

SISTEMA DE ANÁLISIS DE SEÑALES DE VOZ DE APLICACIÓN FONOAUDIOLÓGICA Y LINGÜÍSTICA

H.L. Rufiner*, C. Martínez*, E. Herrera**,
A. Martínez***, J.M. Cornejo***, M. Cadena***.

*UNER, **COLMEX, ***UAM-I.
e-mail: jmcc@xanum.uam.mx

RESUMEN:

En este artículo se presenta un sistema de adquisición, procesamiento y análisis de señales de voz con aplicación en rehabilitación de hipoacusias y patologías relacionadas con el aparato fonador. El sistema está compuesto por hardware y software compatible con PC entorno Windows95, y posibilita la adquisición de señales emitidas por un hablante, su almacenamiento y manipulación a nivel frecuencial o temporal, permitiendo analizar aspectos acústicos y fonéticos. El sistema constituye una herramienta simple en su utilización y a la vez potente, y posibilita, además, realizar estudios comparativos de anomalías producidas en dichos trastornos, y el mejoramiento y/o entrenamiento especial de la voz.

PALABRAS CLAVE:

Procesamiento de la Voz, Análisis del habla, Patologías del Habla y la Audición, Lingüística.

1. INTRODUCCIÓN

La voz, una de las primeras señales que utilizó el hombre para su comunicación, ha sido objeto de numerosas investigaciones. A raíz de ello se han desarrollado múltiples herramientas que permiten el estudio de los aspectos distintivos de este tipo de señal.

Algunos de estos parámetros característicos son los provenientes de análisis tales como FFT, LPC, energía, cruce por ceros, extracción de frecuencia glótica (pitch), etc.

Durante el desarrollo de diversos proyectos relacionados con el análisis y Reconocimiento del Habla surgió la necesidad de contar con una herramienta para el procesamiento digital de las señales de voz que permitiera adquirir y manipular las mismas, registrando de un modo objetivo los eventos acústico-fonéticos, posibilitando el estudio sistemático de las características de la voz y el habla, y facilitando la comprensión y el análisis de los patrones de emisión fonética. Las características de

esta herramienta, por peso propio, motivaron la implementación de un sistema flexible para el análisis de la señal de voz. Dicho sistema fue desarrollado inicialmente en entorno DOS y luego fue migrado a entorno Windows para terminar con una serie de limitantes técnicas.

Este artículo se organizará de la siguiente forma, a continuación se presenta una breve reseña del trabajo realizado en entorno D.O.S. y se explica la migración y continuación de la implementación en entorno Windows95. En la sección siguiente se presentan los distintos tipos de análisis que pueden realizarse con el programa y finalmente se muestran ejemplos de aplicación en el diagnóstico de patologías de la fonación.

2. METODOLOGÍA

Inicialmente se procedió al desarrollo de la herramienta de análisis de voz en entorno D.O.S., realizando la implementación de rutinas de adquisición de señales de voz, procesamiento y etiquetado.

Al respecto, hemos detectado que la implementación de las rutinas en dicho entorno es altamente dependiente de los controladores que el fabricante desarrolla para el modelo particular de placa de sonido. Debido a los múltiples inconvenientes derivados de estos cambios, de los que no se tiene certeza cuando finalizarán, y ante la dificultad para disponer de información técnica que permita realizar los cambios necesarios al software, se optó por la búsqueda de una plataforma que permitiese trabajar con mayor estabilidad y donde los estándares de desarrollo de los controladores fuesen conocidos para poder hacer uso de ellos.

Por estas razones se decidió migrar al entorno Windows. Además, este entorno brinda una interface

excelente con el usuario donde el método de manejo y la presentación es consistente entre las múltiples aplicaciones.

Para la nueva implementación se realizó un análisis de los diversos entornos de desarrollo existentes para la plataforma Windows, seleccionando finalmente Delphi 3. Este ofrece un muy potente compilador, posee la librería de clases Object Pascal de la que se dispone de abundante información técnica, y brinda múltiples recursos para el desarrollo de las interfaces gráficas con el usuario.

Algunas características particulares del programa hacen necesario que se acceda a bajo nivel en el entorno Windows. Para ello se utilizaron las extensiones que posee Windows para multimedia, las que proveen los siguientes servicios:

La Media Control Interface (MCI) para controlar los dispositivos de medios.

Interfaces basadas en cadenas de caracteres y en mensajes para comunicarse con los drivers de dispositivo MCI.

Los dispositivos MCI, que poseen capacidad para reproducir, grabar audio en formatos wave, reproducir archivos MIDI (Musical Instruments Digital Interface) y reproducir discos compactos y audio desde lectoras de CD-ROM.

Soporte para funciones API de bajo nivel para servicios relacionados con multimedia.

Soporte de bajo nivel para la reproducción y grabación de audio de formato wave y dispositivos de audio MIDI.

Soporte de bajo nivel para control preciso de tiempo.

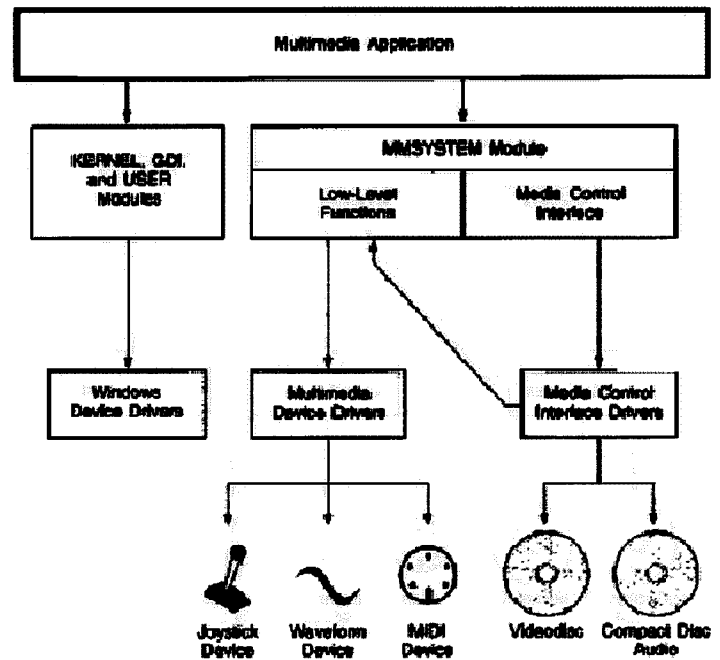
Aunque los servicios de multimedia se incorporan a partir de múltiples archivos, la arquitectura general se puede ver como compuesta de algunos pocos módulos de software, a saber:

La biblioteca MMSYSTEM, que provee los servicios MCI y soporte para las funciones de multimedia de bajo nivel.

Controladores de los dispositivos multimedia, que proveen la comunicación entre las funciones de bajo nivel de la MMSYSTEM y los dispositivos de multimedia tales como reproducción de archivos de sonido, MIDI, y otros.

Manejadores para el control de alto nivel provistos por la MMSYSTEM para acceso a los dispositivos.

La siguiente figura ilustra las múltiples relaciones que existen entre los módulos Windows que proveen servicios de multimedia:



La MCI provee dos interfaces de programación: la interfase basada en cadenas de comandos y la interfase basada en mensajes.

La interfase basada en cadena de comandos provee una serie de comandos que pueden ser utilizados desde cualquier lenguaje de programación de alto nivel o entornos de autor como pueden ser ToolBox.

La interfase basada en mensajes usa la filosofía de envío de mensajes de Windows. En nuestra aplicación, a fin de realizar el procesamiento y análisis de los datos es necesario tener acceso al vector de datos. Las extensiones basadas en cadenas de comandos no permite este acceso por lo que se debió realizar la implementación de las rutinas utilizando la interfase basada en mensajes.

3. Descripción del Sistema

El sistema provee de las siguientes herramientas:

Detección automática de la placa de sonido: Al ejecutarse el programa se realiza una búsqueda hardware de dispositivos de entrada y salida de audio. Este proceso se realiza a fin de conocer las características y capacidades del hardware conectado, muy variable entre distintos equipos.

Adquisición/Reproducción de audio: Se terminó el desarrollo de la unit 'SonidoMM.pas', que permite la grabación de señales de audio y la reproducción de las mismas. Los formatos disponibles incluyen señales de 8 y 16 bits de resolución, con frecuencias de muestreo desde 0 a 44 KHz. Al aprovechar el manejo de memoria plana de 32 bits de Windows95, la longitud

de memoria plana de 32 bits de Windows95, la longitud temporal de las señales que se pueden manejar no tienen la limitación de tamaño máximo 3 seg. que se halla presente en la versión D.O.S. Se dispone de la selección de segmentos de las señales (mediante el mouse), de manera que el usuario tiene la libertad de marcar un tramo de cualquier longitud y adquirir/reproducir únicamente dicha sección.

Lectura de archivos: Se distinguen dos tipos de archivos. Los archivos del primer tipo son señales de voz (formato WAVE), para los cuales la unit 'SonidoMM.pas' implementa la lectura de los mismos, en cualquier formato de los anteriormente explicitados, y con las validaciones y manejo de errores respectivos. El segundo tipo son archivos ASCII, formato que permite leer señales de voz (guardados como datos en un archivo de texto) y señales de análisis tales como la energía, cruces por cero y autocorrelación.

Grabación de archivos: El programa cuenta con la opción de guardar las señales de voz adquiridas (con formato WAVE) y las señales resultantes de los análisis realizados sobre las mencionadas (con formato ASCII). Se completa, de esta manera, el conjunto de operaciones básicas de manejo de señales (adquisición, lectura, reproducción y grabación).

Importación/Exportación de señales: Estas opciones permiten la conversión y grabación de señales de voz en formato ASCII a archivos de onda (WAVE), y viceversa.

Análisis de las señales: Se completaron las rutinas de aplicación de análisis sobre las señales de voz, siendo actualmente funcionales las siguientes: FFT (espectro de Fourier, mostrando las señales de magnitud y fase, o parte real y parte imaginaria), espectrograma, energía, cruces por cero, autocorrelación y ventana de Hamming.

Procesamiento de señales: Se dispone de filtros ARMA (Butterworth de 3er orden): pasa-bajos, pasa-banda, pasa-altos y rechaza-banda; análisis del pitch y se continúa trabajando sobre algoritmos de búsqueda de formantes.

Análisis en tiempo real: Se implementarán los siguientes análisis: sonograma, espectro, pitch y espectrograma (disponibles en la versión D.O.S.)

Sistema de ayuda: Archivos de ayuda hipertexto acerca del manejo del sistema, así también como un tutorial explicativo del manejo del programa. Se continuará

la elaboración del manual del usuario y la guía de referencia rápida.

Edición: Estas herramientas hacen posible, mediante la selección de tramos de señal, el copiado, cortado, borrado e inserción de dichos tramos.

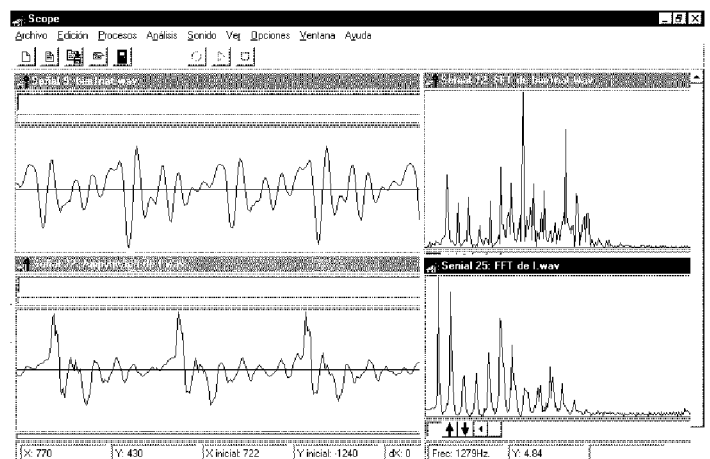
Etiquetado: Las etiquetas se aplican a determinados tramos de la señal, de manera de utilizarlas como referencias acerca del contenido de dichos tramos. De esta manera se logra caracterizar la señal en distintos niveles jerárquicos y obtener patrones etiquetados de la misma.

4. RESULTADOS

El sistema permite obtener diversas señales para el análisis de voces, tales como las mostradas en la figura siguiente: sonograma, energía, espectrograma, AMDF, autocorrelación, espectro de Fourier y otras tales como cruces por cero y pitch. Una salida en pantalla del sistema es la siguiente:

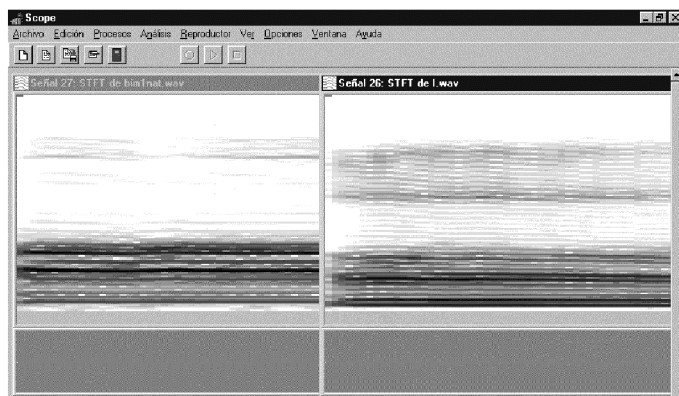
Esta visualización de señales puede servir, por ejemplo, para la comparación de resultados de análisis entre distintas señales de voz. Un ejemplo está dado por el análisis de señales de voces patológicas. El caso presentado aquí muestra una señal de voz bicíclica, también denominada diplofonía o fonación bifurcada, de un hombre diciendo "/a/" en inglés [15]. Esta señal se compara con una "/a/" del inglés pronunciada por un hombre con fonación normal.

En el gráfico siguiente se muestra la señal patológica en la parte superior, junto a su espectro de Fourier, a la derecha; mientras que en la parte inferior se muestra la señal normal (con su espectro de Fourier a la derecha).



En las señales de sonograma se puede ver el período doble que constituye la señal de voz patológica,

contrastada con la señal normal. En los espectros de Fourier se evidencia la diferencia entre ambas señales, lo que se puede también constatar en el espectrograma.



5. DISCUSIÓN

Esta herramienta suscitó el interés de diversos profesionales que se relacionan con el área de la voz, como fonoaudiólogos, terapeutas de la voz (hablada o cantada), otorrinolaringólogos (que trabajan en desordenes de la producción de voz y la audición), y lingüistas.

En el área de la Lingüística su importancia radica en que en el estudio del lenguaje (en sujetos normales) requiere la constatación acústica para apuntalar, verificar o reformular las hipótesis teóricas que esta ciencia hace respecto del lenguaje. Un analizador de voz podría permitir, por ejemplo, el estudio de la intonación en español o en cualquier lengua con el fin de, por citar una aplicación, poder establecer una posible correlación entre cambios suprasegmentales y clase social. También resulta pertinente en estudios de corte vocálicos, o bien en la determinación del timbre de las vocales atonas del español.

6. CONCLUSIONES

Se presentó el desarrollo de un sistema para que cumpla con las necesidades de los campos anteriormente citados, constituyendo una potente herramienta de análisis y diagnóstico en estas áreas, con énfasis en conseguir una interfase de manejo sencillo, pero sin sacrificar la potencia y detalle de los diferentes análisis.

7. REFERENCIAS

1. T.P. Barnwell y S. R. Quackenbush, «An analysis of objectively computable measures for speech quality testing», Proc. ICASSP 1982, Paris, Francia.
2. G. J. Borden y K. S. Harris, «Speech Science Primer», Williams & Wilkins, Baltimore, Londres.
3. A. M. Borzone, «Manual de fonética acústica», Librería Hachette, Argentina, 1980.
4. K. S. Fu, «Syntactic Methods in Pattern Recognition», Academic Press (1974)
5. M. Guirao, A. M. Borzone, «Identification of Argentine Spanish Vowels», Journal of Psycholinguistic Research, Vol 4 No 1, 1975.
6. J. Kelly, Auditory System en E. Kandel, J. Schwartz; Principles of Neural Science. (Elsevier, 1985)
7. L. Rabiner, B. Gold, «Theory and Application of Digital Signal Processing», (Prentice-Hall, 1975)
8. O. Ramos, Análisis de espectros implementado con un microprocesador. Rev. Tel. Electrónica, Junio 1991.
9. L. Rabiner, R. Schafer, «Digital Processing of Speech Signals», (Prentice-Hall, 1987)
10. L. Rocha, «Procesamiento de voz», I Escola Brasileiro-Argentina de Informática, (Kapelusz, 1987)
11. R. Schafer, L. Rabiner, «Digital Representations of Speech Signals» IEEE Proc. Vol 63 No 4 (1975)
12. P. Seligman, Speech-Processing Strategies and their Implementation, 1985.
13. J. Simon, El Reconocimiento de Formas Mediante Algoritmos. (Masson, 1987).
14. A. Waibel, T. Hanazawa, G. Hinton, et al.; Phoneme Recognition Using Time-Delay Neural Networks. IEEE Trans. ASSP Vol. 37, No 3 (1989).
15. A. Alwan, P. Bangayan, C. Long, "Time and Frequency Synthesis Parameters of Severely Pathological Voice Qualities", Dept of Electrical Engineering, UCLA, Proc. of the ICASSP 96.