(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 391 454**

(51) Int. CI.:
G10L 17/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **04030909 .8**
(96) Fecha de presentación: **28.12.2004**
(97) Número de publicación de la solicitud: **1564722**
(97) Fecha de publicación de la solicitud: **17.08.2005**

(54) Título: **Identificación automática de llamadores telefónicos en base a las características de voz**

(30) Prioridad:
12.02.2004 US 777322

(45) Fecha de publicación de la mención BOPI:
26.11.2012

(45) Fecha de la publicación del folleto de la patente:
26.11.2012

(73) Titular/es:
MICROSOFT CORPORATION (100.0%)
ONE MICROSOFT WAY
REDMOND, WA 98052, US

(72) Inventor/es:
PASCOVICI, ANDREI

(74) Agente/Representante:
CARPINTERO LÓPEZ, Mario

DESCRIPCIÓN

Identificación automática de llamadores telefónicos en base a las características de voz

Campo de la invención

La presente invención se refiere a un procedimiento y a un aparato implementado por computadora para identificar automáticamente a los llamadores de las llamadas telefónicas entrantes en base a las características de voz. En particular, la presente invención se refiere a unas técnicas de reconocimiento del habla computerizadas para encaminar y filtrar las llamadas telefónicas entrantes.

Antecedentes de la invención

En los sistemas de comunicaciones telefónicas a menudo son utilizados unos centros de llamadas para encaminar o seleccionar con carácter previo las llamadas en base a las respuestas del llamador a las indicaciones automatizadas. Dichos mecanismos de respuesta a las indicaciones a menudo son retardatarias, dado que el llamador debe navegar a través de un gran número de invitaciones antes de ser encaminado hasta el destinatario de la llamada o hasta la base de datos de información deseados. Así mismo, dichos mecanismos se basan en que el llamador sigue adecuadamente los comandos de invitación. Si el llamador no coopera con los comandos de invitación la llamada no puede ser encaminada de manera precisa. De modo similar, los mecanismos de selección de las llamadas se basan en la cooperación por parte del llamador para que responda sinceramente a las invitaciones de filtración. Esto hace difícil que el llamador y el destinatario encaminen y filtren las llamadas de una manera precisa y eficiente.

Por tanto, los sistemas de reconocimiento del habla, han sido propuestos para contribuir al proceso de encaminamiento de las llamadas. Sin embargo, dichos sistemas de reconocimiento del habla se han basado así mismo en un mecanismo de respuesta a la invitación en el cual el llamador debe responder a unas invitaciones predeterminadas. Por ejemplo, el sistema puede solicitar que el llamador declare el nombre del llamador y / o declare una palabra o secuencia de palabras predeterminada que represente la materia objeto de la llamada o la identidad del destinatario deseado. También aquí, estos sistemas son eficaces únicamente si el llamador es sincero al responder a las invitaciones predeterminadas. Así mismo, los modelos de reconocimiento del habla que son utilizados para determinar el contenido del habla deben ser capaces de segmentar con precisión el contenido, dado el amplio margen de las características de entrada de voz para diferentes llamador s. Dichos sistemas pueden, por tanto, seguir siendo retardatarios o imprecisos y pueden fácilmente ser eludidos por llamadores que no cooperen.

Por tanto, lo que se necesita son unos procedimientos y aparatos mejorados para el prefiltrado y encaminamiento automáticos de las llamadas telefónicas entrantes en base a las características de voz.

El trabajo de Rosenberg, Aaron E. et al.: “Utilización de Carpetas de Mensajes de Correo Electrónico de Voz por Parte de un Llamador que Utiliza un Reconocimiento del Hablante Independiente del Texto” [“Foldering Voice-mail Messages by Caller Using Text Independent Speaker Recognition”,] Proceedings of International Conference of Speech and Language Processing, ICSLP 2000, vol. 2, páginas 474 a 477, se refiere a la utilización de carpetas de mensajes de correo electrónico de voz por el llamador que utiliza un reconocimiento del hablante independiente del texto. El discurso de un mensaje entrante es procesado y puntuado con respecto a los modelos llamador s. Un mensaje cuya puntuación coincidente sobrepase un umbral es archivado en la carpeta de llamadores coincidentes; de no ser así, es etiquetado como “desconocido”. El abonado tiene la capacidad de escuchar un mensaje “desconocido” y archivarlo en la carpeta apropiada, si existe, o crear una nueva carpeta, si no existe. Dichos mensajes etiquetados por el abonado se utilizan para entrenar y adaptar los modelos llamador s.

El trabajo de Carex, M.J. et al.: “Un sistema de verificación del hablante utilizando “alpha - nets” [“A speaker verification system using alpha-nets”], Speech Processing 2, VLSI, International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, US, vol. 2, Conf. 16, páginas 397 a 400, se refiere a un sistema de verificación del hablante que utiliza “alpha - nets”. En la forma más simple de un sistema de verificación se propone que hay dos modelos globales de palabras. Un modelo se deriva de las expresiones del hablante que va a ser verificado. El segundo modelo se deriva de las expresiones de una muestra de la población en general. La puntuación de la prueba de verificación para cada modelo es generada de salida como una medida de la probabilidad de registro a partir de una búsqueda Viterbi sincrónica de tramas. La diferencia entre las puntuaciones para los modelos personal y general es a continuación computada y comparada con un umbral. Si la diferencia sobrepasa el umbral, la persona es aceptada.

El trabajo de Leggetter, Woodland: “Adaptación del Hablante de los HMMs de Densidad Continua Utilizando una Regresión Lineal Multivariada” [“Speaeker Adaptation of Continuous Density HMMs Using Multivariate Linear Regression”, ICSLP 94: 1994 International Conference on Spoken Language Processing, Yokohama, Japón, 18 a 22 de septiembre de 1994 utiliza un conjunto inicial de modelos independientes de hablante satisfactorios y adapta los parámetros modales a un nuevo hablante mediante la transformación de los parámetros medios de los modelos con un conjunto de transformadas lineales. Las transformaciones se encuentran utilizando unos criterios de probabilidad máxima los cuales son implementados de manera similar a los algoritmos de entrenamiento de ML estándar para los HMMs.

Sumario de la invención

Constituye el objetivo de la presente invención reducir el tiempo de entrenamiento de un nuevo modelo acústico de un nuevo llamador.

Este objetivo lo consigue la invención como se reivindica en las reivindicaciones independientes. Formas de realización preferentes se definen en las reivindicaciones dependientes.

Una forma de realización de la presente invención se refiere a un procedimiento de identificación de un llamador de una llamada del llamador a un destinatario. Una entrada de voz es recibida del llamador , y la entrada de voz es dividida en subsecciones. Las características de cada subsección son aplicadas a una pluralidad de modelos acústicos, los cuales comprenden un modelo acústico genérico y modelos acústicos de cualquier llamador identificado con anterioridad, para obtener una pluralidad de puntuaciones acústicas que representan la medida en que las características de cada subsección coinciden con los respectivos modelos acústicos. Para cada subsección, se identifica el modelo acústico que presenta la mejor puntuación acústica de la subsección. El llamador es identificado como uno de los llamadores identificado con anterioridad solo si las mejores puntuaciones acústicas de todas las subsecciones se corresponden con el mismo llamador identificado con anterioridad. En otro caso, el llamador es identificado como un nuevo llamador . Si el llamador es identificado como un nuevo llamador, se genera un nuevo modelo acústico para el nuevo llamador , que es específico del nuevo llamador .

Otro forma de realización de la presente invención se refiere a un sistema para la identificación de un llamador de una llamada del llamador a un destinatario. El sistema incluye un destinatario para la recepción de una entrada de voz del llamador y un repositorio de modelos acústicos para el almacenamiento de una pluralidad de modelos acústicos. La pluralidad de los modelos acústicos incluye un modelo acústico genérico y unos modelos acústicos de cualquier llamador identificado con anterioridad. El sistema incluye así mismo unos medios para la división de la entrada de voz en subsecciones. Además de ello, el sistema comprende unos medios para la aplicación de las características de cada subsección a una pluralidad de modelos acústicos para obtener una pluralidad de puntuaciones acústicas respectivas que representen la medida en que las características de cada subsección coinciden con los respectivos modelos acústicos. Así mismo, el sistema incluye, para cada subsección, unos medios para la identificación del modelo acústico que presente la mejor puntuación acústica de esa subsección y unos medios para la identificación del llamador como uno de los llamadores identificados con anterioridad solo si las mejores puntuaciones acústicas de todas las subsecciones se corresponden con el mismo llamador identificado con anterioridad. Así mismo, en otro caso existen unos medios para la identificación del llamador como un nuevo llamador. Un medio generador de llamador acústico genera un nuevo modelo acústico para el nuevo llamador si el llamador es identificado como un nuevo llamador.

Otra forma de realización de la presente invención se refiere a un medio legible por computadora que comprende unas instrucciones ejecutables por computadora las cuales, cuando son ejecutadas por una computadora, llevan a cabo un procedimiento de identificación de un llamador de una llamada. El procedimiento incluye la recepción de una entrada de voz procedente del llamador y la división de la entrada de voz en subsecciones. Las características de cada subsección son aplicadas a una pluralidad de modelos acústicos, la cual comprende un modelo acústico genérico y unos modelos acústicos de cualquier llamador identificado con anterioridad, para obtener una pluralidad de puntuaciones acústicas respectivas que representen la medida en que las características de cada subsección coinciden con los respectivos modelos acústicos. Para cada subsección, se identifica el modelo acústico que presenta la mejor puntuación acústica de la subsección. El llamador es identificado como uno de los llamadores identificados con anterioridad solo si las mejores puntuaciones acústicas de todas las subsecciones se corresponden con el mismo llamador identificado con anterioridad. Sin embargo, en otro caso, el llamador es identificado como un nuevo llamador . Si el llamador es identificado como un nuevo llamador, se genera un nuevo modelo acústico para el nuevo llamador que es específico para el nuevo llamador.

Breve descripción de los dibujos

La FIG. 1 es un diagrama de un sistema ejemplar para la implementación de la invención mediante una computadora personal convencional, de acuerdo con una forma de realización de la invención.

La FIG. 2 es un diagrama de bloques más detallado de un sistema de módulos para la identificación de un llamador , de acuerdo con una forma de realización de la presente invención.

La FIG. 3 es un diagrama de forma de onda que ilustra una entrada acústica “WAV” recibida de un llamador como una función de tiempo.

La FIG. 4 es un diagrama que ilustra un conjunto vectores característicos generados para la entrada acústica mostrada en la FIG. 3.

La FIG. 5 es un diagrama de estados que ilustra un modelo “Markov” oculto (HMM) básico para una unidad de habla.

La FIG. 6 es un diagrama que ilustra un ejemplo de un modelo de lenguaje simplificado el cual puede ser utilizado en una forma de realización de la presente invención.

La FIG. 7 es un diagrama de flujo que ilustra un proceso implementado por computadora para la identificación de llamadores de llamadas telefónicas a un destinatario, de acuerdo con una forma de realización de la presente invención.

La FIG. 8 es un diagrama de flujo que ilustra la detección de un nuevo llamador o de un llamador identificado con anterioridad dentro del proceso mostrado en la FIG. 7 de acuerdo con una forma de realización de la presente invención.

La FIG. 9 es un diagrama de flujo que ilustra un proceso para el entrenamiento de un modelo de lenguaje específico para detectar un llamador por el contenido de la llamada, de acuerdo con una forma de realización de la presente invención.

Descripción detallada de formas de realización ilustrativas

La FIG. 1 y el análisis relacionado están concebidos para proporcionar una descripción general breve de un entorno informático en el cual la información puede ser implementada. Aunque no se requiere, la invención se describirá, al menos en parte, en el contexto general de las instrucciones ejecutables por computadora, como por ejemplo módulos de programa que son ejecutados por una computadora personal u otro dispositivo informático. En términos generales, los módulos de programa incluyen programas de rutina, objetos, componentes, estructuras de datos, etc. que llevan a cabo tareas específicas o implementan tipos de datos abstractos específicos. Así mismo, los expertos en la materia apreciarán que la invención puede llevarse a la práctica con otras configuraciones de sistemas informáticos, incluyendo dispositivos portátiles, sistemas de multiprocesadores, sistemas electrónicos de consumidor programables o basados en microprocesador, PCs de red, minicomputadoras, computadoras centrales, y similares. La invención puede, así mismo, llevarse a la práctica en entornos informáticos distribuidos en los que los cometidos se llevan a cabo mediante dispositivos de procesamiento a distancia que están unidos mediante una red de comunicaciones. En un entorno informático distribuido, los módulos de programa pueden ser localizados en dispositivos de almacenamientos de memoria tanto locales como remotos.

Con referencia a la FIG. 1, un sistema ejemplar para la implementación de la invención incluye un dispositivo informático de propósito general consistente en una computadora personal 20 que incluye una unidad de procesamiento (UPC) 21, una memoria 22 del sistema, y un bus 23 del sistema que acopla diversos componentes del sistema incluyendo la memoria 22 del sistema y la unidad de procesamiento 21. El bus 23 del sistema puede ser cualquiera de los diversos tipos de estructuras de bus que incluyan un bus de memoria o un controlador de memoria, un bus periférico y un bus local que utilicen cualquiera de las distintas arquitecturas de bus. La memoria 22 del sistema incluye una memoria de solo lectura (ROM) 24 y una memoria de acceso aleatorio (RAM) 25. Un sistema de entrada / salida (BIOS) 26 que contiene la rutina básica que ayuda a transferir información entre los elementos dispuestos dentro de la computadora personal 20, como por ejemplo durante el inicio, es almacenada en la ROM 24. La computadora personal 20 incluye así mismo una unidad 27 de disco duro para la lectura a partir de y la escritura sobre un disco duro (no mostrado), una unidad 28 de disco magnético para la lectura a partir de la escritura sobre un disco magnético extraíble 29, una unidad 30 de disco óptico para la lectura a partir de la escritura sobre un disco óptico extraíble 31, como por ejemplo un CD ROM u otros medios ópticos. La unidad 27 de disco duro, la unidad 28 de disco magnético y la unidad 30 de disco óptico están conectadas al bus 23 del sistema mediante una interfaz 32 con la unidad de disco duro, una interfaz 33 con la unidad de disco magnético y una interfaz 34 con la unidad óptica, respectivamente. Las unidades y los medios legibles por computadora asociados proporcionan un almacenamiento no volátil de las instrucciones legibles por computadora de las estructuras de datos, de los módulos de programa y de otros datos destinados a la computadora personal 20.

Aunque el entorno ejemplar descrito en la presente memoria emplea el disco duro, el disco magnético extraíble 29 y el disco óptico extraíble 31, debe apreciarse por parte de los expertos en la materia que pueden, así mismo, ser utilizados en el entorno operativo ejemplar otros tipos de medios legibles por computadora, los cuales pueden almacenar datos a los que se pueda acceder por una computadora, como por ejemplo casetes magnéticos, tarjetas de memoria “flash”, discos de video digitales, cartuchos Bernoulli, memorias de acceso aleatorio (RAMs), memoria de solo lectura (ROM), y similares.

Una pluralidad de módulos de programa puede ser almacenada en el disco duro, en el disco magnético 29, en el disco óptico 31, en la ROM 24 o en la RAM 25, incluyendo un sistema operativo 35, uno o más programas 36 de aplicación, otros módulos de programa 37 y unos datos de programa 38. Un usuario puede introducir comandos e información en la computadora personal 20 a través de dispositivos de entrada locales, como por ejemplo un teclado 40, un dispositivo señalador 42 y un micrófono 43. Otros dispositivos de entrada (no mostrados), pueden incluir una palanca de mando, un control de mando para juegos, una antena parabólica, un escáner, o similares. Estos y otros dispositivos de entrada están a menudo conectados a la unidad de procesamiento 21 mediante una interfaz de puerto serie 46 que está acoplada al bus 23 del sistema, pero puede estar conectada a otras interfaces, como por ejemplo una tarjeta de sonido, un puerto paralelo, un puerto de juegos o un bus serie universal (USB). Un monitor 47 u otro tipo de dispositivo de visualización está, así mismo, conectado al bus 23 del sistema por medio de una

interfaz, como por ejemplo un adaptador 48 de vídeo. Además del monitor 47, las computadoras personales pueden típicamente incluir otros dispositivos de salida periféricos, como por ejemplo un altavoz 45 y unas impresoras (no mostradas).

La computadora personal 20 puede operar en un entorno de conexión a red utilizando conexiones lógicas con una o más computadoras distantes, como por ejemplo una computadora distante 49. La computadora distante 49 puede ser otra computadora personal, un dispositivo portátil, un servidor, un encaminador, un PC de red, un dispositivo homólogo u otro nodo de red, y típicamente incluye muchos o todos los elementos descritos con anterioridad relacionados con la computadora personal 20, aunque solo se ha ilustrado un dispositivo 50 de almacenamiento de memoria en la FIG. 1. Las conexiones lógicas mostradas en la FIG. 1 incluyen una red de área local (LAN) 51 y una red de área extensa (WAN) 52. Dichos entornos de conexión en red son de uso común en oficinas, redes internas ("Intranets") de ámbito corporativo, e Internet.

Cuando se utiliza en un entorno de conexión en una red de LAN, la computadora personal 20 está conectada a la red de área local 51 por medio de una interfaz o adaptador 53 de red. Cuando se utiliza un entorno de conexión en una red WAN la computadora personal 20 típicamente incluye un módem 54 u otro medio para el establecimiento de comunicaciones a través de la red de área extensa 52, como por ejemplo Internet. El módem 54, el cual puede ser interno o externo, está conectado al bus 23 del sistema por medio de las interfaces de puerto serie 46. En un entorno de conexión en red, unos módulos de programa representados con respecto a la computadora personal 20 o partes de estos, pueden ser almacenados en los dispositivos de almacenamiento de memoria distantes. Debe apreciarse que las conexiones de red mostradas son ejemplares y que pueden ser utilizados otros medios de establecimiento de un enlace de comunicaciones entre las computadoras. Por ejemplo, un enlace de comunicaciones inalámbricas puede establecerse entre una o más partes de la red.

Aunque la FIG. 1 muestra un entorno ejemplar, la presente invención no está limitada a un entorno informático digital. En particular, la presente invención puede ser operada en dispositivos analógicos o en dispositivos de señal mixtos (analógicos y digitales). Así mismo, la presente invención puede ser implementada en un único circuito integrado, por ejemplo. Los módulos pueden ser implementados en hardware o en software o en una combinación de hardware y software.

De acuerdo con lo analizado con anterioridad, la computadora 20 típicamente incluye una diversidad de medios legibles por computadora. Los medios legibles por computadora pueden ser cualquier medio disponible al que se pueda acceder mediante la computadora 20 e incluye medios tanto volátiles como no volátiles, medios extraíbles y no extraíbles. A modo de ejemplo, y sin limitación, los medios legibles por computadora pueden comprender unos medios de almacenamiento en computadora o unos medios de comunicación. Los medios de almacenamiento en computadora incluyen unos medios volátiles y no volátiles, extraíbles y no extraíbles, implementados en cualquier procedimiento o técnica para el almacenamiento de información, como por ejemplo instrucciones legibles por computadora, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento en computadora incluyen, pero no se limitan a, una RAM, una ROM, una EEPROM, una memoria "flash" u otros sistemas técnicos de memoria, un CD-ROM, discos versátiles digitales (DVD) u otros dispositivos de almacenamiento de discos ópticos, casetes magnéticas, cinta magnética, dispositivos de almacenamiento de discos magnéticos u otros dispositivos de almacenamiento magnéticos u otros medios que puedan ser utilizados para almacenar la información deseada y a los cuales se pueda acceder mediante la computadora 20. Los medios de comunicación típicamente incorporan unas instrucciones legibles por computadora, unas estructuras de datos, unos módulos de programa u otros datos en una señal de datos modulada, como por ejemplo una onda portadora u otro mecanismo de transporte e incluyen cualquier medio de suministro de información. El término "señal de datos modulada" significa una señal que presenta una o más de sus características fijadas o modificadas de tal manera que codifiquen información en la señal. A modo de ejemplo, y no de limitación, los medios de comunicación incluyen unos medios por cable, como por ejemplo una red cableada o una conexión de cableado directo, y unos medios inalámbricos como por ejemplo acústicos, de RF, de infrarrojos y otros medios inalámbricos. Combinaciones de cualquiera de los medios expuestos en las líneas anteriores deben, así mismo, quedar incluidos dentro del alcance de los medios legibles por computadora.

La FIG. 2 ofrece un diagrama de bloques más detallado de un sistema de módulos 100 que puede ser implementado dentro del entorno general descrito con referencia a la FIG. 1 para la identificación de un llamador de acuerdo con una forma de realización de la presente invención. El sistema 100 incluye un receptor 102 para la recepción de una señal de habla de entrada de una llamada desde un llamador hasta un destinatario. La señal de habla de entrada puede ser cualquier forma de una señal analógica o de una señal digital. La señal de habla de entrada puede ser transmitida al receptor 102 mediante cualquier procedimiento de comunicación a través de cualquier medio de transmisión. El "destinatario" puede ser cualquier persona concreta, un grupo de individuos, una localización de encaminamiento de llamadas o una base de datos de información, por ejemplo.

El receptor 102 puede incluir cualquier receptor apropiado para la recepción del tipo de señal de entrada de hablaque está siendo transmitido. Por ejemplo, con la llegada de las computadoras personales habilitadas telefónicamente (PCs) y las PCs de Bolsillo con Teléfono Incorporado, el receptor 102 puede incluir un adaptador 53 de red para su acoplamiento a la LAN 51 o una interfaz 46 de puerto serie para su acoplamiento al módem 54 y a la WAN 52.

Si la señal de habla de entrada es una señal analógica, el sistema 100 incluye un convertidor analógico / digital (A / D) 104 para la conversión de la señal a una serie de valores digitales. En una forma de realización, el convertidor de A / D 104 muestrea la señal analógica a 16kHz, creando de esta manera 16 kilobits de datos de voz por segundo. Sin embargo, puede ser utilizada cualquier otra velocidad de muestreo.

Las señales digitales que representan muestras de la señal de habla de entrada son suministradas a la computadora 20. La computadora 20 incluye un módulo de extracción 106 de características, un módulo de reconocimiento del habla (por ejemplo, un descodificador) 107, un módulo 108 de entrenamiento, un módulo de léxico 105, un repositorio 110 de modelos de lenguaje, un repositorio 111 de modelos acústicos, un módulo 112 de identificación de llamador s, un encaminador 113 de llamadas y un módulo 114 de invitación de respuestas. Los elementos de la computadora 20 están acoplados al dispositivo de salida 115 y al dispositivo E / S 116, por ejemplo.

Debe destacarse que el entero sistema 100 o parte del sistema 100 puede ser implementado en el entorno ilustrado en la FIG. 1. El módulo 106 de extracción de características y el módulo 108 de entrenamiento pueden ser, o bien módulos de hardware dispuestos dentro de la computadora 20 o módulos de hardware almacenados en cualquier dispositivo de almacenamiento de información divulgado en la FIG. 1 y a los que se pueda acceder a través de la UCP 21 u otro procesador apropiado. Así mismo, el módulo 105 de almacenamiento de léxico, los módulos acústicos 111 y los modelos de lenguaje 110 están, así mismo, almacenados, de modo preferente, en cualquiera de los dispositivos de memoria apropiados mostrados en la FIG. 1. Así mismo, el motor de búsqueda 107 puede ser implementado en la UCP 21, la cual puede incluir uno o más procesadores o puede ser puesto en práctica mediante un procesador de reconocimiento del habla dedicado empleado por la computadora personal 20. Así mismo, el dispositivo de salida 112 y el dispositivo de S / E 113 pueden incluir cualquiera de los dispositivos de E / S mostrados en la FIG. 1, como por ejemplo el teclado 40, el dispositivo de señalización 43, el monitor 47, una impresora o cualquier dispositivo de memoria mostrado en la FIG. 1, por ejemplo.

Las señales digitales recibidas por el receptor 102 o generadas por el convertidor de A / D 104 son suministradas al módulo 106 de extracción de características. En una forma de realización, el módulo 106 de extracción de características incluye un procesador matricial convencional, el cual lleva a cabo un análisis espectral sobre las señales digitales y computa un valor de magnitud para cada banda de frecuencias de un espectro de frecuencias.

El módulo 106 de extracción de características divide las señales digitales en tramas, cada una de las cuales incluye una pluralidad de muestras digitales. En una forma de realización, cada trama tiene una duración, de manera aproximada, de 10 milisegundos. Las tramas son, a continuación, codificadas en un vector de características que refleja las características espectrales de una pluralidad de bandas de frecuencias. En el caso del modelado de Markov oculto discreto y semicontinuo, el modelo 106 de extracción de características, codifica, así mismo, los vectores de características en una o más palabras clave utilizando técnicas de cuantificación de los vectores y un libro de códigos derivado de los datos de entrenamiento. De esta manera, el módulo 106 de extracción de características proporciona, en su salida, los vectores de características (o palabras clave) para cada expresión emitida. El módulo 106 de extracción de características proporciona, de modo preferente, los vectores de características a una velocidad aproximada de un vector de características cada 10 milisegundos, por ejemplo.

Ejemplos de módulos de extracción de características incluyen los módulos para llevar a cabo una Codificación Predictiva Lineal (LPC), un cepstro derivado de la LPC, una Predicción Lineal Perceptiva (PLP), una Extracción de Características para Modelos de Auditoría, y una extracción de características de Coeficientes de Cepstro de Frecuencias Mel (MFCC). Nótese que la presente invención no está limitada a estos módulos de extracción de características y que pueden ser utilizados otros módulos dentro del contexto de la presente invención.

El flujo de vectores de características producido por el módulo 106 de extracción de características es suministrado al módulo de reconocimiento de voz 107, el cual identifica una secuencia más probable de unidades de habla , como por ejemplo palabras o fonemas, en base al flujo de los vectores de características, en base a uno o más modelos acústicos existentes en el repositorio 111, a uno o más modelos de lenguaje existentes en el repositorio 110 y en base al léxico 105. El módulo 112 de identificación del llamador identifica al llamador como un nuevo llamador o a uno cualquiera de los llamadores identificados con anterioridad, mediante la aplicación de los vectores de características de la entrada de voz a los modelos genéricos y específicos del llamador de las unidades de habla identificadas por el módulo 107 de reconocimiento del habla, las cuales están almacenadas en el repositorio 111. En una forma de realización, el módulo 112 de identificación del llamador utiliza, así mismo, unos modelos de lenguaje genéricos o específicos del llamador, almacenados en el repositorio 110, para contribuir a la identificación. El módulo 112 genera de salida la identidad del llamador y / o el texto de la secuencia más probable de palabras expresadas hacia el encaminador 113 de la llamada o almacena estos resultados dentro de uno de los dispositivos de memoria mostrados en la FIG. 1, por ejemplo. Los resultados pueden, así mismo, ser generados de salida hasta un usuario o un operador a través del dispositivo 115 de E / S. El encaminador 113 de la llamada puede entonces filtrar la llamada o encaminar la llamada hacia uno o más destinos seleccionados en base a la identidad del llamador y / o al contenido de la llamada.

Un modelo acústico es un modelo que indica hasta que punto es probable que una secuencia de vectores de características sea producida mediante una secuencia concreta de unidades acústicas encontradas en una secuencia de unidades de habla hipotéticas. Con arreglo a algunas formas de realización de la presente invención,

cada unidad de habla puede incluir cualquier unidad acústica habitualmente utilizada, como por ejemplo un senono, un fonema, un difono, una sílaba o una palabra. En algunas formas de realización, cada unidad de habla es una combinación de un conjunto de subunidades.

Tal y como se indicó con anterioridad, el repositorio 111 de modelos acústicos incluye al menos un modelo acústico para cada llamador identificado con anterioridad y un modelo genérico que representa las características de voz de una amplia gama de hablantes. Cada modelo acústico incluye un conjunto de modelos, como por ejemplo los modelos de Markov ocultos (HMMs), de una pluralidad de unidades de habla predefinidas que van a ser detectadas. Por ejemplo, cada HMM puede modelar un solo fonema. En una forma de realización, el módulo 107 de reconocimiento de voz aplica los vectores de características recibidos desde el módulo 106 de extracción de características al modelo acústico genérico para determinar un fonema más probable que represente los vectores de características y, por tanto, representa la expresión recibida del llamado.

Un modelo acústico típico es entrenado antes de que sea utilizado para descodificar una secuencia de vectores de características de entrada. Por ejemplo, en la FIG. 2 dicho entrenamiento puede llevarse a cabo por un entrenador 108 en base a un texto de entrenamiento 118, pasando por los parámetros del modelo a partir del modelo acústico y en base al entrenamiento de los vectores de características a partir del extractor 106 de características. En algunas formas de realización de la presente invención, el modelo acústico genérico es entrenado utilizando un texto de entrenamiento genérico representativo de un conjunto genérico de hablantes. Este modelo acústico genérico puede, a continuación, ser utilizado para formar los modelos acústicos específicos del llamador en los cuales los HMMs son actualizados con cada conjunto de vectores de características generado por ese llamador. En una forma de realización, un modelo acústico único puede ser generado para un llamador concreto en base a una sola expresión como por ejemplo una expresión de uno o más fonemas. Cuantas más llamadas y expresiones son recibidas de ese llamador, el correspondiente modelo acústico para ese llamador continúa siendo actualizado.

El motor 107 del módulo de reconocimiento del habla puede, así mismo, acceder a uno o más modelos de lenguaje almacenados en el repositorio 110 para contribuir a la identificación de una palabra o de la secuencia de palabras más probable representada por los datos de entrada. El repositorio 110 puede almacenar un modelo de lenguaje genérico, independiente del llamador y / o una pluralidad de modelos de lenguaje específicos del llamador. En una forma de realización, cada modelo de lenguaje incluye un modelo gramatical carente de texto (CFG) o un modelo estadístico de n-gram, como por ejemplo un trigrama. Un modelo de trigrama determina la probabilidad de una secuencia de palabras en base a las probabilidades combinadas de unos segmentos de 3 palabras de la secuencia. Dicho modelo de lenguaje puede ser modificado para proporcionar un modelo único para cada llamador identificado con anterioridad, de acuerdo con lo analizado con mayor detalle más adelante. Los modelos de lenguaje específicos del llamador pueden ser utilizados para ayudar a una computadora 20 a la identificación de las palabras o de la materia objeto habitualmente utilizada por un llamador específico.

El modelo de lenguaje genérico puede incluir un modelo de lenguaje de trigrama de 60.000 palabras, por ejemplo; modelo derivado de la publicación North American Business News y desarrollado con mayor detalle en la publicación titulada “Modelo de Lenguaje de Texto CSR - III” [“CSR - III Text Lenguage Model”] Universidad de Pensilvania, 1994.

Las FIGS. 3 a 5 ilustran la formación de un conjunto de vectores de características y los detalles de un modelo Markov oculto, el cual puede ser utilizado de acuerdo con una forma de realización de la presente invención. La FIG. 3 es un diagrama de forma de onda que ilustra un “WAV” de entrada acústica recibido del llamador como función del tiempo. Tal y como se indicó con anterioridad, cada entrada acústica está dividida en una pluralidad de tramas de 10 milisegundos, por ejemplo. El módulo 106 de extracción de características genera un conjunto de vectores de características O [k] para cada trama de 10 milisegundos, para k = 1, 2, …, tal y como se muestra en la FIG. 4. Los vectores de características O [k] son en la mayoría de las ocasiones alguna transformación de la Transformada Rápida de de Fourier (FFT) del WAV de entrada acústica, tabicado en ranuras de 10 milisegundos. Los coeficientes de la FFT reflejan las características del habla, como por ejemplo el tono y la cavidad bocal del hablante. Estos vectores de características pueden, a continuación, ser aplicados a los modelos Markov ocultos del respectivo modelo acústico.

La FIG. 5 es un diagrama de estados que ilustra un modelo Markov oculto (HMM) básico para una unidad de habla (por ejemplo un fonema, un senono, un trifono, etc). Un modelo HMM básico es una unidad independiente del lenguaje, la cual representa las unidades acústicas de una unidad de habla. Cada estado puede, o bien permanecer en el presente estado o pasar al siguiente estado dentro del modelo. Cada unidad de habla presenta tres estados, designados como S1, S2 y S3 en la FIG. 5, los cuales representan un estado de “inicio”, un estado “principal” y un estado “final” de la unidad de habla. Cada estado puede permanecer solo en ese estado o pasar al siguiente estado a lo largo de las flechas mostradas en la FIG. 5. La transición de un estado al siguiente presenta una probabilidad P (S2 | S1), la cual representa la probabilidad condicional de la transición del estado S1 al estado S2 teniendo en cuenta el presente estado S1. Cada estado, así mismo, presenta una distribución B [i] de probabilidades, para i = 1 a 3, la cual representa una probabilidad para generar de salida (un número entre 0 y 1) de cualquier vector 0 [k] de características, el cual refleja la probabilidad de observación de cualquiera de los vectores de características posibles. Por ejemplo, las distribuciones de la probabilidad pueden ser distribuciones gaussianas.

Cada modelo acústico existente en el repositorio 111 incluye una colección de dichos modelos Markov ocultos para cada Fonema. Por ejemplo, el Fonema “AX” precedido por el fonema “B” y seguido por el fonema “H” (anotación B - AX + H, como en “bah”) es diferente del mismo exacto “AX” precedido por “L” y seguido por “H” (anotación L - AX + H, como en la última parte de “blah”).

A llevar a cabo el reconocimiento del habla utilizando un modelo acústico existente, el modelo inicial del sistema es S1, con la probabilidad 1, y las probabilidades P [i | j] y las densidades de probabilidad B [i] son conocidas para cada estado del modelo HMM. Al reconocer un fonema, la entrada acústica es convertida en una secuencia de vectores de características o[k], y el reconocedor 107 del habla (mostrado en la FIG. 2) determina qué probabilidad P (O [k] | modelo) se concede al actual modelo HHM.

En otras palabras, el modelo 107 de reconocimiento del habla determina cuál es la probabilidad de que los sonidos representados por la secuencia de los vectores de características de entrada sean de hecho el fonema modelado por el actual HMM objeto de consideración. El fonema modelado por el HMM que presenta la probabilidad más alta es identificado como el fonema expresado.

Cuando se entrena un modelo acústico, como por ejemplo cuando se entrena el modelo genérico o se actualiza un modelo específico de un llamador, se supone que son conocidos los archivos WAV de entrada acústica y, por tanto, la secuencia de los vectores de las características O [k]. el módulo 107 de reconocimiento del habla (o el módulo 112 de identificación del llamador) genera un modelo (P’ [i | j] y B’ [i] para cada estado) que proporciona la probabilidad más alta de observar la secuencia de salida O [k] para cada fonema. Por ejemplo, en una forma de realización, el módulo 112 de identificación del llamador utiliza un procedimiento de reestimación del HMM de Baum - Welch para la actualización o, de no ser así, la adaptación del modelo acústico genérico para reflejar las características de un hablante concreto. Por ejemplo, el modelo acústico para un hablante concreto puede inicialmente incluir los modelos HMM genéricos del modelo acústico genérico y, a continuación, los modelos HMM para los fonemas que se producen en la presente llamada pueden ser actualizados para reflejar las características del habla del llamador mediante el procedimiento de reestimación HMM de Baum - Welch.

La FIG. 6 es un diagrama que ilustra un ejemplo de un modelo de lenguaje simplificado, el cual puede ser utilizado en una forma de realización de la presente invención. En el momento de generar un modelo de lenguaje de trigrama, la primera etapa consiste en recopilar un amplio cuerpo de texto del lenguaje representativo. La segunda etapa consiste en generar unos conteos de frecuencia P1 [W] para cada palabra W, P2 [W |W0] para cada bigrama (par de palabras), y P3 [W | W1, W2] para cada trigrama. El módulo de reconocimiento de voz estará, así mismo, limitado por un diccionario de palabras (WD) que presenta una lista posible de palabras en el respectivo lenguaje. A continuación, se utiliza una estrategia de descuento para generar una probabilidad P [W | W1, W0] para cada palabra del diccionario de palabras. La estrategia de descuento se utiliza para evitar la utilización de todas las secuencias posibles de dos o tres palabras, dado que su número es demasiado grande. Todas las palabras del diccionario de palabras se descomponen en fonemas, los cuales se caracterizan por un HMM de fonemas similar al mostrado en la FIG. 5.

A continuación, se crea un HMM maestro adhiriendo entre sí los HMMs de fonemas y ajustando la probabilidad inicial para introducir cada uno de sus estados iniciales (S1) de acuerdo con la P [W | W1, w0] a partir del modelo HMM. En el ejemplo simplificado mostrado en la FIG. 6, solo dos palabras, “en” y “el” se han observado, y cada una de estas palabras ha sido observada solo una vez. Por tanto, el HMM maestro ofrecerá un estado de inicio distinto S0 con una probabilidad inicial de “1” y unas probabilidades de transición de “0,5” al HMM fonema “AX” y al HMM del fonema “TH”. Dado que hay solo dos palabras, el HMM “AX” tiene una probabilidad de transición de 1,0 al HMM “T”, y el HMM “TH” presenta una probabilidad de transición de 1,0 al HMM “EH”. El HMM “TH” y la transición del HMM “EH” al estado final S3.

En una forma de realización de la presente invención, se crea un único modelo de lenguaje para cada llamador identificado como único mediante la adaptación de un modelo de lenguaje genérico. El modelo de lenguaje genérico es adaptado mediante la utilización de frases reconocidas junto con el “amplio cuerpo de texto” recopilado a partir de las llamadas de entrada del llamador . Este proceso no descubre nuevas palabras, sino más bien nuevas probabilidades P [W | W1, W0] dado que cualquier llamador específico es probable que utilice algunas combinaciones de palabras más que otras. Así mismo, no es necesario recopilar “palabras “en el sentido tradicional, en todas las formas de realización. En una forma de realización, el modelo de lenguaje recopila “palabras” similares al reconocedor MS disponible en Microsoft Corporation con su modelo de lenguaje de dictado. En formas de realización alternativas, el modelo de lenguaje puede simplemente recopilar “fonemas” o pequeños grupos de fonemas como “palabras”, de manera similar al reconocedor MS con su modelo de pronunciación del lenguaje. Las últimas formas de realización presentan la ventaja de proporcionar unas probabilidades útiles de secuencias de fonemas incluso cuando el llamador emite palabras desconocidas, como por ejemplo nombres, pero son menos precisas.

La FIG. 7 es un diagrama de flujo que ilustra un proceso 200 implementado en una computadora que puede ser almacenado en forma de instrucciones, por ejemplo, en un medio legible por computadora y ejecutado por la computadora 20 (mostrada en la FIG. 1). El proceso 200 identifica los llamadores de las llamadas telefónicas a uno o

más destinatarios mediante la generación de unos modelos acústicos para cada llamador identificado, de acuerdo con una forma de realización de la presente invención.

En la etapa 201, una llamada entrante es recibida de un llamador. En la etapa 202, se utiliza un sistema tradicional del ID llamador para captar el número de teléfono de la llamada entrante. Si el número de teléfono coincide con el de un número de teléfono de confianza definido con anterioridad, el sistema genera de salida una señal que indica que ha sido detectado un número de teléfono de confianza, en la etapa 203. El sistema de ID al llamador puede ser utilizado para la identificación de las llamadas que lleguen de una fuente de confianza y para la provisión de una salida temprana del proceso 200 de manera que las llamadas recibidas de esta fuente no experimenten un retraso debido al proceso 200. La señal generada en la etapa 203 puede ser utilizada de cualquier manera pertinente, por ejemplo, mediante el encaminamiento de la llamada entrante hacia un buzón concreto o permitiendo que la llamada sea encaminada hacia el dispositivo telefónico habilitado del destinatario.

Si el número de teléfono entrante no es de confianza o si el sistema no está configurado con la etapa 202, la entrada de voz es aplicada al módulo 107 de reconocimiento del habla (mostrado en la FIG. 2), en la etapa 204 para su segmentación en fonemas conocidos de secuencias. El módulo 106 de extracción de características (también mostrado en la FIG. 2) genera los correspondientes vectores de características desde la entrada de voz y aplica los vectores de características a un módulo genérico gramático carente de contexto (CFG) y al modelo acústico genérico independiente del llamador (denominado “I - AM” en la FIG. 7). El módulo grámatico carente de contexto puede incluir, por ejemplo, un modelo de dictado carente de forma o un modelo estocástico carente de lenguaje. El CFG permite el reconocimiento de cualquier expresión. No es necesario que el CFG genere una forma de texto siempre que produzca una segmentación de fonemas razonablemente precisa.

El modelo acústico genérico independiente del llamador, I - AM, puede incluir un modelo que sea capaz de funcionar para cualquier llamador. Dicho modelo acústico genérico se designa algunas veces como modelo acústico “independiente del género”, el cual funciona para llamadores masculinos, femeninos e infantiles.

Utilizando el CFG el modelo acústico genérico independiente del llamador, I - AM, el módulo de reconocimiento del habla segmenta la entrada de voz en una secuencia de fonemas reconocidos.

Por ejemplo, si la entrada de voz incluye “LLAMO AL …”, el módulo de reconocimiento de voz genera la forma de texto de emisión sonora (“LLAMO AL …”) más la segmentación de fonemas (“IX <sil> AXM <sil> CAX LIX NG”).

En la etapa 205 el módulo 112 de identificación del llamador (FIG. 2) determina si el llamador es un nuevo llamador o un llamador identificado con anterioridad. Este proceso se describe con mayor detalle más adelante con respecto a la FIG. 8. Si el llamador es un nuevo llamador , el proceso 200 pasa a la etapa 206 donde el módulo 112 de identificación del llamador añade un nuevo modelo acústico AM [i] al repositorio 111 de modelos acústicos (FIG. 2) e incrementa un número de modelo variable NUMMODELS (esto es, un número de llamadores identificado con anterioridad) en uno. El módulo 112 de identificación del llamador genera el nuevo modelo acústico AM [i] realizando una copia del modelo acústico genérico AM [0] y, a continuación, utilizando el HMM de cualquier fonema emitido por el llamador en la llamada entrante, de acuerdo con lo descrito con anterioridad.

En la etapa 207 el módulo 112 de identificación del llamador genera de salida una señal que indica un “nuevo llamador”, la cual puede ser utilizada por el modulo 113 de encaminamiento de llamadas (también en la FIG. 2) u otro sistema de gestión de las llamadas para dirigir la llamada de la forma que se desee. El módulo 112 de identificación del llamador puede, así mismo, almacenar un archivo sonoro que represente la entrada de voz y el correspondiente texto (si se ha reconocido en la etapa 204).

En una forma de realización, el módulo 112 de identificación del llamador solicita un examen manual del reconocimiento del llamador / texto a través del dispositivo de E / S 115 (FIG. 2), en la etapa 208. El usuario o el operador del sistema puede examinar el texto de la llamada, escuchar el sonido de la llamada y / o visualizar la identificación del llamador y efectuar cualquier corrección a través del dispositivo de E / S 115. Por ejemplo, el usuario puede examinar y descartar llamadas o aceptar o denegar las clasificaciones efectuadas por el módulo de identificación. Después de la etapa 207, el proceso 200 retorna a la etapa 201 para recibir otra llamada entrante.

Si, en la etapa 205, el módulo 112 de identificación del llamador identifica al llamador como un llamador no nuevo, el proceso 200 pasa a la etapa 210 para identificar cuál de los llamadores identificados con anterioridad ha llamado de nuevo. El módulo 112 de identificación del llamador determina el modelo acústico para el llamador concreto que coincida con mayor exactitud con las características del habla de la emisión sonora de la llamada entrante. En una forma de realización, el módulo 112 de identificación del llamador aplica las características de voz (por ejemplo, los vectores de características) al correspondiente HMM de cada modelo acústico específico del llamador e identifica el modelo acústico AM [j] que presenta la mejor puntuación acústica, para j = 1 en los NUMMODELS, de acuerdo con lo descrito con mayor detalle en la FIG. 8. En la etapa 211, el módulo 112 de identificación del llamador genera de salida una señal que indica que “el Llamador j es Detectado”, donde “j” se corresponde con el modelo acústico que presenta la mejor puntuación acústica en la etapa 210.

La FIG. 8 es un diagrama de flujo que ilustra la detección de un nuevo llamador o de un llamador identificado con anterioridad en la etapa 205 de la FIG. 7, de acuerdo con una forma de realización de la presente invención. El

proceso entra en la etapa 205 con la referencia numeral 300. En la etapa 301, el módulo 112 de identificación del llamador determina si el número de modelos acústicos, NUMMODELS, para los llamadores identificados con anterioridad es mayor de 0. Si no es así, el llamador de la actual llamada entrante es un nuevo llamador , y el proceso 205 sale en la etapa 302. Si el número de llamadores identificado con anterioridad es superior a 0 el llamador actual podría ser un nuevo llamador o uno del los llamadores identificados con anterioridad. El proceso, a continuación, pasa a la etapa 303.

En la etapa 303, el módulo 112 de identificación del llamador calcula la puntuación acústica o "alfa" A [j] para la emisión sonora actual con arreglo a cada uno de los modelos acústicos AM [j] existentes en el repositorio 111 de modelos acústicos, para j = 0 de NUMMODELS, en el que el modelo AM [0] es el modelo genérico independiente del llamador, I - AM. Una puntuación alfa es conocida como una "puntuación de paso hacia delante", la cual es la puntuación acústica resultante de la ejecución del descodificador del modelo de reconocimiento del habla o del árbol de búsqueda sobre la segmentación producida en la etapa 204 de la FIG. 7 (mediante el modelo independiente de la llamada, AM [0]) mientras se utiliza el modelo acústico AM [j]).

En la etapa 304, el módulo 112 de identificación del llamador determina si la puntuación alfa A [0] para el modelo acústico genérico AM[0] presenta la puntuación alfa más alta (o, si no, la puntuación óptima). Si la emisión sonora actual coincide con el modelo acústico genérico mejor que cualquiera de los modelos acústicos específicos del llamador entonces el llamador es identificado como un nuevo llamador , y el proceso sale en la etapa 305. Si la puntuación alfa A [0] para el modelo acústico genérico no es la puntuación alfa mayor, entonces el llamador es identificado como uno de los llamadores identificados con anterioridad, y el módulo 112 del identificación del llamador pasa a la etapa 306 para la identificación del llamador concreto.

Suponiendo que la variable "k" igual al índice en el cual la puntuación alfa A[j] es máxima (k = argmax ([j]), entonces el llamador k es identificado como el llamador y el correspondiente modelo acústico específico del llamador AM [k] para el llamador "k" es actualizado para reflejar las características del habla de la nueva emisión sonora. De esta manera, cada vez que es recibida una llamada entrante por un llamador identificado con anterioridad, el correspondiente modelo acústico para ese llamador es entrenado en mayor medida en base a las unidades acústicas contenidas en la llamada para representar mejor las características del habla del llamador. Después de que el modelo acústico específico del llamador es actualizado, el proceso sale en la etapa 307.

En una forma de realización, el modelo acústico específico del llamador AM [k] puede ser creado o actualizado únicamente con una sola expresión, por oposición al entrenamiento requerido mediante un gran número de expresiones y de repeticiones de expresiones como es habitual en el reconocimiento de habla tradicional o en el software de dictado. El entrenamiento de una expresión única puede llevarse a cabo con un software de reconocimiento de voz actualmente disponible, como por ejemplo el reconocedor Microsoft MS, ya sea repitiendo varias veces la entrada de los sonidos y aplicándola de forma reiterada al reconocedor MS o mediante la reconfiguración del reconocedor MS para entrenar con una expresión de señal. Otros tipos de reconocedores del habla o descodificadores pueden, así mismo, ser utilizados en formas de realización alternativas.

En una forma de realización alternativa, la etapa 304 puede ser refinada aún más mediante la división de la entrada de voz actual en varias subsecciones, como por ejemplo dos subsecciones, y computando las dos puntuaciones alfa A0 [j] y A1 [j] para las dos subsecciones con cada modelo acústico. La etapa 304 devolverá un "NO" (los modelos genéricos no presentan la puntuación acústica más alta) solo cuando tanto A0 [j] y como A1 [j] sean máximas (argmax (AM [k])) sobre el mismo índice K. Este proceso puede ser útil para filtrar las llamadas que incorporan más de un hablante en la entrada de voz para refinar en mayor medida el proceso de identificación.

La FIG. 9 es un diagrama de flujo que ilustra un proceso 400 para el entrenamiento de un modelo de lenguaje específico del llamador ("CFG probabilística") para detectar un usuario con el contenido de la llamada (más que por la acústica). El proceso 400 puede ser utilizado en combinación con el proceso 200 mostrado en la FIG. 7 para incrementar la precisión de la identificación del llamador o como un procedimiento alternativo de identificación del llamador . La llamada entrante puede ser recibida en la etapa 401. En la etapa 402, el proceso 400 obtiene una identificación acústica del llamador mediante la ejecución del proceso de identificación acústica del llamador mostrado en la FIG. 7. En la etapa 403 el proceso 400 añade el "texto" reconocido de la llamada (según ha sido segmentado por el módulo 107 de reconocimiento del habla de la FIG. 2) al repositorio de textos del llamador para el correspondiente modelo de lenguaje específico del llamador. La etapa 403 se corresponde con la etapa de "recopilación de un amplio cuerpo de texto" descrito con referencia a la FIG. 6.

En la etapa 404, el proceso 400 determina si hay palabras suficientes en el repositorio de textos para el llamane concreto para entrenar un modelo de lenguaje, LM (i). Si no es así, el proceso 400 retorna a la etapa 401 para recibir una llamada entrante adicional de ese llamador. Si hay un número suficiente de palabras, el proceso 400 entrena un nuevo modelo de lenguaje, L [i] (para el llamador "i") de acuerdo con el proceso analizado en la FIG. 6 y añade el LM [i] al repositorio 110 de modelos de lenguaje en la etapa 405. El proceso 400 a continuación incrementa el número de modelos de lenguaje específico del llamador, NUMLMMODELS en 1.

En la etapa 406 el proceso 400 genera de salida una señal que indica un "Modelo de Lenguaje Nuevo" y solicita del usuario del sistema un examen manual de la llamada y del reconocimiento del texto en la etapa 407. El usuario

puede examinar y revisar los datos mediante el dispositivo de E / S 115 (mostrado en la FIG. 2). El proceso 400 puede entonces volver a la etapa 401 para recibir otra llamada entrante.

El proceso 400 ilustra la forma en que puede ser utilizado el proceso de identificación del llamador mostrado en la FIG. 7 para generar un modelo de lenguaje correspondiente para cada llamador único. Con el fin de identificar a un llamador que utiliza los modelos de lenguaje, una vez que los modelos de lenguaje han sido entrenados en la medida suficiente, el módulo 112 de identificación del llamador puede simplemente ejecutar el modelo 107 de reconocimiento del habla con el modelo acústico genérico y con el modelo de lenguaje específico del llamador , LM [i] activado a su vez. El modelo de lenguaje que produce el reconocimiento de texto con la mayor probabilidad se corresponde con el llamador actual.

El uso de modelos de lenguaje específicos del llamador específico para identificar un llamador identificará las similitudes semánticas del contenido de la llamada actual con uno de los modelos de lenguaje específicos del llamador, LM [i]. Sin embargo, puede darse el caso de que el actual llamador sea un llamador diferente (no el llamador "i") el cual habla acerca de la misma materia de la que hablaba el llamador "i". Por tanto, los modelos de lenguaje específico del llamador se utilizan de modo preferente, en combinación con los modelos acústicos específicos del llamador para la identificación adecuada de llamadores singulares. Por ejemplo, el proceso de identificación de llamadores acústicos mostrado en la FIG. 7, debe ser ponderado en mayor medida que el proceso de identificación del llamador de modelo de lenguaje entrenado en la FIG. 9 al comunicar un resultado al sistema de identificación. Por ejemplo, si los dos procedimientos de identificación producen resultados diferentes, el resultado de la identificación del modelo de lenguaje se utilizará solo si ofrece una probabilidad mucho más alta que la puntuación acústica específica del llamador del modelo acústico de más alta puntuación. Aquí también, el usuario o el operador del sistema del centro de llamadas puede anular cualquier identificación realizada ya sea por el subsistema de identificación de modelos acústicos o por el subsistema de identificación de modelos del lenguaje.

Los procesos de identificación del llamador mostrados en las FIGS. 7 a 9, pueden crear múltiples modelos acústicos y de lenguaje para los llamadores identificados de forma errónea como "nuevos llamador s". Esto puede producirse, por ejemplo, cuando los fonemas o la materia objeto de dos o más diferentes llamadas procedentes del mismo llamador no se superponen. Cuando los modelos acústicos y del lenguaje continúan siendo entrenados con cada nueva llamada sucesiva procedente de un llamador identificado con anterioridad, los modelos que se corresponden con el mismo llamador empezarán a superponerse unos sobre otros y pueden fusionarse. El sistema de identificación del llamador puede incluir un módulo de fusión que examine de manera periódica todos los modelos específicos del llamador para determinar si cualquier modelo debe ser fusionado en base a criterios definidos de antemano. Estos criterios pueden consistir en la similitud de las señales del modelo para un conjunto determinado de vectores de características, por ejemplo.

Más en concreto, los HMMs de fonemas típicamente probabilidades de transición de los estados utilizando distribuciones gaussianas de múltiples dimensiones (dentro del espacio del Vector de Características) determinadas por un vector Medio y una Matriz de Varianzas. El módulo de fusión podría simplemente agrupar dichos Vectores Medios y / o las matrices de Varianzas para los correspondientes fonemas para cada usuario y ver si están o no lo suficientemente próximos para que se fusionen (utilizando funciones de distancia, como por ejemplo, la distancia Bhattacharya, la cual es la más indicada apara comparar la separación de las funciones de probabilidad, a diferencia de la distancia euclidiana normal).

Así mismo, el sistema de identificación del llamador , (después de saber que, digamos, dos AMs ya entrenados están demasiados próximos uno respecto de otro) almacenar el AM "precursor" (el utilizado como entrada al modulo de entrenamiento en la etapa 306 de la FIG. 8) así como, los archivos WAV utilizados para entrenar (entrada de usuario actual) y solo aplicar el entrenamiento después del "examen manual" (como en la etapa 208 de la FIG. 7) de las entradas de voz de muestra procedentes de los dos llamadores en cuestión. Ello impide la degradación gradual de los AMs específicos del llamador entrenados debido a que son entradas de voz introducidas procedentes de llamadores erróneos. Lo que es con exactitud "demasiado próximo" puede ser cuantificado de manera experimental utilizando cualquier colección disponible de cometidos de Identificación de Usuario (un cuerpo voluminoso de llamadas de teléfonos / archivos WAV pertenecientes a un número lo suficientemente amplio de personas).

Una ventaja de los procesos de identificación del llamador descritos con anterioridad es que el sistema es capaz de la identificación de un llamador solo con una única emisión sonora procedente del llamador. Un nuevo modelo acústico específico del llamador se crea a partir de esa expresión para la identificación de otras llamadas procedentes de ese llamador. Así mismo, el sistema es capaz de identificar a un llamador incluso si el llamador no coopera con ningún mecanismo de respuesta de invitación utilizado para encaminar las llamadas entrantes. Las características acústicas de cualquier emisión sonora, ya sea o no esa emisión sonora una respuesta correcta a una invitación, es modelada para ese llamador. Así mismo, el sistema es capaz de identificar al llamador sin advertir al llamador acerca del proceso de identificación. El proceso puede ser utilizado para filtrar fácilmente llamadas no deseadas de publicitadores telefónicos, por ejemplo, respecto de las llamadas deseadas procedentes de llamadores conocidos.

Así mismo, los grandes centros de llamadas pueden utilizar este sistema para encaminar de una manera más eficiente las llamadas hacia el destinatario o la base de datos de información correctos. Algunos centros de llamadas

requieren que el llamador navegue a través de un largo laberinto de invitaciones antes de ser encaminado hasta el destino correcto. El sistema actual puede proporcionar un llamador identificado con anterioridad con una rápida salida a partir de mecanismo de invitación / respuesta en base a la impresión de voz del llamador y del destinatario o de la materia objeto de las anteriores llamadas. Existen otras muchas aplicaciones para dicho sistema de identificación de llamador s.

Aunque la presente invención ha sido descrita con referencia a formas de realización preferentes, los expertos en la materia advertirán que pueden llevarse a cabo cambios de forma y detalle sin apartarse del alcance de la invención.

REIVINDICACIONES

1.- Un procedimiento de identificación de un llamador de una llamada del llamador a un destinatario, comprendiendo el procedimiento;

a) la recepción de una entrada de voz procedente del llamador;

b) la división de la entrada de voz en subsecciones y la aplicación de las características de cada subsección a una pluralidad de modelos acústicos, la cual comprende un modelo acústico genérico y unos modelos acústicos de cualquier llamador identificado con anterioridad, para obtener una pluralidad de puntuaciones acústicas respectivas que representan en qué medida las características de cada subsección coinciden con los modelos acústicos respectivos;

c) para cada subsección, la identificación del modelo acústico que presenta la mejor puntuación acústica para esa subsección;

la identificación del llamador como uno de los llamadores identificados con anterioridad, solo si las mejores puntuaciones acústicas para todas las subsecciones se corresponden con el mismo llamador identificado con anterioridad; y

en otro caso, la identificación del llamador como un nuevo llamador ; y

d) si el llamador se identifica como un nuevo llamador en la etapa c), la generación de un nuevo modelo acústico para el nuevo llamador, el cual es específico para el nuevo llamador.

2.- El procedimiento de la reivindicación 1, en el que:

la etapa a) comprende la segmentación de la entrada de voz en una secuencia de unidades de habla reconocidas utilizando el modelo acústico genérico;

cada una de la pluralidad de modelos acústicos comprende los modelos de las unidades de habla segmentadas en la etapa a); y

la etapa b) comprende la aplicación de las características a una secuencia de los modelos de las unidades de habla segmentadas en la etapa a) para la pluralidad de modelos acústicos.

3.- El procedimiento de la reivindicación 1, en el que cada uno de la pluralidad de modelos acústicos comprende unos modelos de unidades de habla y en el que el procedimiento comprende así mismo:

e) si el llamador es identificado como uno de los llamadores identificados con anterioridad en la etapa c), la actualización del respectivo modelo acústico para el llamador identificado con anterioridad mediante la identificación de los modelos de las unidades de habla que están incluidas en la entrada de voz, en base a las características.

4.- El procedimiento de la reivindicación 3, en el que la etapa e) comprende la modificación de los modelos de las unidades de habla que están incluidas en la entrada de voz en base a tan poco como a una sola emisión sonora.

5.- El procedimiento de la reivindicación 1, que comprende así mismo:

e) el almacenamiento del nuevo modelo acústico en un repositorio de modelos acústicos con la pluralidad de modelos acústicos, de tal manera que el nuevo modelo acústico se convierte en uno de la pluralidad de modelos acústicos de la etapa b) y el nuevo llamador es incluido como un llamador identificado con anterioridad.

6.- El procedimiento de la reivindicación 1, en el que el modelo acústico genérico comprende unos modelos independientes del llamador de una pluralidad de unidades de habla, y en el que la etapa d) comprende:

d) 1) la generación del nuevo modelo acústico a partir de los modelos independientes del llamador del modelo acústico genérico y la modificación de los modelos independientes del llamador de las unidades de habla que están incluidas en la entrada de voz para representar las características recibidas del nuevo llamador.

7.- El procedimiento de la reivindicación 1, en el que las etapas a) a c) se llevan a cabo sin alertar al llamador durante la llamada de que el llamador está siendo identificado,

8.- El procedimiento de la reivindicación 1, que comprende así mismo:

e) el mantenimiento de un modelo de leguaje específico del llamador para cada uno de los llamadores identificados con anterioridad en base a las entradas de voz procedentes de esos llamador s;

f) la aplicación de las características al modelo acústico genérico y a cada uno de los modelos de lenguaje específicos del llamador para producir una pluralidad de secuencias de unidades de habla reconocidas;

g) la elección de la secuencia de unidades de habla reconocidas que presenta la probabilidad más alta con respecto a las probabilidades de las demás secuencias de unidades de habla reconocidas; y

h) la identificación del llamador en base, en menos en parte, a la secuencia de las unidades de habla reconocidas que presentan la probabilidad más alta.

9.- El procedimiento de la reivindicación 8, que comprende así mismo:

i) si el llamador identificado en la etapa h) es diferente del llamador identificado en la etapa c), la generación de una invitación al usuario para que efectúe un examen manual de al menos uno de los factores siguientes: la entrada de voz, la secuencia de unidades de habla reconocidas, los llamadores identificados, el modelo acústico del llamador identificado en la etapa c), y el modelo de lenguaje específico del llamador del llamador identificado en la etapa h).

10.- El procedimiento de la reivindicación 1, que comprende así mismo:

e) la utilización de una medida de distancia entre la pluralidad de modelos acústicos de los llamadores identificados con anterioridad para señalar determinados modelos acústicos para fusionarlos entre sí.

11.- El procedimiento de la reivindicación 10, en el que la etapa e) comprende la señalización de determinados modelos acústicos con fines de inspección manual.

12.- Un sistema para la identificación de un llamador de una llamada del llamador a un destinatario, comprendiendo el sistema:

un receptor (102) para la recepción de una entrada de voz procedente del llamador ;

un repositorio (111) de modelos acústicos que comprende una pluralidad de modelos acústicos, que incluye un modelo acústico genérico y unos modelos acústicos de los llamadores identificados con anterioridad;

un medio para la división de la entrada de voz en subsecciones;

un medio para la aplicación de las características a cada subsección a una pluralidad de modelos acústicos para obtener una pluralidad de puntuaciones acústicas que representen en qué medida las características de cada subsección coinciden con los modelos acústicos respectivos;

para cada subsección, un medio para la identificación (112) del modelo acústico que presenta la mejor puntuación acústica para esa subsección;

un medio para la identificación del llamador como uno de los llamadores identificados con anterioridad solo si las mejores puntuaciones acústicas para todas las subsecciones se corresponden con el mismo llamador identificado con anterioridad; y

en otro caso, un medio para la identificación del llamador como nuevo llamador ; y

un medio generador de un modelo acústico para la generación de un nuevo modelo acústico para el nuevo llamador si el llamador es identificado como un nuevo llamador .

13.- El sistema de la reivindicación 12, en el que:

el sistema comprende así mismo un módulo de reconocimiento del habla, el cual segmenta la entrada de voz en una secuencia de unidades de habla reconocidas utilizando el modelo acústico genérico;

cada uno de la pluralidad de modelos acústicos comprende unos modelos de las unidades de habla reconocidas por el módulo de reconocimiento del habla; y

el medio para la aplicación comprende un medio para la aplicación de las características a una secuencia de los modelos de las unidades de habla segmentadas por el módulo de reconocimiento del habla para la pluralidad de modelos acústicos.

14.- El sistema de la reivindicación 12, en el que

cada uno de la pluralidad de modelos acústicos comprende unos modelos de unidades de habla;

el sistema así mismo comprende un modelo de actualización del modelo acústico el cual, si el llamador es identificado como uno de los llamadores identificados con anterioridad, actualiza el modelo acústico

respectivo para el llamador identificado con anterioridad mediante la modificación de los modelos de las unidades de habla que están incluidas en la entrada de voz, en base a las características.

15.- El sistema de la reivindicación 14, en el que el módulo de actualización de modelos acústicos es capaz de modificar los modelos de las unidades de habla que están incluidos en la entrada de voz en base a tan poco como una emisión sonora de habla procedente del llamador.

16.- El sistema de la reivindicación 12, en el que el medio generador de modelos acústicos almacena el nuevo modelo acústico en el repositorio de modelos acústicos, de tal manera que el nuevo modelo acústico se convierte en uno de la pluralidad de modelos acústicos y el nuevo llamador es incluido como un llamador identificado con anterioridad.

17.- El sistema de la reivindicación 16, en el que:

el modelo acústico genérico comprende unos modelos independiente del llamador de una pluralidad de unidades de habla; y

el generador de modelos acústicos genera el nuevo modelo acústico a partir de los modelos independientes del llamador del modelo acústico genérico y modifica los modelos independientes del llamador de las unidades de habla que están incluidas en la entrada de voz para representar las características.

18.- El sistema de la reivindicación 12, en el que el sistema está configurado para recibir la entrada de voz e identificar al llamador sin alertar al llamador durante la llamada de que el llamador está siendo identificado.

19.- El sistema de la reivindicación 12, que comprende así mismo:

un repositorio de modelos de lenguaje para el almacenamiento de un modelo de lenguaje específico del llamador para cada uno de los llamadores identificados con anterioridad en base a las entradas de voz procedentes de esos llamador s;

un medio para la aplicación de las características al modelo acústico genérico y a cada uno de los modelos de lenguaje del modelo específico para producir una pluralidad de secuencias de unidades de habla reconocidas; y

un medio para la elección de la secuencia de unidades de habla reconocidas que presente la probabilidad más alta con respecto a las probabilidades de las demás secuencias de unidades de habla reconocidas,

en el que el medio para la identificación identifica al llamador en base, al menos en parte, a la secuencia de unidades de habla reconocidas que presente la probabilidad más alta.

20.- El sistema de la reivindicación 19, en el que el medio para la identificación comprende un medio para la generación de una invitación al usuario para que efectúe un examen manual de al menos uno de los factores siguientes: 1) la entrada de voz, la secuencia de unidades de habla reconocidas que presente la probabilidad más alta, 2) el modelo de lenguaje específico del llamador que produce la secuencia de unidades de habla reconocidas que presente la probabilidad más alta, y 3) el modelo acústico que presente la mejor puntuación acústica, si el modelo de lenguaje específico del llamador que presenta la probabilidad más alta se corresponde con un llamador diferente del modelo acústico que presenta la mejor puntuación acústica en 3).

21.- El sistema de la reivindicación 12, que comprende así mismo:

un medio para la señalización de determinados modelos acústicos para que se fusionen entre sí en base a una medición de distancia entre la pluralidad de modelos acústicos.

22.- El procedimiento de la reivindicación 21, en el que el medio de señalización comprende un medio para la señalización de determinados modelos acústicos a los fines de su inspección manual.

23.- Un medio legible por computadora que comprende unas instrucciones ejecutables por computadora las cuales, cuando son ejecutadas por una computadora, llevan a cabo el procedimiento que comprende:

a) la recepción de una entrada de voz de una llamada de un llamador;

b) la división de la entrada de voz en subsecciones y la aplicación de las características de cada subsección a una pluralidad de modelos acústicos, la cual comprende un modelo acústico genérico y unos modelos acústicos de cualquier llamador identificado con anterioridad, para obtener una pluralidad de puntuaciones acústicas respectivas que representen en qué medida las características de cada subsección coinciden con los modelos acústicos respectivos;

c) para cada subsección, la identificación del modelo acústico que presenta la mejor puntuación acústica para esa subsección;

la identificación del llamador como uno de los llamadores identificados con anterioridad solo si las mejores puntuaciones acústicas para todas las subsecciones se corresponden con el mismo llamador identificado con anterioridad; y

en otro caso, la identificación del llamador como un nuevo llamador ; y

d) si el llamador es identificado como un nuevo llamador en la etapa c), la generación de un nuevo modelo acústico para el nuevo llamador el cual es específico para el nuevo llamador.

24.- El medio legible por computadora de la reivindicación 23, en el que:

la etapa a) comprende la segmentación de la entrada de voz en una secuencia de unidades de habla reconocidas que utilizan el modelo acústico genérico;

cada uno de la pluralidad de modelos acústicos comprende unos modelos de las unidades de habla segmentadas en la etapa a); y

la etapa b) comprende la aplicación de las características a una secuencia de los modelos de las unidades de habla segmentadas en la etapa a) para la pluralidad de modelos acústicos.

25.- El medio legible por computadora de la reivindicación 23, en el que cada uno de la pluralidad de modelos acústicos comprende unos modelos de unidades de habla y en el que el procedimiento comprende así mismo:

e) si el llamador es identificado como uno de los llamadores identificados con anterioridad en la etapa c), la actualización del modelo acústico respectivo del llamador identificado con anterioridad mediante la modificación de los modelos de las unidades de habla que están incluidos en la entrada de voz en base a las características.

26.- El medio legible por computadora de la reivindicación 23, en el que el procedimiento comprende así mismo:

e) el almacenamiento del nuevo modelo acústico en un repositorio de modelos acústicos con la pluralidad de modelos acústicos, de tal manera que el nuevo modelo acústico se convierte en uno de la pluralidad de modelos acústicos de la etapa b), y el nuevo llamador se incluye como un llamador identificado con anterioridad.

27.- El medio legible por computadora de la reivindicación 26, en el que el modelo acústico genérico comprende unos modelos independientes del llamador de una pluralidad de unidades de habla, y en el que la etapa d) comprende:

d) 1) la generación del nuevo modelo acústico a partir de los modelos independientes de la llamada del modelo acústico genérico y la modificación de los modelos genéricos de la llamada de las unidades de habla que están incluidos en la entrada de voz para representar las características.

28.- El medio legible por computadora de la reivindicación 23, en el que el procedimiento comprende así mismo:

e) el mantenimiento de un modelo de lenguaje específico del llamador para cada uno de los llamadores identificados con anterioridad; y

f) la identificación del llamador en base, al menos en parte, a las probabilidades de las secuencias de unidades de habla reconocidas producidas por los modelos de lenguaje específicos del llamador a partir de la entrada de voz.

29.- El medio legible por computadora de la reivindicación 28, en el que el procedimiento comprende:

g) si el llamador identificado en la etapa f) es diferente del llamador identificado en la etapa c), la generación de una invitación al usuario para que efectúe un examen manual de al menos uno de los factores siguientes: la entrada de voz, la secuencia de unidades de habla reconocidas, los llamadores identificados, el modelo acústico del llamador identificado en la etapa c), y el modelo de lenguaje específico del llamador del llamador identificado en la etapa f).

30.- El medio legible por computadora de la reivindicación 23, en el que el procedimiento comprende así mismo:

e) la utilización de una medida de distancia entre la pluralidad de modelos acústicos de los llamadores identificados con anterioridad para señalar determinados modelos acústicos para que se fusionen entre sí.

31.- El medio legible por computadora de la reivindicación 30, en el que la etapa e) comprende la señalización de determinados modelos acústicos con fines de inspección manual.

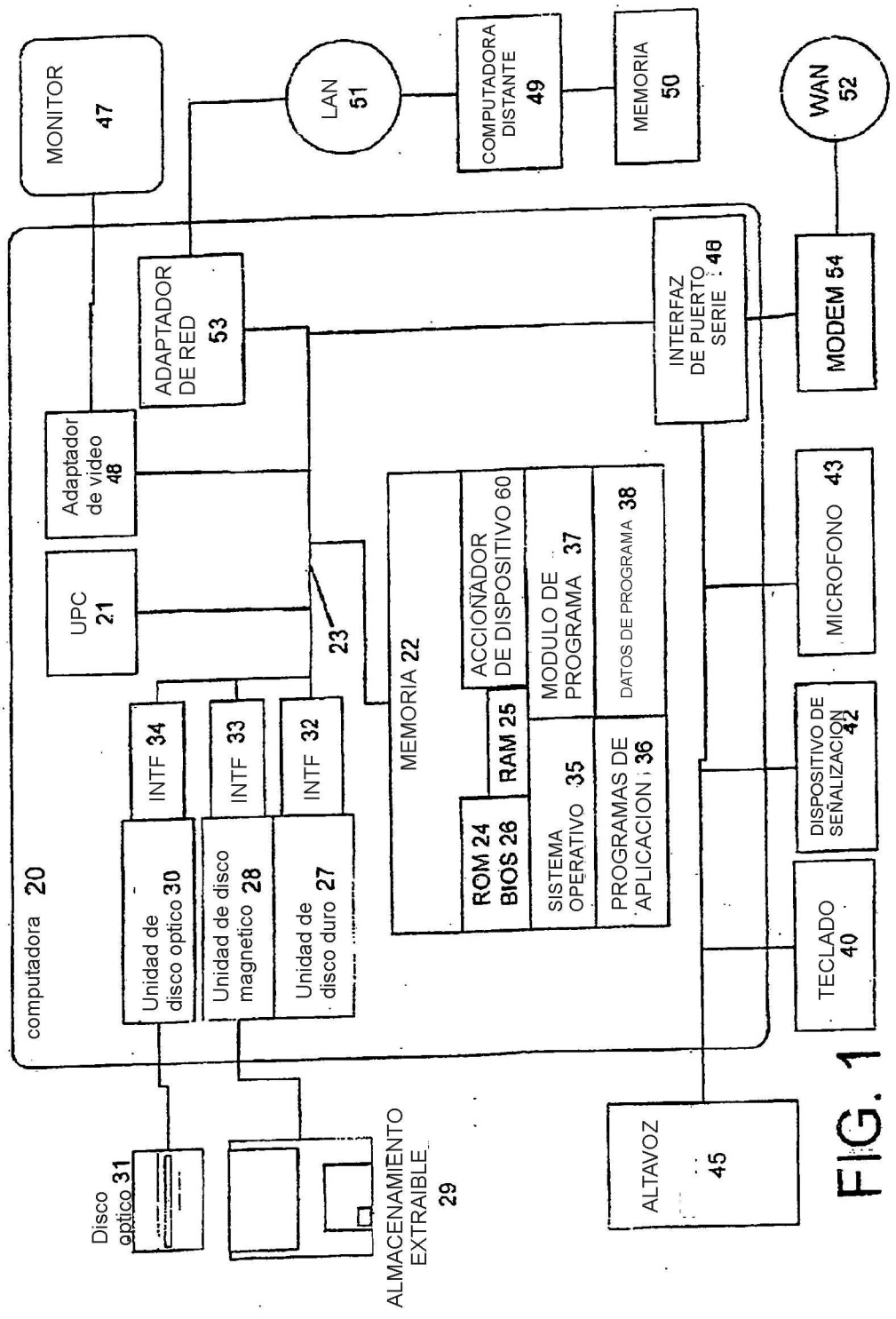

MONITOR 47
LAN 51
COMPUTADORA DISTANTE 49
MEMORIA 50
WAN 52
ADAPTADOR DE RED 53
INTERFAZ DE PUERTO SERIE 46
MODEM 54
Adaptador de video 48
UPC 21
23
MEMORIA 22
ACCIONADOR DE DISPOSITIVO 60
MODULO DE PROGRAMA 37
DATOS DE PROGRAMA 38
RAM 25
ROM 24
BIOS 26
SISTEMA OPERATIVO 35
PROGRAMAS DE APLICACION 36
MICROFONO 43
INTF 34
INTF 33
INTF 32
Unidad de disco optico 30
Unidad de disco magnetico 28
Unidad de disco duro 27
DISPOSITIVO DE SEÑALIZACION 42
TECLADO 40
computadora 20
Disco optico 31
ALMACENAMIENTO EXTRAIBLE 29
ALTAVOZ 45

FIG. 1

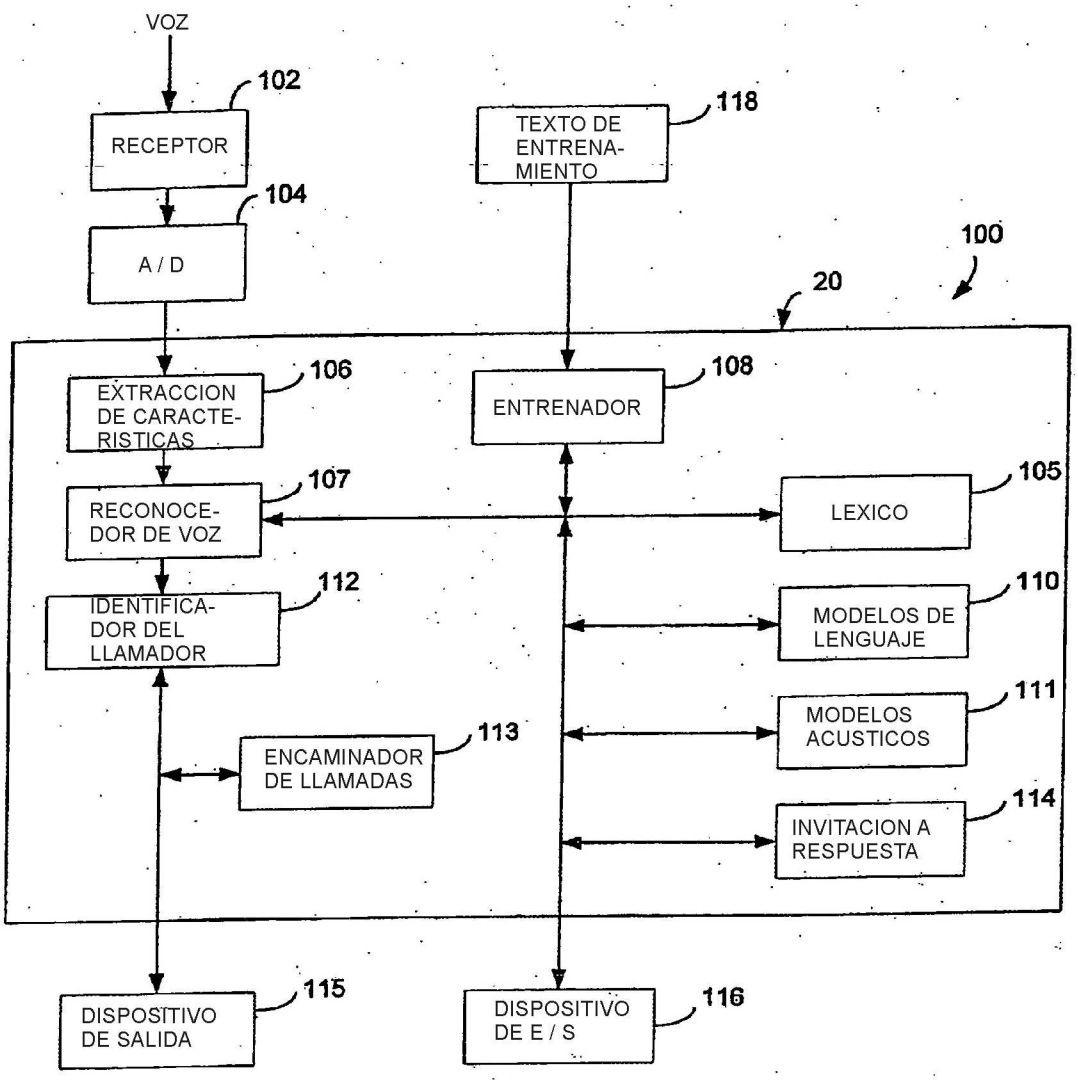

VOZ
102
RECEPTOR
104
A / D
118
TEXTO DE ENTRENA-MIENTO
100
20
106
EXTRACCION DE CARACTE-RISTICAS
108
ENTRENADOR
107
RECONOCE-DOR DE VOZ
105
LEXICO
112
IDENTIFICA-DOR DEL LLAMADOR
110
MODELOS DE LENGUAJE
111
MODELOS ACUSTICOS
113
ENCAMINADOR DE LLAMADAS
114
INVITACION A RESPUESTA
115
DISPOSITIVO DE SALIDA
116
DISPOSITIVO DE E / S

FIG. 2

FIG. 3

FIG. 4

FIG. 5

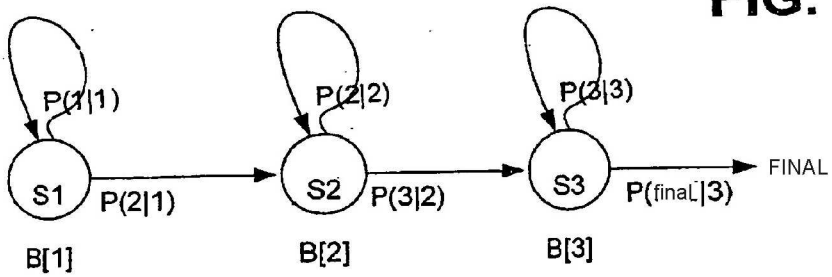
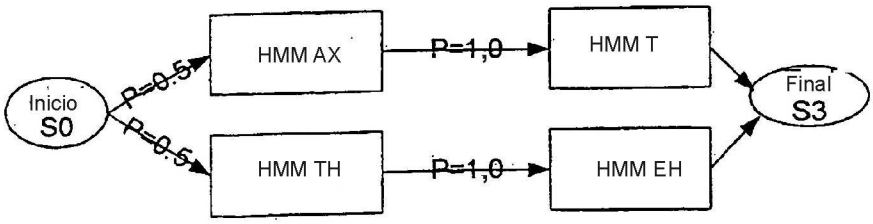


FIG. 6



Palabra “en” vista una vez

palabra “el” vista una vez

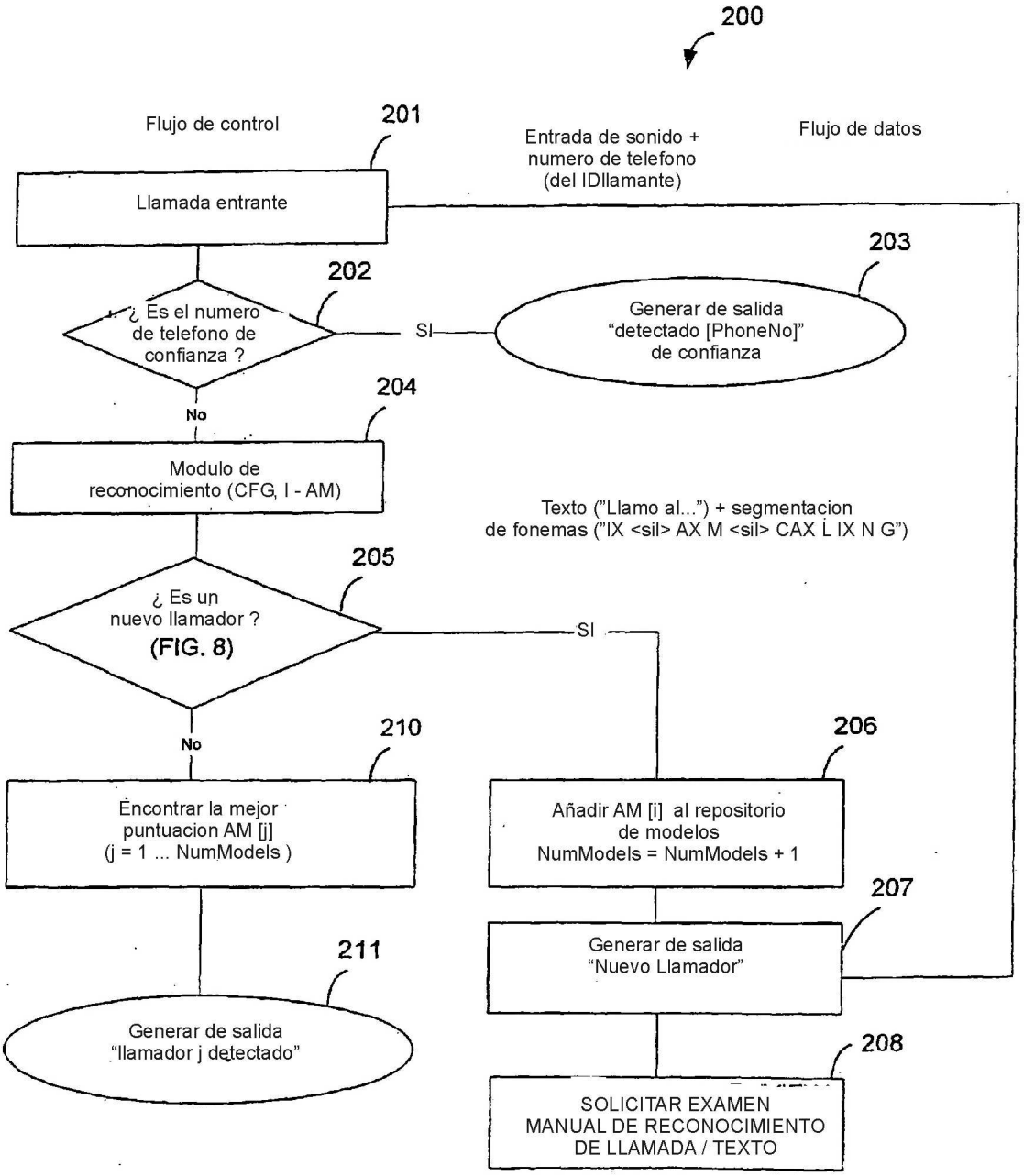

200
Flujo de control
201
Entrada de sonido + numero de telefono (del IDllamante)
Flujo de datos
Llamada entrante
203
202
¿ Es el numero de telefono de confianza ?
SI
Generar de salida "detectado [PhoneNo]" de confianza
No
204
Modulo de reconocimiento (CFG, I - AM)
Texto ("Llamo al...") + segmentacion de fonemas ("IX <sil> AX M <sil> CAX L IX N G")
205
¿ Es un nuevo llamador ?
(FIG. 8)
SI
No
210
206
Encontrar la mejor puntuacion AM [j] (j = 1 ... NumModels)
Añadir AM [i] al repositorio de modelos NumModels = NumModels + 1
207
211
Generar de salida "llamador j detectado"
Generar de salida "Nuevo Llamador"
208
SOLICITAR EXAMEN MANUAL DE RECONOCIMIENTO DE LLAMADA / TEXTO

FIG. 7

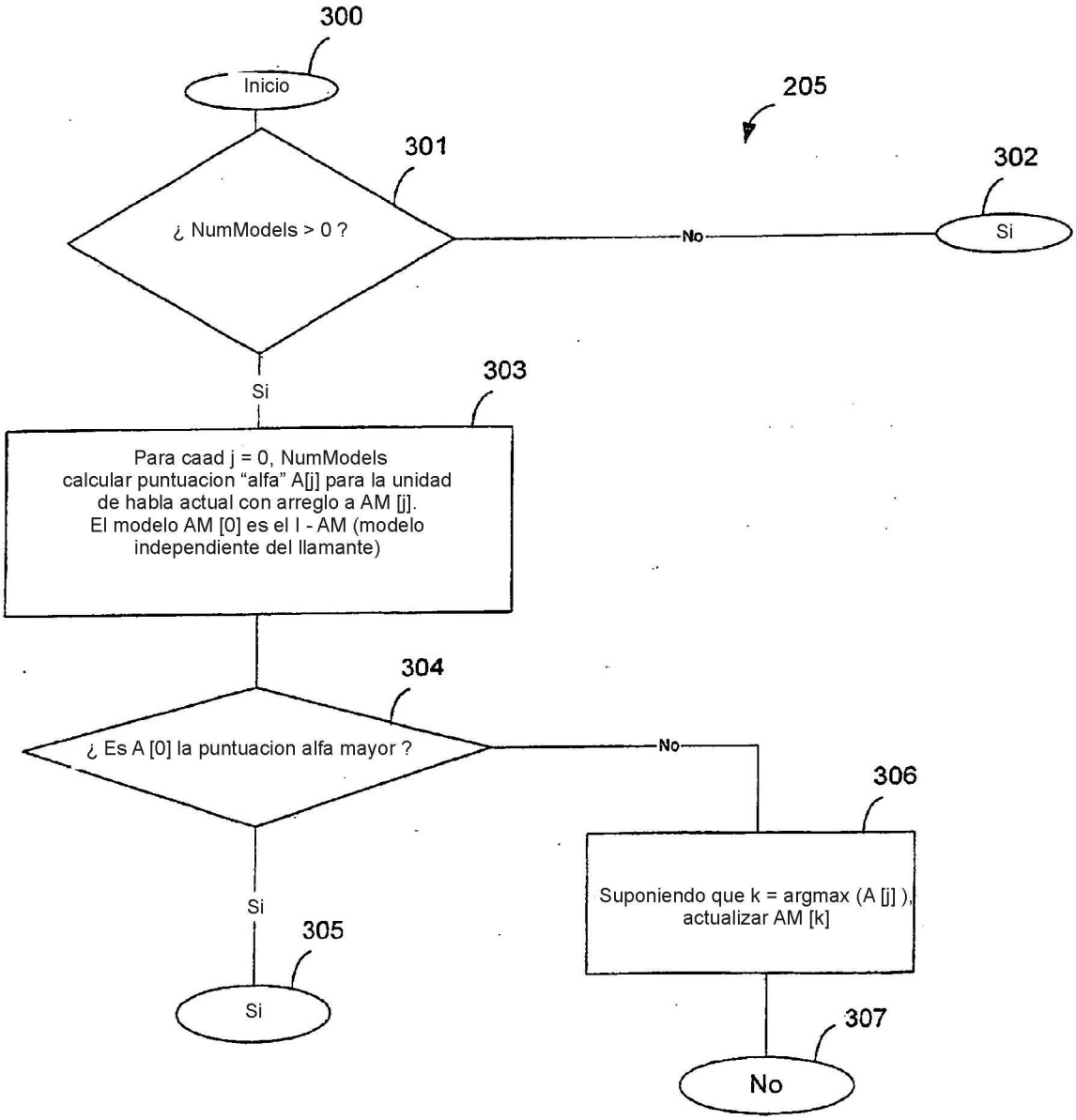

300
Inicio
205
301
¿ NumModels > 0 ?
No
302
Si
Si
303
Para caad j = 0, NumModels
calcular puntuacion "alfa" A[j] para la unidad
de habla actual con arreglo a AM [j].
El modelo AM [0] es el I - AM (modelo
independiente del llamante)
304
¿ Es A [0] la puntuacion alfa mayor ?
No
306
Suponiendo que k = argmax (A [j]),
actualizar AM [k]
Si
305
Si
307
No

FIG. 8

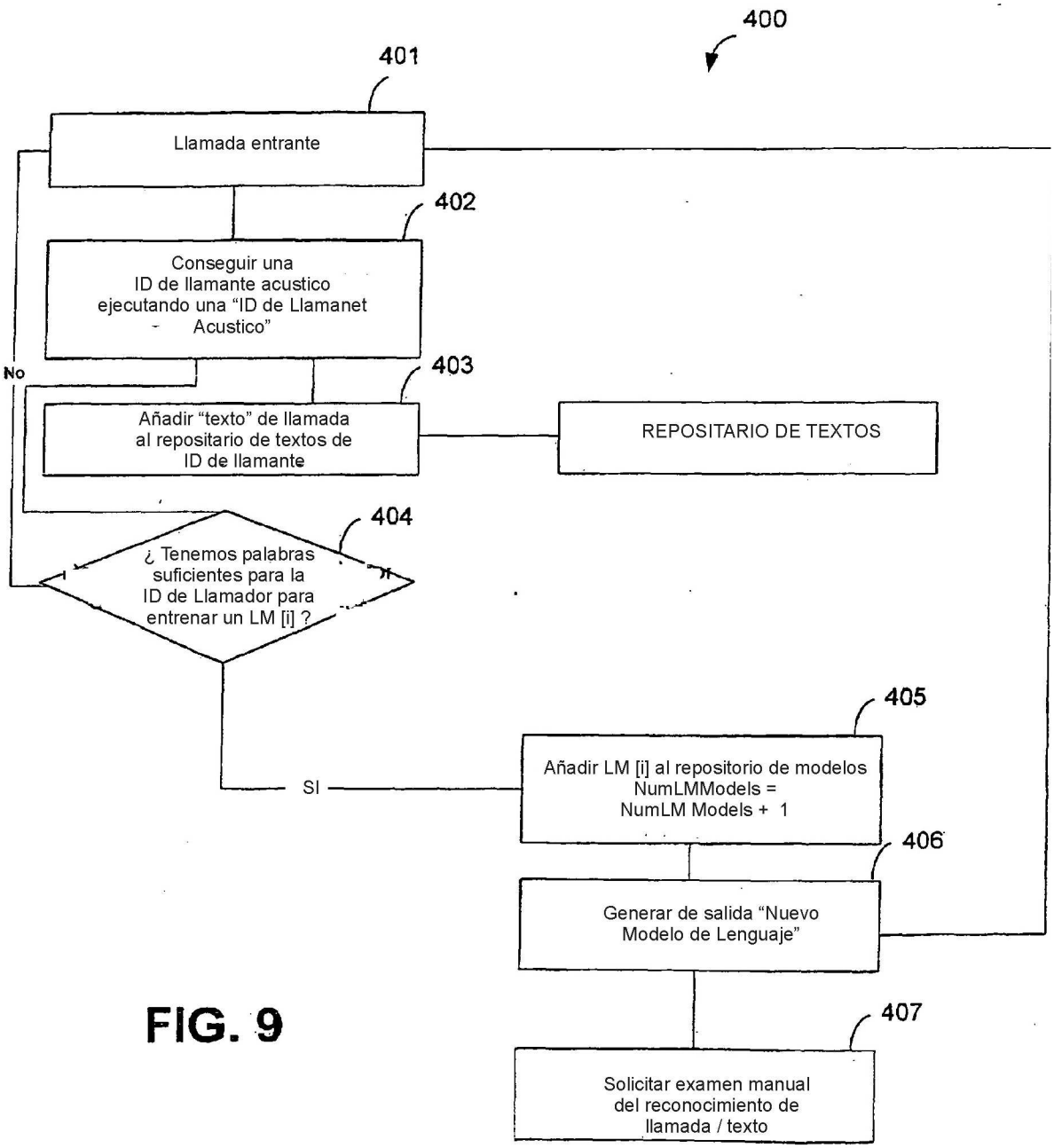

400
401
Llamada entrante
402
Conseguir una
ID de llamante acustico
ejecutando una "ID de Llamanet
- Acustico"
No
403
Añadir "texto" de llamada
al repositario de textos de
ID de llamante
REPOSITARIO DE TEXTOS
404
¿ Tenemos palabras
suficientes para la
ID de Llamador para
entrenar un LM [i] ?
SI
405
Añadir LM [i] al repositorio de modelos
NumLMModels =
NumLM Models + 1
406
Generar de salida "Nuevo
Modelo de Lenguaje"
407
Solicitar examen manual
del reconocimiento de
llamada / texto

FIG. 9