

**UNA HERRAMIENTA BASADA EN PROCESAMIENTO Y
RECONOCIMIENTO DE VOZ PARA LA MEJORA EN
PRONUNCIACIÓN DE VOCALES DE PERSONAS CON
PROBLEMAS AUDITIVOS**



TESIS

MAESTRÍA EN CIENCIAS EN TECNOLOGÍA INFORMÁTICA

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY**

CAMPUS MONTERREY

POR

CRISTINA V. GONZALEZ CÓRDOVA

Mayo de 2003

**UNA HERRAMIENTA BASADA EN PROCESAMIENTO Y
RECONOCIMIENTO DE VOZ PARA LA MEJORA EN
PRONUNCIACIÓN DE VOCALES DE PERSONAS CON
PROBLEMAS AUDITIVOS**

POR

CRISTINA V. GONZÁLEZ CÓRDOVA

TESIS

Presentada a la División de Graduados en Electrónica, Computación, Informática y
Comunicaciones. Este trabajo es requisito parcial para obtener el Título de

Maestra en Ciencias en Tecnología Informática

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES
DE MONTERREY

CAMPUS MONTERREY

Mayo de 2003

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

CAMPUS MONTERREY

DIVISIÓN DE GRADUADOS EN ELECTRÓNICA, COMPUTACIÓN,
INFORMACIÓN Y COMUNICACIONES

PROGRAMA DE POSTGRADO EN ELECTRÓNICA, COMPUTACIÓN,
INFORMACIÓN Y COMUNICACIONES

Los miembros del comité de tesis recomendamos que la presente tesis de la Ing. Cristina V. González Córdova sea aceptada como requisito parcial para obtener el grado académico de Maestría en Ciencias, con especialidad en:

Tecnología Informática

Comité de Tesis

Juan A. Nolazco Flores, Ph. D.
Asesor

Ingrid Kirschning, Ph. D.
Sinodal

Arturo Galván Rodríguez, Ph. D.
Sinodal

David A. Garza Salazar, Ph. D.
Director de los Programas de Postgrado en Electrónica, Computación,
Informática y Comunicaciones

Mayo de 2003

DEDICATORIAS

Principal y especialmente esta tesis esta dedicada a mi “papi” **Salvador Córdova Domínguez** porque me dio todo en la vida, me enseñó a luchar, me enseñó a salir adelante, formó mis valores y mi carácter; aunque él se fué de este mundo, su fuerza y determinación se quedaron en mí.

A mi “mami” Maria Luisa C. de Córdova por ser más que una madre para mí, por apoyarme y amarme incondicionalmente, por estar ahí cuando más la necesité.

A mi madre Margarita Córdova Casanova por el apoyo constante en todos los aspectos y la fe que siempre tuvo en mí.

A mi hermano Erick porque vino a este mundo a traerme la alegría de un hermano que siempre añoré. Que este trabajo le sirva de inspiración para que nunca se de por vencido.

A Sergio Hernández Aniles, por su amor incondicional, porque en todo momento me apoyo, me dio tranquilidad en los momentos más difíciles, me dio un abrazo cuando lo necesitaba.

A mis queridos tíos Noemí, Miguel, Abner, Eunice, Juan y Elia por su amor incondicional y su ayuda en todo momento; porque sé que cuento con ellos como papás y mamás.

A mis primos Alex, Clau, Heco, Lisi, Migue, Vilu, Are, Bebé, Pau, Adri, Ale C., Ale O. y Kike por llenar mi vida de alegría.

A mis amigas y amigos, de todas partes y de todos los tiempo, por haber dejado algo bueno en mí.

A mis maestros de Montemorelos y de Monterrey por su disposición y apoyo en mi formación académica.

¡Muchas Gracias!

AGRADECIMIENTOS

A Dios por darme la vida y poner los medios para que realizara la maestría y esta tesis.

A mi asesor, el Dr. Juan A. Nolzco Flores, por su apoyo, conocimientos, disposición y ayuda en todo momento.

A mis sinodales, la Dra. Ingrid Kirschning y el Dr. Arturo Galván por su disposición, interés y comentarios.

A toda mi familia y amigos por el apoyo en todo momento.

A mis maestros por su guía y disposición.

RESUMEN

En este trabajo de investigación se desarrolló una interfaz para dar retroalimentación visual de la posición articulatoria del tracto vocal a personas con problemas auditivos para mejorar la pronunciación de las cinco vocales del idioma español. La interfaz muestra la posición correcta del tracto vocal contrastándola con la posición realizada por el usuario, calculada a partir de la señal de voz grabada. De esta manera proporciona retroalimentación visual. Se proporcionan patrones estándares de la posición correcta del tracto vocal, pero existe la opción de guardar y usar patrones personalizados. También es capaz de verificar la pronunciación con un reconocedor. Para comprobar que la herramienta de software estuviera ayudando a mejorar la pronunciación, se contó con la participación de cinco personas con sordera profunda, quienes fueron sometidos a un entrenamiento con el software durante ocho semanas, todos los días. Se tomaron muestras de pronunciación antes y después del periodo de entrenamiento, y después de analizar y evaluar los datos, obtuvimos resultados satisfactorios.

TABLA DE CONTENIDO

DEDICATORIAS.....	iv
AGRADECIMIENTOS.....	v
RESUMEN.....	vi
TABLA DE CONTENIDO.....	vii
LISTA DE GRÁFICAS Y FIGURAS.....	ix
LISTA DE TABLAS.....	xi
Capítulo 1 Introducción.....	1
Capítulo 2 Antecedentes.....	4
2.1 Problemas Auditivos	4
2.2 Aparato Fonador Humano.....	7
2.3 Teoría Fonética Articulatoria.....	8
2.4 Modelo Físico de la Propagación del Sonido en el Tracto Vocal.....	14
2.5 Reconocimiento de Voz.....	16
2.6 Interfaz Hombre-Máquina.....	19
Capítulo 3 Estado del Arte.....	22
3.1 Software para Traducción de Lenguaje de Signos.....	23
3.2 Software Comercial y en Desarrollo con Tecnología de Voz.....	24
3.2.1 Speech Viewer de IBM.....	24
3.2.2 CSLU Toolkit.....	26
3.2.3 Sistema Panasonic para Enseñar a Hablar a Niños con Sordera.....	26
3.2.4 Pronunciation Power.....	28
3.2.5 ViaVoice de IBM.....	28
3.2.6 Otros Productos de Software para la Rehabilitación de la Voz.....	28
3.3 Software Desarrollado en México con Tecnología de Voz.....	28
3.3.1 Software para Detectar Trastornos del Habla.....	29
3.3.2 Software para la Enseñanza del Español Hablado en México.....	28
3.3.3 ICATIANI.....	30
Capítulo 4 Trabajo Propuesto.....	31
4.1 Descripción del Sistema.....	32
4.2 Diseño del Sistema.....	32
4.3 Patrones.....	36

Capítulo 5 Pruebas y Evaluación.....	38
5.1 Obtención de Datos.....	39
5.2 Experimentos Tipo I.....	39
5.2.1 Experimento 1, Tipo I.....	41
5.2.2 Experimento 2, Tipo I.....	42
5.2.3 Experimento 3, Tipo I.....	43
5.2.4 Experimento 4, Tipo I.....	45
5.2.5 Experimento 5, Tipo I.....	46
5.2.6 Resultados Generales de Experimentos Tipo I.....	47
5.3 Experimentos Tipo II.....	50
5.3.1 Alumno A03, Vocal E.....	52
5.3.2 Alumno A04, Vocal E.....	53
5.3.3 Alumno A05, Vocal I.....	53
5.3.4 Alumno A06, Vocal I.....	54
5.3.5 Alumno A06, Vocal O.....	54
5.3.6 Alumno A07, Vocal O.....	55
Capítulo 6 Conclusiones y Trabajo Futuro.....	57
Anexo A Algoritmo Levinson-Durbin.....	59
Anexo B Formato de Archivo de Audio WAV.....	60
Anexo C Manual de Usuario de REVISAH.....	62
C.1 Introducción.....	62
C.2 Características Técnicas.....	62
C.3 Requerimientos del Sistema.....	63
C.4 Instalación.....	63
C.5 Operación del Sistema.....	64
Anexo D Cuestionarios.....	68
D.1 Cuestionario Tipo I.....	68
D.2 Cuestionario Tipo II.....	69
Anexo E Resultados de Experimentos Tipo II.....	70
Anexo F Datos de Alumnos Participantes y de Voluntarios.....	71
Referencias Bibliográficas.....	72
Vita.....	77

LISTA DE GRÁFICAS Y FIGURAS

Figura 2.1	Aparato Fonador Humano.....	7
Figura 2.2	Triángulo Vocálico Simplificado.....	10
Figura 2.3	Articulación para la Pronunciación del Fonema <i>a</i>	11
Figura 2.4	Articulación para la Pronunciación del Fonema <i>e</i>	12
Figura 2.5	Articulación para la Pronunciación del Fonema <i>i</i>	12
Figura 2.6	Articulación para la Pronunciación del Fonema <i>o</i>	13
Figura 2.7	Articulación para la Pronunciación del Fonema <i>u</i>	14
Figura 2.8	Tracto Vocal con Área Variable.....	15
Figura 2.9	Tracto Vocal como la Concatenación de Tubos.....	16
Figura 2.10	Proceso de Reconocimiento de Voz.....	18
Figura 3.1	Sistema de Traducción Automática de Lenguaje de Signos.	24
Figura 3.2	Sensores Acústicos y Articulatorios.....	27
Figura 3.3	Interfaz del Sistema ICATIANI.....	30
Figura 4.1	Pantalla Principal del Sistema.....	33
Figura 4.2	Diagrama de Contexto.....	34
Figura 4.3	Diagrama de Flujo de Datos 1, Nivel 1.....	35
Figura 4.4	Diagrama de Flujo de Datos 2, Nivel 1.....	35
Figura 4.5	Diagrama de Flujo de Datos 3, Nivel 1.....	36
Figura 4.6	Graficación de las Doce Áreas.....	37
Figura 5.1	Software de Apoyo para Cuestionario Tipo I.....	40
Gráfica 5.1	Calificaciones de Todas las Vocales para el Alumno A03...	41
Gráfica 5.2	Calificación Promedio de Todas las Vocales para el Alumno A03.....	42
Gráfica 5.3	Calificaciones de Todas las Vocales para el Alumno A04...	42
Gráfica 5.4	Calificación Promedio de Todas las Vocales para el Alumno A04.....	43
Gráfica 5.5	Calificaciones de Todas las Vocales para el Alumno A05...	44
Gráfica 5.6	Calificación Promedio de Todas las Vocales para el Alumno A05.....	44
Gráfica 5.7	Calificaciones de Todas las Vocales para el Alumno A06...	45
Gráfica 5.8	Calificación Promedio de Todas las Vocales para el Alumno A06.....	46
Gráfica 5.9	Calificaciones de Todas las Vocales para el Alumno A07...	46
Gráfica 5.10	Calificación Promedio de Todas las Vocales para el Alumno A07.....	47

Gráfica 5.11 Comparativo General de Calificaciones para la Vocal “A”...	47
Gráfica 5.12 Comparativo General de Calificaciones para la Vocal “E”...	48
Gráfica 5.13 Comparativo General de Calificaciones para la Vocal “I”....	48
Gráfica 5.14 Comparativo General de Calificaciones para la Vocal “O”...	48
Gráfica 5.15 Comparativo General de Calificaciones para la Vocal “U”...	49
Gráfica 5.16 Calificaciones Promedio por cada Vocal.....	49
Gráfica 5.17 Calificaciones Promedio de todas las Vocales de todos los Participantes.....	50
Gráfica 5.18 Calificaciones Promedio de todas las Vocales.....	50
Figura 5.2 Software de Apoyo para Cuestionario Tipo II.....	51
Gráfica 5.19 Gráfico de Antes y Después del Entrenamiento.....	52
Gráfica 5.20 Identificación de la Vocal “E” para A03.....	52
Gráfica 5.21 Identificación de la Vocal “E” para A04.....	53
Gráfica 5.22 Identificación de la Vocal “I” para A05.....	53
Gráfica 5.23 Identificación de la Vocal “I” para A06.....	54
Gráfica 5.24 Identificación de la Vocal “O” para A06.....	55
Gráfica 5.25 Identificación de la Vocal “O” para A07.....	55
Gráfica 5.26 Comparativo de Casos con Mejoría contra Casos sin Mejoría	56
Figura C.1 Pantalla Inicial del Sistema REVISAH.....	64
Figura C.2 Pantalla Principal del Sistema REVISAH.....	65
Figura C.3 Pantalla de Grabación del Sistema REVISAH.....	66
Figura D.1 Hoja #1 del Cuestionario Tipo I.....	68
Figura D.2 Hoja #1 del Cuestionario Tipo II.....	69

LISTA DE TABLAS

Tabla 2.1	Clasificación Tradicional de Fonemas.....	9
Tabla F.1	Alumnos Participantes.....	71
Tabla F.2	Personas Voluntarias.....	71

CAPITULO 1

Introducción

Desde tiempos antiguos la comunicación ha sido parte primordial en la vida cotidiana del ser humano. Como seres racionales, los humanos tienen la necesidad de transmitir sus pensamientos, sentimientos y deseos. La mente humana no solamente es activa por naturaleza sino que busca el diálogo con otras mentes.

Al comenzar su vida, el ser humano encuentra formas primitivas de comunicación con el exterior, por medio de llanto o gestos. Poco a poco aprende a escuchar sonidos y reproducirlos como la forma de comunicación típica del ser humano [25]. La mayoría pasa por este proceso y al crecer no requiere un gran esfuerzo para transmitir lo que quiere decir por medio del habla.

Sin embargo, comunicarse verbalmente no es tan sencillo para personas con problemas de audición, y puede llegar a convertirse en un gran reto. Las personas que carecen por completo de la audición llegan a ser etiquetada como “sordo-mudos”, siendo que fisiológicamente, su aparato fonador es igual al de cualquier persona que escucha. Aunque aprenden a utilizar otros medios de comunicación, el deseo de hacer uso de sus labios y voz se convierte en una necesidad. Cuando no satisfacen esta necesidad pueden llegar a tornarse cohibidos y reprimidos [25].

La deficiente capacidad para comunicarse verbalmente se debe a que carecen de la retroalimentación auditiva que se produce al pronunciar una palabra y ser escuchada por sí mismo. Esta retroalimentación es vital para la adquisición del lenguaje oral, y cuando falta (sordera congénita y demás disfunciones del aparato auditivo) conlleva deformaciones profundas de la actividad fonadora [32].

En México existen pocos profesionistas sordos, pues se les ha brindado un sistema educativo deficiente con limitadas oportunidades de desarrollarse tanto personal como profesionalmente. Además, la ignorancia y prejuicios de la gente limitan su posibilidad de encontrar un empleo. Un estudio realizado en Centros de Atención Múltiple (CAM) y Unidades de Servicio de Apoyo a la Educación Regular (USAER) de México, mostró que únicamente 37% de los alumnos reciben el servicio complementario de terapia del lenguaje como apoyo a la comunicación oral [48].

Afortunadamente, la tecnología permite desarrollar herramientas que ayudan a personas con problemas auditivos [17, 18, 21, 23, 54, 56]. Este trabajo de investigación propone crear un producto de software, basado en procesamiento y reconocimiento de voz, que ayude a mejorar la pronunciación verbal de las vocales del idioma español, sustituyendo la retroalimentación auditiva que comúnmente usamos, por una retroalimentación visual por medio de la computadora. Esto ayudará a mejorar la calidad de vida de personas con discapacidad auditiva.

Esta herramienta tecnológica estará orientada específicamente a las personas con problemas auditivos especialmente los que carecen por completo de la audición. De acuerdo al trabajo de campo realizado en el área metropolitana de Monterrey en instituciones orientadas a ayudar a personas con discapacidad auditiva se encontraron algunos productos de software enfocados a personas con problemas de lenguaje o para personas que quieren aprender a leer. Según las maestras del Instituto Sertoma, este tipo de software es poco adecuado pues está enfocado a otro tipo de usuarios y no satisface plenamente las necesidades de las personas con problemas auditivos. Otras instituciones, por falta de información y recursos, utilizan métodos poco eficientes para sustituir la retroalimentación auditiva, y tratan a las personas con problemas auditivos como si tuvieran trastornos del habla o deficiencia mental.

De acuerdo al estudio realizado, no se cuenta con herramientas de software que ayuden a personas con problemas auditivos a pronunciar correctamente las palabras indicándoles la posición correcta de los labios, dientes y lengua. Los maestros utilizan su propia boca para mostrar la posición adecuada para pronunciar fonemas y palabras, pero este método es poco efectivo pues es difícil mostrar el interior de la boca al pronunciar ciertos fonemas. El maestro no puede enseñar la correcta posición de los órganos internos de la boca y al mismo tiempo enseñar la correcta posición de la vista exterior de la boca, por lo cual el alumno se puede confundir.

Con una interfaz adecuada para la aplicación, y utilizando métodos de procesamiento y reconocimiento de voz, se podría crear una retroalimentación visual que estaría supliendo las necesidades tecnológicas que se tienen en los centros de atención a personas con problemas auditivos para enseñar la correcta pronunciación de las cinco vocales del idioma español.

Por esta razón, en este trabajo se propone probar una herramienta de retroalimentación visual. Esta herramienta es la evolución del proyecto REVISAH [39], a la cuál se le agrega una interfaz de visualización.

Con esta herramienta, la persona que carece de retroalimentación auditiva, podrá visualizar la posición correcta del tracto vocal, incluyendo labios, dientes, lengua y mandíbula, teniendo un patrón predeterminado o ajustable. Tendrá la posibilidad de imitar esta posición grabando la pronunciación que cree correcta. Después de procesar la voz, la interfaz de software desarrollada en este trabajo es capaz de indicarle la posición (articulación) que realizó e indicarle sus posibles errores, mostrándole gráficamente el contraste entre el patrón y lo que la persona realizó. También tiene la opción de mostrar si su voz es reconocida como la vocal que trato de pronunciar, utilizando un reconocedor de vocales. De esta manera la persona obtendrá retroalimentación que le ayudará a mejorar su pronunciación.

Este proyecto está limitado a mejorar la articulación de las cinco vocales del idioma español y no toma en cuenta otras características de la voz como el tono o la intensidad. Soluciones generales a todos los casos y características de la voz no se contemplaron, pues se buscó ser muy específicos en las características articulatorias de las vocales.

El trabajo propuesto tiene, primeramente, una justificación tecnológica. De acuerdo a investigación bibliográfica exhaustiva y de campo, se observó que se carece de herramientas tecnológicas que ayuden a discapacitados auditivos a mejorar la pronunciación de vocales en español, mostrando la articulación correcta. Tampoco se cuenta con herramientas que utilicen retroalimentación visual como medio para identificar deficiencias de pronunciación en la articulación provocadas por la falta de retroalimentación auditiva.

También se tiene una justificación social. Será una herramienta utilizada para ayudar a personas con problemas auditivos a mejorar su comunicación con los demás, permitiéndoles un desarrollo y adaptación más amplios en un mundo de comunicación verbal. A pesar de que poco a poco se están tratando de romper las barreras y discriminación hacia personas con alguna discapacidad en áreas como trabajo, transportación, acceso a edificios y comunicación [19, 23], las personas con problemas auditivos pueden obtener mayores oportunidades en un mundo de hablantes si logran comunicarse efectivamente de forma audible.

Para comprobar que la interfaz de software desarrollada está ayudando a mejorar la pronunciación de las vocales del idioma español se realizó un periodo de entrenamiento con el software. El experimento se llevó a cabo con la participación de cinco personas con sordera profunda, del Centro de Atención Múltiple Instituto Sertoma. El entrenamiento se realizó durante ocho semanas, veinte minutos de lunes a viernes.

Se tomaron muestras de antes y después del periodo de entrenamiento. Se realizaron dos tipos de evaluaciones. La primera consistió en cuestionarios de calificación de la calidad de las pronunciaciones, obteniendo calificaciones de antes y después con una calificación máxima de 5. La segunda consistió en cuestionarios en los cuales el encuestado clasificaba las pronunciaciones de acuerdo a la vocal que él creía que se había pronunciado. Con estos dos tipos de experimentos se obtuvieron resultados que demuestran que la herramienta sí está ayudando mejorar la pronunciación de las vocales.

CAPITULO 2

Antecedentes

De acuerdo a datos recopilados por la organización Mundial de la salud en el año 1996, en México existen 3 millones de personas con algún tipo de discapacidad auditiva. Con respecto a los niños mexicanos que sufren este tipo de discapacidades, el DIF informa que hay alrededor de 35 mil niños con sordera total y 600 mil niños con distintos grados de sordera [48]. En la primera parte de este capítulo explicaremos qué son los problemas auditivos y sus principales causas.

A continuación presentaremos brevemente una descripción del aparato fonador humano, sus partes y el papel fundamental que juega en la producción de sonidos y voz. Posteriormente haremos una revisión de la teoría fonética articulatoria, la cual nos muestra como se relaciona la articulación de la boca con los fonemas.

En la descripción del modelo físico de la propagación del sonido en el tracto vocal expondremos las bases físicas y matemáticas que sustentan la teoría para determinar la articulación realizada analizando la señal de voz.

Finalmente explicaremos el proceso de reconocimiento de voz utilizado en el producto de software desarrollado en este trabajo; y se dará un vistazo a la interfaz hombre-máquina que permite recibir retroalimentación adecuada a los usuarios potenciales del software propuesto.

2.1 Problemas Auditivos

Los problemas auditivos pueden tener varias causas, como: causas congénitas, causas genéticas, el uso de medicamentos ototóxicos, la rubéola prenatal, la anoxia prenatal, infecciones, exposición al ruido excesivo, la presbicusis, el sarampión, la gripe, las infecciones del oído medio, las lesiones en la cabeza, los tumores benignos del nervio auditivo (neuroma acústico), el cáncer (infrecuente), entre otros [59]. Alrededor del 60% de las causas tienen su origen en infecciones virales y bacterianas o en enfermedades como inflamación del oído medio y deshidratación. Esto significa que estos problemas bien podrían haberse evitado por medio de aplicación de vacunas y atención médica oportuna [48]. Por esta razón, hay mayor incidencia de estos problemas en países en vías de desarrollo, como el nuestro.

Existen diferentes tipos de problemas auditivos. Las personas que han perdido completamente la audición se les llama “sordos profundos”. Las personas que conservan restos de audición son llamadas “hipoacúsicos”. La pérdida auditiva parcial puede ser [59]:

- Leve (pérdida de hasta 40 decibeles) – problemas para escuchar una conversación.
- Moderada (40-60 decibeles) – se debe levantar la voz para ser escuchado.
- Severa (más de 60 decibeles) – se debe gritar para ser escuchado.

La pérdida de la audición puede ser provocada por la afectación de distintas partes del oído. El oído está compuesto de tres partes: el oído externo, oído medio, y oído interno. El sonido se desplaza a lo largo del canal de la oreja del oído externo y causa que el tímpano vibre. El oír ocurre cuando las olas de sonido llegan a las estructuras ubicadas dentro de la oreja, y luego el oído convierte esas vibraciones en señales de nervio que el cerebro reconoce como el sonido. Tres huesos pequeños del oído medio (Estríbo, Martillo y Yunque) dirigen la vibración del tambor de la oreja a la cóclea o la recámara auditiva del oído interno [59].

Las dos clases principales de sordera son: sordera de conducción y sordera de nervio. La sordera por conducción es la interrupción de las vibraciones de sonido en el paso del exterior hacia las células del nervio, en el oído interno. Las causas de este problema pueden ser cerilla que bloquea el canal exterior de audición, un tímpano perforado, o fluido en las orejas. La sordera por conducción puede ser corregida con tratamiento médico o quirúrgico [13, 59].

En la sordera de nervio (deficiencia sensorioneural), existe algún defecto en las células sensoriales del oído interno o en el nervio del vestíbulo coclear que impide la transmisión de impulsos de sonidos del oído interno hacia el centro de audición en el cerebro. Las razones comunes de la sordera de nervio son exposición al ruido (ruidos fuertes), los cambios relacionados con la edad, la herencia, drogas ototóxicas (algunos antibióticos, aspirinas en dosis altas), y algunas enfermedades (como la meningitis). La sordera de nacimiento es casi siempre del tipo de sordera de nervio y no es curable por tratamientos médicos [13, 59].

Para personas que no han perdido por completo la audición (hipoacúsicos) existen diferentes tratamientos y técnicas para mejorar la audición. Si el daño es en el oído interno se utilizan audífonos. Estos audífonos, aplicados correctamente, pueden mejorar considerablemente el nivel de vida de quienes lo utilizan. Funcionan amplificando el sonido y son más eficaces en lugares silenciosos. La mayoría de los audífonos comercialmente distribuidos se colocan detrás del oído (BTE), dentro de la oreja (ITE), o en el canal de la oreja (ITC) [59].

El implante coclear es otro tipo de técnica, aunque muy costosa, utilizada para casos más severos. Incluso en personas con sordera profunda es posible recobrar parcialmente la audición. Estos implantes son sistemas electrónicos biomédicos que circunvalen partes deterioradas del oído interno y estimulan el nervio auditivo electrónicamente [10, 59]. Esto se realiza por medio de una antena que capta las vibraciones de sonido y las envía a un procesador de computadora muy pequeño que los transforma en los impulsos eléctricos [8]. Este dispositivo tiene 22 electrodos y cada uno estimula una fibra del nervio auditivo que se encuentra dentro de la cóclea. Hay algunos dispositivos que utilizan software para procesar y transmitir los datos [10].

Actualmente se realizan más de 30 intervenciones anuales de implante coclear en nuestro país [8]. Cabe mencionar que este tipo de implantes han desatado mucha polémica, pues hay quienes piensan que poner este tipo de implantes en niños demasiado pequeños, y sin su autorización, va contra sus derechos [34].

También se han encontrado opiniones encontradas en la filosofía y métodos para educar a las personas sordas. En México como en muchas partes del mundo, los planes educativos no proveen el soporte adecuado para enseñar correctamente a las personas con problemas auditivos, negándoles la oportunidad de una educación de calidad y de sobresalir en el ámbito profesional [48, 50].

Se pueden destacar tres corrientes filosóficas de la educación para sordos: la oralista, la manual, y por otro lado, una tercera corriente mediadora entre las dos posiciones anteriores: la propuesta del bilingüismo [48].

La corriente oralista se basa en la idea de que los sordos son enfermos que deben adaptarse al medio de los “normales”. Propone rescatar lo más posible la audición por medio de aparatos ortopédicos. En este enfoque, las personas con problemas auditivos deben aprender a comunicarse en forma oral con las personas oyentes para poder adaptarse al medio. Esta ha sido una de las corrientes más antiguas y se piensa que tuvo sus inicios en el Congreso de Milán en 1880 [48]. Sarah Fuller fue una de las pioneras norteamericanas en la enseñanza del lenguaje hablado más que el de señas, a los niños sordos. En 1969 Fuller abrió la primera escuela para sordos en Estados Unidos que trabajaba con jornadas de estudio igual que las de una escuela normal [14].

La corriente del enfoque manual se contraponen al enfoque oralista y tomó su mayor fuerza en los años sesenta. Este enfoque promueve el desarrollo de la lengua de señas utilizada por la comunidad sorda para satisfacer sus necesidades comunicativas [48]. Este enfoque ve a las personas sordas, no como enfermas, sino como una comunidad diferente que tiene la necesidad de expresarse de manera distinta a la comunidad de los oyentes. El lenguaje de señas que se utiliza varía de acuerdo al país, pero podemos destacar que en Estados Unidos y parte de Canadá se utiliza el lenguaje visual ASL (American Sign Language) [34] y en México se utiliza el LSM (Lenguaje de Señas Mexicanas) [48].

El enfoque bilingüe ha sido el mejor aceptado, y consiste en tomar el concepto de que las personas con problemas auditivos no son enfermos, sino personas diferentes que deben desarrollar un lenguaje especial (lenguaje de señas) como lengua materna, pero que también reconoce la importancia de la adquisición de la lengua oral como una habilidad para su desenvolvimiento en el ámbito social y profesional. Las personas sordas se desarrollan más integralmente si se vuelven bilingües, aprendiendo la lengua oral como segundo idioma [48]. Consideramos que este enfoque es el más adecuado pues concibe a las personas que no oyen como individuos diferentes, pero no defectuosos. Por otra parte permite desarrollar habilidades de comunicación oral para mejorar la interacción con personas que oyen normalmente. Este enfoque será fundamental para nuestro trabajo propuesto.

Se debe respetar la comunidad de los sordos como una cultura diferente a la de los oyentes. Sin embargo, si las personas con algún problema auditivo aprenden a comunicarse oralmente, lograrán mejores oportunidades, así como un extranjero que aprende la lengua del país donde vive logra tener una mejor calidad de vida. Al comprender la posición de las personas con problemas auditivos podemos tener una mejor visión de las necesidades y características de los usuarios potenciales de la interfaz propuesta en este trabajo de investigación.

2.2 Aparato Fonador Humano

El aparato fonador humano es el que permite la producción de los sonidos, que producen el lenguaje necesario para la comunicación humana. Con la ayuda del flujo de aire producido por los pulmones, el aparato fonador logra generar ondas sonoras que son transmitidas utilizando como medio de propagación el aire (Figura 2.1).

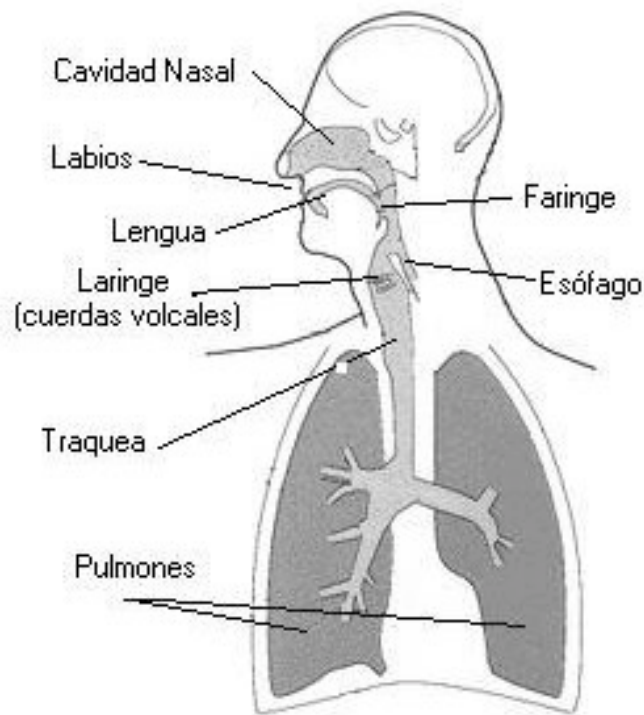


Figura 2.1. Aparato Fonador Humano [47].

Cuando las ondas sonoras llegan al oído humano, éste las conduce por el oído externo (pabellón y canal externo) hacia el oído medio, donde se encuentra el tímpano. Las ondas chocan contra la membrana timpánica y causan el movimiento de los sensores mecánicos del oído (martillo, yunque y estribo), los cuales convierten la onda acústica a

vibraciones mecánicas hacia el oído interno. El oído interno consta del nervio auditivo y la cóclea, cámara llena de líquido segmentado por la membrana basilar. Las vibraciones mecánicas chocan en la ventana ovalada de la entrada de la cóclea creando ondas en el líquido contenido, provocando que la membrana basilar vibre a frecuencias proporcionales a las de la onda acústica de entrada y en un lugar de la membrana que es asociado con estas frecuencias. El nervio auditivo toma esta información y la convierte en impulsos nerviosos que posteriormente son descifrados por el cerebro [32, 45].

El aparato fonador humano se compone de tres elementos fundamentales: sistema subglotal, laringe y sistema supralaringeo [32]. En el sistema subglotal el flujo de aire pasa de los pulmones a la tráquea y posteriormente a la laringe. En la laringe se encuentran las cuerdas vocales que reaccionan elásticamente con la presión del aire, produciendo energía acústica. Finalmente el sistema supralaringeo (nariz, boca y faringe) actúa como filtro variable y transforma en voz la energía acústica producida por la laringe.

En los lenguajes hablados, la unidad de voz más pequeña con significado propio es lo que se conoce como fonema. Todo lenguaje natural tiene su propio conjunto de fonemas, que son del orden de cuarenta fonemas por lenguaje [2].

Para cada fonema existe una posición para cada una de las articulaciones del sistema bucal: las cuerdas vocales, lengua, labios, dientes y mandíbula. Estos fonemas o sonidos son divididos en dos clases: vocales, que permiten el flujo de aire irrestringido en el tracto vocal; y consonantes, las cuales restringen el flujo de aire en algún punto y son más débiles que las vocales [4].

2.3 Teoría Fonética Articulatoria

La palabra “articulación” tiene varios significados en la fonética. Podemos entender como articulación al sonido que resulta de la unión de una o más consonantes con una o más vocales. También hay otro concepto, en el cual articulación se entiende por el movimiento de los órganos para pasar de una posición a otra, cuando se producen dos fonemas inmediatos. Hay un tercer concepto que es el que utilizaremos en la descripción de la teoría fonética articulatoria. Ésta se refiere a la modificación de los elementos del aparato resonador-articulatorio (labios, mandíbula inferior, lengua y velo del paladar), para variar los efectos acústicos del sonido producido por la laringe [7].

Más formalmente, definimos articulación como la actividad coordinada de los músculos del tracto supra-glotal que, en sincronía, producen configuraciones que derivan en diferentes sonidos del habla humano [32]. Estos patrones articulatorios para producir los fonemas, parecen ser producidos por el proceso neuronal de automatización. El cerebro graba cierta posición del tracto vocal para cada fonema aprendido y lo reproduce posteriormente de forma automática.

En la articulación de cualquier fonema entran en acción diferentes órganos. Identificamos como órganos activos a los que pueden moverse, como los labios, la lengua, la mandíbula inferior y el velo del paladar. Los órganos que permanecen estáticos son

llamados órganos pasivos, entre los cuales podemos mencionar los dientes superiores, la protuberancia alveolar y el paladar duro [7].

La teoría fonética tradicional, o teoría fonética articulatoria, fue creada por Melville Bell en 1867 [30]. En su sistema “Visible Speech”, pretendió encontrar un conjunto de configuraciones articulatorias que pudiera enseñar a gente sorda y así habilitarlos para hablar. Estas configuraciones articulatorias no eran mas que instrucciones para producir ciertos fonemas. Bell utilizó muchas hipótesis para especificar la posición de un “punto de articulación” formado por la lengua contra el paladar en la cavidad oral [30, 32]. En dicho lugar los órganos adoptan una configuración especial que produce una resonancia característica, diferente para cada fonema. Para cada realización sonora existe una determinada disposición de los órganos que se denomina “Modo de Articulación”.

El punto de articulación y el modo de articulación tienen gran importancia en la fonética articulatoria, además de ser dos de los rasgos fonéticos tradicionales que determinan las diferentes clasificaciones de fonemas (Tabla 2.1).

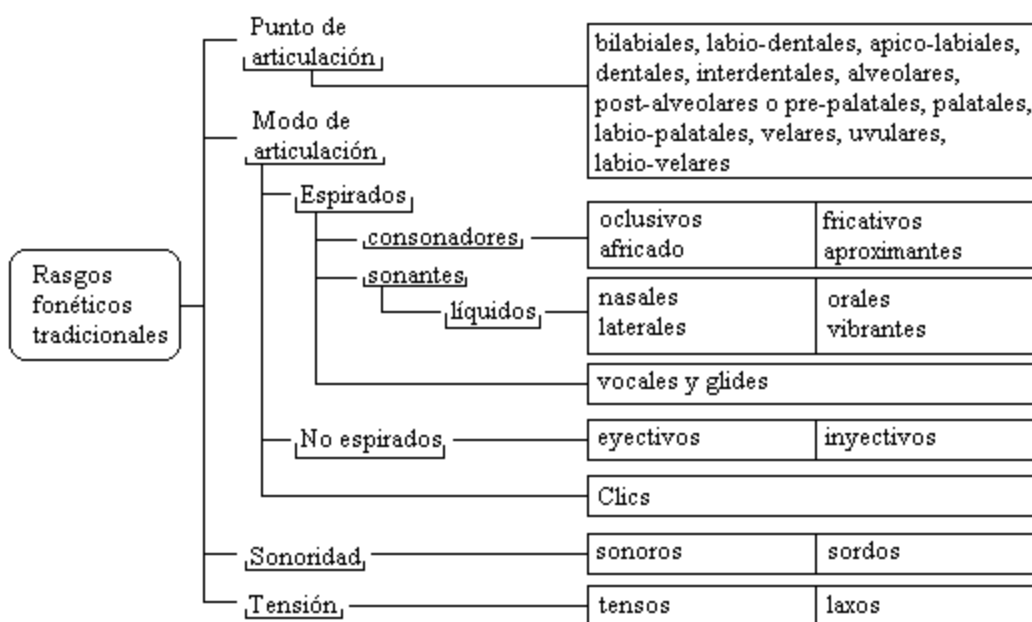


Tabla 2.1. Clasificación Tradicional de Fonemas [32].

Considerando únicamente la posición de la lengua, Helwag ha formado un triángulo ideal para los fonemas de las vocales, denominado triángulo vocálico, y cuyos vértices ocupan los fonemas de posición lingual extrema [7]. En el idioma español se utiliza un subconjunto muy reducido de los fonemas del Alfabeto Fonético Internacional (AFI) [32]. El triángulo vocálico simplificado se muestra en la figura 2.2.

La caracterización de las vocales, visto desde el punto de vista articulatorio, describe la posición de la lengua (anterior, media y posterior) y la apertura (cerrada, semiabierta, abierta). En función de lo anterior se describe el punto de articulación de las vocales [32].

	Anterior	Media	Posterior
Cerrada	i		u
Semiabierta	e		o
Abierta		a	

Figura 2.2. Triángulo Vocálico Simplificado [32].

A pesar de que la teoría de Bell, del punto de articulación, sirvió como base para desarrollar muchas otras teorías fonéticas, esta teoría se basa únicamente en la posición de la cavidad oral, y deja fuera el rol de la cavidad faríngea. Las nuevas teorías acústicas afirman ser mejores que la teoría tradicional articulatoria.

Según algunos estudios realizados con fonemas del idioma inglés con ayuda de rayos X [30], el triángulo vocálico y las teorías de Bell no siempre se cumplen. Se pueden realizar patrones de articulación diferentes y producir el mismo sonido que los descritos en el triángulo vocálico. En general, la mayoría de las personas no generan las vocales utilizando el triángulo vocálico, sino que fijan auditivamente sus propios patrones articulatorios.

No podemos afirmar que un solo patrón articulatorio produzca un cierto fonema, pueden haber muchos patrones que logren producir un mismo fonema. Lo que sabemos es que los patrones definidos por el triángulo vocálico sirven para producir las vocales especificadas. También sabemos que cada persona puede desarrollar su propio patrón fijado auditivamente desde la niñez, y es el que utiliza automáticamente durante toda su vida.

A continuación presentaremos los patrones articulatorios tradicionales para producir cada una de las cinco vocales del idioma español [7].

Fonema: A

Características: Vocal, abierto, medio. Vocal, porque el sonido sale sin obstrucción alguna; abierto, por la posición general de la boca; medio, porque el punto de articulación equidista de los puntos correspondientes a las vocales “i” y “u”. Ver la figura 2.3.

Labios: La separación de los labios es mayor que en la articulación de cualquiera de las otras vocales. Siguen pasivamente el movimiento de los maxilares.

Dientes: Están separados algo más de 1 centímetro.

Lengua: Suavemente extendida en el piso de la boca. Su punta, colocada detrás de los incisivos inferiores, está algo más baja que el borde de éstos. Los bordes de la lengua tocan los molares inferiores, y su dorso está elevado hacia la parte media de la boca. La lengua no desempeña papel activo, ni toca el paladar ni los molares superiores en ninguna parte. El punto de articulación está situado en la parte media.

Velo del paladar: El velo del paladar está levantado, apoyado contra la protuberancia faríngea, impidiendo, en esa forma, el pasaje del aire por las fosas nasales.

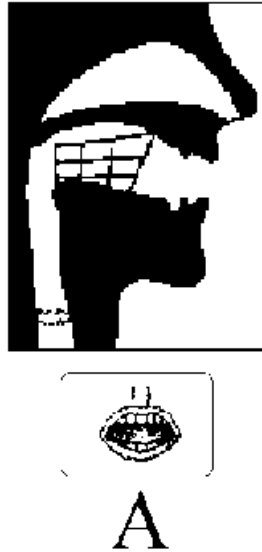


Figura 2.3. Articulación para la Pronunciación del Fonema *a* [7,43].

Fonema: E

Características: Vocal, medio, cerrado, anterior. Ver la figura 2.4.

Labios: Los labios están entreabiertos, y permiten ver los dientes y la lengua. Las comisuras están separadas. Si la boca está bien iluminada, se observa perfectamente la lengua, colocada detrás de los incisivos inferiores.

Dientes: Están separados, pero en distinto plano vertical, es decir: los incisivos inferiores, detrás de los incisivos superiores. La abertura entre los incisivos es, aproximadamente, de unos 6 milímetros.

Lengua: La punta de la lengua se coloca detrás de los incisivos inferiores, y se apoya en la cara interna de éstos; el dorso se eleva, arqueándose y tocando el paladar ampliamente, a ambos lados de su línea media, desde la mitad de los primeros molares hacia atrás. En el centro del paladar, y entre éste y la lengua, queda un canal amplio, por donde pasa el aire sonoro. Este espacio es mayor que el formado cuando se articula la “i”. La lengua descende en su base.

Velo del paladar: Levantado, impidiendo el pasaje del aire por las fosas nasales. Glotis sonora.



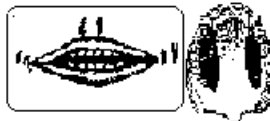
E

Figura 2.4. Articulación para la Pronunciación del Fonema *e* [7,43].

Fonema: I

Características: Vocal, anterior, cerrado. Anterior, porque la lengua ocupa la parte anterior de la cavidad bucal, y cerrado, porque el canal por donde pasa el aire sonoro, entre el dorso de la lengua y el paladar, es pequeño. Ver la figura 2.5.

Labios: La abertura labial es alargada, con las comisuras de los labios un poco retiradas hacia atrás.



I

Figura 2.5. Articulación para la Pronunciación del Fonema *i* [7,43].

Dientes: Los dientes están muy próximos, sin llegar a tocarse. Los incisivos inferiores algo atrás de los superiores, en distinto plano vertical.

Lengua: La punta de la lengua se apoya en la cara posterior de los incisivos inferiores, se arquea fuertemente, levantándose cada vez más, hasta tocar con su dorso, ampliamente, el paladar óseo, a ambos lados, dejando en el centro un canal relativamente estrecho; el contacto de la lengua y el paladar alcanza casi siempre, por delante, hasta los caninos. La estrechez del canal hace que el soplo sonoro, apresado entre el dorso de la lengua y el paladar óseo, comunique a éste vibraciones muy fuertes, que se transmite a todo el cráneo.

Velo del paladar: El velo del paladar está levantado, apoyado contra la protuberancia faríngea, impidiendo, en esa forma, el pasaje del aire por las fosas nasales. Glotis sonora.

Fonema: O

Características: Vocal, posterior, semiabierto, redondeado. Posterior, porque la lengua ocupa la parte posterior de la cavidad bucal, cuando se articula el fonema, en lugar de la parte anterior, como lo hace cuando se emite la “i” y la “e”. Redondeado, porque los labios toman una forma ovalada y ligeramente protuberante. Ver la figura 2.6.

Labios: El músculo orbicular de los labios se contrae, y éstos se aproximan un poco y avanzan, tomando su abertura una forma ovalada. Los labios se separan ligeramente de los dientes, los que no son visibles. Los maxilares se juntan un poco más que en la emisión de la “a”.

Lengua: Se recoge hacia el fondo de la boca; su punta toca la protuberancia alveolar de los incisivos inferiores; su parte anterior está algo deprimida, mientras que el dorso se eleva hacia el velo del paladar, de manera que reduce el canal de pasaje del aire, aunque sin tocar el paladar ni los límites alveolares de la arcada dental superior. No es visible.

Velo del paladar: Se levanta a su altura máxima. La corriente aérea sonora sale por la boca. Glotis sonora.

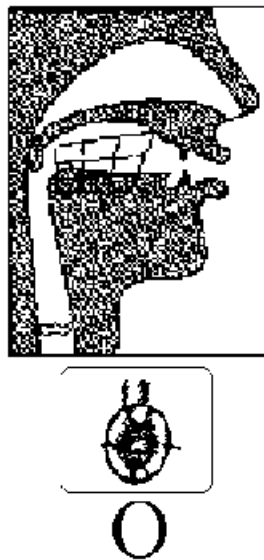


Figura 2.6. Articulación para la Pronunciación del Fonema o [7,43].

Fonema: U

Características: Vocal, posterior, cerrado, redondeado. Cerrado y redondeado por la forma que toman las mandíbulas y los labios. Por la posición de los órganos, es el fonema opuesto a la “i”. Ver la figura 2.7.

Labios: Los labios se aproximan un poco más que cuando se articula la “o”, formando una abertura ovalada, bastante pequeña. Avanzan, separándose de la cara anterior de los dientes, los que no se ven. Los labios llegan casi al máximo de prolongación, en su sentido antero posterior.

Lengua: La punta de la lengua, colocada detrás de los incisivos inferiores, se separa de la protuberancia alveolar de éstos. La lengua se retira hacia el fondo de la cavidad bucal, haciendo más estrecho el canal de pasaje del aire que cuando se articula la “o”, aunque sin llegar a tocar el velo del paladar. No es visible.

Velo del paladar: Se levanta, impidiendo el pasaje del aire a las fosas nasales. Glotis sonora.

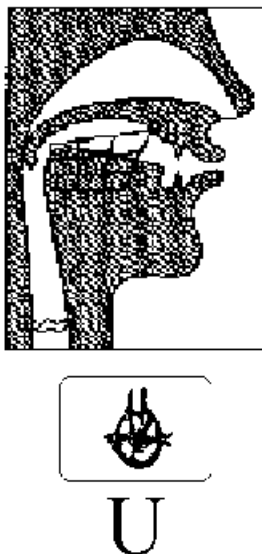


Figura 2.7. Articulación para la Pronunciación del Fonema *u* [7,43].

2.4 Modelo físico de la propagación del sonido en el tracto vocal

El modelo fisiológico nos va a servir para que, dada la señal de voz, podamos reconstruir la posición articulatoria del tracto vocal.

El tracto vocal es un tubo acústico no uniforme. En personas adultas tiene un largo de 17 centímetros aproximadamente. Comienza en las cuerdas vocales que se encuentran en la laringe, y termina en los labios. Su área seccional varía desde cero hasta 20 cm² por el control muscular de las articulaciones de la voz [4].

Para procesar la voz se debe contar con un modelo matemático que se desprenden del análisis de la propagación del sonido en el tracto vocal, producido por el aparato

fonador humano. Se comenzará con un modelo básico, suponiendo que las ondas se propagan de modo plano, y no hay pérdidas por viscosidad y conducción térmica ni en el fluido ni en las paredes [32]. De acuerdo al desarrollo presentado en [32], aplicando las leyes de conservación de la masa y la conservación del momento y la energía, obtenemos las ecuaciones (2.1) y (2.2) que modelan nuestro sistema inicial del tracto vocal (Figura 2.8), con área variable que depende del tiempo y la posición [32].

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial(u/A)}{\partial t} \quad (2.1)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial(pA)}{\partial t} + \frac{\partial A}{\partial t} \quad (2.2)$$

Donde,

$p = p(x,t)$	Presión del sonido en el tubo en la posición x en el instante t .
$u = u(x,t)$	Velocidad del flujo en el volumen en la posición x en el instante t .
ρ	Densidad del aire en el tubo.
c	Velocidad del sonido en el aire.
$A = A(x,t)$	Función de área del tubo.

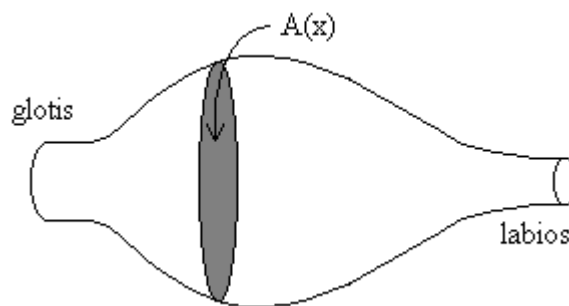


Figura 2.8. Tracto Vocal con Área Variable.

Con el propósito de obtener un algoritmo computacional en tiempo real, se ha supuesto que este modelo puede ser simplificado a un sistema de tubos concatenado de diferente tamaño (Figura 2.9).

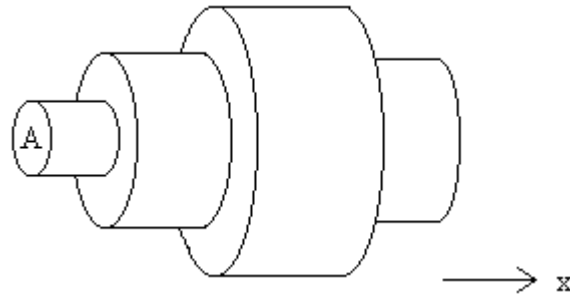


Figura 2.9. Tracto Vocal como la Concatenación de Tubos.

Al pasar la onda de presión, generada desde los pulmones, por este sistema, se puede observar que en cada unión de los tubos parte de la onda de presión se refleja y parte continúa. La cantidad de onda que se refleja se puede cuantificar utilizando un índice de reflexión. El índice de reflexión varía de acuerdo a que tan diferente es el área entre los dos tubos. Entre mayor sea la diferencia de áreas, mayor cantidad de ondas chocan y son reflejadas, y esto hace que el índice de reflexión sea mayor. Estos índices se conocen como coeficientes de reflexión k_i . Para calcular estos coeficientes, se utilizará el método de Levinson-Durbin (Anexo A), por ser el más eficiente.

Mediante una relación con los coeficientes de correlación parcial, se obtienen las razones de área i :

$$\frac{A_{i+1}}{A_i} = \frac{1 - k_i}{1 + k_i} \quad (2.3)$$

Despejando la ecuación (2.3) obtenemos la siguiente expresión para calcular las áreas de cada tubo del modelo de tubos propuesto para representar el tracto vocal:

$$A_{i+1} = \left(\frac{1 - k_i}{1 + k_i} \right) A_i \quad (2.4)$$

Ya que esta área depende de un área inicial (por ejemplo $A_0 = 1$), entonces los valores de A_i estarán normalizados a este valor inicial de área A_0 .

2.5 Reconocimiento de Voz

En la interfaz desarrollada en este trabajo se utiliza el reconocimiento de voz para verificar que la pronunciación de la vocal a practicar este siendo reconocida como tal. Esta

verificación con reconocimiento de voz es opcional en el sistema, pues requiere una conexión a Internet.

El reconocimiento de voz es el proceso de traducir una señal acústica hablada en un mensaje lingüístico [32]. El ser humano aprende a temprana edad a realizar este proceso, y empieza a entender el significado de los sonidos con la ayuda del cerebro.

Las computadoras están teniendo esa evolución, tratando de descifrar las señales acústicas proporcionadas por el ser humano para obtener datos e información con significado que posteriormente podrá procesar y transformar.

En la década de los años 50's se iniciaron los trabajos construyendo reconocedores de palabras y fonemas básicos para un solo hablante. Ya en los años 60's se generaliza el estudio de los sistemas de reconocimiento y se introducen sistemas digitales. Diez años después, en los 70's, destaca Itakura en los Estados Unidos por el uso de coeficientes de predicción lineal, orientado al reconocimiento de palabras aisladas. En la década de los 80's se pasa al reconocimiento de palabras conectadas. Actualmente se realizan enormes esfuerzos para abordar el problema de reconocimiento de discurso continuo [32].

A pesar de que el reconocimiento de voz por medio de la computadora es una tecnología que está en desarrollo, es un área que ha tenido mucho interés y se han logrado grandes avances, como el desarrollo de interfaces hombre-máquina por voz, de manera que la entrada y salida de información a los sistemas de cómputo sea por voz. Su uso intensivo ya es un hecho, como parte de los sistemas de captura de información hablada [32, 33].

Uno de los objetivos principales que se buscan al desarrollar sistemas de reconocimiento de voz, es construir sistemas de reconocimiento del habla continua, que sean independientes del hablante, que contengan grandes vocabularios y que se realicen en tiempo real [32, 55].

Algunas de las técnicas más importantes que se han utilizado para realizar el reconocimiento son: contornos energéticos, alineamiento en tiempo, cuantización vectorial, alineamiento dinámico temporal, modelos ocultos de Markov, redes neuronales artificiales, modelos difusos y máquinas vectoriales de soporte [32, 55].

Uno de los factores más importantes antes de realizar un sistema de reconocimiento de voz es elegir la unidad de reconocimiento. Estas unidades van desde fonemas hasta palabras o frases. El problema más complejo se presenta cuando se requiere reconocimiento de voz continua. En estos casos se requiere dividir la señal de voz en secciones significativas (palabras) y después usar una estructura gramatical para asignarle un significado en conjunto [2]. Para realizar el reconocimiento de palabras se pueden utilizar dos técnicas: segmentación fonética y segmentación por bloques (o lineal) [55].

Antes de realizar el reconocimiento es importante realizar una etapa previa de pre-procesamiento donde se convierte la señal de analógica a digital, se extraen características de la señal y se realiza el entrenamiento de los modelos a utilizar [15, 35].

Ya que se tienen los modelos entrenados se pasa a la etapa de reconocimiento que consta de tres principales tareas: pre-procesamiento digital de señales, comparación de modelos y clasificación [2]. En la figura 2.10 se puede observar un diagrama del proceso completo de reconocimiento automático de la voz.

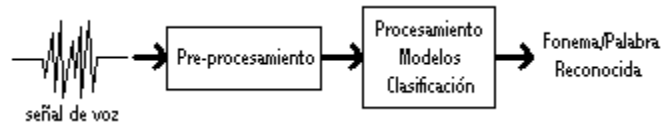


Figura 2.10. Proceso de Reconocimiento de Voz.

El pre-procesamiento de la señal es necesario para obtener una señal de voz con formato adecuado, libre de silencios externos y ruido ambiental. Para atenuar la influencia del ruido se utiliza un filtro pasabanda [2]. También en el pre-procesamiento es necesario aplicar un pre-énfasis a la señal para ecualizar el rango dinámico de las altas y bajas frecuencias y finalmente segmentar la señal y aplicar un ventaneo [2, 35].

Los parámetros característicos que se extraen de cada segmento o bloque de datos son los coeficientes Mel Cepstrales. Estos parámetros son comparados con lo que se tienen en el modelo previamente entrenado y se clasifican tomando como base este modelo. Con esto se obtiene el reconocimiento de voz de la señal de entrada.

El reto del reconocimiento automático de voz poco a poco se ha ido tratando de resolver con la aplicación de técnicas cada vez más complejas, sin embargo, es obvio que aún falta mucho por realizar [3].

El reconocedor utilizado en la interfaz propuesta es el *SPHINX 3*, que fue desarrollado en la *Carnegie Mellon University* [29]. Ya que actualmente la interfaz funciona en ambiente Windows y el reconocedor *SPHINX* funciona en ambiente Linux, se desarrolló una aplicación cliente-servidor para verificar la vocal reconocida a partir de un archivo de voz.

El reconocedor funciona con Modelos Oculto de Markov (HMM) de cinco estados, con 8 Gaussianas por estado. El pre-procesamiento de la señal de voz utiliza los siguiente parámetros: la frecuencia de muestreo de la señal es de 16000 Hz, la frecuencia de muestreo del “frame” es de 100 Hz, la longitud de la ventana Hamming es de 30 ms, se utilizó una FFT de 512 muestras, se utilizaron 31 bancos de filtros, la frecuencia mínima utilizada fue de 100 Hz, la frecuencia máxima utilizada fue de 3500 Hz, se utilizaron 13 coeficientes cepstrales.

Para obtener la base de datos de vocales se grabaron archivos de vocales con la siguiente información: "a-e-i-o-u", "e-o-i-u-a", "i-u-a-e-o", "o-a-u-i-e" y "u-e-o-a-i". Cada voluntario pronunciaba cinco veces cada combinación. En total, cada voluntario aportó 25

archivos de entrenamiento. Se obtuvieron 800 archivos de entrenamiento de 32 personas, de ambos sexos, de entre 17 y 25 años.

El reconocedor tiene un porcentaje de reconocimiento del 98% para los mismos archivos de entrenamiento. Este porcentaje de reconocimiento disminuye en el uso del reconocedor desde la interfaz, pues en la herramienta se crean archivos de voz que contienen una vocal aislada, sin la presencia de silencios.

2.6 Interfaz Hombre-Máquina

La computadora puede ser vista desde diferentes puntos de vista. Hay quienes la ven como un aparato tecnológico incomprensible, temido hasta cierto punto y que solamente las personas jóvenes pueden utilizar. También puede ser vista como un sustituto de máquina de escribir o un aparato de entretenimiento en donde pasar el tiempo jugando.

Desde el punto de vista educacional es concebida desde dos perspectivas. La primera ve la computadora como un conjunto de contenido escolar que debe ser aprendido. En los primeros años en los que las computadoras se integraron en la educación se veía de esta manera, considerándola un fin en sí misma. Entonces se enseñaba la historia de la computación, la teoría, las partes de la computadora y programación para niños [48].

El otro punto de vista ve a la computadora como una herramienta. Una buena interfaz permite comunicar a los seres humanos con las aplicaciones que le ayudarán a realizar actividades escolares, científicas y profesionales en menor tiempo y con mayor precisión. Cuando se tiene esta visión de la computadora, se puede sacar mucho provecho de ella, interactuando de manera natural para obtener resultados satisfactorios.

En los años 70's iniciaban las primeras formas de interacción con las computadoras. Las interfaces humano-computadora consistían en lectores de tarjetas perforadas e impresoras de líneas. La interacción era a un tiempo y espacio distantes. El estudio de la interacción humano-computadora es una disciplina relativamente joven [11].

Uno de los problemas más importantes en el diseño de los sistemas computacionales es la interfaz entre el ser humano y la máquina. Asimismo es un elemento de suma importancia en la puesta en marcha de todo sistema computacional. Es el medio de interacción y comunicación entre el ser humano y la máquina, y es la carta de presentación del sistema. Va más allá de incluir gráficos y sonidos, toma en cuenta conceptos cognitivos y emocionales de la experiencia del usuario y toma en cuenta su deseo de enfrentarse a algo fácil, pero a la vez poderoso. La interfaz de interacción ideal sería un diálogo natural como entre los seres humanos, esto es, una conversación con la computadora como si fuera un ser humano [28, 37, 52].

En la mayoría de los sistemas, la creación de interfaces de usuario amigables comprende la mitad del esfuerzo requerido en la etapa ingenieril del software. Estas interfaces requieren conocimiento detallado en las áreas de la ingeniería humana, despliegue de diseños y gráficos [44].

El diseño de la interfaz puede determinar la aceptación o rechazo del sistema completo. Por tal motivo se le está dando mayor importancia dentro del desarrollo de sistemas. Se estima que del 35% al 45% de los gastos destinados a un proyecto son direccionados al diseño de la interfaz [52].

Detrás de la creación de buenas interfaces encontramos elementos multidisciplinarios, como diseño gráfico, diseño de comunicación visual, diseño industrial, diseño ergonómico y principios psicológicos. Con estos elementos se debe lograr una comunicación efectiva, tanto funcional como estética, así como reducir el tiempo de asimilación por parte del usuario [28, 52].

Los diseñadores de las interfaces deben entender mejor el impacto de sus decisiones, pues de esto dependerá que inhabiliten o faciliten el uso de sus sistemas a todo tipo de usuarios. Los diseñadores deben conocer a sus usuarios potenciales. Deben tener en cuenta su nivel de dominio de computadoras y sus capacidades físicas y mentales. Un criterio común en los desarrolladores de interfaces humano-computadora es “Conocer al Usuario” [5, 11].

Cuando los usuarios son novatos, de edad corta o tienen alguna discapacidad, es importante hacer interfaces simples y uniformes. Simple significa que tenga muy pocas cosas que memorizar para utilizar el sistema y que sea fácil de aprender [23, 27].

Hay otras consideraciones que se deben seguir en el diseño de cualquier interfaz de usuario. Los aspectos más importantes son: que sea interesante y agradable, que tenga tiempo de respuesta rápido, que sea confiable, preciso, usable y libre de errores. También es importante utilizar un vocabulario que sea bien entendido por los usuarios en las instrucciones y botones [5, 11, 27].

Al diseñar interfaces gráficas se deben tomar en cuenta aspectos como: elegancia, simplicidad, escala, contraste, proporción, organización de estructura visual, modularización, representación de imágenes, y guías de estilo. También es importante evitar sobrecargar al usuario de información, ya que es conocido que el ser humano puede manejar pocos conceptos de manera simultánea [41, 52].

Al diseñar una interfaz, se debe involucrar a los usuarios potenciales del sistema y tomar en cuenta sus necesidades, experiencias y capacidades, así como las limitaciones físicas y mentales de las personas que utilizarán el producto de software. Las capacidades de los usuarios son la base para la guía de diseño [41].

La interfaz hombre-máquina puede tener diferentes estilos de interacción: interfaz de preguntas, menú simple, orientada a ventanas, entre otros. Todos ellos enfocados a lograr una interfaz amigable para el usuario [52].

En la interfaz de preguntas la computadora va obteniendo los datos que necesita para procesarlos por medio de preguntas concretas realizadas al usuario. En este tipo de interfaz el humano simplemente reacciona a las preguntas predeterminadas que la interfaz va solicitando. Este tipo de interfaz fue la utilizada en los inicios de la computación, cuando

no se contaba con muchos recursos para hacer una interfaz más amigable. Aunque por parte de la computadora es una forma fácil de interactuar, para el humano puede resultar poco flexible y difícil de entender.

En la interfaz del menú simple, la computadora presenta una serie de opciones que puede realizar, y el humano tiene la libertad de seleccionar la que desea. Aunque es muy parecida a la de preguntas, en ésta se tiene una visión más amplia de todas las funciones del sistema. Este tipo de interacción está conducida por el humano, pero está limitada a las opciones que la interfaz presenta.

Al avanzar el desarrollo y capacidades de las computadoras se lograron mejores interfaces de interacción, como la orientada a ventanas. En este tipo de interacción se pueden manejar una interactividad múltiple, pues el usuario puede estar interactuando con distintas funcionalidades del sistema al mismo tiempo, teniendo una ventana para cada opción. Este tipo de interacción puede confundir a usuarios inexpertos y se pueden perder fácilmente si no está familiarizado. A pesar de ello, este tipo de interfaz es mucho más fácil de utilizar y más amigable. Por su fácil manejo y organización, éste tipo de interfaz será utilizada en el software realizado en este proyecto.

La interfaz no es un elemento aislado en la creación de un producto de software, pues es el principal medio de comunicación entre máquina y hombre. La meta de un buen diseño de interfaz es transmitir un mensaje que pueda ser recibido de manera exacta y a tiempo por el usuario. Es un elemento fundamental y debe ser visto como una parte importante del sistema completo [6, 52].

Del buen diseño de una interfaz depende que una persona conciba la computadora como una herramienta amigable que le haga la vida más fácil, o que conciba a la computadora como un aparato tecnológico que solo le trae dolores de cabeza y que es casi incomprensible.

Es importante mencionar que al diseñar interfaces e interacción en sistemas dirigidos a personas con alguna discapacidad hay que tomar ciertas consideraciones extras. A pesar de que se están haciendo esfuerzos por hacer sistemas comerciales que tengan características adaptables a las diferentes discapacidades es necesario conocer las limitaciones de las personas que usarán el sistema.

La tendencia en interfaces computacionales de usuario parece dirigirse hacia más gráficas y voz y menos texto plano para datos de entrada y salida. Dispositivos de salida de alta calidad en gráficos y síntesis de voz están siendo creados con enorme potencial para las nuevas interfaces de usuario [27].

Para realizar el diseño de la interfaz propuesta, se buscó utilizar gráficos simples y fáciles de entender. Para crear la figura que muestra el corte de perfil, se utilizaron colores primarios (rojo, amarillo y azul) o secundarios (verde, morado y anaranjado) para mantener simplicidad y hacerla atractiva. El objetivo fue crear una interfaz simple, entendible y al mismo tiempo llamativa.

CAPITULO 3

Estado del Arte

Existen varios métodos tradicionales para enseñar a las personas con problemas auditivos a pronunciar correctamente fonemas y palabras del lenguaje español. La metodología empleada varía de acuerdo a la escuela y el maestro.

El educador español Juan Pablo Bonet, fue el primero en publicar una obra sobre el arte de enseñar a los mudos. Su libro titulado “Reducción de las letras, y arte para enseñar a hablar a los mudos”, fue publicado en Madrid en el año 1620. En su libro, se sientan las bases para enseñar los fonemas de forma separada, para luego unirlos y así formar las palabras [12].

Actualmente, las técnicas que el maestro utiliza para enseñar las vocales se fundamentan en sustituir la audición por algún otro medio de captación, como la vista y el tacto, y en los casos de pérdida parcial del oído, aprovechando los restos de audición. La vista juega un papel muy importante cuando se pretende enseñar una correcta pronunciación.

La técnica más común utilizada para enseñar la pronunciación es mostrando al alumno la correcta posición de los labios y dientes utilizando como muestra la boca del maestro. También se utilizan pedazos de cuero u otros materiales para simular la lengua y así mostrar las posiciones de la lengua que no se pueden observar a simple vista [1].

Otros utensilios de ayuda que el maestro utiliza son espejos de 25 a 35 centímetros, ya sea para que el alumno se vea a sí mismo o para que vea al maestro inversamente. En otros casos se ha llegado a utilizar espátulas o cucharas de metal y hasta mondadientes [7].

El alumno tratará de igualar las posiciones aprendidas. Primero se aprenden los fonemas, comenzando por las vocales, que son más fáciles de reproducir. Es importante que sepan distinguir los diferentes fonemas. Después se procede a enseñar palabras simples (una o dos sílabas) uniendo los fonemas ya conocidos. Por último se enseñan aquellas palabras que son más complejas.

Como en muchas otras áreas, las computadoras poco a poco también están llenando necesidades de educación especial. A pesar de esto, los programas computacionales para apoyar la educación especial en las instituciones públicas y privadas de México son pocos y difíciles de conseguir. Muchas escuelas de educación especial, privadas y de gobierno, cuentan con la infraestructura computacional, pero carecen de software adecuado y personal capacitado.

Entre los productos de software que se están utilizando en escuelas para sordos del estado de Nuevo León, como ayuda a los maestros, se encuentran algunos programas orientados a la enseñanza de la lectura para niños normales, sin discapacidad auditiva. Aunque se cuenta con este tipo de software en algunas instituciones, no es la tecnología adecuada, pues está enfocado a enseñar a leer a niños que escuchan perfectamente bien. A pesar de que no están enfocados a educación especial, los maestros utilizan estos programas como un apoyo a los niños, y si bien en algunos casos les ayudan a identificar movimientos de la boca (lectura de labios) o identificación de la grafía de las letras, no son el tipo de programas adecuados para sordos o hipoacúsicos.

Con respecto a proyectos de otros países enfocados a ayudar a las personas con problemas auditivos, encontramos muchas iniciativas de investigación y algunos prototipos que podrían ser una herramienta valiosa en un futuro. Hay una diversidad de áreas que la tecnología está alcanzando, para mejorar el nivel de vida de los discapacitados auditivos.

Por un lado tenemos los proyectos de software que se basan en el lenguaje de señas, tratando de traducir señas a lenguaje escrito o hablado, y viceversa [16, 31, 40]. Otro tipo de proyectos de software enfocados a ayudar a personas con problemas auditivos son los que utilizan tecnología de procesamiento y/o reconocimiento de voz. Estos proyectos tienen como principal objetivo trabajar con aspecto de la voz, mejorando la producción de lenguaje hablado.

3.1 Software para Traducción de Lenguaje de Signos

En Estados Unidos se está desarrollando un sistema interactivo de gráficas computarizadas para el análisis y modelación de movimientos de lenguaje de signos americanos (ASL). Este sistema contiene cuatro componentes.

El primer componente reconstruye movimientos en tres dimensiones y permite segmentar y transformar los datos. El segundo componente permite construir signos sintéticos con funciones de ángulo en un modelo articulado. En el tercer componente es posible manipular la calidad del movimiento independientemente de su ubicación espacial. El último componente permite construir estímulos complejos y secuencias en tiempo real para experimentos psicolingüísticos. La forma de obtener las posiciones tridimensionales de los movimientos es por medio de cámaras (*optoelectronic cameras*) que rastrean las posiciones de diodos de luz colocados en brazos y manos, video grabados en cuartos oscuros [31].

Otro ejemplo que podemos citar en esta área es un prototipo que se está realizando en Japón, en el Laboratorio Central de Investigación de Hitachi. Este sistema utiliza reconocimiento de patrones y síntesis para traducir el Lenguaje de Signos Japonés (JSL). El sistema tiene dos módulos. El primer módulo traduce el lenguaje de signos al japonés. Lee la forma y movimiento de las manos usando DataGlove (guantes especializados utilizados en realidad virtual). Reconoce los movimientos obtenidos y los traduce al lenguaje japonés. El otro módulo traduce el lenguaje japonés al lenguaje de signos japonés. El lenguaje de

signos es desplegado como ademanes animados usando gráficas computacionales tridimensionales [40]. En la figura 3.1 podemos ver un esquema de este sistema.

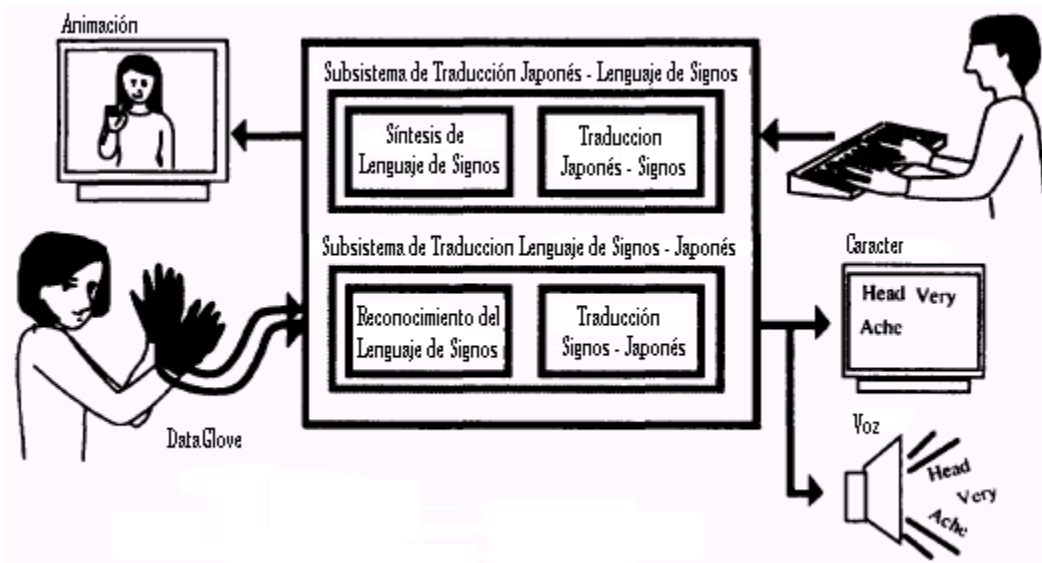


Figura 3.1. Sistema de Traducción Automática de Lenguaje de Signos [40].

3.2 Software Comercial y en Desarrollo con Tecnología de Voz

3.2.1 Speech Viewer de IBM

El producto de software que más se acerca a las necesidades de las personas con problemas auditivos y que es utilizado en centros de rehabilitación, como el DIF de Monterrey y el Centro de Atención Múltiple Instituto Sertoma, es el *Speech Viewer*, producto desarrollado por IBM [21].

Este software es una herramienta en el tratamiento y diagnóstico de trastornos del habla, del lenguaje y de la audición. En este software se manejan atributos del habla tales como: sonoridad, tono, intensidad, precisión en la pronunciación de segmentos del habla y de fonemas sostenidos y cadencia del habla.

El *Speech Viewer* es un producto bien elaborado que contiene muchos módulos para los trastornos del habla, abarcando varias áreas de rehabilitación. Algunos de los módulos especialmente diseñados para mejorar la pronunciación son los siguientes:

- Creación de Técnicas de Precisión de Fonemas: Su objetivo es mejorar la precisión de la producción de fonemas. La visualización de este módulo indica proximidad de un fonema pronunciado respecto a un modelo de fonema objetivo. La proximidad es

medida en una escala de cercanía. La pantalla de acción tiene una escala en diagonal, representada por un gráfico. La parte superior del gráfico representa una producción igual al del modelo objetivo. La parte inferior representa la mayor desviación del modelo objetivo [21].

- Creación de Técnicas de Encadenamiento de Fonemas: Su objetivo es desarrollar habilidades en la pronunciación de una secuencia de dos a cuatro fonemas. La visualización de este módulo indica el momento en que el primer fonema pronunciado se reconoce dentro del umbral de un modelo de fonema objetivo. A continuación, se indica el momento en que se reconoce un segundo fonema con relación a un segundo modelo de fonema objetivo. El procedimiento se puede repetir hasta cuatro fonemas. Este módulo se recomienda utilizar después de dominar el módulo de precisión de fonemas aislados [21].
- Creación de Técnicas de Contraste de Dos Fonemas: El objetivo de este módulo es mejorar la precisión al contrastar la pronunciación de cada fonema en un conjunto de dos. La visualización indica el momento en el que el ejemplo pronunciado se reconoce dentro del umbral de uno de los dos modelos de fonemas objetivo seleccionados [21].
- Creación de Técnicas de Contraste de Cuatro Fonemas: Este módulo es muy parecido al de contraste de dos fonemas, solo que trabaja con cuatro fonemas [21].

Estos cuatro módulos son útiles para tratar de mejorar la pronunciación de fonemas y concatenación de fonemas. El sistema también incluye funciones para fijar los patrones de voces de mujer y hombre, definir modelos de fonemas, definir modelos de segmentos del habla, mantener el registro de cada usuario y tener acceso a estadísticas para monitorear avances de cada usuario. El sistema trabaja con reconocimiento de patrones y el usuario no puede saber sus errores en caso de que el sistema no reconozca los fonemas que pronunció.

El *Speech Viewer* también contiene una sección de módulos de estructuración. En estos módulos si se proporciona retroalimentación visual y auditiva de lo que el usuario está pronunciando, pero no se muestra la articulación para una mejor pronunciación, ni se muestra como mejorar en caso de estar equivocado, sino que trabaja con atributos del habla como: tono, intensidad, sonoridad, ondas, espectrogramas, patrones temporales y espectros. Los módulos que se incluyen en esta sección son:

- Estructuración de Tono e Intensidad: Su objetivo es aumentar la habilidad en la producción de patrones del habla aceptables proporcionando una retroalimentación visual y auditiva sincronizada con una variedad de visualizaciones técnicas [21].
- Estructuración de Ondas y Espectrograma: El objetivo de este módulo es proporcionar retroalimentación auditiva y permitir examinar la prosodia del habla al mismo tiempo que centrarse en detalles del contenido de ondas y de espectros en segmentos seleccionados de un ejemplo del habla [21].

- Estructuración de Espectros: Se busca mejorar la precisión en la producción de fonemas mediante la utilización de espectros dinámicos para proporcionar retroalimentación [21].

Los módulos de reconocimiento de fonemas comparan los espectros de los fonemas pronunciados con los espectros de los modelos. Los modelos están formados por ejemplos del habla que, a juicio del terapeuta, poseen las cualidades de fonemas necesarias para producir el habla objetivo.

Este producto contiene módulos para trastornos del habla y cuenta con visualizaciones gráficas y animaciones atractivas e interesantes. Muestra retroalimentación positiva cuando la tarea se completa y es satisfactoria. En algunos módulos utiliza un tipo de retroalimentación de lo que el usuario pronunció, pero se basa en aspectos que no tienen mucho significado para tratar de igualar. No muestra los errores cometidos por el usuario. No muestra la correcta pronunciación de fonemas ni la posición de la mandíbula, dientes, velo y lengua (articulación).

3.2.2 CSLU Toolkit

El *CSLU toolkit* [26, 54, 56] (*Center for Spoken Language Understanding*) del *Oregon Graduate Institute* es un producto que contiene un conjunto de herramientas que incorporan reconocimiento, síntesis y producción de voz, así como una aplicación de desarrollo y un agente conversacional animado llamado Baldi.

El agente está representado por una cara animada en tercera dimensión que produce expresiones faciales sincronizadas con el sintetizador de voz. Las expresiones de este agente incluyen movimientos de labios, quijada, cejas, ojos y lengua. También puede llegar a expresar emociones y puede ser visualizado con opciones semitransparentes para apreciar el interior de la boca mientras pronuncia las palabras.

Con este toolkit se pueden realizar aplicaciones interesantes para las personas con problemas auditivos, pues pueden leer los labios del agente Baldi y tratar de igualar los movimientos realizados por él. En la sección 3.3.3 veremos una aplicación de este tipo.

3.2.3 Sistema Panasonic para Enseñar a Hablar a Niños con Sordera Profunda

Este proyecto se está desarrollando en el Laboratorio de Tecnología de Voz de Panasonic [23]. Es un sistema para enseñar a hablar a niños con sordera profunda usando acústica sintetizada y patrones articulatorios.

Este proyecto integra un sistema texto-a-voz para sintetizar modelos de producción, incluyendo patrones de contacto de la lengua con el paladar. Con esto se provee un ambiente de entrenamiento mejorado para la producción de voz. Una característica interesante es que el número de articulaciones es ilimitado, proveyendo independencia de entrenamiento al estudiante, al no requerir la supervisión de un experto.

El sistema contiene una integración de múltiples dispositivos de computadora, los cuales utilizan un sintetizador *STLtalk* modificado para proveer un modelo de entrada y salida de voz, y un sistema de sensores acústicos y articulatorios para extraer los parámetros de entrenamiento.

Entre los sensores de entrada podemos encontrar un palatógrafo (*palatograph*) dinámico que extrae los datos de contacto de la lengua con el paladar. El palatógrafo es un paladar artificial que contiene una serie de electrodos que pasan una corriente de amperaje bajo cada vez que son tocados por la lengua. Los demás sensores se muestran en la figura 3.2.

El sistema cuenta con trece diferentes programas computarizados para entrenar la voz. Los programas fueron diseñados como videojuego para hacerlos más atractivos a los niños. La clasificación de los programas es de acuerdo con el aspecto de la voz que trabaja. Tenemos algunos que trabajan con el contacto lengua-paladar, los cuales fundamentalmente utilizan el palatógrafo. Otros trabajan con el tono o con el flujo de aire. También incluye algunos programas que utilizan múltiples parámetros, trabajando varias características de la voz a la vez, como: amplitud, nasalización, tono y sonoridad.

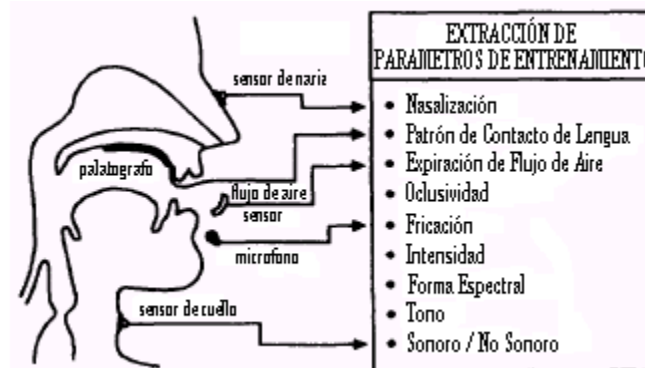


Figura 3.2. Sensores Acústicos y Articulatorios [23].

El sistema crea modelos de parámetros de la voz, que son utilizados para comparar los parámetros de producción de los niños y así evaluar su desempeño. Los dos principales beneficios de proveer modelos articulatorios y acústicos sintetizados son: que los niños pueden practicar sin la supervisión de un maestro, teniendo la posibilidad de extender los periodos de entrenamiento; y que el modelo puede ser adaptado individualmente, pues los patrones pueden ser fijados para cada niño por el maestro.

Este sistema provee una muy buena retroalimentación de los parámetros de la voz que se deben realizar para producir todos los fonemas, aunque en cuestión de articulación sólo provee datos del contacto de la lengua con el paladar.

3.2.4 Pronunciation Power

Existen otras aplicaciones que muestran retroalimentación visual de ondas sonoras, pero no están orientados a ayudar al usuario a mejorar la pronunciación. Tenemos el *Pronunciation Power* [26], orientado a aprender otro lenguaje. Aquí se pueden comparar representaciones gráficas de las ondas sonoras de una pronunciación con un patrón correcto, pero no presenta la comparación en forma de articulación, sino solo para ver si fue igual o no.

3.2.5 ViaVoice de IBM

El *ViaVoice* de IBM [26] presenta retroalimentación en forma de una calificación. Después de grabar una frase, el sistema presenta una calificación de que tan correcto estuvo, pero no puede especificar porque fue el error, ni como corregirlo.

3.2.6 Otros Productos de Software para la Rehabilitación de la Voz

Existen varios productos de software orientados a la rehabilitación de problemas de voz y habla en general, pero no están orientados a la articulación para mejorar la pronunciación de fonemas. Como ejemplos tenemos [54]:

- CSL, Computerized Speech Lab, Kay Elemetrics. Es un sistema para la adquisición, análisis, edición y procesamiento de voz para PC's. Incluye una base de datos que permite ver el comportamiento de parámetros de la voz cuando está presente un desorden de la voz.
- ISA, Intelligent Speech Analyser, Oy Pitchsystems Ltd. Desarrollado para Macintosh. Se enfoca a fonética, foniatría, logopedia, audiología, análisis de la voz, análisis del sonido, análisis del canto, análisis de la música, análisis de instrumentos musicales, investigación sobre gritos de niños, investigación sobre sonidos del pulmón y corazón, análisis de las voces del radio, edición de sonido.
- Signalyze, InfoSignal Inc. Programa interactivo con herramientas para la edición, análisis y manipulación de señales de voz y otros materiales acústicos.
- EXPARM V1.2. Software para el análisis y comparación del habla, enfocado a la extracción y visualización en tiempo real de perfiles acústicos de parámetros de la voz. Para la extracción de los parámetros se aplican filtros digitales y sintetizadores.

3.3 Software Desarrollado en México con Tecnología de Voz

Existen otros trabajos realizados en universidades de México, relacionados al habla utilizando tecnología de procesamiento y reconocimiento de voz en productos de software.

Todos ellos se presentan como investigaciones y prototipos que sientan las bases para productos de software más elaborados en un futuro.

3.3.1 Software para Detectar Trastornos del Habla

En el Instituto Politécnico Nacional se desarrolló un producto de software basado en procesamiento de voz, orientado a detectar el tipo de trastorno del habla y recomendar el tratamiento de rehabilitación. Este software permite la adquisición, procesamiento y visualización de parámetros acústicos de la señal de voz para su interpretación, siendo su principal finalidad el diagnóstico y rehabilitación de pacientes con trastornos de la voz y el habla [18].

Está encaminado a pacientes menores de 10 años. Este proyecto fue desarrollado con la colaboración del Instituto Nacional de Comunicación Humana, quien determinó los parámetros de la voz que ayudan a detectar el tipo de trastorno. Los tipos de trastornos se refieren a alteraciones en tono, timbre e intensidad, así como problemas con la formación de sonidos, dificultades con el tono, volumen o calidad de la voz [18].

Este producto obtiene y analiza parámetros de la voz, pero no está enfocado a la mejora de pronunciación. Su principal objetivo es comparar parámetros correctos de la voz con incorrectos y así diagnosticar que tipo de trastorno se tiene [18].

3.3.2 Software para la Enseñanza del Español Hablado en México

En la Universidad de la Américas se está aplicando el procesamiento y reconocimiento de voz en herramientas tecnológicas para distintas aplicaciones del habla. Se desarrolló un producto de software orientado a la enseñanza del español hablado en México. Cuenta con dos módulos. Uno es un diccionario Inglés-Español accesado por medio de la voz. El segundo módulo utiliza un nuevo método para la verificación de la pronunciación correcta de palabras o frases [26].

Este software utiliza técnicas de reconocimiento de voz y las herramientas del CSLU Toolkit para reconocer la secuencia de sonidos emitidos por el usuario y marcar las partes mal pronunciadas. Su función es tratar de encontrar fonemas mal pronunciados dentro de una frase. Contempla tanto los fonemas del español como los fonemas de la lengua materna de quien trata de aprender español, para captar la mayor cantidad de errores de sustitución. Este software provee una retroalimentación clara y explícita al usuario, pues le muestra el fonema exacto donde se equivocó.

Este producto está orientado a detectar errores causados por la costumbre de pronunciación del idioma materno. Presenta parte de su retroalimentación en forma auditiva [26].

3.3.3 ICATIANI

Otro proyecto realizado en la Universidad de las Américas es un prototipo llamado ICATIANI [56]. Este proyecto fue desarrollado con el CSLU Toolkit, y su principal objetivo es mejorar la pronunciación de vocales en español.

La principal ventaja de este sistema es que hace uso del Agente Conversacional Baldi, quien muestra la correcta pronunciación de los fonemas. De esta manera se puede visualizar la posición de los labios y mandíbula desde diferentes ángulos, aunque la parte interna de la boca es un poco difícil de distinguir. La animación de este agente se complementa con gráficos de la articulación de cada vocal presentando conceptos del triángulo vocálico.

Los usuarios intentan pronunciar las vocales solicitadas y, con ayuda del reconocedor, se puede saber si se pronuncio bien o mal el fonema. El agente Baldi presenta una cara feliz cuando la vocal es reconocida y una cara triste cuando no concuerda con el patrón del reconocedor. Este agente también muestra inicialmente como debe ser la correcta pronunciación (Figura 3.3).

Esta interfaz es muy llamativa y elaborada. El usuario puede saber si pronunció bien o mal una letra, pero a nuestro parecer le ayudaría mucho una retroalimentación de las posiciones articulatorias del tracto vocal.



Figura 3.3. Interfaz del Sistema ICATIANI [56].

CAPITULO 4

Trabajo Propuesto

El objetivo de este trabajo de investigación es desarrollar una interfaz de software que permita a personas con problemas auditivos, mejorar la pronunciación de vocales del idioma español, mejorando así su comunicación verbal. Enseñar a hablar a las personas sordas se vuelve muy complicado, pues deben aprender aspectos articulatorios y acústicos para producir la voz sin contar con el beneficio de la retroalimentación auditiva [23].

Con la ayuda de este trabajo, las personas con problemas auditivos podrán aprender en forma visual la manera correcta de pronunciar las cinco vocales del español, y podrán tener una retroalimentación de la posición articulatoria de los fonemas para verificar el aprendizaje. Para lograr una mejor pronunciación se busca proporcionar un producto de software con una interfaz adecuada y sencilla para ayudar a mejorar la pronunciación de las vocales del lenguaje español a personas con problemas auditivos, utilizando procesamiento de voz, retroalimentación visual y reconocimiento de voz.

Para sustituir la retroalimentación auditiva con la visual se muestra visualmente la correcta posición del tracto vocal para cada una de las cinco vocales. Posteriormente la persona que utiliza esta herramienta intenta reproducir la posición mostrada de una vocal elegida y pronuncia el fonema. Esta pronunciación es tomada por la computadora por medio de un micrófono, se convierte en una señal digital y es procesada para extraer valores que posteriormente sirven para saber la posición del tracto vocal que la persona realizó al pronunciar el fonema.

Al graficar la posición que la persona realizó (articulación), y contrastarla con un patrón predeterminado y opcionalmente personalizado, las persona puede ver gráficamente que tan correcta realizó la pronunciación, en términos de articulación. Esto le permitirá a la persona que no escucha obtener una retroalimentación visual de su propia pronunciación comparándola con la pronunciación correcta. Así logrará ir mejorando la articulación de cada vocal, fijando la posición adecuada de labios, dientes, lengua y mandíbula. Opcionalmente se tiene la posibilidad de utilizar un reconocedor de alto rendimiento que verifica si la pronunciación realizada es reconocida como la vocal que está siendo practicada.

Si a las personas sordas se les provee suficiente información visual articulatoria y acústica de la voz, ellos pueden tener la misma capacidad para aprender a hablar que una persona que si escucha [23]. El gran problema que enfrentan las personas que no oyen es que a pesar de que pueden sustituir la retroalimentación auditiva por la visual, el tiempo que tienen acceso a esta retroalimentación es muchísimo menor que el tiempo que una persona que oye tiene acceso a la retroalimentación auditiva.

Las personas con problemas auditivos pueden obtener muchos beneficios de una retroalimentación visual constante, recibiendo información articulatoria y observando sus propias aproximaciones a las articulaciones correctas [23].

La interfaz propuesta está enfocada para personas con problemas auditivos (sin problemas de aprendizaje) que tenga la madurez suficiente para estar en un programa escolar de primaria o más avanzado. Es decir personas de 6 años en adelante. Los alumnos que participaron en el entrenamiento con el sistema tienen entre 11 y 15 años. Los datos completos de cada uno de ellos se encuentran en el Anexo F.

4.1 Descripción del Sistema

La interfaz de entrenamiento para mejorar la pronunciación de vocales del idioma español fue desarrollada en el lenguaje de programación JAVA. Las tres razones más importantes para elegir este lenguaje de programación son: portabilidad, seguridad y fácil adaptación a Internet [36]. Con la portabilidad de este lenguaje de programación podemos crear código que puede ser usado en varios sistemas operativos y múltiples tipos de computadoras abarcando una mayor cantidad de usuarios. Java proporciona seguridad al limitar los programas al ambiente de ejecución JAVA y no permitir acceso a otras partes de la computadora. Aunque actualmente el código está hecho para funcionar como una aplicación instalada en cada computadora, pensando en el futuro crecimiento y esparcimiento de este proyecto, se tiene la opción de convertir este código en un *applet* que funcione en Internet. Al adaptar el sistema a una página de Internet se da la posibilidad de acceso a mayor cantidad de personas con problemas auditivos.

La interfaz desarrollada se encuentra dentro del proyecto REVISAH (REtroalimentación Visual Interactiva del Sistema Artulatorio del Habla) [39]. El nombre fue elegido porque el sistema es una herramienta de retroalimentación visual de la posición articulatoria del tracto vocal obtenida del procesamiento digital de la señal de voz suponiendo que el tracto vocal se puede modelar como tubos rígidos discretos.

Todo el desarrollo del sistema fue realizado con estrecha vinculación con maestros del CAM Instituto Sertoma, turno matutino.

4.2 Diseño del Sistema

El principal y primer factor que se revisó en el diseño del sistema fue la interfaz. Ésta se diseñó de forma sencilla y fácil de usar, pues los usuarios potenciales del sistema son personas que en la mayoría de los casos tienen poco contacto con la tecnología y manejo de computadoras.

Se decidió mostrar el tracto vocal en un gráfico de un corte de perfil de una persona, siendo este personaje muy sencillo y solo para ejemplificar la vista y dirección de las partes de la boca involucradas.

El estilo de la interfaz es orientado a ventanas. Cuenta con 3 ventanas principales. En la primera ventana se muestra la presentación del sistema, así como las cinco vocales que se pueden practicar. En esta ventana se elige la vocal a practicar, y si se desea algún patrón de articulación. Si no se elige ningún patrón el sistema toma el predeterminado. También en esta ventana se elige si se desea incluir el reconocimiento de fonemas. Hay que tomar en cuenta que para utilizar esta opción se necesita tener una conexión a Internet.

La segunda ventana es la más importante, pues es donde se presenta la retroalimentación visual de la pronunciación realizada por el usuario (Figura 4.1). Esta ventana está dividida en dos partes. Del lado derecho de la ventana se puede observar un gráfico muy general del punto de articulación de la vocal elegida, así como el símbolo gráfico del fonema y una vista frontal de los labios de la pronunciación del fonema. La parte izquierda de la ventana inicialmente aparece en blanco, pero después de hacer la grabación del fonema elegido se puede ver el gráfico que muestra la articulación pronunciada. También se puede observar en la parte superior izquierda un conteo del número de repeticiones que se van realizando. Si se habilito la opción de reconocimiento aparecerá una carita en la parte media superior de la ventana cuando se grafica. Si la carita está feliz y en verde, quiere decir que la vocal pronunciada concuerda con la que se está trabajando. Si la carita está triste y en rojo, quiere decir que la vocal no fue reconocida como la vocal que se está tratando de igualar.

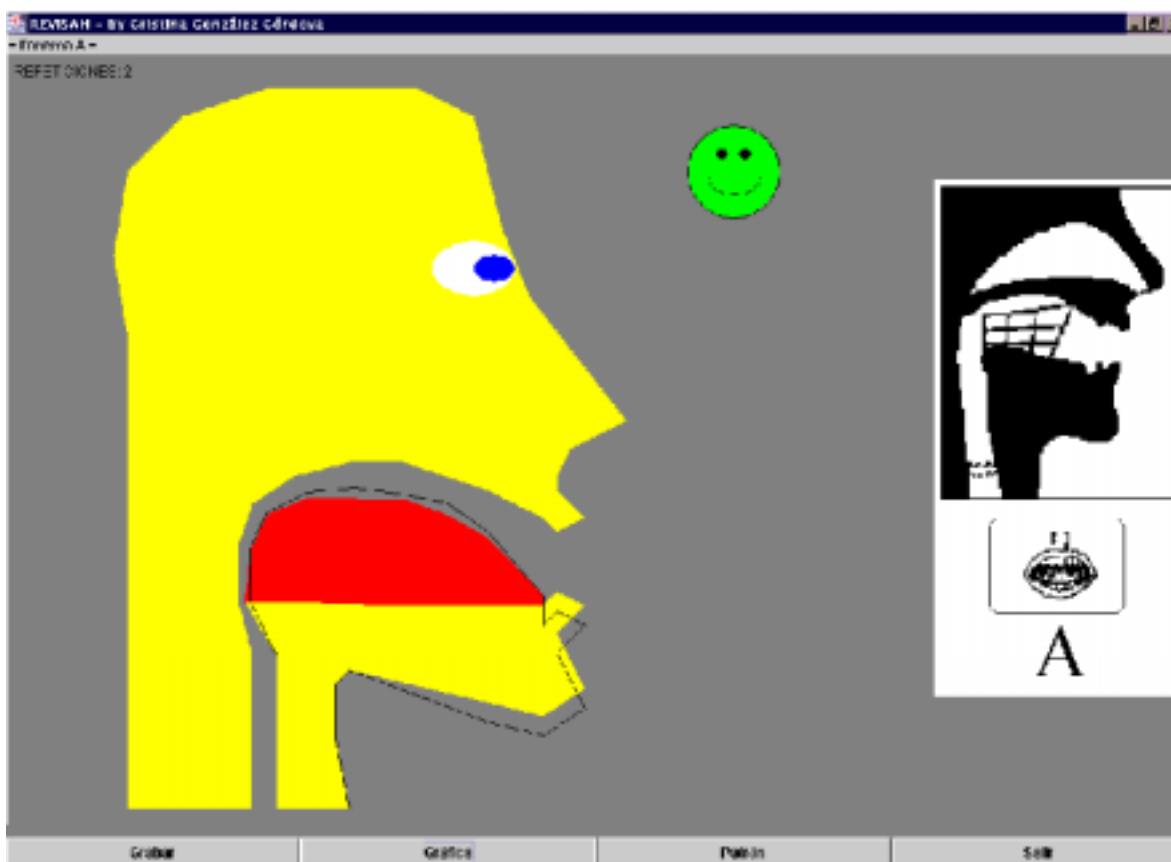


Figura 4.1. Pantalla Principal del Sistema.

Esta ventana contiene cuatro botones. El botón “Grabar” permite abrir la ventana para hacer la grabación del fonema a practicar. El botón “Gráfica” muestra la articulación pronunciada por medio de un gráfico de corte de perfil, también muestra la carita de reconocimiento si esta opción fue habilitada. El botón “Patrón” permite grabar en un patrón la articulación producida por el usuario. El botón “Salir” es el que cierra esta ventana, y si está abierta, también la ventana de grabación de fonemas.

La tercera ventana es llamada con el botón “Grabar” de la ventana anterior. En esta ventana se puede hacer la grabación del fonema que el usuario está tratando de pronunciar correctamente. Contiene tres botones. Con el primer botón se inicia y termina la grabación. El segundo botón permite escuchar el sonido grabado (de utilidad para el maestro), y el tercer botón permite ajustar el umbral de ruido existente en el ambiente. Inicialmente aparece un círculo de color rojo, que indica que no está grabando. Cuando se inicia la grabación se puede observar un círculo de color verde, que le permite al usuario darse cuenta que ya puede iniciar la pronunciación. Al terminar la grabación se muestra un gráfico de la voz, y se indica con un cuadro rojo la parte de voz que se tomará, dejando fuera los segmentos de silencio antes y después de la pronunciación del fonema.

Con respecto al funcionamiento interno del sistema y al flujo de datos, tenemos un sistema muy sencillo con dos niveles. Para representar el flujo de los datos mostramos un Diagrama de Flujo de Datos (DFD). El nivel 0 o diagrama de contexto nos da una vista general de lo que sucede dentro del sistema y se muestra en la figura 4.2.

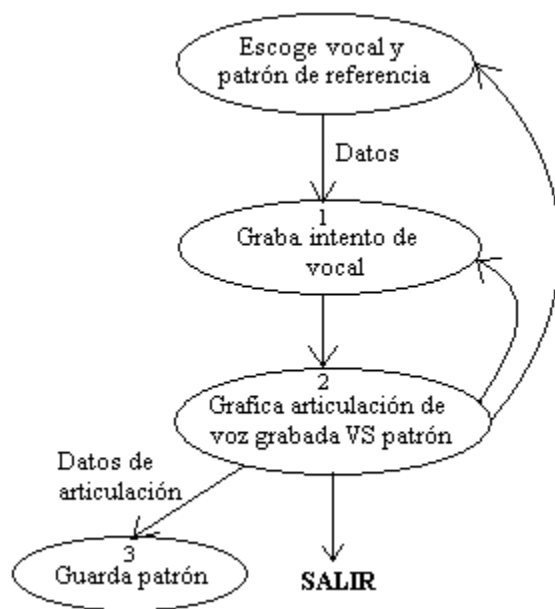


Figura 4.2. Diagrama de Contexto.

En el nivel 1, podemos desglosar el funcionamiento de cada proceso que es llevado a cabo en el diagrama de contexto. El primer diagrama del nivel 1 muestra el proceso por

medio del cual se graba el fonema que el usuario está tratando de mejorar. En la figura 4.3 podemos observar este diagrama.

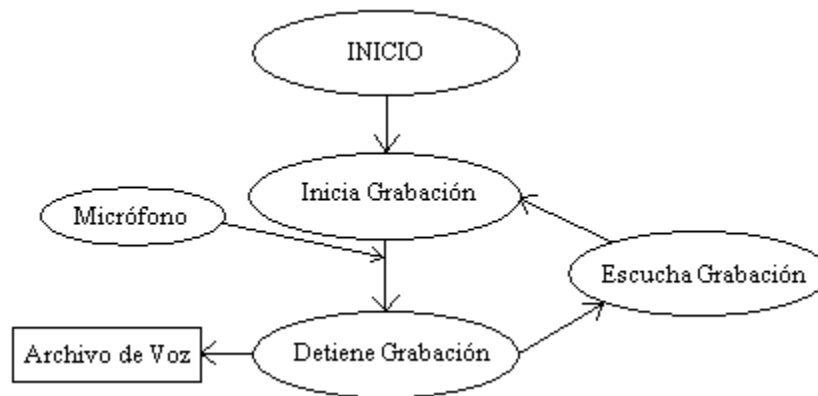


Figura 4.3. Diagrama de Flujo de Datos 1, Nivel 1.

El segundo diagrama del nivel 1 nos muestra como se realiza la graficación de la articulación de la voz pronunciada contrastando con el patrón elegido para ese fonema (Figura 4.4). En este proceso podemos destacar la función que calcula las áreas, pues es la parte medular del sistema. Aquí es donde se realiza el procesamiento de voz, se extraen los parámetros importantes y se calculan las 12 áreas necesarias para representar el tracto vocal, que comienza en la laringe y termina en los labios.

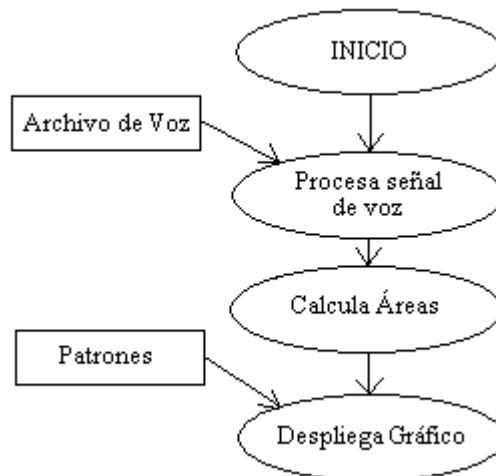


Figura 4.4. Diagrama de Flujo de Datos 2, Nivel 1.

Finalmente, tenemos el tercer Diagrama de Flujo de Datos del nivel 1, que muestra el proceso para guardar un patrón. Podemos observar este diagrama en la figura 4.5.

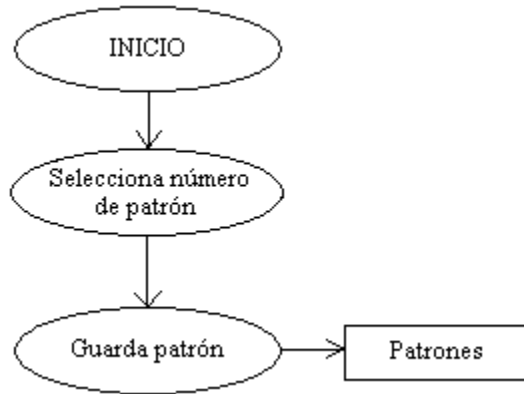


Figura 4.5. Diagrama de Flujo de Datos 3, Nivel 1.

Estos diagramas muestran de forma general como los datos son transportados de una función a otra. Es importante destacar que se utilizaron algunos archivos de voz y de texto que guardan temporalmente la señal de voz o los parámetros calculados.

4.3 Patrones

Aunque se ha comprobado que el punto de articulación no determina la pronunciación de una vocal y que la mayoría de la gente no usa los patrones del triangulo vocálico para pronunciar las vocales [30], se sabe que cada persona fija en la mente ciertos patrones de articulación que auditivamente sabe que suenan correctamente.

El objetivo de presentar patrones predeterminados y patrones ajustables es dar varias opciones para lograr una mejor pronunciación de las vocales. En primer lugar se ofrecen patrones de pronunciación basados en la teoría articulatoria tradicional, y si el usuario se siente cómodo con estos patrones y logra dominarlos puede realizar una pronunciación correcta de las vocales.

Por otra parte, sabemos que algunas veces las personas sordas pueden producir pronunciaciones correctas en cierto momento, pero que tienen dificultad en repetir las y fijarlas como articulaciones mecanizada [23]. De esta manera, los patrones adaptables pueden ser fijados por el maestro o instructor cuando considera que una pronunciación es correcta, y así las personas trabajan sobre su propio patrón tratando de fijarlo como automático en el cerebro.

El sistema tiene una capacidad de guardar hasta seis patrones individuales en cada máquina que tenga instalado el software. En un futuro se podría agregar mas patrones de acuerdo a las necesidades que se puedan presentar.

Cada patrón está integrado por doce valores que representan el área del tracto vocal desde la laringe hasta los labios y un valor que representa la abertura de la mandíbula

inferior. Las doce áreas que representan el tracto vocal son aplicadas desde la mandíbula superior y se dirigen hacia ciertos centros calculados previamente. Esta distribución de las áreas se pensó tomando como ejemplo el modelo estático de Maeda [24] que usa un sistema de coordenadas semipolar. La manera de calcular la gráfica se muestra en la figura 4.6.

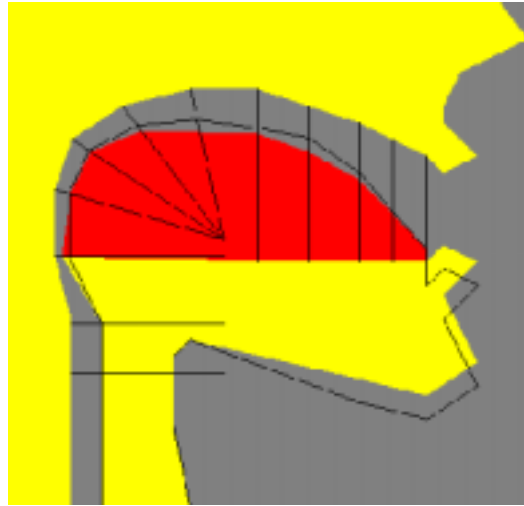


Figura 4.6. Graficación de las Doce Áreas.

Después de calcular las doce áreas, y antes de graficar, se aplica un proceso de suavizado de áreas para evitar picos o configuraciones poco naturales que pudieran resultar de los cálculos realizados.

CAPITULO 5

Pruebas y Evaluación

Para demostrar que la interfaz de software elaborada es una herramienta que ayuda a las personas con problemas auditivos a mejorar la pronunciación de vocales, se desarrollaron dos tipos de experimentos.

Las muestras utilizadas para verificar los experimentos fueron tomadas en el Centro de Atención Múltiple Instituto Sertoma, con la aprobación y el apoyo de la dirección y todo el personal que labora en esa institución.

Se eligieron cinco alumnos de la escuela, sugeridos por la directora, con un problema de sordera total. Los participantes fueron tres hombres y dos mujeres. Fueron sometidos a prácticas con el software. Se tomó una muestra de la pronunciación de las cinco vocales antes del periodo de entrenamientos y otra muestra al término, para cada uno de los alumnos.

Se tuvieron algunas limitaciones con respecto a la asistencia de los alumnos, pues había mucha inasistencia. Los alumnos participantes se encontraban cursando entre el cuarto y el sexto grado de primaria.

La habitación donde se realizaron los entrenamientos fue la sala de maestros, donde se contaba con total privacidad y silencio para no alterar las grabaciones. Se realizaron periodos de entrenamiento de 20 minutos por alumno, y se practicaban las cinco vocales del idioma español, todos los días. El número aproximado de repeticiones por vocal era de diez. El horario que se estuvo utilizando fue de 10:30 a.m. a 12:00 p.m.

La forma de trabajar fue por medio de sesiones diarias guiadas por el instructor, pues los alumnos que participaron en estos experimentos tienen poco contacto con las computadoras. A pesar de esto, comprendieron rápidamente el funcionamiento y pusieron mucho interés en los periodos de entrenamiento. Hubo días en los que alguna maestra estaba presente en los entrenamientos y podía darse cuenta de los avances de los muchachos.

El periodo de entrenamiento de los muchachos fue de ocho semanas. Este periodo de tiempo fue elegido de acuerdo al límite de tiempo para realizar la tesis y a las sugerencias de las maestras del CAM Instituto Sertoma. Se inició el lunes 20 de enero del año 2003 y se finalizó el viernes 14 de marzo del mismo año. Durante estas ocho semanas de entrenamiento se perdieron cuatro días de entrenamiento por factores externos como días libres, suspensión de actividades por parte de la escuela y mal tiempo.

El viernes anterior al inicio del periodo de entrenamiento (17 de enero del 2003) fue utilizado para tomar las muestras de voz iniciales. El lunes posterior al periodo de entrenamiento (17 de marzo del 2003) fue utilizado para tomar las muestras de voz finales.

En el primer tipo de experimento se utilizó un método para evaluar la calidad de la voz llamado “*Mean Opinion Score*” (MOS) [9, 53], se evaluaron las pronunciaciones de antes y después de utilizar el software y se buscó verificar su mejoría. En el segundo tipo de experimentos también se utilizó la opinión pública, pero esta vez para identificar qué fonema se había pronunciado escuchando las grabaciones de antes y después del entrenamiento.

5.1 Obtención de Datos

Para recopilar los archivos de voz que se utilizarían para hacer la comprobación de la mejoría en pronunciación, se utilizó una herramienta desarrollada para grabar archivos de voz con los mismos métodos utilizados en el software.

Antes de iniciar las grabaciones, se llenó una forma por cada alumno participante con sus datos generales (Anexo F) y se le asignó una clave. Posteriormente se le pidió que grabara uno a uno los cinco fonemas de las vocales del idioma español. Para cada vocal se tuvieron tres oportunidades de grabación, tomando la tercera para realizar las evaluaciones de pronunciación. Esos datos fueron guardados, junto con los de los voluntarios, para su uso en las encuestas de opinión.

También se obtuvieron muestras de cinco voluntarios oyentes, para ser utilizados en las pruebas como datos complementarios. La razón por la cual se utilizaron muestras de pronunciaciones de gente oyente fue para darle validez a los experimentos de tipo I y comparar calificaciones comunes con las calificaciones de antes y después del entrenamiento. El procedimiento de grabación fue similar al de los alumnos y se obtuvo la pronunciación de las cinco vocales (Anexo F).

5.2 Experimentos Tipo I

Una de las formas para evaluar la mejora de la pronunciación de los muchachos participantes en los entrenamientos fue utilizando un método para evaluar la calidad de la voz. Este método es conocido como “*Mean Opinion Score*” (MOS) y consiste en escuchar una muestra de voz y asignar una calificación del 1 al 5 de acuerdo a la inteligibilidad y comprensión de acuerdo al criterio de la persona que está evaluando [9, 53].

Para realizar el primer tipo de evaluaciones se colocaron primeramente todos los archivos del fonema “a” en forma aleatoria, tanto los archivos de antes y después de cada alumno como los archivos grabados por los voluntarios. Después, todos los archivos del fonema “e” (también aleatoriamente) y así sucesivamente en orden alfabético. Estos datos fueron utilizados para realizar la evaluación de la calidad de pronunciación de los fonemas.

Se colocaron en forma aleatoria para que el encuestado no estuviera predisposto sabiendo de quien es la pronunciación.

De acuerdo a las especificaciones del método MOS [9, 20, 53], se pidió la participación voluntaria de 32 personas de diferentes edades para contestar los cuestionarios que miden el grado de calidad de las pronunciaciones.

Se preparó un cuestionario de tres hojas de acuerdo a los lineamientos de encuestas de investigación [57, 58, 61]. Este cuestionario describe brevemente el objetivo del estudio y las instrucciones para contestarlo (Anexo D). Los valores numéricos que asignaban a la pronunciación correspondían a [20]:

- 5 = Excelente
- 4 = Buena
- 3 = Aceptable
- 2 = Mediocre
- 1 = Mala

Para poder contestar este cuestionario se preparó un programa de software para escuchar los archivos grabados de las pronunciaciones de las vocales. En este software se presenta la elección de la vocal y 15 opciones que se van eligiendo con el ratón. Para escuchar el archivo requerido en el cuestionario se presiona el botón “PLAY” (Figura 5.1).

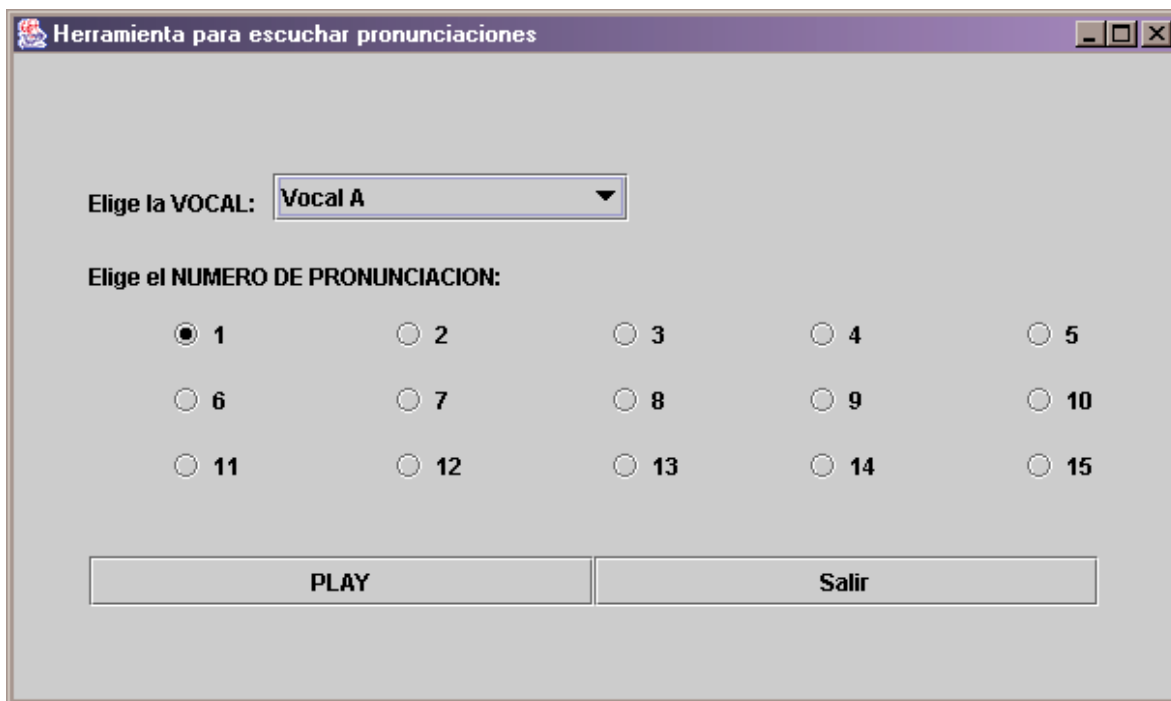


Figura 5.1. Software de Apoyo para Cuestionario Tipo I.

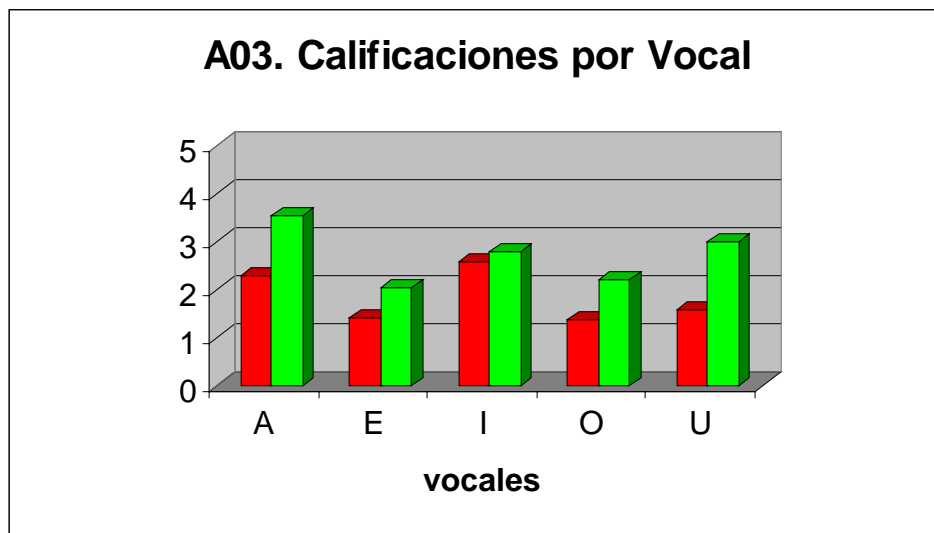
A cada encuestado se le proporcionó una copia en papel del cuestionario, una pluma para contestarlo, una computadora con el programa de apoyo y unos audífonos para una mejor percepción de la pronunciación.

Al tener las 32 calificaciones de cada vocal, de cada persona, se procedió a calcular el promedio asignado a cada pronunciación. Se calcularon las calificaciones tanto para las pronunciaciones de los alumnos sordos de antes y después del entrenamiento, como para los oyentes que participaron como voluntarios. Entre mayor fue la calificación recibida, mejor pronunciación fue percibida por los encuestados.

A continuación presentaremos los resultados de los cinco experimentos tipo I. Se presentan gráficas y explicaciones de cada experimento. Al final se muestra y explica un comparativo de todos los experimentos contrastando con las calificaciones de los voluntarios. En color rojo vemos los resultados de antes del entrenamiento, en color verde los resultados de después del entrenamiento y en color azul las calificaciones de los voluntarios.

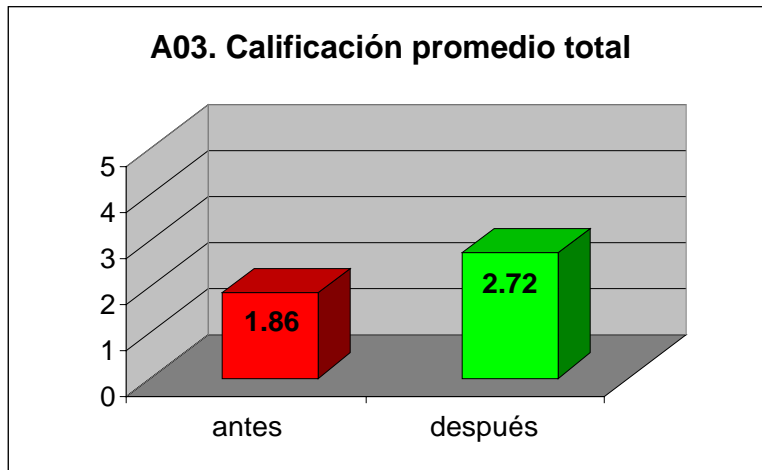
5.2.1 Experimento 1, Tipo I

En la gráfica 5.1 podemos observar las calificaciones recibida por el alumno A03 para las cinco vocales.



Gráfica 5.1. Calificaciones de Todas las Vocales para el Alumno A03.

Podemos observar como esta persona subió su calificación alrededor de 1.5 puntos en las vocales “A” y “U” quedando en calificaciones arriba de 3 puntos. En las vocales “E”, “I” y “O” subió menos y se quedo en calificaciones arriba de 2 puntos. Este alumno tuvo una mejoría en todas las vocales, aunque en diferente grado. El promedio de todas las vocales de antes y después del entrenamiento lo podemos observar en la gráfica 5.2.

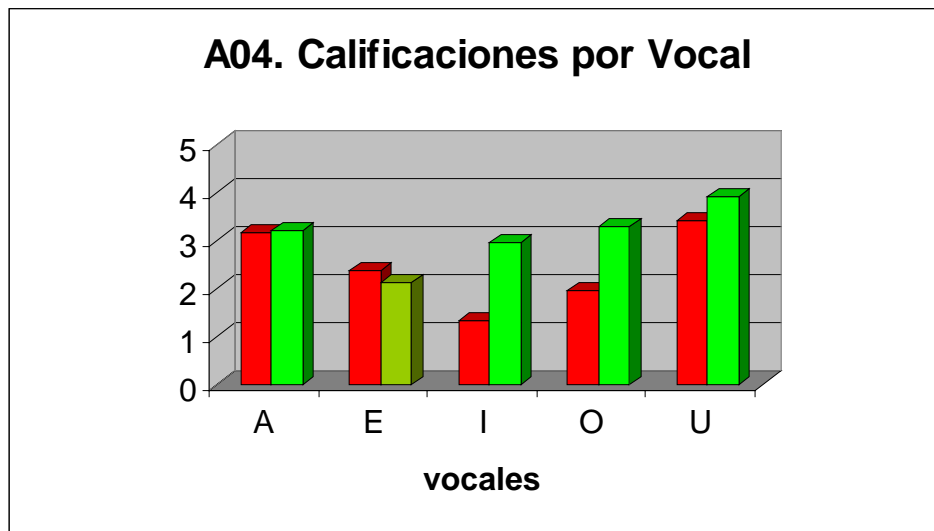


Gráfica 5.2. Calificación Promedio de Todas las Vocales para el Alumno A03.

Este alumno fue el que asistió más veces al entrenamiento (34 días), y en el que se veía mayor entusiasmo y ganas de salir adelante. En los periodos de entrenamiento estaba muy concentrado y atento a la retroalimentación que recibía. En total subió 0.86 puntos en promedio de todas las vocales, siendo el segundo alumno que más subió su calificación.

5.2.2 Experimento 2, Tipo I

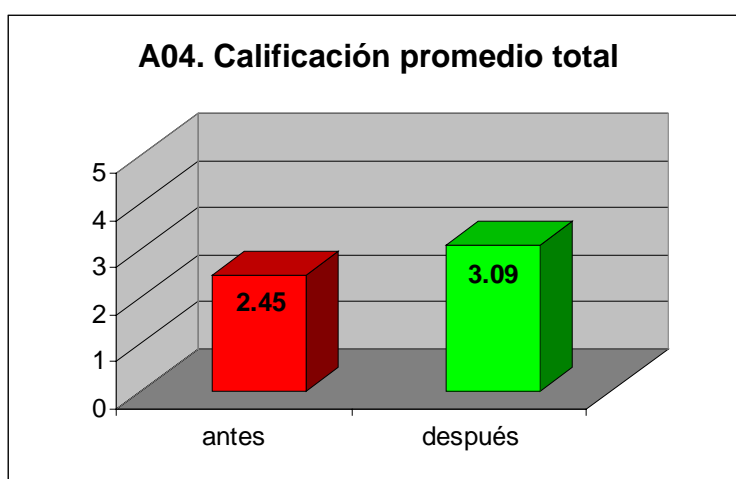
En la gráfica 5.3 podemos observar las calificaciones recibidas por el alumno A04 para las cinco vocales.



Gráfica 5.3. Calificaciones de Todas las Vocales para el Alumno A04.

Podemos observar como esta persona subió su calificación alrededor de 1.5 puntos en las vocales “I” y “O”. En la vocal “U” subió medio punto destacando que en la calificación final llegó casi a los 4 puntos (pronunciación buena). La “A” quedó prácticamente igual. En la “A” y “O” quedó arriba de los 3 puntos y en la “I” muy cerca. La que se le dificultó más fue la “E”, en la prácticamente se mantuvo igual, bajando un pequeño porcentaje en el promedio.

Este alumno subió mucho su calificación en vocales donde estaba muy bajo (menos de 2 puntos). En las vocales que inicialmente había salido alto, no subió mucho. La vocal que se le dificultó más fue la “E”, pues no la dominaba muy bien y fácilmente la confundía con la “A”. En el promedio de todas las vocales subió 0.64 puntos (Gráfica 5.4).

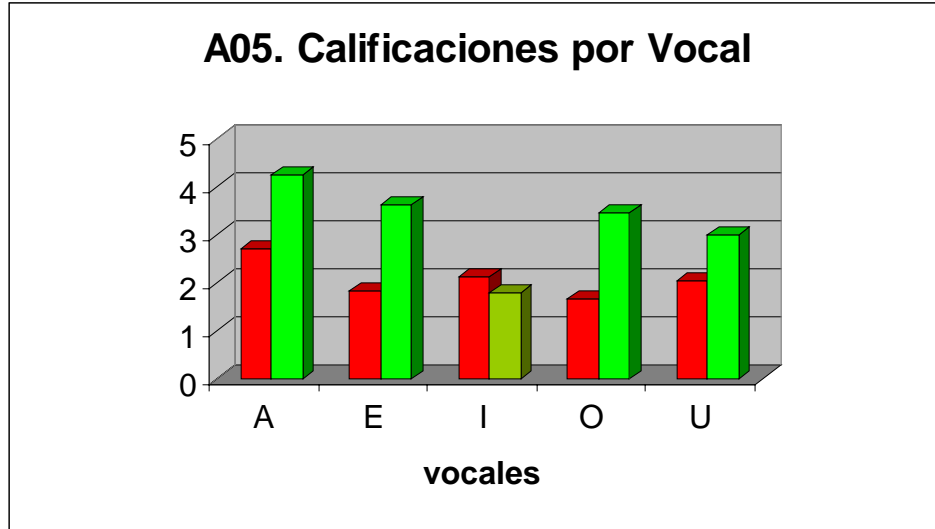


Gráfica 5.4. Calificación Promedio de Todas las Vocales para el Alumno A04.

Este alumno tuvo una asistencia baja (17 días), se distraía en los entrenamientos y no mostraba mucho interés. Quizá hubiera mejorado más si hubiera tenido más constancia y concentración.

5.2.3 Experimento 3, Tipo I

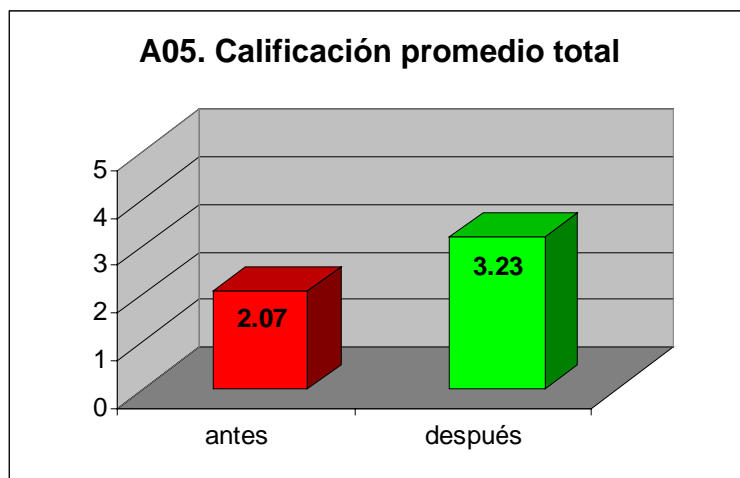
En la gráfica 5.5 podemos observar las calificaciones recibidas por el alumno A05 para las cinco vocales.



Gráfica 5.5. Calificaciones de Todas las Vocales para el Alumno A05.

Esta persona tuvo mejoras substanciales en casi todas las vocales (excepto la “I”). En la “U” subió casi 1 punto, en la “A” subió punto y medio, y en la “E” y “O” casi 2 puntos. Destaca en esta persona que sus puntuaciones finales que mejoraron alcanzaron valores mayores de 3 puntos y en la “A” llegó más arriba de los 4 puntos, siendo una muy buena calificación. La única vocal que prácticamente se quedó igual fue la “I”, presentando un ligero porcentaje menor en la calificación promedio.

Al tratar de pronunciar la “I” la confundía con la “E”, puesto que sus articulaciones son muy parecidas. Su mejora de todas las vocales en promedio fue la mejor de todas, subiendo 1.16 puntos y quedando arriba de los 3 puntos de calificación (Gráfica 5.6).

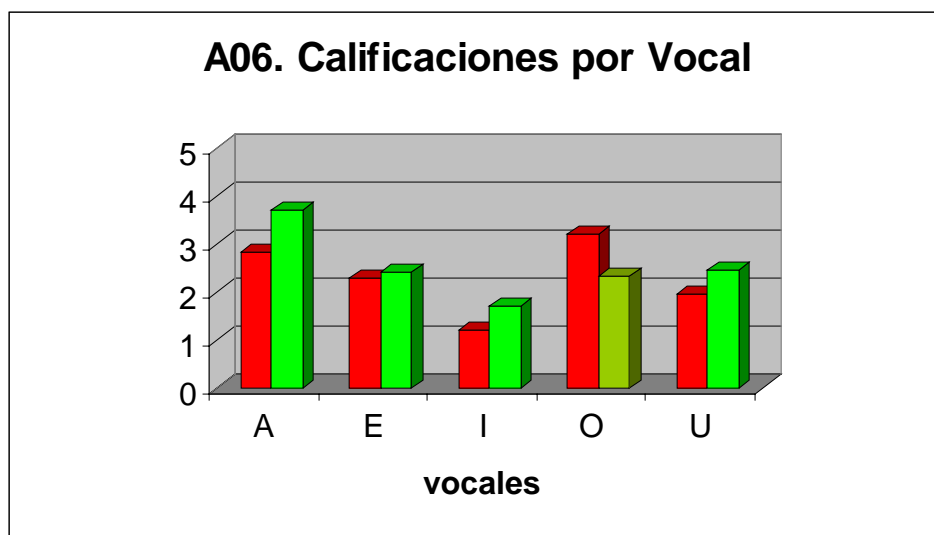


Gráfica 5.6. Calificación Promedio de Todas las Vocales para el Alumno A05.

Este alumno fue el segundo lugar en asistencia (21 días), aunque tuvo muchas menos asistencias que el alumno A03. Su actitud era de superación y todos los días trataba de igualar la articulación del patrón. Le llamaba mucho la atención las gráficas presentadas y trataba de ir todos los días. Le costaba mucho realizar la vocal “E”, la confundía con la “I” puesto que estos dos patrones se parecen mucho.

5.2.4 Experimento 4, Tipo I

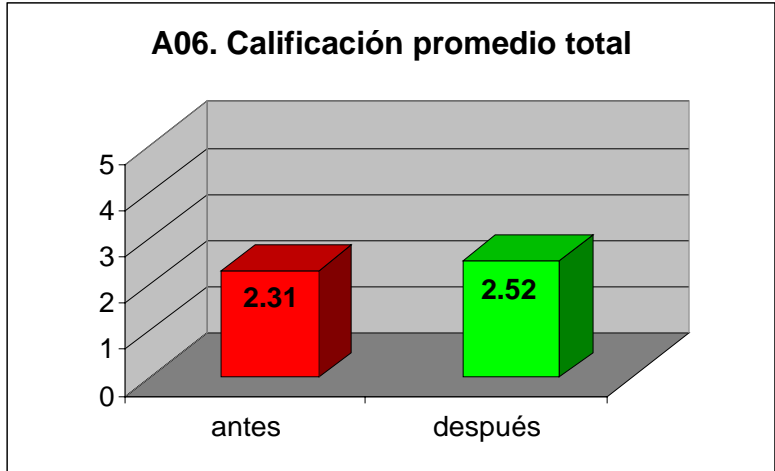
En la gráfica 5.7 podemos observar las calificaciones recibida por el alumno A06 para las cinco vocales.



Gráfica 5.7. Calificaciones de Todas las Vocales para el Alumno A06.

Esta persona fue la que obtuvo menos avances, teniendo una asistencia regular (20 días) y poca constancia. En la vocal que más subió, la “A”, no alcanzó a subir ni 1 punto, aunque quedó arriba de 3 puntos y medio. En la “I” y “U” subió casi medio punto, y en la “E” prácticamente se quedó igual, teniendo como calificación final en estas tres letras alrededor de 2 puntos. Tuvo una sustancial desmejora en la letra “O” de 0.85 puntos, curiosamente la vocal en la que había obtenido el mejor puntaje inicialmente. En la puntuación promedio de todas las vocales tuvo una mejoría de apenas 0.21 puntos, siendo quien mejoró menos en promedio de todas las vocales y quedando en los 2.5 puntos de calificación final (Gráfica 5.8).

A parte de su inconstancia en los entrenamientos, esta persona tenía problemas con su tono de voz y muchas veces realizaba una articulación correcta pero no producía sonido en la glotis. Sus problemas de pronunciación iban más allá del alcance de este proyecto. A pesar de todo logró mejorar en cuatro de cinco vocales, aunque muy poco.



Gráfica 5.8. Calificación Promedio de Todas las Vocales para el Alumno A06.

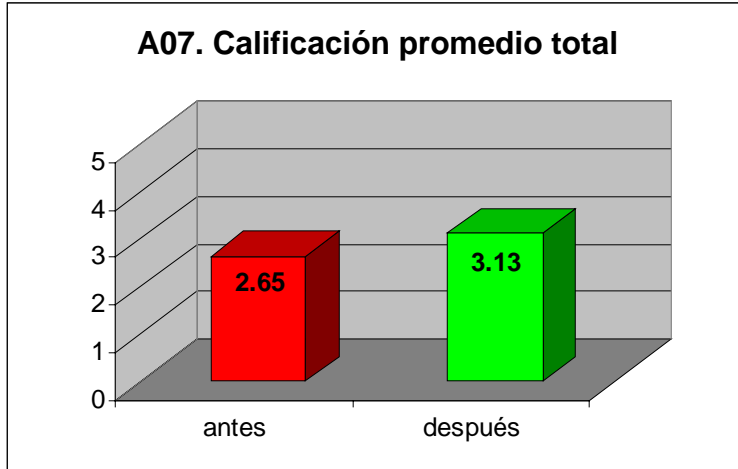
5.2.5 Experimento 5, Tipo I

En la gráfica 5.9 podemos observar las calificaciones recibida por el alumno A07 para las cinco vocales.



Gráfica 5.9. Calificaciones de Todas las Vocales para el Alumno A07.

Esta persona subió en todas las vocales, aunque un porcentaje muy bajo (excepto en la “O”). En la vocal “O”, que había obtenido la menor calificación inicialmente, subió considerablemente mas de 1 punto y medio. Las demás vocales subieron alrededor de 0.20 puntos. Resulta interesante notar que cuatro de cinco vocales tuvieron calificaciones finales de entre 3 y 3.5 puntos. En el promedio de todas las vocales obtuvo una mejora de casi medio punto y quedó arriba de los 3 puntos en la calificación final (Gráfica 5.10).

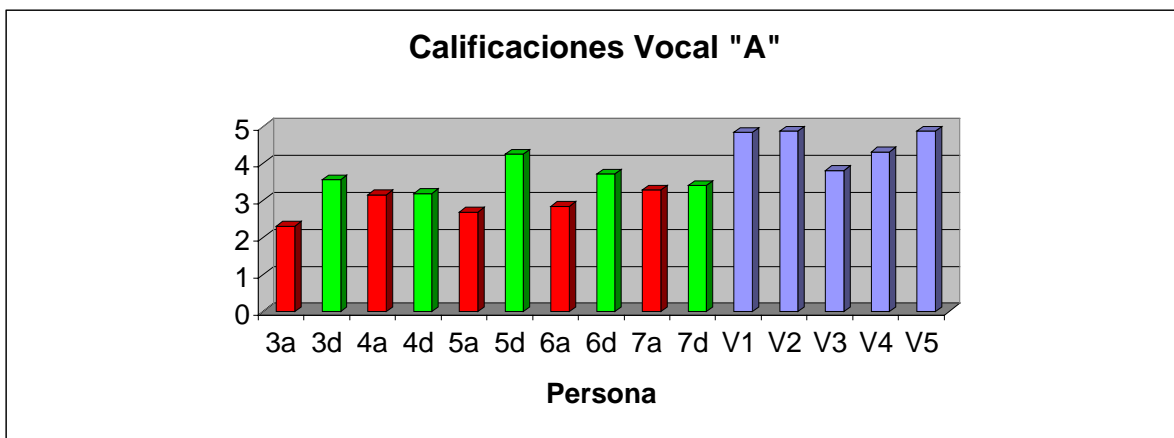


Gráfica 5.10. Calificación Promedio de Todas las Vocales para el Alumno A07.

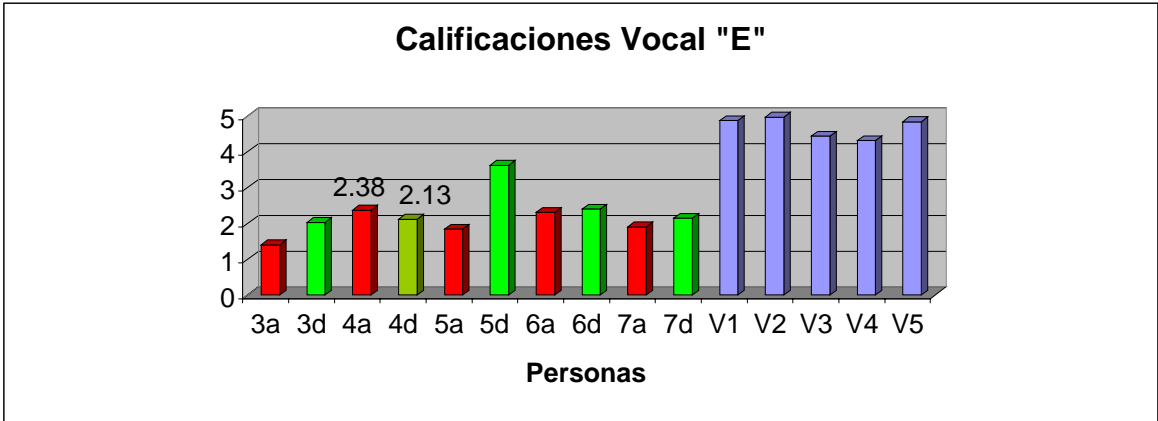
Es importante mencionar que esta persona fue un caso especial en el período de entrenamiento, puesto que solo asistía 1 vez por semana al instituto. A pesar de que tuvo solo 9 asistencias, fue una de las alumnas más constantes, y nunca faltó a una clase. Con ella se realizaban periodos de entrenamiento un poco más prolongados (aproximadamente 30 minutos) y las sesiones eran una vez por semana. Es una persona muy inteligente que rápidamente comprendía la retroalimentación y lograba igualar los patrones. Sus mejoras quizá no fueron tan altas como las de otros, pero se puede notar una tendencia positiva, que con más días de entrenamiento podría ser mayor.

5.2.6 Resultados Generales de Experimentos Tipo I

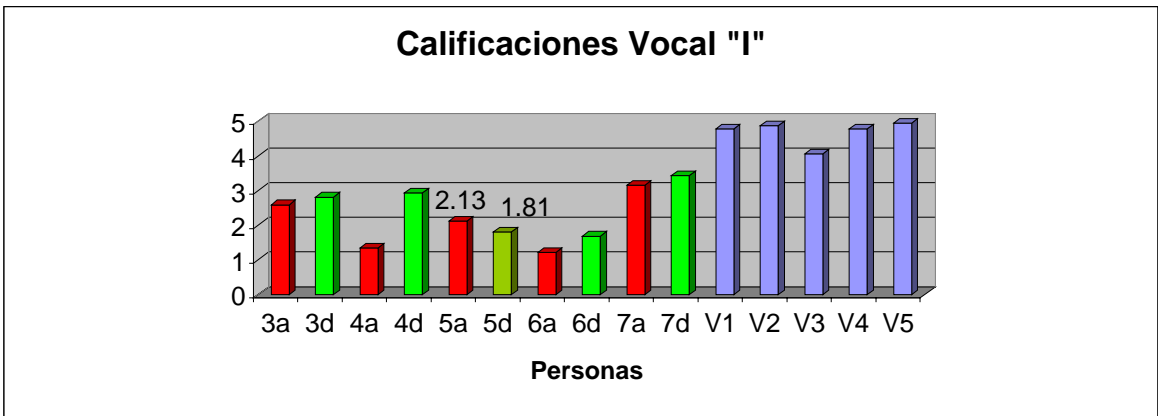
En las gráficas 5.11, 5.12, 5.13, 5.14 y 5.15 podemos ver las gráficas comparativas de cada una de las vocales en general. En color rojo podemos observar las calificaciones iniciales de cada alumno. En color verde tenemos las calificaciones finales de los alumnos y en azul observamos las calificaciones de los voluntarios que no padecen ningún trastorno de la audición.



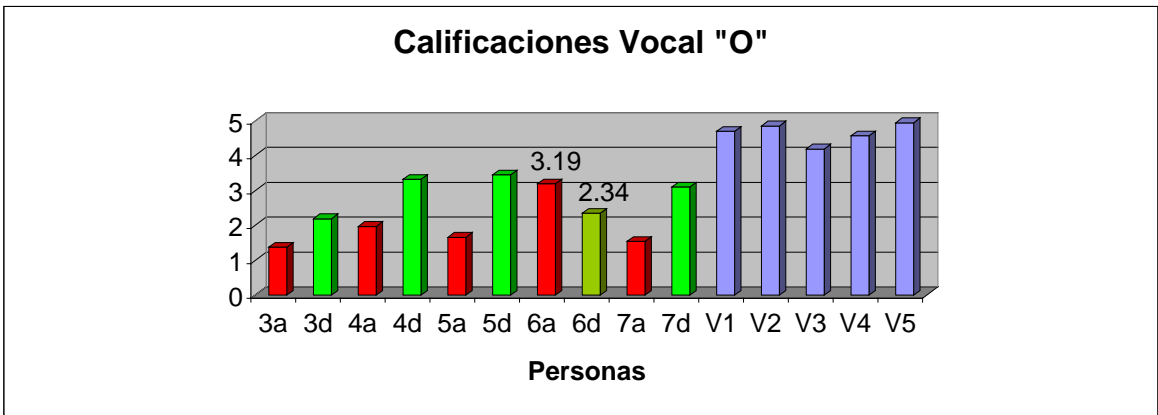
Gráfica 5.11. Comparativo General de Calificaciones para la Vocal "A".



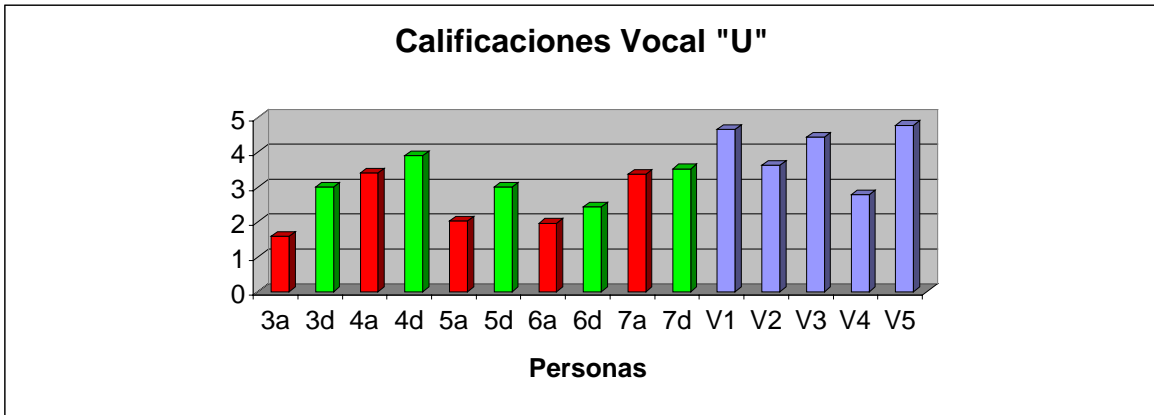
Gráfica 5.12. Comparativo General de Calificaciones para la Vocal "E".



Gráfica 5.13. Comparativo General de Calificaciones para la Vocal "I".



Gráfica 5.14. Comparativo General de Calificaciones para la Vocal "O".

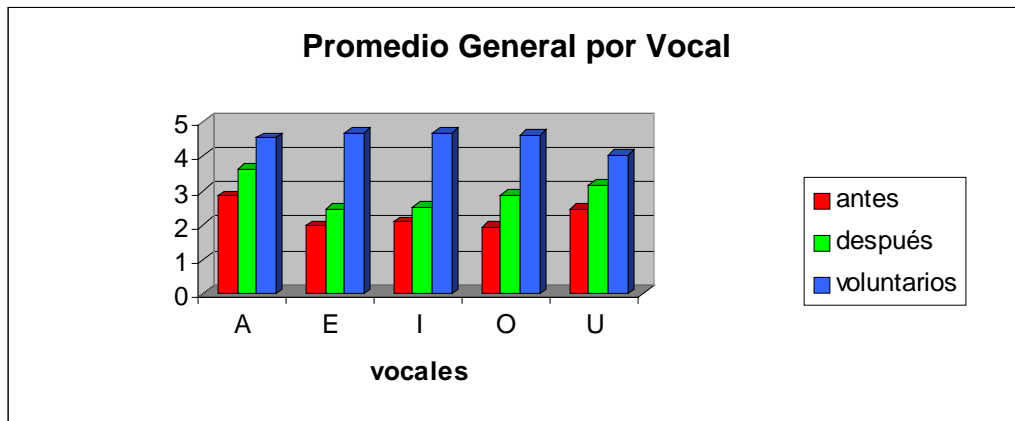


Gráfica 5.15. Comparativo General de Calificaciones para la Vocal "U".

Analizando estas gráficas nos damos cuenta de que las vocales en las que más se mejoró y donde se obtuvieron calificaciones finales más altas son la "A" y la "U". En estas vocales todos mejoraron. De hecho hay alumnos que alcanzan calificaciones finales iguales o superiores a las de algunos voluntarios.

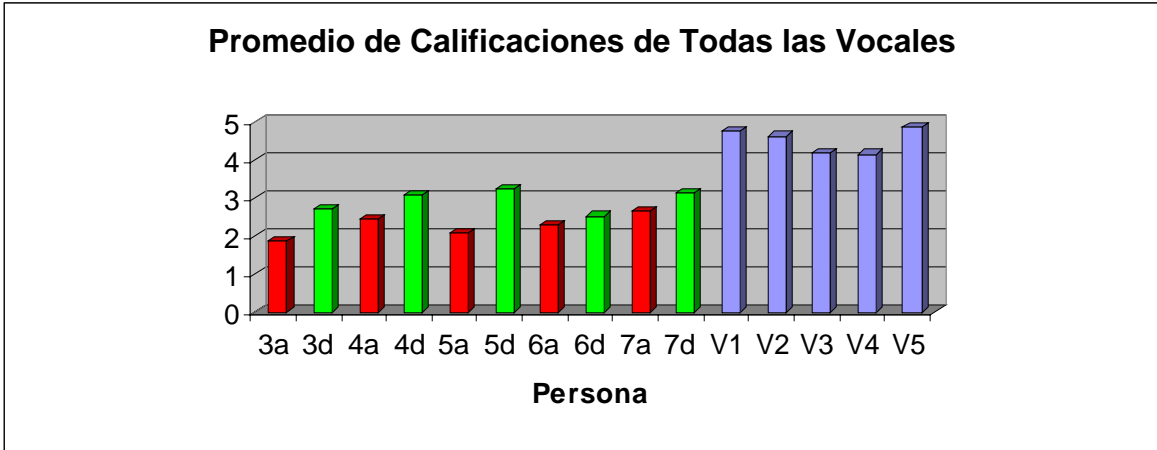
Las vocales que fueron más difíciles de dominar fueron la "E", "I" y "O". Ninguna calificación final estuvo al nivel de las calificaciones de los voluntarios y en cada una hubo alguno de los alumnos que bajo su calificación.

Comparando las calificaciones promedio de todos los alumnos por vocal nos damos cuenta que en todas hubo una mejora, aunque en diferente grado (Gráfica 5.16).



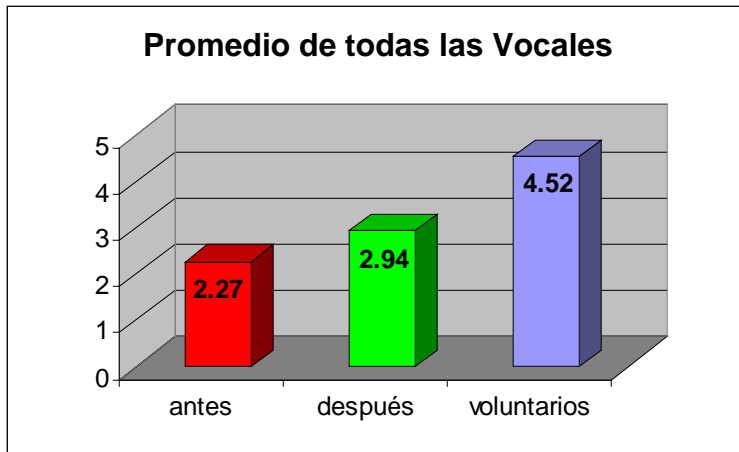
Gráfica 5.16. Calificaciones Promedio por cada Vocal.

En la gráfica 5.17 se presenta un comparativo de los promedios de todas las vocales por alumno de antes y después del entrenamiento y el promedio de todas las vocales de los voluntarios. En la gráfica 5.18 se muestra el promedio de antes y después de todos los alumnos y de todas las vocales, comparándolo con el promedio de todas las vocales de todos los voluntarios.



Gráfica 5.17. Calificaciones Promedio de todas las Vocales de todos los Participantes.

Como vemos en la gráfica 5.17 en promedio no se alcanzan los niveles de los voluntarios, pero si se alcanzan niveles de casi 3 puntos en las calificaciones finales de casi todos los alumnos.



Gráfica 5.18. Calificaciones Promedio de todas las Vocales.

Finalmente en la gráfica 5.18 observamos como hubo una mejoría en general, llegando muy cerca de los 3 puntos en la calificación final y quedando 1 punto y medio debajo del promedio general de los voluntarios.

5.3 Experimentos Tipo II

La segunda forma de evaluar la mejora de la pronunciación fue por medio de otro cuestionario de opinión. En este cuestionario no se le pedía una calificación de calidad de una vocal, sino se pedía que se identificara cual vocal creía el encuestado que se había pronunciado al escuchar una serie de pronunciaciones.

Para realizar el segundo tipo de evaluaciones se colocaron únicamente los archivos grabados por los alumnos antes y después del periodo de entrenamiento. Estos archivos fueron colocados aleatoriamente y sin hacer separaciones por vocal. Aquí se mezclaron vocales y personas, pues se buscaba que el encuestado verificara que vocal de había pronunciado. Los archivos de voz incluidos en este experimento fueron las pronunciaciones de las cinco vocales de los cinco alumnos para antes y después del entrenamiento. En total fueron 50 pronunciaciones que el encuestado debía clasificar dentro de alguna de las cinco vocales (a,e,i,o,u).

El cuestionario fue de 2 páginas (Anexo D). Para contestar este cuestionario también se diseñó un software especial para escuchar las pronunciaciones (Figura 5.2). A cada encuestado se le proporcionó una copia en papel del cuestionario, una pluma para contestarlo, una computadora con el programa de apoyo y unos audífonos para una mejor percepción de la pronunciación.

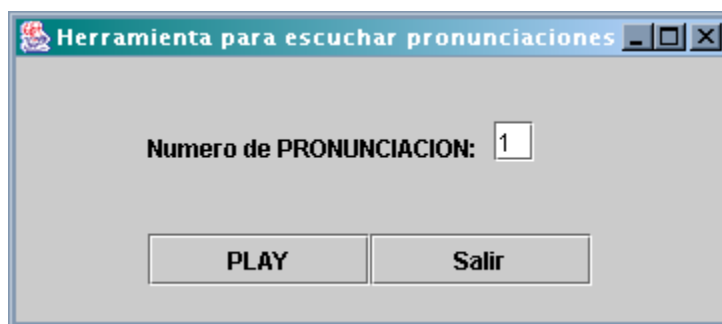


Figura 5.2. Software de Apoyo para Cuestionario Tipo II.

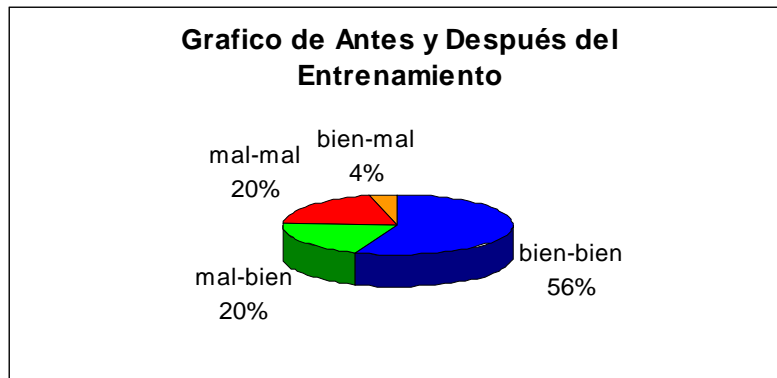
El encuestado iba escuchando cada unas de las 50 pronunciaciones presionando el botón "PLAY". Automáticamente se iba incrementando el número de la pronunciación y si se necesitaba escucharla de nuevo se podía indicar manualmente qué número de pronunciación deseaba escuchar. Por ser una prueba más sencilla se aplicaron 20 encuestas. Este tipo de experimento fue más sencillo de evaluar porque solo se pedía a las personas que clasificaran el sonido escuchado. Esta clasificación entraba dentro de una de las 5 vocales del idioma español.

La mayoría de las pronunciaciones que se realizaron antes del periodo de entrenamiento ya se escuchaban como la vocal indicada, por lo que no se incluirá en esta sección los resultados de todas las pronunciaciones de antes y después del entrenamiento. Tampoco se analizarán los casos específicos en los cuales la pronunciación inicial no era correcta y la final si era correcta. La tabla completa de los resultados se incluyen en el Anexo E.

De 25 casos de pronunciaciones antes y después (Ver Gráfica 5.19):

- 14 casos fueron correctos desde el principio y así se mantuvieron después del periodo de entrenamiento. Esto equivale a un 56%.

- 5 casos fueron incorrectos antes del entrenamiento y fueron correctos después del entrenamiento. Esto equivale a un 20%.
- 5 casos fueron incorrectos antes del entrenamiento y también fueron incorrectos después del entrenamiento. Esto equivale a un 20%.
- 1 caso fue correcto antes del entrenamiento e incorrecto después del entrenamiento. Esto equivale a un 4%.

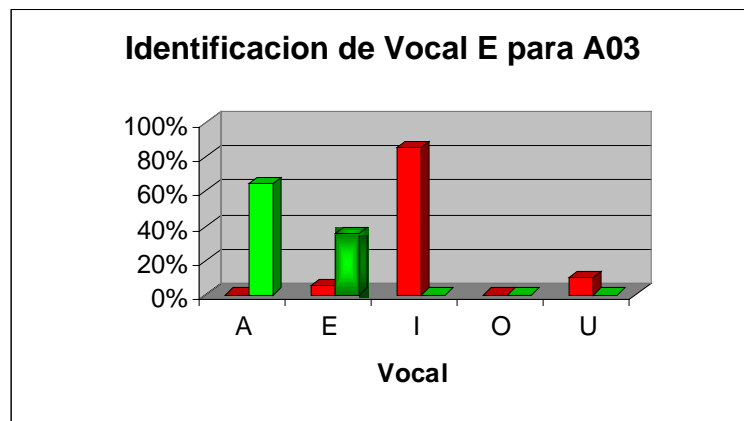


Gráfica 5.19. Gráfico de Antes y Después del Entrenamiento.

A pesar de que en 24% de los casos no se logró totalmente la vocal objetivo, en algunos casos se logró un porcentaje de avance. A continuación se analizará a más detalle los seis casos especiales en los cuales no se logró la vocal objetivo.

5.3.1 Alumno A03, Vocal E

En este caso, según las encuestas, se debía escuchar una “E”, y la mayoría informó que antes del entrenamiento su pronunciación se escuchaba como “I”, y después del entrenamiento se escuchaba como “A”. Analizando a detalle podemos ver que en la muestra de inicio 5% de los encuestados opinó que efectivamente se escuchaba como “E” y un 35% opinó que en la muestra final se escuchaba la “E”. Esto quiere decir que se tuvo un avance del 30% en similitud con la “E”, a pesar de que se está confundiendo mucho con la “A”. En la gráfica 5.20 se muestra el gráfico de este caso.

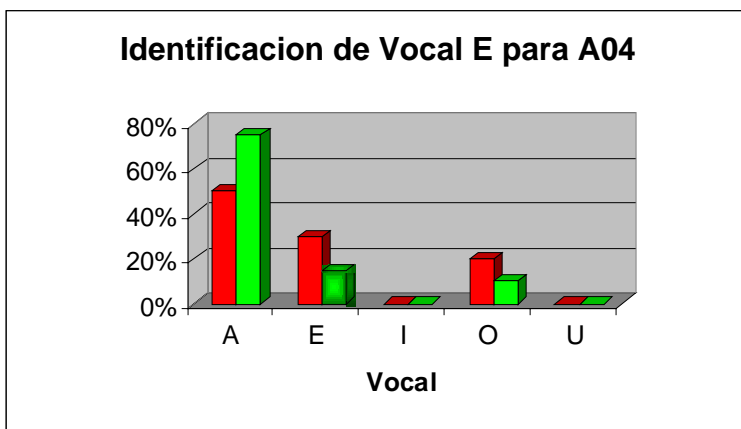


Gráfica 5.20. Identificación de la Vocal “E” para A03.

5.3.2 Alumno A04, Vocal E

En este caso se debía pronunciar una “E”, y lo que se está percibiendo es una “A” en ambos casos (antes y después del entrenamiento). En la gráfica 5.21 se puede apreciar como hay una disminución en el porcentaje de personas que escucharon una “E” al inicio (30%) y las que escucharon una “E” al final (15%).

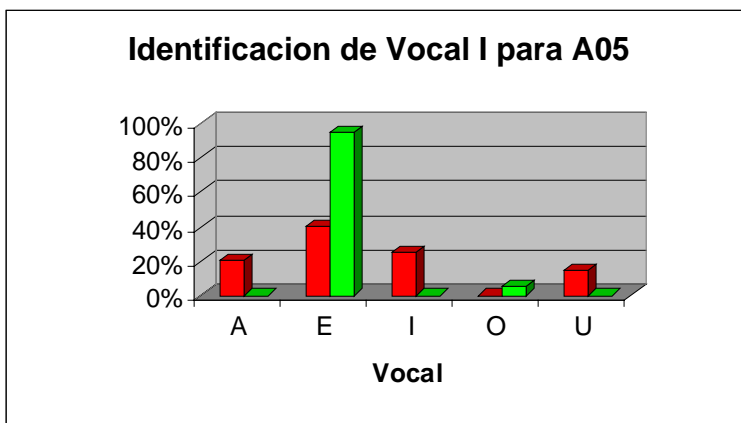
Nos damos cuenta que esta persona, en esta vocal, disminuyó su porcentaje de calificación en la encuesta tipo I. La ausencia de mejoría en este caso puede ser causa de las mismas razones de asistencia y actitud expuestas en la sección 5.2.2.



Gráfica 5.21. Identificación de la Vocal “E” para A04.

5.3.3 Alumno A05, Vocal I

En este caso se debía pronunciar una “I”, y lo que se está percibiendo es una “E” en ambos casos (antes y después del entrenamiento). En la gráfica 5.22 se puede apreciar como hay una disminución en el porcentaje de personas que escucharon una “I” al inicio (25%) y las que escucharon una “E” al final (0%). Quiere decir que en la muestra del final del entrenamiento se pierde por completo la similitud con la vocal objetivo.

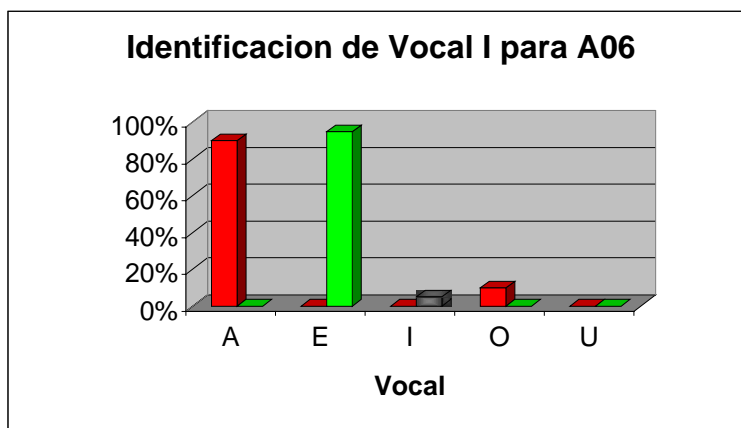


Gráfica 5.22. Identificación de la Vocal “I” para A05.

Nos damos cuenta que esta persona, en esta vocal, también disminuyó su porcentaje de calificación en la encuesta tipo I. La ausencia de mejoría en este caso puede ser causa de las mismas razones de asistencia y actitud expuestas en la sección 5.2.3.

5.3.4 Alumno A06, Vocal I

Este alumno debió pronunciar la “I”, pero de acuerdo a la opinión de los encuestados, en la muestra de inicio se escucha una “A” y en la del final se está escuchando una “E”. Observando la gráfica 5.23 nos damos cuenta que a pesar de que no se obtuvo la vocal objetivo, sí hubo una ligera mejoría de pronunciación. Inicialmente nadie opinó que se escuchara una “I”. En la muestra del final un 5% de los encuestados opinaron que se escuchaba una “I”. Esto significa un pequeño 5% de avance. Además nos damos cuenta que según el triángulo vocálico (Figura 2.2), la “E” (final) está más cerca de la vocal “I” que la “A” (inicial), habiendo también una mejoría en el punto de articulación.

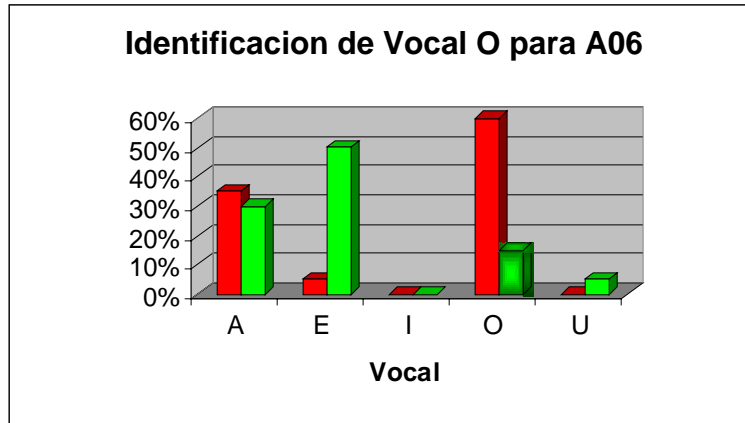


Gráfica 5.23. Identificación de la Vocal “I” para A06.

5.3.5 Alumnos A06, Vocal O

Según apreciamos en la gráfica 5.24, este caso es uno de los peores. Se tuvo una disminución considerable en la pronunciación de la vocal “O”. Si tomamos en cuenta los resultados de las encuestas tipo I, nos damos cuenta que precisamente este caso fue el que bajo más su calificación final. Al parecer esta vocal no pudo ser dominada por esta persona, y quedó comprobado en los dos tipos de experimentos.

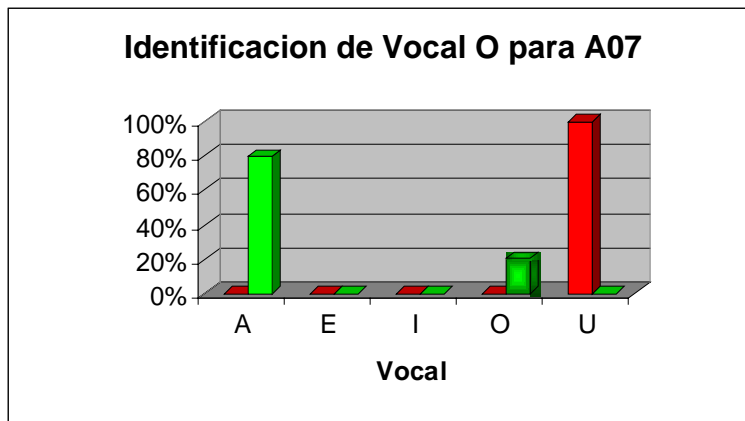
Nos damos cuenta que esta persona, en esta vocal, también disminuyó su porcentaje de calificación en la encuesta tipo I. La desmejora en este caso puede ser causa de las mismas razones de asistencia y actitud expuestas en la sección 5.2.4.



Gráfica 5.24. Identificación de la Vocal "O" para A06.

5.3.6 Alumno A07, Vocal O

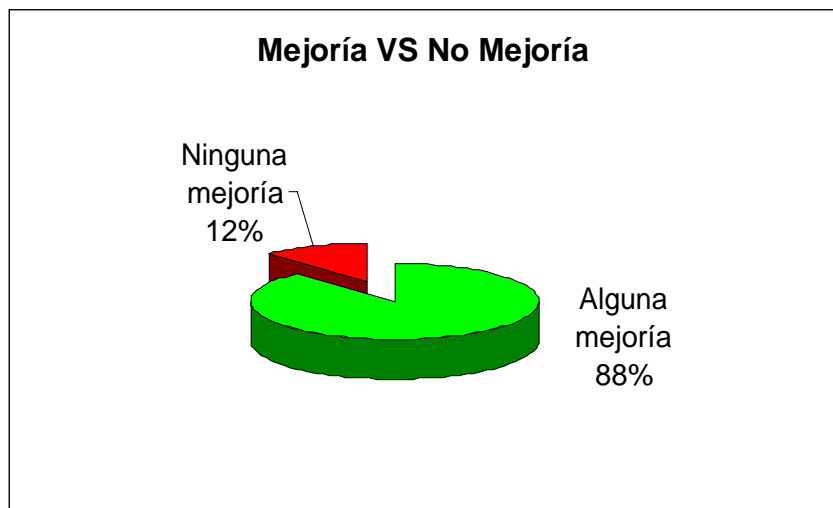
En este último caso, según las encuestas, se debía escuchar una "O", y la mayoría informó que antes del entrenamiento su pronunciación se escuchaba como "U", y después del entrenamiento se escuchaba como "A". En un análisis detallado se puede observar que en la muestra de inicio nadie reportó haber escuchado una "O" (0%) pero en la muestra final un 20% de las personas identificaron la pronunciación como una "O". Esto significa una mejora de pronunciación. En la gráfica 5.25 se muestra el gráfico de este caso.



Gráfica 5.25. Identificación de la Vocal "O" para A07.

Después del análisis de los casos en los cuales no se logró la vocal objetivo, nos dimos cuenta que solo en 3 de ellos no hubo una mejora. Precisamente estos 3 casos corresponden a los únicos 3 casos de los experimentos tipo I en los cuales no hubo mejora de calificación. Esto nos indica que desde dos perspectivas diferentes se puede comprobar que hay casos en los cuales factores externos están evitando una mejora de pronunciación. A pesar de estos únicos 3 casos, nos damos cuenta que en general el producto de software está ayudando a personas con problemas auditivos a mejorar su pronunciación. En la gráfica 5.26 se muestra globalmente el porcentaje de casos en los cuales el software ayudó

a personas con problemas auditivos contrastándolo con el pequeño porcentaje de casos en los cuales, por factores externos, no se mejoró.



Gráfica 5.26. Comparativo de Casos con Mejoría contra Casos sin Mejoría.

CAPITULO 6

Conclusiones y Trabajo Futuro

Poco a poco la tecnología está abarcando más áreas en las cuales ser útil. El desarrollo de herramientas para personas con problemas auditivos no es tan popular y comercial como los sistemas financieros e industriales, pero poco a poco se está tomando conciencia de que la tecnología puede tener su lado humano.

El reconocimiento automático de voz, por su parte, está tomando cada vez más auge. Sus aplicaciones son muy amplias y abarcan sistemas de servicio telefónico, accesos de seguridad, artefactos para minusválidos, comunicación con las computadoras, etcétera [33, 35]. También se está combinando con otras técnicas, como el reconocimiento de imágenes, para obtener información de los labios y así mejorar el reconocimiento [37].

Las personas que no pueden recibir información en forma auditiva no están condenadas a no poderse expresar por medio de la voz. No podemos decir que el lenguaje oral pueda sustituir el lenguaje materno de señas, pero es evidente que en un mundo de hablantes el lenguaje oral es una habilidad que puede ser útil para el desenvolvimiento del discapacitado auditivo. Pero hay que tomar en cuenta que se requiere un proceso distinto de aprendizaje, a través de un canal diferente de comunicación [48].

Este trabajo presentó una interfaz computacional que sustituye la retroalimentación auditiva por retroalimentación visual, ayudando de esta manera a mejorar la pronunciación de vocales a las personas que no pueden escuchar a otros ni escucharse a sí mismos.

Los resultados obtenidos nos muestran que la interfaz les proporciona información y datos que permiten una mejor articulación, y así una mejor pronunciación. En general todos los alumnos participantes en la experimentación lograron mejorar su pronunciación en cierto grado. Esta interfaz computacional, aunada con la dedicación y empeño puesto por los participantes, determinó que tanta mejoría se obtuvo.

Los resultados también nos mostraron que hay vocales en las cuales es más difícil obtener una mejoría. Tal vez se requiera un sistema en tres dimensiones que presente aspecto de la articulación que un gráfico de dos dimensiones no puede mostrar. Una demostración inicial animada de la articulación también podría ser de mucha ayuda.

Este proyecto es solo el principio un sistema más robusto que puede ser desarrollado en los años siguientes utilizando procesamiento y reconocimiento de voz. Por ahora solo se trabajó con vocales del idioma español, en un contexto independiente. Para hacer más eficiente este sistema se puede enseñar las vocales dentro del contexto de una palabra. Para otros trabajos futuros se puede lograr integrar todos los fonemas, posteriormente incluir sílabas, palabras y quizá hasta frases. También se pueden realizar sistemas para personas con problemas auditivos de otros países y lenguajes para dominar.

Un trabajo interesante, que puede partir de este punto, es complementar el sistema con otros atributos de la voz como tono, intensidad, duración, etcétera. Incluir la mayor cantidad de atributos en un trabajo futuro pueda ayudar a mejorar muchísimo más la pronunciación, no solo de vocales, sino de palabras y frases.

Otra forma de complementar la articulación de los fonemas es realizando gráficos tridimensionales y utilizando herramientas físicas como el palatógrafo [23] para detectar el contacto de la lengua con el paladar. Este artefacto podría ser muy útil en fonemas como la “T” y la “E”, en las cuales observamos que se tiene mayor dificultad.

Además de software para personas con problemas auditivos, se pueden realizar aplicaciones para otros campos como aprendizaje de un nuevo idioma, vocalización en cantos, terapia del lenguaje, etcétera.

El procesamiento y reconocimiento de voz son tecnologías que están tomando cada vez más auge. La necesidad de sistemas para ayudar a personas con capacidades diferentes, está latente. Las interfaces que unen las computadoras con el ser humano cada vez son mejores y más efectivas.

Falta mucho por hacer, pero es una realidad que la tecnología está supliendo necesidades de la sociedad que antes eran impensables.

Anexo A

Algoritmo Levinson-Durbin

Para calcular el algoritmo de Levinson-Durbin [32] primero, dada la señal de voz s_n , se requiere calcular los valores de autocorrelación de la señal $R(i)$, que están dados por [51]:

$$R(i) = \sum_{n=0}^N s_n s_{n-i} \quad (\text{A.1})$$

El algoritmo es el siguiente:

$$E_0 = R(0) \quad (\text{A.2})$$

$$k_i = - \left[R(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right] / E_{i-1} \quad (\text{A.2})$$

$$a_i^{(i)} = k_i \quad (\text{A.3})$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)} \quad \text{para } 1 \leq j \leq i-1 \quad (\text{A.4})$$

$$E_i = (1 - k_i^2) E_{i-1} \quad (\text{A.5})$$

Donde,

k_i Es el Coeficiente de Reflexión del tubo i

$a_j^{(i)}$ Es el Coeficientes de Predicción Lineal j cuando se utilizan i tubos

E_i Es el error suponiendo i tubos

con $i = 1, 2, \dots, p$, de donde se obtiene:

$$\alpha_j = a_j^{(p)} \quad i \leq j \leq p \quad (\text{A.6})$$

Anexo B

Formato de archivo de audio WAV

Es el formato nativo de sonido de Windows. Fue creado por Microsoft y se ha convertido en el formato de archivo de audio estándar para computadoras PC con calidad de audio de CD. Es aceptado como formato de audio para intercambio de sonido entre varias plataformas y sistemas operativos.

A diferencia de formatos MPEG y otros formatos comprimidos, los archivos WAV guardan muestras como se van recibiendo y no es requerido un pre-procesamiento o descompresión. Además de guardar los datos (sin comprimir), guarda información valiosa como: si es mono o estéreo, frecuencia de muestreo, etcétera.

Se ha incorporado a la norma MPC2 que exige una tasa de muestreo de 16 bits. Los archivos generados tienen una extensión *.wav [38].

El archivo de audio WAV consta de tres fragmentos (chunks) de información: El segmento RIFF, el cual identifica el archivo como un archivo WAV; el segmento FORMAT el cual identifica parámetros como la frecuencia de muestreo; y el fragmento DATA que contiene los datos (muestras). Cada fragmento contiene los siguientes elementos [49]:

Fragmento RIFF (Longitud total = 12 bytes)

Número de Byte	Descripción
0 - 3	"RIFF" (Caracteres ASCII)
4 - 7	Longitud total del paquete que sigue
8 - 11	"WAVE" (Caracteres ASCII)

Fragmento FORMAT (Longitud total = 24 bytes)

Número de Byte	Descripción
0 - 3	"fmt_" (Caracteres ASCII)
4 - 7	Longitud del segmento FORMAT
8 - 9	Siempre 0x01
10 - 11	Número de canales (Mono = 0x01, Estéreo = 0x02)
12 - 15	Frecuencia de Muestreo
16 - 19	Bytes por segundo

20 – 21	Bytes por muestra (1 = 8 bit mono, 2 = 8 bit estéreo o 16 bit mono, 4 = 16 bit estéreo)
22 – 23	Bits por muestra

Fragmento DATA (Longitud variable)

Número de Byte	Descripción
0 – 3	"data" (Caracteres ASCII)
4 – 7	Longitud de los datos que siguen
8 – fin	Datos (Muestras)

El orden de los bytes predeterminado que se asume para los archivos WAVE es little-endian, en el cual el último byte significativo es guardado primero, es decir los bytes son guardados en orden inverso; este esquema es utilizado principalmente en sistemas Microsoft. Los archivos guardados usando el esquema de ordenamiento de bytes big-endian, en el cual el primer byte significativo es guardado primero (bytes guardados en forma ordenada), son utilizados principalmente en Mac [60], tiene el identificador RIFX en vez de RIFF. Cabe mencionar que al grabar archivos de audio desde un programa realizado en JAVA se utiliza el esquema de ordenamiento de bytes big-endian.

En un archivo WAVE de un solo canal, las muestras son guardadas consecutivamente. Para archivos WAVE estéreo, el canal 0 representa del canal izquierdo, y el canal 1 representa el canal derecho, y son guardados intercalados de acuerdo al tamaño de la muestra.

Cada muestra se representa con un entero. El tamaño de este entero (Integer) es el número menor de bytes requeridos para contener el tamaño de la muestra especificada.

Si se quiere tener muestras de 8-bit estas son guardadas en un solo byte (cada muestra) sin signo, con valores de 0 a 255. Si se desea muestras de 16-bit, estas son guardadas como enteros con signo de complemento 2, con valores de -32768 a 32767 .

Finalmente, en algunos archivos WAV es posible encontrar bytes extras después del fragmento FORMAT y justo antes del fragmento DATA. Estos bytes son conocidos como el *header* o encabezado y son simplemente informativos. El tamaño común de este *header* es de 1024 bytes.

Anexo C

Manual de Usuario de REVISAH

C.1 Introducción

El programa REVISAH es una herramienta de retroalimentación visual de la posición articulatoria del tracto vocal obtenida del procesamiento digital de la señal de voz suponiendo que el tracto vocal se puede modelar como tubos rígidos discretos.

El sistema está diseñado para apoyar a las personas con algún tipo de problema auditivo que deseen obtener retroalimentación visual de su voz y así mejorar la pronunciación de las cinco vocales del idioma español.

La principal ventaja de utilizar este sistema es que proporciona una vista clara de la articulación realizada al pronunciar cierta vocal y se contrasta con un patrón de la posición correcta de los órganos que participan en la pronunciación.

Su funcionamiento es muy sencillo y puede llegar a ser utilizado en forma independiente por las personas que no escuchan (parcial o totalmente). Aún así es recomendable la supervisión de un instructor en las primeras sesiones para fijar los patrones personales (opcional) y mientras la persona se familiariza con el funcionamiento y manejo del sistema. El objetivo de este manual es orientar al usuario en el manejo y empleo del sistema.

C.2 Características Técnicas

El sistema fue desarrollado en JAVA, en una versión gráfica que no necesita un navegador de Internet por el momento. Se cuenta con 17 clases y 1 archivo *.jar que contiene funciones de apoyo (familia GNU getopt) para grabar y reproducir los archivos de voz [46]. Todas las clases y archivos de apoyo como dibujos, gráficas, archivos de sonido, archivos de texto y utilerías están contenidos en un CD. Este CD también cuenta con el JDK de JAVA instalado, permitiendo que se pueda iniciar y utilizar el software sin necesidad de que la computadora tenga instalado JAVA.

Lo único que debe estar grabado en la máquina donde este funcionando el sistema, es una carpeta (que viene en el CD pero debe ser grabada en el disco duro de la computadora), que contiene los archivos de voz y texto que son modificados durante el funcionamiento del sistema. El CD también contiene accesos directos y archivos *.bat que pueden ser copiados al escritorio de la computadora o desde donde se puede iniciar el sistema.

El sistema está diseñado para funcionar correctamente en ambientes WINDOWS, y en un futuro se harán las adecuaciones para que funcione en cualquier plataforma. Si se desea utilizar la opción de reconocimiento es necesario contar con conexión a Internet mientras se esté utilizando el sistema.

Es un sistema muy sencillo de usar y fácil de aprender, pues no contiene comandos complicados o menús sobrecargados. El uso del ratón facilita su uso, pero no es indispensable. Cualquier usuario con conocimientos muy básicos de computación puede hacer uso de él.

C.3 Requerimientos del Sistema

- PC con procesador Pentium o similar.
- Un mínimo de 32 Mb de RAM.
- 300 Kb en disco duro.
- Tarjeta de sonido.
- Micrófono.
- Dispositivo CD-ROM.
- Windows 98, 2000.
- Ratón y bocinas (opcional).
- Conexión a Internet si se desea utilizar la opción de reconocimiento.

C.4 Instalación

La instalación del sistema es muy sencilla. Primero introduzca el CD de REVISAH. Luego presione el botón “Inicio” y elija la opción “Ejecutar...”. Introduzca el siguiente comando:

D:\Instalar.bat

Si desea tener un acceso directo al programa REVISAH puede copiar a su escritorio el acceso directo que viene incluido en el CD. El acceso directo se llama: “REVISAH_1_0”. Si desea puede correr el programa desde el menú “Inicio” opción “Ejecutar...” y teclear el siguiente comando:

D:\REVISAH

Se recomienda revisar que el micrófono este conectado y funcionando con el volumen adecuado, antes de hacer la instalación del software.

C.5 Operación del Sistema

Después que inicie el sistemas REVISAH (ya sea desde el acceso directo o ejecutando el comando del CD), aparecerá la ventana de presentación donde usted podrá ver los datos del programa (Figura C.1).

En esta ventana inicial usted tendrá que seleccionar el fonema que desea practicar, y si previamente guardó un patrón podrá elegirlo. Si es la primera vez que utiliza el sistema se recomienda dejar el patrón “Predeterminado”. Si desea practicar con la vocal elegida presione el botón “Iniciar”. Si desea salir del sistema presione el botón “Salir”.



Figura C.1. Pantalla Inicial del Sistema REVISAH.

También en esta ventana inicial usted tendrá la opción de elegir si desea utilizar reconocimiento o no. Inicialmente está deshabilitada esta opción, pero se puede habilitar dando clic al cuadro blanco que se encuentra antes de la palabra “Reconocimiento”. Es importante que si usted elige esta opción tenga su computadora conectada a Internet.

A continuación aparecerá la pantalla principal del sistema (Figura C.2). Aquí se puede observar inicialmente la identificación del fonema en la parte superior izquierda. Del lado derecho de la pantalla aparece un gráfico general del punto de articulación de la vocal,

así como el símbolo gráfico del fonema y una vista frontal de los labios al pronunciar el fonema elegido. En esta pantalla usted tendrá cuatro botones de interacción. El primer botón (Grabar) es utilizado para abrir la pantalla en la cual usted podrá grabar el fonema a practicar.

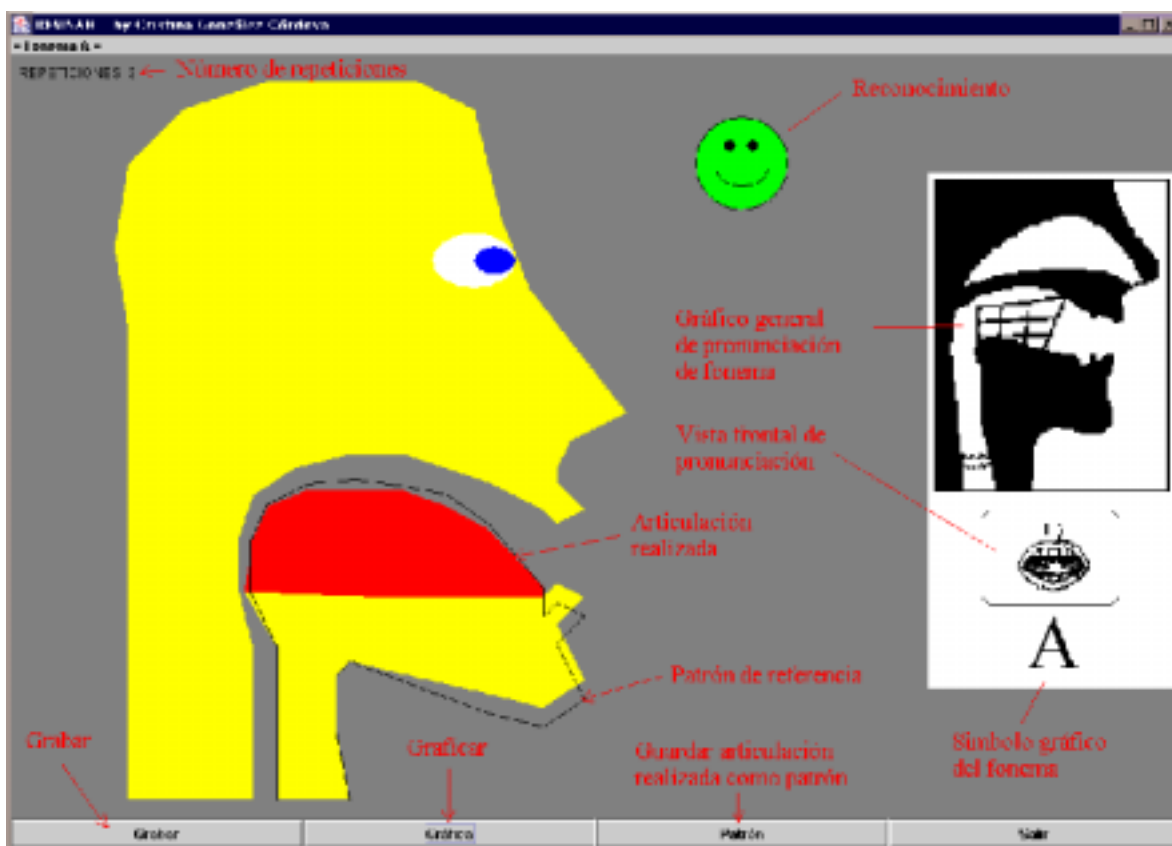


Figura C.2. Pantalla Principal del Sistema REVISAH.

El segundo botón (Gráfica) solo será activado después de que usted grabe el fonema a practicar. Este botón despliega el corte de perfil donde se muestra la articulación que se realizó al pronunciar el fonema y el patrón elegido para ese fonema. La línea negra representa el patrón de referencia. La sección roja representa la lengua y la parte amarilla representa la cara y mandíbula.

Si se eligió la opción de reconocimiento, además de presentar el gráfico del tracto vocal, se presentará una carita en la parte media superior de la ventana. Una carita feliz y de color verde indica que la vocal que se está practicando concuerda con la vocal que se pronunció en el intento. Una carita triste y de color rojo indica que la vocal pronunciada no fue reconocida como la vocal que se está practicando.

Con el tercer botón (Patrón) se tiene la oportunidad de grabar la articulación realizada en un patrón para ser utilizado posteriormente como patrón de referencia. Cuando

presionemos este botón aparecerá un cuadro pequeño en el cual debemos seleccionar en que número de patrón deseamos grabar la articulación realizada y presionar el botón “Grabar Patrón”.

El cuarto botón sirve para salir del fonema activo y regresar a la ventana inicial. También podemos observar en la parte superior izquierda del gráfico un conteo de las veces que se a practicado el fonema.

Finalmente tenemos la ventana donde se realizan las grabaciones del fonema que se está practicando (Figura C.3). En la parte central de la ventana tenemos inicialmente un círculo rojo, que indica que no se está grabando. Para iniciar la grabación presione el botón “Inicia Grabación” y cuando vea el círculo verde pronuncie el fonema. Posteriormente vuelva a presionar el mismo botón para terminar la grabación.

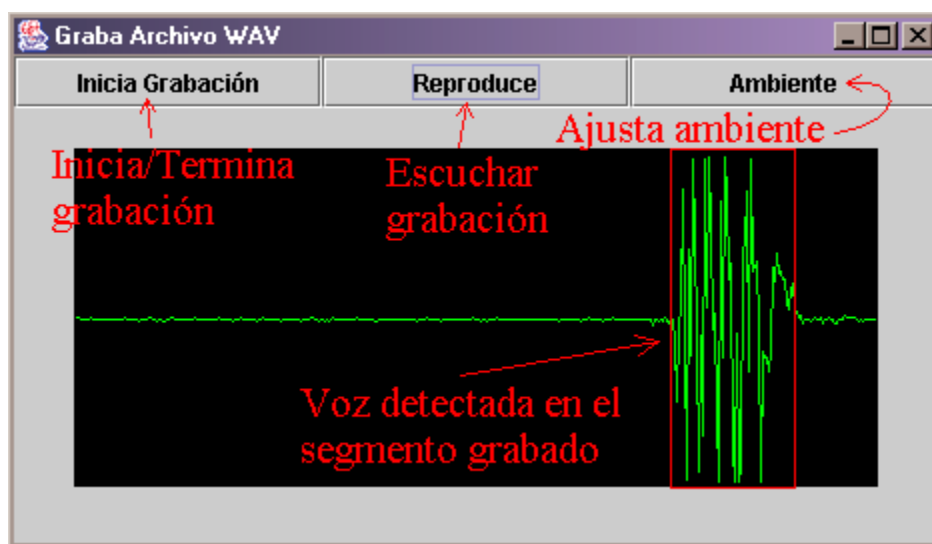


Figura C.3. Pantalla de Grabación del Sistema REVISAH.

Al terminar la grabación se presentará un gráfico de la señal de voz captada por el micrófono. Usted verá un rectángulo rojo en la parte de la grabación que contiene la voz, dejando fuera el silencio inicial y final. Con el botón “Reproduce” usted podrá escuchar el segmento de voz que el cuadro rojo seleccionó.

En caso de que el cuadro rojo no haya seleccionado correctamente el segmento de grabación que contiene el fonema, usted podrá volver a hacer la grabación o ajustar el ambiente. El tercer botón sirve para reajustar el umbral de ruido que existe en el ambiente. Para realizar esta tarea simplemente presione el botón “Ambiente” espere en silencio de dos a tres segundos y vuelva a presionar el mismo botón.

Cuando este satisfecho con el segmento de grabación que se seleccionó regrese a la ventana principal y grafique la articulación que resultó del fonema pronunciado.

Puede realizar varios intentos de cada vocal con el proceso de grabar y graficar. Si desea practicar otro fonema regrese a la ventana inicial, seleccione un nuevo fonema y repita los mismos pasos.

Se recomiendan periodos de práctica de veinte minutos por persona al día, para evitar sobrecarga o cansancio en la persona. Si desea, puede realizar diez repeticiones de cada vocal o bien hacer énfasis en las vocales que se le dificultan más y hacer menos repeticiones de las que ya domina. Se recomienda fijar los patrones personales con la ayuda de un maestro que detecte cuando una vocal se escuche aceptable.

El sistema está diseñado para trabajar correctamente. Si tiene algún problema contacte con su proveedor de sistemas.

Anexo D

Cuestionarios

Se aplicaron dos tipos de cuestionarios, que se muestran a continuación:

D.1 Cuestionario Tipo I

ENCUESTA

El propósito de esta encuesta es conocer su opinión de que tan aceptable se la pronunciación de algunas letras de ciertos nombres. Dentro de las grabaciones que escuchará a continuación se encuentran las de algunas palabras cuyos nombres participarán en un experimento con un software para mejorar la pronunciación. Se está buscando cooperación en su pronunciación de palabras después del período de grabación. Le agradeceremos su cooperación y le pedimos que esta encuesta se realice con honestidad y los datos obtenidos serán analizados en estricta confidencialidad de los datos encuestados.

INSTRUCCIONES: Usted contará con 15 grabaciones de cada vocal (en la computadora presentada). Deberá escuchar cada una de estas grabaciones y dar una calificación del 1 al 5 de acuerdo a que tan entendible e inteligible sea su pronunciación. De acuerdo a su opinión, la calificación se dará según:

- 5 - Excelente
- 4 - Buena
- 3 - Aceptable
- 2 - No buena
- 1 - Mala

Deberá ser en el orden de la encuesta se dará a cada pronunciación. Se podrá seguir una sola calificación a cada pronunciación.

VOCAL A

1. NOMBRE DE IDENTIFICACIÓN 1	5	4	3	2	1
2. NOMBRE DE IDENTIFICACIÓN 2	5	4	3	2	1
3. NOMBRE DE IDENTIFICACIÓN 3	5	4	3	2	1
4. NOMBRE DE IDENTIFICACIÓN 4	5	4	3	2	1
5. NOMBRE DE IDENTIFICACIÓN 5	5	4	3	2	1
6. NOMBRE DE IDENTIFICACIÓN 6	5	4	3	2	1
7. NOMBRE DE IDENTIFICACIÓN 7	5	4	3	2	1
8. NOMBRE DE IDENTIFICACIÓN 8	5	4	3	2	1
9. NOMBRE DE IDENTIFICACIÓN 9	5	4	3	2	1
10. NOMBRE DE IDENTIFICACIÓN 10	5	4	3	2	1
11. NOMBRE DE IDENTIFICACIÓN 11	5	4	3	2	1
12. NOMBRE DE IDENTIFICACIÓN 12	5	4	3	2	1
13. NOMBRE DE IDENTIFICACIÓN 13	5	4	3	2	1
14. NOMBRE DE IDENTIFICACIÓN 14	5	4	3	2	1
15. NOMBRE DE IDENTIFICACIÓN 15	5	4	3	2	1

Figura D.1. Hoja #1 del Cuestionario Tipo I.

D.2 Cuestionario Tipo II

ENCUESTA

El propósito de esta encuesta es clasificar ciertas pronunciaciones dentro de alguna de las 5 vocales del idioma español. Las grabaciones que escuchará a continuación son de algunos estudiantes sordos que participaron en un entrenamiento con un software para mejorar la pronunciación de vocales. Se está buscando comprender si su pronunciación ha mejorado después del periodo de entrenamiento. Le agradecemos su cooperación y le pedimos que esta encuesta es totalmente voluntaria y los datos obtenidos serán analizados en combinación con los de las demás encuestas.

INSTRUCCIONES: Usted escucha un 50 grabaciones de las cinco vocales (comodidad en forma aleatoria (en la computadora proporcionada). Deberá escuchar cada una de estas grabaciones y decir que vocal le parece que fue pronunciada.

Encierre en un círculo la vocal que usted considere que fue pronunciada.

1 - NÚMERO DE PRONUNCIACIÓN 1	A E I O U	9 - NÚMERO DE PRONUNCIACIÓN 9	A E I O U
2 - NÚMERO DE PRONUNCIACIÓN 2	A E I O U	10 - NÚMERO DE PRONUNCIACIÓN 10	A E I O U
3 - NÚMERO DE PRONUNCIACIÓN 3	A E I O U	11 - NÚMERO DE PRONUNCIACIÓN 11	A E I O U
4 - NÚMERO DE PRONUNCIACIÓN 4	A E I O U	12 - NÚMERO DE PRONUNCIACIÓN 12	A E I O U
5 - NÚMERO DE PRONUNCIACIÓN 5	A E I O U	13 - NÚMERO DE PRONUNCIACIÓN 13	A E I O U
6 - NÚMERO DE PRONUNCIACIÓN 6	A E I O U	14 - NÚMERO DE PRONUNCIACIÓN 14	A E I O U
7 - NÚMERO DE PRONUNCIACIÓN 7	A E I O U	15 - NÚMERO DE PRONUNCIACIÓN 15	A E I O U
8 - NÚMERO DE PRONUNCIACIÓN 8	A E I O U	16 - NÚMERO DE PRONUNCIACIÓN 16	A E I O U
		17 - NÚMERO DE PRONUNCIACIÓN 17	A E I O U
		18 - NÚMERO DE PRONUNCIACIÓN 18	A E I O U
		19 - NÚMERO DE PRONUNCIACIÓN 19	A E I O U
		20 - NÚMERO DE PRONUNCIACIÓN 20	A E I O U
		21 - NÚMERO DE PRONUNCIACIÓN 21	A E I O U
		22 - NÚMERO DE PRONUNCIACIÓN 22	A E I O U
		23 - NÚMERO DE PRONUNCIACIÓN 23	A E I O U
		24 - NÚMERO DE PRONUNCIACIÓN 24	A E I O U

Figura D.2. Hoja #1 del Cuestionario Tipo II.

Anexo E

Resultados de Experimentos Tipo II

A continuación se presenta la tabla de resultados de los experimentos tipo II:

Alumno	VOCAL	Antes	Después
A03	A	A	A
A04	A	A	A
A05	A	A	A
A06	A	A	A
A07	A	A	A
A03	E	I	A
A04	E	A	A
A05	E	E	E
A06	E	A	E
A07	E	E	E
A03	I	I	I
A04	I	A	I
A05	I	E	E
A06	I	A	E
A07	I	I	I
A03	O	U	O
A04	O	A	O
A05	O	U	O
A06	O	O	E
A07	O	U	A
A03	U	U	U
A04	U	U	U
A05	U	U	U
A06	U	U	U
A07	U	U	U

Anexo F

Datos de Alumnos Participantes y de Voluntarios

A continuación se presentan los datos de los alumnos participantes en los experimentos y de las personas voluntarias para realizar las pruebas.

Clave	Nombre	Edad	Grado	Sexo	Perdida Audición
A03	Sergio Uriel Ávalos V.	15 años	6to primaria	M	De nacimiento
A04	Lorenzo Antonio Torres E.	13 años	6to primaria	M	De nacimiento
A05	Mario Alberto González R.	15 años	6to primaria	M	De nacimiento
A06	Lucero Bucio R.	14 años	6to primaria	F	1 año
A07	Brenda Judith Beltrán S.	11 años	4to primaria	F	De nacimiento

Tabla F.1. Alumnos Participantes.

CLAVE	NOMBRE	EDAD	SEXO
V01	Adriana Córdova L.	12 años	F
V02	Janeth Ariana Zamora C.	18 años	F
V03	Sergio Alfredo Hernández A.	25 años	M
V04	Alejandra Córdova L.	10 años	F
V05	Carlos Carrasco G.	21 años	M

Tabla F.2. Personas Voluntarias.

Referencias Bibliográficas

- [1] Aguirre, J., *Lenguaje Bimodal*, Universidad Extremadura, Badajoz, España, <http://www.unex.es/interzona/Interzona/Linguistica/Trabajos/28.htm>, 1999.
- [2] Barrón, R., Suárez, S., Moctezuma, C., *Reconocimiento de Comandos Verbales utilizando Cuantización Vectorial y Redes Neuronales*, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, Marzo 1999.
- [3] Barrón, R., Suárez, S., Oropeza, J.L., *Un Sistema para el Desarrollo de Aplicaciones de Voz*, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, Marzo 2000.
- [4] Brea, K., *Adaptación de los Formantes de los Fonemas para un Sintetizador de texto en español a voz*, Tesis de la Maestría en Ciencias con Especialidad en Ingeniería Electrónica, Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Monterrey, Febrero 1995.
- [5] Buxton, W., Foulds, R., Rosen, M., Scadden, L., Shein F., *Human Interface Design and the Handicapped User*, ACM SIGCHI Bulletin, Conference Proceedings on Human Factors in Computing Systems, 1986.
- [6] Carroll, J., *Designing Interaction*, Cambridge University Press, 1991.
- [7] Corredera, T., *Defectos en la Dicción Infantil*, Editorial Kapelusz, Argentina, 1973.
- [8] Cruz, A., *Rescatan a Niños del Silencio Total*, Reforma, México D.F. 2001.
- [9] Deller, J., Proakis, J.G., Hansen, J.H., *Discrete-Time Processing of Speech Signals*, Macmillan Publishing, Estados Unidos, 1993.
- [10] Deslauriers, P., *APL Helps the Deaf to Hear Again*, Proceedings of the International Conference on APL, 1993.
- [11] Edwards, A., Edwards, A., Mynatt., D., *Enabling Technology for Users with Special Needs*, Proceedings of the CHI '94 Conference Companion on Human Factors in Computing Systems, 1994.
- [12] Encyclopædia Británica, *Bonet, Juan Pablo*, Encyclopædia Britannica Online, <http://www.search.eb.com/eb/article?eu=82744>, 2003.
- [13] Encyclopædia Británica, *Deafness*, Encyclopædia Britannica Online, <http://www.search.eb.com/eb/article?eu=30111>, 2003.

- [14] Encyclopædia Británica, *Fuller, Sarah*, Encyclopædia Britannica Online, <http://www.search.eb.com/eb/article?eu=138028>, 2003.
- [15] Espinosa, A., Figueroa, J., *Una Nueva Técnica de Procesamiento y Clasificación de Señales de Voz*, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, 2000.
- [16] Frishberg, N., Corazza, S., Day, L., Wilcox, S., Schulmeister, R., *Sign Language Interfaces*, Conference Proceedings on Human Factors in Computing Systems, 1993.
- [17] García, C., *Revisión de la Tecnología de Síntesis de Voz y Recursos Lingüísticos Existentes para el Idioma Español*, Chihuahua, Instituto Tecnológico de Chihuahua, 2001.
- [18] García, L., Suárez, S., *Uso de Herramientas Visuales para el Diagnóstico y Tratamiento de Voz*, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, Agosto 2000.
- [19] Glinert, E., York, B., *Computers and People with Disabilities*, Communications of the ACM, Volumen 35, Número 5, Mayo 1992.
- [20] Hormazábal M., *Síntesis de Voz para Sistemas de Interacción Hombre-Máquina*, Departamento de Ingeniería Eléctrica, Santiago de Chile, Universidad de Chile, Septiembre 2002.
- [21] IBM Corporation, *Línea de Independencia Speech Viewer II – Guía de Usuario Versión 1.01*, 1993.
- [22] Karchmer, M., Allen, T., *The Functional Assessment of Deaf and Hard of Hearing Students*, American Annals of the Deaf, Volumen 144, Abril 1999, páginas 68-77.
- [23] Keate, E., Javkin, H., Antonanzas-Barroso, N., Zou, R., *A System for Teaching Speech to Profoundly Deaf Children Using Synthesized Acoustic and Articulatory Patterns*, Proceedings of the First Annual ACM Conference on Assistive Technologies, 1994.
- [24] Keller, E., *Fundamentals of Speech and Speech Recognition*, John Wiley & Sons, Suiza, 1994.
- [25] Keller, H., *The Story of my Life*, Doubleday & Company, 1905.
- [26] Kirschning, I., Aguas, N., Ahuactzin, A., *Aplicación de Tecnología de Voz en la Enseñanza del Español*, Grupo de Procesamiento Automático de Voz TLATOA, CENTIA, Puebla, Universidad de las Américas, Agosto 2000.

- [27] Ladner, R., Day, R., Gentry, D., Meyer, K., Rose, S., *A User Interface for Deaf-Blind People*, Department of Computer Science, University of Washington, Estados Unidos, 1987.
- [28] Laurel, B., *The Art of Human Computer Interface Design*, Addison-Wesley, 1993.
- [29] Lenzo, K., CMU Sphinx, Carnegie Mellon University, Pittsburg, Estados Unidos, <http://fife.speech.cs.cmu.edu/sphinx/index.html>, 2002.
- [30] Lieberman, P., Blumstein, S., *Speech Physiology, Speech Perception, and Acoustic Phonetics*, Cambridge University Press, Gran Bretaña, 1988.
- [31] Loomis, J., Poizner, H., Bellugi, U., Blakemore, A., Hollerbach, J., *Computer Graphic Modeling of American Sign Language*, Proceedings of the 10th Annual Conference on Computer Graphics and Interactive Techniques, 1983.
- [32] Llamas, C., Cardeñoso, V., *Reconocimiento Automático del Habla: Técnicas y Aplicación*, Secretariado de Publicaciones e Intercambio Científico Universidad de Valladolid, 1997.
- [33] Menchaca, F.R., *La Voz como Control de Acceso a Edificios*, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, Diciembre 1999.
- [34] Mindess, A., *Reading Between the Signs*, Intercultural Press, Yarmouth, Maine, Estados Unidos, 1999.
- [35] Moctezuma, C., Suárez, S., Rayón, P., Reconocimiento de Palabras Aisladas Utilizando Cuantización Vectorial, Multisección y Redes Neuronales, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, 2000.
- [36] Naughton, P., Schildt, H., *Java 2: The Complete Reference*, McGraw-Hill, Estados Unidos, 1999.
- [37] Nishida, S., *Speech Recognition Enhancement by Lip Information*, ACM SIGCHI Bulletin, Conference Proceedings on Human Factors in Computing Systems, 1986.
- [38] Nistal, G., *Audio o Sonido*, Secretaría del Consejo Superior de Informática, Madrid, España, <http://www.map.es/csi/silice/Mmedia7.html>, 2002.
- [39] Nolzaco, J.A., *REVISAH: Algoritmo de Procesamiento y Reconocimiento de Voz para Apoyar a una Herramienta Visual de Retroalimentación*, Reporte Interno, Departamento de Ciencias Computacionales, ITESM, México, 2001.

- [40] Ohki, M., Sagawa, H., Sakiyama, T., Oohira, E., Ikeda, H., Fujisawa H., *Pattern Recognition and Synthesis for Sign Language Translation System*, Proceedings of the First Annual ACM Conference on Assistive Technologies, 1994.
- [41] Orantes, S., *Patrones de Diseño para Interfaces de Usuario*, Tesis de la Maestría en Ciencias con Especialidad en Computación, Instituto Politécnico Nacional, México D.F., Diciembre 1999.
- [42] O'Shaughnessy, D., *Speech Communication: Human and Machine*, Addison-Wesley, 1990.
- [43] Overholt, M., *The Vowels*, Dartmouth College, Hanover, Estados Unidos, <http://www.dartmouth.edu/artsci/engl/engl18/vowels.html>, 2003.
- [44] Pfauntsch, J., Shub, Ch., *Graphic Animation and Speech Synthesis Applications for the Auditory and Visually Disabled*, University of Colorado, Estados Unidos, 1988.
- [45] Rabiner, L., Biing, J., *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
- [46] Renn, A., *GNU getopt - Java port*, Chicago, Estados Unidos, <http://www.urbanophile.com/arenn/hacking/download.html>, 2002.
- [47] Rodríguez, M.A., Escalada, J.G., Monzón, L., Macarrón, A., *Teoría y Aplicación de la Conversión Texto Voz*, Comunicaciones de Telefónica I+D, Número 4, Diciembre 1991.
- [48] Román, M., *Lectura y Escritura Significativas*, Sistemas Bibliotecarios de Información y Sociedad, Universidad Nacional Autónoma de México, México D.F., 2001.
- [49] Rose, B., *WAV File Format Description*, Communications and Information Technology División, Niagara College of Applied Arts and Technology, Ontario, Canadá, <http://www.technology.niagarac.on.ca/courses/comp630/WavFileFormat.html>, 2003.
- [50] Rose, S., *Inclusion of Students with Hearing Loss in General Education: Fact or Fiction?*, The Teacher Educator, Volumen 37, Número 3, Invierno 2003, páginas 216-229.
- [51] Schroeder, M.R., *Computer Speech: Recognition, Compression, Synthesis*, Springer, Berlin, 1999.
- [52] Sierra, N., Chapa, S.V., *Diseño de Interfaces Visuales*, Centro de Investigación y de Estudios Avanzados del IPN, Instituto Politécnico Nacional, México D.F., 1999.
- [53] Steeneken, H., *The Measurement of Speech Intelligibility*, Gold Line, Connecticut, Estados Unidos, http://www.gold-line.com/pdf/articles/p_measure_TNO.pdf, 2003.

- [54] Suárez, S., *Sistemas de Entrenamiento para Rehabilitación de problemas de Voz y Habla*, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, 1999.
- [55] Suárez, S., Barrón, R., Moctezuma, C., Caballero, F.G., Oropeza, J.L., *Desarrollo de Herramientas para el Procesamiento Digital de Señales: Aplicaciones de Reconocimiento de Voz en Tiempo Real*, Centro de Investigación en Computación, México D.F., Instituto Politécnico Nacional, Enero 2000.
- [56] Toledo, M., *Tecnologías de Voz para el Apoyo a la Adquisición del Lenguaje*, Tesis de la Maestría en Ciencias con Especialidad en Sistemas Computacionales, Universidad de las Américas, Puebla, México, 2002.
- [57] Traugott, M., Lavrakas, P., *Encuestas: Guía para Electores*, Siglo Veintiuno Editores, México D.F., 1997.
- [58] Trochim, W., *The Research Methods Knowledge Base*, Segunda Edición, Atomic Dog Publishing, Estados Unidos, 2001.
- [59] USC University Hospital, *Pérdida Auditiva*, Tenet Healthcare Corporation, Estados Unidos, <http://www.uscuh.com/apps/Library/default.asp?ID=598&viewSpanish=Y>, 2002.
- [60] Verts, W., *An Essay on Endian Order*, University of Massachusetts Amherst, Massachusetts, Estados Unidos, <http://www.cs.umass.edu/~verts/cs32/indian.html>, 1996.
- [61] Zorrilla, S., Torres, M., *Guía para Elaborar la Tesis*, McGraw-Hill, México, 1992.