

Capítulo 1. Introducción.

1.1 Definición del problema.

El mundo de la persona ciega es un mundo falto de visión en el que la información transmitida por otros sentidos cobra una importancia esencial. Si pensamos que se estima que el vidente recibe alrededor del 85% de la información a través de la vista, comprenderemos que su atención se dirige generalmente hacia el análisis de estos estímulos visuales, haciendo obvia la información que recibe de otras vías sensoriales. Pero el caso del ciego es distinto; las sensaciones auditivas, olfativas y de tacto ocupan un lugar fundamental en su experiencia sensorial; por lo tanto perciben el mundo de manera diferente. En lugar de ser un mundo de imágenes es un mundo de sonidos, olores, texturas, etc., donde la información la recibe a través de la actividad de su propio cuerpo y a través de la información verbal [1].

Todas estas reflexiones son importantes a la hora de comunicarse y trabajar en el campo educativo con los invidentes. Uno de los errores que se comete habitualmente con los videntes es la tendencia al "visuocentrismo", es decir, a centrarnos en nuestro modo de representación visual y a considerar al ciego siempre en relación/comparación con el vidente [1]. Las personas que sufren algún tipo de discapacidad se han visto marginadas al acceso de tecnología. Hoy en día se está intentando que esta marginación desaparezca, implementando nuevas herramientas enfocadas a personas con discapacidades.

Específicamente en el caso de los invidentes el uso de las computadoras se ha visto limitado ya que el manejo de los sistemas de información es de tipo visual. En algunos países como Brasil se han implementado sistemas en los que los invidentes puedan tener acceso al uso de computadoras de manera amigable de acuerdo a sus limitantes. La Universidad Federal de Río de Janeiro en Brasil, ha desarrollado un sistema para invidentes llamado Dos-Vox. Este sistema es utilizado por más de 5000 invidentes en Brasil [2], y permite realizar tareas básicas (editar texto, acceder discos, enviar correos, fax, entre otras) comunicándose con el invidente a través de un sintetizador de voz. Entre las ventajas que tiene este software destaca el hecho de que es gratuito y que el instalarlo requiere de poco hardware. El funcionamiento de este sistema se detalla un poco más en el capítulo 2.

Existe en el estado de Puebla un consejo estatal de organizaciones no gubernamentales, el cual tiene un proyecto de servicio social en la Universidad de las Américas Puebla que consiste en traducir el sistema Dos-Vox al español (el cual se llamará MexVox) y una vez terminado capacitar a un grupo de invidentes para trabajar con el sistema.

Desde siempre el habla ha sido el medio de comunicación mas importante. Por esto, el área de reconocimiento de voz (conversión de palabras habladas a palabras escritas) ha tomado gran importancia y por lo tanto las herramientas de tecnología de reconocimiento de voz han tenido un importante desarrollo en los últimos años. El proceso de reconocimiento de voz dota a las máquinas de la capacidad de recibir mensajes orales. Tomando como entrada la señal acústica recogida por un micrófono, el proceso de

reconocimiento de voz tiene como objetivo final decodificar el mensaje contenido en la onda acústica para realizar las acciones pertinentes. Para lograr este fin, un sistema de reconocimiento de voz necesitaría conjugar una gran cantidad de conocimientos acerca del sistema auditivo humano, sobre la estructura del lenguaje, la representación del significado de los mensajes y sobre todo el autoaprendizaje de la experiencia diaria [3].

Básicamente, el reconocimiento del habla es un proceso de clasificación de patrones, cuyo objetivo es clasificar la señal de entrada (onda acústica) en una secuencia de patrones previamente aprendidos y almacenados en unos diccionarios de modelos acústicos y de lenguaje. Este proceso de clasificación supone:

- Que la señal de voz puede ser analizada en segmentos de corta duración y representar cada uno de los segmentos mediante su contenido frecuencial, de forma análoga al funcionamiento del oído.
- Que mediante un proceso de clasificación podemos asignar a cada segmento o conjuntos consecutivos de segmentos una unidad con significado lingüístico.
- Que mediante un procesador lingüístico podemos dar significado a las secuencias de unidades [3].

Este último paso del sistema supone incorporar al sistema de reconocimiento de voz conocimiento acerca de la estructura sintáctica, semántica y pragmática del lenguaje. Sin embargo, los sistemas actuales de reconocimiento de voz solo incorporan estas fuentes de conocimiento sobre tareas muy restringidas y controladas, estando la mayoría de ellos en experimentación en condiciones de laboratorio [3].

1.2 Objetivos

1.2.1 Objetivos Generales

Este proyecto propone el añadir un sistema de reconocimiento de voz gratuito a MexVox usando la Microsoft Speech API y experimentar el impacto que tendrá el dar instrucciones específicas de voz para llevar a cabo tareas que el software realiza normalmente mediante el teclado.

Las tareas en las que nos enfocamos son funciones del menú principal de MexVox, como pueden ser el probar teclado, el editor de textos, la impresora, los juegos, etc; funciones del editor de textos como pueden ser el abrir archivo guardar archivo, cerrar archivo, imprimir archivo, buscar una palabra en un texto, etc; y funciones del lector de documentos y del programa que finaliza MexVox con comandos como salir del sistema, cerrar sesión, etc.

Todas estas funciones se accesan y se ejecutan mediante palabras claves mnemónicas que el sistema reconoce a través de la voz del usuario.

1.2.2 Objetivos Específicos

- 1 Realizar un estudio sobre todas las posibles limitantes del Dos-Vox.
- 2 Participar en la documentación (Ver apéndice A) y traducción del Dos-Vox al español (Ver apéndice B). Esta traducción se llevó a cabo por 4 estudiantes de la Universidad de las Américas Puebla: Carlos Navarro García, Victor M. Reyes

Pedraza, José Manuel Zulaica Rugarcía y Bertha Mariel Márquez Avendaño, los dos últimos, los autores de esta tesis como parte de nuestro proyecto de servicio social.

- 3 Hacer un estudio de Microsoft Speech API (SAPI) para determinar las herramientas a utilizar para el reconocimiento de voz; esto porque el Dos-Vox utiliza herramientas de la SAPI para la síntesis de voz.
- 4 Utilizar un reconocedor de voz gratuito comercial (Microsoft Speech Recognition Engine Versión 4.0.4.2512) e implementarlo por medio de la SAPI al sistema MexVox. Es importante señalar que no se desarrolló el reconocedor sino que solo se adaptó al sistema uno ya existente, entrenándolo y delimitando el diccionario de datos.
- 5 Que al implementar el producto final, éste utilice el mínimo de requerimientos de hardware y de software posibles.

1.3 Alcances

- 1 Que sea utilizado de manera gratuita por los invidentes de la Asociación Cultural Recreativa y proyección social de Invidentes Puebla, A.C. (ACRIP).
- 2 Que pueda ser utilizado en cualquier tipo de PC que tenga tarjeta de sonido y micrófono.

1.4 Limitaciones

1. Nula información escrita en español del sistema Dos-Vox.
2. El reconocedor utilizado es para el idioma inglés, sin embargo lo ocupamos para

un vocabulario de comandos en español (Ver apéndice C). Esto limita a que el reconocimiento tenga un poco más de errores que si utilizáramos uno en español.

3. El reconocedor utilizado es dependiente del locutor, es decir, cada usuario que utilice en sistema debería entrenar con su propia voz al reconocedor, por lo que fue necesario entrenar varias voces como si fuera el mismo usuario (Ver apéndice C). Aunque esto ayuda a resolver el problema de que cada usuario entrene con su propia voz, esto no es lo más óptimo ya que hace que el reconocimiento no sea tan preciso.

1.5 Software utilizado

- Borland Delphi Studio versión 7.0.
- Código fuente Dos-Vox, del cual se tiene el permiso del Mtro. José Antonio Borges, autor original, para su uso y/o modificación. Específicamente para la traducción del DosVox al español se utilizaron los códigos del DosVox (dosvov.dpr), el editor de textos (minied.dpr), la calculadora (calcuvox.dpr), el juego del ahorcado (forcavox.dpr), el lector de documentos (levox.dpr), la impresora común (listavox.dpr), el controlador de volumen (tmix.dpr), el juego letrix (letrix.dpr) y la salida del sistema (fechawin.dpr).

El programa principal del MexVox (mexvox.dpr), el editor de textos (minied.dpr), el lector de textos (levox.dpr) y la salida del sistema (fechawin.dpr) son los programas que utilizan reconocimiento de voz. Esto es sólo una parte del sistema, los demás programas quedaron pendientes para su traducción al español.

- Microsoft Speech API 4.0.

- Reconocedor Microsoft Speech Recognition Engine Versión 4.0.4.2512

1.6 Hardware utilizado.

- El trabajo se realizó sobre una computadora portatil con procesador Celeron 1.8 mhz y memoria de 256 mb; pero se pretende que el software trabaje sobre cualquier procesador.
- Tarjeta de Sonido
- Micrófono
- Audífonos. Esto es porque el Mex-Vox utiliza síntesis de voz que puede ser captada por el reconocedor de voz en caso de no usar audífonos.

1.7 Descripción del documento

Este documento está dividido en 7 capítulos de los que describiremos su contenido a continuación.

El presente capítulo, es una definición general del problema, de los objetivos del proyecto, los alcances y limitaciones, así como los requerimientos de hardware y de software para llevar a cabo el proyecto.

El segundo capítulo, llamado software para invidentes, trata acerca de los sistemas existentes para el que el invidente pueda acceder a la información. Inicia con una historia de los diversos métodos que han existido para que los invidentes puedan tener acceso a ésta

y continúa así hasta llegar a los sistemas actuales para invidentes que permiten el acceso a la información por medio de las computadoras y en particular en el web.

En particular se hace hincapié en el sistema DosVox y su traducción al español llamada MexVox. Se explica en que consiste el sistema y como funciona.

El tercer capítulo trata, por un lado sobre la importancia del lenguaje hablado y por otro, sobre los reconocedores de voz, como se clasifican, las ventajas y desventajas de estos,. Explica de manera general como construir un reconocedor y los tipos de reconocedores que existen.

El cuarto capítulo trata de *Microsoft Speech Application Interface* (SAPI), así como un apartado para el reconocedor que utilizamos que es el *Microsoft Speech Recognition*. En este capítulo explicamos que es la SAPI, y las interfaces con las que trabaja hasta llegar al *Voice Command Control* que son los comandos con los que trabajamos para la implementación del reconocedor al sistema MexVox. Dedicamos un capítulo completo a este tema porque dentro del desarrollo de éste proyecto, gran parte del esfuerzo fue dedicado al estudio de SAPI y consideramos que es importante el conocer este tema.

El quinto capítulo es dedicado a la arquitectura Cliente-Servidor, esto porque dentro de nuestro diseño utilizamos esta arquitectura para comunicar al sistema MexVox con el reconocedor de voz. Explicamos lo que es un cliente, un servidor, las características, las ventajas y desventajas de esta arquitectura y como utilizamos esta arquitectura dentro de nuestro sistema.

El sexto capítulo es acerca del análisis y del diseño del sistema. En este capítulo explicamos de manera detallada en que consistió nuestro trabajo y como fue diseñada la aplicación; así como los problemas a los que nos enfrentamos y la manera en que fueron resueltos.

Finalmente el séptimo capítulo muestra los resultados obtenidos de las pruebas, así como las conclusiones obtenidas al término de este trabajo y las propuestas de trabajos a futuro.