



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 327 522**

51 Int. Cl.:
G10L 15/28 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04718327 .2**

96 Fecha de presentación : **08.03.2004**

97 Número de publicación de la solicitud: **1606796**

97 Fecha de publicación de la solicitud: **21.12.2005**

54 Título: **Procedimiento de reconocimiento de palabra distribuido.**

30 Prioridad: **25.03.2003 FR 03 03614**

45 Fecha de publicación de la mención BOPI:
30.10.2009

45 Fecha de la publicación del folleto de la patente:
30.10.2009

73 Titular/es: **FRANCE TELECOM**
6 place d'Alleray
75015 Paris, FR

72 Inventor/es: **Monne, Jean;**
Petit, Jean-Pierre y
Brisard, Patrick

74 Agente: **Justo Bailey, Mario de**

ES 2 327 522 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 de reconocimiento de palabra distribuido.

5 La presente invención se refiere al sector del mando vocal de aplicaciones, ejercido sobre terminales de usuario, gracias al establecimiento de medios de reconocimiento de palabra. Los terminales usuarios considerados son todos los dispositivos dotados de un medio de captura de la palabra normalmente un micrófono, que poseen capacidades de procesamiento de ese sonido y conectados a uno o a más servidores por un canal de transmisión. Se trata, por ejemplo, de aparatos de mando, de mando a distancia utilizados en aplicaciones domóticas, en automóviles (mando
10 del autorradio o de otras funciones del vehículo), en los PC o en estaciones telefónicas. El campo de las aplicaciones afectadas es esencialmente donde el usuario manda una acción, solicita una información o quiere interaccionar a distancia utilizando un mando vocal. La utilización de mandos vocales no excluye la existencia en el terminal de usuario de otros medios de acción (sistema multimodal), y el retorno de informaciones, de estados o de respuestas puede hacerse igualmente en forma combinada visual, sonora, olfativa y cualquier otro medio humano de percepción.

15 De manera general, los medios para la realización del reconocimiento de palabra comprenden medios de obtención de una señal de audio, medios de análisis acústico que extraen los parámetros de modelización y, por último, medios de reconocimiento que comparan estos parámetros de modelización extraídos de los modelos, y proponen la forma memorizada en los modelos que puede asociarse de la manera más probable a la señal. Opcionalmente, pueden utilizarse medios de detección de actividad vocal VAD (del inglés "Voice Activity Detection"). Aseguran la detección de las secuencias correspondientes a la palabra y antes de ser reconocidas. Extraen la señal de audio de entrada, fuera de los períodos de inactividad vocal, de los segmentos de la palabra, que se tratarán a continuación por los medios de extracción de los parámetros de modelización.

25 Más particularmente, la invención se apoya en las interacciones entre los tres modos de reconocimiento de palabra mencionados, embarcado, centralizado y distribuido.

En un modo de reconocimiento de palabra embarcado, el conjunto de los medios para efectuar el reconocimiento de palabra se encuentran a nivel de terminal de usuario. Las limitaciones de este modo de reconocimiento están, por
30 consiguiente, principalmente ligadas a la potencia de los procesadores embarcados, y a la memoria disponible para almacenar los modelos de reconocimiento de palabra. En cambio, este modo permite un funcionamiento autónomo, sin conexión a un servidor, y de este modo, está condenado a un gran desarrollo ligado a la reducción del coste de la capacidad de procesamiento.

35 En un modo de reconocimiento de palabra centralizado, todo el procedimiento de reconocimiento de palabra y los modelos de reconocimiento se encuentran y se ejecutan en una máquina, llamada generalmente servidor vocal, accesible para el terminal de usuario. El terminal transmite simplemente al servidor una señal de palabra. Este método se utiliza principalmente en las aplicaciones ofertadas por los operadores de telecomunicaciones. Un terminal básico puede así acceder a servicios evolucionados, activados por la voz. En un servidor de reconocimiento de palabra pueden establecerse numerosos tipos de reconocimiento de palabra (robusto, flexible, de muy amplio vocabulario, vocabulario
40 dinámico, palabra continua, mono o multilocutores, varias lenguas, etc.). En efecto, las máquinas centralizadas tienen capacidades de almacenamiento de modelos, tamaños de memoria de trabajo y potencias de cálculos importantes y crecientes.

45 En un modo de reconocimiento de palabra distribuido, los medios de análisis acústico están embarcados en el terminal de usuario, estando los medios de reconocimiento a nivel del servidor. En este modo distribuido, en la fuente se puede realizar, ventajosamente, una función de eliminación de efectos sonoros asociada a los medios de extracción de los parámetros de modelización. Sólo los parámetros de modelización se transmiten, lo que permite una ganancia sustancial en caudal de transmisión, particularmente interesante para las aplicaciones multimodales. Además, la señal a reconocer puede estar mejor protegida contra los errores de transmisión. Opcionalmente, se puede también embarcar la detección de actividad vocal (VAD) para transmitir los parámetros de modelización sólo durante las secuencias de palabras, lo que tiene como ventaja reducir en gran manera el tiempo de transmisión activa. El reconocimiento de palabra distribuido permite además transmitir en el mismo canal de transmisión de las señales de palabra y de datos, principalmente texto, imágenes o vídeos. La red de transmisión puede ser, por ejemplo, de tipo IP, GPRS,
50 WLAN o Ethernet. Este modo permite, igualmente, beneficiarse de procedimientos de protección y de corrección contra las pérdidas de paquetes que constituyen la señal transmitida con destino al servidor. Sin embargo, necesita la disponibilidad de canales de transmisión de datos, con un protocolo estricto de transmisión.

60 La invención propone un sistema de reconocimiento de palabra que incluye los terminales de usuarios y los servidores combinando las diferentes funciones ofertadas por los modos de reconocimiento de palabra embarcado, centralizado y distribuido para ofrecer la máxima eficacia, comodidad y ergonomía a los usuarios de servicios multimodales en los que se utiliza el mando vocal.

65 La Patente US 6 487 534-B1 describe un sistema de reconocimiento de palabra distribuido que incluye un terminal de usuario que dispone de medios de detección de actividad vocal, de medios de extracción de los parámetros de modelización y de medios de reconocimiento. Este sistema comprende además un servidor que dispone igualmente de medios de reconocimiento. El procedimiento descrito consiste en efectuar una primera fase de reconocimiento a nivel de terminal de usuario. Después, en función de los resultados de esta primera fase, los parámetros de modelización

calculados a nivel de terminal pueden, entonces, enviarse con destino al servidor, principalmente a fin de determinar, esta vez gracias a los medios de reconocimiento del servidor, una forma memorizada en los modelos de éste y asociada a la señal enviada.

5 El objetivo pretendido por el sistema descrito en el documento citado es disminuir la carga a nivel de servidor. Sin embargo, resulta que el reconocimiento se efectúa lo mejor posible después del tiempo necesario en la primera fase en el terminal. Cuando debe tener lugar una segunda fase, el tiempo de reconocimiento total es igual al tiempo de reconocimiento de la primera fase más el de la segunda fase.

10 El documento US 6 122 613 describe igualmente un sistema de reconocimiento de palabra en el que se hace un reconocimiento de palabra a partir de una señal de audio en serie o en paralelo a nivel de un servidor y de un terminal.

El documento US2003/00 4720 se refiere, también, a un sistema de reconocimiento de palabra dentro del que se busca aumentar la tolerancia a las faltas debidas a la transmisión de datos.

15 Un objetivo de la presente invención es proponer un procedimiento que pretenda optimizar el tiempo necesario en el reconocimiento de palabra y, por consiguiente, en la realización de la acción solicitada por el usuario.

20 Según un primer aspecto, la invención propone un procedimiento de reconocimiento de palabra distribuido, que incluye al menos un terminal de usuario y al menos un servidor, aptos para comunicar entre ellos por medio de una red de telecomunicaciones, según el cual a nivel de terminal de usuario, se realizan al menos las etapas siguientes:

- obtener una señal de audio a reconocer; y,
- 25 - calcular los parámetros de modelización de la señal de audio a reconocer; e,
- intentar asociar una forma memorizada a los parámetros de modelización; e,
- independientemente de la etapa de intento de asociar una forma memorizada:
- 30 - seleccionar una señal entre al menos la señal de audio a reconocer y una señal que indique los parámetros de modelización, en función de criterios relativos a la carga de los diferentes medios de procesamiento a nivel de terminal y de servidor, y/o en función de los criterios relacionados con la disponibilidad de los canales de transmisión de voz o de datos; y
- 35 - emitir, con destino al servidor, la señal seleccionada;

y, según el cual, a nivel del servidor, se realizan al menos las etapas siguientes:

- 40 - recibir la señal emitida por el terminal de usuario;
- intentar asociar una forma memorizada a la señal recibida.

45 Un procedimiento según la invención permite evitar acumular un tiempo de procesamiento a nivel de terminal y un tiempo de procesamiento a nivel de servidor, y así disminuir el tiempo para dar curso a la solicitud del usuario.

50 Permite también aprovecharse de las ventajas de las características de los medios de reconocimiento de los que disponen respectivamente el terminal y el servidor, principalmente cuando la señal a reconocer es de naturaleza indefinida para determinar lo antes posible la respuesta a efectuar.

En los modos de establecimiento preferidos, la señal emitida por el terminal de usuario con destino al servidor se selecciona entre al menos la señal de audio a reconocer y una señal que indica los parámetros de modelización. Entonces, si la señal recibida es de tipo de audio, el servidor calcula los parámetros de modelización de la señal de audio recibida; e intenta asociar una forma memorizada a los parámetros de modelización de la señal de audio recibida. Si la señal recibida indica los parámetros de modelización, el servidor intenta asociar una forma memorizada a dichos parámetros de modelización.

60 La elección de la señal transmitida - la señal de audio (comprimida o no) o la señal entregada por los medios de cálculo de los parámetros de modelización del terminal - puede definirse por el tipo de aplicaciones en curso, por el estado de la red, o a continuación de una coordinación entre los medios de control respectivos del terminal y del servidor.

65 Ventajosamente, la obtención por el terminal de la señal a reconocer incluye una detección de actividad vocal aplicada a una señal de audio de origen para producir la señal de audio a reconocer eliminando la señal de origen de períodos de inactividad vocal. La señal a emitir se seleccionará, por consiguiente, entre esta señal de audio eliminada de los períodos de inactividad vocal y la señal que indica los parámetros de modelización.

ES 2 327 522 T3

En un modo de establecimiento del procedimiento según la invención, la señal emitida se selecciona de una señal entre al menos la señal de audio de origen, la señal de audio representativa de la señal de origen eliminada de períodos de inactividad vocal después de la detección vocal y la señal que indica los parámetros de modelización.

5 Ventajosamente, si la señal recibida es una señal de audio eliminada de períodos de inactividad vocal, el servidor calcula parámetros de modelización de señal recibida e intenta asociar una forma memorizada a los parámetros de modelización de la señal de audio recibida. Cuando la señal recibida es una señal de tipo audio, pero sobre la cual no se ha realizado detección de actividad vocal, el servidor realiza una detección de actividad vocal aplicada a la señal de audio recibida para producir una señal de audio a reconocer eliminando la señal de origen de períodos de inactividad vocal. Después, calcula parámetros de modelización de la señal de audio a reconocer. Por último, intenta asociar una forma memorizada a los parámetros de modelización.

10 Ventajosamente, se escoge la forma memorizada asociada determinada a nivel de terminal, si existe. Se escoge la forma memorizada asociada determinada en primer lugar. O aún se escoge la forma memorizada mejor asociada según un criterio definido (por ejemplo, probabilidad de verosimilitud).

Según un segundo aspecto, la invención propone un terminal de usuario para establecer el procedimiento de reconocimiento de palabra distribuido descrito anteriormente.

20 Según un tercer aspecto, la invención propone un servidor para establecer el procedimiento de reconocimiento de palabra distribuido descrito anteriormente.

En un modo de realización preferido, al menos ciertos medios para realizar el procesamiento del reconocimiento a nivel de terminal (medios de extracción de parámetros o medios de reconocimiento) se han cargado a distancia por medio de la red de telecomunicaciones. Por ejemplo, pueden ser cargados a distancia por el servidor.

30 Otras características y ventajas de la invención surgirán aún de la lectura de la descripción que viene a continuación. Ésta es puramente ilustrativa y debe leerse con respecto a los dibujos adjuntos en los que la figura única es un esquema que representa un ejemplo de terminal de usuario y un ejemplo de servidor según la invención.

El sistema representado en la figura única incluye un servidor 1 y un terminal de usuario 2, que comunican entre sí por medio de una red (no representada) que dispone de canales para la transmisión de señales de voz y de canales para la transmisión de señales de datos.

35 El terminal 2 incluye un micrófono 4, que recoge la palabra a reconocer de un usuario en forma de una señal de audio. El terminal 2 incluye igualmente los módulos 5, 6 y 7. El módulo de detección de actividad vocal VAD 5 asegura la detección de las secuencias correspondientes a la palabra y debiendo ser reconocidas. Este módulo 5 está pensado, por ejemplo, para detectar rápidamente las palabras de mando. El módulo 6 efectúa un análisis acústico de manera conocida de por sí: calcula los parámetros de modelización, realizando una función de eliminación de efectos sonoros.

40 El módulo 7 ejecuta un algoritmo de reconocimiento de tipo conocido, por ejemplo a base de modelos de Markov ocultos con un vocabulario de tamaño reducido. Este motor 7 de reconocimiento puede funcionar en modo de un solo hablante, y necesita una fase de aprendizaje previa a partir de la voz del usuario.

45 El terminal comprende un controlador 8 adaptado para seleccionar una señal de audio entre la señal de audio de salida del micrófono 4, una señal representativa de los segmentos de palabra extraídos por los medios 5 de detección de actividad vocal y una señal indicativa de los parámetros de modelización 6.

50 El terminal comprende además una interfaz 9 para la emisión en la red, con destino al servidor, de la señal seleccionada por el controlador 8.

El servidor 1 incluye una interfaz 10 de red para recibir las señales que se le dirigen y un controlador 11 que analiza la señal recibida y la dirige a continuación, selectivamente, hacia un módulo de procesamiento del servidor entre varios módulos 12, 13 y 14. El módulo 12 es un detector de actividad vocal, que asegura la detección de los segmentos correspondientes a la palabra de manera similar al módulo 5. No obstante, puede ser diferente del módulo 5, y por ejemplo pensado para detectar rápidamente frases enteras. Su tiempo de reacción puede, por consiguiente, ser diferente al del módulo 5. En el presente ejemplo, su tiempo de reacción será más lento. El módulo 13 asegura el cálculo de parámetros de modelización de manera parecida al módulo 6 de cálculo del terminal. No obstante, el modelo de cálculo puede ser diferente. El módulo 14 ejecuta un algoritmo de reconocimiento de tipo conocido, por ejemplo a base de modelos de Markov ocultos con un vocabulario de tamaño arbitrario, por ejemplo superior a 100.000 palabras. Este motor 14 de reconocimiento compara los parámetros de entrada con modelos de palabra que representan palabras o frases y determina la mejor forma asociada, teniendo en cuenta modelos sintácticos que describen los encadenamientos de las palabras esperadas, los modelos relativos al léxico que precisan las diferentes pronunciaciones de las palabras y los modelos acústicos representativos de los sonidos pronunciados. Estos modelos son, por ejemplo, multilocutores, capaces de reconocer, con una buena fiabilidad, la palabra, independientemente del que habla.

ES 2 327 522 T3

El controlador 11 actúa sobre el módulo de VAD 12, el módulo de cálculo de los parámetros 13 y el motor 14 de reconocimiento de manera que:

- 5 a) cuando la señal recibida por la interfaz 10 de red es de tipo audio y no indica segmentos de palabra obtenidos después de la detección de actividad vocal, activar el módulo 12 de detección de actividad vocal dirigiéndole la señal recibida como señal de entrada, dirigir después los segmentos de palabra extraídos por el módulo 12 al módulo de cálculo de los parámetros 13 de modelización como parámetros de entrada, y después dirigir los parámetros extraídos por este módulo 13 al motor 14 de reconocimiento como parámetros de entrada;
- 10 b) cuando la señal recibida por la interfaz 10 de recepción indica segmentos de palabra después de la detección de actividad vocal, activar el módulo de cálculo de parámetros 13 de modelización dirigiéndole la señal recibida como señal de entrada, y después dirigir los parámetros extraídos por este módulo 13 al motor 14 de reconocimiento como parámetros de entrada;
- 15 c) cuando la señal recibida por la interfaz 10 de recepción indica parámetros de modelización, dirigir dichos parámetros indicados al motor 14 de reconocimiento como parámetros de entrada.

Consideremos una aplicación en la que el usuario expresa: “llamada a Antonio”, figurando Antonio en el repertorio local. La señal de audio correspondiente obtenida por el micrófono 4 del terminal es tratada por el módulo VAD 5, que extrae de ella los segmentos de palabra los cuales se dirigen a su vuelta a los módulos 6, que calcula los parámetros de modelización. Estos parámetros son a continuación dirigidos al motor 7 de reconocimiento para asociarle una forma que provenga del diccionario local. En paralelo, el controlador 8 ha seleccionado una señal a emitir entre la señal de audio de origen, una señal de audio que indica los segmentos de palabra extraídos de la señal de audio de origen después de la detección de actividad vocal y una señal que indica los parámetros de modelización. La señal seleccionada se emite con la ayuda de la interfaz 9 de emisión con destino al servidor.

En el modo de realización considerado, la señal seleccionada por el controlador del terminal 8 es la señal de audio de origen, que ha sido enviada en dirección al servidor tan pronto como ha sido adquirida por el micrófono 4.

- 30 El servidor recibe la señal a reconocer enviada por el terminal y la trata como se ha indicado en el apartado a) anterior.

El proceso de reconocimiento se efectúa así por ambas partes. El terminal determina en un tiempo T1 una forma asociada F1; el servidor determina otra forma asociada F2 en un tiempo T2 diferente de T1. Una de las formas es aceptada según un criterio de elección. El criterio de elección puede ser, por ejemplo, el siguiente: se escoge la forma encontrada más rápidamente y desde que se encuentra, sin esperar la determinación de otra forma. El aplicativo local al terminal pasa entonces a la fase aplicativa siguiente.

Después, el usuario expresa “buscar el mensaje de Josiane”.

La respuesta a nivel de terminal no es consistente y se salda con un rechazo después de un tiempo T1'. El reconocimiento se efectúa en paralelo a nivel de servidor y acaba en un tiempo T2' con la determinación de una forma asociada, que va a permitir interpretar el mensaje solicitado por el usuario.

Un procedimiento de reconocimiento según la invención permite así combinar las ventajas de los sistemas de reconocimiento a nivel de terminal y de servidor. Las palabras cortas son rápidamente determinadas por el motor 7 de reconocimiento del terminal y las frases más complejas son reconocidas rápidamente por el motor 14 de reconocimiento del servidor. El tiempo de procesamiento se optimiza más aún cuando los VAD respectivos presentan características adaptadas, siendo el VAD 5 del terminal, por ejemplo, pensado para detectar, rápidamente, palabras de mando y siendo el VAD 12 del servidor pensado para detectar, rápidamente, frases.

El controlador del terminal 8 determina la señal a transmitir con destino al servidor, por ejemplo en función de criterios de control. Éstos criterios pueden, por ejemplo, estar ligados a la aplicación afectada, o a problemáticas de cargas de los diferentes medios de procesamiento a nivel de terminal y de servidor (los medios de control respectivos pueden cooperar), o aún a problemáticas de disponibilidad de los canales de transmisión de voz o de los canales de transmisión de datos.

Por ejemplo, para ciertos terminales, la señal enviada será sistemáticamente una señal que indique parámetros de modelización. Para otros terminales, la señal enviada dependerá de la aplicación en curso.

En un modo de realización de la invención, en caso de problemas de disponibilidad de canales de datos o del módulo 6 de cálculo del terminal considerado, el controlador 8 se adapta para transmitir la señal de audio (de origen o después de VAD). Esta señal de audio podrá ser transmitida por los canales de transmisión de señales de voz disponibles.

La manera de determinar la forma finalmente aceptada y que será utilizada en el momento de proseguir la aplicación, entre una forma asociada suministrada por el módulo de reconocimiento del servidor y una forma asociada

ES 2 327 522 T3

suministrada por el terminal, puede efectuarse sobre la base de diferentes criterios, que pueden variar de un terminal a otro, pero también de una aplicación a otra o de un contexto determinado a otro.

Estos criterios pueden dar, por ejemplo, prioridad al reconocimiento efectuado a nivel de terminal, o a la forma asociada que presente la mayor tasa de probabilidad, o aún a la forma determinada más rápidamente. Estos criterios de elección pueden estar integrados, por ejemplo, en los controladores 8, 11 del terminal o del servidor.

El servidor según la invención es igualmente apto para efectuar el reconocimiento de palabra sobre una señal transmitida por un terminal que no dispone de medios de extracción de parámetros de modelización, ni de medios de reconocimiento (o en los cuales los medios de extracción o de reconocimiento están inactivos) y que disponen opcionalmente de VAD.

En un modo de realización, el motor 7 de reconocimiento del terminal 2 es un programa ejecutable cargado a distancia, por ejemplo desde el servidor, por medios clásicos de transferencia de datos.

Ventajosamente, para una aplicación determinada del terminal 2, los modelos de reconocimiento del terminal pueden ser cargados a distancia o puestos al día en el transcurso de una sesión aplicativa conectada a la red.

Otros recursos de programas y de sistemas de programación útiles en el reconocimiento de palabra pueden ser, también, cargados a distancia desde el servidor 1, como el módulo 6 de cálculo de parámetros de modelización o el detector de actividad vocal 5.

Podrían describirse otros ejemplos estableciendo, por ejemplo, aplicaciones relacionadas con vehículos, con electrodomésticos y multimedia.

Como lo presentado en los ejemplos de realización anteriormente descritos, un sistema según la invención permite utilizar en un tiempo de procesamiento optimizado para un resultado optimizado los diferentes recursos necesarios para el procesamiento del reconocimiento de palabra y presentes a nivel de terminal (por ejemplo para la carga a distancia) y de servidor.

REIVINDICACIONES

1. Un procedimiento de reconocimiento de palabra distribuido, que incluye al menos un terminal de usuario y al menos un servidor aptos para comunicarse entre ellos por medio de una red de telecomunicaciones, según el cual a nivel de terminal de usuario, se realizan al menos las etapas siguientes:

- obtener una señal de audio a reconocer;
- calcular los parámetros de modelización de la señal de audio a reconocer; e
- intentar asociar una forma memorizada a los parámetros de modelización;

y, según el cual, a nivel de servidor se realizan al menos las etapas siguientes:

- recibir una señal emitida por el terminal de usuario;
- intentar asociar una forma memorizada a la señal recibida;

caracterizado por las etapas siguientes, realizadas a nivel de terminal de usuario:

- independientemente de la etapa de intento de asociar una forma memorizada:
 - seleccionar una señal entre al menos la señal de audio a reconocer y una señal que indique los parámetros de modelización, en función de criterios relativos a la carga de los diferentes medios de procesamiento a nivel de terminal y de servidor, y/o en función de criterios relativos a la disponibilidad de canales de transmisión de voz o de datos; y
 - emitir la señal seleccionada con destino al servidor.

2. Un procedimiento de reconocimiento de palabra distribuido según la reivindicación 1, según el cual la señal emitida por el terminal de usuario con destino al servidor se selecciona entre al menos la señal de audio a reconocer y una señal que indique los parámetros de modelización;

según el cual si la señal recibida es de tipo audio, el servidor calcula los parámetros de modelización de la señal de audio recibida, e intenta asociar una forma memorizada a los parámetros de modelización de la señal de audio recibida;

y según el cual si la señal recibida indica los parámetros de modelización, el servidor intenta asociar una forma memorizada a dichos parámetros de modelización.

3. Un procedimiento según la reivindicación 1 o la reivindicación 2, según el cual la obtención en el terminal de la señal a reconocer incluye una detección de actividad vocal para producir la señal de audio a reconocer en forma de segmentos de palabra extraídos de una señal de audio de origen fuera de períodos de inactividad vocal.

4. Un procedimiento según la reivindicación 3, según el cual la señal emitida es una señal seleccionada entre al menos la señal de audio de origen, la señal de audio a reconocer en forma de segmentos extraídos después de la detección vocal y la señal que indica los parámetros de modelización.

5. Un procedimiento según las reivindicaciones 2 a 4, según el cual cuando la señal recibida es de tipo audio:

- si la señal de audio recibida está en forma de segmentos de palabra extraídos después de detección vocal, el servidor calcula los parámetros de modelización de la señal recibida e intenta asociar una forma memorizada a los parámetros de modelización de la señal de audio recibida;
- si no lo está, el servidor realiza una detección de actividad vocal aplicada a la señal de audio recibida para producir una señal de audio en forma de segmentos de palabras extraídos de la señal de audio recibida fuera de períodos de inactividad vocal, después calcula los parámetros de modelización de la señal de audio e intenta asociar una forma memorizada a los parámetros de modelización.

6. Un procedimiento según una cualquiera de las reivindicaciones precedentes, según el cual se escoge la forma memorizada asociada determinada al nivel de terminal cuando ésta existe.

7. Un procedimiento según las reivindicaciones 1 a 5, según el cual se escoge la forma memorizada asociada determinada lo más rápidamente.

8. Un procedimiento según las reivindicaciones 1 a 6, según el cual se escoge la forma memorizada mejor asociada según un criterio de elección definido.

9. Un terminal de usuario para establecer un procedimiento de reconocimiento de palabra distribuido, según una de las reivindicaciones 1 a 8, que incluye:

- medios de obtención de una señal de audio a reconocer;
- medios de cálculo de parámetros de modelización de la señal de audio; y
- medios de reconocimiento para asociar al menos una forma memorizada a los parámetros de modelización calculados por los medios de cálculo;

caracterizado porque comprende:

- medios de control para seleccionar una señal a emitir con destino al servidor entre la señal de audio a reconocer y una señal que indique los parámetros de modelización calculados en función de criterios relativos a la carga de los diferentes medios de procesamiento a nivel de terminal y de servidor, y/o en función de criterios relativos a la disponibilidad de canales de transmisión de voz o de datos.

10. Un terminal de usuario según la reivindicación 9, en el que los medios de obtención de la señal de audio a reconocer comprenden medios de detección de actividad vocal para producir la señal a reconocer en forma de segmentos de palabra extraídos de una señal de audio de origen, fuera de períodos de inactividad vocal.

11. Un terminal de usuario según la reivindicación 10, en el que los medios de control se adaptan para seleccionar al menos una señal a emitir con destino al servidor entre la señal de audio de origen, la señal de audio a reconocer en forma de segmentos de palabra extraídos por los medios de detección de actividad vocal y la señal que indica los parámetros de modelización calculados.

12. Un terminal de usuario según las reivindicaciones 9 a 11, en el que al menos una parte de los medios de cálculo de parámetros y de los medios de reconocimiento se cargan a distancia desde el servidor.

13. Un terminal de usuario según las reivindicaciones 9 a 12, que incluye medios de determinación de la forma memorizada a escoger entre las formas memorizadas determinadas respectivamente en terminal y en el servidor.

14. Un servidor adaptado para establecer un procedimiento de reconocimiento de palabra distribuido según una de las reivindicaciones 1 a 8 que comprende:

- medios de recepción procedentes de un terminal de usuario de una señal seleccionada en dicho terminal; y
- medios de reconocimiento para asociar al menos una forma memorizada a los parámetros de modelización de entrada;
- medios de cálculo de parámetros de modelización de una señal de entrada;

caracterizado porque comprende:

- medios de control para mandar los medios de cálculo y los medios de reconocimiento de manera que:
 - cuando la señal recibida por los medios de recepción sea de tipo audio, se activen los medios de cálculo de parámetros dirigiéndoles la señal seleccionada como señal de entrada, y se dirijan los parámetros calculados por los medios de cálculo a los medios de reconocimiento como parámetros de entrada, y,
 - cuando la señal seleccionada recibida por los medios de recepción indican parámetros de modelización, se dirijan dichos parámetros indicados a los medios de reconocimiento como parámetros de entrada.

15. Un servidor según la reivindicación 14, que comprende además medios de detección de actividad para producir la señal a reconocer en forma de segmentos de palabra extraídos de una señal de audio de origen fuera de períodos de inactividad vocal y en el que los medios de control se adaptan para mandar los medios de cálculo de parámetros y los medios de reconocimiento cuando la señal recibida es de tipo audio de manera que:

- si la señal recibida de tipo audio está en forma de segmentos de palabra después de la detección de actividad vocal, se activen los medios de cálculo de parámetros dirigiéndoles la señal recibida, como señal de entrada, después se dirijan los parámetros calculados por los medios de cálculo de parámetros a los medios de reconocimiento como parámetros de entrada;
- si no lo está, se activen los medios de detección de actividad vocal del servidor dirigiéndoles la señal recibida como señal de entrada, después se dirijan los segmentos extraídos por los medios de detección de actividad vocal a los medios de cálculo de parámetros como parámetros de entrada, después se dirijan los parámetros calculados por los medios de cálculo de parámetros a los medios de reconocimiento como parámetros de entrada.

ES 2 327 522 T3

16. Un servidor según una de las reivindicaciones 14 a 15, que comprende medios para cargar a distancia recursos de programas y de sistemas de información de reconocimiento vocal por medio de red de telecomunicaciones con destino a un terminal.

5 17. Un servidor según la reivindicación 16, en el que dichos recursos comprenden al menos un módulo entre: un módulo de VAD, un módulo de cálculo de parámetros de modelización de una señal de audio y un módulo de reconocimiento para asociar al menos una forma memorizada a los parámetros de modelización.

10 18. Un servidor según las reivindicaciones 14 a 17, que incluye medios de determinación de la forma memorizada a escoger entre las formas memorizadas determinadas respectivamente en el terminal y en el servidor.

15

20

25

30

35

40

45

50

55

60

65

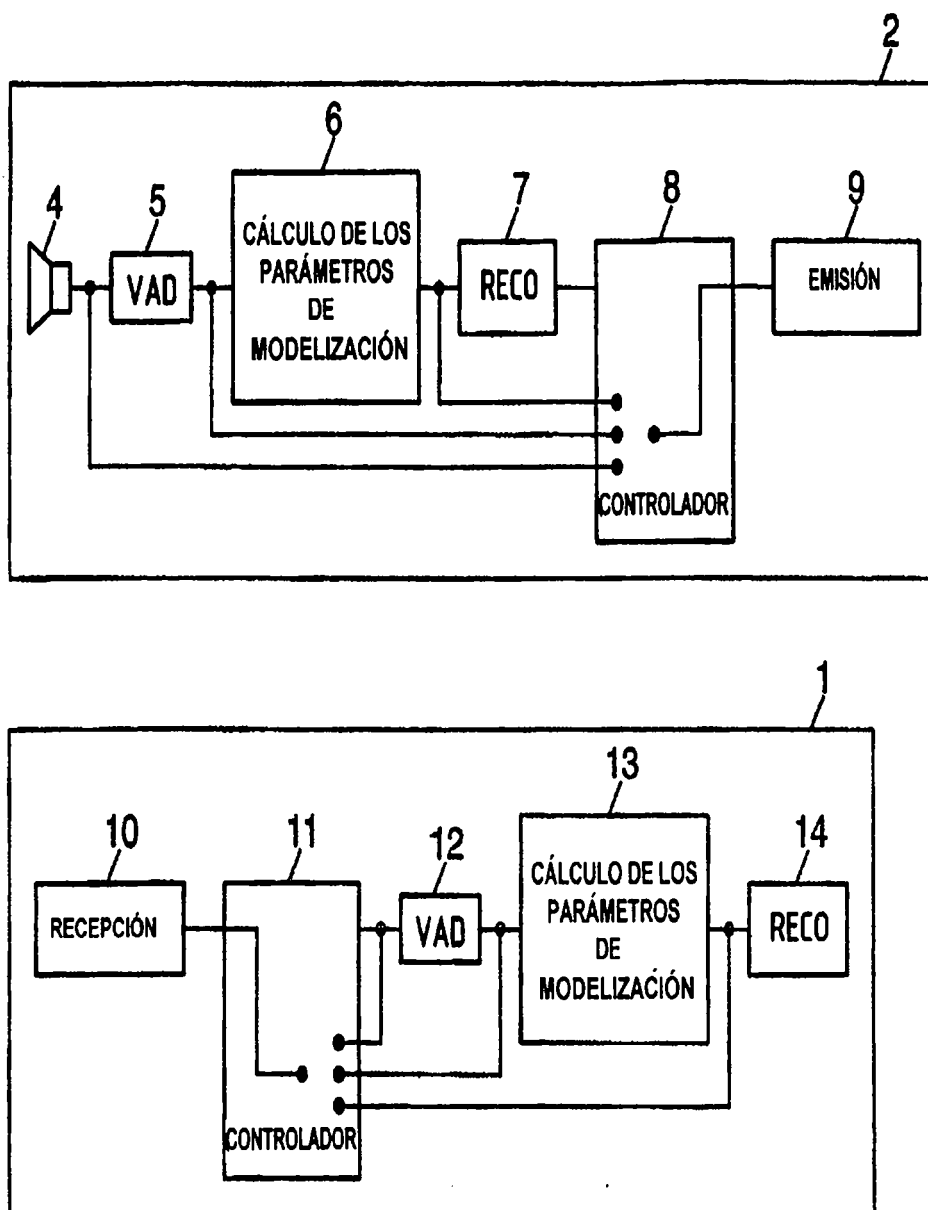


FIG. Única