tras analizar los datos al nivel del servidor remoto, y tras generar informaciones de corrección del modelo de lenguaje que tienen en cuenta al menos una parte de las expresiones no reconocidas.

La invención también se refiere a un dispositivo de reconocimiento de voz que utiliza un modelo de lenguaje, caracterizado porque comprende:

- un analizador adaptado para detectar expresiones no reconocidas;

- un grabador de los datos representativos de al menos una expresión no reconocida;

- un transmisor adaptado para transmitir los datos grabados a un servidor remoto; y

- un receptor de informaciones de corrección que permite corregir el modelo de lenguaje transmitido al dispositivo, y que permite reconocer en el futuro, al menos, ciertas expresiones no reconocidas por el dispositivo, habiendo sido transmitidas las informaciones de corrección por el servidor remoto, tras analizar los datos al nivel del servidor remoto, y tras generar informaciones de corrección del modelo de lenguaje que tienen en cuenta al menos una parte de las expresiones no reconocidas.

La invención también se refiere a un servidor de reconocimiento de voz, ejecutándose el reconocimiento en un conjunto formado por, al menos, un terminal remoto, utilizando un modelo de lenguaje, caracterizado porque comprende los medios siguientes:

- un receptor de datos representativos de, al menos, una expresión no reconocida por al menos un terminal que forma parte del conjunto formado por, al menos, un terminal remoto y que ha detectado la expresión no reconocida al realizar una operación de reconocimiento de voz; y

- un transmisor adaptado para transmitir hacia el conjunto formado, al menos, por un terminal remoto informaciones de corrección obtenidas a partir de un análisis de los datos recibidos a nivel del servidor, permitiendo las informaciones de corrección corregir, para cada uno de los terminales del conjunto, el modelo de lenguaje que permite reconocer en el futuro, al menos, parte de las expresiones no reconocidas.

Las características específicas y las ventajas del módulo, del dispositivo y del servidor de reconocimiento de voz, dado que son similares a las del procedimiento para reconocimiento de voz no se recuerdan en este documento.

Se apreciarán más claramente otras características y ventajas de la invención mediante la lectura de la siguiente descripción de un modo preferido de realización, facilitado a título de mero ejemplo ilustrativo y no limitativo, y las figuras adjuntas, en las cuales:

- la figura 1 presenta un diagrama general de un sistema que incluye un mando de control por voz, en el que puede ejecutarse la técnica de la invención;

- la figura 2 presenta un diagrama del sistema de control por voz del sistema de la figura 1;

- la figura 3 describe un esquema electrónico de un sistema de reconocimiento de voz que ejecuta el diagrama de la figura 2;

La figura 4 presenta un diagrama del servidor del sistema de la figura 1;

La figura 5 representa un organigrama del procedimiento de prueba de expresión y grabación de datos relativos a expresiones no reconocidas, como el ejecutado por el motor de reconocimiento de la figura 2;

La figura 6 representa un organigrama del procedimiento de transmisión de datos relativos a expresiones no reconocidas, como el ejecutado por el módulo de rechazo de la figura 2;

La figura 7 representa un organigrama del procedimiento de recepción de datos de corrección, como el ejecutado por el módulo de carga de los módulos de lenguaje de la figura 2; y

La figura 8 representa un organigrama del procedimiento de recepción y de procesamiento de los datos de corrección, como el ejecutado en el servidor remoto de la figura 4.

Por lo tanto, el principio general de la invención se basa en un reconocimiento de voz ejecutado en los terminales, utilizando el procedimiento para reconocimiento vocal un modelo de lenguaje y/o un conjunto de unidades fonéticas que pueden ser ejecutadas por un servidor remoto, cuando este último lo considere necesario.

En términos generales, cada terminal puede reconocer expresiones (por ejemplo, frases o comandos) formuladas por un interlocutor y ejecutar la correspondiente acción.

No obstante, a menudo se observa que ciertas expresiones, perfectamente comprensibles por un ser humano, no son reconocidas por el dispositivo o el módulo que ejecutan el reconocimiento de voz.

El fracaso del reconocimiento puede deberse a múltiples razones:

- el vocabulario utilizando por el interlocutor no forma parte del modelo de lenguaje;

- pronunciación particular (con acento, por ejemplo);

- giros específicos de expresión no previstos por el dispositivo o módulo para reconocimiento de voz;

- ...

Efectivamente, los modelos de lenguaje y los conjuntos de unidades fonéticas se construyen frecuentemente a partir de datos estadísticos que tengan en cuenta unas muestras de expresiones habitualmente utilizadas por una población típica, no siendo entonces tenidas en cuenta (o no pudiendo serlo) determinadas palabras del vocabulario, pronunciaciones y/o giros de frases.

La invención se basa, ante todo, en la detección de expresiones no reconocidas por el módulo o dispositivo de reconocimiento de voz.

Cuando una expresión no ha sido reconocida, el terminal graba datos representativos de la señal correspondiente a las expresiones no reconocidas (como por ejemplo una grabación digital de voz correspondiente a la expresión) para su transmisión a un servidor remoto.

Al nivel del servidor remoto que centraliza las expresiones no reconocidas de un conjunto de terminales, un operador humano puede entonces analizar las expresiones no reconocidas.

Algunas de ellas podrán resultar incomprensibles y/o no utilizables, por lo que serán descartadas.

En cambio, otras serán perfectamente comprensibles por el operador, que podrá, si lo considera conveniente, mediante la interacción hombre/máquina "traducir" dichas expresiones hasta entonces no reconocidas por los terminales, en un código que sea comprensible para el servidor.

El servidor podrá entonces tener en cuenta estas expresiones con su traducción, para generar informaciones de corrección del modelo de lenguaje y/o del conjunto de unidades fonéticas.

Se observará que la corrección se entiende aquí como:

- modificación del modelo; y/o

- complemento del modelo.

El servidor transmite entonces las informaciones de corrección a cada uno de los terminales, que puede actualizar su modelo de lenguaje y/o conjunto de unidades fonéticas, enriquecido con numerosas expresiones no reconocidas por él o por otros terminales.

De este modo, mejora el reconocimiento de voz de cada uno de los terminales, beneficiándose de la experiencia común a todos los terminales.

De acuerdo con un modo particular de la invención, el análisis no se efectúa a través de un operador, sino por el propio servidor, que puede disponer de recursos mucho más importantes que un simple terminal.

De acuerdo con unos modos específicos de realización, los terminales transmiten al servidor unos datos de contexto (por ejemplo, la hora, la fecha, un comando transmitido manualmente o a través de la voz tras el fracaso de un comando de voz, la ubicación, el tipo de terminal,...) con los datos representativos de la señal correspondiente a las señales no reconocidas.

Esto puede facilitar el trabajo de análisis del operador y/o del servidor.

Se presenta, en relación con la figura 1, un diagrama general de un sistema que incluye un mando de comandos de voz, en el que puede ejecutarse la técnica de la invención.

Este sistema incluye, especialmente:

- un servidor remoto 116 controlado por un operador humano 122; y

- una pluralidad de sistemas de usuarios 114, 117 y 118.

El servidor remoto 116 está conectado a cada uno de los sistemas de usuario 114, 117 y 118 a través de enlaces de comunicaciones descendentes, respectivamente 115, 119 y 120.

Estos enlaces pueden ser permanentes o temporales, y ser de cualquier tipo bien conocido por cualquier experto en la materia. Concretamente, pueden ser del tipo de difusión y estar basados en canales hertzianos, vía satélite o por cable, utilizados por la televisión o de cualquier otro tipo, como por ejemplo, una conexión de tipo Internet.

La figura 1 describe, concretamente, el sistema de usuario 114, que está conectado a través de un enlace 121 de comunicaciones ascendente con el servidor 116. Esta conexión puede ser también de cualquier tipo conocido por los expertos en la materia (concretamente, telefónico, por Internet,...).

El sistema de usuario 114 incluye:

- una fuente de voz 100, que puede estar constituida por un micrófono, destinado a captar una señal de voz generada por un interlocutor;

- un sistema de reconocimiento de voz 102;

- un sistema de control 105 destinado a controlar un aparato 107;

- un aparato controlado 107, como por ejemplo, un televisor, magnetoscopio o terminal de telecomunicaciones móviles;

- una unidad de almacenamiento 109 de las expresiones detectadas como no reconocidas;

- un interfaz 112 que permite las comunicaciones de subida y bajada hacia el servidor 116.

La fuente 100 está conectada al sistema de control de reconocimiento de voz 102, a través de una conexión 101 que le permite transmitir una onda fuente analógica representativa de una señal de voz hacia el sistema de control 102.

El sistema de control 102 puede recuperar las informaciones de contexto 104 (como por ejemplo, el tipo de aparato 107 que puede estar controlado por el sistema de control de comandos 105 o la lista de códigos de comandos) mediante una conexión 104 y transmitir comandos al sistema de control de comandos 105 a través de un enlace 103.

El sistema de control de comandos 105 transmite comandos a través de una conexión 106, por ejemplo, de infrarrojos, hacia el aparato 107, en función de las informaciones que reconoce en función del modelo de lenguaje y de su diccionario.

El sistema de control de comandos 105 detecta las expresiones que no reconoce, y en lugar de limitarse a rechazarlas, emitiendo una señal de falta de reconocimiento, efectúa una grabación de estas expresiones en la unidad de almacenamiento 109 a través de una conexión 108.

La unidad de almacenamiento 109 de las expresiones no reconocidas transmite los datos representativos hacia el interfaz 112 a través de una conexión 111, que los transmite hacia el servidor 116 a través de la conexión 121. Tras una transmisión correcta, el interfaz 110 puede transmitir una señal 110 hacia la unidad de almacenamiento 109, que puede entonces borrar los datos transmitidos.

El sistema de control de comandos 105 recibe, por otra parte, los datos de corrección del interfaz 112 a través de una conexión 113, que ha recibido el propio interfaz 112 del servidor remoto a través de la conexión 115. Estos datos de corrección son tenidos en cuenta por el sistema de control de comandos 105 para la actualización de modelos de lenguajes y/o de conjuntos de unidades fonéticas.

De acuerdo con el modo de realización considerado como origen 100, el sistema de control de reconocimiento de voz 102, el sistema de control de comandos 105, la unidad de almacenamiento 109 y el interfaz 112 forman parte de un mismo dispositivo, y de este modo, las conexiones 101, 103, 104, 108, 11, 110 y 113 son conexiones internas del dispositivo. La conexión 106 es típicamente una conexión inalámbrica.

De acuerdo con una primera variante de realización de la invención descrita en la figura 1, los elementos 100, 102, 105, 109 y 112 están parcial o completamente separados y no forman parte de un mismo dispositivo. En este caso, las conexiones 101, 103, 104, 108, 111, 110 y 113 son conexiones externas inalámbricas o no.

De acuerdo con una segunda variante, la fuente 100, los sistemas de control 102 y 105, la unidad de almacenamiento 109 y el interfaz 112, así como el aparato 107 forman parte de un mismo dispositivo y están conectados entre sí mediante buses internos (conexiones 101, 103, 104, 108, 111, 110, 113 y 106). Esta variante resulta especialmente interesante cuando el dispositivo es, por ejemplo, un teléfono móvil o un terminal de telecomunicaciones portátil.

La figura 2 presenta un organigrama de un sistema de control de comandos de voz como el sistema de control 102 mostrado en relación con la figura 2.

Se observa que el sistema de control 102 recibe del exterior la onda fuente analógica 101 que es procesada por un Decodificador Acústico-Fonético 200 o DAP (denominado “front end” en inglés). El DAP 200 muestrea a intervalos regulares (normalmente cada 10 ms) la onda fuente 101 para generar unos vectores reales o pertenecientes a unos libros de código (o “code books” en inglés) que representan típicamente resonancias bucales transmitidas a través de una conexión 201 hacia un motor de reconocimiento 203. El DAP se basa, por ejemplo, en una PLP (en inglés “Perceptual Linear Prediction [predicción lineal de percepción]”) descrita especialmente en el artículo “Perceptual Linear Prediction (PLP) analysis of speech [predicción lineal de percepción de análisis de voz]” descrito por Hynek Hermansky y publicado en el “Journal of the Acoustical Society of America” Vol. 97, nº 4, 1990, páginas 1738-1752.

Con la ayuda de un diccionario 202, el motor de reconocimiento 203 analiza los vectores reales recibidos utilizando sobre todo modelos de Markov ocultos o HMM (en inglés “Hidden Markov Models”) y modelos de lenguaje (que representan la probabilidad de que una palabra siga a otra palabra). Los motores de reconocimiento se describen en detalle en el libro “Statistical Methods for Speech Recognition”, escrito por Frederick Jelinek y publicado por ediciones MIT Press en 1997.

El modelo de lenguaje permite al motor de reconocimiento 203 (que puede utilizar redes de Markov ocultas) determinar qué palabras pueden seguir a una palabra dada en cualquier expresión utilizable por el interlocutor en una aplicación dada y ofrecer la probabilidad asociada. Las palabras en cuestión pertenecen al vocabulario de la aplicación

que puede ser, independientemente del modelo de lenguaje, de pequeñas dimensiones (normalmente de 10 a 300 palabras) o de grandes dimensiones (por ejemplo con un tamaño superior a 300.000 palabras).

La solicitud de patente PCT/FR00/03329 de fecha 29 de noviembre de 1999 presentada en nombre de THOMSON MULTIMEDIA y publicada con el número de publicación WO-A-01/41125, describe un modelo de lenguaje que incluye una pluralidad de bloques sintácticos. La utilización de la invención que constituye el objeto de la presente solicitud resulta especialmente ventajosa cuando se combina con este tipo de modelo de lenguaje modular, pues los módulos pueden actualizarse por separado, lo que evita la carga remota de archivos con unas dimensiones demasiado grandes.

Los modelos de lenguaje se transmiten mediante un módulo 207 de carga de modelo de lenguaje. El propio módulo 207 recibe modelos de lenguaje, actualizaciones o correcciones de modelo de lenguaje y/o de unidades fonéticas transmitidas desde el servidor a través de la conexión 113.

Cabe señalar que el diccionario 202 pertenece al modelo de lenguaje que hace referencia a palabras del diccionario. De este modo, el propio diccionario 202 puede ser actualizado y/o corregido a través de un modelo de lenguaje cargado por el módulo 207.

Tras la ejecución de una operación de reconocimiento basada en la utilización de un algoritmo de Viterbi, el motor de reconocimiento 203 facilita al módulo de rechazo 211 una lista ordenada de secuencias de palabras acordes con el modelo de lenguaje, que presenta la mejor valoración para la expresión pronunciada.

El módulo de rechazo 211 se ejecuta con posterioridad al motor de reconocimiento 203 y opera de acuerdo con uno o varios de los siguientes principios:

- A veces, por razones propias del algoritmo de Viterbi, este no puede generar una lista coherente, ya que las puntuaciones son tan reducidas que se sobrepasa el límite de las precisiones aceptables para la máquina de cálculo aritmético. Por tanto, no se dispone de una proposición completa coherente. De este modo, cuando el módulo de rechazo 211 detecta una o varias puntuaciones inferiores a un límite aceptable predeterminado, se rechaza la expresión.

Cada elemento de la lista calculada mediante el algoritmo de Viterbi se ha retenido debido a que la puntuación asociada se encontraba entre las puntuaciones relativas más altas de todas las expresiones posibles, de acuerdo con el modelo de lenguaje. Por otra parte, la red de Markov asociada a cada una de estas expresiones permite evaluar la probabilidad intrínseca de que la red en cuestión genere la expresión asociada con la puntuación constatada. El módulo de rechazo 211 analiza esta probabilidad y si esta es inferior a un límite inferior predeterminado de aceptabilidad de la probabilidad, se rechaza la expresión.

De acuerdo con otro método, para las mejores proposiciones, obtenidas mediante el algoritmo de Viterbi, el módulo de rechazo 211 lleva a cabo un procesamiento complementario de las expresiones, utilizando unos criterios que no se habían tenido en cuenta durante el desarrollo del algoritmo de Viterbi. Por ejemplo, verifica que las partes de la señal que deben ser sordas porque están asociadas a vocales, lo son efectivamente. Si las expresiones propuestas no cumplen estas condiciones, son rechazadas.

Cuando el módulo de rechazo 211 rechaza una expresión, como se ha mostrado anteriormente, la expresión se clasifica como no reconocida y se transmite una señal indicadora de la expresión rechazada al motor de reconocimiento 203. Paralelamente, el módulo de rechazo transmite una grabación de la expresión no reconocida a la unidad de almacenamiento 109 a través de la conexión 108.

El motor de reconocimiento 203 se encarga del reconocimiento de las expresiones transmitidas por el DAP 200 en forma de muestras fonéticas. De este modo, el motor de reconocimiento 203 utiliza:

- las unidades fonéticas para construir la representación fonética de una palabra en forma de un modelo de Markov, pudiendo poseer cada palabra del diccionario 202 varias “vocalizaciones” y simultáneamente,

- el modelo de lenguaje en el sentido estricto para reconocer expresiones más o menos complejas.

El motor de reconocimiento 203 facilita expresiones que han sido reconocidas (es decir, no rechazadas por el módulo 211) y que este ha identificado a partir de los vectores recibidos por un medio 205 de conversión de estas expresiones en comandos que puedan ser comprendidos por el aparato 107. Este medio 205 utiliza un procedimiento de traducción mediante inteligencia artificial que él mismo tiene en cuenta y un contexto 104 suministrado por el sistema de control 105 antes de transmitir uno o varios comandos 103 al sistema de control 105.

La figura 3 muestra esquemáticamente un módulo o dispositivo de reconocimiento de voz 102 como el mostrado en relación con la figura 1 y que ejecuta el diagrama de la figura 2.

El sistema de control 102 incluye, conectados entre sí mediante un bus de direcciones y de datos:

- un interfaz de voz 301;

- un convertidor analógico-digital 302;

- un procesador 304;

- una memoria no volátil 305;

- una memoria RAM 306;

- un módulo de recepción 312;

- un módulo de transmisión 313; y

- un interfaz de entrada/salida 307.

Todos los elementos mostrados en la figura 3 son bien conocidos por cualquier experto en la materia. Estos elementos comunes no se describirán aquí.

Por otra parte se observa que la palabra "registro" utilizada en toda la descripción designa en cada una de las memorias mencionadas tanto un área de memoria de poca capacidad (algunos datos binarios) como un área de memoria de gran capacidad (que permite almacenar todo un programa o la totalidad de una secuencia de datos de operaciones).

La memoria no volátil 305 (o ROM) conserva el programa de operación del procesador 304 en un registro "*prog*" 308.

La memoria RAM 306 conserva datos variables y resultados intermedios de procesamiento en registros que por comodidad poseen los mismos nombres que los datos que conservan y que incluyen:

- un registro 309 en el que se conservan grabaciones de expresiones no reconocidas, *Exp_Non_Rec;*

- un contador 310 de frases no reconocidas *Nb_Exp_Non_Rec;* y

- un modelo de lenguaje en un registro 311, *Modèle_Langage.*

Por otra parte se observa que los módulos de recepción 312 y de transmisión 313 son módulos que permiten la transmisión de datos respectivamente a partir de o hacia el servidor remoto 116. Las técnicas de recepción y de transmisión por cable o inalámbricas son bien conocidas por cualquier experto en telecomunicaciones por lo que no serán detalladas.

La figura 4 presenta el servidor 116 del sistema mostrado en relación con la figura 1.

Se observa que el servidor 116 está controlado por un operador humano 122 mediante un interfaz hombre-máquina 404 cualquiera (por ejemplo, del tipo de teclado y pantalla).

El servidor 116 también incluye:

- un receptor 400;

- un analizador 401;

- un módulo 402 de construcción de corrección de modelo de lenguaje y/o de un conjunto de unidades fonéticas; y

- un transmisor 403.

El receptor 400 es compatible con el transmisor 313 de un terminal y puede recibir de cada terminal datos representativos (por ejemplo grabaciones) de expresiones no reconocidas mediante la conexión 121 y eventualmente datos complementarios (por ejemplo contextuales).

El analizador 401 recibe el conjunto de estos datos del receptor 400 a través de una conexión 121 y lo transmite al operador 122 a través de un interfaz que es, por ejemplo, un terminal equipado con:

- una pantalla y un teclado que permiten dialogar con el servidor 116 y controlarlo;

- altavoces o auriculares que permiten la escucha de las grabaciones no reconocidas.

Este interfaz también permite al analizador 401 recibir una información del operador 122 que indique:

- que una expresión no reconocida y no incluida en el modelo de lenguaje sigue siendo incomprensible, no tiene sentido en el contexto de la aplicación al terminal y/o no se refiere al terminal (por tanto no debe incluirse en el

modelo de lenguaje), no teniéndose entonces en cuenta esta expresión para la corrección del modelo de lenguaje y siendo descartada por el analizador 401;

- que una expresión no reconocida pertenece no obstante al modelo de lenguaje en sentido estricto (se trata pues de un problema de reconocimiento puro); en este caso, habrá que modificar las unidades fonéticas y no el modelo de lenguaje en sentido estricto; o

- una traducción, por ejemplo en forma de código de comando, tras la identificación del contenido de una expresión por el operador, no perteneciendo al modelo de lenguaje la expresión no reconocida y teniendo sentido para el terminal al cual estaba destinada; por tanto habrá de corregirse el modelo de lenguaje en sentido estricto.

Es posible una combinación de la segunda y tercera solución; en este caso será necesario modificar simultáneamente las unidades fonéticas y el modelo de lenguaje en sentido estricto.

Este modo de realización corresponde a un procesamiento manual de las expresiones no reconocidas. Según este modo de realización, el operador humano escucha la expresión no reconocida y analiza las razones de su rechazo. El operador 122 determina si la expresión pertenece o no al modelo de lenguaje. En el caso de que la expresión pertenezca al modelo de lenguaje, el operador analiza la expresión para identificar el problema de reconocimiento intrínseco (expresión que pertenece al modelo de lenguaje y que debería haber sido reconocida y que no lo ha sido por otras razones: ruido, acento del interlocutor...).

De acuerdo con una primera variante, el procesamiento es automático y no se produce la intervención de un operador humano. En este caso, el servidor 116, y concretamente el analizador 401, poseen una potencia de cálculo relativamente importante que puede ser mucho mayor que la de un terminal. Según esta variante, el analizador 401 analiza cada expresión no reconocida de forma más adecuada de lo que podría hacerlo un terminal utilizando, por ejemplo, un modelo de lenguaje más rico y/o unos modelos fonéticos más complejos. Al no estar sometido a unas exigencias de cálculo en tiempo real tan estrictas como a las que podría estarlo un terminal (que a menudo precisa un tiempo de reacción rápido a un comando de voz), el analizador 401 también puede, por ejemplo, permitir un reconocimiento que requiera un tiempo de procesamiento más largo que en el caso de un terminal.

De acuerdo con una segunda variante, el procesamiento es semiautomático y la intervención de un operador humano se limita a los casos que no puede resolver el analizador.

La estructura general de un servidor 116 es, según el modo preferido de realización descrito en este documento, similar a la de un terminal como el descrito en relación con la figura 3 e incluye, conectados entre sí mediante un bus de direcciones y de datos:

- un procesador;

- una memoria RAM;

- una memoria no volátil;

- un módulo de transmisión adaptado;

- un módulo de recepción; y

- un interfaz hombre-máquina.

De acuerdo con la figura 5 que representa un organigrama de prueba de expresión y de grabación de datos relativos a expresiones no reconocidas como el ejecutado por el motor de reconocimiento 203 de la figura 2, durante una primera etapa de inicialización 500 el microprocesador 304 comienza la ejecución del programa 308 e inicializa las variables de la memoria RAM 306.

Posteriormente, durante una etapa 501 de espera de expresiones, está a la espera y recibe una expresión transmitida por un interlocutor.

Seguidamente, durante una prueba 502, tras haber ejecutado una operación de reconocimiento de voz de la expresión recibida, determina si la expresión ha sido reconocida o no de acuerdo con uno o varios criterios mostrados con respecto a la descripción del modelo de rechazo 211 de la figura 2.

En caso afirmativo, durante una fase 504 de comando, el terminal 102 tiene en cuenta el resultado del reconocimiento de voz aplicado a la expresión recibida y ejecuta la acción apropiada como por ejemplo un comando.

En caso negativo, durante una fase 503 de grabación de la expresión, la expresión no reconocida se comprime y graba en la unidad de almacenamiento 109 a la espera de una transmisión hacia el servidor remoto 116 como se muestra con respecto a la figura 6.

Al final de una de las etapas 503 o 504, se repite la etapa 501 de espera de expresiones.

La figura 6 representa un organigrama de transmisión de datos relativos a expresiones no reconocidas como el ejecutado por el módulo de rechazo de la figura 2. Durante una primera etapa de inicialización 600, el microprocesador 304 inicia la ejecución del programa 308 e inicializa las variables de la memoria RAM 306.

Posteriormente, durante una etapa 601 de espera de expresiones no reconocidas por el módulo para reconocimiento de voz 102, el microprocesador 304 espera y posteriormente recibe grabaciones de las expresiones no reconocidas.

Más tarde, durante una etapa 602, el terminal 114 se conecta al servidor remoto 116 de acuerdo con unos métodos bien conocidos por cualquier experto en telecomunicaciones.

Seguidamente, durante una etapa 603, las grabaciones de expresiones no reconocidas se formatean y transmiten al servidor remoto 116.

A continuación, durante una etapa 604 de desconexión, el termina se desconecta del servidor remoto 116 y se transmite un señal entre el interfaz 112 con el servidor remoto y la unidad 109 de almacenamiento de los datos correspondientes a las expresiones no reconocidas, indicando la transmisión de las grabaciones de expresiones. Los datos correspondientes a estas expresiones son entonces borrados de la unidad de almacenamiento 109.

Finalmente, se repite la etapa 601.

La figura 7 representa un organigrama de recepción de los datos de corrección, como el ejecutado por el módulo 207 de carga de los modelos de lenguaje de la figura 2.

Tras una primera etapa 700 de inicialización, y durante una fase etapa, el terminal se pone a la espera de los datos de corrección transmitidos por un servidor 116 a una pluralidad de terminales.

Seguidamente, en una etapa 702, el terminal tiene en cuenta los datos de corrección para actualizar el modelo de lenguaje y/o su conjunto de unidades fonéticas utilizadas por el modulo de reconocimiento de voz. En función de la naturaleza de los datos de corrección, estos datos pueden:

- sustituir los datos existentes en el modelo de lenguaje y/o su conjunto de unidades fonéticas;

- modificar los datos existentes;

- completar los datos existentes, y/o

- provocar la supresión de los datos existentes.

Tras la ejecución de la etapa 702, se repite la etapa 703.

La figura 8 representa un organigrama de recepción y procesamiento de los datos de corrección, como el ejecutado en el servidor remoto de la figura 4.

Tras una primera etapa 800 de inicialización de los parámetros y de ejecución de un programa de gestión del servidor, el servidor 116 se pone a la espera de una petición de conexión procedente de un terminal (que ejecuta una etapa 602 mostrada en la figura 6) y establece una conexión con el terminal de acuerdo con unos métodos bien conocidos por cualquier experto en telecomunicaciones.

Posteriormente, durante una etapa 802, el servidor 116 recibe los datos procedentes del terminal conectado que ejecuta la etapa 603 descrita anteriormente. Estos datos contienen principalmente grabaciones de una o varias expresiones rechazadas por el terminal al no haber sido reconocidas por un módulo para reconocimiento de voz ejecutado en el terminal. Cuando se reciben todos los datos, se interrumpe la conexión entre el terminal y el servidor 116.

Seguidamente, durante una etapa 803 de procesamiento de los datos recibidos, al nivel del servidor 116, cada una de las grabaciones de expresiones recibidas se procesa bien manualmente a través del operador 122 o bien automática o semi-automáticamente mediante diferentes variantes mostradas en relación con la figura 4.

Posteriormente, durante una prueba 804, el servidor 116 determina si una o varias de las expresiones recibidas han podido comprenderse y son pertinentes para el terminal que ha transmitido esta expresión o expresiones. En este caso será necesario actualizar los modelos de lenguajes y/o de unidades fonéticas.

En caso contrario, se repetirá la etapa de espera 801.

En caso afirmativo, el servidor 116 construye una corrección del modelo de lenguaje que puede adoptar varias formas que permitan una etapa 607 (mostrada anteriormente) en los terminales tras la recepción de los datos de corrección. Estos datos de corrección incluyen:

- un indicador que precise la naturaleza de la corrección (a saber, sustitución, modificación, complemento o supresión); y

- los datos de las propias correcciones en función del indicador.

Se observa que si el modelo de lenguaje incluye una pluralidad de bloques sintácticos (como es el caso de los modelos de lenguaje descritos en la solicitud internacional WO-A-01/41125 mencionada anteriormente), cada módulo podrá corregirse por separado. En este caso, los datos de corrección también incluyen un indicador del módulo o módulos que deben corregirse).

Después, durante una etapa 806, el servidor 116 transmite los datos de corrección hacia un terminal, o preferiblemente un conjunto de terminales que pueden actualizar su modelo de lenguaje y/o conjunto de unidades fonéticas de acuerdo con una etapa 607.

A continuación se repite la etapa 801.

Por ello el proceso es iterativo y puede repetirse varias veces. También permite que la aplicación evolucione al añadir nuevas solicitudes.

Es evidente que la invención no se limita a los ejemplos de realización mencionados anteriormente.

Concretamente, cualquier experto en la materia podrá introducir cualquier variante en la definición de los terminales que ejecutan la invención, dado que la invención se refiere a cualquier tipo de dispositivo y/o módulo que utilice o pueda utilizar un procedimiento para reconocimiento de voz (por ejemplo ser del tipo terminal multimedia, televisor, magnetoscopio, decodificador digital multimedia (en inglés “set top box”), equipo de audio o vídeo, terminal fijo o móvil...).

Igualmente, la invención se refiere a cualquier tipo de servidor remoto (por ejemplo, servidores de Internet, equipos conectados a transmisores de programas de televisión, equipos conectados a redes de comunicaciones móviles, equipos de proveedores de servicios...).

Además, de acuerdo con la invención, el canal de transmisión de los datos correspondientes a las frases no reconocidas y el canal de transmisión de los datos de corrección de los modelos de lenguajes y/o unidades fonéticas puede ser de cualquier tipo e incluyen:

- las vías de transmisiones por ondas hertzianas;

- las vías de transmisión por satélite;

- los canales de redes de transmisión de televisión;

- los canales de redes de tipo Internet;

- los canales de redes telefónicas;

- los canales de redes móviles; y

- los soportes de medios removibles.

Asimismo, cabe señalar que la invención no se refiere solamente a las frases no reconocidas sino a cualquier tipo de expresión de voz, como por ejemplo una o varias frases, una palabra aislada o no, una locución y un código de voz que permita el diálogo entre una máquina y su usuario. Estas expresiones verbales pueden estar asociadas no solamente a comandos sino a cualquier tipo de datos que puedan constituir el objeto de un diálogo entre una máquina y su usuario, como por ejemplo datos de informaciones que el usuario puede transmitir a la máquina, de configuración, de programación...

También debe señalarse que el método de actualización de los modelos de lenguaje descrito en la patente se aplica no sólo a los procedimientos de reconocimiento de voz en sentido estricto sino que también se aplica a procedimientos de reconocimiento de textos introducidos que soporten faltas de ortografía y/o mecanografía, basados también en procedimientos Markovianos o modelos de lenguaje en sentido estricto como los descritos en la patente.

Se observará que la invención no se limita a una ejecución puramente física, sino que también puede ejecutarse en forma de una secuencia de instrucciones de un programa informático o en cualquier forma que combine una parte física y una parte lógica. En el caso de que la invención se ejecute total o parcialmente en forma de software, la correspondiente secuencia de instrucciones podrá almacenarse en un medio de almacenamiento removible (como por ejemplo un disquete, un CD-ROM o un DVD-ROM) o no, pudiendo dicho medio de almacenamiento ser leído total o parcialmente por un ordenador o un microprocesador. La invención está limitada exclusivamente por el texto de las reivindicaciones adjuntas.

Referencias citadas en la descripción

La lista de referencias citada por el solicitante lo es solamente para utilidad del lector, no formando parte de los documentos de patente europeos. Aún cuando las referencias han sido cuidadosamente recopiladas, no pueden excluirse errores u omisiones y la OEP rechaza toda responsabilidad a este respecto.

Documentos de patente citado en la descripción

- EP 945851 A [0016]
- WO 0141125 A [0077] [0126]
- FR 0003329 W, THOMSON MULTIMEDIA [0077]

Bibliografía de patentes citada en la descripción

- FREDERIK **JELINEK**. Statistical methods for speech recognition. MIT Press, 1997 [0004]
- FREDERICK **JELINEK**. Statistical Methods for Speech Recognition. MIT Press, 1997 [0075]
- HYNEK **HERMANSKY**. Perceptual Linear Prediction (PLP) analysis of speech. *Journal of the Acoustical Society of America*, 1990, vol. 97 (4), 1738-1752 [0074]

REIVINDICACIONES

1. Procedimiento para reconocimiento de voz ejecutado al menos en un terminal (114), utilizando dicho procedimiento para reconocimiento de voz un modelo de lenguaje (311) **caracterizado** porque comprende las siguientes etapas:

- detección (502) de al menos una expresión no reconocida en uno de dichos terminales;

- grabación (503) en dicho terminal de datos representativos de dicha expresión no reconocida (309);

- transmisión (603) de dichos datos grabados desde dicho terminal a un servidor remoto (116), a través de un primer canal de transmisión (121);

- análisis (803), desde dicho servidor remoto, de dichos datos y la información generada (805) para corregir el modelo de lenguaje, teniendo en cuenta al menos una parte de dichas expresiones no reconocidas; y

- transmisión (806) a través de un segundo canal de transmisión (115, 119, 120) de dicho servidor, al menos a un terminal (114, 117, 118), de dicha información de corrección, a fin de permitir reconocer en el futuro al menos algunas de dichas expresiones no reconocidas.

2. Procedimiento de acuerdo con la reivindicación 1, **caracterizado** porque dichos datos representativos de dichas expresiones no reconocidas (309) incluyen una grabación de voz comprimida que representa parámetros descriptivos de la señal acústica.

3. Procedimiento de acuerdo con cualquiera de las reivindicaciones 1 y 2, **caracterizado** porque durante dicha fase de transmisión a través de dicho terminal, este transmite también a dicho servidor al menos una de las informaciones que forman parte de un grupo que comprende:

- informaciones de contexto de utilización de dicho procedimiento para reconocimiento de voz cuando no se ha reconocido una expresión; e

- informaciones relativas al interlocutor que ha pronunciado una expresión no reconocida.

4. Procedimiento de acuerdo con cualquiera de las reivindicaciones 1 a 3, **caracterizado** porque ejecuta un encriptación y/o codificación de dichos datos grabados y/o de dichas informaciones de corrección.

5. Módulo para reconocimiento de voz (102) que utiliza un modelo de lenguaje, **caracterizado** porque incluye:

- un analizador adaptado para detectar las expresiones no reconocidas;

- un grabador de datos representativos de al menos una expresión no reconocida;

- un transmisor adaptado para transmitir dichos datos grabados a un servidor remoto; y

- un receptor de informaciones de corrección que permite corregir dicho modelo de lenguaje transmitido hacia dicho módulo y que permite reconocer en el futuro al menos algunas de dichas expresiones no reconocidas por dicho módulo, habiendo sido transmitidas las informaciones de corrección por el servidor remoto después de analizar en dicho servidor remoto dichos datos y tras generar informaciones de corrección de dicho modelo de lenguaje, teniendo en cuenta al menos una parte de las expresiones no reconocidas.

6. Dispositivo de reconocimiento de voz (102) que utiliza un modelo de lenguaje, **caracterizado** porque incluye:

- un analizador adaptado para detectar las expresiones no reconocidas;

- un grabador de datos representativos de al menos una expresión no reconocida;

- un transmisor adaptado para transmitir dichos datos grabados a un servidor remoto; y

- un receptor de informaciones de corrección que permite la corrección de dicho modelo de lenguaje transmitido hacia dicho módulo y que permite reconocer en el futuro al menos algunas de dichas expresiones no reconocidas por dicho módulo, habiendo sido transmitidas las informaciones de corrección por el servidor remoto después de analizar en dicho servidor remoto dichos datos y tras generar informaciones de corrección de dicho modelo de lenguaje teniendo en cuenta al menos parte de las expresiones no reconocidas.

7. Servidor de reconocimiento de voz (116), ejecutándose dicho reconocimiento en un conjunto formado al menos por un terminal remoto, utilizando un modelo de lenguaje, **caracterizado** porque incluye los medios siguientes:

- un receptor de datos representativos de al menos una expresión no reconocida por al menos un terminal que forma parte de dicho conjunto y que ha detectado dicha expresión no reconocida durante una operación de reconocimiento de voz; y

- un transmisor adaptado para transmitir a dicho conjunto formado, al menos, por un terminal remoto, informaciones de corrección obtenidas a partir de un análisis de dichos datos recibidos por dicho servidor, permitiendo dichas informaciones de corrección corregir dicho modelo de lenguaje por parte de cada uno de los terminales de dicho conjunto, y permitiendo dicho modelo de lenguaje reconocer en el futuro al menos parte de las expresiones no reconocidas.

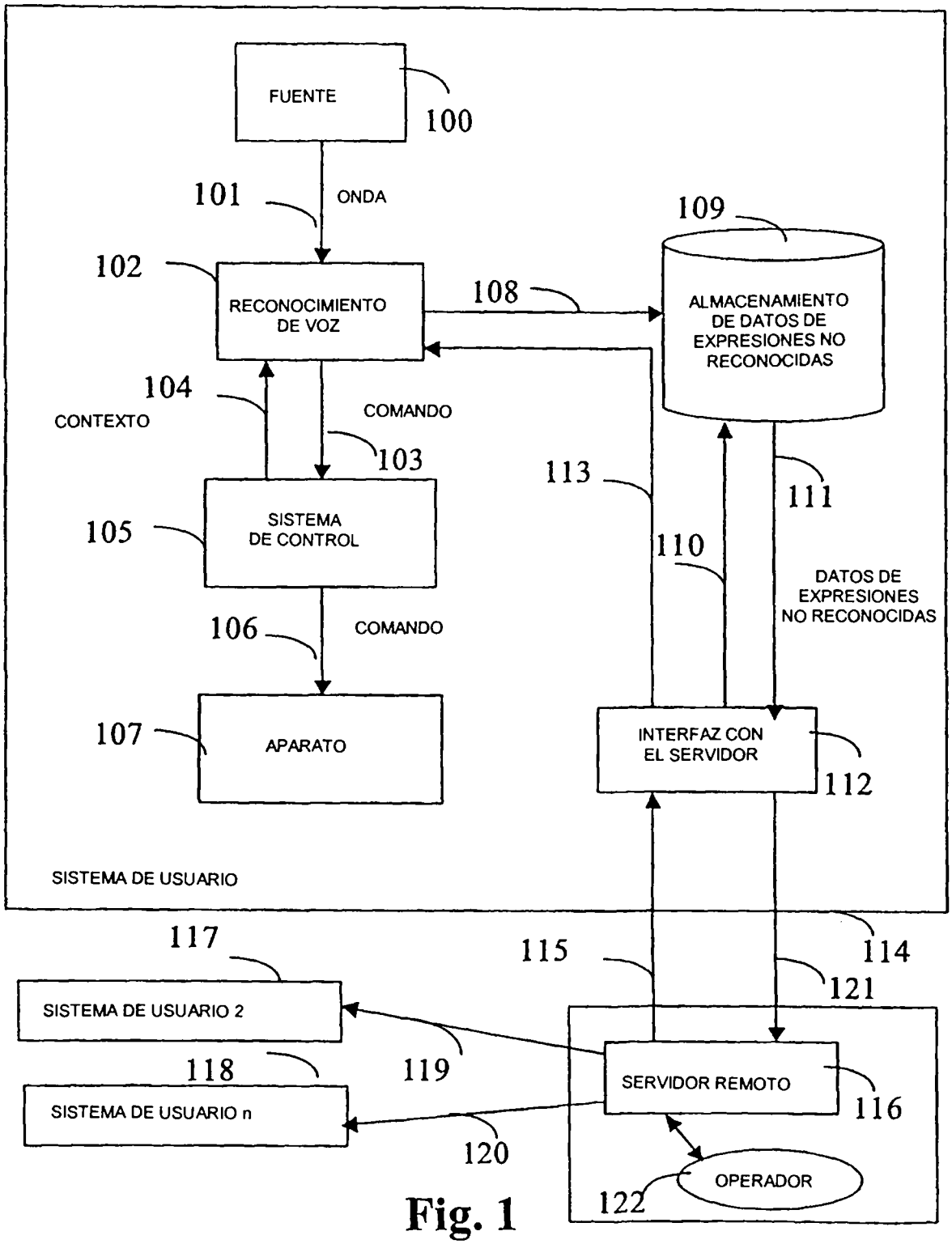
FUENTE
100
101
ONDA
102
RECONOCIMIENTO DE VOZ
108
109
ALMACENAMIENTO DE DATOS DE EXPRESIONES NO RECONOCIDAS
104
CONTEXTO
COMANDO
103
113
111
105
SISTEMA DE CONTROL
110
DATOS DE EXPRESIONES NO RECONOCIDAS
106
COMANDO
107
APARATO
INTERFAZ CON EL SERVIDOR
112
SISTEMA DE USUARIO
117
115
114
121
SISTEMA DE USUARIO 2
118
119
SERVIDOR REMOTO
116
SISTEMA DE USUARIO n
120
OPERADOR
122

Fig. 1

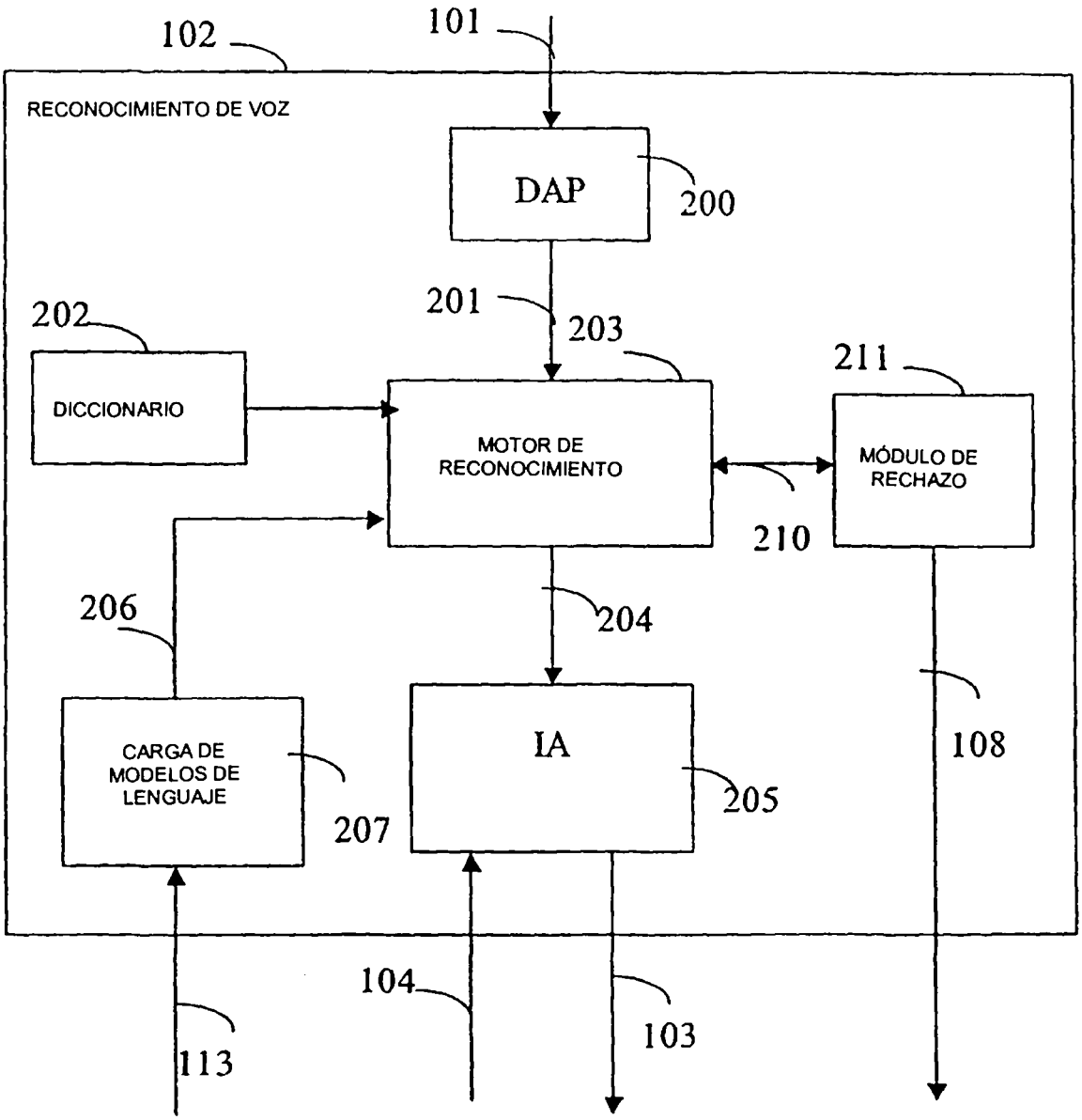
102
101
RECONOCIMIENTO DE VOZ
DAP
200
201
202
203
211
DICCIONARIO
MOTOR DE RECONOCIMIENTO
MÓDULO DE RECHAZO
210
206
204
CARGA DE MODELOS DE LENGUAJE
IA
108
207
205
104
103
113

Fig. 2

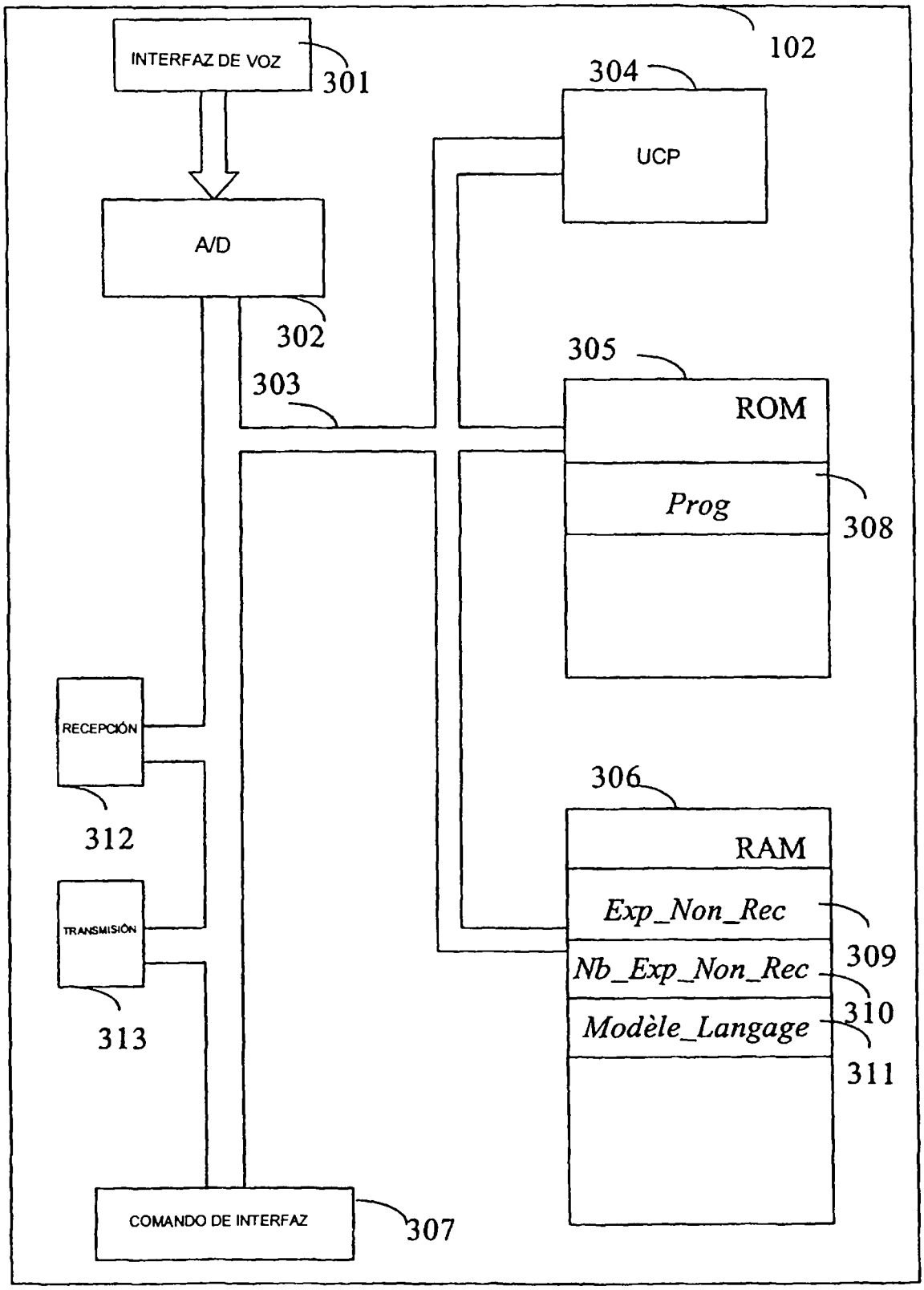
INTERFAZ DE VOZ
301
102
304
UCP
A/D
302
303
305
ROM
Prog
308
RECEPCIÓN
312
306
RAM
Exp_Non_Rec
309
Nb_Exp_Non_Rec
310
TRANSMISIÓN
313
Modèle_Langage
311
COMANDO DE INTERFAZ
307

Fig. 3

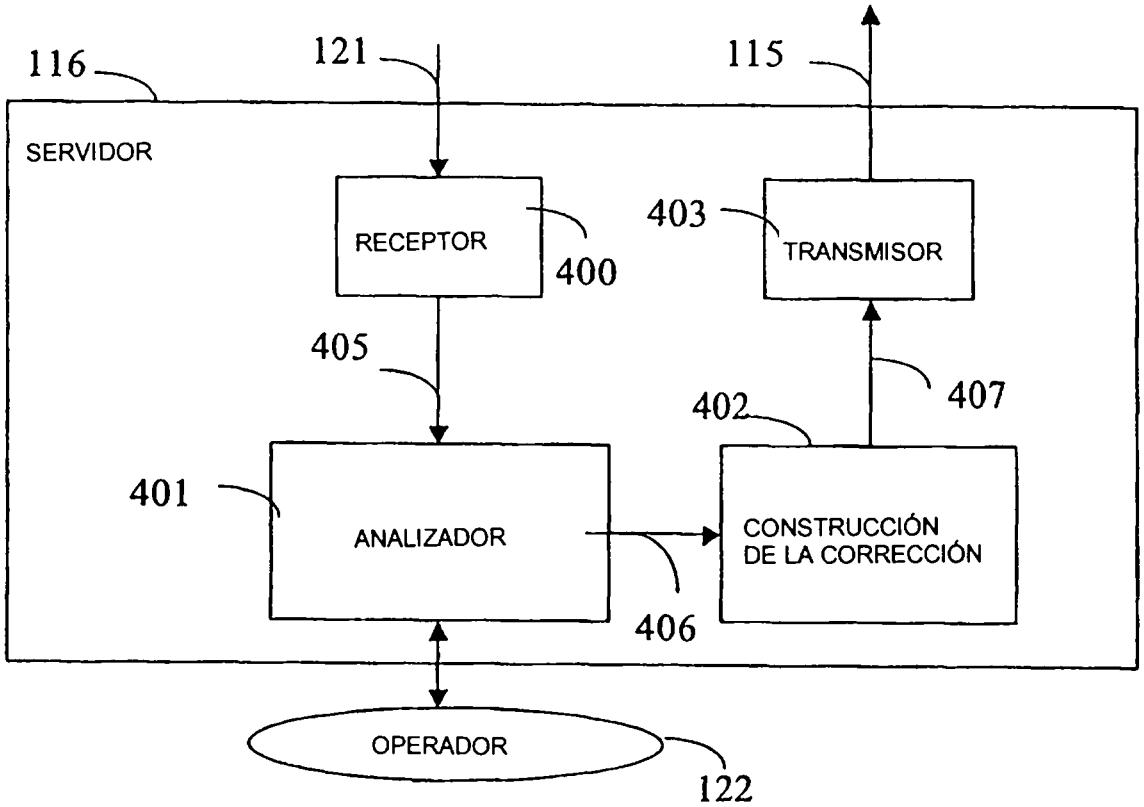
116
121
115
SERVIDOR
RECEPTOR
400
403
TRANSMISOR
405
407
402
401
ANALIZADOR
CONSTRUCCIÓN
DE LA CORRECCIÓN
406
OPERADOR
122

Fig. 4

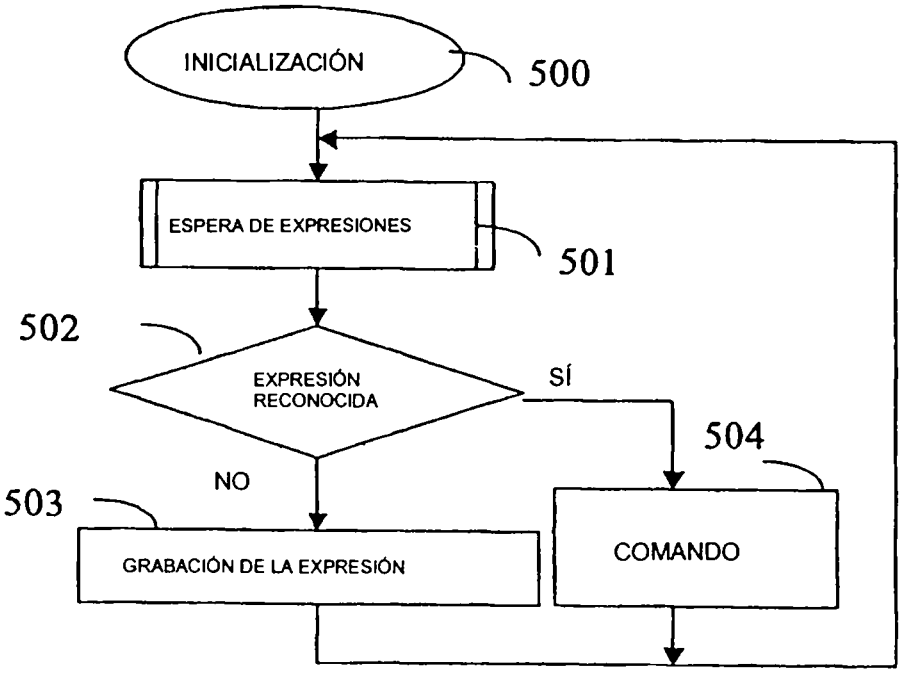
INICIALIZACIÓN
500
ESPERA DE EXPRESIONES
501
502
EXPRESIÓN
RECONOCIDA
SÍ
504
NO
503
GRABACIÓN DE LA EXPRESIÓN
COMANDO

Fig. 5

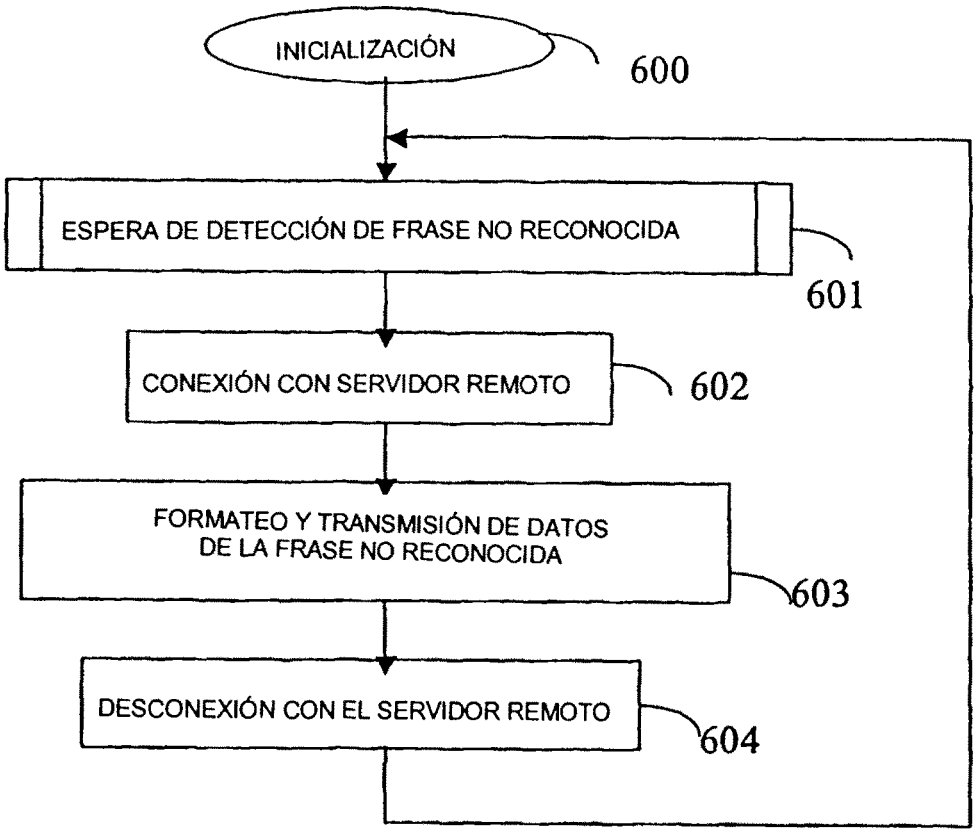
INICIALIZACIÓN
600
ESPERA DE DETECCIÓN DE FRASE NO RECONOCIDA
601
CONEXIÓN CON SERVIDOR REMOTO
602
FORMATEO Y TRANSMISIÓN DE DATOS DE LA FRASE NO RECONOCIDA
603
DESCONEXIÓN CON EL SERVIDOR REMOTO
604

Fig. 6

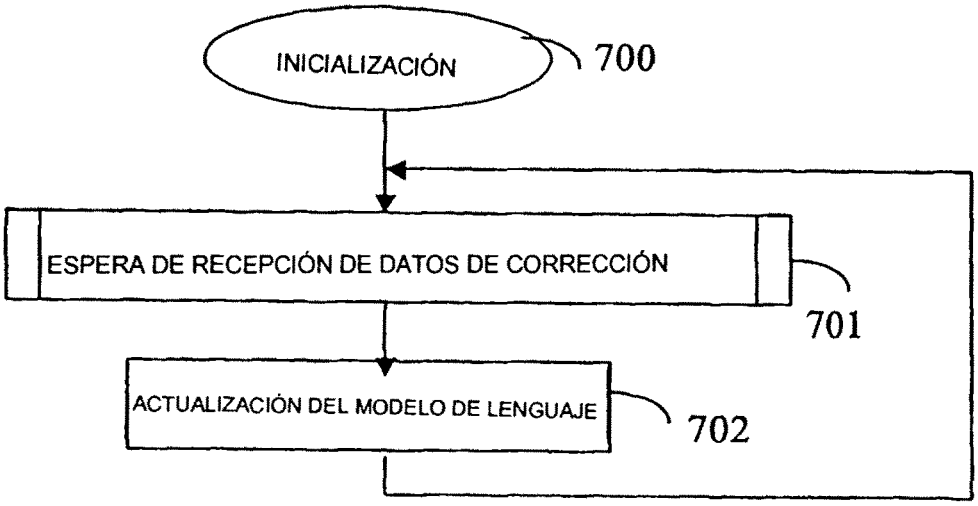
INICIALIZACIÓN
700
ESPERA DE RECEPCIÓN DE DATOS DE CORRECCIÓN
701
ACTUALIZACIÓN DEL MODELO DE LENGUAJE
702

Fig. 7

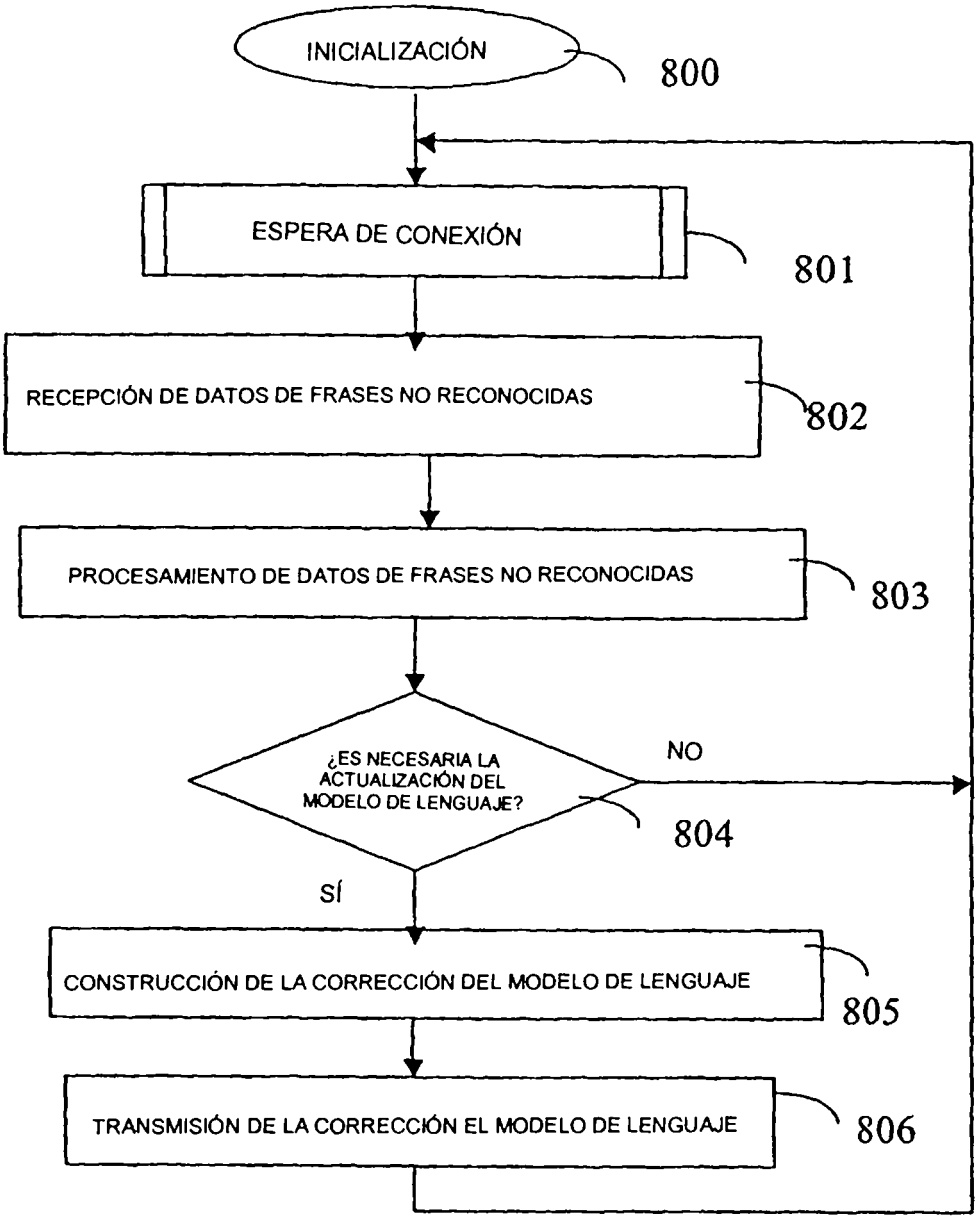
INICIALIZACIÓN
800
ESPERA DE CONEXIÓN
801
RECEPCIÓN DE DATOS DE FRASES NO RECONOCIDAS
802
PROCESAMIENTO DE DATOS DE FRASES NO RECONOCIDAS
803
¿ES NECESARIA LA ACTUALIZACIÓN DEL MODELO DE LENGUAJE?
NO
804
SÍ
CONSTRUCCIÓN DE LA CORRECCIÓN DEL MODELO DE LENGUAJE
805
TRANSMISIÓN DE LA CORRECCIÓN EL MODELO DE LENGUAJE
806

Fig. 8