(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 276 674**

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

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **00911864 .7**
(86) Fecha de presentación : **18.02.2000**
(87) Número de publicación de la solicitud: **1183680**
(87) Fecha de publicación de la solicitud: **06.03.2002**

(54) Título: **Sistema automatizado de transcripción y método que usa dos modelos de conversión de voz y corrección asistida por computadora.**

(30) Prioridad: **19.02.1999 US 120997 P**

(45) Fecha de publicación de la mención BOPI:
01.07.2007

(45) Fecha de la publicación del folleto de la patente:
01.07.2007

(73) Titular/es: **Custom Speech USA, Inc.**
Suite B365, 360 North Court Street
Crown Point, Indiana 46307, US

(72) Inventor/es: **Kahn, Jonathan;**
Qin, Charles y
Flynn, Thomas, P.

(74) Agente: **Carvajal y Urquijo, Isabel**

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

Sistema automatizado de transcripción y método que usa dos modelos de conversión de voz y corrección asistida por computadora.

Antecedentes de la invención

1. **Campo de la invención**

La invención se refiere en lo general a sistemas de computadora de reconocimiento de voz y en particular a un sistema y método para automatizar la transcripción al texto del dictado por voz de varios usuarios finales.

2. **Técnica anterior**

Programas de reconocimiento de voz son bien conocidos en la técnica. Aunque estos programas finalmente son útiles para convertir automáticamente la voz a texto, muchos usuarios se disuaden de usar estos programas porque se requiere que cada usuario gaste una cantidad importante de tiempo entrenándose para el sistema. Usualmente este entrenamiento empieza haciendo que cada usuario lea una serie de materiales preseleccionados durante aproximadamente 20 minutos. Luego, cuando el usuario continua usando el programa, como las palabras se transcriben de manera impropia se espera que el usuario se detenga y entrene al programa de modo que la palabra pretendida llegue a la última exactitud del modelo acústico. Desafortunadamente la mayoría de dos profesionales (doctores, dentistas, veterinarios, abogados) y ejecutivos de negocios, no tienen voluntad para gastar el tiempo desarrollando el modelo acústico necesario para beneficiarse realmente de la transcripción automática.

Por lo tanto es un objeto de la presente invención proveer un sistema que ofrezca entrenamiento transparente del programa de reconocimiento de voz a los usuarios finales.

Hay sistemas para usar computadoras para la transcripción con ruta desde un grupo de usuarios extremos. Los más frecuente estos sistemas se usan en lugares de muchos usuarios, tales como hospitales. En esos sistemas un usuario con su voz dicta en una computadora de propósito general u otro dispositivo de registro y el texto resultante se transfiere automáticamente a una transcripción humana. La transcripción humana transcribe el texto que luego se regresa al "autor" original para revisarlo. Estos sistemas tienen la condición perpetua de emplear un número suficiente de transcripciones humanas para transcribir todos los textos dictados.

Por lo tanto es otro objeto de la presente invención proveer un medio automatizado de transmitir la voz al texto, donde siempre sea adecuado llevar a un mínimo el número de transcripciones humanas necesaria para transcribir los textos audibles que vienen en el sistema.

Es un objeto asociado al anterior el proveer un medio simplificado para proveer transcripciones de texto Verbatim para el entrenamiento de un modelo acústico del usuario para una parte de reconocimiento de voz del sistema.

Es otro objeto asociado de la presente invención, automatizar un programa de reconocimiento de voz preexistente para llevar aun mínimo el numero de operadores necesarios para operar el sistema.

Estos y otros objetos se harán evidentes a los técnicos con los dibujos, las especificaciones y las reivindicaciones.

Sumario de la invención

La presente descripción se refiere a un sistema y un método para los servicios de transcripción básicamente automáticos para uno o más usuarios de voz. En particular este sistema incluye usar dos casos de convertir voz para facilitar el establecimiento de un texto de transcripción verbatim con un mínimo de transcripción humana.

El sistema incluye medios para recibir un texto de dictado de voz de un usuario actual. El texto de dictado de voz se alimenta a un primer medio para convertir automáticamente le texto de dictado de voz en un primer texto descrito y luego a segundos medios para automáticamente convertir el texto de dictado de voz en un segundo texto escrito.

Los medios primero y segundo tiene variable de conversión en juegos primero y segundo respectivamente. Estos juegos primero y segundo de conversión variables, tienen cuando menos una diferencia.

Por ejemplo, cuando los medios de conversión de voz automáticos primero y segundo comprenden cada uno un programa de reconocimiento de voz preexistente, los programas ellos mismos pueden ser diferentes uno del otro. Varios programas de reconocimiento de voz tienen inherentemente diferentes vías de conversión voz a texto, así, resultan en una conversión diferente en casos de voz difíciles, los cuales a su vez pueden usarse para establecer el texto verbatim. Entre los medios de conversión de voz preexistentes disponibles están Naturally Speaking de Dragon Systems, Via Voice de IBM y Magic Speech de Philips Corporation.

En otra vía o modalidad, los juegos primero y segundo de las variables de conversión pueden cada uno comprender un modelo de lenguaje (esto es, un modelo lenguaje general o especializado) que de nuevo resultaría en conversiones diferentes en casos difíciles llevando a un establecimiento más fácil del texto verbatim. Alternativamente uno o mas juegos asociados con el programa o programas de reconocimiento de voz preexistentes, que se usen podría modificarse.

En otro intento, el texto de dictado con voz puede procesarse previamente a su aplicación a uno o ambos de los medios de conversión automáticos. De esta manera las variables de conversión (esto es el tamaño de la palabra digital, la tasa de muestra y el retiro de rangos armónicos particulares) pueden diferirse entre los casos de conversión de voz.

El sistema incluye además medios para editar manualmente una copia de los textos escritos primero y segundo para crear el texto verbatim del texto o presentación del dictado con voz. En una modalidad, el primer texto escrito se sincroniza cuando menos temporalmente a la presentación de dictado con voz. En este caso, el medio de edición manual incluye medios para comparar una copia en secuencia de los textos escritos primero y segundo, dando como resultado una lista en secuencia de palabras no acopladas tomadas del primer texto escrito. El medio de edición manual incluya además medios para buscar crecientemente una palabra no acoplada corriente contemporáneamente dentro de un primer esquema asociado con el primer medio de conversión automática, que contenga el primer texto escrito y un segundo esquema asociado con la lista en secuencia.

Los medios de edición manuales también incluyen medios para corregir la palabra no acoplada corrien-

te en el segundo esquema. El medio corrector incluye medios para presentar la palabra no acoplada corriente de una manera substancialmente aislada visualmente del otro texto, en el primer texto escrito y medios para presentar una porción del registro de dictado de voz sincronizado del primer esquema asociado con la palabra no acoplada corriente. En una modalidad los medios de edición incluyen además medios para ver alternativamente la palabra no acoplada corriente en el contexto dentro de la copia del primer texto escrito.

El sistema también puede incluir medios de entrenamiento para mejorar la exactitud del programa de reconocimiento de voz.

La aplicación también presenta un método para los servicios de transcripción automatizados para uno o más usuarios con voz en un sistema que incluya cuando menos un programa de reconocimiento de voz. El método incluye (1) recibir un texto de dictado con voz desde un usuario actual con voz, (2) crear automáticamente un primer texto escrito desde el texto dictado con voz con un programa de reconocimiento escrito usando un primer juego de variables de conversión; (3) crear automáticamente un segundo texto escrito desde el texto de dictado con voz con un programa de reconocimiento de voz usando un segundo juego de variables del conversión; (4) establecer manualmente un texto verbatim a través de la comparación de los textos escritos primero y segundo, (5) regresar el texto verbatim al usuario actual. Establecer un texto verbatim que incluya (6) comparar en secuencia una copia del primer texto escrito con el segundo texto escrito resultando en una lista en secuencia de palabras no acopladas tomadas de la copia del primer texto escrito, la lista en secuencia tiene un principio; un final y una palabra actual no acoplada, la palabra actual no acoplada se avanza sucesivamente desde el principio al fin, (7) se busca de manera creciente para la palabra no acoplada actual contemporáneamente dentro de un primer esquema asociado con cuando menos un programa de reconocimiento de voz que contenga el primer texto escrito y un segundo esquema asociado con la lista en secuencia. (8) presentar la palabra no acoplada actual de una nueva manera substancialmente aislada visualmente del otro texto en la copia del primer texto escrito y presentar una parte del dictado con voz sincronizado registrada desde el primer esquema asociado con la palabra no acoplada actual, y (9) corregir la palabra no acoplada actual para que sea una representación verbatim de la porción del registro de dictado con voz sincronizado.

Breve descripción de los dibujos

La Figura 1 de los dibujos es un diagrama en bloque de una modalidad potencial del presente sistema para automatizar substancialmente servicios de transcripción para uno o más usuarios con voz;

La Figura 1b de los dibujos es un diagrama de bloque de una computadora de propósito general la cual puede usarse como estación de dictado, una estación de transcripción y el medio de control dentro del presente sistema;

La Figura 2a de los dibujos es un diagrama de flujo del lazo principal del medio de control del presente sistema;

La Figura 2b de los dibujos es un diagrama de flujo de la parte de la etapa de inscripción de los medios de control del presente sistema,

La Figura 2c de los dibujos es un diagrama de flujo de la parte de la etapa de entrenamiento del medio de control del presente sistema;

La Figura 2d de los dibujos es un diagrama de flujo de la parte de la etapa de automatización del medio de control del presente sistema;

La Figura 3 de los dibujos es una estructura de directorio usada por el medio de control en el presente sistemas;

La Figura 4 de los dibujos es un diagrama de bloque de una parte de una modalidad preferida de los medios de edición manuales;

La Figura 5 de los dibujos en una vista en elevación de lo restante de una modalidad preferida del medio de edición manual; y

La Figura 6 de los dibujos es una ilustración del arreglo del sistema que presenta sistema de transcripción automatizado y método que usa casos de conversión de dos voz y corrección asistida por computadora.

Mejores métodos para realizar la invención

Aunque la presente invención puede materializarse en muchas formas diferentes, se muestra aquí en los dibujos y se discute a continuación algunas modalidades con el entendimiento que la presente discusión ha de considerarse únicamente como un ejemplo de los principios de la invención y no pretende limitar la invención a las modalidades ilustradas.

La figura 1 de los dibujos muestra en lo general una modalidad potencial del presente sistema para los servicios de transcripción substancialmente automatizados para uno o más usuarios con voz. El presente sistema debe incluir algún medio para recibir un texto de dictado con voz desde un usuario actual. Estos medios receptores del texto de dictado con voz pueden ser un registro audio - digital, un registro audio - analógico, o medios normales para recibir textos de computadora en medios magnéticos o por medio de una conexión de datos.

Como se muestra, en una modalidad el sistema 100 incluye las estaciones de registro digital múltiples 10, 11, 12 y 13. Cada estación de registro digital tiene cuando menos un registro audio digital y medios para identificar el usuario con voz actual. Preferentemente cada una de estas estaciones de registro digitales, está asociada en una computadora de propósito general (tal como la computadora 20) aunque una computadora especializada puede desarrollarse para este propósito específico. La computadora de propósito general aunque tiene la ventaja adicional de ser adaptable a varios usos además de funcionar dentro del presente sistema 100. En general, la computadora del propósito general debe tener entre otros elementos, un microprocesador (tal como el de Intel Corporation PENTIUM, Cyrix K6 ó Motorola 6800) memoria volátil y no volátil, uno o más dispositivos de almacenamiento en masa, esto es, HDD (no mostrado) floppy drive 21, y otros dispositivos retirables 22 tal como un impulsor CD-ROM DITTO, ZIP o JAZ, de (Iomega Corporation y similares), varios dispositivos de entrada de usuario, tal como un ratón (mouse) 23, un tablero 24, o un micrófono 25; y un sistema de presentación en vídeo 26. En una modalidad la computadora de propósito general está controlada por el sistema de operación WINDOS 9.x. Sin embargo se considera que el sistema presente trabajaría igualmente bien usando una computadora MACINTOSH o algún otro sistema

operativo, tal como WINDOWS CE, UNIX o JAVA para citar únicamente unos cuantos.

A parte de la plataforma de computadora particular en una modalidad utilizando una entrada audio analógica (vía micrófono 25) la computadora de propósito general debe incluir una tarjeta de sonido (no mostrada) por su puesto, en una modalidad con una entrada digital no sería necesaria una tarjeta de sonido.

En la modalidad mostrada en la figura 1, las estaciones de grabación de audio digital 10, 11, 12 y 13 están cargadas y configuradas para correr un software de grabación de audio digital en un sistema computacional en base a PENTIUM operando de acuerdo con WINDOWS 9.x o de otros vendedores tales como The Programmers' Consortium, Inc. De Oakton Virginia (VOICEDOC). Syntrillium Corporation de Phoenix, Arizona (COOL EDIT) o Dragon Systems Corporation (Dragon Naturally Speaking Professional Edition). Estos diferentes programas de software producen un archivo de dictado de voz en forma de un archivo "WAV". Sin embargo como es conocido para los expertos en la técnica, otros formatos de archivo de audio tales como MP3 o DSS, también podrían usarse para formatear el archivo de dictado de voz sin salirse del espíritu de la presente invención. En una modalidad en la cual se usa el software VOICEDOC ese software también asigna un archivo para que maneje el archivo WAV, sin embargo aquellos con un conocimiento ordinario en la técnica saben guardar un archivo de audio en un sistema de computadora usando métodos de manejo de archivos de sistema operativo estándar.

Otros medios para recibir un archivo de dictado de voz es una grabadora digital especial 14, tal como la Olympus Digital Voice Recorder D-1000 fabricada pro la Olympus Corporation. Así, si el usuario de voz actual está más cómodo con un tipo más convencional pude continuar usando una grabadora digital especial 14. Con el fin de obtener el archivo de texto de audio digital, al terminar la grabación, la grabadora digital especial 14 se conectara operativamente a una de las estaciones de grabado de audio digital tal como la 13, para descargar el archivo de audio digital en una computadora de uso general. Con este método, por ejemplo no se requeriría tarjeta de audio.

Otra alternativa de recibir el archivo de dictado de voz puede consistir en el uso de una forma u otra de medios magnéticos en una de las estaciones de grabación de audio digital para cargar el archivo de audio en el sistema.

En algunos casos puede ser necesario procesar previamente los archivos de audio para hacerlos aceptables para el procesamiento por parte del software de reconocimiento de voz. Por ejemplo, un formato de archivo DSS que puede tener que ser cargado en un formato de archivo WAV, o la velocidad de muestreo de un archivo de audiodigital deber ser aumentada o reducida. Por ejemplo al usar la grabadora de voz digital Olympus, la velocidad de Dragon Naturally Speaking, Olympus de 8 MHz debe ser transformada a 11 MHz. Software para lograr ese pre-procesamiento está disponible de una variedad de fuentes incluyendo Syntrillium Corporation y Olympus Corporation.

El otro aspecto de las estaciones de grabación de audio digitales es un medio para identificar al usuario de voz actual. Los medios de identifiación pueden incluir un teclado 24 sobre le cual el usuario (u otro operado) puede introducir el código de identificación único del usuario actual. Es claro que la identificación del usuario puede introducirse usando una variedad de dispositivos de entrada de computadora, tal como dispositivos de apunte (por ejemplo ratón 23), pantalla de tacto (no mostrada), un apuntador de luz (no mostrado), una lectora de código de barras (no mostrada) o sonidos de audio por medio de un micrófono 25, por nombrar solo algunos.

En el caso de un usuario por primera vez los medios de identificación también pueden asignar al usuario un numero de identificación después de recibir la información potencialmente identificadora de ese usuario, incluyendo: (1) nombre; (2) dirección; (3) ocupación; (4) acento o dialecto vocal, etc. Como se menciona en asociación con los medios de control, en base a esta información de entrada, se establece un perfil de usuario de voz y un sub-directorio dentro de los medios de control. Así sin importar los medios de identificación particulares usados, una identificación de usuario debe establecerse para cada usuario de voz y subsecuentemente proporcionarse con un archivo de audio digital correspondiente para cada uso de tal forma que medios de control pueden enrutar apropiadamente y el sistema finalmente transcribir el audio.

En una modalidad de la presente invención, los medios identificadores también pueden buscar la selección manual de un vocabulario especial. Se contempla que los valores del vocabulario especial puedan ser generados para diferentes usuarios tales como médicos (por ejemplo radiología, cirugía ortopédica, ginecología) y legal (por ejemplo corporativa, patentes, litigios) o altamente específicos tales que dentro de cada especialidad los parámetros de vocabulario pudieran limitarse posteriormente en base a las circunstancias particulares de un archivo de dictado particular. Por ejemplo si el usuario de voz actual es un radiólogo que dicta las lecturas de una exploración CAT abdominal la nomenclatura es altamente especializada y diferente de la nomenclatura de un ultrasonido renal. Al segmentar angostamente cada valor de vocabulario seleccionable es factible un aumento en la precisión del convertidor de voz automático.

Como se muestra en la figura 1, las estaciones de grabación de audio digitales pueden estar conectadas de manera operativa al sistema 100 como parte de una red de computadoras 30, o alternativamente puede estar conectada operativamente al sistema por medio de un anfitrión de Internet 15. Como se muestra n la figura 1b, la computadora para propósitos generales puede estar conectada tanto al enchufe de la red 27 como al del teléfono. Con el uso de un anfitrión de Internet, puede lograrse la conexión al enviar por correo electrónico el archivo de audio por medio del Internet. Otro método para completar tal conexión es por medio de conexión directa a módem por medio de un software de control remoto tal como PC ANYWHERE, que es distribuido por Symantec Corporation de Cupertino California. También es posible, si se conoce la dirección IP de la estación de grabado de audio digital 10 o el anfitrión de Internet 15, el transferir el archivo de audio usando el protocolo de transferencia de archivos básico. Así, como puede observarse a partir de lo anterior, el presen-

te sistema permite gran flexibilidad para los usuarios de voz para proporcionar una entrada de audio en el sistema.

Los medios dé control 200 controlan el flujo del archivo de dictado de voz de acuerdo con el estado de entrenamiento del usuario de voz actual. Como se muestra en las figuras 2a,2b,2c,2d, los medios de control 200 comprenden un programa de software que opera en una computadora para todo uso 40. En particular, el programa se inicializa en la etapa 201 en donde se fijan las variables, los acumuladores se vacían y en particular se carga la configuración particular de esta instalación particular de los medios de control. Los medios de control monitorean continuamente un directorio objetivo (tal como el "actual" (mostrado en la figura 3)) para determinar si un nuevo archivo ha sido movido en el objetivo, etapa 202. Una vez que se encuentra un nuevo archivo (tal como el "6723.id" (mostrado en la figura 3)), se realiza una determinación de si el usuario actual 5 (mostrado en la figura 1) es un usuario nuevo, etapa 203.

Para cada nuevo usuario (como lo indica la existencia de un archivo ".pro" en el subdirectorio "actual"), se establece un nuevo subdirectorio, la etapa 204 (tal como el subdirectorio "usern" (mostrado en la figura 3)). Este subdirectorio se usa para almacenar todos los archivos de audio ("xxxx.wav"), texto escrito ("xxxx.wrt"), texto verbatim ("xxxx.vb"). texto de transcripción ("xxxx.txt") y perfil de usuario ("usern.pro") para el usuario en particular. Cada trabajo en particular es asignado a un número único "xxxx" tal que todos los archivos asociados con un trabajo pueden asociarse por es enumero. Con esta estructura de directorio, el número de usuarios se limita prácticamente por el espacio de almacenamiento dentro de la computadora de uso general 40.

Ahora que el subdirectorio de usuario ha sido establecido, el perfil de usuario se mueve al subdirectorio, etapa 205. El contenido de este perfil de usuario puede variar entre sistemas. El contenido de un perfil de usuario potencia se muestra en la figura 3 conteniendo; el nombre del usuario, la dirección, la ocupación y el nivel de entrenamiento. Además de la variable del nivel de entrenamiento, que se necesita, los otros datos son útiles para enrutar y transcribir los archivos de audio.

Los medios de control que han seleccionado un grupo de archivos con una manija, determinan la identidad del usuario actual al comparar el archivo ".id" con su "user.tbl", etapa 206. Ahora que el usuario es conocido, el perfil de usuario puede ser recuperado del subdirectorio del usuario y determinarse el nivel de entrenamiento actual, etapa 207. Las etapas 208-211 indican que el nivel de entrenamiento actual es: inscripción, entrenamiento, automatización y detener automatización.

La inscripción es la primera etapa en los servicios de transcripción automáticos. Como se muestra en la figura 2h, el archivo de usuario se envía a la transcripción, etapa 301. En particular, el archivo "xxxx.wav" se transfiere a las estaciones de transcripción 50 y 51. En una modalidad preferida, ambas se estaciones son computadoras de propósito general, que hacen funcionar tanto un reproductor de audio y medios de entrada manuales. El reproductor de audio es probablemente un reproductor de audio digital, aunque es posible que un archivo de audio analógico pueda ser transferido a las estaciones. Existen varios reproductores de audio incluyendo una utilidad en el sistema operativo WINDOWS 9.x y varios de terceros tales como The Programmers' Consortium, Inc. De Oakton, Virginia (VOICESCRIBE). Sin imitar el reproductor de audio usado para reproducir el archivo de audio, medios de entrada manuales corren en la computadora al mismo tiempo. Estos medios de entrada manual pueden consistir de cualquier editor de textos o procesador de palabras (tal como MS WORD, WordPerfect, AmiPro o Word Pad) en combinación con un teclado, ratón, u otro dispositivo de interfaz con el usuario. En una modalidad de la presente invención, estos medios de entrada manuales puede por si mismo ser software de reconocimiento de voz, tal como Naturally Speaking de Dragon Systems de Newton, Massachusetts, Via Voice de IBM Corporation de Armonk, Nueva York, o Speech Magic de Philips Corporation de Atlanta Georgia. El trascriptor humano 6 escucha el archivo de audio creado por el usuario actual 5 como se conoce introduce manualmente el contenido percibido del teto grabado, estableciendo así el archivo transcrito, etapa 302. Siendo humano, el transcriptor humano 6 es probable que aplique su experiencia, educación y tendencias en el texto y así no introduzca una transcripción verbatim del archivo de audio. Al completar la transcripción humana, el transcriptor humano 6 guarda el archivo e indica que está listo para ser transferido al subdirectorio actual de Usuario como "xxxx.txt", etapa 303.

En tanto que este actuario actual esta solo en la etapa de inscripción, un operador humano tendrá que escuchar el archivo de audio y compararlo manualmente con el archivo transcrito y crear un archivo verbtim, etapa 304. El archivo verbatim "xxxx.vb" también se transfiere al subdirectorio de usuario actual, etapa 305. Ahora que el texto verbatim esta disponible, los medios de control 200 indican los medios de conversión de texto automáticos, etapa 306. Estos medios automáticos de conversión de voz pueden ser un programa pre-existente, tal como el Naturally Speaking de Dragon System's, el Via Voice de IBM o Speech Magic de Philips, por nombrar solo algunos. Alternativamente, podría ser un programa único que está diseñado para realizar específicamente el reconocimiento de voz automático.

En una modalidad preferida, el Naturally Speaking de Dragon Systems ha sido usado al correr un ejecutable simultáneamente con Naturally Speaking que alimenta los golpes de tecla y operaciones con el ratón fantasma a través de WIN32API, de tal forma que Naturally Speaking cree que esta interactuando con un ser humano, cuando de hecho está siendo controlado por medios de control 200. Esas técnicas son bien conocidas en la técnica de las pruebas del software de computación y no será detallado. Debe ser suficiente le decir que al ver el flujo de la aplicación de cualquier programa de reconocimiento de voz, puede crearse un ejecutable para imitar las etapas manuales interactivas.

Si el usuario actual es un nuevo usuario, el programa de reconocimiento de voz no necesitara establecer al nuevo usuario, etapa 307. Los medios de control proporcionan la información necesaria a partir del perfil de usuario encontrado en el subdirectorio de usuario actual. Todo reconocimiento de voz requiere entrenamiento suficiente para establecer un modelo acústico de un usuario particular. En el caso de Dragon, inicialmente le programa busca duran-

te aproximadamente 20 minutos de audio usualmente obtenidos por el usuario que lee un texto enlatado proporcionado por Dragon Systems. También existe una funcionalidad incluida en Dragon que permite el "entrenamiento móvil". Usando esta característica, el archivo verbatim y el archivo de audio se alimentan en el programa de reconocimiento de voz para iniciar el entrenamiento del modelo acústico para ese usuario, etapa 308. A pesar de la longitud del archivo de audio, los medios de control 200 cierran el programa de reconocimiento de voz al terminar el archivo, etapa 309.

Como en la etapa de inscripción es demasiado temprana para usar el texto creado automáticamente, una copia del archivo transcrito se envía al usuario actual usando la información de dirección contenida en el perfil de usuario, etapa 310. Esta dirección puede ser una dirección normal o una dirección de correo electrónica. Después de esa transmisión, el programa regresa al ciclo principal de la figura 2a.

Después de que se han realizado un cierto numero de minutos de entrenamiento para un uso en particular, le nivel de entrenamiento del usuario puede cambiarse de inscripción a entrenamiento. El limite de esto es subjetivo pero tal ve una buena regla es cuando Dragon parece estar creando texto escrito con una precisión del 80% o mas, puede realizarse la conmutación entre los estados. Así para un usuario de ese tipo el siguiente evento de transcripción enviara a los medios de control 200 al estado de entrenamiento. Como se muestra en la figura 2c, las etapas 401-403 son las mismas etapas de transcripción humana que las 301-303 en la fase de inscripción. Una vez que el archivo transcrito se establece, los medios de control 200 inician los medios de conversión de voz automática (o el programa da reconocimiento de voz) y selecciona al usuario actual, etapa 404. El archivo de audio se alimenta en el programa de reconocimiento de voz y el texto escrito se establece dentro del acumulador del programa, etapa 405. En el caso de Dragon, este acumulador se le asigna el mismo archivo en el mismo caso del programa. Así el acumulador puede ser copiado fácilmente usando comandos estándares del sistema operativo y la edición manual puede empezar, etapa 406.

En una modalidad particular utilizando el sistema VOICEWARE de The Programmers' Consortium, Inc. De Oakton, Virginia, el usuario introduce el audio en el programa VOICEDOC del sistema VOICEWARE, creando así un archivo ".wav". Además antes de liberar este archivo ".wav" en el servidor VOICEWARE, el usuario selecciona un "transcriptor". Esta "transcripción" puede ser un transcriptor humano particular o puede ser el "transcriptor computarizado". Si el usuario selecciona un "transcriptor computarizado" también puede seleccionar si la transcripción se maneja local o remotamente. Al archivo se le asigna un número de trabajo por parte del servidor VOICEWARE, que enruta el trabajo a la porción VOICESCRIBE del sistema. Normalmente, VOICESCRIBE se usa por el transcriptor humano para recibir y reproducir el archivo de audio del trabajo (".wav"). Además el archivo de audio es tomado por los medios de conversión de voz automáticos. En esta modalidad del sistema VOICEWARE, al colocar a VOICESCRIBE en la "modalidad automática" los nuevos trabajos (esto es un archivo de audio recién creado por VOICEDOC) se descargan automáticamente del servidor VOICEWARE y una ventana VOICESCRIBE que tiene un titulo de ventana formado por el número de trabajo del archivo ".wav" actual. Un archivo ejecutable, que corre en el fondo "ve" que la ventana VOICESCRIBE está abierta y usando el WIN32APIU determina el numero de trabajo a partir del titulo de la ventana VOICESCRIBE. El archivo ejecutable entonces activa los medios de conversión de voz automáticos. En el Naturally Speaking del Sistema Dragon, por ejemplo existe incluida una función para realizar el reconocimiento de voz en un archivo ".wav" pre-existente. El programa ejecutable alimenta los datos del teclado fantasma a Naturally Speaking para abrir el archivo ".wav" del directorio "actual" (ver figura 3) que tiene el número de trabajo del trabajo actual.

En esta modalidad, después de que Naturally Speaking ha terminado de transcribir automáticamente el contexto del archivo ".wav", el archivo ejecutable prosigue con la operación al seleccionar todo el texto en la ventana abierta de Naturally Speaking y copiarla en el tablero del sistema operativo WINDOWS 9.x. Entonces usando la utilidad de tablero, guardar el tablero como archivo de texto usando el número de trabajo actual con un sufijo "dmt". El archivo ejecutable entonces "oprime" el botón "terminar" en el VOICESCRIBE para regresar el archivo "dmt" al servidor VOICEWARE. Como aquellos con experiencia común en la técnica el procedimiento anterior puede realizarse utilizando otro software de grabación digital y otros medios de conversión de voz automáticos. Adicionalmente, existe funcionalidad análoga la tablero WINDOWS en otros sistemas operativos. También es posible requerir la intervención humana para activar o repetir una o varias de las etapas anteriores. Además aunque los diferentes programas que ejecutan las diferentes etapas podrían correr en un numero de computadoras interconectadas (por medio de LAN, WAN, conexión a Internet, correo electrónico y similares), también se contempla que todo el software necesario puede correr en una sola computadora.

Otra modalidad alternativa también se contempla en la cual el usuario dicta directamente en los medios de conversión de voz automáticos y el servidor VOICEWARE toma una copia en la dirección inversa. Esta modalidad funciona de la siguiente manera: sin realmente grabar voz, le usuario oprime el botón "terminar" en VOICEDOC creando un archivo ".wav" vacío. Este archivo vacío sin embargo recibe un numero de trabajo único por parte del servidor VOICEWARE. El usuario (o un archivo ejecutable que corre en el fondo) entonces despliega los medios de conversión de voz automáticos y el usuario dicta directamente en el programa, de la misma manera previamente usada en asociación con esos medios de conversión de voz automáticos. Al terminar de dictar, el usuario presiona un botón titulado "retorno" (generado por un archivo ejecutable de fondo), el ejecutable entonces inicializa un macro que toma el número de trabajo actual de VOICEWARE (de la manera descrita antes), selecciona todo el texto en el documento y la copia en el tablero. EL tablero entonces se guarda en el archivo "<jobnumer>.dmt," como se describe antes. El ejecutable entonces "oprime" el boten "terminar" (por medio de WIN32API) en el VOICESCRIBE, que regresa efectivamente el archivo de texto transcrito automáticamente al servidor VOICEWARE, que a su vez regresa la transcripción completa al usuario del VOICESCRIBE. Notablemente, aunque varios programas

ejecutando varias etapas podrían estar corriendo en varias computadoras interconectadas (vía un LAN, WAN, conexión de Internet, correo electrónico y similares), también se contempla que todo el software necesario este corriendo en una misma computadora. Como lo entenderán aquellos con experiencia normal en la técnica, el procedimiento anterior puede realizarse utilizando otro software de grabación digitales y otros medios de conversión automáticos de voz. Adicionalmente, funcionalidad análoga al tablero WINDOWS existe en otros sistemas operativos. También es posible requerir la intervención humana para activar una o mas de las etapas anteriores.

La edición manual no es una tarea fácil. Los seres humanos tienden a equivocarse. Así la presente invención también incluye medios para mejorar esa tarea. Como se muestra en la figura 4, el archivo transcrito ("3333.txt") y la copia del texto escrito ("3333.wrt") Se comparan secuencialmente palabra por palabra 406ª con el fin de establecer una lista secuencial de palabras que no coinciden 406b que son seleccionadas de la copia del texto escrito. Esta lista tiene un apuntador inicial y uno final 406c a la palabra no coincidente actual. Como base de esta lista secuencial se encuentra otra lista de objetos que contiene las palabras no coincidentes originales, así como las palabras inmediatamente anterior y posterior a esa palabra que no coinciden, el punto inicial en la memoria de cada palabra que no coincide en la lista secuencial de palabras sin coincidir 406b y la longitud de la palabra que no coincide.

Como se muestra en la figura 5, la palabra no coincidente señalada por el apuntador 406c de la lista 406b se muestra en un aislado visual substancial del otro texto de la copia del texto escrito en un monitor de computadora estándar 500 en una ventana activa 501. Como se muestra en la figura 5, el contexto de la palabra no coincidente puede seleccionarse por el operador para que sea mostrado dentro de la frase en la que reside, palabra por palabra o en el contexto de la frase, al oprimir sobre los botones 514, 515 y 516, respectivamente.

Asociada a esta ventana activa 501 se encuentra una ventana de fondo 502 que contiene la copia del archivo de texto escrito. Como se muestra en la ventana de fondo 502, una búsqueda creciente ha localizado (ver apuntador 503) la siguiente ocurrencia de una palabra no coincidente "cash". AL mismo tiempo, dentro de la ventana 505 que contiene al acumulador del programa de reconocimiento de voz. La misma búsqueda creciente ha localizado (ver apuntador 506) la siguiente ocurrencia de la palabra no coincidente. Un usuario humano que probablemente solo estará viendo la ventana activa 501, activara la reproducción de audio del programa de reconocimiento de voz al oprimir el botón "reproducir" 501, que reproduce el audio sincronizado al apuntador de texto 506. En base a esa parte de texto que puede ser repetido una y otra vez al oprimir sobre el botón reproducir, el usuario humano puede introducir manualmente la corrección a la palabra no coincidente por medio del teclado, el ratón o hasta ordenes auditivas en otro programa de reconocimiento de voz que corre en esta ventana.

En el presente ejemplo, aun si la selección del contexto aislado ofrecido por los botones 514, 515 y 516, puede ser un difícil de determinar la palabra verbatim correcta fuera del contexto, de acuerdo con esto existe un botón de cambio de ventana 513 que moverá la ventana de fondo 502 a la parte superior con el apuntador visible 503 indicando la localización actual dentro de la copia del texto escrito. El usuario entonces puede regresar a la ventana activa e introducir la palabra correcta, "trash". Esta cambio solo afectara la copia del texto escrito en la ventana de fondo 502.

Cuando el operador está listo para la siguiente palabra no coincidente, el operador oprime el botón de avance 511, que avanza el apuntador 406c en la lista de palabras no coincidentes y activa la búsqueda creciente en ambas ventanas 502 y 505. Esta palabra sin coincidir ahora se muestra aislada y el operador puede reproducir el texto sincronizado del programa de reconocimiento de voz y corregir esta palabra también. Si en algún punto en la operación, el operador quisiera regresar a la palabra no coincidente previa, el operador oprime en el botón de inversa 512, que mueve el apuntador 406c de regreso a la palabra en la lista y provoca un retroceso en la búsqueda creciente. Esto se logra usando la lista inferior de objetos que contiene las palabras no coincidentes originales. Esta lista es atravesada objeto por objeto, pero alternativamente cada una de las grabaciones podría ser arreglada de tal forma que cada punto tiene el mismo tamaño de palabra para ayudar al desplazamiento bidireccional de la lista. Como las palabras no coincidentes en esta lista solo se leen solo es posible regresar a la palabra no coincidente original de tal forma que el operador puede determinar si una corrección diferente debió haber sido realizada.

Finalmente la copia del texto escrito se corrige finalmente dando como resultado una copia verbatim, que se guarda en el subdirectorio del usuario. El archivo verbatim también se pasa al programa de reconocimiento de voz para el entrenamiento, etapa 407. El nuevo (y mejorado) modelo acústico se guarda, etapa 408, y el programa de reconocimiento de voz se cierra, etapa 409. Como el sistema aun esta en entrenamiento, el archivo transcrito se regresa al usuario, como en la etapa 310 de la fase de inscripción.

Como se muestra en la figura 4, el sistema puede también incluir medios para determinar la precisión de la salida de los medios comparadores secuenciales. Específicamente, al contra el número de palabras en el texto escrito y el número de palabras en la lista 406b la proporción de palabras en esa lista secuencial a palabras en el texto escrito puede determinarse, proporcionando un porcentaje de precisión. Como antes, depende de la selección cuando hace avanzar a los usuarios de una etapa a otra. Una vez que se ha logrado el objetivo, el perfil de usuario se cambia en la siguiente etapa, etapa 211.

Se provee una mejora potencial o funcionalidad derivativa al determinar el porcentaje de precisión. En una modalidad, este porcentaje puede usarse para evaluar la pericia de un transcriptor humano. En particular, al usar cualquier archivo verbatim conocido o un usuario bien establecido, el archivo ".wav" asociado será reproducido para el transcriptor humano y la anterior comparación se realizara en el texto transcrito contra el archivo verbatim creado por medio del proceso anterior. De esta manera puede proporcionarse funcionalidad adicional por medio del presente sistema.

Como se entiende actualmente, los fabricantes de los programas de reconocimiento de voz usados para grabar idiomas extranjeros, dicciones, etc. con archivos verbatim establecidos manualmente a los mode-

los de voz del programa. Debe ser fácilmente evidente que el método anterior de establecer texto verbatim, podría usarse en el desarrollo de esos archivos de voz simplificando grandemente este proceso.

Una vez que el usuario ha llegado a la etapa de automatización, los mayores beneficios del presente sistema pueden obtenerse. El software de reconocimiento de voz se inicializa, etapa 600, y el usuario actual se selecciona, etapa 601. Si se desea puede seleccionarse un vocabulario particular, etapa 602. Entonces la conversión automática del archivo de audio digital grabado por el usuario actual, puede comenzar, etapa 603. Cuando se completa, el archivo escrito se transmite al usuario en base a la información contenida en el perfil del usuario, la etapa 604 y el programa regresa al ciclo principal.

Desafortunadamente, existen casos en los cuales los usuarios de voz no pueden usar la transcripción automática durante un periodo de tiempo (durante una enfermedad, después de un trabajo dental, etc.) debido a que su modelo acústico ha sido alterado temporal (o aun permanentemente). En ese caso, el administrador del sistema puede fijar la variable de nivel de entrenamiento para detener el estado de automatización en el cual solo se realiza las etapas 301, 302, 303, 305 y 310 (ver la figura 2b).

La figura 6 de los dibujos muestra otro arreglo potencial de varios elementos asociados con la presente invención. En este arreglo, como antes, un usuario dicta verbalmente un documento que desea transcribir, que se guarda como un archivo de dictado de voz 700 en una de las maneras descritas antes. En esta modalidad, mas que tener un transcriptor humano que produzca un archivo transcrito - el archivo de dictado de voz se convierte automáticamente en un texto escrito cuando menos dos veces.

Después de la doble conversión de texto automática, los resultantes archivos de texto escrito primero y segundo se comparan entre si usando medios de edición de copias manual (como se describe antes en asociación con las figuras 4 y 5) facilitando el que un operador humano corrija rápida y manualmente el segundo archivo de texto escrito.

De esta manera, se cree que el servicio de transcripción puede proporcionarse con un esfuerzo menor por parte del transcriptor humano. La clave para obtener un texto escrito lo suficientemente preciso para entregarlo al usuario final debe diferenciarse la conversión voz a texto de alguna manera entre las corridas primera y segunda. En particular, entre la etapa de conversión primera y segunda el sistema puede cambiar uno o mas de los siguientes:

(1) programas de reconocimiento de voz (por ejemplo el sistema Naturally Speaking de Dragon, Via Voice de IBM o Magic Speech de Philips Corporation);

(2) modelos de idioma dentro de un programa de reconocimiento de voz particular (por ejemplo ingles general versus un vocabulario especializado (por ejemplo medico, legal));

(3) valores ajustados dentro de u programa de reconocimiento de voz (por ejemplo “mas preciso” versus “velocidad”); y/o

(4) el archivo de dictado de voz al pre-procesarlo con un procesador de señal digital (tal como Cool Edit de Syntrillium Corporation de Phoenix, Arizona o un CI DSP56000 programado de Motorola, Inc.) al cambiar el tamaño de palabras digitales, la velocidad de muestreo, retirar rangos armónicos y otras modificaciones de potencial.

Al cambiar una o más de las anteriores “variables de conversión” se cree que la segunda conversión voz a texto producirá un texto escrito ligeramente diferente que la primera conversión voz a texto y que al comparar los dos textos escritos resultantes usando los nuevos medios de edición manual descritos aquí, un operador humano puede revisar las diferencias en la forma mencionada antes y producir rápidamente un texto verbatim para presentarlo al usuario final. Así de esta manera, se cree que una transcripción completamente automática puede lograrse con menos intervención humana que en los otros métodos descritos.

Este sistema y el método básico se ilustran en la figura 6. Debe observarse que aunque se muestran dos medios de conversión de voz automáticos 702 y 703, puede haber un solo caso de un programa de reconocimiento de voz corriendo en una sola computadora, pero usando diferentes variables de conversión entre las iteraciones de conversión del archivo de dictado de voz. Es claro que es igualmente posible el tener múltiples casos de programas de reconocimiento de coa corriendo en una sola máquina o aun en máquinas separadas interconectadas por medio de una red computarizada (LAN, WAN, par a par, o similares) como lo saben aquellos con un conocimiento ordinario en la técnica.

Similarmente cuando los medios de edición manuales 705 se muestran como separados de los medios de conversión de voz automáticos también pueden ser implementados en la misma computadora como uno o ambos casos de medios de conversión automáticos. De igual manera los medios de edición manuales pueden implementarse también en una computadora separada, así como interconectados con otras computadoras en una red computarizada.

Finalmente, el procesador de señal digital 701 se muestra para ilustrar que un método para cambiar las variables de conversión es el de alterar i la entrada al archivo de dictado de voz a uno o ambos de los casos de los medios de conversión automática de voz. Otra vez, este procesador de señal digital puede ser implementado en la misma computadora como cualquiera o todos los bloques funcionales anteriores en una computadora separada interconectada con las otras computadoras por medio de una red computarizada.

Se contempla que el caso anterior en el cual se usan dos iteraciones de conversión voz a texto podría extrapolarse a un caso en el cual se realizan aun más iteraciones de conversión, usando cada uno varios grupos de variables de conversión, realizándose la comparación de texto entre pares únicos de salidas de texto escritas y después entre si, teniendo como resultado un aumento en la precisión de la transcripción automática dejando pocas palabras para ser consideradas en la edición manual.

La descripción y los dibujos anteriores solo explican e ilustran la invención y la invención no debe ser limitada a ellos. Aquellos expertos en la técnica que tengan ante ellos la descripción podrán hacer modificaciones y variaciones sin salirse del alcance de

la presente invención. Por ejemplo, es posible implementar todos los elementos del presente sistema en una sola computadora de propósito general al esencialmente compartir la maquina entre el usuario de voz, el transcriptor y el programa de reconocimiento de voz. Los ahorros de costo resultantes hacen que este sistema sea accesible a mas tipos de situaciones de oficinas no solo a grandes clínicas medicas, hospitales, bufetes legales u otras empresas grandes.

REIVINDICACIONES

1. Un sistema para automatizar los servicios de transcripción para uno o más usuarios de voz, que comprende:

- medios para recibir un archivo de dictado de voz por parte de un usuario actual, siendo este usuario actual el usuario de voz o uno de los varios usuarios de voz;

- primeros medios para convertir automáticamente el mencionado archivo de dictado de voz en un primer texto escrito, los primeros medios automáticos de conversión tienen un primer grupo de variables de conversión;

- segundos medios para convertir automáticamente el mencionado archivo de dictado de voz en un segundo texto escrito, los segundos medios automáticos de conversión tienen un segundo grupo de variables de conversión, los grupos primero y segundo de variables de conversión tienen al menos una diferencia; y

- medios para editar manualmente una copia de los anteriormente mencionados textos escritos primero y segundo para crear un texto verbatim del archivo de dictado de voz;

  - en donde el primer texto escrito está al menos temporalmente sincronizado con el archivo de dictado de voz, y los mencionados medios de edición manual comprenden:

  - medios para comparar secuencialmente una copia del primer texto escrito con el segundo texto escrito dando como resultado una lista secuencial de palabras no coincidentes obtenidas de la copia del primer texto escrito, la lista secuencial tiene un inicio, un final y una palabra no coincidente actual, la palabra no coincidente actual avanza sucesivamente desde dicho inicio hasta el final;

  - medios para buscar incrementalmente la palabra no coincidente actual simultáneamente con un primer acumulador, que contiene el primer texto escrito, asociado con los primeros medios de conversión automáticos y un segundo acumulador asociado con la lista secuencial; y

  - medios para corregir la palabra no coincidente actual en el segundo acumulador, los medios de corrección incluyen medios para mostrar la palabra no coincidente actual en una manera sustancialmente aislada visualmente del otro texto en la copia del primer texto escrito, y medios para reproducir una porción de la grabación del dictado de voz sincronizada asociada con la palabra no coincidente actual del primer acumulador.

2. El sistema, de conformidad con la reivindicación 1, en donde los medios de edición además incluyen medios para ver alternativamente la palabra no coincidente actual en el contexto dentro de la copia del primer texto escrito.

3. El sistema, de conformidad con la reivindicación 1, en donde los medios de conversión de voz automáticos primero y segundo comprenden, cada uno, un programa de reconocimiento de voz preexistente para el uso interactivo humano, cada uno de esos medios convertidores de voz automáticos primero y segundo incluyen medios para responder automáticamente a una serie de preguntas interactivas del programa de reconocimiento de voz preexistente.

4. El sistema, de conformidad con la reivindicación 3, en donde la diferencia entre los grupos primero y segundo de variables de conversión es ofrecida por los mencionados medios automáticos primero y segundo de conversión de voz, que incluyen diferentes programas de reconocimiento de voz preexistentes.

5. El sistema, de conformidad con la reivindicación 4, en donde los medios de conversión de voz automáticos se seleccionan dentro del grupo que consiste esencialmente de Naturally Speaking de Dragon Systems, Via Voice de IBM y Magic Speech de Philips Corporation.

6. El sistema, de conformidad con la reivindicación 1, en donde la diferencia entre el grupo primero y segundo de variables de conversión comprende un modelo de lenguaje usado en asociación con el programa de reconocimiento de voz preexistente.

7. El sistema, de conformidad con la reivindicación 6, en donde un modelo de lenguaje generalizado se utiliza en el primer grupo de variables de conversión y un modelo de lenguaje especializado se usa en el segundo grupo de variables de conversión.

8. El sistema, de conformidad con la reivindicación 1, en donde la diferencia entre los grupos primero y segundo de variables de conversión consiste por lo menos en un grupo asociado con el programa de reconocimiento de voz preexistente.

9. El sistema, de conformidad con la reivindicación 1, en donde la diferencia entre el grupo primero y segundo de variables de conversión comprende medios para procesar previamente el audio antes de ser introducido a los primeros medios de conversión automáticos.

10. El sistema, de conformidad con la reivindicación 9, en donde la diferencia entre los grupos primero y segundo de variables de conversión comprende medios para procesar previamente el audio antes de ser introducido a los segundos medios de conversión automáticos, en donde dicha variable primera y segunda de procesamiento previo es diferente.

11. El sistema, de conformidad con la reivindicación 10, en donde las variables de procesamiento previo se seleccionan del grupo consistente esencialmente de tamaño de palabra digital, velocidad de muestreo y retiro de intervalos armónicos particulares.

12. El sistema, de conformidad con la reivindicación 1, en donde la diferencia entre los grupos de variables de conversión primero y segundo comprende un modelo de lenguaje usado en asociación con el programa de reconocimiento de voz preexistente.

13. El sistema, de conformidad con la reivindicación 12, en donde un modelo de lenguaje generalizado se utiliza en el primer grupo de variables de conversión y un modelo de lenguaje especializado se usa en el segundo grupo de variables de conversión.

14. El sistema, de conformidad con la reivindicación 1, en donde la diferencia entre los grupos primero y segundo de variables de conversión comprende medios para procesar previamente el audio antes de

su introducción a los primeros medios de conversión automáticos.

15. El sistema, de conformidad con la reivindicación 14, en donde la diferencia entre el grupo primero y segundo de variables de conversión comprende medios para procesar previamente el audio antes de ser introducido a los segundos medios de conversión automáticos, siendo diferente la variable primera y segunda de procesamiento previo.

16. El sistema, de conformidad con la reivindicación 1, que además incluye medios para formar a dichos medios de conversión de voz automáticos para obtener una mayor precisión con el archivo de voz del usuario actual.

17. El sistema, de conformidad con la reivindicación 16, en donde los medios de formación comprenden una porción de formación preexistente de un programa de reconocimiento de voz previamente existente destinado al uso interactivo humano, los medios de formación incluyen medios para responder automáticamente a una serie de preguntas interactivas de la porción de formación previamente existente de dicho programa de reconocimiento de voz preexistente.

18. Un método para la automación de los servicios de transcripción para uno o más usuarios de voz para un sistema que incluya por lo menos un programa de reconocimiento de voz, que consiste de las siguientes etapas:

- recibir un archivo de dictado de voz de un usuario de voz actual;

- crear automáticamente un primer texto escrito a partir del archivo de dictado de voz con un programa de reconocimiento de voz usando un primer grupo de variables de conversión;

- crear automáticamente un segundo texto escrito a partir del archivo de dictado de voz con un programa de reconocimiento de voz usando un segundo grupo de variables de conversión;

- establecer manualmente un archivo verbatim por medio de la comparación de los textos escritos primero y segundo; y

- devolver el archivo verbatim al usuario actual,

- en donde dicha etapa de establecimiento manual de un archivo verbatim incluya las siguientes subetapas:

- comparar secuencialmente una copia del primer texto escrito con el segundo texto escrito dando como resultado una lista secuencial de palabras no coincidentes obtenidas de la copia del primer texto escrito, la lista secuencial tiene un principio, un final y una palabra no coincidente actual, la palabra no coincidente actual avanza sucesivamente desde el inicio al final;

- buscar incrementalmente la palabra no coincidente actual simultáneamente dentro de un primer acumulador, que contiene el primer texto escrito, asociado con al menos un programa de reconocimiento de voz, y un segundo acumulador asociado con la lista secuencial; y

- mostrar la palabra no coincidente actual de una forma sustancialmente aislada visualmente de otro texto en la copia del primer texto escrito y reproducir la porción de la grabación de dictado de voz sincronizada asociada con la palabra no coincidente actual desde el primer acumulador; y

- recibir un entrada manual de la corrección de la palabra no coincidente actual para que sea una representación verbatim de la porción de la grabación de dictado de voz sincronizada.

19. El método, de conformidad con la reivindicación 18, que además comprende:

- seleccionar el primer grupo de variables de conversión a partir de los programas de reconocimiento de voz preexistentes disponibles; y

- seleccionar de manera diferente el segundo grupo de variables de conversión a partir de los programas de reconocimiento de voz preexistentes disponibles.

20. El método, de conformidad con la reivindicación 18, que además comprende:

- seleccionar el primer grupo de variables de conversión a partir de modelos de lenguaje disponibles; y

- seleccionar de manera diferente el segundo grupo de variables de conversión a partir de modelos de lenguaje disponibles.

21. El método, de conformidad con la reivindicación 18, que además comprende procesar previamente el archivo de dictado de voz antes de crearse automáticamente un primer texto escrito, el procesamiento previo forma al menos una parte del primer grupo de variables de conversión.

22. El método, de conformidad con la reivindicación 21, que además incluye el proceso previo del archivo de dictado de voz de una manera diferente al primer grupo de variables de conversión de procesamiento previo, antes de crear automáticamente un segundo texto escrito, el procesado previo forma al menos una parte del segundo grupo de variables de conversión.

FIG. 1

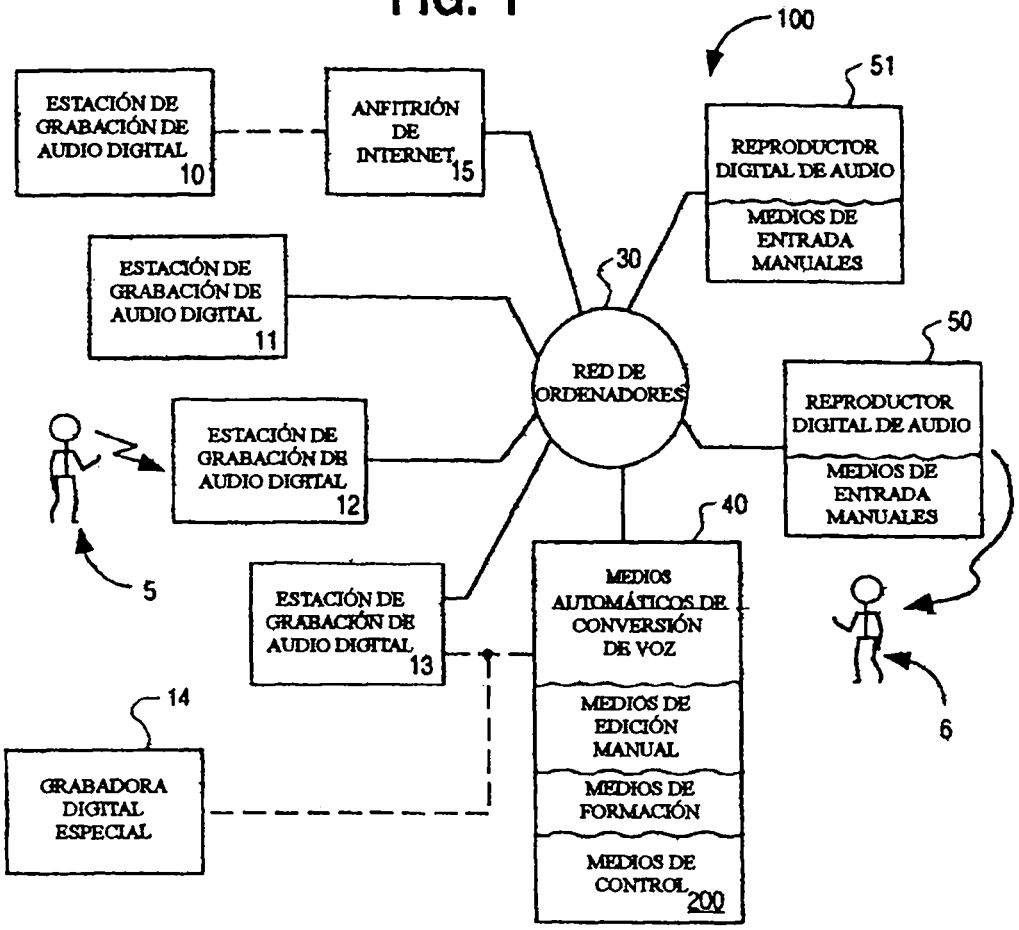

100
51
ESTACIÓN DE GRABACIÓN DE AUDIO DIGITAL 10
ANFITRIÓN DE INTERNET 15
REPRODUCTOR DIGITAL DE AUDIO
MEDIOS DE ENTRADA MANUALES
30
ESTACIÓN DE GRABACIÓN DE AUDIO DIGITAL 11
50
RED DE ORDENADORES
REPRODUCTOR DIGITAL DE AUDIO
MEDIOS DE ENTRADA MANUALES
ESTACIÓN DE GRABACIÓN DE AUDIO DIGITAL 12
5
40
ESTACIÓN DE GRABACIÓN DE AUDIO DIGITAL 13
MEDIOS AUTOMÁTICOS DE CONVERSIÓN DE VOZ
MEDIOS DE EDICIÓN MANUAL
MEDIOS DE FORMACIÓN
MEDIOS DE CONTROL 200
6
14
GRABADORA DIGITAL ESPECIAL

FIG. 1b

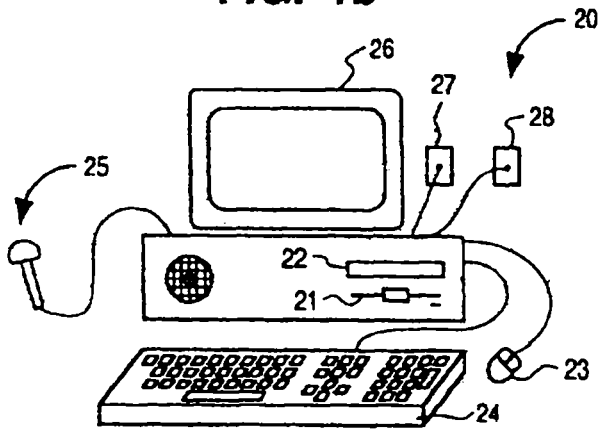

20
26
27
28
25
22
21
23
24

FIG. 2a

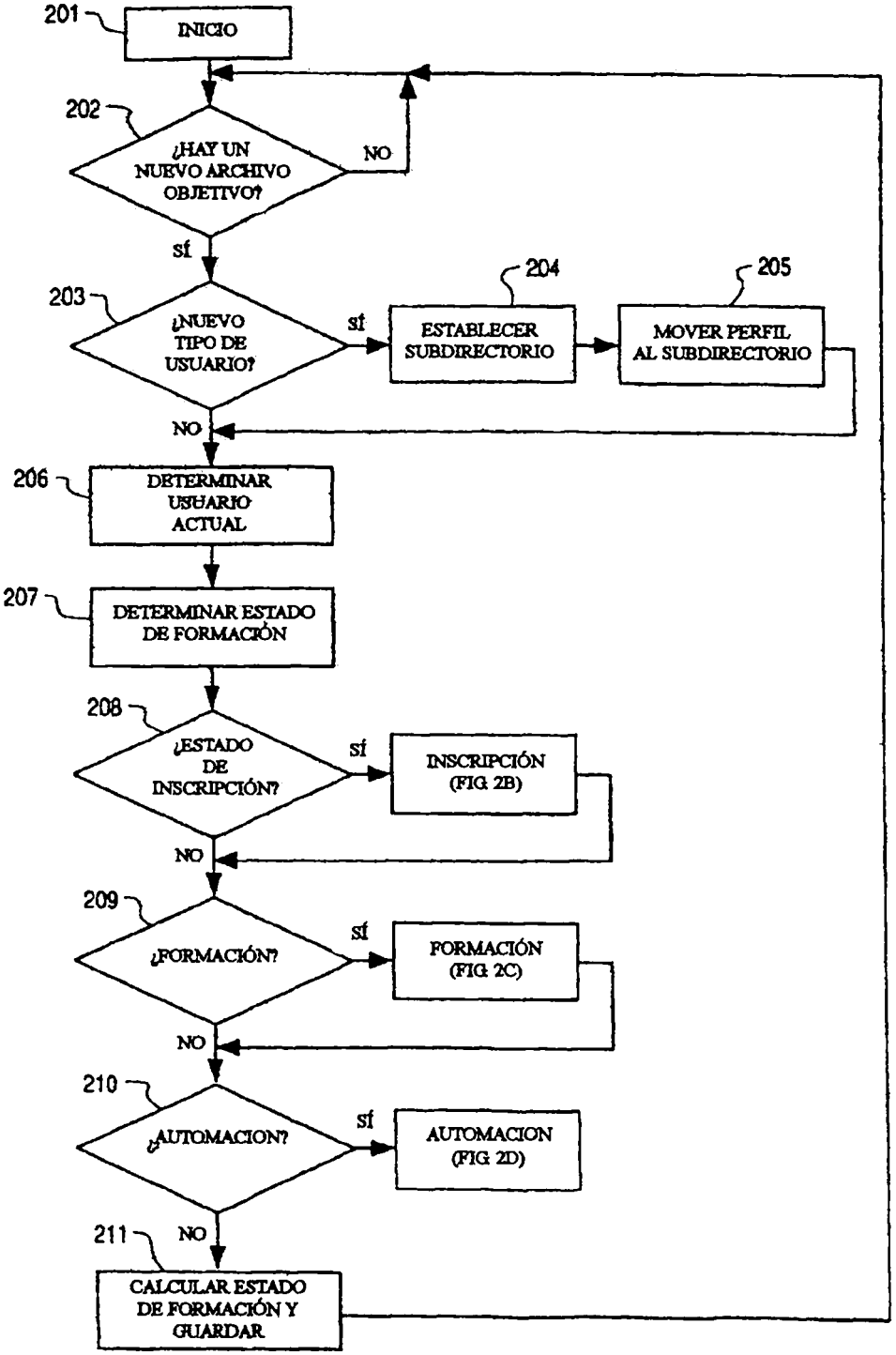

201
INICIO
202
¿HAY UN NUEVO ARCHIVO OBJETIVO?
NO
SÍ
203
¿NUEVO TIPO DE USUARIO?
SÍ
204
ESTABLECER SUBDIRECTORIO
205
MOVER PERFIL AL SUBDIRECTORIO
NO
206
DETERMINAR USUARIO ACTUAL
207
DETERMINAR ESTADO DE FORMACIÓN
208
¿ESTADO DE INSCRIPCIÓN?
SÍ
INSCRIPCIÓN (FIG 2B)
NO
209
¿FORMACIÓN?
SÍ
FORMACIÓN (FIG 2C)
NO
210
¿AUTOMACION?
SÍ
AUTOMACION (FIG 2D)
NO
211
CALCULAR ESTADO DE FORMACIÓN Y GUARDAR

FIG. 2b

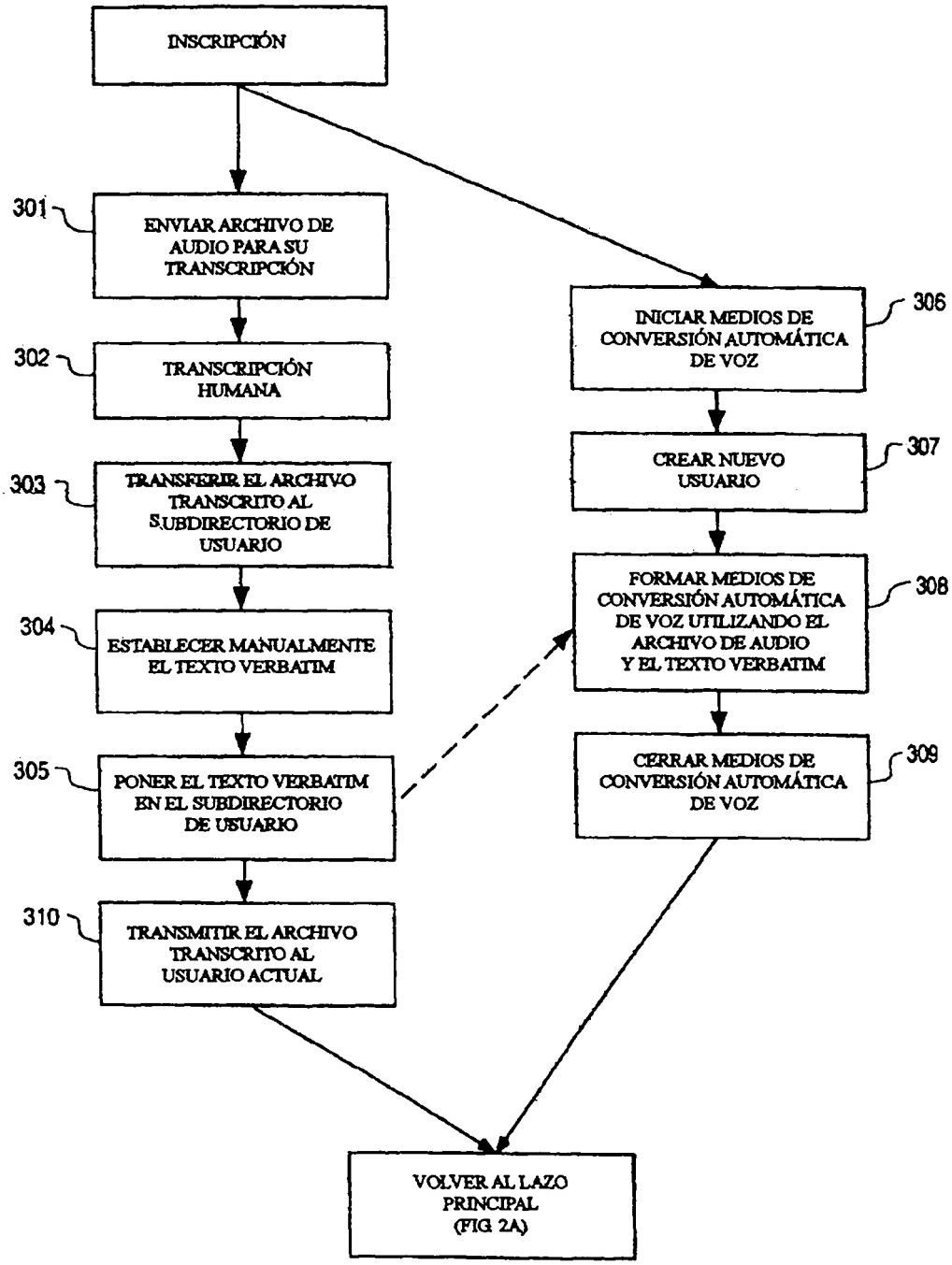

INSCRIPCIÓN
301
ENVIAR ARCHIVO DE AUDIO PARA SU TRANSCRIPCIÓN
302
TRANSCRIPCIÓN HUMANA
303
TRANSFERIR EL ARCHIVO TRANSCRITO AL SUBDIRECTORIO DE USUARIO
304
ESTABLECER MANUALMENTE EL TEXTO VERBATIM
305
PONER EL TEXTO VERBATIM EN EL SUBDIRECTORIO DE USUARIO
310
TRANSMITIR EL ARCHIVO TRANSCRITO AL USUARIO ACTUAL
306
INICIAR MEDIOS DE CONVERSIÓN AUTOMÁTICA DE VOZ
307
CREAR NUEVO USUARIO
308
FORMAR MEDIOS DE CONVERSIÓN AUTOMÁTICA DE VOZ UTILIZANDO EL ARCHIVO DE AUDIO Y EL TEXTO VERBATIM
309
CERRAR MEDIOS DE CONVERSIÓN AUTOMÁTICA DE VOZ
VOLVER AL LAZO PRINCIPAL (FIG. 2A)

FIG. 2c

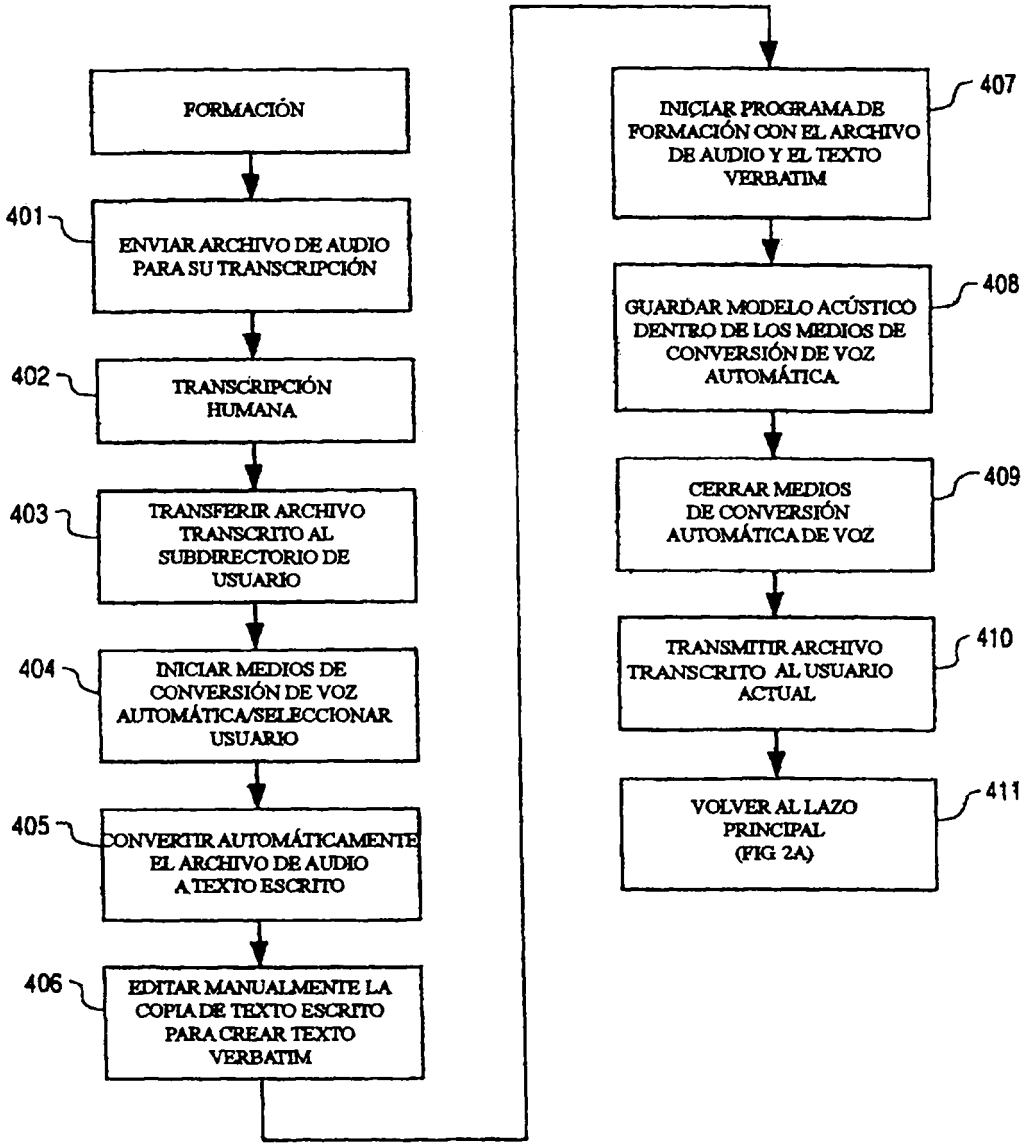
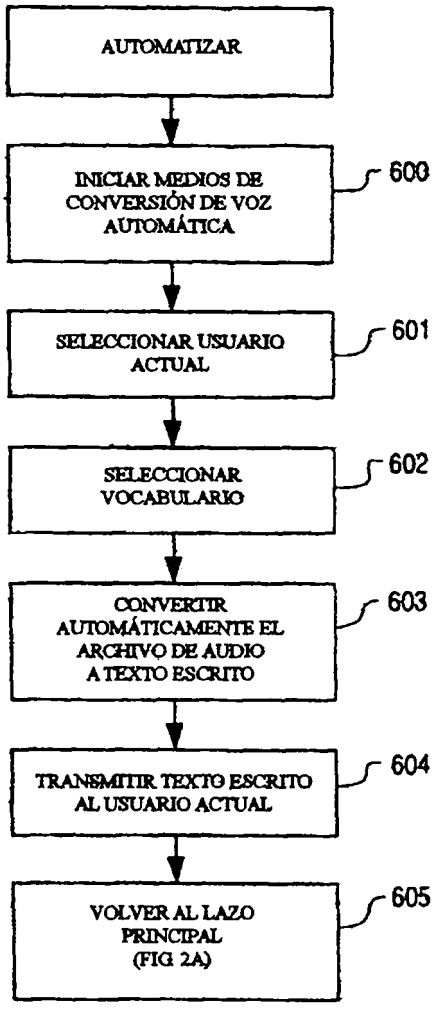

FORMACIÓN
401
ENVIAR ARCHIVO DE AUDIO PARA SU TRANSCRIPCIÓN
402
TRANSCRIPCIÓN HUMANA
403
TRANSFERIR ARCHIVO TRANSCRITO AL SUBDIRECTORIO DE USUARIO
404
INICIAR MEDIOS DE CONVERSIÓN DE VOZ AUTOMÁTICA/SELECCIONAR USUARIO
405
CONVERTIR AUTOMÁTICAMENTE EL ARCHIVO DE AUDIO A TEXTO ESCRITO
406
EDITAR MANUALMENTE LA COPIA DE TEXTO ESCRITO PARA CREAR TEXTO VERBATIM
407
INICIAR PROGRAMA DE FORMACIÓN CON EL ARCHIVO DE AUDIO Y EL TEXTO VERBATIM
408
GUARDAR MODELO ACÚSTICO DENTRO DE LOS MEDIOS DE CONVERSIÓN DE VOZ AUTOMÁTICA
409
CERRAR MEDIOS DE CONVERSIÓN AUTOMÁTICA DE VOZ
410
TRANSMITIR ARCHIVO TRANSCRITO AL USUARIO ACTUAL
411
VOLVER AL LAZO PRINCIPAL (FIG 2A)

FIG. 2d


AUTOMATIZAR
INICIAR MEDIOS DE CONVERSIÓN DE VOZ AUTOMÁTICA
600
SELECCIONAR USUARIO ACTUAL
601
SELECCIONAR VOCABULARIO
602
CONVERTIR AUTOMÁTICAMENTE EL ARCHIVO DE AUDIO A TEXTO ESCRITO
603
TRANSMITIR TEXTO ESCRITO AL USUARIO ACTUAL
604
VOLVER AL LAZO PRINCIPAL (FIG 2A)
605

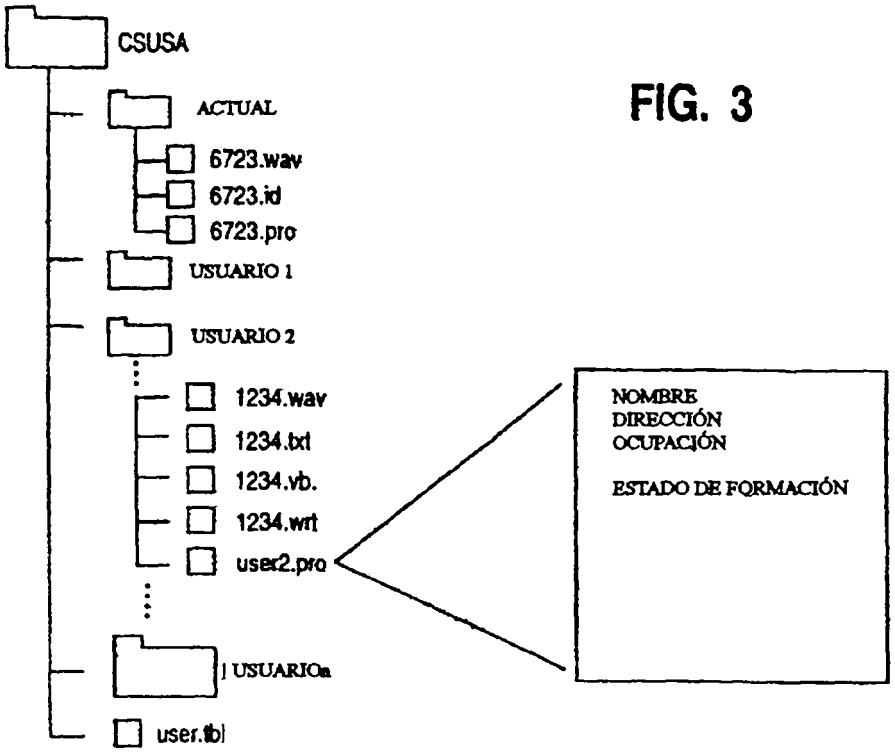

CSUSA
ACTUAL
6723.wav
6723.id
6723.pro
USUARIO 1
USUARIO 2
1234.wav
1234.txt
1234.vb.
1234.wrt
user2.pro
USUARIOn
user.tbl
NOMBRE
DIRECCIÓN
OCUPACIÓN
ESTADO DE FORMACIÓN

FIG. 3

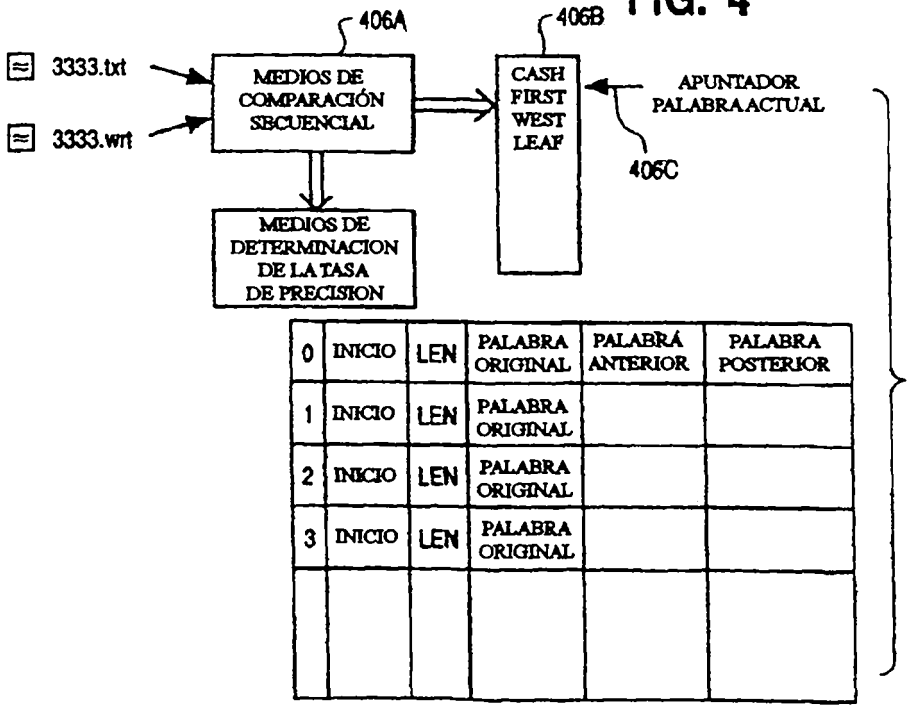

406A
406B
3333.txt
3333.wrt
MEDIOS DE COMPARACIÓN SECUENCIAL
CASH FIRST WEST LEAF
APUNTADOR PALABRA ACTUAL
406C
MEDIOS DE DETERMINACION DE LA TASA DE PRECISION
0 INICIO LEN PALABRA ORIGINAL PALABRÁ ANTERIOR PALABRA POSTERIOR
1 INICIO LEN PALABRA ORIGINAL
2 INICIO LEN PALABRA ORIGINAL
3 INICIO LEN PALABRA ORIGINAL

FIG. 4

FIG. 5

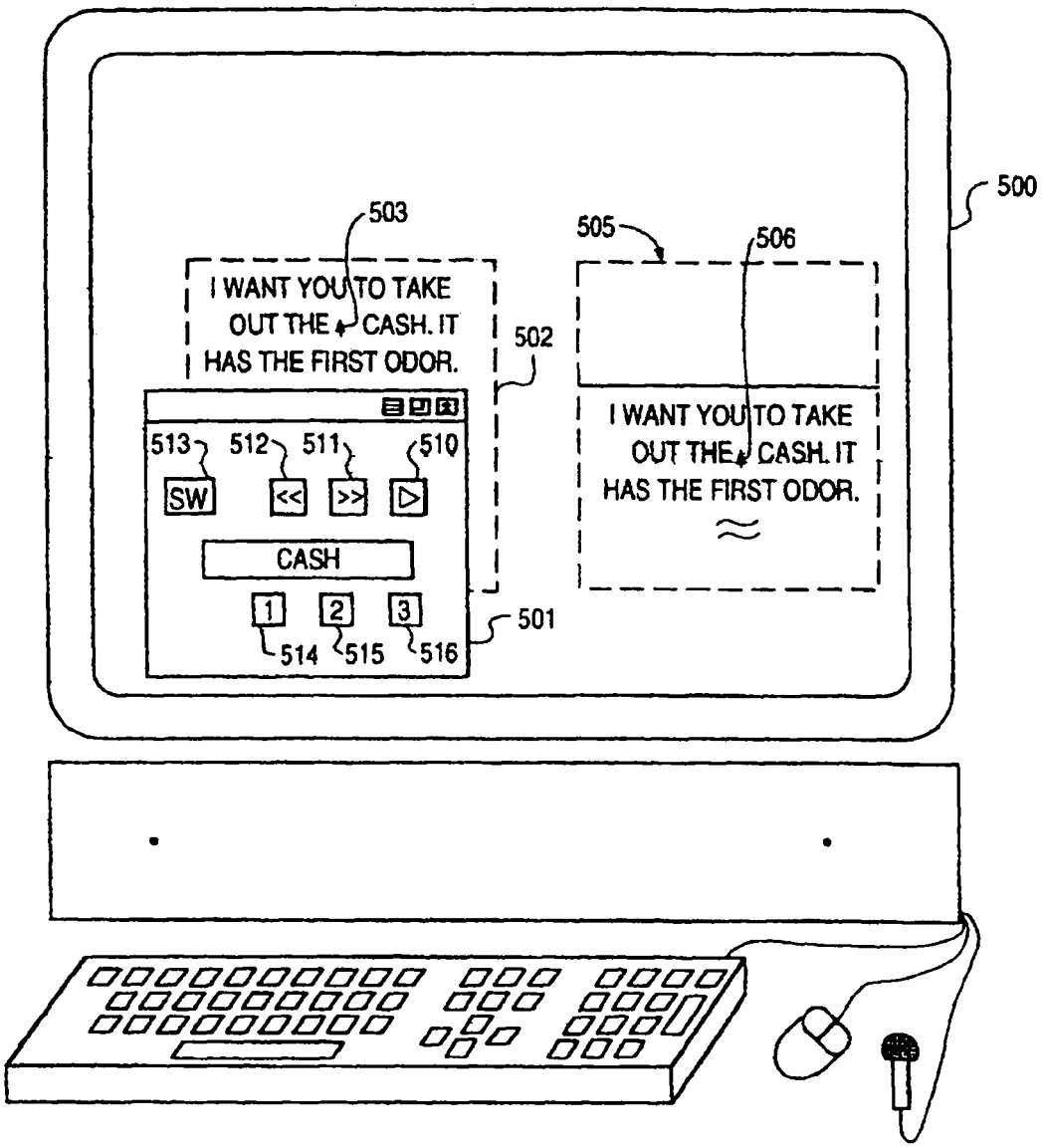
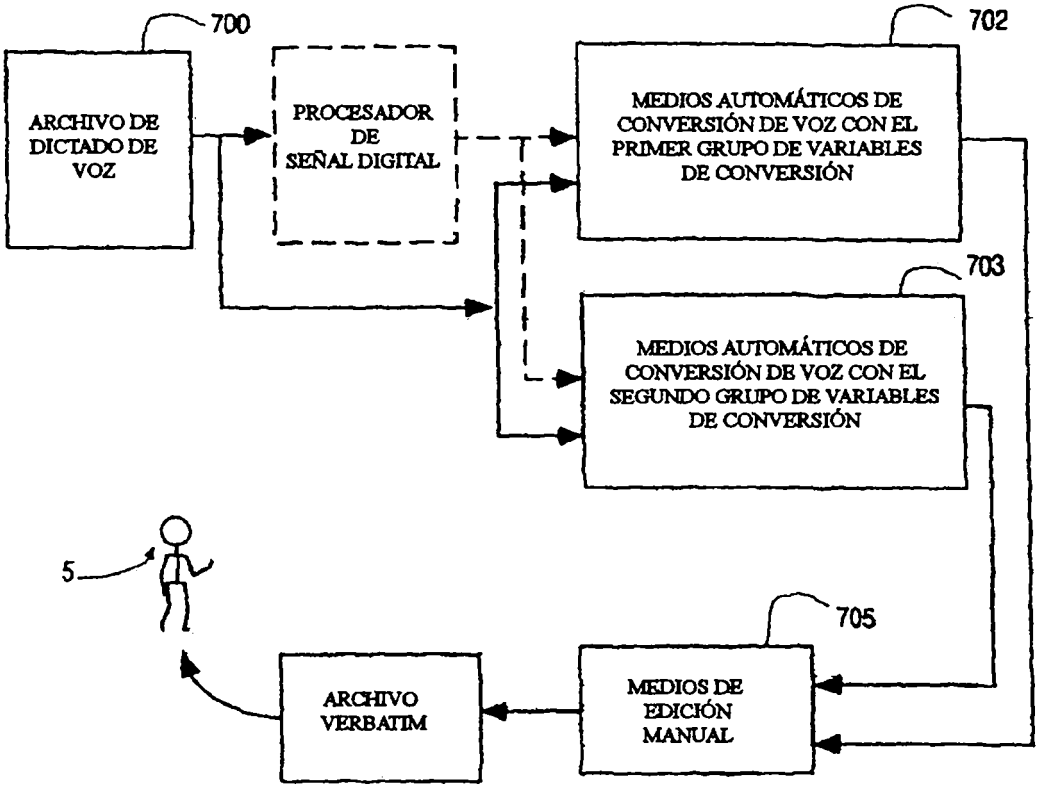

500
503
505
506
I WANT YOU TO TAKE
OUT THE CASH. IT
HAS THE FIRST ODOR.
502
513
512
511
510
SW
<<
>>
CASH
1
2
3
514
515
516
501
I WANT YOU TO TAKE
OUT THE CASH. IT
HAS THE FIRST ODOR.

FIG. 6


700
ARCHIVO DE DICTADO DE VOZ
PROCESADOR DE SEÑAL DIGITAL
702
MEDIOS AUTOMÁTICOS DE CONVERSIÓN DE VOZ CON EL PRIMER GRUPO DE VARIABLES DE CONVERSIÓN
703
MEDIOS AUTOMÁTICOS DE CONVERSIÓN DE VOZ CON EL SEGUNDO GRUPO DE VARIABLES DE CONVERSIÓN
5
705
MEDIOS DE EDICIÓN MANUAL
ARCHIVO VERBATIM