(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 331 698**

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

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **04718324 .9**
(96) Fecha de presentación : **08.03.2004**
(97) Número de publicación de la solicitud: **1606795**
(97) Fecha de publicación de la solicitud: **21.12.2005**

(54) Título: **Sistema de reconocimiento de voz distribuido.**

(30) Prioridad: **25.03.2003 FR 03 03615**

(45) Fecha de publicación de la mención BOPI:
13.01.2010

(45) Fecha de la publicación del folleto de la patente:
13.01.2010

(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**

DESCRIPCIÓN

Sistema de reconocimiento de voz distribuido.

La presente invención se refiere al campo del control vocal de aplicaciones, ejercido sobre terminales de usuario, gracias al empleo de medios de reconocimiento de la voz. Los terminales de usuario considerados son todos los dispositivos dotados de un medio de captura de la voz, habitualmente un micrófono, que posee capacidades de tratamiento de este sonido y conectados a uno o más servidores mediante un canal de transmisión. Se trata, por ejemplo, de aparatos de control, de control a distancia utilizados en aplicaciones domésticas, en automóviles (control de auto-radio o de otras funciones del vehículo), en PC o en terminales de teléfono. El campo de las aplicaciones concernidas es esencialmente aquel en el que el usuario ordena una acción, solicita una información o quiere interactuar a distancia utilizando una orden de voz. La utilización de órdenes de voz no excluye la existencia en el terminal de usuario de otros medios de acción (sistema multi-modal), y el retorno de informaciones, de estados o de respuestas también se puede realizar de forma combinada visual, sonora, olfativa y mediante cualquier otro medio perceptible por el ser humano.

De manera general, los medios para la realización del reconocimiento de voz comprenden medios de obtención de una señal de audio, medios de análisis acústico que extraen parámetros de modelización y finalmente medios de reconocimiento que comparan estos parámetros de modelización calculados con modelos y proponen la forma memorizada en los modelos que puede estar asociada a la señal de la forma más probable. Opcionalmente se pueden utilizar medios de detección de actividad vocal VAD ("Voice Activation Detection"). Estos aseguran la detección de secuencias correspondientes a la voz y que deben ser reconocidas. Estos extraen de la señal de audio entrante, fuera de periodos de inactividad vocal, segmentos de voz que a continuación serán tratados mediante los medios de cálculo de los parámetros de modelización.

Más particularmente, la invención se refiere a las interacciones entre los tres modos de reconocimiento de voz llamados embarcado, centralizado y distribuido.

En un modo de reconocimiento de voz embarcado, el conjunto de los medios para realizar el reconocimiento de voz se encuentra al nivel del terminal de usuario. Las limitaciones de este modo de reconocimiento están, por lo tanto, vinculadas a la potencia de los procesadores embarcados y a la memoria disponible para almacenar los modelos de reconocimiento de voz. Como contrapartida, este modo permite un funcionamiento autónomo, sin conexión a un servidor y como tal es susceptible a un fuerte desarrollo vinculado a la reducción del coste de la capacidad de tratamiento.

En un modo de reconocimiento de voz centralizado, todo el procedimiento de voz y los modelos de reconocimiento se encuentran y se ejecutan en una máquina, llamada generalmente servidor vocal, accesible a través del terminal de usuario. El terminal transmite simplemente al servidor una señal de voz. Este método se utiliza particularmente en las aplicaciones ofrecidas por los operadores de telecomunicaciones. De este modo, un terminal básico puede acceder a terminales evolucionados, activados mediante la voz. Muchos tipos de reconocimiento de voz (robusto, flexible, vocabulario muy amplio, vocabulario dinámico, voz continua, mono- o multi-locutor, varios idiomas, etc.) se pueden implementar en un servidor de reconocimiento de voz. En efecto, las máquinas centralizadas tienen capacidades de almacenamiento de modelos, tamaños de memoria de trabajo y potencias de cálculo grandes y crecientes.

En un modo de reconocimiento de voz 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, una función de eliminación de ruido asociada a los medios de cálculo de los parámetros de modelización se puede realizar ventajosamente en la fuente. Solamente se transmiten los parámetros de modelización, lo que permite un aumento sustancial del caudal de transmisión, particularmente interesante para las aplicaciones multimodales. Además, la señal a reconocer puede estar mejor protegida contra los errores de transmisión. Opcionalmente, también se puede embarcar la detección de actividad vocal (VAD) para transmitir los parámetros de modelización solamente durante las secuencias de voz, lo que tiene la ventaja de reducir de manera importante el periodo de transmisión activa. El reconocimiento de voz distribuido permite además transmitir por el mismo canal de transmisión señales de voz y de datos, particularmente texto, imágenes o videos. La red de transmisión puede ser por ejemplo de tipo IP, GPRS, WLAN o Ethernet. Este modo también permite beneficiarse de procedimientos de protección y de corrección contra las pérdidas de paquetes que constituyen la señal transmitida con destino al servidor. Sin embargo, requiere la disponibilidad de canales de transmisión de datos, con un protocolo estricto de transmisión.

La invención propone un sistema de reconocimiento de voz que comprende terminales de usuario y servidores que combinan las diferentes funciones ofrecidas por los medios de reconocimiento de voz embarcado, centralizado y distribuido, para ofrecer la máxima eficacia, comodidad y ergonomía a los usuarios de servicios multi-modales en los que se utiliza el control vocal.

La patente US 6487534 B1 describe un sistema de reconocimiento de voz distribuido que comprende un terminal de usuario que dispone de medios de detección de actividad vocal, medios de cálculo de los parámetros de modelización y medios de reconocimiento. Este sistema comprende además un servidor que también dispone de medios de reconocimiento. El principio descrito es la realización de al menos una primera fase de reconocimiento a nivel del terminal de usuario. En una segunda fase opcional, los parámetros de modelización calculados a nivel del terminal se

envían con destino al servidor, para determinar particularmente, esta vez gracias a los medios de reconocimiento del servidor, una forma memorizada en los modelos de éste y asociada a la señal enviada.

El objeto pretendido por el sistema descrito en el documento mencionado es reducir la carga a nivel del servidor. Sin embargo, de esto se deriva que el terminal debe realizar el cálculo de los parámetros de modelización de forma local antes de transmitirlos eventualmente con destino al servidor. Ahora bien, existen circunstancias en las que, por razones de gestión de carga o por razones de aplicación, es preferible realizar este cálculo a nivel del servidor.

De esto se deriva también que los canales utilizados para la transmisión de los parámetros de modelización a reconocer, en un sistema de acuerdo con el documento mencionado, deben ser imperativamente canales adecuados para transmitir este tipo de datos. Ahora bien, dichos canales de protocolo muy estricto no están disponibles forzosamente de forma permanente en la red de transmisión. Es por ello que es interesante poder utilizar canales clásicos de transmisión de señales de audio, para no retardar o bloquear el proceso de reconocimiento iniciado a nivel del terminal.

Un objeto de la presente invención, tal como se define mediante las reivindicaciones 1, 7 y 11, es proponer un sistema distribuido que resulte menos afectado por las limitaciones mencionadas anteriormente.

De este modo, según un primer aspecto, la invención propone un sistema de reconocimiento de voz distribuido, que comprende al menos un terminal de usuario y al menos un servidor adecuados para comunicarse entre sí por medio de una red de telecomunicaciones, en el que el terminal de usuario comprende:

- medios de obtención de una señal de audio a reconocer,

- primeros medios de cálculo de parámetros de modelización de la señal de audio, y

- primeros medios de control para seleccionar al menos una señal a emitir con destino al servidor entre la señal de audio a reconocer y una señal que indica los parámetros de modelización calculados;

y en el que el servidor comprende:

- medios de recepción de la señal seleccionada procedente del terminal de usuario,

- segundos medios de cálculo de parámetros de modelización de una señal de entrada,

medios de reconocimiento para asociar al menos una forma memorizada a parámetros de entrada, y

- segundos medios de control para controlar los segundos medios de cálculo y los medios de reconocimiento para:

  - cuando la señal seleccionada recibida por los medios de recepción es de tipo audio, activar los segundos medios de cálculo de parámetros remitiéndoles la señal seleccionada como señal de entrada y remitir los parámetros calculados por los segundos medios de cálculo a los medios de reconocimiento como parámetros de entrada, y

  - cuando la señal seleccionada recibida por los medios de recepción indica parámetros de modelización, remitir dichos parámetros indicados a los medios de reconocimiento como parámetros de entrada.

De este modo, el sistema de acuerdo con la invención permite transmitir desde la terminal de usuario con destino al servidor ya sea la señal de audio (comprimida o no) ya sea la señal emitida por los medios de cálculo de los parámetros de modelización del terminal. La elección de la señal transmitida se puede definir o bien por el tipo de aplicación en curso, o bien por el estado de la red, o bien según una coordinación entre los medios de control respectivos del terminal y del servidor.

Un sistema de acuerdo con la invención dota de capacidad al terminal de usuario para realizar, en función por ejemplo de parámetros de entrada de los que disponen los medios de control en un momento dado, el cálculo de los parámetros de modelización a nivel del terminal o a nivel del servidor. Este cálculo también se puede realizar en paralelo a nivel del terminal y a nivel del servidor.

Un sistema de acuerdo con la invención permite realizar el reconocimiento de voz desde terminales de diferentes tipos que coexisten en una misma red, por ejemplo:

- terminales que no disponen de ningún medio de reconocimiento local (o cuyo medio de reconocimiento local está inactivo), en cuyo caso la señal de audio es enviada para su reconocimiento con destino al servidor,

- terminales que disponen de medios de detección de actividad vocal sin medios de cálculo de parámetros de modelización, ni medios de reconocimiento (o cuyos medios de cálculo de parámetros y los medios de reconocimiento están inactivos) y que transmiten al servidor para su reconocimiento una señal de audio de origen o una señal de audio representativa de segmentos de voz extraídos de la señal de audio fuera de periodos de inactividad vocal, y

- servidores que disponen por ejemplo únicamente de medios de reconocimiento, sin medios de cálculo de parámetros de modelización.

Ventajosamente, los medios de obtención de la señal de audio del terminal de usuario pueden comprender además medios de detección de actividad vocal para extraer de la señal de audio de origen, fuera de periodos de inactividad vocal, segmentos de voz. Los medios de control del terminal seleccionan entonces al menos una señal a emitir con destino al servidor, entre una señal de audio representativa de los segmentos de voz y la señal que indica los parámetros de modelización calculados.

Ventajosamente, los medios de control del terminal son adecuados para seleccionar al menos una señal a emitir con destino al servidor entre al menos la señal de audio de origen, la señal de audio que indica los segmentos de voz extraídos de la señal de audio de origen y la señal que indica parámetros de modelización calculados. A nivel del servidor, los medios de control son adecuados para controlar los medios de cálculo y los medios de reconocimiento para, cuando la señal seleccionada recibida por los medios de recepción es representativa de los segmentos de voz extraídos por los medios de detección de actividad vocal del terminal, activar los medios de cálculo de parámetros del servidor remitiéndoles la señal seleccionada como señal de entrada y remitir los parámetros calculados por estos medios de cálculo a los medios de reconocimiento como parámetros de entrada.

En una realización preferida, el servidor comprende además medios de detección de actividad vocal para extraer de una señal recibida de tipo audio, fuera de los periodos de inactividad vocal, segmentos de voz. En este caso, a nivel del servidor, los medios de control son adecuados para controlar los medios de cálculo y los medios de reconocimiento para:

- cuando la señal seleccionada recibida por los medios de recepción es de tipo audio:

  - si la señal recibida de tipo audio es representativa de segmentos de voz después de la detección de actividad vocal, activar los segundos medios de cálculo de parámetros remitiéndoles la señal seleccionada como señal de entrada y después remitir los parámetros calculados por los segundos medios de cálculo de parámetros a los medios de reconocimiento como parámetros de entrada;

  - de lo contrario, activar los medios de detección de actividad vocal del servidor remitiéndoles la señal seleccionada como señal de entrada y después remitir los segmentos extraídos por los medios de detección de actividad vocal a los segundos medios de cálculo de parámetros como parámetros de entrada y después remitir los parámetros calculados por los segundos medios de cálculo de parámetros a los medios de reconocimiento como parámetros de entrada;

- cuando la señal seleccionada recibida por los medios de recepción indica parámetros de modelización, remitir dichos parámetros indicados a los medios de reconocimiento como parámetros de entrada.

Ventajosamente, el terminal de usuario comprende además medios de reconocimiento para asociar al menos una forma memorizada a parámetros de entrada.

En este último caso, los medios de control del terminal pueden ser adecuados para seleccionar una señal a emitir con destino al servidor en función del resultado proporcionado por los medios de reconocimiento del terminal. Y el terminal de usuario puede comprender además medios de almacenamiento adecuados para almacenar una señal a nivel del terminal, para poder, en el caso en que el resultado del reconocimiento local en el terminal no sea satisfactorio, enviar la señal para el reconocimiento por parte del servidor.

Ventajosamente, los medios de control del terminal pueden ser adecuados para seleccionar una señal a emitir con destino al servidor independientemente del resultado proporcionado por los primeros medios de reconocimiento.

Es preciso observar que los medios de control de un terminal pueden pasar de uno a otro de los dos modos descritos en los dos párrafos anteriores, en función por ejemplo del contexto de la aplicación o del estado de la red.

Preferiblemente, los medios de control del servidor cooperan con los medios de control del terminal. De este modo, el terminal puede evitar enviar con destino al servidor por ejemplo una señal de audio si ya hay una carga importante a nivel de los medios de cálculo de parámetros del servidor. En una posible realización, los medios de control del servidor están configurados para cooperar con los medios del terminal para adaptar el tipo de señales enviadas por el terminal en función de las capacidades respectivas de la red, del servidor y del terminal.

Los medios de cálculo y de reconocimiento del terminal pueden ser normalizados o privados.

En una realización preferida, al menos algunos de entre los medios de reconocimiento y de cálculo de parámetros a nivel del terminal, se le proporcionan mediante descarga, en forma de código ejecutable por el procesador del terminal, por ejemplo desde el servidor.

De acuerdo con un segundo aspecto, la invención propone un terminal de usuario para implementar un sistema de reconocimiento de voz distribuido de acuerdo con la invención.

De acuerdo con un tercer aspecto, la invención propone un servidor para implementar un sistema de reconocimiento de voz distribuido de acuerdo con la invención.

Otras características y ventajas de la invención surgirán también con la lectura de la siguiente descripción. Ésta es puramente ilustrativa y se debe leer teniendo en cuenta los dibujos adjuntos, en los que:

- la figura única es un esquema que representa un sistema en una realización de la presente invención.

El sistema representado en la figura única comprende 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.

El terminal 2 comprende un micrófono 4, que recibe la voz a reconocer de un usuario en forma de una señal de audio. El terminal 2 también comprende un módulo de cálculo de parámetros de modelización 6, que realiza de forma conocida por sí misma un análisis acústico que permite extraer los parámetros pertinentes de la señal de audio y que eventualmente puede realizar ventajosamente una función de eliminación de ruidos. El terminal 2 comprende un controlador 8, que selecciona una señal entre la señal de audio y una señal indicativa de los parámetros calculados por el módulo de cálculo de parámetros 6. El terminal comprende además una interfaz 10 para la emisión en la red de la señal seleccionada, con destino al servidor.

El servidor 1 comprende una interfaz de red 12 para recibir las señales que le son remitidas, un controlador 14 que analiza la señal recibida y la dirige a continuación selectivamente hacia un módulo de tratamiento entre varios módulos 16, 18, 20. El módulo 16 es un detector de actividad vocal, que asegura la detección de los segmentos que corresponden a la voz y que deben ser reconocidos. El módulo 18 asegura el cálculo de parámetros de modelización de forma semejante al módulo de cálculo 6 del terminal. Sin embargo, el modelo de cálculo puede ser diferente. El módulo 20 ejecuta un algoritmo de reconocimiento de tipo conocido, por ejemplo a base de modelos de Markov ocultos con un vocabulario por ejemplo superior a 100.000 palabras. Este motor de reconocimiento 20 compara los parámetros de entrada con modelos de voz que representan palabras o frases y determina la mejor forma asociada, teniendo en cuenta modelos sintácticos que describen las sucesiones de palabras esperadas, modelos léxicos que precisan las diferentes pronunciaciones de las palabras y modelos acústicos representativos de los sonidos pronunciados. Estos modelos son por ejemplo multi-locutores, capaces de reconocer con buena fiabilidad, la voz, independientemente del locutor.

El controlador 14 controla el módulo de VAD 16, el módulo de cálculo de parámetros 18 y el motor de reconocimiento 20 para:

a) cuando la señal recibida por la interfaz de recepción 12 es de tipo audio y no indica segmentos de voz obtenidos mediante detección de actividad vocal, activar el módulo de VAD 16 remitiéndole la señal recibida como señal de entrada y después remitir los segmentos de voz extraídos por el módulo de VAD 16 al módulo de cálculo de parámetros 18 como parámetros de entrada y después remitir los parámetros calculados por estos medios de cálculo de parámetros 18 al motor de reconocimiento 20 como parámetros de entrada,

b) cuando la señal recibida por la interfaz de recepción 12 es de tipo audio e indica segmentos de voz después de la detección de actividad vocal, activar el módulo de cálculo de parámetros 18 remitiéndole la señal recibida como señal de entrada y después remitir los parámetros calculados por este módulo de cálculo de parámetros 18 al motor de reconocimiento 20 como parámetros de entrada,

c) cuando la señal recibida por la interfaz de recepción 12 indica parámetros de modelización, remitir dichos parámetros indicados al motor de reconocimiento 20 como parámetros de entrada.

Por ejemplo, en el caso en que el usuario del terminal 1 utilice una aplicación que permita solicitar informaciones sobre la bolsa y dice “cotización al cierre de los tres últimos días del valor Lambda”, la señal de audio correspondiente es capturada por el micrófono 4. En la realización del sistema de acuerdo con la invención, esta señal es tratada a continuación, por defecto, por el módulo de cálculo de parámetros 6 y después se envía una señal, que indica los parámetros de modelización calculados, hacia el servidor 1.

Cuando surgen, por ejemplo, problemas de disponibilidad de canales de datos o del módulo de cálculo 6, es la señal de audio a la salida del micrófono 4 la que selecciona el controlador 8 para transmitirla con destino al servidor 1.

El controlador también puede ser adecuado para enviar sistemáticamente una señal que indica los parámetros de modelización.

El servidor recibe la señal con la interfaz de recepción 12 y después realiza, para efectuar el reconocimiento de voz en la señal recibida, el tratamiento indicado en a) o b) si la señal enviada por el terminal 1 es de tipo audio o el tratamiento indicado en c) si la señal enviada por el terminal 1 indica parámetros de modelización.

El servidor de acuerdo con la invención también es adecuado para realizar el reconocimiento de voz en una señal transmitida por un terminal que no dispone de medios de cálculo de parámetros de modelización, ni de medios de reconocimiento y que dispone eventualmente de medios de detección de actividad vocal.

Ventajosamente, en una realización de la invención, el sistema puede comprender además un terminal de usuario 22, que comprende un micrófono 24 similar al del terminal 2, un módulo 26 de detección de actividad vocal. La función del módulo 26 es semejante a la del módulo de detección de actividad vocal 16 del servidor 1. Sin embargo, el modelo de detección puede ser diferente. El terminal 22 comprende un módulo de cálculo de parámetros de modelización 28, un motor de reconocimiento 30 y un controlador 32. El terminal comprende una interfaz 10 para la emisión en la red, con destino al servidor, de la señal seleccionada por el controlador 32.

El motor de reconocimiento 30 del terminal puede tratar por ejemplo un vocabulario de menos de 10 palabras. Este motor puede funcionar en modo mono-locutor y no necesita una fase de aprendizaje previa a partir de la voz del usuario.

El reconocimiento de voz se puede realizar de diferentes maneras:

- exclusivamente a nivel del terminal, o

- exclusivamente a nivel del servidor, o

- parcial o totalmente a nivel del terminal e igualmente, de manera alternativa o simultánea, parcial o totalmente a nivel del servidor.

Cuando se debe realizar la elección de la forma finalmente conservada, entre una forma asociada proporcionada por el módulo de reconocimiento del servidor y una forma asociada proporcionada por los del terminal, se puede realizar en base a diferentes criterios, que pueden variar de un terminal al otro, pero también de una aplicación a otra o de un contexto dado a otro. Estos criterios pueden dar por ejemplo prioridad al reconocimiento realizado a nivel del terminal o a la forma asociada que presente la mayor tasa de probabilidad o también a la forma que se determine más rápidamente.

La forma en la que se realiza este reconocimiento se puede fijar a nivel del terminal en un modo dado. O puede variar en función particularmente de criterios vinculados a la aplicación en cuestión, a problemáticas de carga de los diferentes medios a nivel del terminal y del servidor o también a problemáticas de disponibilidad de canales de transmisión de voz o de datos. Los controladores 32 y 14 situados respectivamente a nivel del terminal y del servidor traducen la forma en la que se debe realizar el reconocimiento.

El controlador 32 del terminal es adecuado para seleccionar una señal entre la señal de audio de origen a la salida del micrófono 24, una señal de audio representativa de segmentos de voz extraídos por el módulo de VAD 26 y una señal que indica parámetros de modelización 28. Según el caso, el tratamiento a nivel del terminal continuará o no más allá de la etapa de tratamiento del terminal que suministra la señal a emitir.

Por ejemplo, se puede considerar una realización en la que el módulo de VAD 26 del terminal está diseñado por ejemplo para detectar rápidamente palabras de mando y el módulo de VAD 16 del servidor puede ser más lento, pero está diseñado para detectar frases completas. Una aplicación, en la que el terminal 22 realiza un reconocimiento local y de forma simultánea conlleva la realización de un reconocimiento por parte del servidor a partir de la señal de audio transmitida, permite particularmente sumar las ventajas de cada módulo de detección vocal.

Consideraremos ahora una aplicación en la que el reconocimiento se realiza exclusivamente de forma local (terminal) o exclusivamente a distancia (servidor centralizado) en base a palabras clave que permitan la conmutación:

El reconocimiento en curso es en primer lugar local: el usuario dice “llamar a Antoine”, figurando Antoine en el directorio local. A continuación dice “mensajería”, palabra clave que es reconocida de forma local y que hace bascular a reconocimiento por parte del servidor. El reconocimiento es ahora a distancia. Dice “buscar el mensaje de Josiane”. Cuando dicho mensaje se ha escuchado, dice “terminado”, palabra clave que hace bascular de nuevo a la aplicación a reconocimiento local.

La señal transmitida al servidor, para realizar allí el reconocimiento, era de tipo señal de audio. En otra realización, ésta podría indicar los parámetros de modelización calculados en el terminal.

Consideraremos ahora una aplicación en la cual el reconocimiento a nivel del terminal y el reconocimiento a nivel del servidor son alternos. El reconocimiento se realiza en primer lugar a nivel del terminal 22 y la señal después de la detección vocal se almacena. Si la respuesta es consistente, es decir si no hay rechazo del módulo de reconocimiento 30 y si la señal reconocida es válida desde el punto de vista de la aplicación, la aplicación local en el terminal pasa a la siguiente fase de la aplicación. En caso contrario, la señal almacenada es enviada al servidor para realizar el reconocimiento en una señal que indica segmentos de voz después de la detección de actividad vocal en la señal de audio (en otra realización, son los parámetros de modelización los que podrían almacenarse).

De este modo, el usuario dice “llamar a Antoine”; el conjunto del tratamiento a nivel del terminal 22 se realiza con almacenamiento de la señal. La señal es reconocida con éxito de forma local. Dice entonces “buscar el mensaje de Josiane”; el reconocimiento a nivel del terminal fracasa; la señal almacenada se transmite entonces al servidor. La señal es reconocida correctamente y el mensaje solicitado se reproduce.

En otra aplicación, el reconocimiento se realiza simultáneamente a nivel del terminal y también, y esto independientemente del resultado del reconocimiento local, a nivel del servidor. El usuario dice “llamar a Antoine”. El reconocimiento se desarrolla a los dos niveles. Como el tratamiento local interpreta la orden, el resultado a distancia no es tenido en cuenta. Después el usuario dice “buscar el mensaje de Josiane” que genera un fracaso de forma local y que es reconocido correctamente a nivel del servidor.

En una realización, el motor de reconocimiento 30 del terminal 22 es un programa ejecutable descargado desde el servidor mediante medios clásicos de transferencia de datos.

Ventajosamente, para una aplicación dada del terminal 22, pueden descargarse o actualizarse modelos de reconocimiento del terminal, durante una sesión de la aplicación conectada a la red.

Otros recursos lógicos útiles para el reconocimiento de voz también se pueden descargar desde el servidor 1, como el módulo 6, 28 de cálculo de parámetros de modelización o el detector de actividad vocal 26.

Se podrían describir otros ejemplos, que emplean por ejemplo aplicaciones vinculadas a coches, a electrodomésticos o multimedia.

Como se presenta en los ejemplos de realización descritos anteriormente, un sistema de acuerdo con la invención permite utilizar de forma optimizada los diferentes recursos necesarios para el tratamiento del reconocimiento de voz y presentes a nivel del terminal y del servidor.

REIVINDICACIONES

1. Sistema de reconocimiento de voz distribuido, que comprende al menos un terminal de usuario y al menos un servidor adecuados para comunicarse entre sí por medio de una red de telecomunicaciones, en el que el terminal de usuario comprende:

- medios de obtención de una señal de audio a reconocer,

- primeros medios de cálculo de parámetros de modelización de la señal de audio, y

- primeros medios de control para seleccionar al menos una señal a emitir con destino al servidor entre la señal de audio a reconocer y una señal que indica los parámetros de modelización calculados, en función del contexto de la aplicación del terminal;

y en el que el servidor comprende:

- medios de recepción de la señal seleccionada procedente del terminal de usuario,

- segundos medios de cálculo de parámetros de modelización de una señal de entrada,

- medios de reconocimiento para asociar al menos una forma memorizada a parámetros de entrada, y

- segundos medios de control para controlar los segundos medios de cálculo y los medios de reconocimiento para:

  - cuando la señal seleccionada recibida por los medios de recepción es de tipo audio, activar los segundos medios de cálculo de parámetros remitiéndoles la señal seleccionada como señal de entrada y remitir los parámetros calculados por los segundos medios de cálculo a los medios de reconocimiento como parámetros de entrada, y

  - cuando la señal seleccionada recibida por los medios de recepción indica parámetros de modelización, remitir dichos parámetros indicados a los medios de reconocimiento como parámetros de entrada.

2. Sistema de acuerdo con la reivindicación 1, en el que los primeros medios de control seleccionan una señal a emitir en función, además, del estado de la red y/o según una coordinación entre los medios de control respectivos del terminal y del servidor.

3. Sistema de acuerdo con la reivindicación 1, 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 extractos de una señal de audio de origen, fuera de periodos de inactividad vocal.

4. Sistema de acuerdo con la reivindicación 3, en el que los primeros medios de control son adecuados para seleccionar la señal a emitir con destino al servidor entre al menos la señal de audio de origen, la señal de audio a reconocer en forma de segmentos extraídos por los medios de detección de actividad vocal y la señal que indica parámetros de modelización calculados por los primeros medios de cálculo de parámetros.

5. Sistema de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que:

- el servidor comprende además medios de detección de actividad vocal para extraer de una señal de tipo audio, fuera de periodos de inactividad vocal, segmentos de voz, y

- los segundos medios de control son adecuados para controlar los segundos medios de cálculo y los medios de reconocimiento cuando la señal seleccionada recibida por los medios de recepción es de tipo audio, para:

si la señal de tipo audio es representativa de segmentos de voz después de la detección de actividad vocal, activar los segundos medios de cálculo de parámetros remitiéndoles la señal seleccionada como señal de entrada y después remitir los parámetros calculados por los segundos medios de cálculo de parámetros a los medios de reconocimiento como parámetros de entrada;

de lo contrario, activar los medios de detección de actividad vocal del servidor remitiéndoles la señal recibida como señal de entrada y después remitir los segmentos extraídos por los medios de detección de actividad vocal a los segundos medios de cálculo de parámetros como señal de entrada y después remitir los parámetros calculados por los segundos medios de cálculo de parámetros a los medios de reconocimiento, como parámetros de entrada.

6. Sistema de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que el terminal de usuario comprende además medios de almacenamiento adecuados para almacenar la señal de audio reconocer o los parámetros de modelización calculados por los primeros medios de cálculo de parámetros.

7. Terminal de usuario para implementar un sistema de reconocimiento de voz distribuido de acuerdo con una de las reivindicaciones 1 a 6, que comprende:

- medios de obtención de una señal de audio a reconocer,

- medios de cálculo de parámetros de modelización de la señal de audio, y

- primeros medios de control para seleccionar al menos una señal a emitir con destino al servidor entre la señal de audio a reconocer y una señal que indica los parámetros de modelización calculados, en función del contexto de la aplicación del terminal.

8. Terminal de usuario de acuerdo con la reivindicación 7, en el que los primeros medios de control seleccionan una señal a emitir en función, además, del estado de la red y/o según una coordinación entre los medios de control respectivos del terminal y del servidor.

9. Terminal de usuario de acuerdo con la reivindicación 7 u 8, en el que al menos una parte de los medios de cálculo de parámetros se descarga desde el servidor.

10. Terminal de usuario de acuerdo con la reivindicación 7 u 8, en el que al menos una parte de los medios de reconocimiento se descarga desde el servidor.

11. Servidor para implementar un sistema de reconocimiento de voz distribuido de acuerdo con una de las reivindicaciones 1 a 6, que comprende:

- medios de recepción, procedente de un terminal de usuario, de una señal seleccionada en dicho terminal,

- medios de cálculo de parámetros de modelización de una señal de entrada,

- medios de reconocimiento para asociar al menos una forma memorizada a parámetros de entrada, y

- medios de control para controlar los segundos medios de cálculo y los medios de reconocimiento para:

  - cuando la señal seleccionada recibida por los medios de recepción es de tipo audio, activar los medios de cálculo de parámetros remitiéndoles la señal seleccionada como señal de entrada y remitir 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 indica parámetros de modelización, remitir dichos parámetros indicados a los medios de reconocimiento como parámetros de entrada.

12. Servidor de acuerdo con la reivindicación 11, que comprende medios para descargar por medio de la red de telecomunicaciones, con destino a un terminal, al menos una parte de los primeros medios de cálculo de parámetro o de los medios de reconocimiento del terminal.

13. Servidor de acuerdo con la reivindicación 12, que comprende medios para descargar recursos lógicos de reconocimiento de voz por medio de la red de telecomunicaciones con destino a un terminal.

14. Servidor de acuerdo con la reivindicación 13, en el que dichos recursos comprenden al menos un módulo de 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 parámetros de modelización.

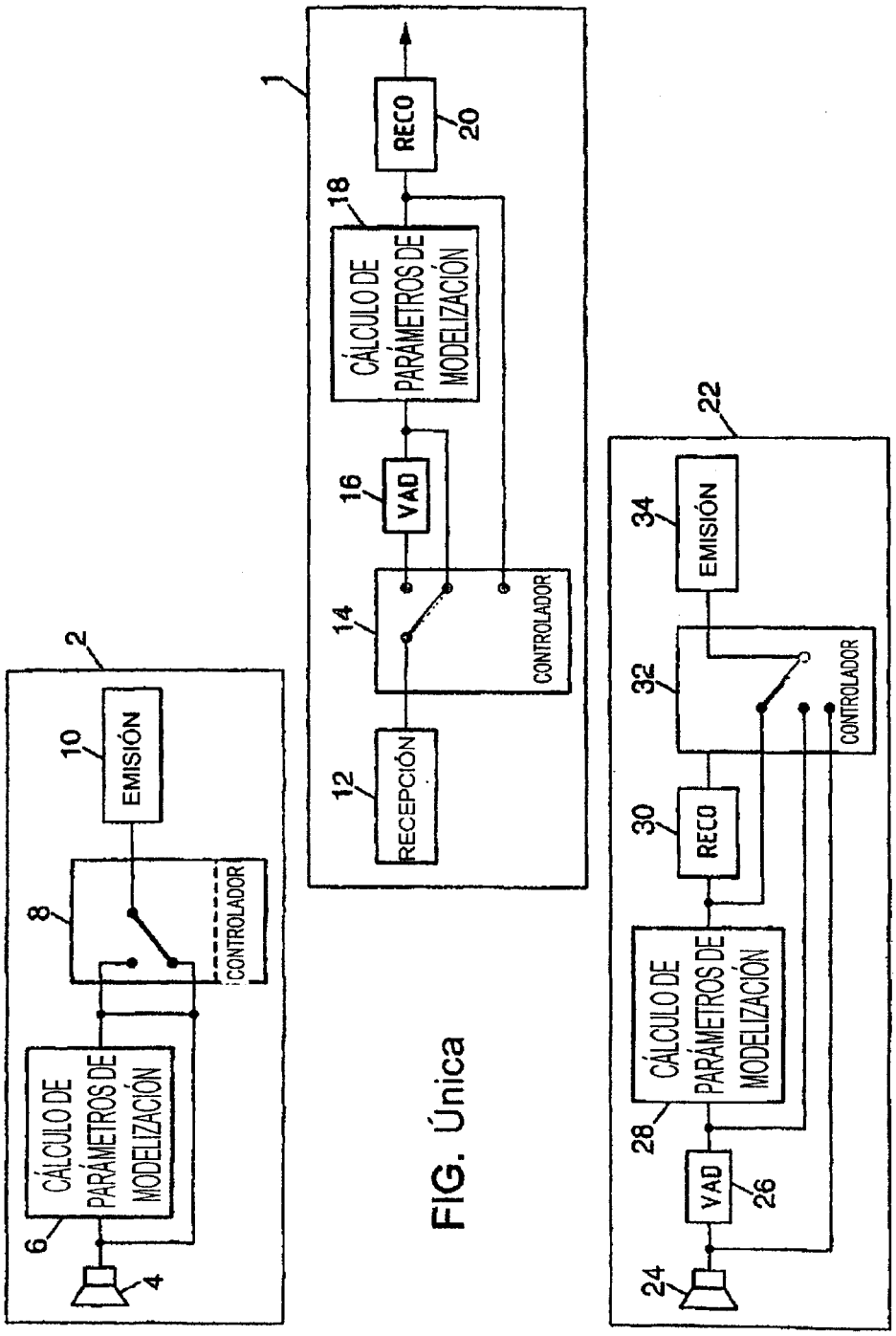
1
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
EMISIÓN
CONTROLADOR
CÁLCULO DE PARÁMETROS DE MODELIZACIÓN
RECEPCIÓN
VAD
RECO

FIG. Única