



①9



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

①1 Número de publicación: **2 279 489**

⑤1 Int. Cl.:
G10L 15/18 (2006.01)
G10L 15/22 (2006.01)

①2

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑧6 Número de solicitud europea: **05290726 .8**
⑧6 Fecha de presentación : **01.04.2005**
⑧7 Número de publicación de la solicitud: **1585110**
⑧7 Fecha de publicación de la solicitud: **12.10.2005**

⑤4 Título: **Sistema de aplicación de voz.**

③0 Prioridad: **02.04.2004 FR 04 03511**

④5 Fecha de publicación de la mención BOPI:
16.08.2007

④5 Fecha de la publicación del folleto de la patente:
16.08.2007

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

⑦2 Inventor/es: **Paillet, Eric;**
Dubois, Dominique y
Larreur, Danielle

⑦4 Agente: **Lehmann Novo, María Isabel**

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

Sistema de aplicación de voz.

5 La presente invención se refiere a los sistemas automáticos de reconocimiento de voz capaces de desencadenar una acción en función de una frase pronunciada por un usuario.

Tales sistemas se utilizan particularmente en los servidores de voz de redes de telecomunicaciones.

10 Estos servidores de voz se utilizan en el seno de aplicaciones de voz interactivas en las cuales se entabla un diálogo entre un usuario y un autómata con el fin de determinar lo que se espera de este usuario.

15 Están compuestos de un sistema de reconocimiento de voz que proporciona una frase en bruto pronunciada por el usuario y un sistema de análisis semántico que descompone esta frase en una serie de palabras clave. Además, este posee un conjunto de reglas que asocian a un conjunto de palabras clave una acción por ejecutar. El analizador semántico busca entonces la regla, o las reglas, cuyas palabras clave esperadas se encuentran en la frase pronunciada por el usuario.

20 Si son seleccionadas así varias reglas, el analizador semántico selecciona la regla más pertinente utilizando criterios tales como una ponderación probabilista, el contexto en el cual ha sido pronunciada la frase. Esto ha sido divulgado en el documento de Tsuboi y cols.: "A real-time task-oriented speech understanding system using keyword spotting", ICASSP1992, págs 197-200.

25 Una vez elegida la regla, la acción que define es ejecutada por un sistema de gestión de diálogos. A menudo, en los servidores de voz, la acción corresponde a la generación de una frase pregrabada que aporta la respuesta esperada por el usuario o bien que formula una pregunta que permite delimitar mejor lo que se espera de éste.

30 Las técnicas utilizadas actualmente por los analizadores semánticos funcionan según el principio de una correspondencia estricta entre las palabras encontradas en el enunciado y las palabras esperadas en la regla.

35 Así, cuando una palabra clave está presente en el enunciado, aunque no sea determinante para el sentido general, debe encontrarse en la regla para que ésta sea aceptada.

Ahora bien, este tipo de funcionamiento está poco adaptado a las frases que se dan habitualmente en el lenguaje oral, en particular porque estas frases son ruidosas, gramaticalmente incorrectas, están mal construidas, e incluyen a menudo dudas o información redundante que no han sido previstas a la hora de escribir las reglas.

40 Esta sensibilidad extrema obliga entonces al creador a escribir todas las reglas posibles teniendo en cuenta todos los errores de sintaxis imaginables.

Este inconveniente limita entonces en gran medida el uso de tales sistemas.

45 El objetivo de la invención es entonces el de proponer un sistema de aplicación de voz que reconozca con facilidad las reglas aplicables a pesar de los ruidos e imperfecciones de la frase pronunciada.

Según la invención, un sistema es reivindicado según lo enunciado en la reivindicación 1, un procedimiento según la reivindicación 5, y un programa según la reivindicación 9.

50 El objeto de la invención es entonces un sistema de aplicación de voz que incluye unos medios de adquisición de al menos una frase pronunciada por al menos un usuario, conectados a unos medios de análisis semántico que incluyen unos medios de reconocimiento de palabras clave que pertenecen a la frase pronunciada y están preparados para generar una lista ordenada de palabras clave, denominada enunciado, de la frase pronunciada, estando dichos medios de reconocimiento conectados a unos medios de asociación, en forma de reglas, entre al menos una palabra clave predeterminada y una acción determinada y unos medios de selección de al menos una acción determinada cuando un conjunto de palabras clave contenidas en la regla correspondiente están presentes en la frase pronunciada, caracterizado porque los medios de selección recorren para identificación el conjunto de reglas y buscan, para cada regla dada, la presencia, en la frase pronunciada, de un conjunto de palabras clave de esta regla para seleccionar la acción determinada correspondiente relativa a la regla así resuelta e identificada.

60 Según otras características de la invención:

- el conjunto de palabras clave de una regla está formado de subconjuntos ordenados por palabras clave denominadas expresiones, estando cada palabra clave o expresión combinada con otras palabras clave o expresiones de manera que al menos dos palabras clave o expresiones sean, bien intercambiables, bien estén presentes en un orden de aparición determinado, o bien estén presentes en un orden cualquiera;
- para una regla dada que incluye un conjunto de expresiones, los medios de selección seleccionan la acción correspondiente cuando la regla ha sido resuelta completamente,

si no buscan en la expresión corriente la primera palabra clave del enunciado y,

si se encuentra la primera palabra clave, buscan el resto de palabras clave de la expresión en el enunciado y

si esta última búsqueda resulta infructuosa, la expresión corriente queda invalidada para esta primera palabra clave y se inicia una nueva búsqueda,

si no la regla se resuelve y la acción correspondiente queda seleccionada, y si no se encuentra la primera palabra clave, se retoma la búsqueda entre el resto de palabras clave;

- los medios de análisis semántico incluyen además unos medios de conexión preparados para determinar la acción que debe ejecutarse entre el conjunto de acciones seleccionadas.

La invención se comprenderá mejor tras la lectura de la descripción que figura a continuación, aportada únicamente a modo de ejemplo, y hecha en referencia a los dibujos adjuntos, en los cuales

- la fig. 1 es un esquema del conjunto de la invención;

- la fig. 2 es un organigrama de un servidor de voz que utiliza la invención;

- la fig. 3 es un organigrama general de la invención; y

- la fig. 4 es un organigrama detallado de la invención para una regla.

Un sistema de aplicación de voz según la invención incluye, figura 1, unos medios 1 de adquisición de frases pronunciadas por un usuario.

De forma clásica, estos medios de adquisición se componen de un micrófono, por ejemplo el de un teléfono combinado, conectado a una tarjeta electrónica que transforma la señal analógica generada por el micrófono en una serie de datos numéricos representativa de la señal recibida.

Estos medios 1 de adquisición están conectados a unos medios 2 de reconocimiento de voz.

Estos medios 2 de reconocimiento utilizan, de forma clásica, unas tecnologías de tipo N-gram muy conocidas. Compañías como Nuance y Scansoft proponen comercialmente tales tecnologías particularmente bien adaptadas al habla continua. Otras tecnologías de reconocimiento de voz son igualmente factibles sin que ello modifique la invención.

Los medios 2 de reconocimiento de voz transforman así la serie de datos numéricos recibida de los medios de adquisición 1 en una frase en bruto.

Los medios 3 de análisis semántico, o analizador semántico, incluyen unos medios 8 de reconocimiento de palabras clave que transforman la frase en bruto en un conjunto ordenado de palabras clave reconocidas, o enunciado.

Incluyen además unos medios 4 de asociación entre unas palabras clave y unas acciones. Estos medios de asociación se presentan preferentemente en forma de reglas del tipo: <palabra clave 1> <palabra clave 2> ... <palabra clave N> → acción.

Los medios 3 de análisis semántico incluyen igualmente unos medios 5 de selección que comparan el conjunto ordenado de palabras clave reconocidas en la frase pronunciada con las diferentes reglas 4. Las reglas 4 que corresponden al conjunto de palabras clave definen así el conjunto de acciones potenciales por efectuar.

El analizador semántico 3 incluye igualmente unos medios 6 de conexión. Estos medios 6 de conexión son utilizados cuando varias reglas han sido seleccionadas con el fin de determinar cuál contiene la acción que debe ser ejecutada.

Una vez elegida la acción, ésta es efectuada por unos medios 9 de diálogo que generan una frase adaptada y la transmiten al usuario como respuesta a la frase que él ha emitido.

Esta frase puede ser una respuesta o bien una pregunta que permite precisar lo que se espera del cliente y crea así un diálogo entre el usuario y el servidor.

Las acciones generadas pueden corresponder igualmente a órdenes de un autómata. Por ejemplo, un sistema de control/mando de un proceso puede utilizar un sistema de aplicación de voz según la invención para recoger las órdenes de un operador, en lugar de, o como complemento de, interfaces más tradicionales como son el teclado y la pantalla.

El analizador semántico 3 será ahora descrito de forma más precisa según su modo de funcionamiento.

ES 2 279 489 T3

Como se ha indicado más arriba, cada acción está asociada a una serie de palabras clave ordenadas, correspondiendo el todo a una regla.

El conjunto de reglas, figura 2, está almacenado en un analizador semántico, por ejemplo en forma de fichero. Un modo preferente de realización consiste en reunir las reglas en un fichero de texto que contenga una regla por línea.

Las palabras clave son entonces ordenadas utilizando tres operadores.

El primer operador, denominado &, corresponde al operador Y ordenado. Así, A & B indica que las palabras clave A y B deben estar presentes y que B sigue a A en el orden del enunciado.

El segundo operador, denominado #, corresponde al operador Y no ordenado. A # B indica que las palabras clave A y B deben estar presentes y que el orden de aparición de A y de B en la frase no tiene importancia: AB y BA son reconocidos como pertenecientes a esta regla.

El tercer operador, denominado |, corresponde al operador O. A|B indica que el enunciado debe contener A o B, alternativamente. Las palabras clave A y B son por tanto intercambiables.

Estos tres operadores son combinables entre ellos y unos paréntesis permiten definir grupos de palabras clave.

Por ejemplo (A|B) & (C#D) indica que la regla es válida para un enunciado que empieza por las palabras clave A o B seguidas de CD o de DC.

Según la forma preferente de realización de la invención, la acción por realizar correspondiente a la regla está escrita al final de la línea, después de las palabras clave, y figura entre llaves.

El analizador semántico 3 recibe figura 3, etapa 10, en entrada una frase en forma de una serie ordenada de palabras clave, o enunciado, y posee un conjunto de reglas en forma de un fichero.

Lee una primera regla en 11 y busca la palabra clave esperada por ésta. Una regla se marca como válida en 12 cuando la serie de palabras clave que define se encuentra en el enunciado.

No obstante, puede ocurrir que haya palabras esperadas en la regla que estén separadas en el enunciado por otras palabras no previstas. Estas son entonces eliminadas y consideradas como ruido no pertinente.

El analizador semántico intenta sin embargo sistemáticamente comprobar la conformidad de la frase con la regla.

A continuación, una vez agotadas todas las posibilidades de concordancia o una vez se ha descubierto que la regla se aplica, el analizador busca la regla siguiente en 13. Si ésta existe, es analizada como se ha indicado más arriba, si no, el analizador semántico transmite en 14 a los medios 6 de conexión del conjunto de reglas válidas.

Así, de forma particularmente ventajosa, el analizador semántico 3 permite ignorar algunas palabras clave del enunciado y considerar como información no determinante, es decir, como ruido, todo lo que se encuentra entre dos palabras no esperadas.

Con el fin de permitir una exploración completa de las posibilidades del enunciado con respecto a la lista de palabras clave de la regla, el analizador semántico utiliza el siguiente recorrido recursivo, figura 4:

1. Si la expresión ha sido completamente resuelta en 20, se obtiene una regla correcta en 21 aunque queden palabras clave no sometidas a prueba;

2. Si no, busca en 22 en la expresión de la primera palabra de la lista de palabras clave;

3. Si la palabra ha sido encontrada en 23, de la misma forma se inicia una búsqueda en 24 con el resto de palabras clave:

a. Si la búsqueda con el resto de palabras clave ha fracasado en 25, se invalida en 27 la subexpresión que haya permitido encontrar la primera palabra, para esa primera palabra y únicamente para ella (se la considera como ruido) y se retoma de nuevo la búsqueda. El resultado final es entonces el resultado de esta nueva búsqueda.

b. Si la búsqueda con el resto de palabras ha sido satisfactoria en 25, se ha encontrado una regla correcta en 26.

4. Si no se ha encontrado la palabra en 23, se la considera como ruido en 28 y se inicia la búsqueda en 22 entre el resto de palabras clave. El resultado final es entonces el resultado de esta nueva búsqueda.

ES 2 279 489 T3

Esto permite una vuelta atrás si una subexpresión iniciada ha fracasado y si quedan alternativas en la regla que no han sido exploradas.

Con el fin de que se comprenda mejor este funcionamiento, supongamos, a modo de ejemplo, que el enunciado es

[Móvil] [Limitar] [Cantidad] [Pagar] [Descuento] [Pagar] [Cosa] [Caro]

y que la regla define la expresión

((Descuento # (Limitar & Cantidad) # (Pagar & Caro)) # Móvil).

El desarrollo del algoritmo se hace de la siguiente forma:

1.- Búsqueda de la palabra [Móvil] en la expresión, búsqueda satisfactoria.

2.- Búsqueda de [Limitar] satisfactoria, se inicia la subexpresión (Limitar & Cantidad).

3.- Búsqueda de [Cantidad] en la subexpresión iniciada, satisfactoria. La subexpresión (Limitar & Cantidad) queda resuelta.

4.- Búsqueda de [Pagar], satisfactoria y subexpresión (Pagar & Caro) iniciada.

5.- Búsqueda de [Descuento] en la subexpresión iniciada. Fracaso de la búsqueda. Se considera [Descuento] como ruido y se continúa.

6.- Búsqueda del segundo [Pagar] en la subexpresión iniciada. Nuevo fracaso de la búsqueda. Se considera el segundo [Pagar] como ruido y se continúa.

7.- [Cosa] tampoco se encuentra en la expresión iniciada. [Cosa] se considera como ruido.

8.- La palabra clave [Caro] se busca en la expresión iniciada. Búsqueda satisfactoria de la palabra [Caro], no obstante, no quedan más palabras clave y la expresión no ha sido resuelta completamente. Se regresa al punto 7 con un fracaso de la resolución de la regla.

7.1- Dado que [Cosa] no ha sido encontrada, se regresa al punto 6.

6.1- Dado que [Pagar] ha sido considerada como ruido se regresa al punto 5.

5.1- Ocorre lo mismo con [Descuento], se regresa al punto 4.

4.1- Al haber encontrado [Pagar], se invalida la subexpresión (Pagar & Caro) para la búsqueda de este primer [Pagar], pero sigue siendo accesible para la búsqueda del segundo [Pagar]. Esta subexpresión ya no se considera como iniciada. Se busca de nuevo el primer [Pagar]. Esta vez la búsqueda fracasa dado que la subexpresión (Pagar & Caro) es inaccesible. Se considera el primer [Pagar] como ruido y se continúa.

5.2- Búsqueda de [Descuento] que se encuentra dado que esta vez no se ha iniciado subexpresión alguna.

6.2- Búsqueda del segundo [Pagar] que se encuentra, la subexpresión (Pagar & Caro) se inicia de nuevo.

7.2- Búsqueda de [Cosa], fracaso de la búsqueda, se la considera por tanto como ruido y se continúa.

8.1- Búsqueda de [Caro] satisfactoria. La expresión se resuelve completamente, por tanto se ha logrado encontrar una regla correcta.

Así, la invención permite, de forma particularmente ventajosa, al sistema de reconocimiento de voz, reconocer las reglas aplicables a pesar de los ruidos e imperfecciones de la frase pronunciada.

REIVINDICACIONES

1. Sistema de aplicación de voz que incluye unos medios (1, 2) de adquisición de al menos una frase pronunciada por al menos un usuario, conectados a unos medios (3) de análisis semántico que incluye unos medios (8) de reconocimiento de palabras clave que pertenecen a la frase pronunciada y están preparados para generar una lista ordenada de palabras clave, denominada enunciado, de la frase pronunciada, estando dichos medios de reconocimiento conectados a unos medios (4) de asociación, en forma de reglas, entre al menos una palabra clave predeterminada y una acción determinada y unos medios (5) de selección de al menos una acción determinada cuando un conjunto de palabras clave contenidas en la regla correspondiente está presente en la frase pronunciada, **caracterizado** porque los medios (5) de selección recorren para identificación el conjunto de reglas y buscan, para cada regla dada, la presencia, en la frase pronunciada, de un conjunto de palabras clave de esta regla para seleccionar la acción determinada correspondiente relativa a la regla así resuelta e identificada.

2. Sistema de aplicación de voz según la reivindicación 1, **caracterizado** porque el conjunto de palabras clave de una regla está formado por subexpresiones ordenadas de palabras clave denominadas expresiones, estando cada palabra clave o expresión combinada con otras palabras clave o expresiones de forma que al menos dos palabras clave o expresiones sean, bien intercambiables, bien estén presentes siguiendo un orden de aparición determinado, o bien estén presentes siguiendo un orden cualquiera.

3. Sistema de aplicación de voz según la reivindicación 2, **caracterizado** porque, para una regla dada que incluye un conjunto de expresiones, los medios de selección seleccionan la acción correspondiente cuando la regla ha sido completamente resuelta,

si no buscan en la expresión corriente la primera palabra clave del enunciado y,

si se encuentra la primera palabra clave, buscan el resto de palabras clave de la expresión en el enunciado y

si esta última búsqueda resulta infructuosa, la expresión corriente queda invalidada para esta primera palabra clave y se retoma la búsqueda,

si no la regla se resuelve y se selecciona la acción correspondiente, y si no se encuentra la primera palabra clave, la búsqueda se reanuda entre el resto de palabras clave.

4. Sistema de aplicación de voz según una cualquiera de las reivindicaciones precedentes, **caracterizado** porque los medios (3) de análisis semántico incluyen además unos medios (6) de conexión preparados para determinar la acción por ejecutar entre el conjunto de acciones seleccionadas.

5. Procedimiento de reconocimiento de voz que incluye una etapa previa de asociación, en forma de reglas, entre al menos una palabra clave predeterminada y una acción determinada, e incluye además las etapas:

- de adquisición de al menos una frase pronunciada por al menos un usuario,

- de análisis semántico incluyendo una subetapa de reconocimiento de palabras clave que pertenecen a la frase pronunciada y una subetapa de generación de una lista ordenada de palabras clave, denominada enunciado, de la frase pronunciada, y

- de selección de al menos una acción determinada cuando un conjunto de palabras clave contenidas en la regla correspondiente están presentes en la frase pronunciada,

caracterizada porque durante la etapa de selección, se recorre para identificación el conjunto de reglas y se busca, para cada regla dada, la presencia, en la frase pronunciada, de un conjunto de palabras clave de esta regla para seleccionar la acción determinada correspondiente relativa a la regla así resuelta e identificada.

6. Procedimiento según la reivindicación 5, **caracterizado** porque el conjunto de palabras clave de una regla está formado de subconjuntos ordenados de palabras clave denominadas expresiones, estando cada palabra clave o expresión combinada con otras palabras clave o expresiones de forma que al menos dos palabras clave o expresiones sean, bien intercambiables, bien estén presentes en un orden de aparición determinado, o bien estén presentes en un orden cualquiera.

7. Procedimiento según la reivindicación 6, **caracterizado** porque para una regla dada que incluye un conjunto de expresiones, durante la etapa de selección,

se selecciona la acción correspondiente cuando la regla ha sido completamente resuelta,

si no se busca en la expresión corriente la primera palabra clave del enunciado y,

si se encuentra la primera palabra clave, se buscan el resto de palabras clave de la expresión en el enunciado y

ES 2 279 489 T3

si esta última búsqueda resulta infructuosa, la expresión corriente queda invalidada para esta primera palabra clave y la búsqueda se reanuda,

5 si no la regla se resuelve y se selecciona la acción correspondiente, y si no se encuentra la primera palabra clave, se reanuda la búsqueda entre el resto de palabras clave.

10 8. Procedimiento según una cualquiera de las reivindicaciones 5 a 7, **caracterizado** porque la etapa de análisis semántico incluye además una subetapa de determinación de la acción por ejecutar entre el conjunto de acciones seleccionadas.

9. Programa de ordenador que incluye instrucciones de programa adaptadas para implementar un procedimiento de reconocimiento de voz según una cualquiera de las reivindicaciones 5 a 8, cuando dicho programa es ejecutado por un sistema informático.

15 10. Soporte de información legible por un ordenador, en el cual se encuentra almacenado un programa de ordenador según la reivindicación 9.

20

25

30

35

40

45

50

55

60

65

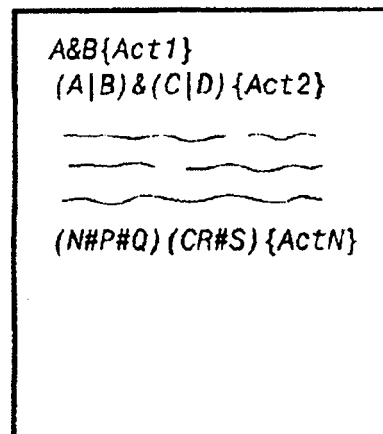
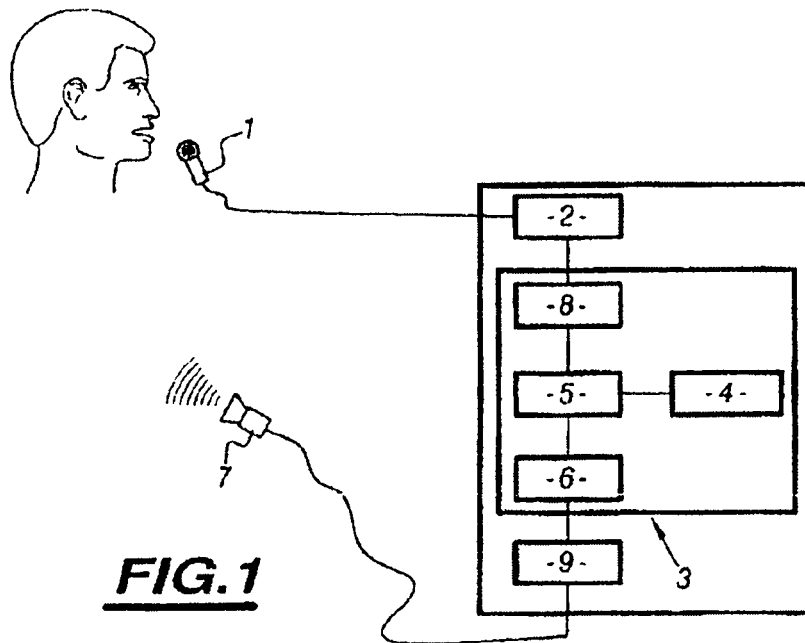


FIG. 2

FIG.3

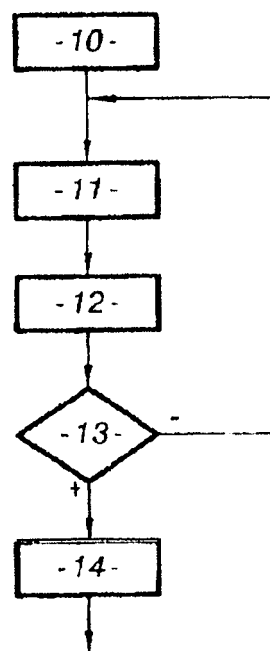


FIG.4

