

Título: Impacto de las Nuevas Tecnologías de Reconocimiento de Voz en El Manejo, por Personas con Discapacidad Física, de Aplicaciones de Diseño Asistido por Ordenador.

Área temática: V. La superación de discapacidades físicas/marginación y exclusión sociales mediante la educación virtual

Ponentes:

Eugenio Lázaro Cañedo-Argüelles
Dpto. de Sistemas Informáticos.
Esc. Téc. Sup. de Ingeniería (ICAI)
Universidad Pontificia Comillas (U.P.Co.)
Alberto Aguilera, 23. / 28008 - Madrid.
Tfno.: 91-548-44-07; Fax.: 91-541-18-60.
e_lazaro_ca@einf.upco.es

Manuel Domínguez Somonte.
Dpto. de Ingeniería de Fabricación
Esc. Téc. Sup. de Ingenieros Industriales
Universidad Nacional de Educación a Distancia
Apdo. de correos 60149 / Madrid 28080
Tfno.: 91-398-82-32
mdominguez@ind.uned.es

1. Resumen

La evolución del diseño industrial en los últimos años ha sido ciertamente importante. Se ha pasado de las representaciones 2D en papel a visualizaciones fotorealísticas 3D, con posibilidades como: perspectivas automáticas, cambios de escala de forma continua, sección automatizada de dibujos, etc.

Todo esto se ha conseguido mediante la gran evolución que han sufrido tanto el hardware como el software de diseño, llegándose a aplicaciones bastante completas pero muy complejas.

En la actualidad aparece un nuevo reto: el ordenador debe comunicarse con el usuario mediante un lenguaje familiar al usuario, no mediante un lenguaje artificial de ordenador. Este tipo de manejo de las aplicaciones abre nuevas y prometedoras posibilidades a usuarios con capacidad limitada en el manejo de teclados, ratones, y demás dispositivos.

Con estos nuevos sistemas, la posibilidad de uso de las aplicaciones de diseño asistido por ordenador deberá concretarse mediante el uso de la voz. Debe ser posible controlar, aunque sea parcialmente, estos programas a través de órdenes verbales entendidas por el ordenador y ejecutadas con la misma precisión que con teclados, ratones, digitalizadores, etc.

Las posibilidades para operar en nuevos entornos de trabajo, como sería el de las personas con discapacidades físicas, obligan a un profundo estudio de sus posibilidades, para comprobar sus ventajas e inconvenientes.

Este trabajo pretende mostrar la situación actual del manejo de dichas aplicaciones con sistemas de voz, haciendo hincapié en sus limitaciones y posibilidades, así como las perspectivas futuras hacia las que debieran evolucionar para conseguir una mayor eficacia sin restar potencia o posibilidades de trabajo.

2. El manejo de sistemas de diseño asistido por ordenador.

Desde su desarrollo a partir de los años 70, se vio muy claro las capacidades de estos equipos en el campo del diseño. Su versatilidad y rapidez ha hecho que el clásico diseño, mediante útiles y mesa de dibujo, haya sido desplazado por estaciones de trabajo con una pantalla, un teclado, un ratón y otros dispositivos.

Dado que el nivel de informatización ha alcanzado hasta niveles de pequeñas empresas (hasta hace poco el tema de la informatización estaba limitado a empresas con grandes recursos), hoy en día no se concibe una oficina de diseño sin sus correspondientes equipos de mayor o menor potencia. Como consecuencia de ello, esta actividad ha sufrido un gran vuelco, no tanto en su contenido como en sus herramientas y metodologías de trabajo; así, se han abierto nuevas y variadas posibilidades que antes, por su complejidad o tiempo requerido, no se podían llevar a cabo.

Los problemas de manejar una determinada aplicación informática varían mucho con el tipo de ésta, pero en el caso de las aplicaciones de diseño asistido por ordenador (**DAO**), al ser muy complejas y con posibilidades casi infinitas, las dificultades se multiplican. Tanto es así que el tiempo requerido para aprender su manejo, hasta alcanzar un nivel aceptable, puede llegar a ser hasta un año, dependiendo del propio programa, las opciones de éste y el campo de trabajo en que se desarrolle la tarea.

Estas dificultades, que se dan en todo tipo de aplicaciones, han hecho que se busquen nuevas formas de comunicación con los ordenadores.

3. Reconocimiento de voz y procesamiento del lenguaje natural.

Desde el momento en que los ordenadores han llegado a ser parte importante en muchos aspectos de la sociedad actual se ha visto muy claro que mucha gente tiene serias dificultades en trabajar con ellos... y no solo las personas discapacitadas.

Con demasiada frecuencia el usuario es incapaz de determinar qué debe hacer para conseguir algo que necesita en un determinado momento. Aunque parezca mentira, el aspecto críptico de la informática no se ha perdido con el paso de los años, lo que ocurre es que éste queda enmascarado por el hecho de que los usuarios de las nuevas generaciones se desenvuelven mejor que las anteriores generaciones, ya que nacen rodeados por este entorno de máquinas a las que se habitúan. Sin embargo para personas no nacidas bajo ese ambiente a veces resulta francamente difícil trabajar con estos equipos.

Si hablamos de personas con discapacidad física... el problema se incrementa hasta niveles impensables. Pensemos en alguien que se vea imposibilitado de usar sus manos ¿cómo acceder a las posibilidades de un ordenador? ¿Cómo podría trabajar con diferentes aplicaciones? El problema es enorme pero tiene solución.

Los actuales sistemas informáticos pretenden resultar lo más “amigables” posible, pero a veces resultan poco intuitivos y requieren del usuario una formación y comprensión que a veces no están a su alcance. Como complicación adicional, la comunicación con el ordenador ha de hacerse a través de dispositivos tales como el teclado, el ratón y otros aún menos intuitivos. Si, además, el usuario presenta dificultades adicionales a la hora de manejar estos dispositivos... ¿cómo podrá trabajar?

Este panorama lleva a la necesidad de crear nuevos entornos de comunicación que sean más naturales, sencillos e intuitivos, pudiendo relacionarlos con las características del trabajo que se ha de hacer y que permitan prescindir del manejo a través de las manos. Con relación a esto y si se observan las formas de comunicación que usa el ser humano se ve claramente que el sistema conversacional es el más sencillo, claro y rápido; entonces, ¿porqué no aplicar esto a la comunicación con los ordenadores? ¿Es que no sería posible aplicar las técnicas de comunicación oral a estas máquinas?

Se disponen de micrófonos y captadores de sonido amén de analizadores de sonidos y otros artilugios; luego, ¿podría ser posible transformar los sonidos en órdenes precisas para los equipos informáticos?

En principio no parece una labor demasiado difícil dada la posibilidad de distinguir unos sonidos de otros; sin embargo, lo que se desea no es algo tan simple como el reconocimiento de unos cuantos “sí” o “no”, sino tener un completo abanico de posibles comandos compuestos de frases, cuyo significado debe comprender el ordenador; éste debe entender lo que se le dice y responder a ello ejecutando la correspondiente orden.

No hay que olvidar que el ser humano es capaz de decir lo mismo (en la misma lengua) de múltiples formas. Ante esto el ordenador debe reaccionar ejecutando lo que se le indica, sea con una u otra frase. Se podría decir que lo que se busca no es ya que el ordenador ejecute lo que “se le dice” sino “lo que se le quiere decir”. O sea, debe ser capaz de “comprender” la orden dada... y obrar en consecuencia.

El campo de la informática controlada mediante la voz abre expectativas nuevas que van a revolucionar su operatoria en un futuro no muy lejano, permitiendo prescindir de teclados y ratones; también, va a permitir, por un lado habilitar el uso para personas con discapacidades físicas, y por otro desprenderse de la atadura de un puesto físico fijo permitiendo una movilidad hasta ahora impensable. Realmente se podrá llegar a usar un teléfono móvil para ejecutar tareas a distancia desde cualquier punto. Es más, adelantándose al futuro, lo lógico será que el ordenador, mediante un sintetizador de voz, conteste verbalmente, ya sea a las preguntas que se le hagan, ya sea a las tareas, para confirmar su ejecución.

Este nuevo entorno facilitará labores que podrán organizarse y controlarse al mismo tiempo que el usuario trabaja sobre otras actividades y, lo que es más importante, pondrá la informática más accesible a determinados colectivos de discapacitados (invidentes y minusválidos físicos). Si se tiene en cuenta que el ordenador puede controlar multitud de tareas en los entornos actuales (viviendas, oficinas, fabricas, almacenes, etc.), no es necesario empezar a imaginar sus posibilidades.

En cada campo habrá que distinguir los diferentes tipos de trabajo que se pueden hacer: administrativo, ejecutivo, investigador, operador, programador, diseñador, etc. Existen tantas posibilidades como aplicaciones se han ido creando... y son muchas. Así, no es lo mismo un puesto administrativo dedicado a la escritura de cartas y memorándums que un puesto de programador; o un puesto de operaciones financieras en la bolsa que un diseñador gráfico. Cada tipo de trabajo va a presentar requisitos de funcionamiento muy diferentes, con características marcadas por las necesidades del trabajo.

En el caso de un diseñador habida cuenta de los cientos de ordenes que permiten los programas de diseño, se debería poder ejecutar estas pidiéndoselo verbalmente al ordenador la mayoría de las cosas.

Es decir, la problemática habrá de ser estudiada en cada caso, viendo sus necesidades y posibilidades. En determinadas circunstancias se puede llegar a la conclusión de que el uso de órdenes verbales puede dificultar, en lugar de facilitar, el trabajo; si así fuera habría que prescindir de este entorno y buscar otro.

Vamos a ver someramente lo que son cada una de las aplicaciones por separado (Reconocimiento de Voz y Diseño Asistido por Ordenador) y, seguidamente, el conjunto de ambas. Vamos a ver las ventajas e inconvenientes que supone, así como las posibilidades que permiten. Por último trataremos de dar un salto al futuro para “ver” que se haría (o se debería hacer) para conseguir ese entorno amigable e intuitivo que debiera extenderse a todas las facetas de la informática, independientemente de las dificultades del trabajo en que se aplique.

4. Aplicaciones de diseño asistido por ordenador.

4.1. Definición y características.

Son programas informáticos que permiten trabajar en el campo del diseño mediante equipos informáticos. Estos equipos sustituyen el dibujo clásico de líneas y curvas sobre el papel con un trazado sobre la pantalla de un ordenador, con todas las posibilidades que ello implica.

Su sistema de trabajo es mediante la creación de los diversos elementos gráficos que componen el diseño. Estos elementos van creándose en función de las necesidades y suelen ser muy sencillos: puntos, líneas, rectángulos, arcos, elipses, polilíneas, etc. que se van añadiendo en “tablero” de dibujo informático. Cada uno de ellos tiene sus propias características (algunas comunes y otras no), como son: posición (en forma de coordenadas), color, grosor, estilo, etc. El conjunto de elementos sencillos puede agruparse en grupos y subgrupos de manera que se van creando elementos cada vez más complejos.

La manipulación de los diversos elementos es, gracias a la informática, muy sencilla; se puede cambiar su posición, tamaño, aspecto, color, etc. Los grupos que se hayan creado pueden manejarse en su conjunto o “desagrupándolos”, se pueden repetir en diferentes posiciones, mover, etc. Las posibilidades son casi infinitas.

Una característica importante en este tipo de programas es que los elementos se definen mediante sus características (tipo, posición, aspecto, etc.), de forma que al guardar cada proyecto (que es como se define cada trabajo individual) lo que se almacena son realmente sus características; esto implica que a la hora de realizar modificaciones de ampliación, movimiento y reducción de tamaño el margen de juego, sin variar la calidad, es inmenso; en el caso del otro tipo de dibujos (por píxeles) la calidad de presentación en pantalla puede ser mejor (y esto, con los actuales ordenadores y las nuevas técnicas, es discutible), pero sus posibilidades en cuanto a cambios de escala, posicionamiento y tamaño están muy limitadas.

Una vez terminado el diseño, éste se guarda como una unidad de la que se pueden extraer cuantas vistas se quieran, secciones, perspectivas, con diversas escalas, añadiendo sombreados, transparencias, cortes, etc. Es decir la posibilidad de obtención de información es enorme, pudiendo adaptarlas a las necesidades específicas del momento.

La posterior salida de la información a través de dispositivos de impresión de gran tamaño (trazadores) completa unos sistemas de diseño muy versátiles.

4.2. Manejo.

Este tipo de aplicaciones, instaladas en ordenadores con entorno gráfico (sistemas Windows y similares), suelen manejarse mediante menús y con la ayuda del ratón. Hasta hace poco también se usaba el lápiz digitalizador, aunque éste ya se encuentra prácticamente fuera de uso. Así, para dibujar una línea es posible escoger una opción en el correspondiente menú y, luego, mediante el ratón señalar entre qué puntos se desea trazar, o directamente con los diferentes botones del ratón trazarla señalando puntos y tipos de opciones.

También es posible trabajar desde el teclado usando comandos aunque de forma limitada; para ello las aplicaciones suelen ofrecer una ventana en la que es posible ir escribiendo comandos (con una sintaxis muy exacta), que se ejecutan mediante la tecla "Intro". Es más, la mayoría de las opciones que se ejecutan mediante los menús y el ratón, tiene su reflejo en la ventana de comandos, en la que va apareciendo su comando equivalente.

La multitud de comandos u opciones que presentan estas aplicaciones hace que requieran una formación de cierto nivel incluso para usuarios tipo estudiante. Además conforme crece la potencia de los nuevos ordenadores, crece la complejidad de la aplicación al incorporarse nuevas posibilidades y efectos muy sofisticados como: texturas de superficie, iluminación variable, sombreado, reflejos, etc.

Estos programas incorporan diversas formas de ayuda interactiva muy variada. Ésta suele servir de consulta para temas puntuales y nunca para aprendizaje (Figura 1). Otra forma de ayuda suele ser mediante posibles consultas a través de Internet a unas direcciones fijadas de antemano. Además se incorporan los denominados "tutoriales". Estos suelen ser pequeños programas que sirven para ejercitar y aprender el uso mediante diversos temas o capítulos que el alumno va siguiendo paso a paso, si bien su nivel no es demasiado elevado y van dirigidos, mayoritariamente, a usuarios de muy bajo nivel.

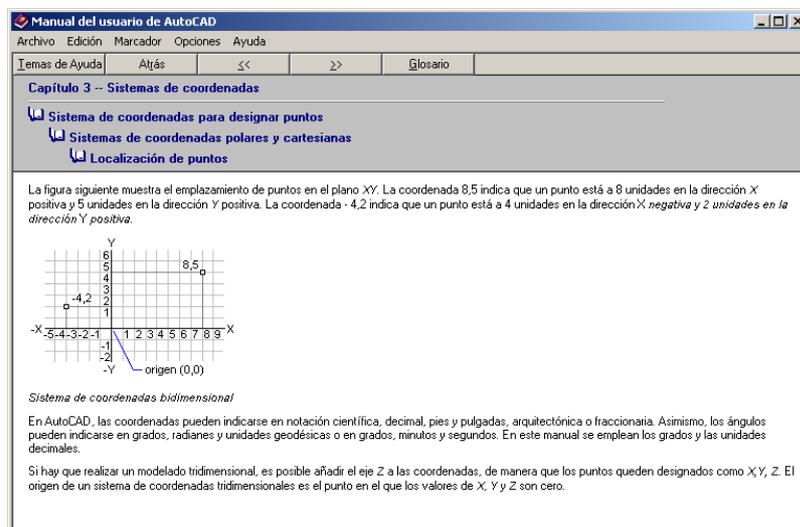


Fig. 1 -Una pantalla de ayuda de Autocad R.14®.

4.3. Requisitos de hardware.

Los requisitos de hardware para trabajar con este tipo de programas son bastante exigentes:

- Un potente microprocesador con una buena velocidad de trabajo.
- Abundante memoria RAM lo más rápida posible. En este aspecto trabajar con menos de 128 Mbytes se hace muy incomodo.
- Espacio abundante en unidades de almacenamiento (discos duros y CD-ROM's). Hay que tener en cuenta que cualquier proyecto de diseño suele implicar gran cantidad de elementos y el volumen de cada proyecto puede ocupar bastantes Mbytes.
- Una buena pantalla, a ser posible de gran tamaño (trabajar con una de 14 pulgadas podría resultar de chiste); su calidad de resolución debe ser alta y la velocidad de refresco igual. Por supuesto para ello se requeriría una "excelente" tarjeta gráfica incorporada al ordenador.
- Como requisitos adicionales habría que añadir la posibilidad de usar periféricos específicos para imprimir como son los modernos "trazadores", ya sea para uso privado del ordenador de trabajo, ya sea para diversos puestos mediante acceso por una red

4.4. Conclusiones con relación a las aplicaciones de diseño asistido por ordenador.

En el momento actual es cierto que la mesa clásica del delineante prácticamente ha desaparecido; en su lugar hay un magnífico y potente ordenador con una unidad de impresión. Este ordenador, además, se encuentra conectado en red a una Base de Datos gráfica en la que se almacenan todos los "planos" de la empresa, pudiendo consultarse, modificarse y extraer información seleccionada.

Esta es la realidad actual y el punto de partida a nuevas posibilidades que se pueden desarrollar como: el acceso mediante Internet a la información gráfica, creación de prototipos tridimensionales mediante equipos de "Generación Rápida de Prototipos", etc.

Entonces ¿hacia donde debe evolucionar este campo de "diseño informático"? ¿Las cosas se van a reducir a buscar entornos más "amigables" únicamente?

No parece que los cambios vayan a reducirse a eso, sobre todo teniendo en cuenta lo que implican las nuevas aplicaciones de "Reconocimiento de Voz" y de "Procesamiento del Lenguaje Natural", en especial esta última cuyas características vamos a tratar de ver en el siguiente punto.

Si se habla de inteligencia artificial (y parece crudo eso de hablar de ordenadores inteligentes y cosas así) se podría entender que lo que se busca es que el ordenador haga nuestro trabajo, no que tome nuestras decisiones; en este aspecto sería importante conseguir que la máquina "entendiera" (puesto entre comillas) lo que se le quiere decir, no lo que se le dice. Es ahí donde estaría la auténtica "inteligencia" de la máquina. Todo esto aplicado al mundo del diseño conduciría a un manejo mucho más sencillo del ordenador.

Es curioso que antes de la aparición de los ordenadores, para ser un buen delineante antes se requiriese únicamente buenos conocimientos de dibujo y diseño gráfico; ahora eso no basta, se requiere un conocimiento profundo de la informática y sus dispositivos. Las dificultades aumentan en lugar de reducirse. No deja de ser curioso este cambio, se busca mayor eficacia, nuevas posibilidades, más información, pero se consigue aumentando la dificultad de manera exponencial. A esto hay que añadir una velocidad en los cambios que hacen muy difícil estar al día continuamente.

Hay que volver a la sencillez de antaño (sin sacrificar las mejoras conseguidas) y esto puede hacerse a través de ese nuevo entorno que es el reconocimiento de voz.

5. Aplicaciones de reconocimiento de voz.

5.1. Definición y características.

Son programas informáticos que permiten trabajar con el ordenador relacionando el lenguaje oral con los comandos del ordenador. Permiten sustituir las interfaces como el teclado y el ratón por un micrófono, a través del cual se pueden dictar las correspondientes órdenes al sistema. Esto implica la posibilidad de usar la voz en los diferentes aspectos del trabajo en que se usa el ordenador.

Este control mediante la voz puede ser parcial o total. La primera es la más habitual y la segunda podría ir dirigida a que personas discapacitadas físicas pudieran usar los ordenadores sin limitaciones impuestas por sus propias características.

El tema del reconocimiento de voz lleva poco tiempo desarrollándose. En el año 1996 IBM presentó, ya en fase de comercialización, el primer programa aunque no con demasiado éxito. Al año siguiente sacó una versión mejorada que prácticamente sería el actual ViaVoice® de IBM aunque aún demasiado lento. A partir de entonces el número de aplicaciones ha crecido exponencialmente y ya existen en el mercado gran cantidad de posibilidades de elección con precios muy variados.

5.2. Estructura de las aplicaciones de Reconocimiento de Voz.

Para estudiar estas aplicaciones conviene distinguir entre dos facetas muy conectadas entre sí pero absolutamente diferenciadas; lo que los profesionales llaman: **Reconocimiento de Voz (RV)** y **Procesamiento del Lenguaje Natural (PLN)**. Realmente son dos fases muy diferenciadas de lo que sería el proceso total; las dos están muy relacionadas pero trabajan de manera diferente con objetivos diferentes.

5.2.1. Reconocimiento de Voz (RV)

En inglés se denomina Voice or Speech Recognition (o con sus siglas **SR**): es la fase de la aplicación que recibe los sonidos, mediante el clásico y conocido micrófono (de mejor o peor calidad) transformándolos, luego de un largo proceso interno, en conjuntos de palabras conocidas, o sea, en una presunta frase. Esta frase puede tener o no sentido, pero es la interpretación que el módulo hace de los sonidos recibidos vía micrófono.

5.2.2. Procesamiento del Lenguaje Natural (PLN)

Este segundo módulo, denominado en inglés Natural Language Processing (o con sus siglas **NLP**) recibe la frase captada por el módulo **RV** y trata de “comprender la frase”; así se puede poner un ejemplo.

Se dicta la frase:

Esta parte recibe los sonidos mediante un micrófono

Sin embargo el módulo (por defectos en la pronunciación del usuario) obtiene:

En esta parte recibe y los sonidos mediante un micrófono

Ahora el módulo **PLN** debe recoger la frase y “comprender” que la partícula “y” sobra; debe interpretar la frase con absoluta claridad ya sea comenzando “En esta parte...” o “Esta parte...”

Es decir, lo que recibe es un juego de palabras con más o menos sentido y lo que devuelve es una frase con un sentido preciso y real. Estas frases se forman mediante un análisis basado en el contexto de trabajo y en patrones gramaticales y morfológicos de la conversación.

La realidad es que ambos módulos no trabajan de manera independiente sino que se relacionan entre sí solicitándose mutuamente aclaraciones. Así, el módulo **RV** puede solicitar aclaraciones al **PLN** para entender ciertas palabras cuyo sonido es prácticamente el mismo; a su vez, el módulo **PLN** seleccionara la palabra analizando el contexto y la morfología de la frase. De la misma manera el módulo **PLN** puede consultar al **RV** pidiendo aclaraciones sobre la enfatización de determinadas palabras en una determinada frase. por ejemplo:

; Eso no me gusta !

La distinta entonación que se hace de las diferentes palabras podría conducir a dos interpretaciones sutilmente diferentes: se puede enfatizar un disgusto en general o algo específico que no nos gusta. El sentido habría que extraerlo del entorno de la conversación (lo debería hacer el sistema de reconocimiento de voz).

Todo esto no hace sino mostrar las serias dificultades que presentan los programas de reconocimiento de voz, pues han de hacer hincapié no ya en reconocer las palabras sino en la comprensión de frases.

5.3. Módulo de Reconocimiento de voz (RV).

La forma de trabajar de este es la siguiente. Se comienza por una recogida del sonido con todas sus características: frecuencia, intensidad, modulación, etc.

Estos sonidos se analizan mediante diversos algoritmos (hay muchos y cada programa usa uno u otro) en los que intervienen el cálculo de la frecuencia y de la intensidad hasta reducir el sonido completo (el juego de ondas) a un conjunto de sonidos básicos (fonemas) que son la base de la dicción. Estos fonemas se conjuntan hasta formar “palabras”, palabras que se van a consultar en la Base de datos que incorpora el programa, seleccionando aquella que se parezca más al conjunto de sonidos captados.

Obsérvese la dificultad en la captación teniendo en cuenta que una misma palabra no la emite igual un adulto o un niño; si añadimos las diferentes entonaciones entre diferentes regiones, personas con mala pronunciación, intensidad variable de la voz con los sentimientos, etc. el problema se multiplica de manera exponencial.

Sin embargo superadas todas las dificultades se consigue una frase primaria que se traspaasa al siguiente módulo.

5.4. Módulo de Procesamiento del Lenguaje Natural.

Una vez recibida la frase desde el módulo **RV** hay que darle un significado. Aquí los problemas crecen aún más. A parte de todas las reglas morfológicas y gramaticales, el programa debe tener en cuenta el contexto en que se trabaja para decidir qué frase tiene sentido (o cual es el auténtico sentido de la frase). No se debe perder de vista la dificultad de interpretación de una frase cuando su idea puede ser dicha de diferentes formas por diversas personas; traducir al ordenador todo el conjunto de emociones que producen unos sonidos hasta llegar a su correcta interpretación parece una labor inalcanzable.

5.5. Una conclusión.

Todas estas dificultades hacen que la principal característica de los programas actuales de reconocimiento de voz sea que solo trabajan con el primer módulo; es decir interpretan sonidos para dar lugar a palabras. Este reconocimiento se hace mediante los sonidos emitidos al hablar por el micrófono; el conjunto de sonidos que forman una palabra es buscado en un diccionario que incorpora la aplicación, el más similar acústicamente hablando, señala la palabra seleccionada.

Esto implica que hay sonidos (correspondientes a palabras del usuario) que no son reconocidos, ya sean por mala pronunciación o por no tenerlas almacenadas, dando como resultado que el programa la sustituye por otra, normalmente la que acústicamente se parezca más. Por supuesto es posible corregir los errores y almacenar palabras nuevas, así como “educar” al sistema para que el reconocimiento sea cada vez más preciso.

6. Manejo de los sistemas actuales de reconocimiento de voz.

Dado el todavía nivel alcanzado en este campo, el trabajo con estos sistemas se reduce a dos formas diferentes: como dictado o como comandos. Se salta entre ambas formas de trabajo en función de las necesidades del usuario. La distinción ha de hacerse mediante el micrófono, de manera que o se dicta texto “puro y duro” o se dicta un comando. En función de ello el programa se limitará a escribir “algo en algún sitio” o tratará de ejecutar alguna orden.

La diferenciación entre ambas formas de trabajo puede hacerse de maneras diferentes según el sistema que se use; así en algunos de los sistemas, hay una orden específica para comenzar a dar órdenes (valga la redundancia), mientras que en otros se diferencia por la forma de dictar.

Por ejemplo, la siguiente frase “Hacer clic en Inicio” puede ser interpretada de dos formas:

- 1- Como texto a escribir en un documento.
- 2- Como orden para desplegar el menú **Inicio** de **Windows**.

Para la primera forma, se debe dictar diferenciando claramente las palabras, mientras que para la segunda la frase se dicta seguida sin interrupciones, de esta forma el sistema lo considera una orden y su resultado es desplegar el menú **Inicio** para seleccionar alguna opción.

Frases de este estilo hay muchas y así se tienen: “Hacer clic en Programas”, “Presionar Tabulador”, “Nueva línea”, etc. Cada una de ellas marca una orden para el ordenador que puede ejecutarse independientemente del contexto de trabajo del usuario.

Poniéndonos en el contexto de una aplicación de diseño asistido por ordenador habría frases como: “Hacer clic en Dibujo”, “Línea”, “Inicio coma centro coma fin”, “100,100”, “50”, etc. De estas, algunas serían opciones seleccionadas en los menús y otras se interpretarían como valores que irían apareciendo en la ventana de comandos de la aplicación.

Sin embargo, cuando se trata de seleccionar opciones u órdenes cuyos términos de los menús no las recoge el diccionario del sistema de reconocimiento de voz se presentan serias dificultades, requiriendo un previo “aprendizaje” del sistema para poder reconocer esas voces. Así, el término “no AutoCAD” no es reconocido hasta en tanto el usuario no lo incorpore al diccionario personal del sistema de reconocimiento de voz.

Si los términos, encima, están en inglés (como ocurre con algunas aplicaciones) o se usan juegos de caracteres específicos que no corresponden con ninguna palabra, como por ejemplo “Objeto OLE”, las dificultades crecen considerablemente.

6.1. El mercado.

En la actualidad existen en el ámbito comercial diversos programas de este tipo aunque reúnen una serie de características comunes. Quizás los más comunes sean el **ViaVoice® Gold** de **IBM** (Figura 6) y **Dragon Naturally Speaking®** de **Dragon Systems, Inc.** aunque hay otros como **IN CUBE Voice Command®** de **Command Corp, Inc.**

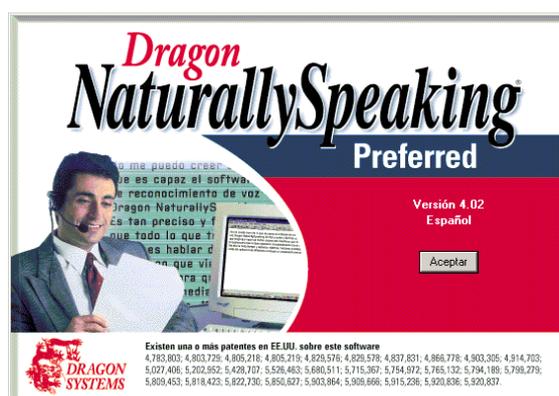


Figura 2 – Pantalla de Dragon Naturally Speaking®.

Todos ellos son parecidos con requisitos hardware relativamente fuertes aproximadamente así:

- Mínimo: processor Pentium® 133 MHz - compatible Windows® 95, PCWindows® 98, Windows NT® 4.0 o Windows 2000®
- Memoria: mínimo 64 Mbytes (es posible hacer pruebas con 32 Mbytes pero el trabajo se hace muy incómodo). Si se trabaja con sistemas operativos de red, tipo Windows NT posiblemente se requiera más memoria aún.
- Disco duro: dependiendo de las utilidades que se instalen unos 100 Mbytes libres, aunque pueden llegar a necesitarse hasta 250 Mbytes.
- Tarjeta de sonido estándar de 16-bit o similar.
- Altavoces requeridos para la ayuda multimedia interactiva y tutoriales.

Estos programas incorporan diversas utilidades, además de tutoriales para aprendizaje, y su correspondiente micrófono y auricular de cabeza que permiten trabajar con más eficacia.

6.2. Conclusiones con relación a los sistemas de reconocimiento de voz.

Este tipo de aplicaciones se encuentran ya en una fase de comercialización sin que se haya alcanzado realmente un estado definido como “Reconocimiento Conversacional”.

Algunas de estas aplicaciones hechas en plan de prueba, para entornos muy determinados (que hay que señalarlos al comienzo) permiten augurar resultados prometedores, pero la potencia de los equipos en los que se han probado es superior a la habitual de los PC, de hecho estos sistemas trabajan en entornos de arquitectura distribuida con equipos de gama media-alta.

Sin embargo se abren maravillosas expectativas y las grandes empresas de desarrollo admiten que es el futuro de la informática (en todos sus campos) va por ahí. En una nota del ABC, (1 de febrero de 2000) se afirma que “Los dos grandes magnates de la industria del ocio y la información, Stephen Case - impulsor de la fusión AOL/Warner - y el fundador de Microsoft, Bill Gates, ... admiten que el próximo punto de investigación ... será los sistemas de reconocimiento de voz”.

7. Uso de aplicaciones de reconocimiento de voz (RV) en el entorno del diseño asistido por ordenador.

Hasta aquí se han descrito de manera somera ambas aplicaciones, “reconocimiento de voz” y “diseño asistido por ordenador”, así como sus principales características. Sin embargo el manejo de ambas al tiempo presenta unos aspectos muy específicos que es aconsejable analizar.

Así se va a estudiar el funcionamiento de ambas aplicaciones en paralelo, estudiando sus ventajas e inconvenientes; en qué aspectos se facilita el trabajo y en cuáles no lo facilita o su manejo no supone una ventaja sustancial.

7.1. Inicio del programa.

Las características de la aplicación que controla el sistema de reconocimiento de voz variarán dependiendo del fabricante, nosotros vamos a hacer referencia a la de **Dragon Naturally Speaking**, versión **4.02 Preferred**.

Este sistema permite trabajar en forma de dictado al tiempo que admite una larga lista de órdenes que permiten manejar **Windows** con bastante amplitud.

Dado que el sistema de reconocimiento de voz no va a reconocer en principio el nombre de la aplicación de diseño, es necesario que añadamos ese término, escribiéndolo y vocalizándolo para su almacenamiento en el archivo personal de voces. Para ello, una vez arrancado el sistema hay que entrar en la opción que permite añadir términos y entrenarlos con la voz (véase la figura 3).

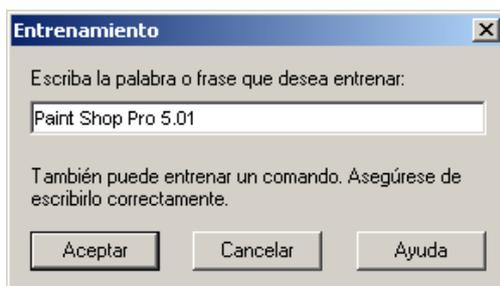


Figura 3 – Entrenamiento de nuevos términos.

La grabación del nuevo término se hace mediante un cuadro de diálogo que muestra el texto y presenta un botón para comenzar la grabación (véase la figura 4).

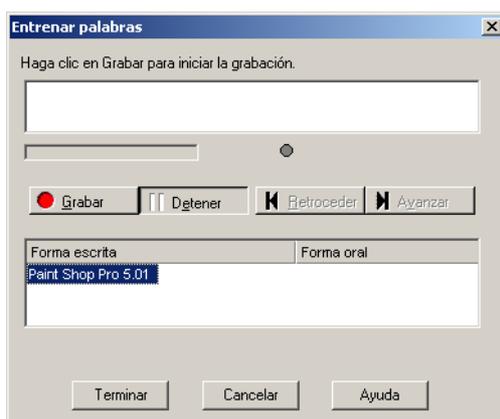


Figura 4 – Grabación de nuevos términos.

Una vez finalizada el sistema busca en su diccionario y si no encontrase alguna de las palabras (por ejemplo “Paint”) requiere que, previamente sea añadida a la lista. Esto se hace con otro cuadro de diálogo (véase la figura 5) que permite incluir términos que no están en el diccionario.

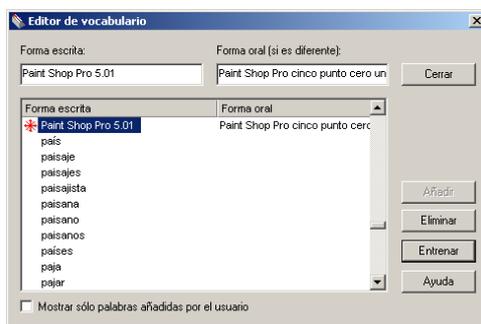


Figura 5 – Editor de vocabulario.

A partir de aquí el nuevo término respondería al llamarlo mediante la voz “Hacer clic en Paint Shop Pro Cinco Punto Cero Uno” poniéndose en marcha este programa.

7.2. Trabajando a través del reconocimiento de voz.

Si el término “AutoCAD” es añadido, para ponerlo en marcha basta con dar las siguientes órdenes verbales.

Hacer clic en menú Inicio

Hacer clic en AutoCAD

Una vez en el programa de diseño, es posible manejar los diferentes menús, sus opciones y subopciones, así como insertar datos numéricos (para coordenadas de los elementos del dibujo) hasta llegar a dibujar lo que se desee. Es decir, se tiene acceso a los menús, sub-menús y opciones de estos mediante órdenes verbales. También se puede controlar los cuadros de diálogo de las opciones que los tengan. Algunas de las órdenes son de tipo “estándar”:

Nueva línea: supone pulsar el correspondiente botón Aceptar.

Escape o Cancelar: suele ser el botón Cancelar.

También hay órdenes para manejar listas y diversas teclas:

Presionar Tabulador <número>: indica cuantas pulsaciones de Tab debe ejecutar (cosa muy interesante para moverse por los diversos elementos de los cuadros de diálogo).

En determinados momentos es posible variar el modo de trabajo de manera que lo que se dicte pueda ser interpretado como texto, por ejemplo, en el caso de dar un nombre de archivo para guardar un documento, o en el caso de acceder a la ventana de comandos de un programa de diseño para indicar datos (coordenadas de un punto o cosas así). Para ello el programa de Reconocimiento de Voz incorpora órdenes específicas:

Letras: hace que se pueda dictar letra por letra (para dar nombres de archivos y cosas así).

Números: lo mismo que el anterior pero para ir dictando dígitos.

Así cuando se ejecuta la orden de dibujar una línea en un programa como AutoCAD R.14, se pronunciarían las siguientes órdenes (se parte del funcionamiento normal mediante comandos):

- **Hacer clic en Dibujo:** se accede al menú correspondiente y se despliega.

Cada orden dada verbalmente y que haya sido reconocida por el sistema aparecerá en un pequeño recuadro de fondo amarillo (véase la figura 6).

- **Hacer clic en Línea** (o simplemente **Línea**): se escoge la opción del menú Dibujo.
- **100 coma**: para poder dictar el primer dato de las coordenadas del primer punto.
- **Sin espacio 100**: esto añade el valor 100 detrás de la coma sin ningún espacio.
- **Nueva línea**: ejecuta el primer punto.
- **200 coma**: para poder dictar el primer dato de las coordenadas del segundo punto.

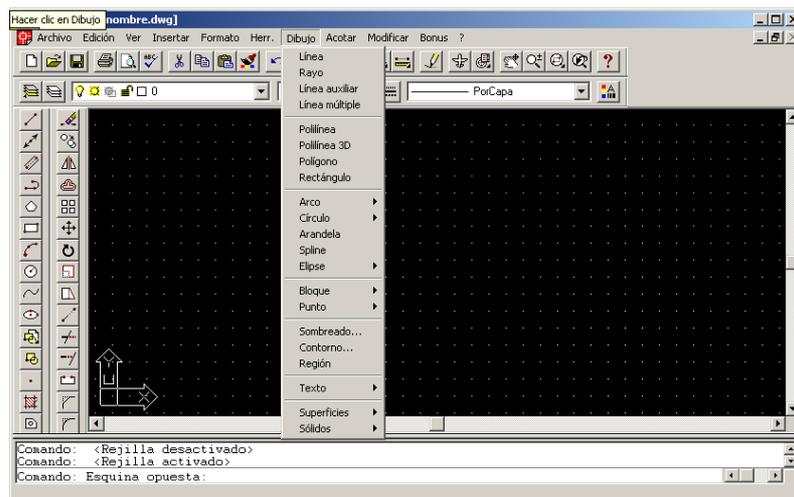


Fig. 6 - Orden “Hacer clic en Dibujo” en la barra de título de AutoCAD®

- **Sin espacio 100**: esto añade el valor 100 detrás de la coma sin ningún espacio como final de las coordenadas del segundo punto.
- **Nueva línea**: dibuja la línea.
- **Presionar escape**: termina el dibujo de la línea.

El resultado de esta pequeña prueba puede verse en la figura 7, con su reflejo en las órdenes que aparecen en la ventana de comandos de AutoCAD® 14.

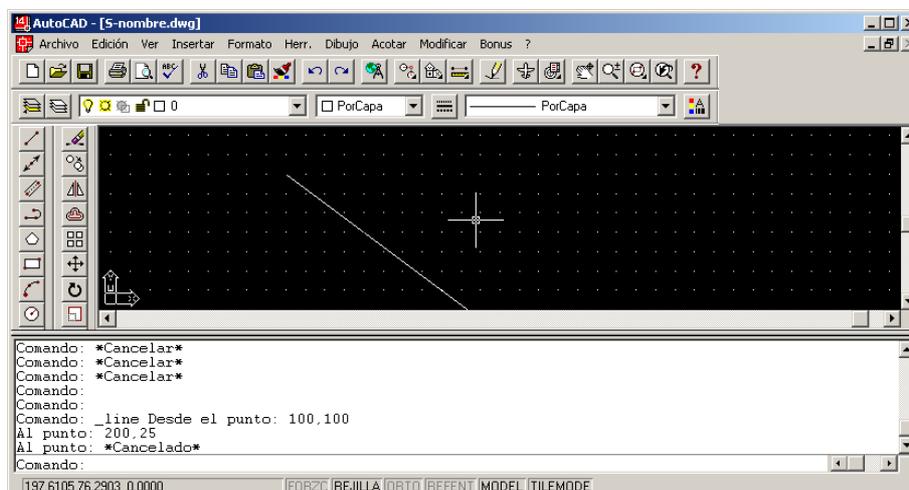


Figura 7 - Dibujo de una línea en AutoCAD® 14 mediante órdenes verbales con Dragon Naturally Speaking®

Esta ha sido una pequeña muestra de lo que sería actualmente el uso de un programa de diseño mediante otro de reconocimiento de voz. Quizás no parece muy prometedor, pero hay que reconocer que es un entorno nuevo todavía poco explorado.

7.3.Resultados.

Este trabajo solo ha incluido una pequeñísima demostración de en qué consistiría el uso del reconocimiento de voz para controlar una aplicación de diseño asistido por ordenador.

Sin embargo, no hay que engañarse. Ese uso varía mucho con relación a la labor que se pretende llevar a cabo, y no ya con el objetivo global del diseño sino con cada una de las tareas individuales que componen el diseño objetivo.

Es decir, para dibujar un plano de un diseño hay que realizar múltiples tareas individuales diferentes: dibujar líneas, círculos, arcos, cambiar formatos y características, copiar y pegar, acotar, etc. Cada una de estas tareas individuales se maneja, con el sistema de reconocimiento de voz, de manera diferente.

Teniendo en cuenta la siguiente clasificación de órdenes en un programa de diseño es posible hablar específicamente de las características del sistema en cada clase. Este orden¹ puede ser el de la tabla 1.

Entidades	Capas	Edición de entidades	Visualización	Acotado	Herramientas de ayuda	Bloques	Archivos
Líneas.	Codificación.	Modificación de dimensiones	Modificación de escala.	De entidades.	Rayado de partes.	Creación de bloques.	Abrir
Arcos, círculos, elipses.	Características.	Modificación de escala	Acercamiento.	De conjuntos.	Rejilla imantada o no.	Librerías.	Guardar
Polígonos.	Presentación.	Modificación de forma	Alejamiento.		Puntos imantados.	Listas de materiales.	Importar, exportar.
Polilíneas.		Eliminación.	Punto de vista.		Paralelas, perpendiculares, tangentes.		
Texto.		Copia, imagen especular.	Vistas.		Distancias.		
		Traslación	Impresión.		Aplicaciones programadas.		
		Giros					
		Agrupar, desagrupar.					

Tabla 1 – Organización en clases de los comandos de diseño asistido por ordenador.

Ahora, analizando algunos de los comandos del grupo **Entidades** se podrían extraer ciertas conclusiones. El resultado se muestra en las siguientes tablas.

Nombre del comando	Menú/Opción	Comando	Subcomando	Cuadro diálogo	Ratón imprescindible	Comando de "voz" SUGERIDO
Líneas auxiliares	Dibujo - Línea auxiliar	Hor; Ver; Ang		No		Línea auxiliar <tipo> <punto> <ángulo>
		Bisectriz		No	Sí	Bisectriz <recta 1> <recta 2> <sentido>
		Desfasar		No	Sí	Paralela <recta> <distancia> <+/->
Rayo	Dibujo - Rayo	rayo		No		Rayo <punto1> <punto 2>
Línea múltiple	Dibujo - Línea múltiple	lineam		No		Línea doble a <distancia> <punto 1>, <punto 2>, ... >cerrar>

Para dibujar líneas auxiliares o rayos nos encontramos con la misma problemática que en el caso de líneas, y para las líneas múltiples la necesidad de definición de características es aun mayor.

Nombre del comando	Menú/Opción	Comando	Subcomando	Cuadro diálogo	Ratón imprescindible	Comando de "voz" SUGERIDO
Polilínea	Dibujo - Polilínea	pol	cerrar			Polilínea <punto1> <punto 2> ... cierra

El comando Polilínea exige multitud de parámetros, todos seguidos, indicando las características de cada segmento de la polilínea: coordenadas (única característica que puede hacerse con el ratón), grosores, tipo de

¹ Diseño y Dibujo Asistido por Computadora, Cap. "Dibujo tradicional, dibujo asistido", Manuel Domínguez, pags. 7 y siguientes.

segmento (arco o línea), etc. Si, además, alguno de los segmentos es un arco el número de parámetros a definir es bastante alto. Posiblemente sea de los comandos más complicados y su paso a voz implicaría mucho trabajo.

Nombre del comando	Menú/Opción	Comando	Subcomando	Cuadro diálogo	Ratón imprescindible	Comando de "voz" SUGERIDO
Rectángulo	Dibujo Rectángulo	Rectáng				Rectángulo <tipo> <Coordenadas de esquinas>
Arandela	Dibujo Arandela	Arandela				Arandela <Centro> <Radio 1> <Radio 2>
Elipse	Dibujo - Elipse	elipse				Elipse <ejes> <inclinación>; Elipse <focos> <distancia>; Elipse >ángulo del arco elíptico>

Este tipo de comandos no implicaría grandes problemas, salvo si se, por ejemplo, quisiera hacer una elipse tangente a otra figura... y cosas así.

Nombre del comando	Menú/Opción	Comando	Subcomando	Cuadro diálogo	Ratón imprescindible	Comando de "voz" SUGERIDO
Círculo	Dibujo - Círculo	circulo	centro-radio	No		Círculo <centro> <radio>
			centro-diámetro	No		Círculo <centro> <diámetro>
			2 puntos	No		Círculo <punto 1> <punto 2>
			3 puntos	No		Círculo <punto 1> <punto 2> <punto 3>
			Tan, Tan, Radio	No		Círculo <elemento 1> <elemento 2> <radio o diámetro>
			Tan, Tan, tan	No		Círculo <elemento 1> <elemento 2> <elemento 3>
Arco	Dibujo Arco		3 puntos			Arco <punto 1> <punto 2> <punto 3>
			Inicio, Centro, Fin			Arco <punto 1> <punto 2> <punto 3>
			Inicio, Centro, Ángulo			Arco <punto 1> <punto 2> <ángulo>
			Inicio, Centro, Longitud			Arco <punto 1> <punto 2> <longitud>
			... etc			... etc
			Continuar		Sí	Arco Mas

Las dificultades que plante el dibujo de círculos y arcos con el funcionamiento actual, son muy grandes... dependiendo de la opción escogida. Si esta es **3 Puntos** no parece que sea difícil, pero en el caso de longitudes, direcciones, tangencias, etc. resulta muy difícil, sino imposible, prescindir del ratón.

Como se ve, pese a la reducida relación de comandos, los problemas planteados son muchos y su solución no está en manos del usuario, tanto es así que cuesta creer que alguien quiera usar el reconocimiento de voz en el manejo de una aplicación de diseño asistido por ordenador, salvo que sea absolutamente necesario por las circunstancias del trabajo (discapacitados sobre todo). De todo esto se pueden sacar abundantes conclusiones que veremos a continuación.

En general, distinguiendo por las clases de órdenes se pueden hacer los siguientes comentarios:

1 – **Entidades**: siempre que las entidades se tracen mediante coordenadas conocidas, no hay ningún problema. Si se hace necesario señalar alguna otra entidad o característica que ha de señalarse con el ratón, entonces se hace más difícil pues este dispositivo es bastante poco ágil con el sistema de voz.

2 - **Capas**: trabajar en diversas capas no ofrece serias dificultades pues basta seleccionar y limitarse a trabajar. Su creación o eliminación tampoco implica grandes dificultades. La tarea de modificar características de entidades se haría en algunos casos muy pesada.

3 - **Edición de entidades**: es aquí donde el usuario discapacitado se topa con grandes dificultades. La necesidad de señalar entidades para su edición es muy difícil. Que duda cabe que es aquí donde las necesidades de cambio en el manejo se hacen más patentes.

4 - **Visualización**: algunas de las herramientas de visualización pueden manejarse relativamente bien (zoom, ventanas), otras resultan más difíciles (vistas).

5 - **Acotado**: ocurre como en la edición de entidades, muchas dificultades para señalar objetos.

6 - **Herramientas de ayuda**: todas las que se manejen mediante menús y opciones no ofrecen grandes dificultades pues es posible acceder con relativa facilidad a ellos.

7 - **Bloques**: vuelve la dificultad de señalar entidades y darles nombre, así como reconocerlas a posteriori llamándolas por su nombre.

8 - **Archivos**: las opciones de abrir, guardar, guardar como, etc. Se manejan sin ningún problema.

7.4. Características exigibles a los sistemas de reconocimiento de voz.

Hay tres características que todo sistema de reconocimiento de voz debiera cubrir Estas serían: **Fiabilidad, Rapidez y Efectividad**.

1 - La **Fiabilidad**: podría medirse por el número de veces que el sistema de reconocimiento de voz “interpreta mal” las órdenes del usuario. Este es un punto vital y en el que hoy en día se tiene que avanzar todavía mucho. Los sistemas de reconocimiento de voz basan este reconocimiento en patrones fonéticos. La consecuencia es que cualquier usuario acatarrado... ¡no puede trabajar! ¡A menos que vuelva a “educar” al sistema... o cree un nuevo usuario (que se podría denominar **Yo acatarrado**). Realmente el número de errores es todavía muy grande. Para colmo de males, el ruido ambiental influye muchísimo, y por supuesto otra persona cercana hablando lo inhabilita totalmente. Es posible aumentar la fiabilidad a base de reducir la **Velocidad**.

2 - La **Velocidad**: se podría medir por el tiempo que tarda en reaccionar o ejecutarse la correspondiente orden una vez dictada verbalmente. Ésta característica esta relacionada de manera inversa con la fiabilidad. De manera que a mayor velocidad, menos fiabilidad y viceversa.

3 - La **Efectividad**: esta podría medirse individualmente para cada tarea tratando de deducir qué sistema es más efectivo (¿podríamos decir **ergonómico?**), el sistema mediante voz o el convencional con teclado y ratón. Hay que volver a diferenciar personas discapacitadas y las que no lo están. Para las primeras volvemos a lo mismo: no habría otra alternativa; pero para las segundas sí que la habría... y ya se sabe cual es. No usar el sistema de reconocimiento de voz.

Se ha de tener en cuenta que no se valoraría igual cada característica para una persona discapacitada que para otra que no lo es. Para la primera, la alternativa sería sencillamente no usar el sistema, cosa inaceptable; pero para una persona no discapacitada la alternativa sería volver a trabajar con teclado y ratón y olvidarse del sistema de voz.

8. Conclusiones

Las conclusiones de este trabajo hacen referencia a las necesidades de modificar el entorno de trabajo para conseguir más eficacia. Las necesidades se refieren tanto a los sistemas de reconocimiento de voz como a las aplicaciones de diseño asistido por ordenador, si bien estas necesidades de cambio son mucho mayores para las segundas que para los primeros. Es decir, son los fabricantes de aplicaciones de diseño los que van a tener que soportar más cambios.

8.1. Necesidades de cambio en los sistemas de reconocimiento de voz.

Con relación a las aplicaciones **RV** qué duda cabe que han de mejorar en sus prestaciones en general. El reconocimiento debe ser mucho más claro y preciso y no puede estar sometido a tantos avatares. Es decir, la voz de una persona debe tener un cierto margen de juego; posiblemente estos programas ya estén dando un cierto margen, pero a la vista de las pruebas hechas éste se debería ampliar... sin que ello lleve a provocar interpretaciones “raras” por parte de la aplicación. Es decir, se necesita mucha más precisión para reconocer la voz del usuario y pasarla correctamente a texto.

La otra parte del reconocimiento de voz, que es el procesamiento del lenguaje natural, queda aparcado. Posiblemente pasen décadas hasta tener sistemas absolutamente fiables.

8.2. Necesidades en las aplicaciones de diseño asistido por ordenador.

Son estas aplicaciones las que sería más sencillo adaptar para que trabajasen con los sistemas actuales de reconocimiento de voz. Para ello deberían reorganizar las órdenes dándoles otras posibilidades de textos más sencillos y claramente diferenciables unos de otros, de manera que el usuario no se vea en la necesidad de “enseñar” un montón de órdenes a su sistema.

Otra necesidad es la clara diferenciación que se ha de hacer entre datos y órdenes. A veces algunas órdenes se traducen en datos y viceversa

Como conclusión final conviene señalar que, las aplicaciones de diseño asistido deben evolucionar hacia este nuevo contexto, y lo mismo que en su momento variaron su aspecto para adaptarse al entorno gráfico tipo “Windows” y al manejo mediante un ratón, ahora deberán prever cambios en sus sistemas de trabajo para adaptarse a estas nuevas expectativas que se están creando conforme el desarrollo del reconocimiento de voz avance. Si no lo hacen, desaparecerán y serán sustituidas por nuevas aplicaciones mejor adaptadas. Que duda cabe que será la posibilidad de evolucionar lo que marcará su supervivencia.

Así mismo, la enseñanza de estas aplicaciones deberá evolucionar a la par y prepararse para esos nuevos entornos “amigables” que todos prometen... aunque no acaban de verse. La evolución en la potencia de los ordenadores y el interés por estos nuevos desarrollos hacen esperar próximamente nuevos cambios. Habrá que estar pendientes de ellos.

9. Bibliografía.

[1] “Reconocimiento de Voz y Fonética Acústica”/ Bernal, Jesús y otros/ Editorial Rama / Enero-2000 / ISBN: 84-789-7398-2.

[2] “Diseño y Dibujo Asistido por Computadora”/ Domínguez, Manuel y otros/ Universidad Nacional de Educación a Distancia / 1995 / ISBN: 84-362-3282-8.

[3] “Del ratón a la palabra”/ Baker, Stephen / Madrid Business Week - Dinero nº. 802./ 9 de sept. de 1999.

[4] “Voice Recognition” / Richard L. Klevans, Robert D. Rodman (Contributor) / Artech House Telecommunications Library/ October 1997/ ISBN: 0890069271.

[5] “Fundamentals of Speech Recognition”/ Lawrence Rabiner, Biing-Hwang Juang, Bilng-Hwang Juang (Contributor) / Prentice Hall/ 1ª edition - April 12, 1993/ ISBN: 0130151572.

[6] “How to build a Speech Recognition Application”/ Bruce Balentine, David P. Morgan/ Enterprise Integration Group Inc./ 1ª edition - April, 1999/ ISBN: 0-9671278-1-5

[7] “AutoCAD® 14 – Fundamentos”/ AutoDesk®/ Editorial Paraninfo/ 1998 / ISBN: 84-283-2449-2.

- AutoCAD® R 14/ Manual del usuario.

- Dragon Naturally Speaking® - Preferred/ Manual del Usuario/ Dragon Systems Inc./ <http://www.dragonsys.com/>

- IBM Via Voice®/ Manuales del usuario/ <http://www.software.ibm.com/is/voicetype/>